A compilation of global bio-optical *in situ* data for ocean-colour satellite applications – version ~~three~~four

Salem Ibrahim Salem[1,2], André Valente[3,4], Shubha Sathyendranath[1,5], Andrei Chuprin[1], Théo Renard[6], Jaime Pitarch[7], Ruth Airs[1], Moreno Ana[8], Maria-Manuel Angelico[8], David Antoine[9,10], Robert Arnone[11], Florinda Artuso[12], Paulo O. Oliveira[8], William Balch[13], Kathryn Barker[14], Ray Barlow[15], Simon Bélanger[16], Marco Bellacicco[7], Jean-François Berthon[17], Şükrü Beşiktepe[18], Francesco Bignami[7], Yngve Borsheim[19], Astrid Bracher[20,21], Federica Braga[7], Vittorio Brando[7], Robert J. W. Brewin[22], Vanda Brotas[4,1], Elisabetta Canuti[17], Francisco P. Chavez[23], Andrés Cianca[24], Hervé Claustre[10], Lesley Clementson[14], Simone Colella[7], Raul Costa[25,26], Cristina Nunes[8], Richard Crout[27], Emmanuel Devred[29], Annalisa Di Cicco[7], Federico Falcini[7], Ivan Farace[7], Afonso Ferreira[3,28], Scott Freeman[30], Robert Frouin[31], Carlos García-Soto[32,33], Isabella Giardina[12], Stuart W. Gibb[34], Arthur Gleason[35], Ralf Goericke[31], Richard Gould[27], Michael Grant[1,36], Nathalie Guillocheau[37], Hiroto Higa[38], Stanford B. Hooker[30], Chuanmin Hu[39], Joji Ishizaka[40], Thomas Jackson[36,1], Mati Kahru[31], Milton Kampel[41], Bokyung Kim[42], Holger Klein[43], Piotr Kowalczuk[44], Susanne Kratzer[45], Raphael Kudela[46], Antonia Lai[12], Rosario Lavezza[19], Jesus Ledesma[47], Huiwon Lee[48], Steven Lohrenz[49], Hubert Loisel[50], Vivian Lutz[51], Antonio Mannino[30], Francesca Margiotta[52], Victor Martinez-Vicente[1], Salvatore Marullo[7], Luca Massi[53], Patricia Matrai[13], Atsushi Matsuoka[54], David McKee[55], Chiara Melillo[53], Carlos B. Mendes[28,25], Brian G. Mitchell[31], Tiffany Moisan[56,†], Enrique Montes[57,58], Frank Muller-Karger[39], Aimee Neeley[30], Michael Novak[30], Leonie O'Dowd[59], Michael Ondrusek[60], Emanuele Organelli[7], Heeyoon Park[61], Trevor Platt[1,†], Francesca Polonelli[53], Alex Poulton[62], Mariella Ragni[52], Michel Repecaud[63], Maurizio Ribera d'Alcalà[52], Tamara Rodriguez-Ramos[64], Rüdiger Röttgers[65], Manuel Ruiz-Villarreal[64], Michela Sammartino[66], Rosalia Santoleri[7], Gian Marco Scarpa[7], Thomas Schroeder[14], Valeria Segura[51], Timothy Smyth[1], Denise Smythe-Wright[67], Heidi M. Sosik[68], Joanna Stoń-Egiert[44], Malcolm Taberner[36], Jing Tan[31], Rob Thomas[59], Crystal Thomas[30], Gavin H. Tilstone[1], Andreia Tracana[3], Michael Twardowski[69], Vincenzo Vellucci[10], Gianluca Volpe[7], Kenneth Voss[35], Jeremy Werdell[30], Marcel Wernand[70,†], Bozena Wojtasiewicz[71], Simon Wright[72], Sunmi Yang[48], Giuseppe Zibordi[17], Maria Laura Zoffoli[73], Steve Groom[1]

[1] Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, United Kingdom.
[2] Faculty of Engineering, Alexandria University, Al Horreya Avenue, Alexandria 11432, Egypt.
[3] Atlantic International Research Centre (AIR Centre), Caminho do Lameirinho, Angra do Heroísmo 9700-117, Portugal.
[4] MARE - Marine and Environmental Sciences Centre/ARNET - Aquatic Research Network, Faculdade de Ciências, Universidade de Lisboa, Campo Grande 016, 1749-016 Lisboa, Portugal.
[5] National Centre for Earth Observation, Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3DH, United Kingdom.
[6] Faculty of Sciences, University of Liège, Allée du Six Août 11, Liège 4000, Belgium.
[7] Institute of Marine Sciences, National Research Council of Italy (CNR-ISMAR), Via del Fosso del Cavaliere 100, Rome 00133, Italy.
[8] Instituto Português do Mar e da Atmosfera (IPMA), Rua C do Aeroporto, Lisboa 1749-077, Portugal.
[9] Remote Sensing and Satellite Research Group, School of Earth and Planetary Sciences, Curtin University, Kent Street, Bentley, WA 6102, Australia.
[10] Sorbonne Université, CNRS, Institut de la Mer de Villefranche, 181 Chemin du Lazaret, Villefranche-sur-Mer 06230, France.
[11] School of Ocean Science and Engineering, The University of Southern Mississippi, 1020 Balch Boulevard, Stennis Space Center, MS 39529, United States of America.
[12] Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Via Enrico Fermi 45, Frascati (Rome) 00044, Italy.
[13] Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, ME 04544, United States of America.
[14] The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia.
[15] Bayworld Centre for Research and Education, Beach Road, Gqeberha 6013, South Africa.
[16] Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC G5L 3A1, Canada.

[17] European Commission, Joint Research Centre, Via Enrico Fermi 2749, Ispra 21027, Italy.
[18] Institute of Marine Sciences and Technology, Dokuz Eylül University, Baku Bulvarı No. 100, İzmir 35340, Türkiye.
[19] Institute of Marine Research, Nordnesgaten 50, Bergen 5005, Norway.
[20] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, Bremerhaven 27570, Germany.
[21] Institute of Environmental Physics, University of Bremen, Otto-Hahn-Allee 1, Bremen 28359, Germany.
[22] Centre for Geography and Environmental Science, University of Exeter, Treliever Road, Penryn TR10 9FE, United Kingdom.
[23] Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039, United States of America.
[24] Oceanic Platform of the Canary Islands (PLOCAN), Carretera de Taliarte s/n, Telde 35214, Spain.
[25] Federal University of Rio Grande (FURG), Avenida Itália, km 8, Rio Grande 96203-900, Brazil.
[26] National Oceanography Centre, European Way, Southampton SO14 3ZH, United Kingdom.
[27] U.S. Naval Research Laboratory, 1005 Balch Boulevard, Stennis Space Center, MS 39529, United States of America.
[28] Departamento de Biologia, Faculdade de Ciências, Universidade de Lisboa, Campo Grande 016, 1749-016 Lisboa, Portugal.
[29] Fisheries and Oceans Canada, Bedford Institute of Oceanography, 1 Challenger Drive, Dartmouth, NS B2Y 4A2, Canada.
[30] National Aeronautics and Space Administration, Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, MD 20771, United States of America.
[31] Scripps Institution of Oceanography, University of California San Diego, 8622 Kennel Way, La Jolla, CA 92037, United States of America.
[32] Spanish Institute of Oceanography, Spanish National Research Council (IEO-CSIC), Corazón de María 8, Madrid 28002, Spain.
[33] Plentzia Marine Station, University of the Basque Country (PIE-UPV/EHU), Areatza Pasealekua s/n, Plentzia 48620, Spain.
[34] Environmental Research Institute, University of the Highlands and Islands, Castle Street, Thurso KW14 7JD, United Kingdom.
[35] Department of Physics, University of Miami, 1320 Campo Sano Avenue, Coral Gables, FL 33146, United States of America.
[36] European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), Eumetsat-Allee 1, Darmstadt 64295, Germany.
[37] Earth Research Institute, University of California Santa Barbara, 6832 Ellison Hall, Santa Barbara, CA 93106-3060, United States of America.
[38] Faculty of Urban Innovation, Yokohama National University, 79-1 Tokiwadai, Yokohama 240-8501, Japan.
[39] College of Marine Science, University of South Florida, 140 Seventh Avenue South, St. Petersburg, FL 33701, United States of America.
[40] Institute for Space–Earth Environmental Research, Nagoya University, Furo-cho, Nagoya 464-8601, Japan.
[41] National Institute for Space Research (INPE), Avenida dos Astronautas 1758, São José dos Campos 12227-010, Brazil.
[42] Division of Exhibition and Education, Cheongju National Ocean Science Museum, 52 Millenieom 1-ro, Cheongju-si 28353, Republic of Korea.
[43] Federal Maritime and Hydrographic Agency, Bernhard-Nocht-Straße 78, Hamburg 20359, Germany.
[44] Institute of Oceanology, Polish Academy of Sciences, Powstańców Warszawy 55, Sopot 81-712, Poland.
[45] Department of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91 Stockholm, Sweden.
[46] University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, United States of America.
[47] Instituto del Mar del Perú, Esquina Gamarra y General Valle s/n, Callao 07021, Peru.
[48] Marine Convergence Department, Underwater Survey Technology 21 Co., Ltd., 12 Gaetbeol-ro, Incheon 21999, Republic of Korea.
[49] University of Massachusetts Dartmouth, 285 Old Westport Road, Dartmouth, MA 02747, United States of America.
[50] Université du Littoral Côte d'Opale, Université de Lille, CNRS, 32 Avenue Foch, Wimereux 62930, France.
[51] Instituto Nacional de Investigación y Desarrollo Pesquero (INIDEP), Paseo Victoria Ocampo 1, Mar del Plata B7602HSA, Argentina.
[52] Stazione Zoologica Anton Dohrn, Villa Comunale, Naples 80121, Italy.
[53] University of Florence, Via Micheli 1, Florence 50121, Italy.
[54] Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, 8 College Road, Durham, NH 03824, United States of America.
[55] University of Strathclyde, 16 Richmond Street, Glasgow G1 1XQ, United Kingdom.
[56] National Aeronautics and Space Administration, Goddard Space Flight Center, Wallops Flight Facility, Wallops Island, VA 23337, United States of America.
[57] Cooperative Institute for Marine and Atmospheric Studies, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, United States of America.
[58] National Oceanic and Atmospheric Administration, Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, FL 33149, United States of America.
[59] Marine Institute, Rinville, Oranmore H91 R673, Ireland.

[60] National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, Center for Satellite Applications and Research, 5830 University Research Court, College Park, MD 20740, United States of America.
[61] National Ocean Satellite Center, Korea Hydrographic and Oceanographic Agency, 351 Haeyang-ro, Busan 49111, Republic of Korea.
[62] Heriot-Watt University, Boundary Road, Edinburgh EH14 4AS, United Kingdom.
[63] French Research Institute for Exploitation of the Sea (IFREMER), Centre de Brest, 1625 Route de Sainte-Anne, Plouzané 29280, France.
[64] Spanish Institute of Oceanography, Spanish National Research Council (IEO-CSIC), Paseo Marítimo Alcalde Francisco Vázquez 10, A Coruña 15001, Spain.
[65] Helmholtz-Zentrum Hereon, Max-Planck-Straße 1, Geesthacht 21502, Germany.
[66] Institute of Marine Sciences, National Research Council of Italy (CNR-ISMAR), Calata Porta di Massa, Naples 80133, Italy.
[67] National Oceanography Centre, European Way, Southampton SO14 3ZH, United Kingdom.
[68] Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole, MA 02543-1050, United States of America.
[69] Harbor Branch Oceanographic Institute, Florida Atlantic University, 5600 US Highway 1 North, Fort Pierce, FL 34946, United States of America.
[70] Royal Netherlands Institute for Sea Research (NIOZ), Landsdiep 4, 't Horntje 1797 SZ, The Netherlands.
[71] Australian Institute of Marine Science, Indian Ocean Marine Research Centre, Crawley, WA 6009, Australia.
[72] Institute for Marine and Antarctic Studies, University of Tasmania, 20 Castray Esplanade, Battery Point, TAS 7004, Australia.
[73] Institute of Marine Sciences, National Research Council of Italy (CNR-ISMAR), Area Science Park Basovizza - Edificio Q2 - Strada Statale 14, km 163.5, Trieste 34149, Italy.
† Deceased.

Correspondence to: Salem I. Salem (sis@pml.ac.uk)~~A. Valente (adovalente@fc.ul.pt)~~

Abstract. A global *in situ* ~~data set~~dataset for validation of ~~ocean-colour~~satellite products from the ESA Ocean Colour Climate Change Initiative (OC-CCI) is presented. This version of the compilation, with data starting in ~~1997~~1996, ~~now~~ extends to ~~2021~~2025, which is important for the validation of recent algorithms and satellite products. The ~~data set~~dataset comprises *in situ* observations of the following variables: spectral remote-sensing reflectance (R~~r~~rs), ~~concentration of~~ chlorophyll-a concentration ("chla"; HPLC and fluorometric), spectral inherent optical properties (IOPs: algal pigment absorption "aph", detritus plus gelbstoff absorption "adg", and particle backscattering "bbp"), spectral diffuse attenuation coefficient (kd) and total suspended matter (tsm). Data were obtained from multiple ~~-project~~ archives acquired via open internet services, or from individual projects, acquired directly from data providers. Compared to the previous version (Valente et al., 2022)~~,~~ this release incorporates updates from existing sources and introduces data from 13 additional sources. Notably~~For instance~~, AERONET-OC coverage has been significantly expanded (39 sites compared to 11 in v2022)~~,~~ and MOBY Platinum data have been included from 2024. Methodologies were implemented for homogenisation, quality control and merging of all data. Minimal changes were made ~~on~~ to the original data, other than conversion to a

standard format, elimination of some points after quality control and averaging of observations that were close in time and space. The harmonisation procedures also include the application of a bidirectional reflectance distribution function (BRDF) correction to "r~~Rr~~rs". The result is a merged table available in text format. Overall, ~~the size of~~ the data~~-~~set increased ~~grew~~ by ~115 %, ~~with 148,432~~reaching 319,183 rows, with each row representing a unique station in space and time (cf 148,432 ~~136,250~~ rows in Valente et al., 2022~~previous version; Valente et al., 20~~22~~19~~).~~-~~ R~~C~~ecords of chlaovera increased by ~31 % (82,543 to 107,922);~~). The number of~~ "rrs" observations nearly doubled +94 % (68,641 → 133,325);~~.~~ ~~Inherent optical properties~~IOPs showed substantial expansion, with aph increasing by ~150 % (4,265 to 10,655) and adg by 78 % (1,654 to 2,948). The most pronounced ~~increase~~change was ~~observed~~ for bbp, which ~~expanded~~increased by ~10,258 % (792 to 82,033), while moderate ~~growth~~increases were observed for kd (+30 %; 2,454 to 3,197) and ~~tsmTSM~~"tsm" (+51 %; 1,546 to 2,334).~~Observations of remote-sensing reflectance increased to 68,641 (cf 59,781 in previous version; Valente et al., 2019). There was also a near tenfold increase in chlorophyll data since 2016.~~ Metadata of each *in situ* measurement (original source, cruise or experiment, principal investigator) are included in the final table. By making the metadata available, provenance is better documented, and it is also possible to analyse each set of data separately. The compiled data are available at ~~https://doi.org/10.1594/PANGAEA.941318~~ (~~Valente~~ Salem et al., ~~2022~~2026).

1 Introduction

Data collected by satellite ocean~~-~~colour sensors provide synoptic observations on ~~ocean~~ productivity and ~~the~~ variability of the marine environment, at high spatial and temporal resolutions. Ocean colour data, recognized as an Essential Climate Variable (ECV)~~s~~ by the Global Climate Observation System (GCOS, 2022)~~,~~ are invaluable to address key issues, such as the detection of marine ecosystem changes over short and long scales, ~~modifications due to climate change,~~ the study of the global carbon cycle and the assessment of coastal water quality degradation~~s~~ (IOCCG, 2008; McClain, 2009). A main goal of the European Space Agency ~~ESA~~ Ocean Colour Climate Change Initiative (ESA OC-CCI) ~~was~~ is to generate a suite of ocean ~~-~~colour products for use in climate studies (Sathyendranath et al., 2019). For this purpose, the existing major data streams for ocean colour ~~were~~ are blended into a coherent ocean ~~-~~colour data record. Currently, data from ~~five~~ six ocean ~~-~~colour sensors are ~~being~~ merged: the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), ~~of~~ a NASA-funded mission operated by Orbital Sciences Corporation, the Medium Resolution Imaging Spectrometer (MERIS) of ESA, the MODerate resolution Imaging Spectro-radiometer (MODIS) onboard the Aqua platform of NASA, the Visible Infrared Imaging Radiometer Suite (VIIRS) of NASA and NOAA, and the Ocean and Land Colour Instrument (OLCI) ~~of~~ onboard ESA Sentinel-3A and Sentinel-3B. For the validation of the ESA OC-CCI satellite products, a compilation of *in situ* bio-optical data was produced, which is presented in ~~.~~ t~~T~~his paper ~~presents that compilation~~.

~~There~~ In this regard, there are several sets of *in situ* bio-optical data, worldwide, suitable for validation of ocean ~~-~~colour satellite ~~data~~products. While some are managed by the data producers, others are in international repositories with contributions from multiple scientists. Many have rigid quality controls and are built specifically for ocean colour validation. The use of ~~only any~~just one of these ~~data set~~datasets would limit the ~~amount~~ number and distribution of data in validation exercises. It is, therefore, vital to merge all these *in situ* ~~data set~~datasets to maximize the number of matchups available for validation, with wider distribution in time and space, and, consequently, to reduce uncertainties in the validation exercise. However, merging several ~~data set~~datasets together can be a challenging ~~complicated~~ task. ~~First~~First, it is necessary to acquire and harmonize all ~~data set~~datasets into a single standard format. Second, during the merging, duplicates between ~~data set~~datasets must be identified and removed. Third, the metadata should be propagated throughout the process and made available in the final merged ~~data set~~dataset. Ideally, the compiled merged ~~data set~~dataset would be made available as a simple text table, to facilitate ease of access and manipulation. In this work, such unification of multiple ~~data set~~datasets ~~is~~~~are~~is presented. This was done for the validation of the ESA OC-CCI ocean ~~-~~colour products, but with the intent to also serve the broader user community.

A merged ~~data set~~dataset is not without drawbacks: it is likely to be large (with hundreds of thousands of observations) and so not always easy to manipulate; because the merging is done on pre-existing, processed databases, it is not possible to have full control of the whole processing chain; the ~~data set~~dataset would be a collection of observations collected by several investigators using different instruments, sampling methods and protocols, which might eventually have been modified by the processing routines used by the repositories or archives. ~~-~~To minimise these potential drawbacks, this work has~~we have~~, for the most part, incorporated only ~~data set~~datasets that have emerged from the long-term efforts of the ocean ~~-~~colour and biological oceanography~~oceanographical~~ communities to provide scientists with high-quality *in situ* data~~, and implemented additional~~. Additional quality checks were implemented on the data, to enhance confidence in the quality of the merged product. Nevertheless,~~it is still recognized that different and unpredictable uncertainties~~ uncertainties arising from the diversity of instruments, methods, and processing schemes may still affect the available data~~from the diverse sources due to the use of a variety of field/laboratory instruments, methods and data reduction schemes~~.

This is the fourth version of the compilation. The first, second and third versions were described in Valente et al. (2016), Valente et al. (2019) and Valente et al. (2022), respectively. Compared to the previous version (Valente et al., 2022), the present version contains more *in situ* measurements for all target variables. The data additions arise from updates to nine existing data sources (MOBY, AERONET-OC, AWI, CALCOFI, HOT, SeaBASS, IMOS, BATS, and BIOCHEM), together with the incorporation of 13 newly added data sources (see Sections 2.3.1 and 2.3.28–2.3.39; magenta data source names in Table 1). These updates enhance both the temporal continuity and the geographic representativeness of the dataset, strengthening its suitability for satellite ocean colour validation, algorithm development, and global bio-optical analyses. Overall, the main objective of the present version was to populate the compilation with more recent data. Methodologies for data harmonization and integration (described below) have not been altered relative to the last version.

Methodologies used for data harmonization and integration, ~~as well as~~ together with a description of the acquired individual ~~data set~~datasets, are provided in Section 2. Geographic distribution and other characteristics of the final merged ~~data set~~dataset are shown in Section 3 while Section 4 provides an overview of the data.

2 Data and methods

2.1 Dataset overview and variables~~Pre-processing and merging~~

The compiled global *in situ* bio-optical ~~data set~~dataset described in this work includes observations from 40 data sources (Table 1), covering a wide range of marine environments and trophic regimes, from oligotrophic to eutrophic waters ~~has an emphasis, though not exclusively, on open-ocean data~~. It comprises the following variables: remote_~~-~~sensing reflectance ("rrs"), chlorophyll-a concentration ("chla"), algal pigment absorption coefficient ("aph"), detrital plus gelbstoff "CDOM" absorption coefficient ("adg"), particle backscattering coefficient ("bbp"), diffuse attenuation coefficient for downward irradiance ("kd") and total suspended matter ("tsm~~”~~"). ~~—~~The variables "rrs", "aph", "adg", "bbp" and "kd" are spectrally dependent, and this dependence is implied~~,~~ hereafter~~, implied~~.

The data were compiled from ~~27 sources (~~MOBY "Gold and Platinum", BOUSSOLE, AERONET-OC, SeaBASS, NOMAD, MERMAID, AMT, ICES, HOT, GeP&CO, AWI, ARCSSPP, BARENTSSEA, BATS, BIOCHEM, BODC, CALCOFI, CCELTER, CIMT, COASTCOLOUR, ESTOC, IMOS, MAREDAT, PALMER, SEADATANET, TPSS, ~~and~~ TARA, CNR_GOS, DFO, GOAL, iFADO, INIDEP, IOPAN, ISEE, KHOA, MEDGOOS, UCSD, UNH, and YNU. ~~): e~~Each ~~one~~ dataset is described in ~~Sect. 2.2~~ Section 2.3 and data contributors are listed in Table 3.

The data sources in this work should also be viewed as groups of data that were acquired from a specific source, standardized with a specific method and later merged into the compilation. The compiled *in situ* observations are essentially surface (i.e., depth ≤ 10 m~~no information depending on depth~~), have a global distribution and cover the period ~~1997~~ 1996 to ~~2021~~2025. The listed variables, with the exception of total suspended matter, were chosen as they are the operational satellite ocean-colour products of ESA OC-CCI project.

~~The compilation is provided in the format of three, two-dimensional, main tables that relate to each other via one unique key identifying each row. The format of the tables is described in Appendix B. Despite being provided in three main tables, the compilation should still be viewed conceptually as one unique table and, as such, it is still described in that way. The data set contains two flags: "flag_time" and "flag_chl_method". The first is because three data sources were used (ESTOC, MAREDAT and TPSS) where information on time (hour of the day) was not available. The time for these observations was set to 12:00:00 (UTC) and the observations were flagged with "1" in column "flag_time". A second flag was necessary, because in two data sources (ARCSSPP and SEADATANET) there was uncertainty on whether the compiled chlorophyll concentrations were measured using fluorometric, spectrophotometric or HPLC methods. The compiled chlorophyll observations from these two data sources were flagged with "1" in column "flag_chl_method" and were marked as "chla_fluor".~~

2.1.1 processing and normalization of remote sensing reflectance (rrs)

~~This is the third version of the compilation. The first and second versions were described in Valente et al. (2016) and Valente et al. (2019), respectively. Compared to the previous version (Valente et al., 2019), the present version contains more measurements of "rrs", "chla" and "aph". The "rrs" stations increased by ~15% (i.e., from 59,781 to 68,641), resulting from updates of AERONET-OC, BOUSSOLE, MOBY, MERMAID and AWI. The new stations are mainly for the period of 2019-2021 (previous version had "rrs" data until 2018). Regarding "chla", a major increase in the number of recent observations was obtained. The previous version had "chla" data until 2017, with 533 stations for the period 2016-2017. The current version has 5,140 stations for 2016-2021, which constitutes a near tenfold (964 %) increase since 2016. The new "chla" data originates from updates of BOUSSOLE, MERMAID, SeaBASS, HOT, AMT, PALMER, CCELTER, CALCOFI, AWI and IMOS. As for the number of "aph" stations, it~~

~~increased by ~30 % (i.e., from 3,293 to 4,265), with most of the data between 2012-2020 (previous version finished in 2012). The new "aph" data comes from updates of SeaBASS and AWI.~~ ~~Overall, the main objective of the present version was to populate the compilation with more recent data. Methodologies for data harmonization and integration (described below) have not been altered relative to the last version.~~

Remote-sensing reflectance is a primary ocean colour product defined as "rrs = Lw/Es", where "Lw" is the upward (above) water-leaving radiance and "Es" is the total downward irradiance at sea level. Another quantity that is often required is the "normalized" water-leaving radiance ("nLw") (Gordon and Clark, 1981), which is related to remote-sensing reflectance via "rrs = nLw/Fo", where "Fo" is the top-of-the-atmosphere solar irradiance. If not directly available, remote-sensing reflectance was calculated through the equations described above, depending on the format of the original data. The original data were acquired in an advanced form (e.g., time-averaged, extrapolated to surface) from nine data sources designed for ocean ~~-~~colour validation and applications (MOBY, BOUSSOLE, AERONET-OC, SeaBASS, NOMAD, MERMAID, COASTCOLOUR, TARA, AWI), therefore, only requiring the conversion to a common format. In processing by space agencies, the quantity "rrs" is normalized to a single Sun-viewing geometry (Sun at zenith and nadir viewing) ~~taking in account the~~ to account for bidirectional reflectance effects through the application of a bidirectional reflectance distribution function (BRDF) correction. In previous versions of this compilation, this correction followed ~~as described in~~ Morel and Gentili (1996) and Morel et al. (2002). In the v2026 database, the BRDF correction follows Park and Ruddick (2005), (hereafter, PR05). The transition to PR05 ensures consistency with the atmospheric correction used in the OC-CCI processing chain, which relies on the Polymer processor (Steinmetz et al., 2011), where the PR05 BRDF correction is implemented. The adoption of the same BRDF formulation harmonizes the normalization applied to *in situ* and satellite remote-sensing reflectance products. Thus, for consistency with satellite "rrs" product, the ~~latter~~ PR05 normalization was applied to the *in situ* "rrs". However, the BRDF correction was not applied to NOMAD, MERMAID, and some SeaBASS data (i.e., 698 samples from v2022), either because the "rrs" values were already fully normalized via M02 or due to a lack of information on viewing geometry.

2.1.2 Compilation of chlorophyll-a (chla)

Chla ~~Chlorophyll-a concentration~~ is a proxy measure for phytoplankton biomass and one of the most-widely used satellite ocean ~~-~~colour products (IOCCG, 2008). To validate satellite-derived ~~chlorophyll-a concentration~~chla, two different variables were compiled: one of these represents chlorophyll-a measurements made through fluorometric or spectrophotometric methods, referred to hereafter as "chla_fluor" and the other is the chlorophyll concentration derived from HPLC (High-Performance Liquid Chromatography) measurements, referred to hereafter as "chla_hplc". The chlorophyll data were compiled from the following ~~25~~ 37 data sources: BOUSSOLE, SeaBASS, NOMAD, MERMAID, AMT, ICES, HOT, GeP&CO, AWI, ARCSSPP, BARENTSSEA, BATS, BIOCHEM, BODC, CALCOFI, CCELTER, CIMT, COASTCOLOUR, ESTOC, IMOS, MAREDAT, PALMER, SEADATANET, TPSS, ~~and~~ TARA, DFO, UNH, YNU, ISEE, iFADO, IOPAN, UCSD, INIDEP, GOAL, CNR_GOS, MEDGOOS, and KHOA. One requirement for "chla_fluor" measurements was that they were made using in vitro methods (i.e., based on extractions of chlorophyll-a). Although this severely decreased the number of observations, since in vivo fluorometry (e.g., fluorometers mounted on CTD's) is widely available in oceanographic databases, it was decided to exclude such data because of potential problems with the calibration of *in situ* fluorometer data. The variable "chla_hplc" was calculated by summing all reported chlorophyll-a derivatives, including divinyl chlorophyll-a, epimers, allomers, and chlorophyllide-a. The two chlorophyll variables are retained separately in the database to facilitate their use. HPLC measurements offer ~~could be considered of~~ higher sensitivity and accuracy, while~~quality~~ ~~, but~~ fluorometric measurements are more numerous. Thus, one option for users is to use "chla_fluor" only when there are no "chla_hplc" measurements available. To be consistent with satellite-derived chlorophyll values, which are derived from the light emerging from the upper layer of the ocean, all chlorophyll observations in the top 10 meters (replicates at the same depth, or measurements at multiple depths) were averaged if the coefficient of variation among observations was less than 50 %, otherwise they were discarded. The averages were then assigned to the surface. The depth of 10 m was chosen as a compromise between clear oligotrophic and turbid eutrophic waters. Other methods, such as chlorophyll depth-averages using local attenuation conditions (Morel and Maritorena, 2001), require observations at multiple depths, which, given our decision to use only in vitro measurements, would have reduced considerably the final number of observations.

2.1.3 Compilation of other variables

If the ~~inherent optical properties~~IOPs ("aph", "adg" and "bbp") had not already been calculated and provided in the contributed ~~data set~~datasets, they were computed from related available variables ~~that were available~~: particle absorption ("ap"), detrital absorption ("ad"), coloured dissolved organic matter ~~(CDOM)~~ absorption ("ag"), total backscattering ("bb"). The following equations were used "adg = ad + ag", "ap = aph + ad", and "bb = bbp + bbw". For the latter equation, the variable "bbw" was computed using "bbw = bw/2", where "bw" is the scattering coefficient of seawater derived from Zhang et al. (2009). The diffuse attenuation coefficient for downward irradiance ("kd") did not require any conversion and was compiled as originally acquired. The methodology and depth interval used for its estimation were not consistently reported across the contributing datasets. Consequently, the compiled "kd" values may reflect different integration depths and estimation protocols depending on the original data source. Observations of ~~inherent optical properties~~IOPs and kd were acquired in total from ~~six~~ 12 data sources (SeaBASS, NOMAD, MERMAID, AWI, COASTCOLOUR, TPSS, DFO, UNH, YNU, ISEE, CNR_GOS, and MEDGOOS). Concerning total suspended matter "tsm", these data were compiled from 7 sources (SeaBASS, NOMAD, MERMAID, COASTCOLOUR, ETOC, CNR_GOS, KHOA, and UNH).

~~The merged data set was compiled from 27 sets of *in situ* data, which were obtained individually either from archives that incorporate data from multiple contributors (SeaBASS, NOMAD, MERMAID, ICES, ARCSSPP, BIOCHEM, BODC, COASTCOLOUR, MAREDAT, SEADATANET), or from particular contributors, measurement programs or projects (MOBY, BOUSSOLE, AERONET-OC, HOT, GeP&CO, AMT, AWI, BARENTSSEA, BATS, CALCOFI, CCELTER, CIMT, ESTOC, IMOS, PALMER, TPSS, TARA) and were subsequently, homogenized and merged. Data contributors are listed in Table 2 and in the auxiliary material.~~

2.2 ~~x~~Pre-processing, quality control, and data structure ~~xxx~~

The compilation is provided in the format of three, two-dimensional, main tables that relate to each other via one unique key identifying each row. The format of the tables is described in ~~Appendix B~~Code and data availability section. Despite being provided in three main tables, the compilation should still be viewed conceptually as one unique table and, as such, it is still described in that way.

The ~~data set~~dataset contains two flags: "flag_time" and "flag_chl_method". The first is because three data sources were used (ESTOC, MAREDAT and TPSS) where information on time (hour of the day) was not available. The time for these observations was set to 12:00:00 (UTC) and the observations were flagged with "1" in column "flag_time".~~–~~ A second flag was necessary, because in two data sources (ARCSSPP and SEADATANET) there was uncertainty on whether the compiled chlorophyll concentrations were measured using fluorometric, spectrophotometric or HPLC methods. The compiled chlorophyll observations from these two data sources were flagged with "1" in column "flag_chl_method" and were marked as "chla_fluor".

There were methodological differences between ~~data set~~datasets. Therefore, after acquisition, and prior to any merging, each set of data was pre-processed for quality control and converted to a common format. During this process, data were discarded if they had: 1) unrealistic or missing, date and geographic coordinate fields; 2) poor quality (e.g., original flags) or measurement method ~~of observation~~ that did not meet the criteria for the ~~data set~~dataset (e.g., *in situ* fluorescence for chlorophyll concentration); and 3) spuriously high or low data. For the last, the following limits were imposed: for "chla_fluor" and "chla_hplc" [0.001-~~100~~500] mg m^{-3}; for "rrs" [0-0.15] sr^{-1}; for "aph", "adg" and "bbp"~~–~~ [0.0001-~~10~~30] m^{-1}; for "tsm" [0-1000] g m^{-3}; for "kd" [(aw(λ)-~~10~~20] m^{-1}, where "aw" are the pure water absorption coefficients derived from Pope and Fry (1997). Also, during this stage, three metadata strings were attributed to each observation: "dataset", "subdataset" and "contributor". The "dataset" contains the name of the original set of data, and can only be one of the following: "aoc", "boussole", "mermaid", "moby", "nomad", "seabass", "hot", "ices", "amt", "gepco", "arcsspp", "awi", "barentssea", "bats", "biochem", "bodc", "calcofi", "cc", "ccelter", "cimt", "estoc", "imos", "maredat", "palmer", "seadatanet", "tpss", "tara", "cnr_gos", "dfo", "goal", "ifado", "inidep", "iopan", "isee", "khoa", "medgoos", "ucsd", "unh", "ynu". The "subdataset" starts with the "dataset" identifier and is followed by additional information about the data, as <dataset>_<cruise/station/site>) (e.g., "seabass_car81"). The "contributor" contains the name of the data contributor. An effort was made to homogenize the names of data contributors from the different sets of data. These three metadata are the link to trace each observation to its ~~origin~~origin, and they were propagated throughout the processing stage. Finally, ~~this processing stage ended with~~ each set of data ~~being~~ was scanned for replicated variables ~~data~~ and replicate stations if found ~~data~~, the~~se~~ data ~~which when found,~~ were averaged if the coefficient of variation was less than 50 %, otherwise they were discarded. Replicates were defined as multiple observations of the same variable, with the same date, time, latitude, longitude and depth. Replicate station data were defined as multiple measurements of the same variable, with the same date, time, latitude and longitude. For the latter case, a search window of 5 minutes in time and 200 meters in distance was given to account for station drift. A small number of observations that were identified as replicates had a different "subdataset" identifiers (i.e., different cruise names). These observations were considered suspicious if the values were ~~different, and~~different and discarded. If the values were the same, one of the observations was retained. This possibly originated from the same group of data being contributed to an archive by two different data contributors.

Once a set of data was homogenized, its data were integrated into a unique table. This final merging focused on the removal of duplicates between the sets of data. Although some duplicates are known (e.g., MOBY, BOUSSOLE, AERONET-OC and NOMAD data are found in SeaBASS and MERMAID), others are unknown (e.g., how many of GeP&CO, ICES, AMT, HOT are within NOMAD, SeaBASS and MERMAID). Therefore, duplicates were identified using the metadata ("dataset" and "subdataset") when possible, and by considering the temporal-spatial matches between different datasets (i.e., within 5 minutes and 200 meters). ~~, as an additional precaution. For temporal-spatial matches, several thresholds were used, but typically 5 minutes and 200 meters were taken to be sufficient to identify most duplicated data, which reflected small differences in time, latitude and longitude, between the different sets of data.~~

Larger thresholds were used in some cases as a precautionary procedure. This was the case when searching for NOMAD data in other ~~data set~~datasets, because NOMAD includes a few cases where merging of radiometric and pigment data was done with large spatial-temporal thresholds (Werdell and Bailey, 2005). A large temporal threshold was also used when integrating observations from the three data sources that did not have time available (ESTOC, MAREDAT and TPSS). In regard to all data, if duplicates were found, data from the NOMAD ~~data set~~dataset were selected first, followed by data from individual projects or contributors (MOBY, BOUSSOLE, AERONET-OC, AMT, HOT,GeP&CO, AWI, BARENTSSEA, BATS, CALCOFI, CCELTER, CIMT, ESTOC, IMOS, PALMER, TPSS and TARA) and finally for the remaining ~~data set~~datasets (SeaBASS, MERMAID, ICES, ARCSSPP, BIOCHEM, BODC, COASTCOLOUR, MAREDAT, ~~and~~ SEADATANET, DFO, UNH, YNU, ISEE, iFADO, IOPAN, UCSD, INIDEP, GOAL, CNR_GOS, MEDGOOS, and KHOA). This procedure was chosen to preserve the NOMAD ~~data set~~dataset as a whole, since it is widely used in ocean ~~-~~colour validation. It should be noted that, by this procedure, data from individual projects or contributors may be listed under NOMAD (e.g., some PALMER data are found in NOMAD with metadata string "nomad_palmer_lter"). After ~~giving priority to~~ NOMAD, the priority was generally given to data from individual projects or contributors, but due to an incremental approach, where only new data are added to previous versions of the compilation, some data from individual projects or contributors (BATS, CALCOFI, CIMT, PALMER and TPSS) added in later stages, may be found under other data sources. This occurs mainly for BATS and CALCOFI, which have their earlier chlorophyll data in SeaBASS with metadata strings "seabass_bats*" and "seabass_cal*", and CIMT which has some of its data under COASTCOLOUR. After all data from a given source were free of duplicates, they were merged consecutively by variable in the final table. During this process, ~~we also searched for~~ rows (stations) that were close in space and time were merged. ~~separated from each other by time differences less than 5 minutes and horizontal spatial differences of less than 200 meters.~~ When such rows were found, the observations in those rows were merged into a single row. The compiled merged data were compared with the original sets to certify that no errors occurred during the merging. As a final step, a water-column (station) depth was recorded for each observation, which was the closest water column depth from the ETOPO1 global relief model (National Geophysical Data Center ETOPO1; Amante and Eakins, 2009). For observations where the closest water depth was above sea level (e.g., data collected very near the coast), it was given the value of zero.

Data processing thus included two major steps: pre-processing and merging. The first step was related to the processing of each of the ~~27~~ 40 contributing ~~data set~~datasets and aimed to identify problems and convert the data of interest to a standard format. The second step dealt with the integration of all the contributing sets of data into a unified ~~data set~~dataset and included the elimination of duplicated data between the individual sets of data. In the next subsections a brief overview of each original set of data is provided.

~~2.2~~2.3 Pre-processing of ~~each~~ individual ~~sets of data~~dataset

~~2.2~~2.3.1 Marine Optical BuoY (MOBY)

MOBY is a fixed mooring system operated ~~by the National Oceanic and Atmospheric Administration (~~NOAA~~)~~ that provides a continuous time series of water-leaving radiance and surface irradiance in the visible region of the spectra since 1997. The site is located a few kilometres west of the Hawaiian Island of Lanai where the water depth is about 1200 m. Since its deployment, MOBY measurements have been the primary basis for the on-orbit vicarious calibrations of the SeaWiFS and MODIS ocean colour sensors. A full description of the MOBY system and processing is provided in Clark et al. (2003). Data are freely available for scientific use at the MOBY Gold directory. The products of interest are the "Scientific Time Series" files, which refer to MOBY data averaged over sensor-specific wavelengths and particular hours of the day (around 20-23 UTC). For this work, the satellite band-average products for SeaWiFS, MODIS AQUA, MERIS, VIIRS-SNPP, VIIRS-~~JPSS~~ NOAA20 (also known as VIIRS-JPSS~~NOAA-20 VIIRS~~), VIIRS-NOAA21, OLCI-S3A and OLCI-S3B were compiled from the "R2017 Reprocessing".

~~In addition to the MOBY Gold, ~~T~~t~~he v2026 database also incorporates data from ~~the ~~MOBY Platinum, which represents the continuation of the MOBY time series with improved instrumentation. The Platinum configuration employs a dual in-line spectrometer (DILS) system that records radiance and irradiance concurrently, in contrast to the MOBY Gold system where these quantities were measured sequentially. The Platinum products also include a Best Lw Product flag that identifies the preferred estimate of water-leaving radiance for use in analyses. In addition to the historical MOBY site near Hawaii, the Platinum system also includes measurements from the MarONET site off Western Australia, thereby extending the geographic coverage of the time series. The MOBY Gold data record extends until 31 May 2022, whereas the Platinum measurements begin in May 2024.

The "inband" average subproduct was used, and to maintain the highest quality, only data determined from the upper two arms ("Lw1") and flagged "good" quality were acquired. Data from the MOBY203 deployment were discarded due to the absence of surface irradiance data. The compiled variable was the remote-sensing reflectance, "rrs", which was computed from the original water-leaving radiance ("Lw") and surface irradiance ("Es"). The water-leaving radiances were corrected for the bidirectional nature of the light field (Park and Ruddick, 2005) ~~(Morel and Gentili, 1996; Morel et al., 2002) using the same look-up table and method as that used in the SeaWiFS Data Analysis System (SeaDAS) processing code~~. The MOBY Gold data were reprocessed in 2017 ("MOBY R2017 Reprocessing") to include various improvements in the calibration of the instrument and post processing, which include: 1) a new method to extrapolate the upwelling radiance attenuation coefficient to the surface (Voss et al., 2017); 2) an increase in arm depth by 0.234 m; and 3) a single pixel shift in the data for the red spectrograph collected at a bin factor of 384. Only the last two changes were included in present compilation. The first change uses model results to improve Lw at wavelengths above 575 nm, by correcting the diffuse upwelling radiance attenuation coefficient for inelastic effects. ~~Thus~~Thus, for wavelengths above 575 nm, the Lw21 product, in the Gold directory should be ~~investigated~~used. As mentioned before, the MOBY data compiled in this work are satellite-sensor -specific. Therefore, ~~attention ~~it is necessary to use the correct

MOBY data when validating a particular sensor. The way MOBY data are stored in the final merged table is consistent with the original wavelengths; however, these wavelengths can differ from what is sometimes expected to be the central wavelength of a given band and sensor (Table 3). Irrespective of the wavelength where MOBY data are stored in the final table, for validation of bands 1-6 of SeaWiFS, MOBY data stored in the final merged table at 412, 443, 490, 510, 555 and 670 nm, respectively, should be used. For validation of bands 1-9~~7~~ of MODIS AQUA, MOBY data stored in the final merged table at 416, 442, 489, 530, 547, 665, ~~and~~ 677~~-~~, 747, and 865 nm, respectively, should be used. For validation of bands 1-10 of MERIS, MOBY data stored in the final merged table at 410.5, 440.4, 487.8, 507.7, 557.6, 617.5, 662.4, 679.9, 706.2 and 752.5 nm, respectively, are the appropriate data. For validation of bands 1-12 of OLCI-S3A, MOBY data stored in the final merged table at 400.3032, 411.8453, 442.9626, 490.493, 510.4676, 560.4503, 620.4092, 665.2744, 674.0251, 681.5705, 709.1149 and 754.1813, respectively, are the appropriate data. For validation of bands 1-12 of OLCI-S3B, MOBY data stored in the final merged table at 400.5947, 411.9509, 442.9882, 490.3991, 510.4022, 560.3664, 620.284, 665.1312, 673.8682, 681.3856, 708.9821 and 754.0284, respectively, are the appropriate data. For validation of bands 1-5 of VIIRS-SNPP, MOBY data stored in the final merged table at 412.9, 444.5, 481.2, 556.3 and 674.6 nm, respectively, are the appropriate data. ~~Finally, for~~For the validation of bands 1-5 of VIIRS-~~JSPP~~NOAA20, MOBY data stored in the final merged table at 411, 445, 489.01, 556 and 667 nm, respectively, are the appropriate data. For the latter sensor, the original value was 489 nm, but it was changed to 489.01 nm to differentiate from the 489 nm of MODIS AQUA. Finally, for bands 1–6 of VIIRS-NOAA21, MOBY data were stored at the corresponding sensor cent~~er~~re wavelengths of 411.051, 444.895, 488.363, 555.333, 671.366, and 747.317 nm. A summary of the satellite band ~~centers~~centres and the MOBY wavelengths stored in the final merged table is provided in Table 4. ~~-~~The look-up table to fully normalize "rrs" only covers the range 413-660 nm; compared to the previous versions of the compilation, in the present version, the "rrs" MOBY at wavelengths outside this range were not discarded and fully normalized using the closest entry of the lookup table (i.e., at 413 nm or 660 nm).

~~2.2~~2.3.2 BOUée pour l'acquiSition de Séries Optiques à Long termE (BOUSSOLE)

The BOUSSOLE Project started in 2001 with the objective of establishing a time series of bio-optical properties in oceanic waters to support the calibration and validation of ocean ~~-~~colour satellite sensors (Antoine et al., 2006). The project consists of a monthly cruise program and a permanent optical mooring (Antoine et al., 2008). The mooring collected~~s~~ radiometry and inherent optical properties (IOPs) in continuous mode every 15 minutes at 2 depths (4 and 9 m nominally). M~~The m~~onthly cruises we~~a~~re devoted to the mooring servicing, to the collection of vertical profiles of radiometry and IOPs, and to water sampling at 11 depths from the surface down to 200 m, for subsequent analyses including phytoplankton pigments, particulate absorption, CDOM absorption and suspended particulate matter load. The BOUSSOLE mooring wa~~i~~s in the Western Mediterranean Sea at a water depth of 2400m. All pigment (2001-2019) and radiometric (two subsets: 2003-2012 and 2015-2019) data were provided by the Principal Investigators. The first radiometric subset was obtained from measurements made with multispectral Satlantic OCI-200 radiometers; the second radiometric subset was obtained from measurements made with hyperspectral Satlantic OCR radiometers, convolved with spectral response function of Sentinel3 OLCI-A bands. The compiled variables were "rrs" and "chla_hplc". Remote-sensing reflectance was computed from the original "fully-normalized" water-leaving radiance ("nLw_ex"), which is the "normalized" water-leaving radiance ("nLw" previously described), with a correction for the bidirectional nature of the light field (Morel and Gentili, 1996; Morel et al., 2002). The solar irradiance ("Fo") was computed from two available variables in the original set of data: the normalized water-leaving radiance ("nLw") and the remote-sensing reflectance ("rrs"), using the equation "Fo = nLw/rrs". Only radiometric observations that me~~e~~t the following criteria were used: 1) tilt of the buoy was less than 10°; 2) the buoy was not lowered by more than 2 m as compared to its nominal water line (to ensure the Es reference sensor is above water and exempt from sea spray); and 3) the solar irradiance was within 10 % of its theoretical clear-sky value (determined from Gregg and Carder, 1990). The latter criterion was used to select clear skies only. An additional quality control was to remove observations that were 50 % higher or lower than the daily average. This removed a small number of "spikes" in the time series. The final quality control step was to remove days where the standard deviation was more than half of the daily average. This was meant to identify days with high variability. Very few days (N = 2) were removed with this test. These quality control criteria were applied per wavelength, which resulted in some observations with an incomplete spectrum.

~~2.2~~2.3.3 AErosol RObotic NETwork-Ocean Color (AERONET-OC)

AERONET-OC is a component of AERONET, including sites where sun-photometers operate with a modified measurement protocol leading to the determination of the fully-normalized water-leaving radiance (Zibordi et al., 2006; Zibordi et al., 2009). As a result of a collaboration between the Joint Research Centre (JRC) and NASA to develop (Hooker et al., 2000) and exploit (Zibordi et al., 2002) the technology, this component has been specifically developed for the validation of ocean ~~-~~colour radiometric products. The strength of AERONET-OC is "the production of standardised measurements that are performed at different sites with identical measuring systems and protocols, calibrated using a single reference source and method, and processed with the same codes" (Zibordi et al., 2006; Zibordi et al., 2009). All ~~high~~ high-quality ~~data~~ ("Level-2") data from "version 3" reprocessing of AERONET-OC (Zibordi et al., 2021) were acquired from the project website, for 39 sites (compared with 11 sites in ~~the previous version of the database~~Valente et al. 2022)~~: Abu_Al_Bukhoosh (~25° N, ~53° E) , COVE_SEAPRISM (~36° N, ~75° W), Gloria (~44° N, ~29 °E), Gustav_Dalen_Tower (~58 °N, ~17 °E), Helsinki Lighthouse (~59 °N, ~24° E), LISCO (~40° N, ~73° W), Lucinda (~18° S, ~146° E), MVCO (~41° N, ~70° W), Palgrunden (~58° N, ~13° E; Philipson et al., 2016), Venice (~45° N, ~12° E) and WaveCIS_Site_CSI_6 (~28° N, ~90° W)~~. The compiled variable was "rrs". Remote-sensing reflectance was computed from the original "fully-normalized" water-leaving radiance (see ~~Sect. 2.2~~ Section 2.3.2 for definition). The solar irradiance ("Fo"), which is not part of the AERONET-OC data, was computed from the Thuillier (2003) solar spectrum irradiance, by averaging "Fo" over a wavelength-centred 10 nm window.

For each AERONET-OC site, the exact wavelengths of each record can change over time depending on the specific instrument deployed. In previous versions of the compilation, AERONET-OC "rrs" values were stored in the database at the exact wavelengths reported for each measurement. In the v2026 database, the AERONET-OC "rrs" values are stored at 23 standard wavelengths (340, 380, 400, 412, 440, 443, 490, 500, 510, 531, 532, 551, 555, 560, 620, 667, 675, 681, 709, 779, 865, 870, 1020). The exact wavelengths associated with each measurement are preserved in separate columns named Exact_Wavelengths(um)_xxx, where xxx corresponds to each of the 23 standard bands.~~Data were compiled for the exact wavelengths of each record, which can change over time for a given site depending on the specific instrument deployed. In comparison with the previous version of the compilation, the present OC-CCI data set version 3, now uses the "version 3" reprocessing of AERONET-OC data (Zibordi et al., 2021).~~

~~2.2~~2.3.4 SeaWiFS Bio-optical Archive and Storage System (SeaBASS)

SeaBASS is one of the largest archives of *in situ* marine bio-optical data (Werdell and Bailey, 2003) with a long-established inventory (Hooker et al. 1994). It is maintained by NASA's Ocean Biology Processing Group (OBPG) and includes measurements of optical properties, phytoplankton pigment concentrations, and other related oceanographic and atmospheric data. The SeaBASS database consists of *in situ* data from multiple contributors, collected using a variety of measurement instruments with consistent, community-vetted protocols, from several marine platforms such as fixed buoys, hand-held radiometers and profiling instruments. Quality control of the received data includes a rigorous series of protocols that range from file format verification to inspection of the geophysical data values (Werdell and Bailey, 2003). Radiometric data were mostly acquired through the "Validation" search tool, which provided *in situ* data with matchups for particular ocean ~~-~~colour sensors (Bailey and Werdell, 2006). The criterion in the search-query was defined to have the minimal flag conditions in the satellite data, to retrieve a greater number of matchups and, therefore, *in situ* data. Regarding phytoplankton pigment data, the majority were acquired through the "Pigment" search tool, which provided pigment data directly from the archives.

As was stated in the SeaBASS website http://seabass.gsfc.nasa.gov/seabasscgi/search.cgi, the "Pigment" search tool was originally designed to return only in vitro fluorometric measurements, which is consistent with our approach, but over time chlorophyll-a measurements made using other methods (e.g., in vivo fluorometry) were included in the retrieved pigment data. In the pigment data used in this work, a large number of *in situ* fluorometric measurements from continuous underway instruments were identified and discarded, as such measurements may be subject to calibration uncertainties and are not directly comparable with in vitro methods. These data were initially identified from cruises with more than 50 observations per ~~day, and~~day and then re-checked in the SeaBASS website to confirm whether indeed they were continuous underway measurements. A total of 120,412 such measurements were identified and discarded. Given the large volume of this group of data, it is possible that some chlorophyll-a observations from *in situ* methods may have escaped the scrutiny and persisted into the final merged ~~data set~~dataset. The "Pigment" search tool was recently discontinued and, instead, the "File" search tool can be used, which was also used here to acquire chlorophyll, as well as radiometric observations, for more recent years. For the present version (v2026), data retrieved using the "File" search tool were constrained to observations archived between 1 January 2022 and 1 September 2025, with data type "all", and depths between 0 and 100 m; however, measurements deeper than 10 m were subsequently excluded during post-processing. The remote sensing reflectance acquired from the "File" search tool were corrected for the bidirectional effects using PR05 normalization, where geometry information was available and where the data were not already normalized ~~(Morel and Gentili, 1996; Morel et al., 2002)~~. The compiled variables from SeaBASS data were: "rrs", "chla_hplc", "chla_fluor", "aph", "adg", "bbp", "kd".

~~2.2~~2.3.5 NASA bio-Optical Marine Algorithm ~~Data set~~Dataset (NOMAD)

NOMAD is a publicly-available ~~data set~~dataset compiled by the NASA OBPG ~~at the Goddard Space Flight Center~~. It is a high-quality global ~~data set~~dataset of coincident radiometric and phytoplankton pigment observations for use in ocean ~~-~~colour algorithm development and satellite-data product-validation activities (Werdell and Bailey, 2005). The source bio-optical data is the SeaBASS

~~archive,~~archive; therefore, many dependencies exist between these two ~~data set~~datasets, which were addressed during the merging. The current version (Version 2.0 ALPHA, 2008) includes data from 1991 to 2007 and an additional set of observations of inherent optical properties. The current version was used in this work, but with an additional set of columns of remote-sensing reflectance corrected for the bidirectional effects (Morel and Gentili, 1996; Morel et al., 2002; Park and Ruddick, 2005). This additional set of columns was provided directly by the NOMAD creators. The compiled variables were "rrs", "chla_hplc", "chla_fluor", "aph", "adg", "bbp", "kd". Conversion was necessary only for "aph", "adg" and "bbp", and followed the procedures described in ~~Sect.~~ Section 2.1. For the calculation of "bbp" the variable "bb" was used with a smooth fitting to remove noise. A portion of NOMAD data were optically weighted (for methods see Werdell and Bailey, 2005). These data are not consistent with the protocols chosen in this work, but these observations were retained since NOMAD is a ~~widely-used~~widely used ~~data set~~dataset in ocean ~~-~~colour validation.

~~2.2~~2.3.6 MERIS Match-up *In situ* Database (MERMAID)

MERMAID provides *in situ* bio-optical data matched with concurrent and comparable MERIS Level 2 satellite ocean ~~-~~colour products (Barker, 2013a; Barker, 2013b). The MERMAID *in situ* database consists of data from multiple contributors, measured using a variety of instruments and protocols, from several marine platforms such as fixed buoys, hand-held radiometers and profiling instruments. Comprehensive quality control and protocols are used by MERMAID to integrate all the data into a common and comparable format (Barker, 2013a; Barker, 2013b). Access to MERMAID data is limited to the MERIS Validation Team, the MERIS Quality Working Group and to the *in situ* data contributors. For this work, access has been granted to the MERMAID database, through a signed Service Level Agreement. ~~The~~ MERMAID ~~data~~ includes sub-sets of several ~~data set~~datasets used in this compilation (MOBY, AERONET-OC, BOUSSOLE, NOMAD~~). These observations~~) and were removed ~~from the MERMAID data set~~ to avoid duplication (as discussed in ~~Sect.~~Section 2.1). The compiled variables were "rrs", "chla_hplc", "chla_fluor", "aph", "adg", "bbp", "kd" and "tsm". Remote-sensing reflectance was calculated by dividing by π the original "fully-normalized" water-leaving reflectance ("Rw_ex"), which is the water-leaving reflectance (Rw= π Lw / Es), with a correction for the bidirectional nature of the light field (Morel and Gentili, 1996; Morel et al., 2002). Conversion was also necessary for "aph", "adg" and "bbp", and followed the procedures described in ~~Sect.~~ Section 2.1.

In comparison with the previous version of the compilation, a set of "chla_fluor" observations from MERM~~I~~AID were considered suspicious and excluded from the compilation (N=3241, from mermaid_MAREL-carnot, mermaid_MAREL-itroise and mermaid_MAREL-vilaine).

~~2.2~~2.3.7 Hawaii Ocean Time-series (HOT)

The HOT programme provides repeated comprehensive observations of the hydrography, chemistry and biology of the water column at ~~a~~ stations located 100 km north of Oahu, Hawaii, since October 1988 (Karl and Michaels, 1996). The HOT dataset used in this work originates from four sites (ALOHA, Kahe Point, Kaena, and HALE ALOHA),~~This site~~ ~~is~~ which are representative of the North Pacific subtropical gyre. ~~Cruises are made approximately once a month to the deep-water~~ Station ALOHA (A Long-Term Oligotrophic Habitat Assessment; 22° 45′ N, 158° 00′ W) is sampled approximately once a month. Pigment data ("chla_hplc" and "chla_fluor") were extracted directly from the project website. Radiometric measurements from the HOT project are also available, but observations of "rrs" and "kd" from the HOT project were acquired in this work as part of the SeaBASS ~~data set~~dataset.

The present version incorporates observations from three additional stations: Kahe Point (21°20.6′ N, 158°16.4′ W), Kaena (21°50.8′ N, 158°21.8′ W), and Hale ALOHA (22°27.5′ N, 158°7.9′ W). These stations extend the spatial coverage of the dataset in the waters surrounding the Hawaiian Islands while maintaining consistency with the long-term HOT observational framework. The observations from these additional stations include "chla_flour".

~~2.2~~2.3.8 Geochemistry, Phytoplankton, and Color of the Ocean (GeP&CO)

GeP&CO wa~~i~~s part of the French PROOF programme and aimed~~s~~ to describe and understand the variability of phytoplankton populations, and to assess its consequences on the geochemistry of the oceans (Dandonneau and Niang, 2007). It wa~~i~~s based on the quarterly travels of the merchant ship Contship London from France to New Caledonia in the Pacific. A scientific observer sailed on each trip and operated the sampling for surface water, filtration, various measurements and checking at several times of each day. The experiment started in October 1999 and finished in July 2002. Pigment data were extracted from the project website. Additional pigment data obtained during the OISO-4 cruise in the south Indian Ocean onboard R/V Marion-Dufresne (Jan-Feb 2000) were added. The samples were measured by Yves Dandonneau following the method used in the GeP&CO project. The compiled variable was "chla_hplc" and "chla_fluor".

~~2.2~~2.3.9 Atlantic Meridional Transect (AMT)

AMT is a multidisciplinary programme, which undertakes biological, chemical and physical oceanographic research ~~during~~ on transects between the UK and destinations in the South Atlantic (Robinson et al., 2006). The programme was established in 1995 (e.g., Robins et al. 1996 and Aiken et al. 1998) and since then has completed ~~29~~ 31 research cruises. Pigment data between 1997 (AMT5) and 2018 (AMT28) were mostly provided by the British Oceanographic Data Centre (BODC) following a specific request for discrete observations of chlorophyll-a concentration since 1997. The AMT data were isolated by searching for the string "AMT" in the "Cruise" columns and the respective Principal Investigators were then searched individually in a separated metadata file. Data not flagged with highest quality or without method of measurement were not used. For ~~any~~ interest in the original data, BODC is the point of contact, which ensures that if there are any updates, the most recent data are supplied. In the case of AMT 26, 27 and 28, data were provided to the OC-CCI project by Gavin Tilstone, whereas in the case of AMT 20 and 23, data were provided by Robert J. W. Brewin. The compiled variables are "chla_hplc" and "chla_fluor".

~~2.2~~2.3.10 International Council for the Exploration of the Sea (ICES)

ICES is a network of more than 4000 scientists from almost 300 institutes, with 1600 scientists participating in activities annually. The ICES Data Centre manages a number of large ~~data set~~dataset collections related to the marine environment covering the Northeast Atlantic, Baltic Sea, Greenland Sea and Norwegian Sea. Most of data originate from national institutes that are part of the ICES network of member countries. Data were provided (on 2014-04-28) from the ICES database on the marine environment ~~(Copenhagen, Denmark)~~ following a specific request. The ICES data were made available under the ICES data policy and if there is any conflict between this and the policy adopted by the users, then the ICES policy applies. The compiled variables were "chla_hplc" and "chla_fluor".

~~2.2~~2.3.11 Arctic System Science Primary Production (ARCSSPP)

ARCSSPP database is a synthesis of observations between 1954 and 2006, from the Arctic Ocean and northern Seas (Matrai et al., 2013). The observations were acquired from data repositories, publications or provided by individual investigators. The database includes quality-controlled observations of productivity and chlorophyll~~-~~-a, photosynthetically available radiation and hydrographic parameters. This collection of data was acquired at http://www.nodc.noaa.gov/cgi-bin/OAS/prd/accession/download/63065. For the present work, only observations of chlorophyll-a concentration with known time zones were used. The compiled chlorophyll observations were from discrete samples, but the exact method (either "chla_fluor" or "chla_hplc") was not available for all observations. Thus, the ARCSSPP chlorophyll observations were marked as "chla_fluor", although some might have been from HPLC measurements, and were flagged with "1" in a column "flag_chla_method". The compiled variable was "chla_fluor".

~~2.2~~2.3.12 Data provided by Astrid Bracher, Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research (AWI)

In this work, the AWI data source refers to the group of observations that were provided to OC-CCI project by Astrid Bracher. These are bio-optical observations collected during several cruises across the globe. The observations span the Atlantic, Pacific, Indian, Arctic, and Southern Oceans and include coastal and shelf environments such as the North Sea, with a particular emphasis on the Atlantic Ocean. All data a~~we~~re available through the PANGAEA repository. The compiled variables include chla_hplc, rrs, aph, kd, and adg (~~. adg was~~ computed as the sum of ad and ag)~~.~~ . ~~Observations of concentration of chlorophyll-a, and 1nm spectrally resolved remote sensing reflectances and algal pigment absorption coefficient were considered.~~ The methods for these observations are described by Álvarez et al. (2022), Bracher et al. (2015), Taylor et al. (2011a), Liu et al. (2018) and Tilstone et al. (2020). A comprehensive list of the AWI datasets used in this compilation is provided in Table B1 of Appendix B.~~For chlorophyll, data from the following cruises were used: ANT-XXIV/1, ANT-XXIV/4, ANT-XXVI/4 and MSM18/3 (doi.pangaea.de/10.1594/PANGAEA.847820), SO202/2 (doi.pangaea.de/10.1594/PANGAEA.820607), ANT-XXVII/2 (doi.pangaea.de/10.1594/PANGAEA.848590), ANT-XXV/1 (doi.pangaea.de/10.1594/PANGAEA.819099), ANT-XXVIII/3 and SO218 (doi.pangaea.de/10.1594/PANGAEA.848591), ANT23-1 (doi.org/10.1594/PANGAEA.871713), MSM9-1 (doi.org/10.1594/PANGAEA.873070), M91 (doi.org/10.1594/PANGAEA.864786), SO234+235 (doi.org/10.1594/PANGAEA.898929), SO243 (doi.org/10.1594/PANGAEA.898920), PS93.2 (doi.org/10.1594/PANGAEA.894872), HE462 (doi.org/10.1594/PANGAEA.899043), PS99.1 (doi.org/10.1594/PANGAEA.905502), PS99.2 (doi.org/10.1594/PANGAEA.894874), PS103 (doi.org/10.1594/PANGAEA.898941), PS107 (doi.org/10.1594/PANGAEA.894860), PS113 (doi.org/10.1594/PANGAEA.911061). Concerning remote sensing reflectances, the observations taken during cruises ANT-XXIV/4 and ANT-XXVI/4 (doi.pangaea.de/10.1594/PANGAEA.847820), ANT-XXV/1 (doi.pangaea.de/10.1594/PANGAEA.819099) and ARK26-3 (doi.pangaea.de/10.1594/PANGAEA.884528) were gathered. The remote sensing reflectances were corrected for the bidirectional nature of the light field (Morel and Gentili, 1996; Morel et al., 2002). The absorption coefficients were taken during cruises SO202/2 (doi.pangaea.de/10.1594/PANGAEA.820607), ANT_XXV/1 (doi.pangaea.de/10.1594/PANGAEA.819099), ANT-XXVI/3 and ANT-XXVIII/3 (doi.pangaea.de/10.1594/PANGAEA.819617), ARK26-3 (doi.org/10.1594/PANGAEA.885246), PS93.2~~

~~(doi.org/10.1594/PANGAEA.907605), PS99.2 (doi.org/10.1594/PANGAEA.907648) and PS107 (doi.org/10.1594/PANGAEA.907419). The compiled variables were "chla_hplc", "rrs" and "aph".~~

~~2.2~~2.3.13 Bermuda Atlantic Time-series Study (BATS)

BATS is a long-term study by the Bermuda Institute of Ocean Sciences based on regular cruises in the western Atlantic Ocean (Sargasso Sea) since 1988. The cruises at BATS site (~ 31º 40'N, 64º 10'W) sample ocean temperature and salinity, but are focused on biogeochemical variables such as nutrients, dissolved inorganic carbon, oxygen, HPLC of pigments, primary production and sediment trap flux. In this work all the phytoplankton pigment data available from the BATS website (http://bats.bios.edu/bats-data/) were considered, which also included regional and transect cruises not specific to the nominal BATS site. The compiled variables were "chla_hplc" and "chla_fluor".

~~2.2~~2.3.14 Data provided by Knut Yngve Børsheim (BARENTSSEA)

The BARENTSSEA data source refers to a group of observations that were provided to OC-CCI project by Knut Yngve Børsheim. This collection was developed using data from the archives of the Institute of Marine Research (Norway). It comprises observations of temperature, salinity and chlorophyll-a routinely collected by cruises, mainly in the North Sea, the Norwegian Sea and the Barents Sea between 1997 and 2013. The chlorophyll-a concentration was measured by filtering and extraction using Turner fluorometers. The compiled variable was "chla_fluor".

~~2.2~~2.3.15 The Fisheries and Oceans Canada database for biological and chemical data (BIOCHEM)

BioChem is an archive of marine biological and chemical data maintained by Fisheries and Oceans Canada (DFO, 2018; Devine et al., 2014). The available observations are from department research initiatives and collected in areas of Canadian interest. Available parameters include pH, nutrients, chlorophyll, dissolved oxygen and other plankton data (species and biomass). Chlorophyll measurements from in vitro fluorometric methods were acquired (from http://www.dfo-mpo.gc.ca/science/data-donnees/biochem/index-eng.html) with close guidance by the BioChem helpdesk, confirming quality and methods. The used data span from 1997 to 2014 and were mainly from the Gulf of St. Lawrence (western North Atlantic). The compiled variable was "chla_fluor".

~~2.2~~2.3.16 British Oceanographic Data Centre (BODC)

BODC is the designated marine science data centre for the United Kingdom. The data used in this work derive from a specific request for discrete observations of chlorophyll-a concentration since 1997. Initially, this request was used to compile AMT data (see section ~~2.2~~2.3.9). The remaining data comprising observations of chlorophyll-a concentration from fluorometric and HPLC methods, mostly sampled in the North Atlantic, were analysed and added (the "dataset" string for this data source is "bodc"). Data not flagged with highest quality or without method of measurement were discarded. The compiled variables were "chla_hplc" and "chla_fluor".

~~2.2~~2.3.17 California Cooperative Oceanic Fisheries Investigations (CALCOFI)

CalCOFI is a partnership of the California Department of Fish & Wildlife, National Oceanic & Atmospheric Administration Fisheries Service and Scripps Institution of Oceanography. CalCOFI has conducted quarterly cruises off southern and central California since 1949. Data collected in the upper 500 meters include: temperature, salinity, oxygen, nutrients, chlorophyll, primary productivity, plankton biodiversity, and biomass. For this work, only observations of chlorophyll-a concentration derived from fluorometric methods flagged with highest quality were used. Data were acquired from the file " CalCOFI_Database_194903-202105_csv_16October2023.zip~~CalCOFI_Database_194903-201911_csv_10Jul2020.zip~~". The compiled variable was "chla_fluor".

~~2.2~~2.3.18 California Current Ecosystem Long-Term Ecological Research (CCELTER)

CCELTER investigates the California Current coastal pelagic ecosystem, with a focus on long-term forcing. The CCELTER data includes primary and derived measurements from both Process and CalCOFI-augmented cruises, as well other time series. CCELTER data include variables from the physical environment, biogeochemistry and biological populations/communities. For this work chlorophyll observations measured from discrete bottle samples from CCELTER Process cruises determined by extraction and bench fluorometry (doi:10.6073/pasta/bbb278091dee3c96972087b7dee3673c) were used. The compiled variable was "chla_fluor".

~~2.2~~2.3.19 Center for Integrated Marine Technologies (CIMT)

CIMT was a non-operational program where marine scientists from different disciplines and institutions combined their efforts on observations directed towards understanding the central California upwelling system. The CIMT archived data includes coastal ocean observations from satellites, shipboard data, moorings and large marine animal movements. For this work, pigment data from discrete bottle samples taken during CIMT monthly cruises were used. Data were acquired from the project website (https://cimt.ucsc.edu/data_portal.htm). The compiled variable was "chla_fluor".

~~2.2~~2.3.20 CoastColour Round Robin (COASTCOLOUR)

COASTCOLOUR ~~data set~~datasets were designed to evaluate the performance of ocean colour satellite algorithms in the retrieval of water quality parameters in coastal waters (Nechad et al., 2015). Three types of COASTCOLOUR ~~data set~~datasets are available: 1) a match-up ~~data set~~dataset where *in situ* bio-optical observations are available simultaneously with a cloud-free MERIS product; 2) an *in situ* reflectance ~~data set~~dataset where an *in situ* reflectance is available simultaneously with an *in situ* measurement of chlorophyll-a concentration and/or total suspended matter; and 3) a simulated ~~data set~~dataset where reflectances were generated by a radiative transfer model. This work used the match-up ~~data set~~dataset, which includes most of the *in situ* measurements, and is available at https://doi.pangaea.de/10.1594/PANGAEA.841950 (Nechad et al., 2015). The match-up ~~data set~~dataset provides optical, biogeochemical and physical data collections at 17 sites across the globe. From this ~~data set~~dataset, observations of reflectance, chlorophyll-~~-~~a, total suspended matter and IOPs were compiled. The remote sensing reflectances were corrected for the bidirectional nature of the light field (Morel and

Gentili, 1996; Morel et al., 2002). The compiled variables were "rrs", "chla_hplc", "chla_fluor", "aph", "adg", "bbp" and "tsm".

~~2.2~~2.3.21 European Station for Time series in the Ocean, Canary Islands (ESTOC)

ESTOC is an open-ocean monitoring site located in the eastern North Atlantic subtropical gyre. ESTOC was initiated in 1991 with particle flux measurements, and in 1994 began standard observations of the water column, in addition to the deployment of a current meter mooring. The core parameters measured at ESTOC include salinity, temperature, current speed, nutrients, chlorophyll, inorganic carbon, particulate organic carbon and nitrogen, and sinking particle flux (Neuer et al., 2007). For this work measurements of chlorophyll~~-~~a concentration from monthly cruises from 1994 to 2011 were used. These data were provided to OC CCI following a specific request. The time of day was unavailable and was set to 12:00:00 (UTC). These observations were flagged with "1" in column "flag_time". The compiled variable was "chla_fluor".

~~2.2~~2.3.22 Australia's Integrated Marine Observing System (IMOS)

IMOS is enabled by Australia's National Collaborative Research Infrastructure Strategy (NCRIS) funded by the Australian Government. Since 2006, IMOS has been~~is~~ operating a wide range of observing equipment throughout the coastal and open ocean around Australia, making all data openly available to the scientific community, other stakeholders and users. In this work, the IMOS data contribution refers to two ~~data set~~datasets. One is a data collection entitled 'IMOS National Reference Station (NRS) - Phytoplankton HPLC Pigment Composition Analysis', which was acquired from the Australian Ocean Data Network (AODN) portal (https://portal.aodn.org.au). This ~~data set~~dataset comprises of phytoplankton pigment composition measured by HPLC collected with small vessels on monthly basis at nine National Reference Stations as part of the IMOS National Mooring Network. The other chlorophyll-a ~~data set~~dataset measured by HPLC and fluorometry methods, is a subset (2015-2021) of the IMOS Bio-optical Database also available through the AODN portal. This database comprises of a suite of bio-optical parameters from samples collected during research voyages in Australian waters and is used by the IMOS Ocean Colour Sub-Facility to assess the accuracy of satellite ocean colour products in Australian coastal and open ocean waters (Schroeder et al., 2016). The previous data compilations include an earlier subset of HPLC chlorophyll~~-~~-a concentration from the IMOS Bio-optical Database that was acquired through the SeaBASS archive. These data can be found under "dataset" string "seabass" and Lesley Clementson as data contributor. The compiled variables for IMOS were "chla_hplc" and "chla_fluor".

~~2.2~~2.3.23 MARineEcosytem DATa (MAREDAT)

MAREDAT database is a global assemblage of pigments measured by HPLC (Peloquin et al., 2013) from combination of 136 independent field ~~data set~~datasets, solicited from investigators and databases. The database provides high quality measurements of taxonomic pigments including chlorophylls a and b, 19’-butanoyloxyfucoxanthin, 19’-hexanoyloxyfucoxanthin, alloxanthin, divinyl chlorophyll~~-~~-a, fucoxanthin, lutein, peridinin, prasinoxanthin, violaxanthin and zeaxanthin. The database is available through PANGAEA (~~http://~~doi.pangaea.de/10.1594/PANGAEA.793246, Pitarch et al., 2013). For this work only measurements of Total ~~Chlorophyll~~ chlorophyll-a flagged with high quality were used. The time of day was unavailable and was set to 12:00:00 (UTC). These observations were flagged with "1" in column "flag_time". The compiled variable was "chla_hplc".

~~2.2~~2.3.24 Palmer Station Long-Term Ecological Research (PALMER)

PALMER is a monitoring station located in the western Antarctic Peninsula. The Palmer station investigates the marine ecology of the Southern Ocean with focus on the pelagic marine ecosystem, including sea ice habitats, regional oceanography and nesting sites of seabird predators. The PALMER data include measurements of meteorological, oceanographic, sea ice, predators, nutrients and biogeochemistry, pigments, primary production, zooplankton and microbes parameters. This work used the measurements of chlorophyll analysed by HPLC and~~—~~ fluorometry taken at the Palmer Station

(~~doi:~~doi.org/10.6073/pasta/09a1f2cc150b547e3c9b20c39e10cfc2 and ~~doi:~~doi.org/10.6073/pasta/6bbce1e3264571463c0354874dba88cf, Antarctica LTER et al., 2020a, b) and from the annual cruises off the coast of the Western Antarctica Peninsula (~~doi:~~doi.org/10.6073/pasta/4d583713667a0f52b9d2937a26d0d82e and doi.org/~~doi:~~10.6073/pasta/ec55e3d0d7260e1df98c9156f9becdeb, Palmer Station Antarctica LTER et al., 2018, 2020c). The compiled variables were "chla_hplc", "chla_fluor".

~~2.2~~2.3.25 SeaDataNet archive (SEADATANET)

SeaDataNet is a Pan-European infrastructure for ocean and marine data management. It aims to develop a standardised system for managing large and diverse ~~data set~~datasets collected by oceanographic cruises and automatic observation systems. For this work, discrete chlorophyll-a concentration observations with an "access restriction" set to "academic" and "unrestricted" were acquired from the SeaDataNet platform with guidance from the helpdesk. Only data from the "Institute of Marine Research - Norwegian Marine Data Centre (NMD), Norway", which comprised most of the acquired data, were used. All chlorophyll observations were from discrete samples measured by fluorometric, spectrophotometric or HPLC methods, but the exact method was not given. Thus, the observations were marked as "chla_fluor", although some were possibly from HPLC measurements, and were flagged with "1" in a column "flag_chla_method". The compiled variables were "chla_fluor".

~~2.2~~2.3.26 Data provided by Trevor Platt and Shubha Sathyendranath (TPSS)

In this work, the TPSS data source refers to a group of observations that were provided to this compilation by Trevor Platt and Shubha Sathyendranath. This is a collection of bio-optical *in situ* data collected during cruises predominantly in the Northwest Atlantic, but also from the Indian Ocean, South Pacific and Central Atlantic (see Sathyendranath et al. 2009 for additional details regarding the cruises). It comprises measurements of phytoplankton pigments and algal pigment absorption coefficients. The time of day was unavailable and was set to 12:00:00 (UTC). These observations were flagged with "1" in column "flag_time". The compiled variables were "chla_hplc", "chla_fluor" and "aph".~~sssss~~

~~2.2~~2.3.27 Bio-optical data from Tara expeditions (TARA)

The Tara expeditions consist of several cruises around the world, some with durations of several years, designed to study and understand the distribution of planktonic organisms in the world ocean. The discrete observations of remote sensing reflectance and chlorophyll-a concentration from HPLC measurements taken during the Tara "Oceans" (2009-2013) and "Mediterranean" (2014) expeditions were considered in this work. These data were provided to ESA OC-CCI project by Emmanuel Boss and were available in the SeaBASS archive. The remote sensing reflectances were corrected for the bidirectional nature of the light field (Park and Ruddick 2005)~~Morel and Gentili, 1996; Morel et al., 2002~~). The compiled variables were "chla_hplc" and "rrs".

2.3.28 National Research Council – Group of Satellite Oceanography (CNR_GOS)

The CNR_GOS dataset consists of bio-optical observations collected in the Mediterranean Sea between 1999 and 2024 through multiple research cruises conducted by the Group of Satellite Oceanography (GOS) of the Italian National Research Council (CNR). Measurements are station-based and include remote-sensing reflectance and diffuse attenuation coefficient derived from in-water radiometric profiles, as well as chlorophyll-a concentration and, more sporadically, inherent optical properties and total suspended matter. Radiometric measurements were acquired using free-falling profilers following community protocols, providing high-quality data suitable for satellite validation. Chlorophyll-a was determined initially by spectrofluorometry and, since 2012, by HPLC. The compiled variables were "rrs", "kd", "chla_hplc", "tsm", "aph", "adg", and "bbp".

2.3.29 Department of Fisheries and Oceans Canada (DFO)

The DFO measurements, provided by Emmanuel Devred, were collected in the Northwest Atlantic Ocean between 2000 and 2022, primarily within long-term monitoring programs such as the Atlantic Zone

Monitoring Program (AZMP) and Atlantic Zone Off-Shelf Monitoring Program (AZOMP). These observations cover regions including the Scotian Shelf and the Labrador Sea, encompassing a wide range of oceanographic conditions from coastal to open-ocean environments. The dataset includes chla_hplc, chla_fluor, and phytoplankton absorption (aph), derived from water samples collected during multiple research cruises over more than two decades. Phytoplankton absorption coefficients were measured for a subset of stations and spectrally resolved following standard laboratory procedures.~~Compiled variables?~~

2.3.30 Grupo de Oceanografia de Altas Latitudes (GOAL)

Measurements from the GOAL program (Federal University of Rio Grande, Brazil) were acquired during austral summer expeditions to the northern Antarctic Peninsula between 2008 and 2020 (excluding 2011–2012). The dataset consists of "chla_hplc" measurements collected across multiple oceanographic campaigns aboard Brazilian research vessels. Seawater samples were filtered onto GF/F filters and analysed using HPLC following a C8 reverse-phase method with a pyridine-containing mobile phase. Pigment identification and quantification were based on calibrated standards and photodiode array detection, with quality control procedures applied to ensure reliability of low-concentration measurements.

2.3.31 Innovation in the Framework of the Atlantic Deep Ocean (iFADO)

The iFADO project provides chlorophyll-a measurements derived from High-Performance Liquid Chromatography (HPLC), following standardized pigment analysis protocols (Zapata et al., 2000; Mendes et al., 2007). Total chlorophyll-a (TChla) wa~~i~~s computed as the sum of chlorophyll-a, its derivatives, and divinyl chlorophyll-a, consistent with current recommendations for satellite validation (Canuti et al., 2025). The data were collected in coastal and shelf waters along the Spanish and Portuguese Atlantic coasts.

2.3.32 Instituto Nacional de Investigación y Desarrollo Pesquero (INIDEP)

Measurements provided by IOPAN were collected in the Baltic Sea between 2013 and 2020. The dataset includes many coincident "chla_hplc" and "chla_fluor" measurements from multiple field campaigns. conducted in different seasons. The Baltic Sea is an optically complex coastal environment characterized by strong gradients in phytoplankton biomass, coloured dissolved organic matter, and suspended materials.

Methodology of "chla_hplc" measurements was based on an extraction of pigments from phytoplankton cells using 90% acetone solution by mechanical grinding and sonication during 2 hours in darkness at 4 ºC (Parsons et al. 1984). The pigments separation, qualification, and quantification was performed according to methodology presented in Stoń-Egiert and Kosakowska 2005. The chromatographic system was equipped with a diode array absorbance detector, a fluorescence detector, and a C18 analytical column.

Spectrophotometric measurements of chlorophyll-a concentration were based on pigment extraction in 100% ethanol during 24 hours in darkness. Absorbance in the spectrophotometric reading was at 665 nm, corrected by cell-to-cell blank and absorbance at 750 nm (Lorenzen, 1967: Guideline for Baltic Monitoring Program, 1988).

2.3.33 Institute of Oceanology, Polish Academy of Sciences (IOPAN)

Measurements provided by IOPAN were collected in the Baltic Sea between 2013 and 2020. The dataset includes many coincident "chla_hplc" and "chla_fluor" measurements from multiple field campaigns. The Baltic Sea is an optically complex coastal environment characterized by strong gradients in phytoplankton biomass, coloured dissolved organic matter, and suspended materials.

2.3.34 Institute for Space-Earth Environmental Research (ISEE)

Measurements provided by the Institute for Space-Earth Environmental Research (ISEE) were collected in coastal waters of Japan, primarily in Ariake Sea and Ise–Mikawa Bay. The dataset spans nearly two decades (2001–2019), providing one of the longest continuous *in situ* records in optically complex coastal environments. The observations include chla_fluor, rrs, and inherent optical properties ("aph", "adg", "bbp"), with substantial temporal overlap between radiometric and biogeochemical measurements. The concurrent availability of "rrs" and IOPs makes this dataset particularly valuable for the development and validation of ocean colour algorithms, as well as for satellite product evaluation in turbid and productive waters.

2.3.35 Korea Hydrographic and Oceanographic Agency (KHOA)

Field measurements provided by KHOA originate from dedicated surveys carried out between 2022 and 2025 in the regional seas surrounding the Korean Peninsula, including the Yellow Sea, East Sea, Korea Strait, and East China Sea. The dataset comprises chlorophyll-a concentrations derived from both HPLC and fluorometric methods, together with "tsm".

2.3.36 Mediterranean Global Ocean Observing System (MEDGOOS)

The MedGOOS dataset includes bio-optical observations collected during cruises in the Mediterranean Sea between 2004 and 2008, the University of Florence aboard the CNR's R/V Urania. Chlorophyll-a concentrations were determined from seawater samples collected at approximately 5 m depth and analysed by HPLC following standard protocols, with total chlorophyll-a defined as the sum of chlorophyll-a and divinyl-chlorophyll-a. Radiometric measurements of downwelling irradiance were acquired using a calibrated spectroradiometer under stable sky conditions, and spectral diffuse attenuation coefficients (Kd) were derived from irradiance profiles following NASA ocean optics protocols. Quality control procedures included spectral smoothing and comparison with expected ranges based on optical models and previous studies.

2.3.37 University of California San Diego (UCSD)

Measurements provided by UCSD were collected during oceanographic campaigns in the Pacific Ocean in 2015 and 2022, including the OUTPACE and SOKOWASA cruises. The dataset comprises "chla_hplc" and spectral "rrs" observations acquired at discrete stations across oligotrophic and mesotrophic waters. "rrs" measurements were obtained using HyperPro profiling radiometers following standard protocols, with multiple casts performed at each station and processed using ProSoft software-. Chlorophyll-a and "rrs" measurements were generally collected at the same stations, although not always simultaneously, with small temporal and spatial offsets due to ship drift and sequential sampling.

2.3.38 University of New Hampshire (UNH)

Measurements provided by the UNH were collected in the Great Bay Estuary (USA) between 2022 and 2024, a coastal system characterized by highly turbid and optically complex waters influenced by strong absorption and scattering processes. The dataset includes rrs, chla_hplc, tsm, and inherent optical properties ("aph" and adg). Remote-sensing reflectance was measured using RAMSES radiometers deployed near the water surface, with multiple casts per station used to derive mean values and associated uncertainties. Absorption coefficients were determined following IOCCG protocols using laboratory spectrophotometric methods with integrating sphere measurements. Chlorophyll-a concentrations were derived from HPLC pigment analysis, and suspended matter was obtained from filtered water samples. The dataset provides a consistent set of coincident optical and biogeochemical measurements.

2.3.39 Yokohama National University (YNU)

Measurements provided by Yokohama National University were obtained from coastal and inland water environments in Japan, including Tokyo Bay, the Seto Inland Sea, and Lake Kasumigaura. These water bodies are characterized by optically complex conditions influenced by high phytoplankton biomass, suspended sediments, and coloured dissolved organic matter. The observations span the period 2010–2021, covering a wide range of optical conditions typical of productive waters. The dataset includes chlorophyll-a concentrations derived from both HPLC and fluorometric methods, as well as spectral "rrs" and inherent optical properties ("aph", adg, and bbp). Observations were collected during field campaigns partially funded by the Japan Aerospace Exploration Agency (JAXA).

3 Results

3.1 Database overview and updates~~Summary~~

In this work several sets of bio-optical *in situ* data were acquired, homogenised and merged into a single unified ~~data set~~dataset. The ~~data set~~dataset comprises *in situ* observations between ~~1997~~ 1996 and ~~2021~~2025, with a global distribution, and includes the following variables: "rrs", "chla", "aph", "adg", "bbp", "kd" and "tsm". All observations were processed in such a way that they can be compared directly with satellite-derived ocean ~~-~~colour data. The compiled ~~data set~~dataset corresponds to a table with a total of 319,183~~151,673~~ rows~~ and 3,458 columns~~. Each row represents a unique station in space and time, separated from the rest by at least 5 minutes and 200 meters. For each variable at a given station, three metadata strings are provided: "dataset", "subdataset" and "contributor". The columns of the table take the form described in Table 2~~1~~. The data contributors are indicated in Table 3~~2~~. Regarding spectral variables, all original wavelengths were preserved, which required many unique wavelengths to be maintained in the database. No band shifting was performed (though some archived data in some data sources may have been merged with nearby wavelengths) and no minimum number of wavelengths per observation was imposed. This allowed further manipulation of the ~~data set~~dataset for different purposes.

Compared to the previous release (Valente et al., 2022), the present version extends the temporal coverage by more than four years (2022 → 2025, Table 1) and increases the overall volume of observations across all variables by ~115 % (from 148,432 to 319,183 records). Variable-specific increases are as follows: "chla" +31% (from 82,543 to 107,922), "rrs" +94 % (from 68,641 to 133,325), "aph" +150 % (from 4,265 to 10,655), "adg" +78 % (from 1,654 to 2,948), "bbp" +10,258% (from 792 to 82,033), "kd" +30 % (from

2,454 to 3,197), and "tsm" +51 % (from 1,546 to 2,334). In addition to extending the temporal record, the new contributions improve the spatial distribution of the compilation, particularly in regions that were previously sparsely represented, including waters surrounding Japan and Korea, the Mediterranean Sea, and the Southern Ocean.

In the following paragraphs, the final group of observations is described in terms of each variable and the corresponding contributing ~~data set~~datasets; however, it is important to note that the numbers reported here do not reflect the original numbers in each contributing ~~data set~~dataset, since observations close in time and space were averaged and quality controls were applied. Furthermore, duplicates across contributing ~~data set~~datasets were removed (e.g., NOMAD and others, such as MOBY, were removed from MERMAID; also, data of individual projects, such as PALMER and AMT, can be listed under NOMAD). Nevertheless, the reported numbers still give a general view of the contributions from each ~~data set~~dataset and provide~~s~~ users with valuable information for analysing each set of data separately.

3.2 Distribution and characteristics of remote sensing reflectance~~Rrs~~

Observations of remote-sensing reflectance are available at ~~948~~ 3,008 unique wavelengths (i.e., columns), between 301.22 ~~313~~ nm and 1022.1 nm (Fig. 1). In total there are 133,325 ~~68,641~~ observations (i.e., rows) of remote-sensing reflectance. The total number of observations are partitioned per contributing ~~data set~~datasets as follows: AERONET-OC (96,538~~34,551~~), BOUSSOLE (22,620), MOBY (6,959~~6,034~~), NOMAD (3,326), MERMAID (895), SeaBASS (3,871~~730~~), AWI (71), COASTCOLOUR (307), ~~and~~ TARA (107), cnr_gos (624), Isee (568), Ucsd (70), Unh (45) and Ynu (173). Data from AERONET-OC, BOUSSOLE and MOBY correspond to continuous time series, and, hence, ~~the~~ provide higher number of observations. In comparison with the previous version (Valente et al., ~~2019~~2022), which had reflectance data until 2018, the number of stations ~~increased~~ increases by ~94 ~~15~~% (i.e., from 68,641~~59,781~~ to 133,325~~68,641~~). The new data points are mainly from recent years (2019-~~2021~~2025) and from updates of AERONET-OC, BOUSSOLE, MOBY, MERMAID, ~~and~~ AWI, CNR_GOS, ISEE, UCSD, UHN, and

YNU. The new data extend~~ed~~ the temporal coverage towards more recent years, but the statistical distribution of values and the spatial coverages (discussed below) remain ~~have essentially remained~~ the same when compared to the previous version (Valente et al., ~~2019~~2022). This is ~~explained by~~because most of the new observations ~~coming~~ are from continuous time series at fixed ~~the~~ locations (AERONET-OC, BOUSSOLE, and MOBY Gold and Platinum).~~–~~

The magnitude distribution of the remote sensing reflectance~~s~~ at 44X nm and 55X nm is provided in Fig. 2a and b, respectively. Data were selected within a ±8 nm window around~~first searched at~~ 445 and 555 nm~~, and then with a search window up to 8 nm, to include also data at 547 nm~~. Median values at 44X nm ranged from 0.0008~~0.003~~ ~~m~~sr^{-1} (YNU~~AERONET-OC~~) to~~and~~ 0.009 ~~m~~sr^{-1} (MOBY), whereas at 55X nm the median values lie between 0.0013 ~~m~~sr^{-1} (AWI) to~~and~~ 0.0109~~0.007~~ sr~~m~~$^{-1}$ (COASTCOLOUR). The observations ~~remain~~ are unevenly distributed ~~between each~~by month of the year in both hemispheres, with the summer months having higher data representation ~~–~~(Figure 3). The Northern Hemisphere has also more data than the Southern Hemisphere (Fig. 3). As a quality control indicator, reflectance band ratios were plotted against each other (490:555 versus 412:443, Fig. 4). Most points are within the boundaries of the NOMAD ~~data set~~dataset, but some scattered points ~~were~~ are found. These points were retained to allow further manipulation with different quality control criteria. The geographic distribution of the remote-sensing reflectance stations (Fig. 5) still shows a higher number of observations in some coastal regions, such as those of North America, ~~and~~ Northern Europe, Mediterranean Sea and East Asia. Away from continental margins, the Atlantic Ocean has the highest density of observations. The b~~B~~est geographic coverage is provided by the NOMAD database. Data from SeaBASS ~~is~~ are also well dispersed in space~~ but fewer in number~~. Data from MERMAID are mainly located along the coasts of Europe, North America, and the central region of the North Atlantic Ocean. The observations from AERONET-OC, BOUSSOLE, COASTCOLOUR and MOBY are ~~concentrated~~ from ~~in~~ specific sites around the world, while AWI data are available for the Atlantic and Arctic Oceans. TARA data are spread across several regions, with highest data density in the Mediterranean Sea. Additional regional datasets further enhance the spatial coverage, including CNR_GOS in the Mediterranean Sea, ISEE, YNU and KHOA around Japan and Korean Peninsula.

3.3 Chlorophyll-a characteristics

~~Chla~~

Observations of chlorophyll-a concentration were divided into those measured by fluorometric or spectrophotometric methods ("chla_fluor"), and HPLC methods ("chla_hplc"):~~. A~~ comparison of the two ~~types of measurements~~ when available at the same station (Fig. 6), shows good agreement (Trees et al., 1985). No data were filtered for this analysis and the good correlation can be explained, in part, by the quality control measures implemented by the data providers and curators of repositories such as NOMAD and SeaBASS (Werdell and Bailey, 2005). The total number of stations with concurrent observations of "chla_fluor" and "chla_hplc" is 10,200 ~~5,953~~, with contributions from SeaBASS (24.9 ~~39~~%), DFO (24.7 %), IOPAN (10.1 %), TPSS (9.3 ~~16~~%), PALMER (~~14~~8.5 %), NOMAD (6.3 ~~11~~%), BATS (5.5 %), COASTCOLOUR (2.6 4%), MERMAID (2.1 4%), HOT (2.3 4%), AMT + GeP&CO + BODC + BIOCHEM ~~+ CCELTER + CALCOFI~~+ IMOS + KHOA (3.7 %). The "chla_fluor" observations are available in 80,790~~61,317~~ stations (rows), with values ~~limited to the~~ range between 0.001 to 309.2~~100~~ mg m^{-3} (Fig. 7). They are from NOMAD (2,350), SeaBASS (19,482~~18,575~~), MERMAID (480), ICES (5,421), HOT (974~~755~~), AMT (396), ARCSSPP (189), BARENTSSEA (7,188), BATS (657~~356~~), BIOCHEM (11,527~~4,592~~), BODC (895), CALCOFI (7,074~~5,396~~), COASTCOLOUR (3,322), CCELTER (468), CIMT (204.), ESTOC (100), GEPCO (56), IMOS (4,424~~1136~~), PALMER (3,237), SEADATANET (5,403), ~~and~~ TPSS (1000), CNR_GOS(144), DFO (2,944), INIDEP(918), IOPAN(1,220), ISEE (458), KHOA (396) and YNU (78).~~–~~ The total number of "chla_hplc" observations is 37,909 ~~27,215~~, ranging from 0.002 to 383.6~~99.8~~ mg m^{-3} (Fig. 7), with contributions from NOMAD (1,309), SeaBASS (12,489~~10,257~~), MERMAID (707), ICES (2,994), HOT (248~~222~~), GeP&CO (1,536), BOUSSOLE (577), AMT (1,359), AWI (2,723~~2,343~~), BATS (334), BODC (735), COASTCOLOUR (848), IMOS (2,231~~340~~), MAREDAT (1,024), PALMER (1,525), TPSS (1,002), ~~and~~ TARA (161),

CNR_GOS (349), DFO (3,024), GOAL (607), iFADO (292), IOPAN (1,177), KHOA (119), MEDGOOS (53), UCSD (35), UNH (44), and YNU (36). Compared to ~~the previous version (~~Valente et al., (~~2019~~2022), the "chla_hplc" observations increased by ~~~16~~39% (~~23,550~~27,215 to 37,909~~27,215~~). ~~As for the~~ The "chla_fluor" observations~~, they have decreased~~ increased by ~32% (from 61,317 ~~61,525~~ to 80,790~~61,317~~)~~, which is explained by the added observations (N=3033) being less than the removed stations due to quality control (N=3241; see section 2.2.6)~~. The new data points come from updates of ~~BOUSSOLE, MERMAID,~~ SeaBASS, HOT, BATS, BIOCHEM, ~~AMT, PALMER, CCELTER,~~ CALCOFI, AWI and IMOS, as well as adding new dataset of CNR_GOS, DFO, GOAL, iFADO, INIDEP, IOPAN, ISEE, KHOA, MEDGOOS, UCSD, UNH, and YNU.

The combined chlorophyll ~~data set~~dataset (all chlorophyll data considered, but for a given station, HPLC data were selected if available), has a total of 107,922~~82,543~~ observations which represents a major ~~n~~ increase of ~31 ~~4~~% (i.e., from 82,543~~79,731~~ to 107,922~~82,543~~) when compared to ~~the previous version (~~Valente et al.~~,~~ (~~2019~~2022). ~~The present version represents a major increase in the number of recent observations. For the combined chlorophyll data set, 533 stations were available in previous version for the period 2016-2017 (previous version had chlorophyll data until 2017). Now, there are 5,140 stations for the period 2016-2021, which represents an increase of ~964 % for the period of 2016 onwards.~~ Overall, data distribution and spatial coverage remain the same between present and previous versions. Approximately 10 %, 50% and 40 % of observations are from oligotrophic (<0.1 mg m^{-3}), mesotrophic (0.1 - 1 mg m^{-3}), and eutrophic (>1 mg m^{-3}) waters, respectively. When compared with the proportions of the world ocean in these trophic classes, 56 % oligotrophic, 42 % mesotrophic and 2 % eutrophic (Antoine et al., 1996), oligotrophic waters are still under-represented relative to eutrophic waters in the compilation. The combined chlorophyll ~~data set~~dataset is also still unevenly distributed geographically, with higher coverage in the Northern Hemisphere (Fig. 3). The spatial distribution of the chlorophyll values for the combined ~~data set~~dataset (Fig. 8) shows a good agreement with known biogeographical features, such as lower chlorophyll values in the subtropical gyres, and higher values in temperate, coastal and upwelling regions. Many regions show a good spatial coverage (e.g., Atlantic and Pacific Ocean), while others are less well sampled (e.g., Southern and Indian Oceans). ~~Of~~ Among the contributing ~~data set~~datasets, SeaBASS provides one of the most extensive global spatial ~~coverage~~coverages (Fig. 9). Other ~~data set~~datasets also provide broad coverage across multiple ocean basins~~from several locations across the globe~~ (NOMAD, GEPCO, MAREDAT, TARA). The ICES, MERMAID and BODC data are mainly located along the coastal regions of Europe. The AMT ~~and many AWI data mostly~~ cover the Atlantic Ocean and~~.~~ AWI observations are more regionally concentrated, mainly in the North Atlantic and European seas, with some additional measurements in the Southern Ocean.~~Other AWI data cover the Amundsen to Bellinghausen Sea of the Southern Ocean, the North Sea, the Arctic Ocean, the Indian Ocean and the subtropical and tropical Pacific.~~ Coverage for the Arctic region and northern seas of the North Atlantic is provided by SEADATANET, ARCSSPP and BARENTSEA ~~data set~~datasets. Observations from BIOCHEM and TPSS are mostly from the Northwest Atlantic, whereas CALCOFI, CCELTER and CIMT provide data for the western coast of North America, together with DFO and iFADO datasets, which provide extensive coverage in the Northwest and Northeast Atlantic, respectively. The data from IMOS mainly cover~~s~~ the coastal Australian waters. In the Mediterranean Sea, coverage is strengthened by the addition of CNR_GOS, and MEDGOOS datasets. IOPAN contributes to the spatial coverage in the Baltic and northern European seas. ISEE and YNU provide observations in optically complex coastal and inland waters of Japan, while KHOA contributes data in the regional seas surrounding the Korean Peninsula. The remaining ~~data set~~datasets provide observations for fixed locations: PALMER (western Antarctic peninsula), COASTCOLOUR (17 coastal sites across the world), BATS (Bermuda, North Atlantic), BOUSSOLE (Mediterranean), HOT (Hawaii, North Pacific), ESTOC (Canaries, North Atlantic). Figure 9 shows all data sources that contribute with chlorophyll observations, but many overlap each other, especially around Europe and North America. For additional analysis and as an example of the applications of the compiled dataset, the combined chlorophyll data ("chla_fluor" and "chla_hplc") were partitioned into 5º x 5º boxes and for each box the number of observations, average value and standard deviation were computed (Fig. 10 a, b and c, respectively). The number of observations can be very high (>1000) in some boxes along the European and North American coastlines and relatively low (<20) in oceanic regions. Compared with previous version, coverage has improved in several regions, including the Mediterranean Sea, North Pacific coastal areas, East Asia, and parts of the Southern Ocean. The well-known global biogeographical features, such as the lower chlorophyll in the subtropical gyres and higher values in coastal and upwelling areas, clearly emerge in the average value map (Fig. 10 b). There is a close correspondence between the spatial patterns of the average and standard deviation maps (Fig. 10 b and c), which may be an indicator of the data quality.

~~Coincident Rrs and Chla~~3.4 Coincident rrs and chla

~~Coincident~~ Concurrent observations of chlorophyll-a concentration and remote-sensing reflectance are available at 4,523 ~~3,645~~ stations. These observations are ~~mostly~~ mainly from NOMAD (~~80~~ 64.5 %), CNR_GOS (10.7 %) MERMAID (~~9~~ 7.5 %), ISEE (5.5 %), COASTCOLOUR (~~6~~5 %), and SeaBASS (~~3~~ 2.7 %). The maximum of three selected band ratios of remote-sensing reflectance is plotted against chlorophyll-a concentration (Fig. 11). The "chla" values used are the combined HPLC and fluorometric chlorophyll-a and for the "rrs", the closest spectral observation within 2 nm was used. The maximum band ratios were calculated as the maximum of [rrs(443)/rrs(555), rrs(490)/rrs(555), rrs(510)/rrs(555)] or [rrs(443)/rrs(560), rrs(490)/rrs(560), rrs(510)/rrs(560)] if rrs(555) was not available. The relationship between maximum band ratio and chlorophyll is generally consistent with ~~close to~~ the NASA OC4 and OC4E (OC4 for MERIS) v6 standard algorithm (https://www.earthdata.nasa.gov/apt/documents/chlor-a/v1.0~~http://oceancolor.gsfc.nasa.gov/cms/atbd/chlor_a~~)~~.~~. This agreement is expected, as these algorithms were calibrated using datasets that include a substantial contribution from NOMAD, which is also well represented in the present compilation. For comparison, the regional MedOC4 algorithm (Volpe et al., 2007, 2012, 2019), developed for Mediterranean waters, is also shown in Fig. 11, illustrating deviations associated with region-specific optical properties (e.g., the CNR_GOS dataset).~~similarly based on maximum band ratios, providing confidence in the quality of the compiled data. Compared to the previous~~

~~version (Valente et al., 2019), the relations between maximum band ratio and chlorophyll are not altered by the additional number of concurrent observations (N=13).~~

3.5 Inherent optical properties characteristics

~~IOPs~~

The inherent optical properties ("aph", "adg" and "bbp") show different spectral coverage, where aph comprises 10,655 observations at 3,134 unique wavelengths between 290 and 850 nm, adg comprises 2,948 observations at 1,035 unique wavelengths between 300 and 850 nm, and bbp comprises 82,033 observations at 77 unique wavelengths between 365 and 730 nm. Compared with the previous version, the IOP data have been expanded through updates to two existing data sources (AWI and SeaBASS) and the inclusion of five new sources (CNR_GOS, DFO, ISEE, UNH, and YNU). ~~are available at 550 unique wavelengths between 300 and 850 nm. There is a total of 4,265, 1,654 and 792 observations, for "aph", "adg" and "bbp", respectively.~~ For "aph" the total number of observations is distributed among NOMAD (1,190), TPSS (966), COASTCOLOUR (593), AWI (2,263~~991~~), SeaBASS (2,264~~453~~), ~~and~~ MERMAID (72), CNR_GOS(28), DFO (2747), INIDEP(249), ISEE (164), UNH (45) and YNU (105). For "adg" the contributions are as follows: NOMAD (1,079), COASTCOLOUR (531), SeaBASS (11) and MERMAID (33). The "bbp" observations come from NOMAD (371), COASTCOLOUR (154), SeaBASS (81,105 ~~32~~), ~~and~~ MERMAID (235), CNR_GOS (85), ISEE (35), and YNU (48). ~~Compared to previous version (Valente et al., 2019), only "aph" was updated, resulting in a ~30 % increase (i.e., from 3,293 to 4,265). Most of the new observations fall within the period 2012-2020, thus increasing the temporal coverage (previous version had "aph" until 2012).~~ Data distribution of "aph", "adg" and "bbp" at 44X nm and 55X nm for each ~~data set~~dataset is provided in Fig. 12 a-f. Median values of "aph", "adg" and "bbp" at 44X and 55X nm for each ~~data set~~dataset are summarized in Table 5~~3~~. As a quality indicator, the following band ratios for the absorption coefficients were calculated: aph(490)/aph(443), aph(412)/aph(443), adg(443)/adg(490) and adg(412)/adg(443). Data within 2 nm of the wavelengths were used to maximize the number of points. The distribution of the ratios is shown in Fig. 13. Several observations were found to be outside the thresholds used in the International Ocean-Colour Coordinating Group (IOCCG) report 5 for quality control (IOCCG, 2006; see dotted vertical black lines in Fig. 13). These points are highlighted here for information, but retained in the database, since these were mostly from NOMAD and there was an interest to preserve this ~~data set~~dataset as a whole. Also, not discarding these data allows further manipulation with different quality control criteria. On the annual scale, the observations of ~~the inherent optical properties~~IOPs continue to be strongly underrepresented in the Southern Hemisphere where there is a complete absence of data during the austral winter (Fig. 3). The new "aph" data in the present version have ~~primarily~~ increased primarily the spatial coverage in the Arctic region, with additional contributions in several coastal and shelf regions~~only increased the spatial coverage in the Arctic region~~. Overall, the geographic coverage for observations of "aph", "adg" and "bbp" (Fig. 14) remains sparsely sampled~~is poor~~, with most open ocean regions not being sampled, except for the Atlantic Ocean. Small clusters of data are in specific coastal regions, such as the western coast of North America, with improved coverage also evident in regions such as the Mediterranean Sea, northern European waters, and parts of the Northwest Atlantic. The number of bbp observations has increased substantially from 792 to 82,033 samples, mainly along transects in the Pacific Ocean.

3.6 Diffuse attenuation coefficient characteristics

~~Kd~~

~~Finally, for t~~The samples of diffuse attenuation coefficient for downward irradiance ("kd"~~, not updated in present version~~) increase by ~30 % (from 2,454 to 3,197) there are 652 ~~25~~ unique wavelengths between 313~~405~~ and 952~~709~~ nm. The total of 3,197 ~~2,454~~ observations ~~is divided between~~sourced from 6 datasets NOMAD (2,266), SeaBASS (~~118~~193), ~~and~~ MERMAID (70), AWI (2), CNR_GOS (627), and MEDGOOS (40), with the increase in this version arising from updates to AWI and SeaBASS and the addition of CNR_GOS and MEDGOOS. Data distribution of "kd" at 44X nm and 55X nm for each ~~data set~~dataset is shown in Fig. 12g and 12h. ~~No "kd" data at these wavelengths were available for the~~

~~SeaBASS data set (only at 490 nm).~~ Median values of "kd" at 44X nm and 55X are su~~a~~mmaris~~z~~ed ~~at~~in Table 5.~~span between 0.08 m^{-1} (NOMAD) and 0.1 m^{-1} (MERMAID), whereas at 55X nm the "kd" values are approximately 0.1 m^{-1} (NOMAD and MERMAID).~~ The best geographical coverage is provided by NOMAD (Fig. 15), with a higher coverage in the Atlantic, compared with other oceans. Except for the coastal regions of North America, Mediterranean Sea and the Japan Sea, most coastal regions are not sampled. In the Northern Hemisphere, "kd" is distributed evenly across all months of the year, but in the Southern Hemisphere there are few data points during the austral winter (Fig. 3).

3.7 Total suspended matter characteristics

~~TSM~~

For ~~total suspended matter (~~"tsm"~~; not updated on present version),~~ ~~there is~~ a total of 2,334~~1546~~ observations ~~are~~is ~~a~~available, with the largest contribution from ~~divided between~~ COASTCOLOUR (1199), followed by KHOA (397), ~~and~~ MERMAID (347) and SeaBASS (295). ~~The~~ Most observations of "tsm" are ~~available in a greater number~~ in the Northern Hemisphere (Fig. 3)~~) and are~~ distributed across several coastal regions around Europe, Mediterranean Sea, China Sea, Indonesia and Australia (Fig. 15). The increase in observations relative to previous versions is mainly associated with contributions from SeaBASS, CNR_GOS, KHOA, and UNH.

3.8 Bio-optical relationships

Although most of the stations with concurrent variables are from the NOMAD ~~data set~~dataset, for completeness, an examination of bio-optical relationships is provided (Fig. 16). The relation between "aph" at 443 nm and chlorophyll-a (Fig. 16 a) ~~agrees with Bricaud et al. (2004).~~ follows the general power-law behaviour described by Bricaud et al. (1995, 2004), with observations generally closer to the Bricaud et al. (1995) relationship. The Bricaud et al. (2004) formulation tends to overestimate "aph", particularly at moderate to high chlorophyll concentrations.

A total of 8,583~~3,387~~ points exists with these two variables available (distributed among multiple datasets including DFO, AWI, SeaBASS, NOMAD, TPSS, COASTCOLOUR and others~~29 % from NOMAD, 28 % from TPSS, 22 % from AWI, 10% from COASTCOLOUR and remaining 11 % from MERMAID and SeaBASS~~). The relation between the sum of "aph" and "adg" at 443 nm and "rrs" at 443 nm (Fig. 16 b), shows a dispersion similar, except for some scattered points, to an equivalent analysis on the IOCCG report 5 (IOCCG, 2006; see their Fig. 2.3). Again, the scattered data were retained in the final table to preserve the NOMAD ~~data set~~dataset. A total of 1268~~1,112~~ points exists for which these three variables are available (85~~97~~ % from NOMAD). The relation between the ratio rrs(490)/rrs(555) and kd(490) (Fig. 16c) shows a good agreement with the NASA KD2S standard algorithm (https://oceancolor.gsfc.nasa.gov/resources/atbd/kd/ ~~http://oceancolor.gsfc.nasa.gov/cms/atbd/kd_490~~). A total of 2,282~~2,280~~ points exists for which these three variables are available (93 ~~-~~% from NOMAD). The relation between the ratio rrs(490)/rrs(555) and "bbp" at 555 nm (Fig. 16 c) shows a good agreement with the relation suggested by Tiwari and Shanmugam (2013). A total of 498~~365~~ points exists for which these three variables are available (65~~89~~ % from NOMAD).

4 Summary and conclusions

In this work, a compilation of bio-optical *in situ* data is presented, resulting from the acquisition, homogenization and unification of several sets of ~~data~~ measurements obtained from different sources. The compiled data have a global coverage and span the period from ~~1997~~ 1996 to ~~2021~~2025, encompassing ~~which corresponds to~~ the period of the~~a~~ continuous satellite ocean ~~-~~colour ~~data~~ observations ~~record~~since September 1997. Minimal changes were made on the original data, other than conversion to standard format, data ~~reductions~~ filtering in time and space, and quality control. *In situ* measurements of the following variables were compiled: remote-sensing reflectance, chlorophyll-a concentration, algal pigment absorption coefficient, detrital and coloured dissolved organic matter absorption coefficient, particle backscattering coefficient, diffuse attenuation coefficient for downward irradiance and total suspended matter.

The final set of data consists of a substantial number of *in situ* observations, available in a simple text format, and processed in a way that ~~are~~ can be used directly for the evaluation of satellite-derived ~~ocean-colour~~ data. The major advantages of this compilation are that it merges six commonly-used data sources in ocean ~~-~~colour validation (MOBY, BOUSSOLE, AERONET-OC, SeaBASS, NOMAD, MERMAID), four data sources developed for ocean ~~-~~colour applications (AWI, COASTCOLOUR, TPSS and TARA) and 29~~17~~ additional ~~set~~datasets ~~of chlorophyll-a concentration data~~ (AMT, ICES, HOT, GeP&CO, ARCSSPP, BARENTSSEA, BATS, BIOCHEM, BODC, CALCOFI, CCELTER, CIMT, ESTOC, IMOS, MAREDAT, PALMER, SEADATANET, CNR_GOS, DFO, GOAL, iFADO, INIDEP, IOPAN, ISEE, KHOA, MEDGOOS, UCSD, UNH, YNU) free of duplicated observations, which will be invaluable ~~-and many of which may not have been used previously~~ for satellite comparison and ~~/~~validation. This ~~data set~~dataset was initially created with the intention of evaluating the quality of the satellite ocean ~~-~~colour products from the ESA OC-CCI project, but it can also be used for other purposes, including the validation of retrievals from recent satellite missions such as ~~Landsat 8 and Sentinel 2. It may also be useful in the preparation of future sensors like~~ NASA PACE and VIIRS-NOAA21. In addition, it is likely one of the largest collections of bio-optical observations (including chlorophyll-a, inherent optical properties, and radiometric measurements) ~~chlorophyll-a concentrations~~ ever assembled, making it useful for the climate and biological scientific communities. The objective of publishing the compilation is to make it easily accessible by the broader community.

In comparison with previous versions, the main advantage of the present version (~~version 3~~v2026) is that it includes more recent data (especially from ~~2016~~ 2022 onwards). These new ~~data~~ measurements are key for the validation of the most recent ocean ~~-~~colour missions ~~(e.g., Sentinel 2B and Sentinel 3B)~~ and for other activities such as System Vicarious Calibration which is undertaken using an entire mission dataset. Future improvements of this data collection could be made by continuing to analyse the available data from the projects, cruises and archives described in the present work (namely SeaBASS archive which hosts many bio-optical *in situ* data) and find new data sources, while making sure that the already compiled ~~data set~~datasets are the most updated ones following scientific advances and improved quality control measures.

Author contribution

~~AV~~ Salem Ibrahim Salem compil~~i~~ed the database, carried out the integration and quality checking and ~~drafted~~ updated the manuscript. The~~ first eight~~ second to sixth ~~authors, and as well as the~~and last authors~~s~~ ~~are part of the ESA OC-CCI team and contributed to~~ contributed to the design of the compilation, and to the quality checking, as well as contributing data. The remaining authors are listed alphabetically and are data contributors (see their respective ~~data set~~dataset on Table 3~~2~~) or individuals responsible for the development of a particular ~~data set~~dataset (e.g., Jeremy Werdell for NOMAD and Kathryn Barker for MERMAID). All data contributors (listed in Table 3~~2~~) were contacted for authorization of data publishing and offered co-authorship. In the case of the ICES ~~data set~~dataset the permission for publishing was given by the ICES team. All the authors have critically reviewed the manuscript.

Code and data availability

The compiled data are available at https://doi.pangaea.de/10.1594/PANGAEA.995869 (Salem et al., 2026), and the Python scripts used for database compilation, quality control, and figure generation are available from the corresponding author upon reasonable request. The database is composed of three main tables: table "insitudb_chla_v2026.csv" with the observations of "chla_fluor" and "chla_hplc"; table "insitudb_rrs_v2026.csv" with observations of "rrs"; and table "insitudb_iopskdtsm_v2026.csv" with remaining observations ("aph", "adg", "bbp", "kd" and "tsm"). The rows within the three tables relate to each other via a unique key (column "idx"). The three tables can be viewed conceptually as one table with all data. To help with data manipulation, six auxiliary tables derived from the previous three main tables are provided. The table "insitudb_metadata_v2026.csv" contains all available metadata and helps, for example, to find rows (i.e., "idx") with multiple variables (e.g., "rrs" and "chla_fluor"). The remaining four tables ("insitudb_rrs_satbands_2nm_v2026.csv", "insitudb_rrs_satbands_6nm_v2026.csv", "insitudb_iopskdtsm_satbands_2nm_v2026.csv" and "insitudb_iopskdtsm_satbands_6nm_v2026.csv") contain the spectral data of the main tables (i.e., "insitudb_rrs_v2026.csv" and "insitudb_iopskdtsm_v2026.csv") aggregated within ±2 nm and ±6 nm, respectively, of SeaWiFS, MODIS AQUA, MERIS, VIIRS-SNPP, VIIRS-NOAA20, VIIRS-NOAA21, OLCI-S3A and OLCI-S3B sensor bands. The tables are generated by assigning, in each row of the main tables (i.e., "insitudb_rrs_v2026.csv" and "insitudb_iopskdtsm_v2026.csv"), the closest spectral observation within 2 nm (or 6 nm) of a sensor band. The centre-wavelengths of each band and sensor used in the generation of the files are the following: SeaWiFS bands 1-8 were centred at [412, 443, 490, 510, 555, 670, 765, 865] nm, respectively; MODIS-AQUA bands 1-9 were centred at [412, 443, 488, 531, 547, 667, 678, 748, 869] nm, respectively; MERIS bands 1-13 were centred at [412, 442, 490, 510, 560, 620, 665, 681, 709, 753, 779, 865, 885] nm, respectively; VIIRS-SNPP bands 1-6 were centred at [410, 443, 486, 551, 671, 746] nm, respectively; VIIRS-NOAA20 and VIIRS-NOAA21 bands 1-6 were centred at [411, 445, 489, 556, 667, 746] nm, respectively; OLCI-S3A and OLCI-S3B bands 1-15 were centred at [400, 412, 443, 490, 510, 560, 620, 665, 674, 681, 709, 754, 779, 865, 885] nm. An exception to this procedure was made to confirm that the correct MOBY data are stored in the files (see Section 2.3.1. for discussion on how MOBY wavelengths are stored in the main file). In addition, the exact wavelengths of the AERONET-OC data were used when constructing the aggregated ±2 and ±6 nm "rrs" products. Finally, a "readme" file is provided to help the user. Table 2 shows how the compiled data looks like. It is given the example of a query for available chlorophyll data from subdataset "seabass_car81".

Table 6: Example of how the compiled data looks like. It is shown the result if the compilation is queried for the chlorophyll data from subdataset "seabass_car81".

idx	time	lat	lon	chla_fluor	chla_fluor_dataset	chla_fluor_subdataset	chla_fluor_contributor
31095	2002-08-06T09:02:00Z	10.5	-64.67	0.185	seabass	seabass_car81	Frank_Muller-Karger

APPENDIX A: Notation

ad	Detrital absorption coefficient (m^{-1})
adg	Detrital plus CDOM absorption coefficient (m^{-1})
AERONET-OC	AErosol RObotic NETwork-Ocean Color
ag	CDOM absorption coefficient (m^{-1})
AMT	Atlantic Meridional Transect
ap	Particle absorption coefficient (m^{-1})
aph	Algal pigment absorption coefficient (m^{-1})
ARCSSPP	Arctic System Science Primary Production
AWI	Data collection from Astrid Bracher
aw	Pure water absorption coefficient (m^{-1})
BARENTSSEA	Data collection from Knut Yngve Børsheim
BATS	Bermuda Atlantic Time-series Study
bb	Total backscattering coefficient (m^{-1})
bbp	Particle backscattering coefficient (m^{-1})
bbw	Backscattering coefficient of seawater (m^{-1})
BIOCHEM	The Fisheries and Oceans Canada database for biological and chemical data
BODC	British Oceanographic Data Centre
BOUSSOLE	Bouée pour l'acquisition d'une Série Optique à Long Terme
BRDF	Bidirectional Reflectance Distribution Function
CALCOFI	California Cooperative Oceanic Fisheries Investigations
CCELTER	California Current Ecosystem Long Term Ecological Research
CDOM	~~Coulored~~Coloured Dissolved Organic Matter
chla	Chlorophyll~~-~~a concentration (mg m^{-3})
chla_fluor	Chlorophyll~~-~~a concentration determined from fluorometric or spectrophotometric methods (mg m^{-3})
chla_hplc	Total chlorophyll~~-~~a concentration determined from HPLC method (mg m^{-3})
CIMT	Center for Integrated Marine Technology
CNR_GOS	National Research Council – Group of Satellite Oceanography

COASTCOLOUR	Compilation of data in several coastal sites
DFO	Department of Fisheries and Oceans Canada
Es	Surface irradiance (or above-water downwelling irradiance) (mW cm^{-2} μm^{-1})
ESA	European Space Agency
ESTOC	Estación Europea de Series Temporales del Oceano
Fo	Top-of-the-atmosphere solar irradiance (mW cm^{-2} μm^{-1})
GeP&CO	Geochemistry, Phytoplankton, and Color of the Ocean
GOAL	Grupo de Oceanografia de Altas Latitudes
KHOA	Korea Hydrographic and Oceanographic Agency
HOT	Hawaii Ocean Time-series
HPLC	High-Performance Liquid Chromatography
ICES	International Council for the Exploration of the Sea
idx	Unique observation identifier
iFADO	Innovation in the Framework of the Atlantic Deep Ocean
IMOS	Integrated Marine Observing System
INIDEP	Instituto Nacional de Investigación y Desarrollo Pesquero
IOPAN	Institute of Oceanology, Polish Academy of Sciences
ISEE	Institute for Space-Earth Environmental Research
kd	Diffuse attenuation coefficient for downward irradiance (m^{-1})
Lw	water-leaving radiance (or above-water upwelling radiance) (mW cm^{-2} μm^{-1} sr^{-1})
MAREDAT	Compilation of data in several coastal sites
MEDGOOS	Mediterranean Global Ocean Observing System
MERIS	Medium Resolution Imaging Spectrometer
MERMAID	MERIS Match-up *In situ* Database
MOBY	Marine Optical Buoy
MODIS	Moderate Resolution Imaging Spectro-radiometer
MVCO	Martha's Vineyard Coastal Observatory
NASA	National Aeronautics and Space Administration
nLw	Normalized water-leaving radiance (mW cm^{-2} μm^{-1} sr^{-1})

nLw_ex	nLw with a correction for bidirectional effects (mW cm^{-2} µm^{-1} sr^{-1})
NOMAD	NASA bio-Optical Marine Algorithm ~~Data set~~Dataset
OC-CCI	Ocean Colour Climate Change Initiative
OLCI	Ocean and Land Colour Instrument
PALMER	Palmer station Long-Term Ecological Research
PR05	Park and Ruddick (2005) BRDF correction method
rrs	Remote-sensing reflectance (sr^{-1})
Rw	Irradiance reflectance (dimensionless)
SeaBASS	SeaWiFS Bio-optical Archive and Storage System
SEADATANET	Archive of *in situ* marine data
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
TARA	Data collection from global transects
TPSS	Data collection from Trevor Platt and Shubha Sathyendranath
UCSD	University of California San Diego
UNH	University of New Hampshire
YNU	Yokohama National University
VIIRS	Visible Infrared Imaging Radiometer Suite

~~APPENDIX B: Data availability~~

~~The compiled data are available at https://doi.org/10.1594/PANGAEA.941318 (Valente et al., 2022). The database is composed of three main tables: table "insitudb_chla.csv" with the observations of "chla_fluor" and "chla_hplc"; table "insitudb_rrs.csv" with observations of "rrs"; and table "insitudb_iopskdtsm.csv" with remaining observations ("aph", "adg", "bbp", "kd" and "tsm"). The rows within the three tables relate to each other via a unique key (column "idx"). The three tables can be viewed conceptually as one table with all data. To help with data manipulation, six auxiliary tables derived from the previous three main tables are provided. The table "insitudb_metadata.csv" contains all available metadata and helps, for example, to find rows (i.e., "idx") with multiple variables (e.g., "rrs" and "chla_fluor"). The table~~

~~"auxiliary_table_contributors.csv" contains the number of observations per data contributor, variable and dataset. The remaining four tables ("insitudb_rrs_satbands2.csv", "insitudb_rrs_satbands6.csv", "insitudb_iopskdtsm_satbands2.csv" and "insitudb_iopskdtsm_satbands6.csv") contain the spectral data of the main tables (i.e., "insitudb_rrs.csv" and "insitudb_iopskdtsm.csv") aggregated within ±2 nm and ±6 nm, respectively, of SeaWiFS, MODIS AQUA, MERIS, VIIRS-SNPP, VIIRS-JPSS, OLCI-S3A and OLCI-S3B sensor bands. The tables are generated by assigning, in each row of the main tables (i.e., "insitudb_rrs.csv" and "insitudb_iopskdtsm.csv"), the closest spectral observation within 2 nm (or 6 nm) of a sensor band. The centre-wavelengths of each band and sensor used in the generation of the files are the following: SeaWiFS bands 1-8 were centred at [412, 443, 490, 510, 555, 670, 765, 865] nm, respectively; MODIS-AQUA bands 1-9 were centred at [412, 443, 488, 531, 547, 667, 678, 748, 869] nm, respectively; MERIS bands 1-13 were centred at [412, 442, 490, 510, 560, 620, 665, 681, 709, 753, 779, 865, 885] nm, respectively; VIIRS-SNPP bands 1-6 were centred at [410, 443, 486, 551, 671, 746] nm, respectively; VIIRS-JPSS bands 1-6 were centred at [411, 445, 489, 556, 667, 746] nm, respectively; OLCI-S3A and OLCI-S3B bands 1-15 were centred at [400, 412, 443, 490, 510, 560, 620, 665, 674, 681, 709, 754, 779, 865, 885] nm. An exception to this procedure was made to confirm that the correct MOBY data are stored in the files (see Sect. 2.2.1. for discussion on how MOBY wavelengths are stored in the main file). Finally, a "readme" file is provided to help the user. Table 1 shows how the compiled data looks like. It is given the example of a query for available chlorophyll data from subdataset "seabass_car81".~~

~~idx~~	~~time~~	~~lat~~	~~lon~~	~~chla_fluor~~	~~chla_fluor_dataset~~	~~chla_fluor_subdataset~~	~~chla_fluor_contributor~~
~~31095~~	~~2002-08-06T09:02:00Z~~	~~10.5~~	~~-64.67~~	~~0.185~~	~~seabass~~	~~seabass_car81~~	~~Frank_Muller-Karger~~

~~Table B1: Example of how the compiled data looks like. It is shown the result if the compilation is queried for the chlorophyll data from subdataset "seabass_car81".~~ **APPENDIX B: Comprehensive list of AWI datasets**

Table B1: AWI datasets included in the OC-CCI compilation, listing full DOI links, associated variables, and the version in which each dataset was incorporated (v2019, v2022, and v2026).

Link (DOI)	Variable	Database Version	Reference
https://doi.org/10.1594/PANGAEA.879224	rrs	v2019	Bracher, 2017a
https://doi.org/10.1594/PANGAEA.879225	rrs	V2019	Bracher, 2017b
https://doi.org/10.1594/PANGAEA.879226	rrs	V2019	Bracher, 2017c
https://doi.org/10.1594/PANGAEA.847820	chla_hplc, rrs	v2019	Bracher et al., 2015c
https://doi.org/10.1594/PANGAEA.820607	chla_hplc, aph	v2019	Zindler et al., 2013
https://doi.org/10.1594/PANGAEA.848590	chla_hplc	v2019	Bracher, 2015a
https://doi.org/10.1594/PANGAEA.819099	chla_hplc, rrs, aph	v2019	Taylor et al., 2011b
https://doi.org/10.1594/PANGAEA.848591	chla_hplc	v2019	Soppa et al., 2014
https://doi.org/10.1594/PANGAEA.871713	chla_hplc	v2022	Bracher et al., 2015b
https://doi.org/10.1594/PANGAEA.873070	chla_hplc	v2022	Bracher and Taylor, 2017d
https://doi.org/10.1594/PANGAEA.864786	chla_hplc	v2022	Hepach et al., 2016
https://doi.org/10.1594/PANGAEA.898929	chla_hplc	v2022	Bracher et al., 2019c
https://doi.org/10.1594/PANGAEA.898920	chla_hplc	v2022	Bracher, 2019a
https://doi.org/10.1594/PANGAEA.894872	chla_hplc	v2022	Liu et al., 2018b
https://doi.org/10.1594/PANGAEA.899043	chla_hplc	v2022	Bracher and Wiegmann, 2019d
https://doi.org/10.1594/PANGAEA.905502	chla_hplc	v2022	Liu et al., 2019a
https://doi.org/10.1594/PANGAEA.894874	chla_hplc	v2022	Liu et al., 2018c
https://doi.org/10.1594/PANGAEA.898941	chla_hplc	v2022	Bracher, 2019b
https://doi.org/10.1594/PANGAEA.894860	chla_hplc	v2022	Liu et al., 2018a
https://doi.org/10.1594/PANGAEA.911061	chla_hplc	v2022	Bracher et al., 2020
https://doi.org/10.1594/PANGAEA.982433	chla_hplc	v2026	Xi et al., 2025
https://doi.org/10.1594/PANGAEA.941011	chla_hplc	v2026	Bracher and Wiegmann, 2022b
https://doi.org/10.1594/PANGAEA.884528	rrs	v2022	Bracher et al., 2018
https://doi.org/10.1594/PANGAEA.819617	aph	v2019	Soppa et al., 2013
https://doi.org/10.1594/PANGAEA.885246	aph	v2022	Gonçalves-Araujo et al., 2018
https://doi.org/10.1594/PANGAEA.907605	aph	v2022	Wiegmann et al., 2019
https://doi.org/10.1594/PANGAEA.907648	aph	v2022	Liu et al., 2019e
https://doi.org/10.1594/PANGAEA.907419	aph	v2022	Liu et al., 2019d

https://doi.org/10.1594/PANGAEA.983241	aph	v2026	Bracher et al., 2025f
https://doi.org/10.1594/PANGAEA.937971	aph	v2026	Bracher and Taylor, 2021r
https://doi.org/10.1594/PANGAEA.937982	aph	v2026	Bracher et al., 2021s
https://doi.org/10.1594/PANGAEA.938153	aph	v2026	Bracher et al., 2021i
https://doi.org/10.1594/PANGAEA.938193	aph	v2026	Bracher and Liu, 2021b
https://doi.org/10.1594/PANGAEA.938260	aph	v2026	Bracher et al., 2021j
https://doi.org/10.1594/PANGAEA.983161	adg	v2026	Bracher et al., 2025e
https://doi.org/10.1594/PANGAEA.938383	ag	v2026	Bracher et al., 2021m
https://doi.org/10.1594/PANGAEA.938384	ag	v2026	Bracher et al., 2021n
https://doi.org/10.1594/PANGAEA.971524	ag	v2026	Bracher et al., 2025a
https://doi.org/10.1594/PANGAEA.938491	ag	v2026	Bracher et al., 2021c
https://doi.org/10.1594/PANGAEA.938494	ag	v2026	Bracher et al., 2021d
https://doi.org/10.1594/PANGAEA.971552	ag	v2026	Bracher et al., 2025b
https://doi.org/10.1594/PANGAEA.938467	ag	v2026	Bracher et al., 2021e
https://doi.org/10.1594/PANGAEA.938468	ag	v2026	Bracher et al., 2021f
https://doi.org/10.1594/PANGAEA.971585	ag	v2026	Bracher et al., 2025c
https://doi.org/10.1594/PANGAEA.971587	ag	v2026	Bracher et al., 2025d
https://doi.org/10.1594/PANGAEA.938399	ag	v2026	Bracher et al., 2021o
https://doi.org/10.1594/PANGAEA.938400	ag	v2026	Bracher et al., 2021p
https://doi.org/10.1594/PANGAEA.938472	ag	v2026	Bracher et al., 2021k
https://doi.org/10.1594/PANGAEA.938473	ag	v2026	Bracher et al., 2021l
https://doi.org/10.1594/PANGAEA.938152	ad	v2026	Bracher et al., 2021g
https://doi.org/10.1594/PANGAEA.907646	ad	v2026	Liu et al., 2019c
https://doi.org/10.1594/PANGAEA.938196	ad	v2026	Bracher and Liu, 2021a
https://doi.org/10.1594/PANGAEA.907408	ad	v2026	Liu et al., 2019b
https://doi.org/10.1594/PANGAEA.938185	ad	v2026	Bracher et al., 2021q
https://doi.org/10.1594/PANGAEA.938262	ad	v2026	Bracher et al., 2021h
https://doi.org/10.1594/PANGAEA.93148	kd	v2026	Knoll and Müller, 2003
https://doi.org/10.1594/PANGAEA.939927	kd	v2026	Bracher et al., 2022a

Acknowledgements

This paper acknowledges funding from the ESA OC-CCI project, whose products are developed and maintained through support from ESA, the Copernicus Climate Change Service (C3S), and the NERC Earth Observation Data Analysis and AI Service (NEODAAS), as well as from~~,~~ the ESA PHYTOplankton biomass and diversity Climate Change Initiative (PHYTO-CCI, 4000147645/25/I-LR), EUMETSAT "Multi-mission Ocean Colour Algorithm Prototyping" (OMAPS), Simons Foundation (grant number 549947) and the European Union's Horizon 2020 research and innovation programme under grant agreement 810139: Project Portugal Twinning for Innovation and Excellence in Marine Science and Earth Observation – PORTWIMS. This work is also a contribution to project PEst-OE/MAR/UI0199/2014 and to the Horizon Europe project BioEcoOcean (GA 101136748). We would like to thank the efforts of the teams responsible for collection of the data in the field and of the teams responsible for processing and storing the data in archives, without which this work would not be possible. We thank Tamoghna Acharyya at Plymouth Marine Laboratory for his initial contribution to this work. We thank the NOAA (US) for making available the MOBY data, and Yong Sung Kim for the help in questions about MOBY data. BOUSSOLE is supported and funded by the European Space Agency (ESA), the Centre National d'Etudes Spatiales (CNES), the Centre National de la Recherche Scientifique (CNRS), the Institut National des Sciences de l'Univers (INSU), the Sorbonne Université (SU), and the Institut de la Mer de Villefranche (IMEV). We thank ACRI-ST, ARGANS and ESA for access to the MERMAID Database (http://hermes.acri.fr/mermaid). We thank Annelies Hommersom, Pierre Yves Deschamps and David Siegel for allowing the use of MERMAID data for which they are Principal Investigators. The AMT is funded by the UK Natural Environment Research Council through its National Capability Long-term Single Centre Science Programme, Climate Linked Atlantic Sector Science (grant number NE/R015953/1). This study used computing resources funded through the NERC Earth Observation Data Analysis and AI Service (NEODAAS). This study contributes to the international IMBeR project and is contribution number 375 of the AMT programme. We thank the British Oceanographic Data Centre (BODC) for access to AMT data and in particular to Polly Hadziabdic and Rob Thomas for their help in questions about the AMT ~~data set~~dataset. We thank Arwen Bargery, Denise Cummings, Giorgio Dall'Olmo, Ella Darlington,, Victoria Hill, Patrick Holligan, Gerald Moore, Emilio Suarez and Glen Tarran for the use of AMT data for which they are the data contributors. We thank Sam Ahmed, Hui Feng, Alex Gilerson, Brent Holben and Sherwin Ladner for allowing the use of the AERONET-OC data for which they are Principal Investigators. We also thank the Principal Investigators and site managers and site support for AERONET-OC for sustaining the measurements, and the respective national and international funding bodies for financial support. We also thank Giuseppe Zibordi for data quality control and Brent Holben and his team from NASA Goddard Space Flight Centre for making the data available on-line. The AWI ~~data-set~~dataset was supported by the captain, crew and other scientists at the specific German RVs (Meteor, Heincke, Maria S. Merian, Polarstern, Sonne) expeditions and funding by the Helmholtz Association (HGF Innovative Network Funds Phytooptics, Helmholtz Infrastructure Initiative FRAM), the European Union (Seventh Framework Programme SHIVA-226224FP7-ENV-2008- 1020 1), the BMBF (OASIS FK03G0235A and ASTRA FK03G0243A), the DFG (#268020496—TRR 172 project C03 within the Transregional Collaborative Research Center ArctiC Amplication: Climate Relevant Atmospheric 1023 and SurfaCe Processes, and Feedback Mechanisms (AC)3), the collaborative project OLCI-PFT (ACRI-AWI Offer #209-180104) and the ESA 656 S5P+Innovation Theme 7 Ocean Colour (S5POC) project (No 4000127533/19/I-NS), the Copernicus Marine Innovation Service Evolution R&D Project ML-PhyTAO (23138L03D-COP-INNO SCI-9000), and the ESA Phyto-CCI project (4000147645/25/I-LR). The Australian Integrated Marine Observing System (IMOS) and CSIRO are acknowledged for funding the Lucinda AERONET-OC site. Data from IMOS was sourced from Australia's Integrated Marine Observing System (IMOS) - IMOS is enabled by the National Collaborative Research Infrastructure Strategy (NCRIS). We thank Janet Anstee, Joey Crosswell, Britta Schaffelke, Bernadette Sloyan, Paul Thomson and Tom Trull for the use of the IMOS data that they are Principa~~le~~ Investigators. We thank Bob Bidigare, Matthew Church, Ricardo

Letelier and Jasmine Nahorniak for making the HOT data available, and the National Science Foundation for support of the HOT research (grant OCE 09-26766). We thank Yves Dandonneau for allowing the use of GeP&CO data. We thank ICES database on the marine environment (Copenhagen, Denmark, 2014) for allowing the use of their archived data, and Marilynn Sørensen for the help with questions about the ICES ~~data set~~dataset. We thank all ICES contributors for their data. We thank Eric Zettler and SEA Education Association. This study includes data collected by the US Marine Biodiversity Observation Network (MBON) funded under the US National Ocean Partnership Program through NASA grant NNX14AP62A "National Marine Sanctuaries as Sentinel Sites for a Demonstration Marine Biodiversity Observation Network (MBON)". The CARIACO Ocean Time-Series program also provided significant decade-long bio-optical information used in this study. These data were obtained from NOMAD and SeaBASS. We thank NASA, SeaBASS and the Ocean Biology Processing Group (OBPG) for access to SeaBASS and NOMAD data. We thank NASA for project funding for data collection. We thank Chris Proctor from SeaBASS for his valuable and prompt help in a variety of questions. We are deeply thankful to the data contributors of NOMAD and SeaBASS:~~–~~ James Allen, Kevin Arrigo, Dirk Aurin, Mike Behrenfeld, Kelsey Bisson, Emmanuel Boss, Chris Brown, Dylan Catlett, Mary Luz Canon, Douglas Capone, Ken Carder, Carlos Del Castillo, Alex Chekalyuk, Jay-Chung Chen, Dennis Clark, Javier Concha, Jorge Corredor, Glenn Cota, Yves Dandonneau, Heidi Dierssen, David Eslinger, Piotr Flatau, Alex Gilerson, Joaquim Goes, Gwo-Ching Gong, Adriana Gonzalez-Silvera, Jason Graff, Nils Haentjens, Larry Harding, Jon Hare, Sung-Ho Kang, Grace Kim, Gary Kirkpatrick, Oleg Kopelevich, Sasha Kramer, Sam Laney, Pierre Larouche, Zhongping Lee, Ricardo Letelier, Marlon Lewis, Stephane Maritorena, John Marra, Chuck McClain, Christophe Menkes, Mark Miller, Allen Milligan, Ru Morrison, James Mueller, Ruben Negri, James Nelson, Norman Nelson, Mary Jane Perry, David Phinney, John Porter, Collin Roesler, Joe Salisbury, David Siegel, Mike Sieracki, Jeffrey Smart, Raymond Smith, James Spinhirne, Dariusz Stramski, Rick Stumpf, Ajit Subramaniam, Lynne Talley, Chuck Trees, Ryan Vandermeulen, Toby Westberry, Ronald Zaneveld, Eric Zettler and Richard Zimmerman. For the BIOCHEM data we thank the Fisheries and Oceans Canada and the following data contributors: Diane Archambault, Hughes Benoit, Esther Bonneau, Eugene Colbourne, Alain Gagne, Yves Gagnon, Tom Hurlbut, Catherine Johnson, Pierre Joly, Maurice Levasseur, Jean-Francois Lussier, Sonia Michaud, Patrick Ouellet, Jacques Plourde, Stephane Plourde, Luc Savoie, Michael Scarratt, Philippe Schwab, Michel Starr and François Villeneuve. We also thank Laure Devine for the help in processing the BIOCHEM ~~data set~~dataset. CalCOFI research is supported by contributions from the participating agencies: The California State Department of Fish and Wildlife, NOAA, National Marine Fisheries Service, Southwest Fisheries Science Center, and the University of California, Integrative Oceanography Division at the Scripps Institution of Oceanography, UCSD. The authors would like to thank the Oceanic Platform of the Canary Islands (PLOCAN) and its staff for making freely available the use of this ESTOC ~~data set~~dataset. We thank the following MAREDAT data providers: Robert Bidigare, Denise Cummings, Giacomo DiTullio, Chris Gallienne, Ralf Goericke, Patrick Holligan, David Karl, Michael Landry, Michael Lomas, Michael Lucas, Jean-Claude Marty, Walker Smith, Rick Stumpf, Emilio Suarez, Koji Suzuki, Maria Vernet and Simon Wright. We thank Oscar Schofield, Raymond Smith and Maria Vernet for allowing the use of the PALMER data. Data from the Palmer LTER data repository were supported by Office of Polar Programs, NSF Grants OPP-9011927, OPP-9632763 and OPP-0217282. We thank the SeaDataNet Pan-European infrastructure for ocean and marine data management (http://www.seadatanet.org). We thank Emmanuel Boss for the TARA data. Funding for the collection and processing of the TARA ~~data set~~dataset was provided by NASA Ocean Biology and Biogeochemistry program under grants NNX11AQ14G, NNX09AU43G, NNX13AE58G and NNX15AC08G to the University of Maine. The CARBO-ACID cruise of iFADO dataset was funded by the Eurofleets+ project, under the EU H2020 research and innovation programme (Grant Agreement No 824077). We thank Tae-Ho Kim (Underwater Survey Technology 21 Inc., Republic of Korea) for his efforts in compiling, organizing, and providing the KHOA dataset, and for ensuring the consistency and completeness of the associated metadata. We acknowledge the AERONET-OC program and thank the Principal Investigators, Co-Investigators, and their staff for establishing and maintaining the 39 AERONET-OC sites used in this study. We would like to honour the memory of Marcel Wernand, Tiffany Moisan and Trevor Platt, authors who contributed to the previous versions.

References

Aiken, J., Cummings, D.G., Gibb, S.W., Rees, N.W., Woodd-Walker, R., Woodward, E.M.S., Woolfenden, J., Hooker, S.B., Berthon, J-F., Dempsey, C.D., Suggett, D.J., Wood, P., Donlon, C., Gonzalez-Benitez, N., Huskin, I., Quevedo, M., Barciela-Fernandez, R., de Vargas, C. and McKee, C.: AMT-5 Cruise Report. NASA Tech. Memo. 1998–206892, Vol. 2, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 113pp, 1998.

Álvarez, E., Losa, S.N., Bracher, A., Thoms, S. and Völker, C.: Phytoplankton light absorption impacted by photoprotective carotenoids in a global ocean spectrally-resolved biogeochemistry model, J. Adv. Model. Earth Syst., 14, e2022MS003126, 2022.

Amante, C. and Eakins, B.W.: ETOPO1, 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA, 2009.

Antoine, D., André, J. M. and Morel, A.: Oceanic primary production: 2. Estimation at global scale from satellite (CZCS) chlorophyll. Global Biogeochemical Cycles, 10, 57 – 70, 1996.

Antoine, D., Chami, M., Claustre, H., D'Ortenzio, F., Morel, A., Bécu, G. , Gentili, B., Louis, F., Ras, J., Roussier, E., Scott, A.J. , Tailliez, D., Hooker, S. B.,Guevel, P., Desté, J.-F., Dempsey, C. and Adams, D.: BOUSSOLE : a joint CNRS-INSU, ESA, CNES and NASA Ocean Color Calibration And Validation Activity. NASA Technical memorandum N° 2006 - 214147, 61 pp, 2006.

Antoine, D., Guevel, P., Desté, J.-F., Bécu, G., Louis, F., Scott, A. and Bardey, P.: The "BOUSSOLE" Buoy—A New Transparent-to-Swell Taut Mooring Dedicated to Marine Optics: Design, Tests, and Performance at Sea. J. Atmos. Oceanic Technol., 25, 968–989, 2008.

Bailey, S.W. and Werdell, P.J.: A multi-sensor approach for the on-orbit validation of ocean color satellite data products. Rem. Sens. Environ., 102, 12-23, 2006.

Barker, K.: In-situ Measurement Protocols. Part A: Apparent Optical Properties, Issue 2.0, Doc. no: CO-SCI-ARG-TN-0008, ARGANS Ltd., p. 126, 2013a.

Barker, K.: In-situ Measurement Protocols. Part B: Inherent Optical Properties and in-water constituents, Issue 1.0, Doc. no: CO-SCI-ARG-TN-0008, ARGANS Ltd., p. 39, 2013b.

Bracher, A.: Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXVII/2, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.848590, 2015a.

Bracher, A., Taylor, M. H., Taylor, B. B., Dinter, T., Röttgers, R., and Steinmetz, F.: Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXIII/1, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.871713, 2015b.

Bracher, A., Taylor, M. H., Taylor, B. B., Dinter, T., Röttgers, R., and Steinmetz, F.: Phytoplankton pigments, hyperspectral downwelling irradiance and remote sensing reflectance during POLARSTERN cruises ANT-XXIII/1, ANT-XXIV/1, ANT-XXIV/4, ANT-XXVI/4, and Maria S. Merian cruise MSM18/3, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.847820, 2015c.

Bracher, A.: Remote sensing reflectance during POLARSTERN cruise ANT-XXIV/4, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.879224, 2017a.

Bracher, A.: Remote sensing reflectance during POLARSTERN cruise ANT-XXV/1, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.879225, 2017b.

Bracher, A.: Remote sensing reflectance during POLARSTERN cruise ANT-XXVI/4, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.879226, 2017c.

Bracher, A., and Taylor, B. B.: Phytoplankton pigment concentrations measured by HPLC during Maria S. Merian cruise MSM9/1, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.873070, 2017d.

Bracher, A., Gonçalves-Araujo, R., Dinter, T., and Cherkasheva, A.: Remote sensing reflectance obtained from spectral underwater upwelling radiance and solar downwelling irradiance measurements during POLARSTERN cruise ARK-XXVI/3 (PS78), Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.884528, 2018.

Bracher, A.: Phytoplankton pigment concentrations during RV Sonne cruise SO243, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.898920, 2019a.

Bracher, A.: Phytoplankton pigment concentrations in the Southern Ocean during RV POLARSTERN cruise PS103 in Dec 2016 to Jan 2017, Alfred Wegener Institute, Helmholtz Centre for Polar and

Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.898941, 2019b.

Bracher, A., Cheah, W., and Wiegmann, S.: Phytoplankton pigment concentrations in the tropical Indian Ocean in July and August 2014 during RV Sonne cruises SO234 and SO235, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.898929, 2019c.

Bracher, A., and Wiegmann, S.: Phytoplankton pigment concentrations in the North Sea and Sogne Fjord from 29 April to 7 May 2016 during RV HEINCKE cruise HE462, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.899043, 2019d.

Bracher, A., Wiegmann, S., Xi, H., and Dinter, T.: Phytoplankton pigment concentration and phytoplankton groups measured on water samples obtained during POLARSTERN cruise PS113 in the Atlantic Ocean, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.911061, 2020.

Bracher, A., and Liu, Y.: Spectrophotometric measurements of absorption coefficients by non-algal particles in the Atlantic Southern Ocean during RV POLARSTERN cruise PS103 in Dec 2016 to Jan 2017, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938196, 2021a.

Bracher, A., and Liu, Y.: Spectrophotometric measurements of absorption coefficients by phytoplankton in the Atlantic Southern Ocean during RV POLARSTERN cruise PS103 in Dec 2016 to Jan 2017, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938193, 2021b.

Bracher, A., Liu, Y., Hellmann, S., Bensi, M., Kovacevic, V., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter from North Sea to Fram Strait measured at fixed stations with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS99.1, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938491, 2021c.

Bracher, A., Liu, Y., Hellmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter from North Sea to Fram Strait measured underway with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS99.1, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938494, 2021d.

Bracher, A., Liu, Y., Oelker, J., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter across the South Atlantic Ocean measured at fixed stations with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS103, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938467, 2021e.

Bracher, A., Liu, Y., Oelker, J., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter across the South Atlantic Ocean measured underway with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS103, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938468, 2021f.

Bracher, A., Liu, Y., and Wiegmann, S.: Spectrophotometric measurements of absorption coefficients by non-algal particles during during HEINCKE cruise HE462 in the North Sea and Sogne Fjord from 29 April to 7 May 2016, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938152, 2021g.

Bracher, A., Liu, Y., and Wiegmann, S.: Spectrophotometric measurements of absorption coefficients by non-algal particles during RV POLARSTERN cruise PS121 from 11 Aug to 10 Sep 2019, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938262, 2021h.

Bracher, A., Liu, Y., and Wiegmann, S.: Spectrophotometric measurements of absorption coefficients by phytoplankton during HEINCKE cruise HE462 in the North Sea and Sogne Fjord from 29 April to 7 May 2016, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938153, 2021i.

Bracher, A., Liu, Y., and Wiegmann, S.: Spectrophotometric measurements of absorption coefficients by phytoplankton during RV POLARSTERN cruise PS121 from 11 Aug to 10 Sep 2019, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938260, 2021j.

Bracher, A., Liu, Y., Wiegmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter (CDOM) from North Sea to Fram Strait measured at fixed stations with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS121, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938472, 2021k.

Bracher, A., Liu, Y., Wiegmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter (CDOM) from North Sea to Fram Strait measured underway with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS121, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938473, 2021l.

Bracher, A., Liu, Y., Wiegmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter obtained at fixed stations with a Liquid Waveguide Capillary Cell system during HEINCKE cruise HE462 in the North Sea and Sogne Fjord, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938383, 2021m.

Bracher, A., Liu, Y., Wiegmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter obtained underway with a Liquid Waveguide Capillary Cell system during HEINCKE cruise HE462 in the North Sea and Sogne Fjord, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938384, 2021n.

Bracher, A., Liu, Y., Wiegmann, S., Xi, H., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter across the Atlantic Ocean measured at fixed stations with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS113, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938399, 2021o.

Bracher, A., Liu, Y., Wiegmann, S., Xi, H., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter across the Atlantic Ocean measured underway with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS113, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938400, 2021p.

Bracher, A., Liu, Y., Xi, H., and Wiegmann, S.: Spectrophotometric measurements of absorption coefficients by non-algal particles during POLARSTERN cruise PS113 along an Atlantic Transect, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.938185, 2021q.

Bracher, A., and Taylor, B. B.: Phytoplankton absorption during POLARSTERN cruise ANT-XXVI/4 (PS75), PANGAEA [data set], https://doi.org/10.1594/PANGAEA.937971, 2021r.

Bracher, A., Taylor, B. B., and Cheah, W.: Phytoplankton absorption during SONNE cruise SO218, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.937982, 2021s.

Bracher, A., Taylor, M.H., Taylor, B., Dinter, T., Röttgers, R. and Steinmetz, F.: Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations, Ocean Sci., 11, 139–158, 2015.

Bracher, A., Dinter, T., and von Appen, W. J.: Mean spectral diffuse attenuation coefficients from 320 to 917 nm from high spectrally resolved radiometric underwater light transmission profiles obtained by an undulating platform towed behind the ship, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.939927, 2022a.

Bracher, A., and Wiegmann, S.: Phytoplankton pigment concentrations during POLARSTERN cruise PS121 from North Sea to Fram in August to September 2019, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.941011, 2022b.

Bracher, A., Liu, Y., Hellmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter from the East Greenland Sea measured at fixed stations with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS99.2, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.971524, 2025a.

Bracher, A., Liu, Y., Hellmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter from the East Greenland Sea measured underway with a Liquid Waveguide Capillary

Cell system during POLARSTERN cruise PS99.2, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.971552, 2025b.

Bracher, A., Liu, Y., Wiegmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter from the East Greenland Sea measured at fixed stations with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS107, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.971585, 2025c.

Bracher, A., Liu, Y., Wiegmann, S., and Röttgers, R.: Absorption coefficients by coloured dissolved organic matter from the East Greenland Sea measured underway with a Liquid Waveguide Capillary Cell system during POLARSTERN cruise PS107, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.971587, 2025d.

Bracher, A., Soppa, M. A., Banks, A. C., Xi, H., Chaikalis, S., and Röttgers, R.: Absorption coefficients by coloured detrital and dissolved organic matter from a global surface water in-situ collection at the first eight bands of and matched to the Ocean Land Colour Imager (OLCI), PANGAEA [data set], https://doi.org/10.1594/PANGAEA.983161, 2025e.

Bracher, A., Soppa, M. A., Banks, A. C., Xi, H., Chaikalis, S., and Röttgers, R.: Absorption coefficients by phytoplankton at the first eight Ocean Land Colour Instrument (OLCI) bands from a global in-situ collection of open ocean, coastal and inland surface waters matched to OLCI, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.983241, 2025f.

Bricaud, A., Babin, M., Morel, A., and Claustre, H.: Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization, J. Geophys. Res. Oceans, 100, 13321–13332, 1995.

Bricaud, A., Claustre, H., Ras, J., and Oubelkheir, K.: Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size sctructure of algal populations. J. Geophys. Res., 109, C11010, doi:10.1029/2004JC002419, 2004.

Clark, D. K., Yarborough, M. A., Feinholz, M. E., Flora, S., Broenkow, W., Kim, Y. S., Johnson, B. C., Brown, S. W., Yuen, M. and Mueller, J. L.: MOBY, A Radiometric Buoy for Performance Monitoring and Vicarious Calibration of Satellite Ocean Colour Sensors: Measurements and Data Analysis Protocols. In Ocean Optics Protocols for Satellite Ocean Colour Sensor Validation, NASA Technical Memo. 2003-211621/Rev4,— _Vol VI, 3-34 (Eds J. L. Muller, G. Fargion and C. McClain). Greenbelt, MD.: NASA/GSFC, 2003.

Dandonneau, Y. and Niang, A.: Assemblages of phytoplankton pigments along a shipping line through the North Atlantic and Tropical Pacific. Prog. Oceanogr., 73, 2007.

Devine, L., Kennedy, M.K., St-Pierre, I., Lafleur, C. , Ouellet, M. and Bond, S.: BioChem: the Fisheries and Oceans Canada database for biological and chemical data. Can. Tech. Rep. Fish. Aquat. Sci., 3073: iv + 40 pp,– http://waves-vagues.dfo-mpo.gc.ca/Library/351319.pdf, 2014.

DFO.: BioChem: database of biological and chemical oceanographic data. Department of Fisheries and Oceans, Canada. http://www.dfo-mpo.gc.ca/science/data-donnees/biochem/index-eng.html. Database accessed on April 2015, 2018

GCOS, 2022. GCOS 2022 implementation plan, World Meteorological Organization (WMO). Geneva.

Gordon, H. R. and Clark, D.K.: Clear water radiances for atmospheric correction of coastal zone color scanner imagery, Applied Optics, 20, 4175-4180, 1981.

Gonçalves-Araujo, R., Wiegmann, S., and Bracher, A.: Phytoplankton absorption coefficient spectra during POLARSTERN cruise ARK-XXVI/3 (PS78, TRANSARC), PANGAEA [data set], https://doi.org/10.1594/PANGAEA.885246, 2018.

Gregg, W.W. and Carder, K. L.: A simple spectral solar irradiance model for cloudless maritime atmospheres, Limnol. Oceanogr., 35, 1657-1675, 1990.

Hepach, H., Quack, B., Tegtmeier, S., Engel, A., Bracher, A., Fuhlbrügge, S., Galgani, L., Atlas, E. L., Lampel, J., Frieß, U., and Krüger, K.: Pigment measured on water bottle samples during METEOR cruise M91, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.864786, 2016.

Hooker, S.B., McClain, C.R., Firestone, J.K., Westphal, T.L., Yeh, E-n. and Ge, Y.: The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS), Part 1. NASA Tech. Memo. 104566, Vol. 20, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 40 pp, 1994.

Hooker, S.B., Zibordi, G., Berthon, J-F., Bailey, S.W. and Pietras, C.M.: The SeaWiFS Photometer Revision for Incident Surface Measurement (SeaPRISM) Field Commissioning. NASA Tech. Memo. 2000–206892, Vol. 13, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 24pp, 2000.

IOCCG: Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications. Lee, Z-P. (eds), Reports of the International Ocean-Colour Coordinating Group, No. 5, IOCCG, Dartmouth, Canada, 2006.

IOCCG: Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology. Platt, T., Hoepffner, N., Stuart, V. and Brown, C. (eds.), Reports of the International Ocean-Colour Coordinating Group, No. 7, IOCCG, Dartmouth, Canada, 2008.

Karl, D.M. and Michaels, A.F.: The Hawaiian Ocean Time-series (HOT) and Bermuda Atlantic Time-series Study (BATS)—Preface . Deep-Sea Res. II, 43, 127–128, 1996.

Knoll, M., and Müller, T. J.: Physical oceanography at station POS212/2-CTD-821_038, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.93148, 2003.

Liu, Y., Boss, E., Chase, A. P., Xi, H., Zhang, X., Röttgers, R., Pan, Y., and Bracher, A.: Phytoplankton pigment concentration measured by HPLC during POLARSTERN cruise PS107, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.894860, 2018a.

Liu, Y., Boss, E., Chase, A. P., Xi, H., Zhang, X., Röttgers, R., Pan, Y., and Bracher, A.: Phytoplankton pigment concentration measured by HPLC during POLARSTERN cruise PS93.2, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.894872, 2018b.

Liu, Y., Boss, E., Chase, A. P., Xi, H., Zhang, X., Röttgers, R., Pan, Y., and Bracher, A.: Phytoplankton pigment concentration measured by HPLC during POLARSTERN cruise PS99, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.894874, 2018c.

Liu, Y., Hellmann, S., Wiegmann, S., and Bracher, A.: Phytoplankton pigment concentrations measured by HPLC during POLARSTERN cruise PS99.1, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.905502, 2019a.

Liu., Y., Roettgers, R., Ramírez-Pérez, M., Dinter, T., Steinmetz, F., Noethig, E.-M., Hellmann, S., Wiegmann, S., Bracher A.: Underway spectrophotometry in the Fram Strait (European Arctic Ocean): a highly resolved chlorophyll *a* data source for complementing satellite ocean color. Optics Express, 26(14), A678-A698; https://doi.org/10.1364/OE.26.00A678, 2018

Liu, Y., Wiegmann, S., and Bracher, A.: Absorption coefficient spectra (median) of non-algal particles during POLARSTERN cruise PS107, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.907408, 2019b.

Liu, Y., Wiegmann, S., and Bracher, A.: Absorption coefficient spectra (median) of non-algal particles during POLARSTERN cruise PS99, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.907646, 2019c.

Liu, Y., Wiegmann, S., and Bracher, A.: Spectrophotometric measurements of absorption coefficients and optical density by total particles, phytoplankton and non-algal particles during POLARSTERN cruise PS107, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.907419, 2019d.

Liu, Y., Wiegmann, S., and Bracher, A.: Spectrophotometric measurements of absorption coefficients and optical density by total particles, phytoplankton and non-algal particles during POLARSTERN cruise PS99, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.907648, 2019e.

Matrai, P. A., Olson, E., Suttles, S., Hill, V. J., Codispoti, L. A., Light, B. and Steele, M.: Synthesis of primary production in the Arctic Ocean: I. Surface waters, 1954-2007. Prog. Oceanogr., 110, 93-106, doi:10.1016/j.pocean.2012.11.004, 2013

McClain, C. R.: A decade of satellite ocean color observations. Annu. Rev. Marine Sci. Vol. 1, pp 19-42, 2009

Morel, A. and Gentilli, B.: Diffuse Reflectance of Oceanic Waters. 3. Implications of Bidirectionality for the Remote-Sensing Problem. Applied Optics, 35, 4850-4862, 1996.

Morel, A., Antoine, D. and Gentilli, B.: Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle scattering phase function. Applied Optics, 41(30), 6289-6306, 2002.

Morel, A. and Maritorena, S.: Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical Research, 106, 7163 – 7180, 2001.

Nechad, B., Ruddick, K., Schroeder, T., Oubelkheir, K., Blondeau-Patissier, D., Cherukuru, N., Brando, V., Dekker, A., Clementson, L., Banks, A. C., Maritorena, S., Werdell, J., Sá, C., Brotas, V., Caballero de Frutos, I., Ahn, Y.-H., Salama, S., Tilstone, G., Martinez-Vicente, V., Foley, D., McKibben, M., Nahorniak, J., Peterson, T., Siliò-Calzada, A., Röttgers, R., Lee, Z., Peters, M., and Brockmann, C.: CoastColour Round Robin datasets: a database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters. Earth Syst. Sci. Data Discuss., 8, 173-258, doi:10.5194/essdd-8-173-2015, 2015.

Neuer, S., Cianca, A., Helmke, P., Freudenthal, T., Davenport, R., Meggers, H., Knoll, M., Santana-Casiano, J.M., González-Davila, M., Rueda, M.-J. and Llinás, O.: Biogeochemistry and hydrography in the eastern subtropical North Atlantic gyre. Results from the European time-series station ESTOC. Progress in Oceanography, 72, 1–29, doi:10.1016/j.pocean.2006.08.001, 2007.

Palmer Station Antarctica LTER, Schofield, O., Vernet, M., and Prezelin, B.: Photosynthetic pigments of water column samples analyzed using High Performance Liquid Chromatography (HPLC), collected aboard Palmer LTER annual cruises off the coast of the Western Antarctic Peninsula, 1991–2016, version 8, Environmental Data Initiative [dataset], https://doi.org/10.6073/pasta/4d583713667a0f52b9d2937a26d0d82e, 2018.

Palmer Station Antarctica LTER, Schofield, O., Vernet, M., and Smith, R.: Chlorophyll and phaeopigments from water column samples, collected at selected depths at Palmer Station Antarctica, during the Palmer LTER field seasons, 1991–2019, version 7, Environmental Data Initiative [dataset], https://doi.org/10.6073/pasta/09a1f2cc150b547e3c9b20c39e10cfc2, 2020a.

Palmer Station Antarctica LTER, Schofield, O., Vernet, M., and Smith, R.: Chlorophyll and phaeopigments from water column samples, collected at selected depths aboard Palmer LTER annual cruises off the coast of the Western Antarctic Peninsula, 1991–2019, version 10, Environmental Data Initiative [dataset], https://doi.org/10.6073/pasta/6bbce1e3264571463c0354874dba88cf, 2020b.

Palmer Station Antarctica LTER, Schofield, O., Vernet, M., and Prezelin, B.: Photosynthetic pigments of water column samples analyzed using High Performance Liquid Chromatography (HPLC), sampled during the Palmer LTER field seasons at Palmer Station, Antarctica, 1991–2015, version 6, Environmental Data Initiative [dataset], https://doi.org/10.6073/pasta/ec55e3d0d7260e1df98c9156f9becdeb, 2020c

Park, Y.-J., Ruddick, K., 2005. Model of remote-sensing reflectance including bidirectional effects for case 1 and case 2 waters. Appl. Opt. 44, 1236–1249.

Peloquin, J., Swan, C., Gruber, N., Vogt, M., Claustre, H., Ras, J., Uitz, J., Barlow, R., Behrenfeld, M., Bidigare, R., Dierssen, H., Ditullio, G., Fernandez, E., Gallienne, C., Gibb, S., Goericke, R., Harding, L., Head, E., Holligan, P., Hooker, S., Karl, D., Landry, M., Letelier, R., Llewellyn, C. A., Lomas, M., Lucas, M., Mannino, A., Marty, J.-C., Mitchell, B. G., Muller-Karger, F., Nelson, N., O'Brien, C., Prezelin, B., Repeta, D., Jr. Smith, W. O., Smythe-Wright, D., Stumpf, R., Subramaniam, A., Suzuki, K., Trees, C., Vernet, M., Wasmund, N., and Wright, S.: The MAREDAT global database of high performance liquid chromatography marine pigment measurements, Earth Syst. Sci. Data, 5, 109-123, https://doi.org/10.5194/essd-5-109-2013, 2013.

Philipson, P., Kratzer, S., Ben Mustapha, S., Strömbeck, N. and Stelzer, K.: Satellite-based water quality monitoring in Lake Vänern, Sweden. International Journal of Remote Sensing, 37:16, 3938-3960, doi:10.1080/01431161.2016.1204480, 2016.

Pope, R., and Fry, E.: Absorption spectrum (380 - 700nm) of pure waters: II. Integrating cavity measurements, Appl. Opt. 36, 8710-8723, 1997.

Pitarch, J., Vanhellemont, Q., Röttgers, R., Borges, A. V., and Babin, M.: Match-up data set for remote sensing of inherent optical properties in the Arctic Ocean [dataset]. PANGAEA, https://doi.org/10.1594/PANGAEA.793246, 2013.

Robins, D.B., Bale, A.J., Moore, G.F., Rees, N.W., Hooker, S.B., Gallienne, C.P., Westbrook, A.G., Marañon, E., Spooner, W.H. and Laney, S.R.: AMT-1 Cruise Report and Preliminary Results. NASA Tech. Memo. 104566, Vol. 35, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 87pp, 1996.

Robinson, C., Poulton, A. J., Holligan, P. M., Baker, A. R., Forster, G., Gist, N., Jickells, T. D., Malin G., Upstill-Goddard, R., Williams, R. G., Woodward, E. M. S. and Zubkov, M. V.: The Atlantic Meridional Transect (AMT) Programme: a contextual view 1995-2005. Deep-Sea Research II, 53, 1485-1515, doi: 10.1016/j.dsr~~2.2~~2.3006.05.015, 2006.

Salem, S. I., Valente, A., Sathyendranath, S., Chuprin, A., Renard, T., Pitarch, J., Airs, R., Moreno, A., Angelico, M.-M., Antoine, D., Arnone, R., Artuso, F., Balch, W. M., Barker, K., Barlow, R., Bélanger, S., Bellacicco, M., Berthon, J.-F., Besiktepe, S., Bignami, F., Borsheim, Y., Bracher, A., Braga, F., Brando, V. E., Brewin, R. J. W., Brotas, V., Canuti, E., Chavez, F. P., Cianca, A., Claustre, H., Clementson, L., Colella, S., Costa, R. R., Crout, R., Devred, E., Di Cicco, A., Falcini, F., Farace, I., Ferreira, A., Freeman, S., Frouin, R., García-Soto, C., Giardina, I., Gibb, S. W., Gleason, A., Goericke, R., Gould, R., Grant, M., Guillocheau, N., Higa, H., Hooker, S. B., Hu, C., Ishizaka, J., Jackson, T., Kahru, M., Kampel, M., Kim, B., Klein, H., Kowalczuk, P., Kratzer, S., Kudela, R. M., Lai, A., Lavezza, R., Ledesma, J., Lee, H., Lohrenz, S., Loisel, H., Lutz, V. A., Mannino, A., Margiotta, F., Martinez-Vicente, V., Marullo, S., Massi, L., Matrai, P. A., Matsuoka, A., McKee, D., Melillo, C., Mendes, C. R. B., Mitchell, B. G., Moisan, T., Montes, E., Muller-Karger, F. E., Neeley, A., Novak, M. G., Nunes, C., O'Dowd, L., Oliveira, P. O., Ondrusek, M., Organelli, E., Park, H., Platt, T., Polonelli, F., Poulton, A. J., Ragni, M., Repecaud, M., Ribera d'Alcalà, M., Rodriguez-Ramos, T., Röttgers, R., Ruiz-Villarreal, M., Sammartino, M., Santoleri, R., Scarpa, G. M., Schroeder, T., Segura, V., Smyth, T. J., Smythe-Wright, D., Sosik, H., Stoń-Egiert, J., Taberner, M., Tan, J., Thomas, R., Thomas, C. S., Tilstone, G. H., Tracana, A., Twardowski, M. S., Vellucci, V., Volpe, G., Voss, K., Werdell, J., Wernand, M. R., Wojtasiewicz, B., Wright, S., Yang, S., Zibordi, G., Zoffoli, M. L., and Groom, S.: A compilation of global bio-optical in situ data for ocean-colour satellite applications – version four, PANGAEA [dataset], https://doi.pangaea.de/10.1594/PANGAEA.995869, in review.

Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., Forget, M.-H., Maass, H. and Platt, T.: Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. Mar. Ecol. Prog. Ser., 383: 73–84, doi: 10.3354/meps07998, 2009.

Sathyendranath, S., Brewin, R. J. W., Brockmann, C., Brotas, V., Ciavatta, S., Chuprin, A., Couto, A. B., Dowell, M., Franz, B., Grant, M., Groom, S., Horseman, A., Jackson, T., Krasemann, H., Lavender, S., Martinez Vicente, V., Melin, F., Platt, T., Regner, P., Roy, S., Steinmetz, F., Swinton, J., Thompson, A., Valente, A., Werdell, J., Zuhlk, M., Brando, V. E., Frouin, R., Gould, R. W., Hooker, S., Kahru, M., Mitchell, B. G., Muller-Karger, F., Sosik, H. M. and Voss, K. J.: An ocean-colour time series for use in climate studies: the experience of the ocean-colour climate change initiative (OC-CCI). Sensors, 19, 4285, 2019.

Schroeder, T., Lovell, J., King, E., Clementson, L. and Scott, R.: IMOS Ocean Colour Validation Report 2015-16, Report to the Integrated Marine Observing System (IMOS), CSIRO Oceans and Atmosphere, 33 pp., 2016.

Soppa, M. A., Dinter, T., Taylor, B. B., and Bracher, A.: Particulate and phytoplankton absorption during POLARSTERN cruises ANT-XXVI/3 and ANT-XXVIII/3, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.819617, 2013.

Soppa, M. A., Hirata, T., Silva, B., Dinter, T., Peeken, I., Wiegmann, S., and Bracher, A.: Phytoplankton pigment concentrations in the South Atlantic Ocean, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.848591, 2014.

Steinmetz, F., Deschamps, P.-Y., Ramon, D., 2011. Atmospheric correction in presence of sun glint: application to MERIS. Opt. Express 19, 9783–9800. https://doi.org/10.1364/OE.19.009783

Taylor, B. B., Torrecilla, E., Bernhardt, A., Taylor, M. H., Peeken, I., Röttgers, R., Piera, J., and Bracher, A.: Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance. Biogeosciences, 8, 3609-3629, https://doi.org/10.5194/bg-8-3609-2011, 2011a.

Taylor, B. B., Torrecilla, E., Bernhardt, A., Taylor, M. H., Peeken, I., Röttgers, R., Piera, J., and Bracher, A.: Phytoplankton pigments, composition, hyperspectral light field data and biooptical properties during POLARSTERN cruise ANT-XXV/1, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.819099, 2011b.

Thuillier, G., Hersé, M., Labs, D., Foujols, T., Peetermans, W., Gillotay, D., Simon, P. C. and Mandel, H.: The solar spectral irradiance from 200 nnm to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS 1-2-3 and EURECA missions. Solar Physics, 214:1–22, 2003.

Tilstone, G., Dall'Olmo, G., Hieronymi, M., Ruddick, K., Beck, M., Ligi, M., Costa, M., D'Alimonte, D., Vellucci, V., Vansteenwegen, D., Bracher, A., Wiegmann, S., Kuusk, J., Vabson, V., Ansko, I., Vendt, R., Donlon, C., Casal, T.: Field intercomparison of radiometer measurements for ocean colour validation, *Remote Sens.*, *12*(10), 1587, https://doi.org/10.3390/rs12101587, 2020

Tiwari, S. P., and Shanmugam, P.: An optical model for deriving the spectral particulate backscattering coefficients in oceanic waters. Ocean Science, 9 (6), 987-1001, 2013.

Trees, C. C., Kennicutt II, M. C. and Brooks, J. M.: Errors associated with the standard fluorimetric determination of chlorophylls and phaeopigments. Marine Chemistry, 17: 1-12, 1985.

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., Beşiktepe, Ş., 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.

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., Beşiktepe, Ş., Borsheim, Y., Bracher, A., Brando, V., Canuti, E., Chavez, F., Cianca, A., Claustre, H., Clementson, L., Crout, R., Frouin, R., García-Soto, C., Gibb, S. W., Gould, R., Hooker, S. B., Kahru, M., Kampel, M., Klein, H., Kratzer, S., Kudela, R., Ledesma, J., Loisel, H., Matrai, P., McKee, D., Mitchell, B. G., Moisan, T., Muller-Karger, F., O'Dowd, L., Ondrusek, M., Platt, T., Poulton, A. J., Repecaud, M., Schroeder, T., Smyth, T., Smythe-Wright, D., Sosik, H. M., Twardowski, M., Vellucci, V., Voss, K., Werdell, J., Wernand, M., Wright, S., and Zibordi, G.: A compilation of global bio-optical in situ data for ocean-colour satellite applications – version two, Earth Syst. Sci. Data, 11, 1037–1068, https://doi.org/10.5194/essd-11-1037-2019, 2019.

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Jackson, T., Chuprin, A., Taberner, M., Airs, R., Antoine, D., Arnone, R., Balch, W. M., Barker, K., Barlow, R., Bélanger, S., Berthon, J.-F., Beşiktepe, Ş., Borsheim, Y., Bracher, A., Brando, V., Brewin, R. J. W., Canuti, E., Chavez, F. P., Cianca, A., Claustre, H., Clementson, L., Crout, R., Ferreira, A., Freeman, S., Frouin, R., García-Soto, C., Gibb, S. W., Goericke, R., Gould, R., Guillocheau, N., Hooker, S. B., Hu, C., Kahru, M., Kampel, M., Klein, H., Kratzer, S., Kudela, R., Ledesma, J., Lohrenz, S., Loisel, H., Mannino, A., Martinez-Vicente, V., Matrai, P., McKee, D., Mitchell, B. G., Moisan, T., Montes, E., Muller-Karger, F., Neeley, A., Novak, M., O'Dowd, L., Ondrusek, M., Platt, T., Poulton, A. J., Repecaud, M., Röttgers, R., Schroeder, T., Smyth, T., Smythe-Wright, D., Sosik, H. M., Thomas, C., Thomas, R., Tilstone, G., Tracana, A., Twardowski, M., Vellucci, V., Voss, K., Werdell, J., Wernand, M., Wojtasiewicz, B., Wright, S., and Zibordi, G: : A compilation of global bio-optical in situ data for ocean-colour satellite applications - version 3. PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.941318, 2022

Volpe, G., Santoleri, R., Vellucci, V., Ribera d'Alcalà, M., Marullo, S., and D'Ortenzio, F.: The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates, Remote Sens. Environ., 107, 625–638, https://doi.org/https://doi.org/10.1016/j.rse.2006.10.017, 2007.

Volpe, G., Colella, S., Forneris, V., Tronconi, C., and Santoleri, R.: The Mediterranean Ocean Colour Observing System – system development and product validation, Ocean Sci., 8, 869–883, https://doi.org/10.5194/os-8-869-2012, 2012.

Volpe, G., Colella, S., Brando, V. E., Forneris, V., La Padula, F., Di Cicco, A., Sammartino, M., Bracaglia, M., Artuso, F., and Santoleri, R.: Mediterranean ocean colour Level 3 operational multi-sensor processing, Ocean Sci., 15, 127–146, https://doi.org/10.5194/os-15-127-2019, 2019.

Voss, K. J., Gordon, H. R., Flora, S., Johnson, B. C., Yarbrough, M., Feinholz, M., Houlihan, T.: A method to extrapolate the diffuse upwelling radiance attenuation coefficient to the surface as applied to the Marine Optical Buoy (MOBY), J. Atm. and Ocean. Tech., 34, 1423-1432 (2017), DOI: 10.1175/JTECH-D-16-0235.1, 2017.

Werdell, P.J., Bailey, S., Fargion, G., Pietras, C., Knobelspiesse, K., Feldman, G. and McClain, C.: Unique data repository facilitates ocean color satellite validation. EOS Transactions AGU, 84(38), 379, 2003.

Werdell, P.J. and Bailey, S. W.: An improved bio-optical ~~data set~~dataset for ocean color algorithm development and satellite data product validation. Remote Sensing of Environment, 98(1), 122-140, 2005.

Wiegmann, S., Liu, Y., and Bracher, A.: Spectrophotometric measurements of absorption coefficients by total particles, phytoplankton and non-algal particles during POLARSTERN cruise PS93.2, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.907605, 2019.

Xi, H., Wiegmann, S., Hohe, C., Schmidt, I., and Bracher, A.: A validation data set of phytoplankton pigment concentrations and phytoplankton groups measured on water samples collected from various expeditions, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.982433, 2025.

Zemp, M., Chao, Q., Han Dolman, A., Herold, M., Krug, T., Speich, S., Suda, K., Thorne, P., Yu, W., 2022. GCOS 2022 implementation plan. Global Climate Observing System GCOS 244, 85.

Zibordi, G., Hooker, S. B., Berthon, J. F., and D'Alimonte, D.: Autonomous above-water radiance measurements from an offshore platform: a field assessment experiment. Journal of Atmospheric and Oceanic Technology, 19(5), 808-819, 2002.

Zibordi, G., Holben, B.N., Hooker, S.B., Mélin, F., Berthon, J.-F., Slutsker, I., Giles, D., Vandemark, D., Feng, H., Rutledge, K., Schuster, G. and Al Mandoos, A.: A network for standardized ocean color validation measurements. EOS Trans. Am. Geophys. Union, 87, 30, 293, 297, 2006.

Zibordi, G., Holben, B.N., Slutsker, I., Giles, D., D'Alimonte, D., Mélin, F., Berthon, J.-F., Vandemark, D., Feng, H., Schuster, G., Fabbri, B.E., Kaitala, S. and Seppälä, J.: AERONET-OC: A network for the validation of ocean color primary radiometric products. J. Atmos. Ocean. Tech., 26, 1634-1651, 2009.

Zibordi, G., Holben, B. N., Talone, M., D'Alimonte, D., Slutsker, I., Giles, D. M. and Sorokin, M. G.: Advances in the Ocean Color Component of the Aerosol Robotic Network (AERONET-OC). J. Atmos. Ocean. Tech. 38(4), 725–746, 2021.Zhang, X., Hu, L. and He, M.-X.: Scattering by pure seawater: Effect of Salinity. Optics Express, 17, 5698-5710, 2009.

Zibordi, G., Holben, B.N., Talone, M., D'Alimonte, D., Slutsker, I., Giles, D.M., Sorokin, M.G., 2021. Advances in the Ocean Color Component of the Aerosol Robotic Network (AERONET-OC). J. Atmos. Ocean. Technol. 38, 725–746. https://doi.org/https://doi.org/10.1175/JTECH-D-20-0085.1

Zindler, C., Bracher, A., Marandino, C. A., Taylor, B. B., Torrecilla, E., Kock, A., and Bange, H. W.: Sulphur compounds, methane, and phytoplankton during SONNE cruise SO202/2 (Transbrom Sonne), PANGAEA [data set], https://doi.org/10.1594/PANGAEA.820607, 2013.

TABLES & TABLE CAPTIONS

Table 1: Data source temporal coverage between previous database version (v2022) and current version (v2026). Colored cells indicate the years for which observations are available from each data source. Orange-filled cells represent data included in the v2022, while blue-filled cells indicate newly added observations in the v2026. The color of the data source names identifies the update status of each dataset: black denotes data sources already included in previous versions with no additional data updates, green indicates existing sources that have been updated with new observations, and magenta marks newly added data sources introduced in the v2026 compilation.

ID	Data Source	Parameters							
1	MOBY Gold	Rrs							
2	AERONET-OC	Rrs							
3	AWI	Rrs	Chla_hplc		aph	adg		kd	
4	CALCOFI			Chla_fluor					
5	HOT		Chla_hplc	Chla_fluor					
6	PALMER		Chla_hplc	Chla_fluor					
7	SeaBASS	Rrs	Chla_hplc	Chla_fluor	aph	adg	bbp	kd	TSM
8	IMOS		Chla_hplc	Chla_fluor					
9	AMT		Chla_hplc	Chla_fluor					
10	BOUSSOLE	Rrs	Chla_hplc						
11	NOMAD	Rrs	Chla_hplc	Chla_fluor	aph	adg	bbp	kd	TSM
12	MERMAID	Rrs	Chla_hplc	Chla_fluor	aph	adg	bbp	kd	TSM
13	ICES		Chla_hplc	Chla_fluor					
14	GeP&CO		Chla_hplc	Chla_fluor					
15	ARCSSPP			Chla_fluor					
16	BARENTSSEA			Chla_fluor					
17	BATS		Chla_hplc	Chla_fluor					
18	BIOCHEM		Chla_hplc	Chla_fluor					
19	BODC		Chla_hplc	Chla_fluor					
20	CCELTER			Chla_fluor					
21	CIMT			Chla_fluor					
22	COASTCOLOUR	Rrs	Chla_hplc	Chla_fluor	aph	adg	bbp		TSM
23	ESTOC			Chla_fluor					
24	MAREDAT		Chla_hplc						
25	SEADATANET			Chla_fluor					
26	TPSS		Chla_hplc	Chla_fluor	aph				
27	TARA	Rrs	Chla_hplc						
28	MOBY Platinum	Rrs							
29	CNR_GOS	Rrs	Chla_hplc	Chla_fluor	aph		bbp	kd	TSM
30	DFO		Chla_hplc	Chla_fluor	aph				
31	GOAL		Chla_hplc						
32	iFADO		Chla_hplc						
33	INIDEP			Chla_fluor	aph				
34	IOPAN		Chla_hplc	Chla_fluor					
35	ISEE	Rrs		Chla_fluor	aph	adg	bbp		
36	KHOA		Chla_hplc	Chla_fluor					TSM
37	MEDGOOS		Chla_hplc					kd	
38	UCSD	Rrs	Chla_hplc						
39	UNH	Rrs	Chla_hplc		aph	adg			TSM
40	YNU	Rrs	Chla_hplc	Chla_fluor	aph	adg	bbp		

Table 2: The standard variables, nomenclatures and units in the final table.

Variable/Column	Description and units
idx	Unique key identifying each row
time	UTC, <YYYY-MM-DD>T<HH:MM:SS>Z
lat	Decimal degree, -90:90, South Negative
lon	Decimal degree, -180:180, West Negative
depth_water	Sampling depth (m) – all assigned to zero
chla_hplc	Total chlorophyll a concentration determined from HPLC method (mg m^{-3})
chla_fluor	Chlorophyll a concentration determined from fluorometric or spectrophotometric methods (mg m^{-3})
rrs_<band>	Remote-sensing reflectance (sr^{-1})
aph_<band>	Algal pigment absorption coefficient (m^{-1})
adg_<band>	Detritus plus gelbstoff "CDOM" absorption coefficient (m^{-1})
bbp_<band>	Particle backscattering coefficient (m^{-1})
kd_<band>	Diffuse attenuation coefficient for downward irradiance (m^{-1})
tsm	Total suspended matter (g m^{-3})
etopo1	Water depth from ETOPO1 (m)
chla_hplc_dataset	Metadata string for "chla_hplc"
chla_hplc_subdataset	Metadata string for "chla_hplc"
chla_hplc_contributor	Metadata string for "chla_hplc"
chla_fluor_dataset	Metadata string for "chla_fluor"
chla_fluor_subdataset	Metadata string for "chla_fluor"
chla_fluor_contributor	Metadata string for "chla_fluor"
rrs_dataset	Metadata string for "rrs"
rrs_subdataset	Metadata string for "rrs"
rrs_contributor	Metadata string for "rrs"
aph_dataset	Metadata string for "aph"
aph_subdataset	Metadata string for "aph"
aph_contributor	Metadata string for "aph"
adg_dataset	Metadata string for "adg"
adg_subdataset	Metadata string for "adg"
adg_contributor	Metadata string for "adg"
bbp_dataset	Metadata string for "bbp"
bbp_subdataset	Metadata string for "bbp"
bbp_contributor	Metadata string for "bbp"
kd_dataset	Metadata string for "kd"
kd_subdataset	Metadata string for "kd"
kd_contributor	Metadata string for "kd"
tsm_dataset	Metadata string for "tsm"
tsm_subdataset	Metadata string for "tsm"
tsm_contributor	Metadata string for "tsm"
Exact_Wavelengths(um)_< band>	Exact wavelength (µm) associated with each AERONET-OC rrs band
flag_time	"1" if observation without time (set to 12:00:00 UTC)
flag_chl_method	"1" if observation as unknown chlorophyll method

Table 3: Original sets of data and data contributors in the final table.

Data Source	Description	Data contributors
Marine Optical Buoy (MOBY) Gold	Daily observations of remote-sensing reflectance, measured by a fixed mooring system, located west of the Hawaiian Island of Lanai. Data compiled between 1997-2022. Data were obtained from the MOBY website. Compiled standard variable: "rrs".	Paul DiGiacomo, Kenneth Voss
Marine Optical Buoy (MOBY) Platinum	MOBY Platinum extends the MOBY time series with improved instrumentation from May 2024 onward, including measurements from the MOBY (Hawaii) and MarONET (Western Australia) sites. It uses a dual in-line spectrometer (DILS) system for simultaneous radiance and irradiance measurements, with a "Best Lw Product" indicator to select the optimal estimate. The compiled variable was remote-sensing reflectance ("rrs").	Kenneth Voss
Bouée pour l'acquisition d'une Série Optique à Long Terme (BOUSSOLE)	High frequency (15 min) observations of remote-sensing reflectance, from a fixed mooring system, located in the Western Mediterranean Sea. Measurements of chlorophyll-a concentration are also available at the mooring locations. Remote-sensing reflectance and chlorophyll-a data were compiled between 2003-2019 and 2001-2020, respectively. Data were provided by David Antoine. Compiled standard variables: "rrs", "chla_hplc".	David Antoine, Vicenzo Vellucci
AErosol RObotic NETwork-Ocean Color (AERONET-OC)	Daily observations of remote-sensing reflectance, measured by modified sun-photometers. Data compiled between 2002-2024 from 39 sites, including: AAOT, ARIAKE TOWER, Abu Al Bukhoosh, Bahia Blanca, Banana River, Blyth NOAH, COVE SEAPRISM, Casablanca Platform, Chesapeake Bay, GOT Seaprism, Gageocho Station, Galata Platform, Gloria, Grizzly Bay, Gustav Dalen Tower, Helsinki Lighthouse, Ieodo Station, Irbe Lighthouse, Kemigawa Offshore, LISCO, Lake	Ana Inés Dogliotti, Burton Jones & Matthew Ragan, Dimitry Van der Zande, Dirk Aurin, Douglas Vandemark & Hui Feng & Heidi M. Sosik, Giuseppe Zibordi, Giuseppe Zibordi & Marco Talone, Hiroto Higa, Hubert Loisel & Paula Pratolongo & Robert Frouin, Jae-Seol Shim & Joo-Hyung Ryu, Joji Ishizaka & Kohei Arai, Marco Talone, Nima Pahlevan, Pawan Gupta & Elena Lind, Rodney Forster, Sam Ahmed & Alex Gilerson, Sherwin Ladner,

	Erie, Lake Okeechobee, Lake Okeechobee N, Lucinda, MVCO, PLOCAN Tower, Palgrunden, RdP-EsNM, Sacramento River, San Marco Platform, Section-7 Platform, Socheongcho, South Green Bay, Thornton C-power, USC SEAPRISM, USC SEAPRISM 2, Venise, WaveCIS Site CSI 6, and Zeebrugge-MOW1. Data were obtained from the AERONET-OC website. Compiled standard variable: "rrs".	Susanne Kratzer, Thomas Schroeder, Tim Moore & Steve Ruberg & Menghua Wang, Young-Je Park, Young-Je Park & Hak-Yeol You
SeaWiFS Bio-optical Archive and Storage System (SeaBASS)	Global archive of in situ marine data from multiple contributors. Bio-optical global data between 1997-2025 were extracted from the SeaBASS website. Pigment data were mostly extracted using "Pigment Search" tool, which provides data directly from the archives. Radiometric data were extracted using "Validation" tool, which only provides in situ data with matchups for ocean colour sensors. Compiled standard variables: "rrs", "chla_hplc", "chla_fluor", "aph", "adg", "bbp", "kd".	Robert Arnone, James Allen, Kevin Arrigo, Dirk Aurin, William Balch, Ray Barlow, Mike Behrenfeld, Sukru Besiktepe, Kelsey Bisson, Emmanuel Boss, Chris Brown, Dylan Catlett, Douglas Capone, Ken Carder, Carlos Del Castillo, Francisco Chavez, Alex Chekalyuk, Jay-Chung Chen, Dennis Clark, Herve Claustre, Lesley Clementson, Javier Concha, Jorge Corredor, Glenn Cota, Yves Dandonneau, Heidi Dierssen, David Eslinger, Piotr Flatau, Scott Freeman, Robert Frouin, Carlos Garcia, Alex Gilerson, Joaquim Goes, Gwo-Ching Gong, Adriana Gonzalez-Silvera, Rick Gould, Jason Graff, Nils Haentjens, Larry Harding, Jon Hare, Stanford B. Hooker, Chuanmin Hu, Milton Kampel, Sung-Ho Kang, Grace Kim, Gary Kirkpatrick, Oleg Kopelevich, Sasha Kramer, Sam Laney, Pierre Larouche, Jesus Ledesma, Zhongping Lee, Ricardo Letelier, Marlon Lewis, Steven Lohrenz, Mary Luz Canon, Antonio Mannino, Stephane Maritorena, John Marra, Chuck McClain, Christophe Menkes, Mark Miller, Allen Milligan, Greg Mitchell, Ru Morrison, James Mueller, Frank Muller-Karger, Ruben Negri, James Nelson, Norman Nelson, Michael Novak, Mary Jane Perry, David Phinney, John Porter, Collin Roesler, Joe Salisbury, David Siegel, Mike Sieracki, Jeffrey Smart, Raymond Smith, Heidi Sosik, James Spinhirne, Dariusz Stramski, Rick Stumpf, Ajit Subramaniam, Lynne Talley, Chuck Trees, Michael Twardowski, Ryan Vandermeulen, Kenneth Voss, Marcel Wernand, Toby Westberry, Ronald

		Zaneveld, Eric Zettler, Giuseppe Zibordi, Richard Zimmerman
NASA bio-Optical Marine Algorithm Data set (NOMAD)	High-quality global data set of coincident bio-optical in situ data. The data set was built upon SeaBASS archive. The current version (Version 2.0 ALPHA, 2008) was used, with an additional set of columns of remote-sensing reflectance corrected for the bidirectional nature of the light field, provided by NOMAD creators. Data compiled between 1997-2007. Compiled standard variables: "rrs", "chla_hplc", "chla_fluor", "aph", "adg", "bbp", "kd".	Robert Arnone, Kevin Arrigo, William Balch, Ray Barlow, Mike Behrenfeld, Chris Brown, Douglas Capone, Ken Carder, Francisco Chavez, Dennis Clark, Herve Claustre, Jorge Corredor, Glenn Cota, David Eslinger, Piotr Flatau, Robert Frouin, Rick Gould, Larry Harding, Stanford B. Hooker, Oleg Kopelevich, Marlon Lewis, Antonio Mannino, John Marra, Mark Miller, Greg Mitchell, Tiffany Moisan, Ru Morrison, Frank Muller-Karger, James Nelson, Norman Nelson, David Siegel, Raymond Smith, Timothy Smyth, James Spinhirne, Dariusz Stramski, Rick Stumpf, Ajit Subramaniam, Kenneth Voss
MERIS Match-up In situ Database (MERMAID)	Global database of in situ bio-optical data matched with concurrent MERIS Level 2 satellite ocean colour products The "Extract matchup" tool to acquire data was used. Data was compiled between 2002-2012. Access has been granted through a signed Service Level Agreement. Compiled standard variables: "rrs", "chla_hplc", "chla_fluor", "aph", "adg", "bbp", "kd", "tsm".	Simon Belanger, Jean-Francois Berthon, Vanda Brotas, Elisabetta Canuti, Pierre Yves Deschamps, Annelies Hommersom, Mati Kahru, Holger Klein, Susanne Kratzer, Hubert Loisel, David McKee, Greg Mitchell, Michael Ondrusek, Michel Repecaud, David Siegel, Gavin Tilstone, Giuseppe Zibordi
Atlantic Meridional Transect (AMT)	Multidisciplinary programme that makes biological, chemical and physical oceanographic measurements during an annual voyage between the United Kingdom and destinations in the South Atlantic. It has compiled observations of chlorophyll-a concentration between 1997 (AMT5) and 2018 (AMT28). Data were provided by the British Oceanographic Data Centre (BODC) and directly from data contributors. Compiled standard variables: "chla_hplc", "chla_fluor".	Ruth Airs, Arwen Bargery, Ray Barlow, Robert J. W. Brewin, Denise Cummings, Giorgio Dall'Olmo, Ella Darlington, Afonso Ferreira, Stuart Gibb, Victoria Hill, Patrick Holligan, Victor Martinez-Vincente, Gerald Moore, Leonie O'Dowd, Alex Poulton, Emilio Suarez, Glen Tarran, Andreia Tracana, Rob Thomas, Gavin Tilstone
International Council for the Exploration of the Sea (ICES)	Database of several collections of data related to the marine environment. It has compiled observations of chlorophyll-a concentration in the northern European Seas, between 1997-2012. Data were provided by the ICES database on the marine environment (2014, Copenhagen,	Not Available

	Denmark). Compiled standard variables: "chla_hplc", "chla_fluor".	
Hawaii Ocean Time-series (HOT)	Multidisciplinary programme that makes repeated biological, chemical and physical oceanographic observations near Oahu, Hawaii. Measurements of chlorophyll-a concentration between 1997-2023 were extracted from the project website. Compiled standard variables: "chla_hplc", "chla_fluor".	Bob Bidigare, Matthew Church, Ricardo Letelier, Jasmine Nahorniak
Geochemistry, Phytoplankton, and Color of the Ocean (GeP&CO)	Program of in situ data collection aboard merchant ship from France to New Caledonia, between 1999 and 2002. Measurements of chlorophyll-a concentration were obtained from the project website. Compiled standard variables: "chla_hplc", "chla_fluor".	Yves Dandonneau
ARCSSPP	"Arctic System Science Primary Production" database. Available from NODC FTP site. Compiled standard variable: "chla_fluor".	Patricia Matrai
AWI	Several 2007-2023 cruises in Atlantic, Pacific and Southern Ocean from Astrid Bracher's group at AWI. Provided by Astrid Bracher. Available from PANGAEA. Compiled standard variables: "chla_hplc", "rrs", "aph", "adg", "kd".	Astrid Bracher, Rüdiger Röttgers
BARENTSSEA	Data collection from cruises of the Institute of Marine Research (Norway) mainly around the Barents Sea. Provided by Knut Yngve Børsheim. Compiled standard variable: "chla_fluor"	Knut Yngve Børsheim
BATS	Data collection from the "Bermuda Atlantic Time-series Study". Available from BATS website. Compiled standard variables: "chla_fluor", "chla_hplc"	Not Available
BIOCHEM	The Fisheries and Oceans Canada database for biological and chemical data. Mostly data from Gulf of St. Lawrence. Available from BIOCHEM website. Compiled standard variable: "chla_fluor" and "chla_hplc".	Diane Archambault, Hughes Benoit, Esther Bonneau, Eugene Colbourne, Alain Gagne, Yves Gagnon, Tom Hurlbut, Catherine Johnson, Pierre Joly, Maurice Levasseur, Patrick Ouellet, Jacques Plourde, Luc Savoie, Michael Scarratt, Philippe Schwab, Michel Starr, François Villeneuve,
BODC	"British Oceanographic Data Centre". Mainly European Seas. Provided by BODC. Compiled standard variables: "chla_fluor", "chla_hplc"	Not Available
CALCOFI	Cruise data from the "California Cooperative	Ralf Goericke

	Oceanic Fisheries Investigations" program. Available from CalCOFI website. Compiled standard variable: "chla_fluor".	
CCELTER	Cruise data from "California Current Ecosystem Long Term Ecological Research". Available from CCELTER website. Compiled standard variable: "chla_fluor".	Ralf Goericke
CIMT	Sampling from the "Center for Integrated Marine Technology" (California). Available from CIMT website. Compiled standard variable: "chla_fluor".	Raphael Kudela
COASTCOLOUR	Quality controlled compilation of bio-optical data in several coastal sites. Available from PANGAEA. Compiled standard variables: "chla_fluor", "chla_hplc", "rrs", "aph", "adg", "bbp", "tsm".	Not Available
ESTOC	Sampling from the "Estación Europea de Series Temporales del Oceano" Canary Islands. Provided by Andrés Cianca. Compiled standard variable: "chla_fluor".	Octavio Llinas and Andres Cianca
IMOS	Australian National Reference Stations – Phytoplankton HPLC Pigment Composition Analysis. Available from the Australian Ocean Data Network (AODN). Compiled standard variable: "chla_hplc". Bio-optical Database of Australian Waters. Available from the Australian Ocean Data Network (AODN). Compiled standard variables: "chla_hplc", "chla_fluor"	Janet Anstee, David Antoine, Russ Babcock, Ina Barton, Philip Boyd, Nagur Cherukuru, Lesley Clementson, Joey Croswell, Diana Davies, Peter Davis, Arnold Dekker, Michelle Devlin, Martina Doblin, Ruth Eriksen, Rene Gruber, Christine Hanson, Christina Howley, John Keesing, Rudy Kloser, Leonardo Laiolo, Richard Matear, James McLaughlin, Dan Moran, Dominique O'Brien, John Parslow, Eric Raes, Charlotte Robinson, Barbara Robson, Christina Schallenberg, Thomas Schroeder, Elizabeth Shadwick, Bernadette Sloyan, Monika Soja-Wozniak, Andy Steven, Pete Strutton, David Suggett, Hugh Sweatman, Paul Thomson, Bronte Tilbrook, Tom Trull, John Volkman, Jane Waterhouse, Karen Wild-Allen, Bozena Wojtasiewicz, Simon Wright, Cathryn Wynn-Edwards, Jock Yong.
MAREDAT	Quality controlled global compilation of chla	Ray Barlow, Robert Bidigare, Herve

	HPLC. Available from PANGAEA. Compiled standard variable: "chla_hplc".	Claustre, Denise Cummings, Giacomo DiTullio, Chris Gallienne, Ralf Goericke, Patrick Holligan, David Karl, Michael Landry, Michael Lomas, Michael Lucas, Jean-Claude Marty, Walker Smith, Denise Smythe-Wright, Rick Stumpf, Emilio Suarez, Koji Suzuki, Maria Vernet, Simon Wright
PALMER	"Palmer station Long-Term Ecological Research" (Antarctica). Available from PALMER website. Compiled standard variables: "chla_fluor", "chla_hplc".	Oscar Schofield, Raymond Smith, Maria Vernet
SEADATANET	Global archive of in situ marine data. Available from SEADATANET website. Compiled standard variable: "chla_fluor".	Not Available
TPSS	Compilation of bio-optical data predominantly from the Northwest Atlantic, but also from the Indian Ocean, South Pacific and Central Atlantic. Provided by Trevor Platt and Shubha Sathyendranath. Compiled standard variables: "chla_hplc", "chla_fluor", "aph".	Trevor Platt, Shubha Sathyendranath.
TARA	Data collection from the TARA global transects. Provided by Emmanuel Boss. All data available in SeaBASS. Compiled standard variables: "chla_hplc", "rrs".	Emmanuel Boss
National Research Council – Group of Satellite Oceanography (CNR_GOS)	Bio-optical observations collected in the Mediterranean Sea (1999–2024) from multiple research cruises. Measurements include radiometry, chlorophyll-a, and inherent optical properties derived following standard protocols. Compiled variables: "rrs", "kd", "chla_hplc", "tsm", "aph", "adg", "bbp".	Jaime Pitarch, Marco Bellacicco, Francesco Bignami, Federica Braga, Vittorio E. Brando, Simone Colella, Annalisa Di Cicco, Federico Falcini, Ivan Farace, Salvatore Marullo, Emanuele Organelli, Michela Sammartino, Rosalia Santoleri, Gian Marco Scarpa, Gianluca Volpe, Maria Laura Zoffoli, Florinda Artuso, Isabella Giardina, Antonia Lai, Rosario Lavezza, Francesca Margiotta, Mariella Ragni, Maurizio Ribera d'Alcalà, Vincenzo Vellucci
Department of Fisheries and Oceans Canada (DFO)	Long-term observations from the Northwest Atlantic (2000–2022; AZMP and AZOMP), including chlorophyll-a (HPLC and fluorometric) and phytoplankton absorption from multi-decadal monitoring programs. Compiled variables were	Emmanuel Devred

	"chla_hplc", "chla_fluor", and "aph".	
Grupo de Oceanografia de Altas Latitudes (GOAL)	Chlorophyll-a observations from austral summer campaigns in the northern Antarctic Peninsula (2008–2020), providing consistent coverage of high-latitude waters. Compiled variables: "chla_hplc"	Carlos Rafael Borges Mendes, Raul Rodrigo Costa
Innovation in the Framework of the Atlantic Deep Ocean (iFADO)	The iFADO project provides chlorophyll-a measurements derived from HPLC, collected in coastal waters along the Spanish and Portuguese Atlantic coasts and provided via Vanda Brotas. Compiled variable: "chla_hplc".	Vanda Brotas, Andreia Tracana, Manuela Ruiz-Villarreal, Tamara Rodríguez-Ramos, María-Manuel Angélico, Paulo B. Oliveira, Ana Moreno, Cristina Nunes
Instituto Nacional de Investigación y Desarrollo Pesquero (INIDEP)	Observations from multiple research cruises in the southwestern Atlantic (2001–2022), including chlorophyll-a and phytoplankton absorption measurements across diverse oceanographic conditions. Compiled variables: "chla_fluor" and "aph".	Valeria Segura, Vivian Lutz
Institute of Oceanology, Polish Academy of Sciences (IOPAN)	Chlorophyll-a observations from the Baltic Sea (2013–2020), including largely consistent HPLC and fluorometric measurements in optically complex waters. Compiled variables: "chla_hplc" and "chla_fluor".	Piotr Kowalczuk, Joanna Stoń-Egiert, Katarzyna Dragańska-Deja, Justyna Meler, Monika Zabłocka
Institute for Space-Earth Environmental Research (ISEE)	Long-term observations (2001–2019) from Japanese coastal waters (Ariake Sea and Ise–Mikawa Bay), including co-located Rrs and IOPs suitable for algorithm development and satellite validation in optically complex environments. Compiled variables: chla_fluor, rrs, aph, adg, bbp	Joji Ishizaka
Korea Hydrographic and Oceanographic Agency (KHOA)	Field survey data from Korean regional seas (2022–2025), including chlorophyll-a (HPLC and fluorometric) and total suspended matter with associated station metadata. Compiled variable: "chla_hplc", "chla_fluor", "tsm"	chla_hplc, chla_fluor, tsm
Mediterranean Global Ocean Observing System (MEDGOOS)	Data collected in the Mediterranean Sea (2004–2008) during cruises conducted by the University of Florence aboard the CNR's R/V Urania. Chlorophyll-a was measured by HPLC and diffuse attenuation coefficients were derived from radiometric profiles following standard	Emanuele Organelli, Luca Massi, Chiara Melillo, Francesca Polonelli

	protocols. Compiled variable: "chla_hplc", "kd"	
University of California San Diego (UCSD)	Measurements from Pacific Ocean campaigns (2015–2022; OUTPACE and SOKOWASA), including rrs from HyperPro profiling and chlorophyll-a observations at corresponding stations. Compiled variables were "chla_hplc" and "rrs".	Robert Frouin , Jing Tan
University of New Hampshire (UNH)	Measurements from the Great Bay Estuary (USA; 2022–2024), including coincident rrs, inherent optical properties, chlorophyll-a (HPLC), and suspended matter in optically complex estuarine waters. Compiled variables were "chla_hplc", "rrs", "tsm", "aph", and "adg".	Atsushi Matsuoka
Yokohama National University (YNU)	Measurements from coastal and inland waters in Japan (Tokyo Bay, Seto Inland Sea, Lake Kasumigaura) collected between 2010–2021, including chlorophyll-a, Rrs, and inherent optical properties in optically complex environments. Compiled variable: "chla_hplc", "chla_fluor", "rrs", "aph", "adg", "bbp"	Hiroto Higa

Table 4. Central wavelengths of the satellite ocean-colour bands and the corresponding MOBY wavelengths mapped in the final merged table (nm) of this database compilation.

Sensor	Satellite band centers (nm)	MOBY mapped wavelengths in merged database (nm)	No. of MOBY Mapped bands
SeaWiFS	412, 443, 490, 510, 555, 670, 765, 865	412, 443, 490, 510, 555, 670	1–6
MODIS-AQUA	412, 443, 488, 531, 547, 667, 678, 748, 869	416, 442, 489, 530, 547, 665, 677	1–7
MERIS	412, 442, 490, 510, 560, 620, 665, 681, 709, 753, 779, 865, 885	410.5, 440.4, 487.8, 507.7, 557.6, 617.5, 662.4, 679.9, 706.2, 752.5	1–10
VIIRS-SNPP	410, 443, 486, 551, 671, 746	412.9, 444.5, 481.2, 556.3, 674.6	1–5
VIIRS-NOAA20	411, 445, 489, 556, 667, 746	411, 445, 489.01, 556, 667	1–5
VIIRS-NOAA21	411, 445, 489, 556, 667, 746	411.051, 444.895, 488.363, 555.333, 671.366, 747.317	1–6
OLCI-S3A	400, 412, 443, 490, 510, 560, 620, 665, 674, 681, 709, 754, 779, 865, 885	400.3032, 411.8453, 442.9626, 490.493, 510.4676, 560.4503, 620.4092, 665.2744, 674.0251, 681.5705, 709.1149, 754.1813	1–12
OLCI-S3B	400, 412, 443, 490, 510, 560, 620, 665, 674, 681, 709, 754, 779, 865, 885	400.5947, 411.9509, 442.9882, 490.3991, 510.4022, 560.3664, 620.284, 665.1312, 673.8682, 681.3856, 708.9821, 754.0284	1–12

Table 5. Summary of median values for "aph", "adg", "bbp" and "kd" at 44X and 55X nm for each data set (as shown in Fig. 12 a-h). Data was first searched at 445 and 555 nm, and then with a search window up to 8 nm, to include data at 547 nm.

	Median "aph"		Median "adg"		Median "bbp"		Median "kd"	
	44x nm	55x nm	44x nm	55x nm	44x nm	55x nm	44x nm	55x nm
SeaBASS	0.0586	0.01	0.6281	0.1318	0.0051	0.0043	0.1990	0.1327
MERMAID	0.0282	0.0052	0.1149	0.0286	0.0080	0.0052	0.1394	0.1197
NOMAD	0.0353	0.0046	0.0515	0.0112	0.0030	0.0022	0.0837	0.1060
COASTCOLOUR	0.0665	0.0096	0.1259	0.0175	0.0047	0.0037	–	–
MEDGOOS	–	–	–	–	–	–	0.0685	0.1125
AWI	0.0241	0.0046	0.0432	0.0074	–	–	0.0428	0.0874
TPSS	0.0454	0.0071	–	–	–	–	–	–
CNR_GOS	0.011	0.0021	–	–	0.0028	0.0019	0.0509	0.0834
DFO	0.0368	0.0039	–	–	–	–	–	–
INIDEP	0.0401	0.0066	–	–	–	–	–	–
ISEE	0.2426	0.0533	0.3987	0.112	0.0002	0.0001	–	–
UNH	0.1166	0.0306	0.8225	0.1937	–	–	–	–
YNU	0.6393	0.1699	0.623	0.1862	0.0308	0.0247	–	–

FIGURES AND CAPTIONS

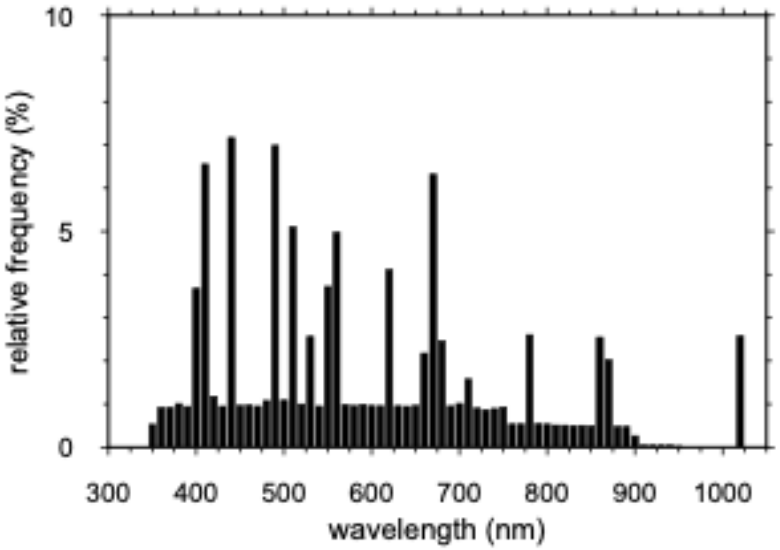

Figure 1. Relative spectral frequency of remote-sensing reflectance in the final table, using 10 nm wide class intervals, defined as the ratio of the number of observations at a particular waveband to the total number of observations at all wavebands, multiplied by 100 to report results in percentage. Data at a total of 2,994 unique wavelengths, between 301.22 nm and 1020.0 nm, were compiled.

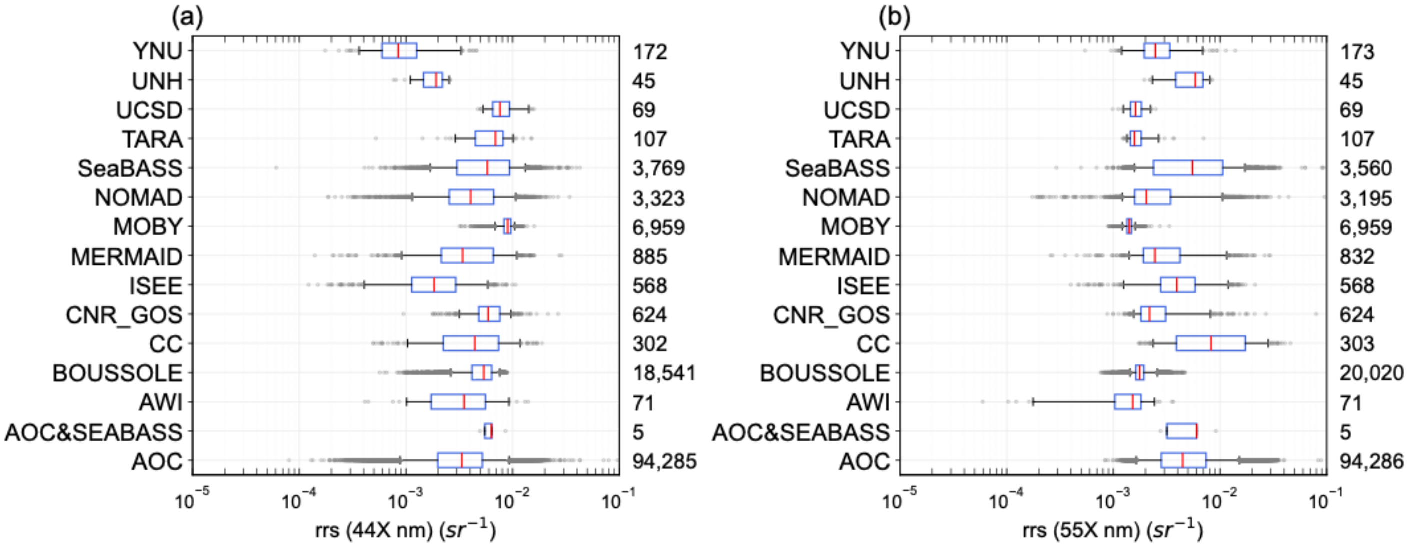

Figure 2. The distribution of (a) "rrs" at 44X nm and (b) "rrs" at 55X nm. Data were first searched at 445 and 555 nm, and then with a search window of up to 8 nm, to include data at 547 nm. The black boxes delimit the percentiles 0.25 and 0.75 of the data and the black horizontal lines show the extension of up to percentiles 0.05 and 0.95. The red line represents the median value and the black circles the values below (and above) the percentile 0.05 (0.95). The number of measurements of each data set is reported on the right axis of the graph.

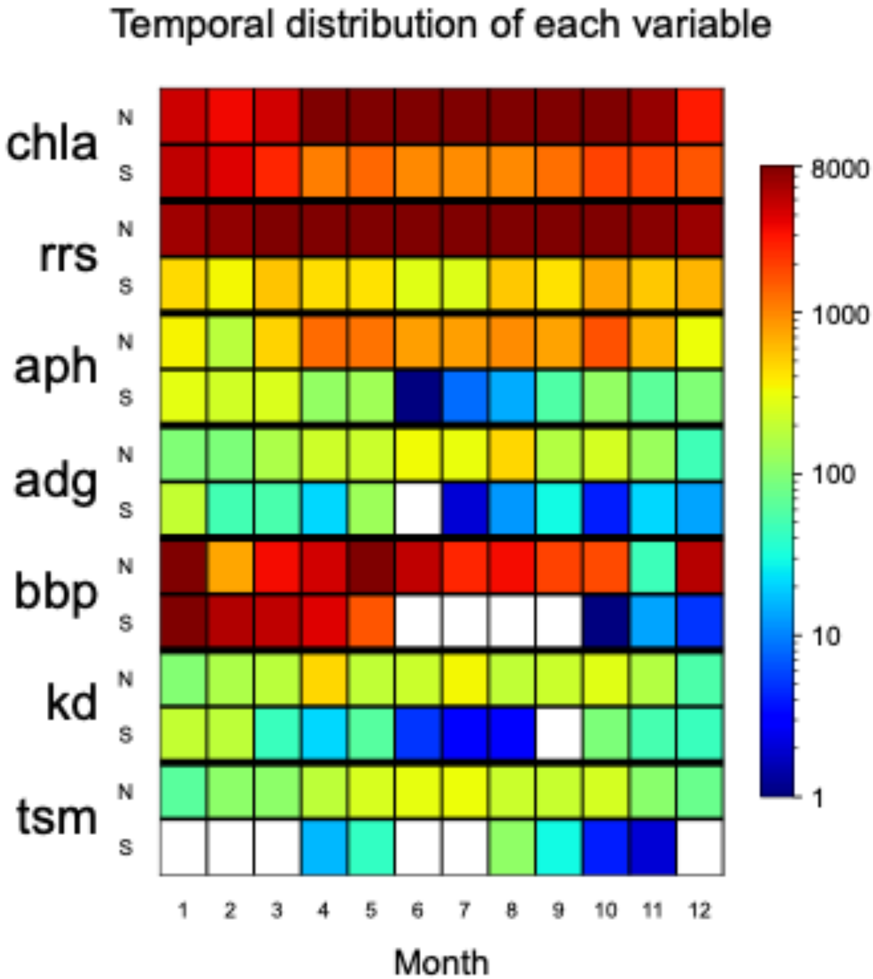

Figure 3. Temporal distribution of chlorophyll-a concentration ("chla"), remote-sensing reflectance ("rrs"), algal pigment absorption coefficient ("aph"), detrital plus CDOM absorption coefficient ("adg"), particle backscattering coefficient ("bbp"), the diffuse attenuation coefficient for downward irradiance ("kd") and total suspended matter ("tsm") in the final table. All chlorophyll data were considered, but for a given station, HPLC data were selected if available. Colours indicate the number of sampling events available for each variable, as a function of month and hemisphere of data acquisition ("N" - Northern Hemisphere; "S" - Southern Hemisphere). The empty (white) squares indicate no data for that month.

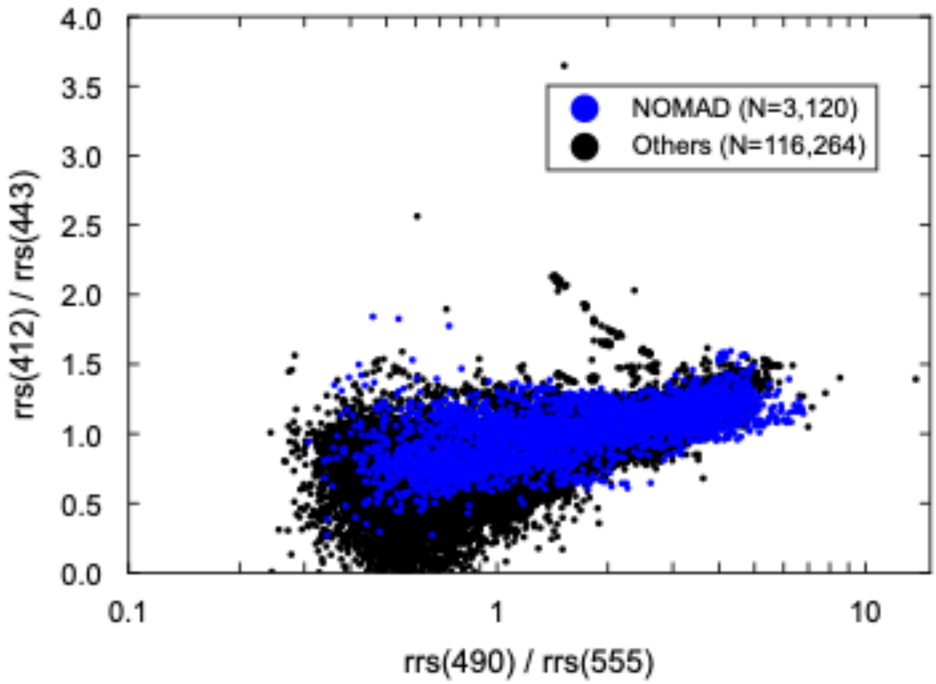

Figure 4. Ranges of remote-sensing reflectance band ratios (412:443 and 490:555) for all data. The points from the NOMAD data set are shown in blue for reference. To maximize the number of ratios per data set, a search window up to 12 nm was used, when the four wavelengths (412, 443, 490, 555) were not simultaneously available. The effect of different search windows was negligible in the ratio distribution.

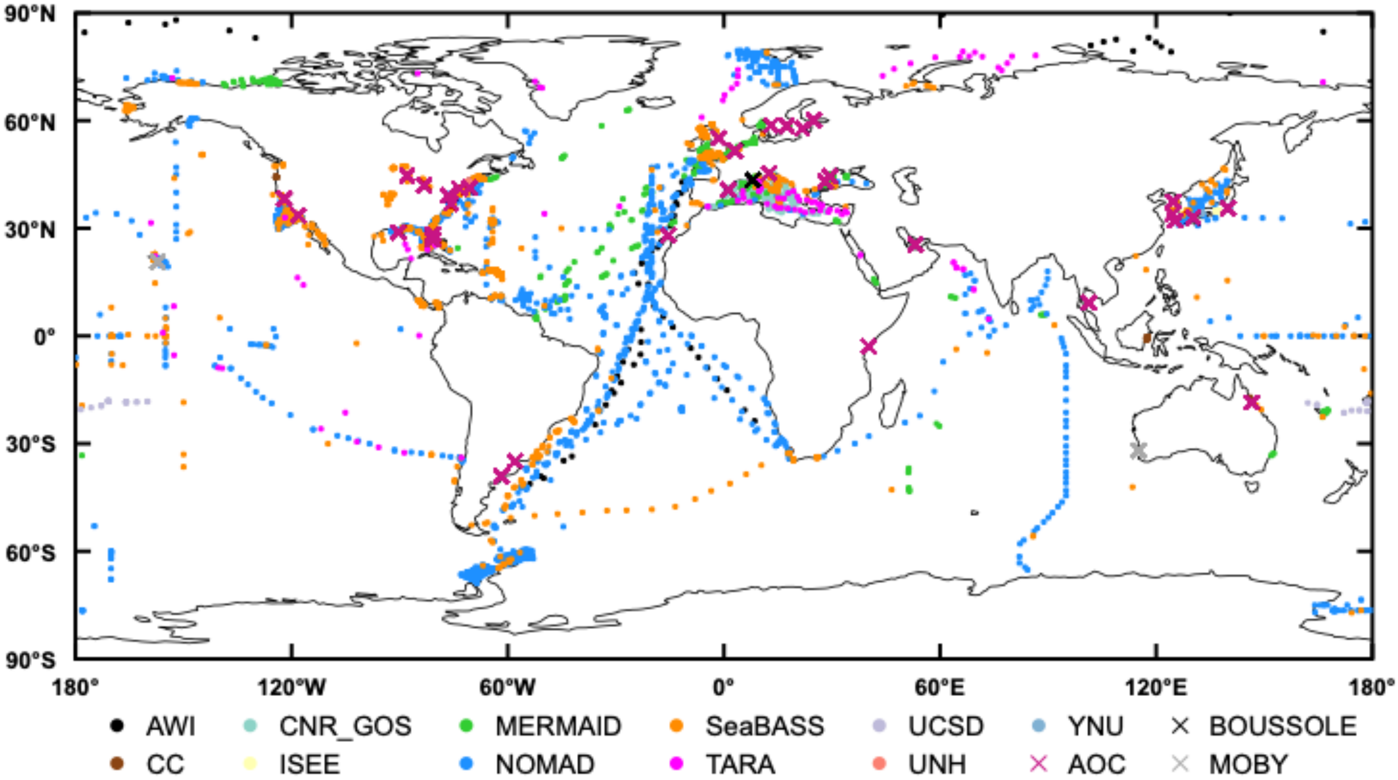

Figure 5. Global distribution of remote-sensing reflectance per data set in the final table. The data sources are identified with different colours. Points show locations where at least one observation is available. Crosses show sites from where time series data of remote-sensing reflectance are available.

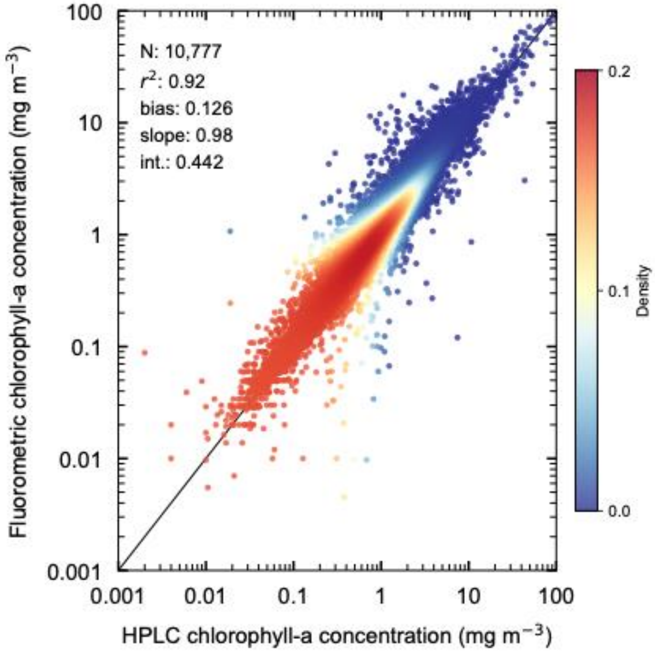

Figure 6. Comparison of coincident observations of chlorophyll-a concentration derived with different methods ("chla_hplc" and "chla_fluor"). The data were transformed prior to regression analysis to account for their log-normal distribution.

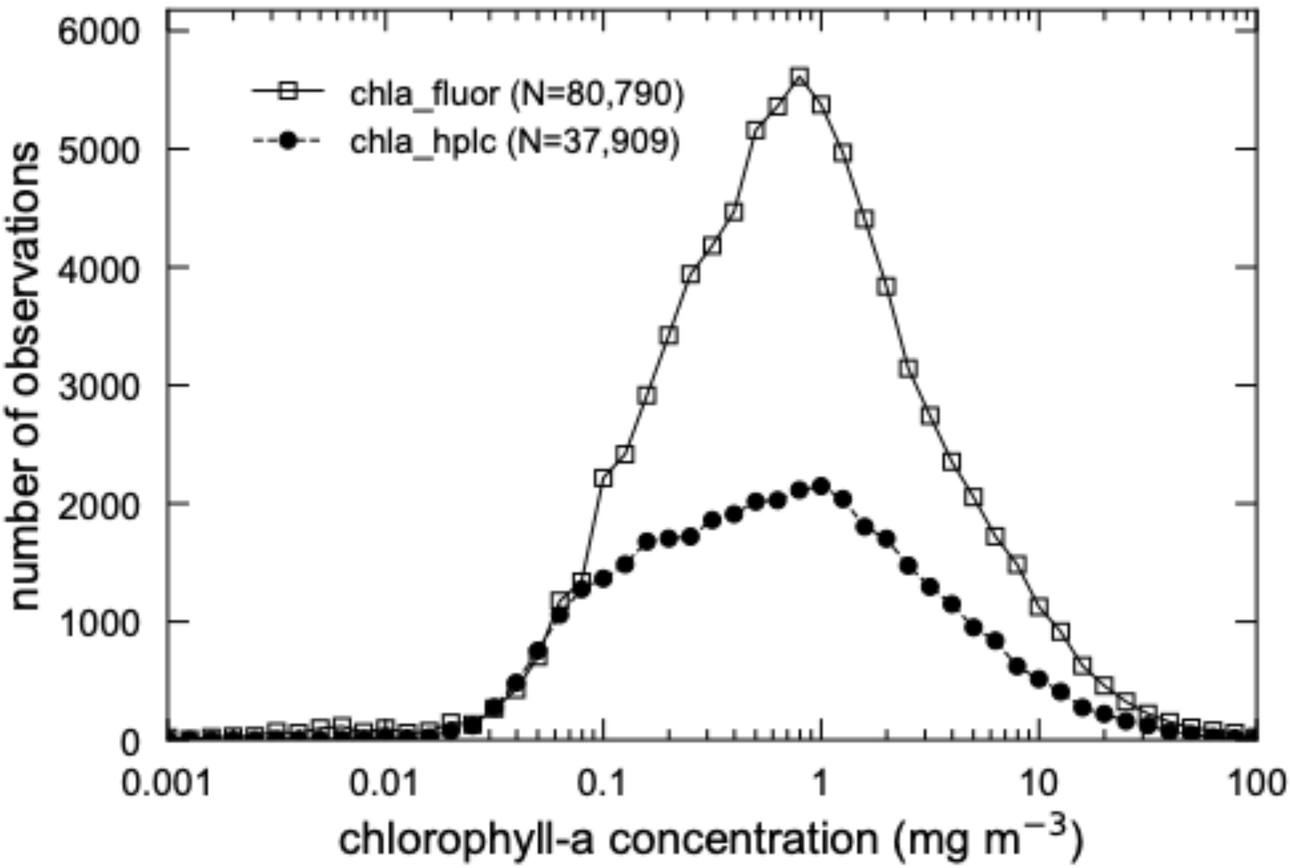

Figure 7. Number of observations per chlorophyll-a concentration acquired with different methods ("chla_fluor" and "chla_hplc").

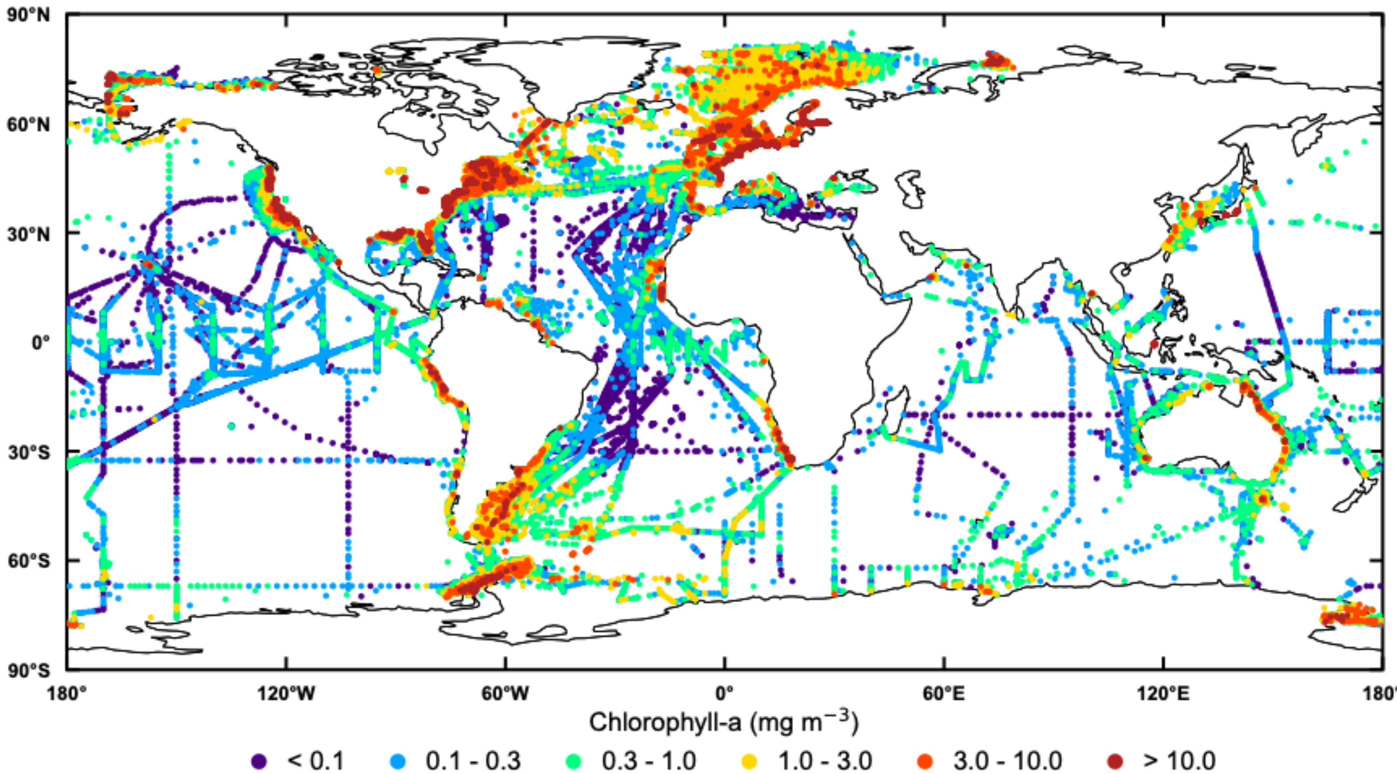

Figure 8. Global distribution of chlorophyll-a concentration per interval of the observed value. All chlorophyll-a data were considered, but for a given station, HPLC data were selected if available.

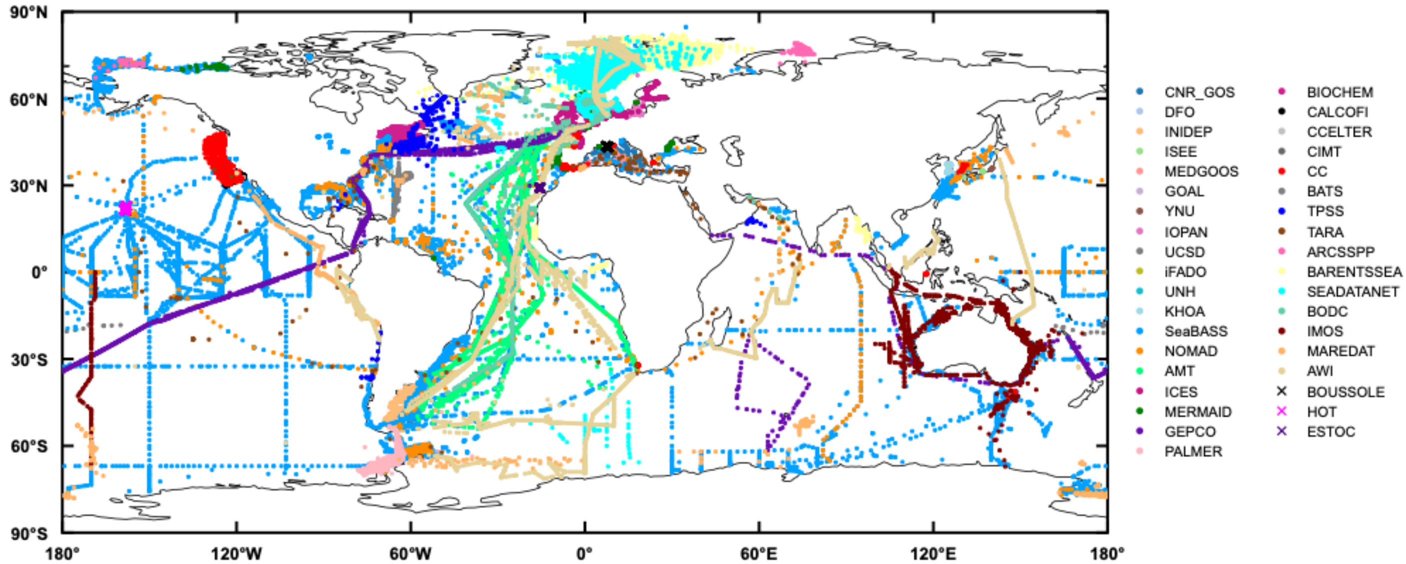

Figure 9. Global distribution of chlorophyll-a concentration per data set in the final table. All chlorophyll-a data were considered, but for a given station, HPLC data were selected if available. Crosses show sites from where data of chlorophyll-a are available in a specific geographic location.

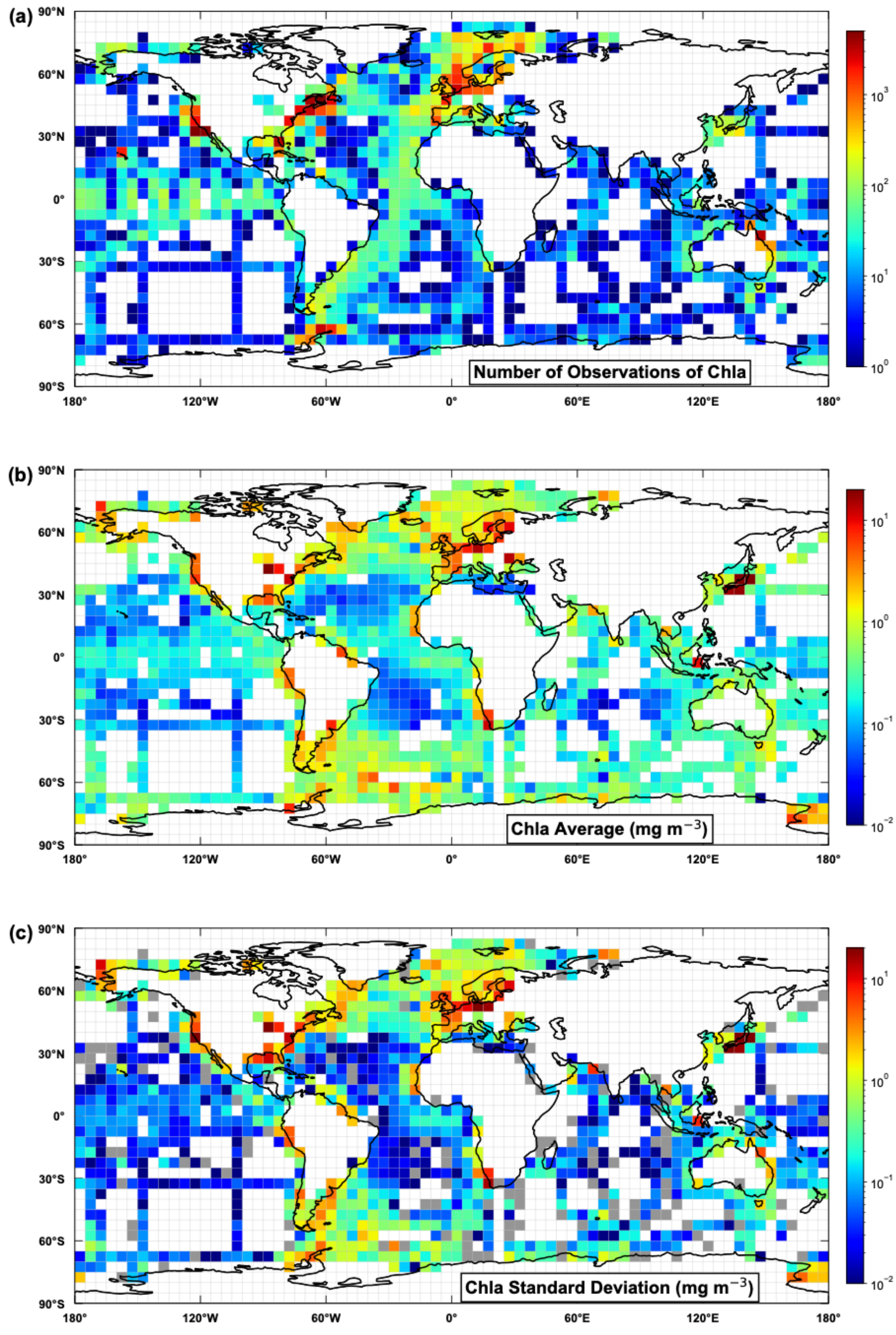

Figure 10. The chlorophyll-a (mg m^{-3}) data partitioned into 5º x 5º boxes showing: (a) number of observations, (b) average value and (c) standard deviation in each box. All chlorophyll-a data were considered, but for a given station, HPLC data were selected if available. In the standard deviation plot, grey colour boxes represent zero standard deviation (i.e., one observation).

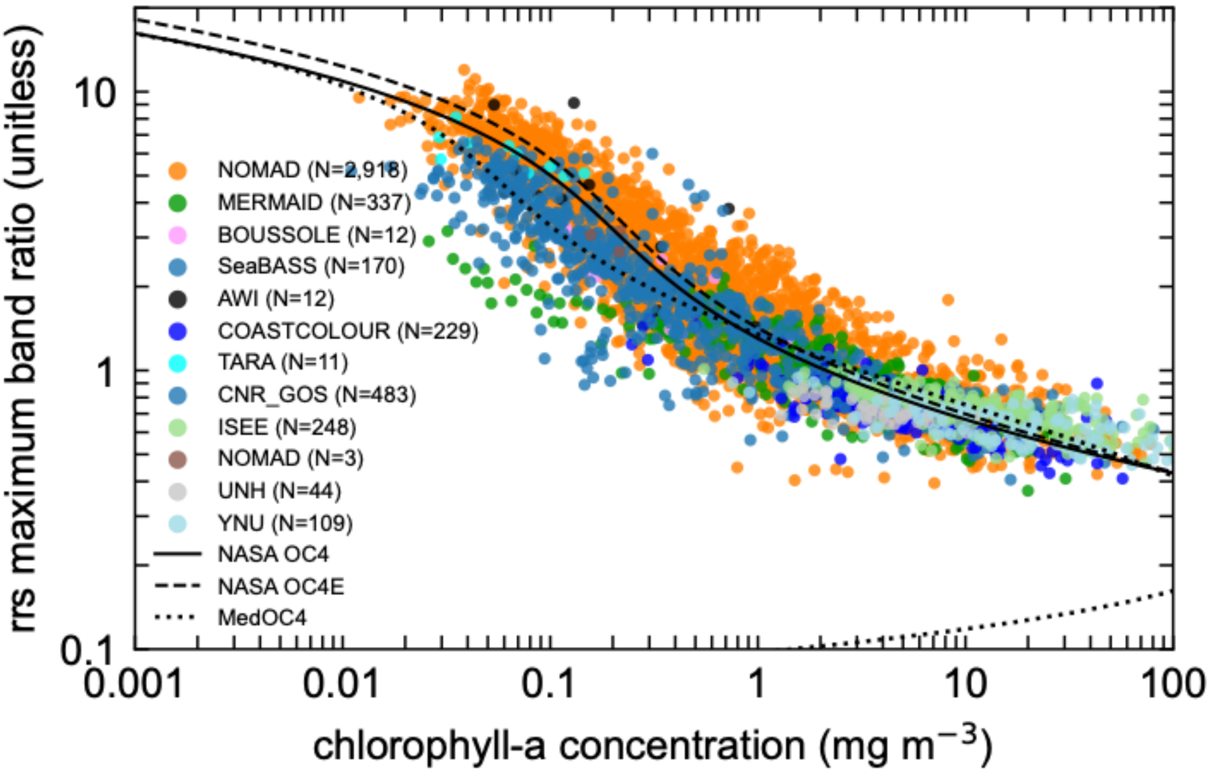

Figure 11. A remote-sensing reflectance maximum band ratio (as defined in text) ([443,490,510]/555 or [443,490,510]/560 if 555 not available) as a function of chlorophyll-a concentration. All chlorophyll data were considered, but for a given station, HPLC data were selected if available. Data within 2 nm of the wavelengths were used. For reference, the solid and dotted lines show the NASA OC4 and OC4E v1.1 standard algorithms, respectively (https://www.earthdata.nasa.gov/apt/documents/chlor-a/v1.0; accessed : 15 March 2026). The dotted line shows the regional MedOC4 algorithm for Mediterranean waters (Volpe et al., 2007, 2012, 2019). The total number of points was 4,523, of which 64.5% were from NOMAD.

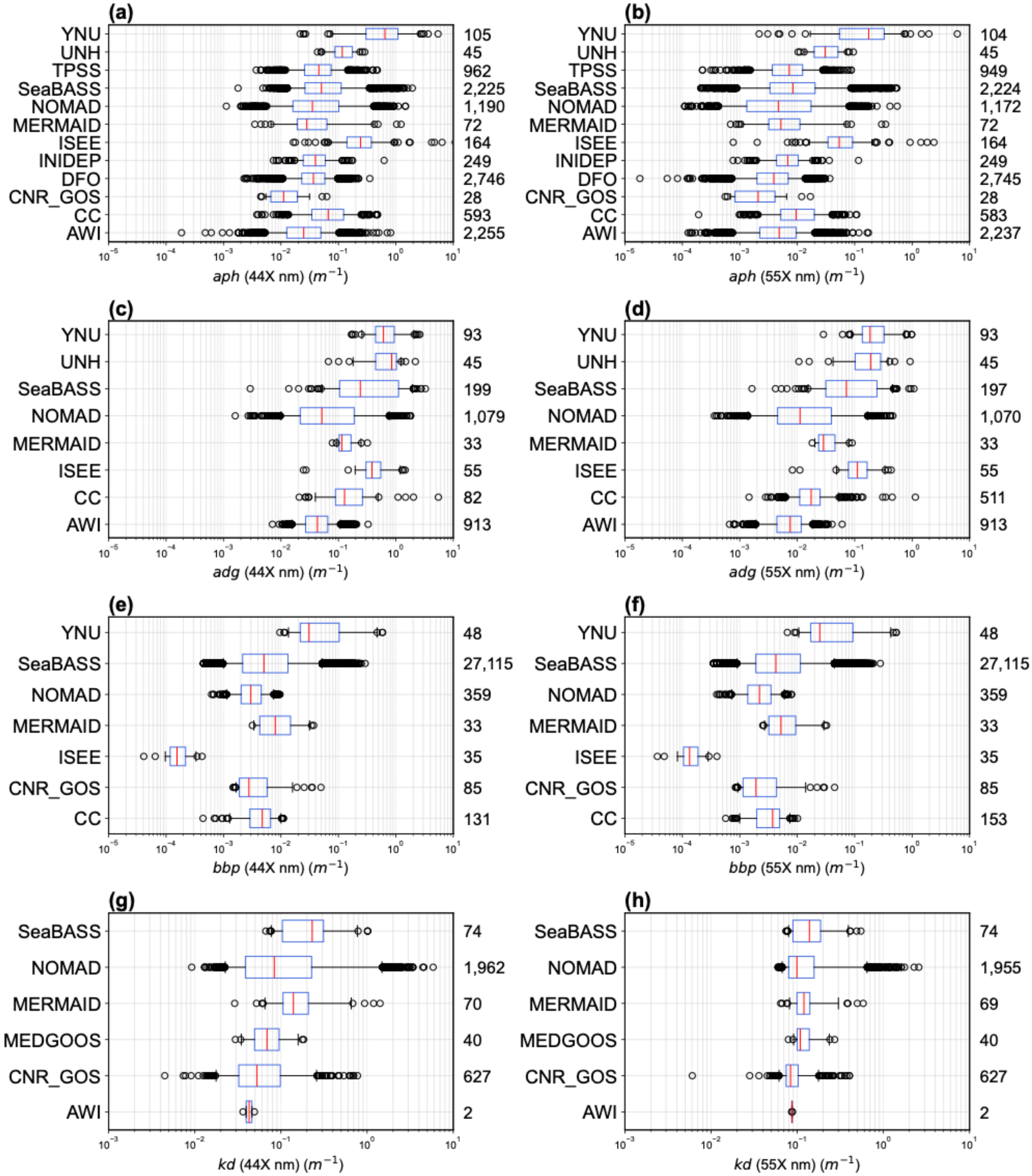

Figure 12. The distribution of: (a) "aph" at 44X nm; (b) "aph" at 55X; (c) "adg" at 44X nm; (d) "adg" at 55X; (e) "bbp" at 44X nm; (f) "bbp" at 55X; (g) "kd" at 44X nm; (h) "kd" at 55Xnm. Data were first searched at 445 and 555 nm, and then with a search window up to 8 nm, to include data at 547 nm. The graphical convention is identical to Fig. 2.

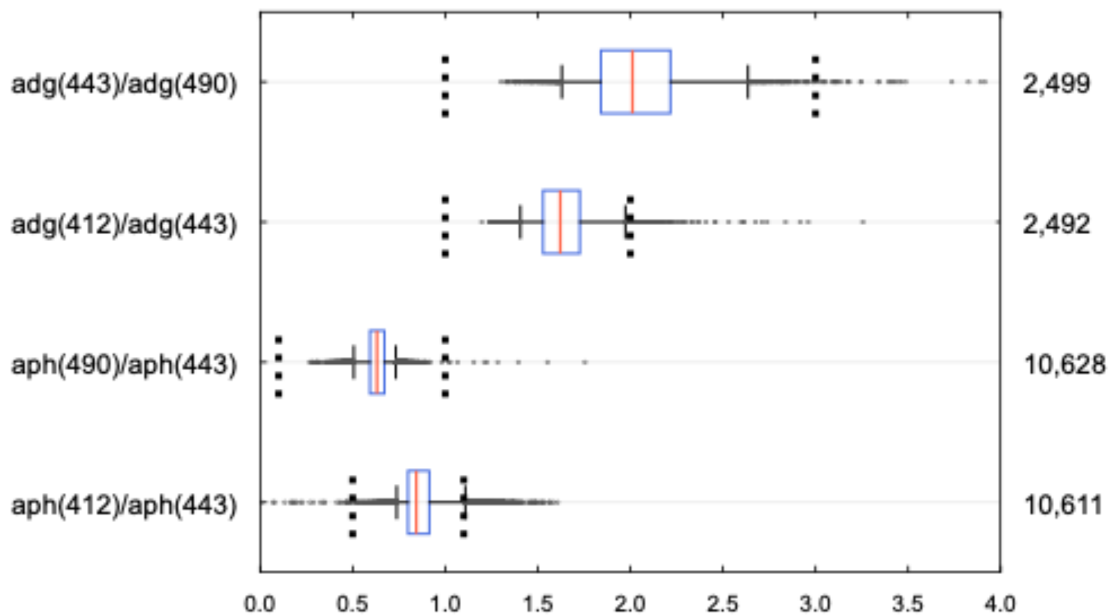

Figure 13. The distribution of absorption coefficients band ratios: adg(443)/adg(490), adg(412)/adg(443), aph(490)/aph(443) and aph(412)/aph(443). Data within 2 nm of the wavelengths were used. The graphical convention is identical to Fig. 2. The vertical dashed lines show the lower and upper thresholds used for quality control in the IOCCG report 5.

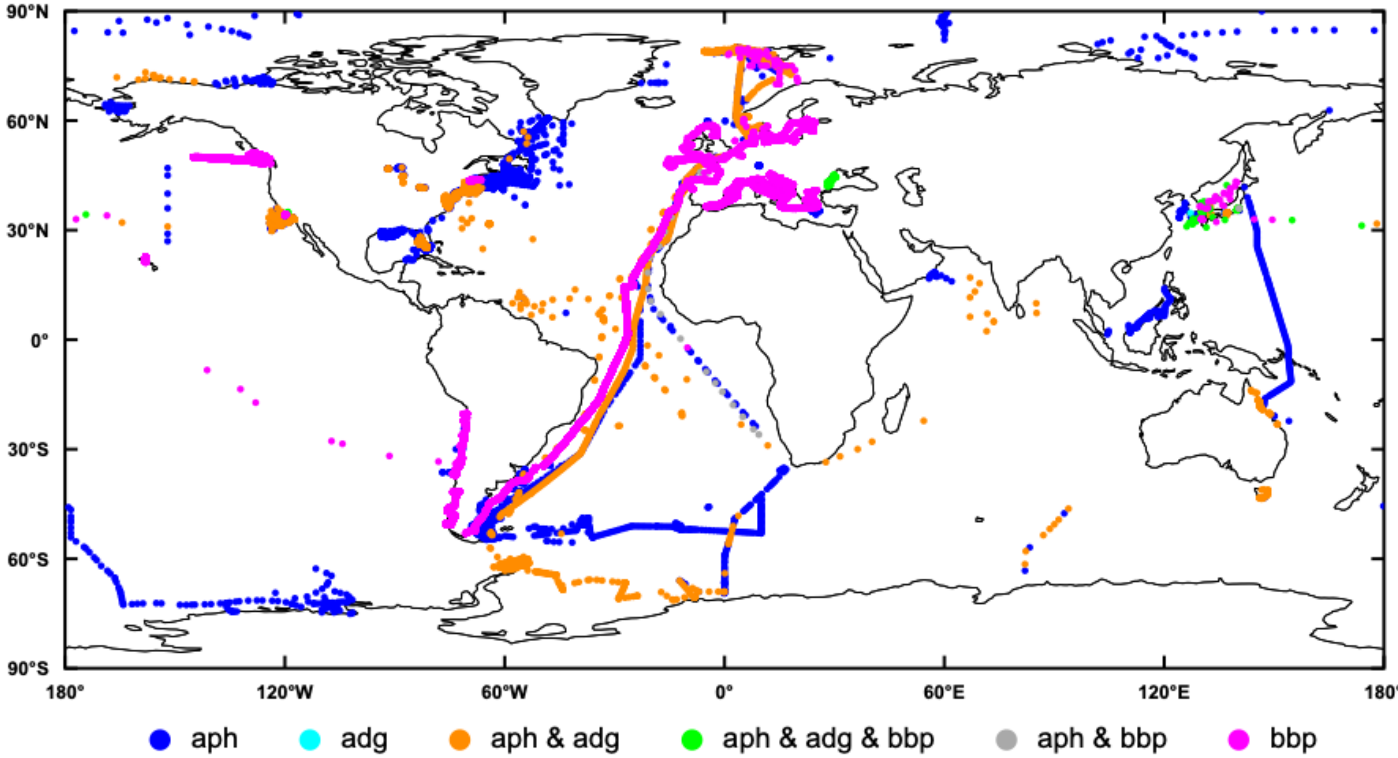

Figure 14. Global distribution of observations of inherent optical properties (algal pigment absorption coefficient "aph", detritus plus gelbstoff "CDOM" absorption coefficient "adg" and particle backscattering coefficient "bbp") in the final table.

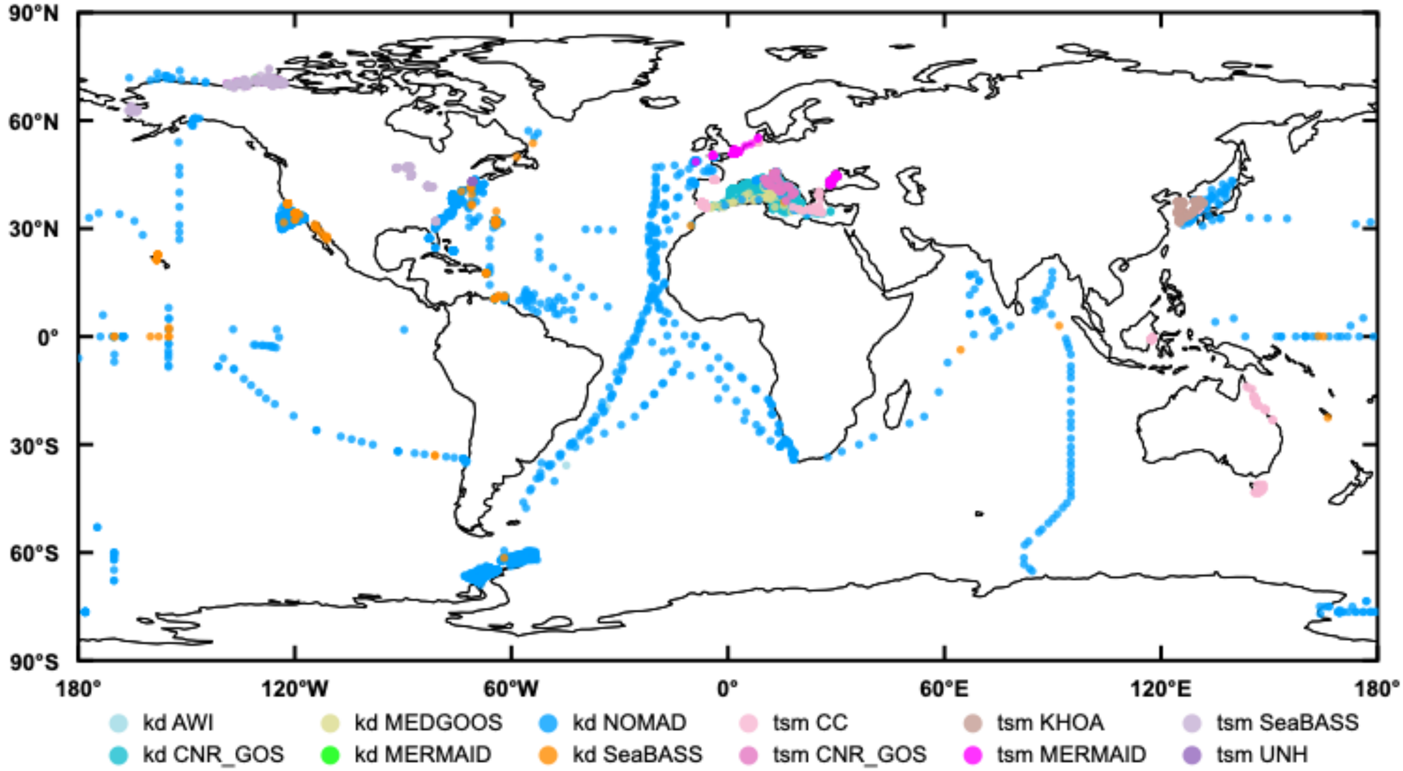

Figure 15. Global distribution of diffuse attenuation coefficient for downward irradiance ("kd") and total suspended matter ("tsm") per data set in the final table. The "tsm" and "kd" points from MERMAID overlap each other in west Black Sea (~40 ºN 30 ºE) and Arctic (~70 ºN 120 ºW).

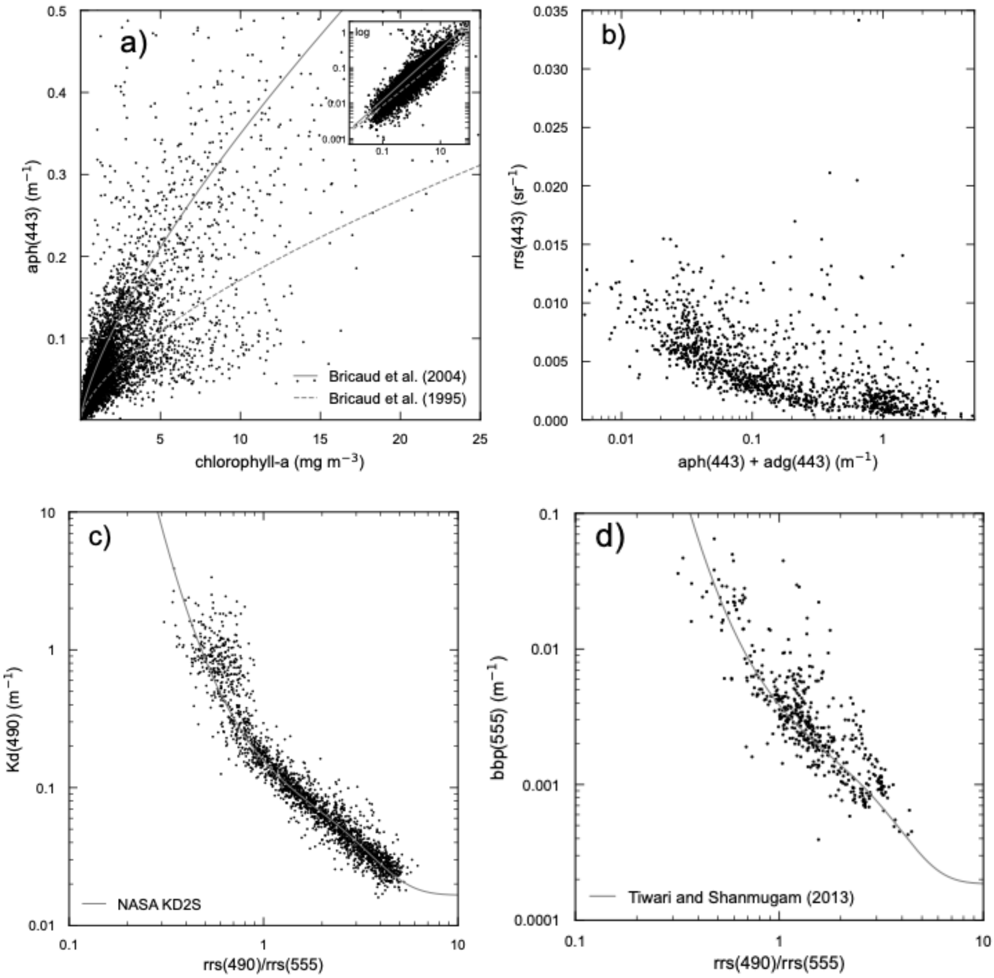

Figure 16. Examples of bio-optical relationships in the final merged table: (a) aph(443) versus chlorophyll-a; a log–log representation is shown in the top-right inset for reference. Total number of points (8,583) is divided between AWI (1,495), COASTCOLOUR (335), MERMAID (214), NOMAD (991), SeaBASS (1,426), TPSS (955), DFP (2,719), INIDEP (247), YNU (74), ISEE (69), UNH (44), and CNR_GOS (27) . For reference the solid line show the regression from Bricaud et al. (2004). (b) [aph(443) + adg(443)] versus rrs(443). Total number of points (1,268) is divided between MERMAID (33), NOMAD (1,079), YNU (82), UNH (45), ISEE (29). (c) [rrs(490)/rrs(555)] versus kd(490). The total number of points (2,282) is divided between MERMAID (62), NOMAD (2,117) and SeaBASS (103). For reference the solid line show the NASA KD2S standard algorithm (https://oceancolor.gsfc.nasa.gov/resources/atbd/kd/). (d) [rrs(490)/rrs(555)] versus bbp(555). The total number of points (498) is divided between MERMAID (33), NOMAD (324), COASTCOLOUR (4) SeaBASS (4), CNR_GOS (85), and YNU (48). For reference the solid line show the relation proposed by Tiwari and Shanmugam (2013). A search window of 2 nm was used for (a) and (b), and a search window of 5 nm was used for (c) and (d) to include data at 560 nm when not available at 555 nm.