The COASTLOOC ~~project~~ dataset

Authors (no particular order and to be validated)

- Philippe Massicotte[1]
- Marcel Babin[4, 1]
- Frank Fell[2]
- Vincent Fournier-Sicre[3, 5]
- David Doxaran[4]

Figures: [Click here for the figures](#)

Table 1: [table1](#)

(1) Takuvik International Research Laboratory (IRL 3376) ULaval - CNRS, Biology department, Laval University, Quebec G1V 06A, Canada

(2) Informus GmbH, 13187 Berlin, Germany

(3) ACRI-ST, 260 Route du Pin Montard, 06904 Sophia-Antipolis, France ~~Phone: +33 4 92 96 75 00~~

(4) Laboratoire d'Océanographie de Villefranche, CNRS / Sorbonne Université, 06230, Villefranche-sur-mer, France

(5) [current address]~~adresse] Data Processing Systems Procurement Manager~~ European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), EUMETSAT-Allee 1, 64295 Darmstadt, Germany ~~Office: +49 6151 807 7336 Mobile: +49 151 174 38520~~

Abstract

Coastal Surveillance Through Observation of Ocean Color (COAST𝓁OOC) oceanographic expeditions were conducted in 1997 and 1998 to examine the

relationship between the optical properties of seawater and related biological and chemical properties across the coastal-to-open ocean gradient in various European seas~~along the European coasts~~. A total of 379 stations were visited along the coasts of the Gulf of Lion in the Mediterranean Sea (n = 61), Adriatic Sea (n = 39), Baltic Sea (n = 57), North Sea (n = 99), English Channel (n = 85) and Atlantic Ocean (n = 38). Particular emphasis was ~~A particular emphasis has been~~ dedicated to the collection of a comprehensive set of apparent (AOPs) and inherent (IOPs) optical properties to support the development of ocean color remote sensing algorithms. The data were collected in situ using traditional ship-based sampling, but also from a helicopter, which is a very efficient means for that type of coastal sampling. The dataset collected during the COASTℓOOC campaigns is ~~still today quite~~ unique in that it is fully consistent in terms of operators, protocols, and instrumentation. This rich and historical dataset is still today frequently requested and used by other researchers. Therefore, we present the result of an effort to compile and standardize a dataset which will facilitate its use~~their reuse~~ in future development and evaluation of new bio-optical models adapted for optically-complex waters. The dataset is available at https://doi.org/10.17882/93570.

1. Introduction

Since the launch of the Coastal Zone Color Scanner (CZCS) by NASA in 1978, ocean color remote sensing has been used to monitor the state of, and the changes in global marine ecosystems both in time and space. In open oceans, the main component that affects the variations in the inherent (IOPs) and apparent (AOPs) optical properties of seawater is phytoplankton, which is usually represented by the concentration of chlorophyll a (Morel1977). Such open ocean waters are traditionally termed Case 1 waters. Many ~~simple~~ empirical spectral band ratio algorithms have been developed to link changes in remotely-sensed ocean color (OC), measured as reflectance, to ~~the~~ variations in chlorophyll a concentration (see OReilly2019 for an extensive evaluation of OC band ratio algorithms). Because these algorithms perform generally~~surprisingly~~ well, a plethora of studies have been conducted based on those~~it~~, notably about phytoplankton phenology (e.g., Vargas2009) and phytoplankton primary production (see Carr2006 and references therein).

Space-borne monitoring of aquatic ecosystems is more challenging for optically complex Case 2 waters (Morel1977) often found in coastal areas. In contrast to Case 1 waters, the optical properties of Case 2 waters are determined by several types of constituents occurring in individually highly variable concentrations. Rivers draining ~~Inland rivers typically drain~~ large catchment areas ~~that~~ deliver important quantities of optically significant substances such as chromophoric dissolved organic matter (CDOM) and suspended particulate matter (SPM) to the coastal waters (Hedges1997, Cole2007, Massicotte2017). In shallow~~these~~ areas, bottom reflectance and resuspension of sediments can also alter the signal of water-leaving reflectance (Lee1998). Because CDOM and SPM do not necessarily covary with chlorophyll a and can mask the presence of~~signal from~~ phytoplankton (Sathyendranath2000), bio-optical algorithms developed for Case 1 waters are

generally not appropriate for optically-complex Case 2 waters (Gordon1983). Hence, concomitant in situ measurements of optically active components (CDOM, SPM, chlorophyll *a*) and related radiometric quantities (AOPs and IOPs) are, even to this day, critically needed to develop and improve bio-optical algorithms for coastal areas.

The objective of the Coastal Surveillance Through Observation of Ocean Color (COAST*l*OOC) oceanographic expeditions was to acquire a comprehensive set of AOPs, IOPs, and concentrations and IOPs along the European coasts. Concentrations of the optically significant components along the European coasts were also collected to help the development of new remote sensing algorithms in optically complex waters. Back in 1997 and 1998, the COAST*l*OOC campaigns were among the very first concerted efforts to fill the knowledge gaps in coastal marine optics by establishing a large, consistenthomogenous, and comprehensivecomplete optical dataset for coastal waters. A total of 379 stations were visited along the coasts of the Mediterranean, Adriatic, Baltic and North Sea, in the English Channel, and in the Atlantic Ocean, mostly in coastal, but also in open oceanic waters. While this unique set of data has been used or referenced over the years in numerous peer-reviewed publications (see list in Appendix A), a reference final dataset, with well-documented quality controls, has never been published. Even though the COAST*l*OOC campaigns were carried out more than 20 years ago, the rich and historical dataset that has been collected still has great potential to contribute to the development and evaluation of new bio-optical models adapted for optically-complex waters. The goal of this paper is to formally present a quality-controlled and final reference version of the COAST*l*OOC dataset, archived in a public repository where users can download it readily.

2. Study area and sampling overview

2. 1 Study area and general sampling strategy

During the six COASTLOOC campaigns, further referred to as C1 to C6, a total of 379 locations were visited (Fig. 1). The stations were located along the coasts of the Gulf of Lion in the Mediterranean Sea (n = 61), Adriatic Sea (n = 39), Baltic Sea (n = 57), North Sea (n = 99), English Channel (n = 85), and Atlantic Ocean (n = 38). Within each area, the stations were generally distributed along across-shore or along-shore transects to capture the land-to-sea gradients and document river plumes (Fig. 1B). COASTLOOC campaigns were either ship-based (C1, C4, C5) or used a helicopter as a sampling platform (C2, C3, C6) and took place between 1997-04-02 and 1998-09-25 (Fig. 2A). Compared to traditional ship-based sampling, the helicopter platform allowed efficient sampling to efficiently sample stations close to the coast in areas difficult to access by boat (some samples were collected in waters as shallow as 1 m, see Babin2003b). Combining both ship and airborne sampling approaches allowed covering the whole inshore to open-ocean aquatic continuum. The bathymetry (GEBCO2020) varied greatly across the stations, where it averaged ca. 10 meters in the Adriatic Sea and ca. 800 2600 meters in the Case 1 Atlantic OceanMediterranean Sea (Fig. 2B).

2.2 Water sampling and processing

2.2 Data collection

At each site, between six and eight liters of water just below the surface were collected using either Niskin or polyethylene bottles for laboratory analyses. Water samples collected onboard the ship were immediately processed, or within two hours if sampled using the helicopter (see Babin2003 for further

details).

2.3 Suspended particulate matter and pigments

At each site, between six and eight liters of water just below the surface were collected using either Niskin or polyethylene bottles for laboratory analyses. On board the ship, water samples were analyzed and/or stored immediately after collection. When using a helicopter as a sampling platform, water samples were kept in ~~six liters~~ polyethylene containers for no longer than two hours after sampling (see Babin2003 for further details). Filtration was conducted at low vacuum onto 25 mm glass fiber filters (Whatman, GF/F) to collect particles for subsequent analyses, as described below. Three separate subsamples (up to two liters each) were filtered for each sample with the following purposes. To determine the concentration of suspended particulate matter (SPM; g m^{-3}), the particles collected onto a pre-weighed GF/F filter were dried and stored in a freezer at -80 °C until the dry weight was determined less than two months later in the laboratory (VanDerLinde1998). For pigment analysis, the filter with collected particles was inserted into a cryotube and kept in liquid nitrogen until analysis less than three months later. High-performance liquid chromatography (HPLC) was used as described by Vidussi et al. (1996) to determine liposoluble pigment concentrations. Total chlorophyll a is here defined as the sum of chlorophyll a, divinyl-chlorophyll a, chlorophyll a isomer and epimer, chlorophyllids a, and phaeopigments. The concentrations of particulate organic carbon and nitrogen (POC and PON, respectively) were determined using a Carlo Erba NCS 2500 elemental analyzer as described in Ferrari et al. (2003). These determinations were made on particles collected on~~onto~~ a GF/F filter precombusted at 450 °C for 2 h and stored in the freezer at -80 °C until analysis less than two months later. Further details concerning laboratory analyses can be found in Babin2003, Ferrari2000, and Ferrari2003). To

determine the concentration of dissolved organic matter (DOC, M), seawater samples were filtered on a 0.22 mm Millipore membrane. The final filtrate was transferred to a 100-mL amber-glass bottle, stored in the refrigerator and analyzed within three~~3~~ weeks. DOC was determined by high-~~-~~temperature catalytic oxidation using a Carlo Erba 480 analyzer. The methodological details can be found in Ferrari (2000).

2.~~3~~4 Optical measurements

At each station, an optical package was deployed in the water column from the surface, down to one to two ~~a few~~ meters above the ocean bottom in shallow waters, and to maximum depths of ca. 100 m in deep waters. The ~~deployed~~ package included a CTD (conductivity, temperature, and depth) as well as an array of optical instruments to simultaneously measure ~~the coefficients of~~ beam attenuation, absorption and scattering coefficients, as well as~~, and~~ irradiance in the water column. The Modular Data and Power System (MODAPS) data-gathering system from Wetlabs Inc., USA was used to combine the time-stamped data from the different instruments to produce a data matrix with a common depth grid. For further information, please refer to the COASTlOOC final report (Coastlooc2000).

A specific aspect of COASTlOOC~~COASTLOOC~~ is the use of a helicopter as a sampling platform, which has enabled high-frequency visiting of sites in very shallow waters (see section 2.1). An additional advantage is due to the fact that helicopter-based optical measurements are not influenced by the sampling platform itself, unlike ship-based measurements unavoidably affected by the presence of the ship hull, most prominently close to the surface. A drawback of this approach is the difficulty to operate a helicopter stationary at a constant height, especially at low or gusty wind conditions: while the former requires substantial flight heights to avoid the helicopter downdraft impacting the optical measurements, rendering sampling less

efficient, the latter ~~varying wind speeds. This~~ can result in unwanted vertical movements of the instrument package in water on the order of 1-2 m s^{-1}. To ensure the availability of data of sufficient quality close to the surface, several consecutive up- and downcasts were therefore performed at each "helicopter" site. Among ~~Across all~~ the measured vertical profiles at each site, the most suitable casts for depth merging and surface extrapolation (which are not necessarily the same, see the end of section 2.3.3) were selected using a combination of objective criteria and visual inspection.

2.~~3~~4.1 Irradiance measurements

The SeaWiFS Profiling Multichannel Radiometer (SPMR, Satlantic Inc, Canada) was used to measure downward (Ed) and upward (Eu) spectral irradiances (W m-2 μm-1) in the water column. Irradiance was measured at 13 wavelengths matching the MERIS channels of direct relevance for ocean observations (411, 443, 456, 490, 509, 531, 559, 619, 664, 683, 706, 779, 866 nm, except for COASTlOOC 1 operating at 590 nm instead of 619 nm), ranging from the blue part of the spectrum to the near-infrared at an acquisition rate of 6 Hz. The actual wavelengths differ slightly between upward and downward observations.

The SPMR was fixed to the IOP instrument frame, except for the COASTlOOC 5 mission, where it was used in free-falling mode. ~~SeaWiFS Profiling Multichannel Radiometer (SPMR, Satlantic Inc, Canada) was used to measure the downward (E_d, W m^{-2} μm^{-1}) and upward (E_u, W m^{-2} μm^{-1}) irradiances in the water column. E_d and E_u at a total of 13 wavelengths from the blue part of the spectrum (ca. 410 nm) to the near-infrared (ca. 865 nm) were measured at a rate of 6 Hz . The actual wavelengths differ slightly between upward and downward observations. Also, the set of nominal wavelengths were partly modified after campaign no. 1 (i.e., for campaigns 2 to 6) to better match the MERIS satellite channels (see below).~~ For the ship-based measurements,

concomitant observations of the in-air downward spectral irradiance at the sea surface were performed to account for changes in the incoming solar radiation during the in situ profiling (e.g., due to clouds). This approach was not achievable for the helicopter-based measurements, since there was no suitable place to mount the radiometer. For these campaigns, measurements of E_d taken by the profiler while in the air (i.e., ~~air~~ (before and after the in-water measurements) were used to assess E_d on the sea surface. In order to minimize the time lag between the in-water measurements and the Ed values on the sea surface, our processing code automatically chose the air measurements closest in time. Under stable atmospheric conditions, no significant error is expected to result from this approach, while it will contribute to increased variability in the retrieved parameters when sampling under varying cloud cover. However, the practical impact of this is limited due to the short duration of helicopter-based sampling, typically lasting between one and three minutes per profile.

2.3.2 Water-leaving reflectance

~~2.4.2 Irradiance surface extrapolation~~

To calculate the water-leaving reflectance, accurate measurements of both E_d and E_u just below the water surface (z=0⁻) are needed. In case no spectrally matching in-air E_d reference data are available (as is the case for the helicopter-based campaigns), this was done by extrapolating both $E_d(z)$ and $E_u(z)$ vertical profiles toward the sea surface by fitting an exponential model for the depth dependence of the measured spectral irradiance values (equation 1):

$$\Gamma(\lambda, z) = ¿\Gamma¿$$

where $\Gamma(\lambda, z)$ is either E_d or E_u measured at wavelength λ and at depth z, $\Gamma¿$ is E_d or E_u estimated just below ~~at~~ the air-water interface, and K_Γ is ~~either~~ the

estimated diffuse attenuation coefficient for either downward (K_{Ed}, m⁻¹) or upward (K_{Eu}, m⁻¹) irradiance. Note that a correction to account for instrument self-shading of the upwelling irradiance was not applied. We assume the impact of this omission is relatively small except for highly absorbing waters (depending on constituent concentration and/or observation wavelength), where it may lead to an underestimation of near-surface upwelling irradiances.

In case concomitant spectrally matching in-air E_d (0⁺) reference data are available (as is the case for most ship-based campaigns), E_d (0⁻) is calculated from E_d (0⁺) by applying a constant factor of 0.943 to account for losses of the downwelling irradiance (direct + diffuse) due to Fresnel reflection at the rough sea surface. The ~~channel-specific air-sea transmission factors. The such~~ derived E_u (0⁻) and E_d (0⁻) values were thereafter~~are then~~ used to calculate the water reflectance R (0⁻) just below the sea surface:

$$R ¿$$

Vertical irradiance profiles in the red and near-infrared channels at wavelengths beyond ~~above~~ 600 nm were fitted using a sum of two exponential functions of which one represents the diffuse attenuation due to absorption and elastic scattering for the channel under consideration, while the other accounts for ~~to account for~~ Raman scattering from the corresponding (shorter) excitation wavelengths significantly contributing ~~especially~~ to the upward light field ~~at these wavelengths~~ (e.g., Sugihara et al. 1984~~Sugihara1984~~). Not doing so, i.e., using just a single exponential function for surface extrapolation, leads to ~~will lead to a~~ substantial underestimation of the extrapolated subsurface values due to significant Raman contribution from excitation wavelengths with significantly lower K values (see Appendix 1). Figure A1 showing the near-surface depth dependence of E_u at 779 nm demonstrates why the two-function approach is required. Below a depth of

ca. 1.5 m, the signal is strongly dominated by Raman scattering and is therefore characterized by a diffuse attenuation coefficient very close to that of the corresponding excitation wavelengths.

2.34.3 Irradiance depth merging

Depth-merged irradiance profiles were derived for all sites by applying a standardized approach. The underlying strategy for depth merging was to provide users with the information they need to do their analysesown analyses, while modifying the original data as little as possible. Processing was therefore limited to procedures that could be applied equally to all COASTlOOC campaigns, whether helicopter- or ship-based. For example, variations of the incoming solar radiation were not considered in the depth merging process, since such observations were not available for the helicopter-based campaigns. However, if measured while operating from research vessels, such information is included in the depth-merged profile data to allow users doingto do their own subsequent analyses.

Typically, several up- and downcast vertical profiles were performed, especially when using the helicopter as a sampling platform, to increase the potential availability of high-quality observations. The actual merging process was executed as follows. First, any observation not meeting both of the two following conditions was discarded: (1) SPMR instrument tilt smaller than 20° and (2) SPMR vertical speed between 0.1 and 2.0 m s^{-1}. To automatically identify the best-suited cast for profile generation, the observations meeting the above two conditions were then distributed cast-wise over 0.2 m wide vertical bins, starting at zero depth. Finally, the cast with the highest number of bins containing at least one valid observation was selected for depth profile generation, if it fulfilled two further conditions: (3) at least 2 m vertical distance between the highest and the lowest bin and (4) at least 50% of the bins containingcontain at least one observation. No depth merged data is

provided for stations where there is no cast meeting all of the above four conditions. Combining several casts taken at one of a station has not proven successful due to potentially large discontinuities in the aggregated values at depth levels where the number of available casts changes, especially in case of unstable atmospheric conditions with the corresponding short-term irradiance fluctuations.

Actual depth merging was done by taking the median of all observations of a particular irradiance parameter within each depth bin. Spectral irradiance observations were only considered for values above 0.01 W m^{-2} µm^{-1} to reduce the impact of residual impacts of potentially inaccurate dark current and correction and significantly increasing radiometric noise. Finally, an attempt was made to identify and mask out remaining outliers in the depth-merged irradiance profiles. This was done by applying Grubb's test at a 5% significance level (Guthrie2012) to the residual of a second-order polynomial fit of the logarithmic-transformed spectral irradiance observations for a 2 m depth window, sliding down in steps of 1 m. While this procedure has successfully identified individual spikes, some obvious outliers extending over several depth bins remain in the merged dataset, and users will need to make their own judgment on how to identify and remove those. Examples of depth-merged irradiance profiles are shown in Fig. 3 for one clear-water site in the NE Atlantic and one CDOM-rich site in the Baltic Sea.

Examples of depth-merged irradiance profiles are shown in Fig. 3 for one clear-water site in the NE Atlantic (Fig. 3A) and one CDOM-rich site in the Baltic Sea (Fig. 3B). Figure 3A features a double deep chlorophyll maximum with the corresponding irradiance increase due to chlorophyll fluorescence at water depths between ca. 35 m and 45 m, most prominently visible at 683 nm, but also affecting the neighboring channels at 664 nm and 706 nm. Diffuse attenuation in the blue to green wavelengths is correspondingly

enhanced as reflected by the change in slope at a depth of ca. 34 m. Fig. 3B features Baltic Sea waters rich in CDOM, resulting in strong absorption in the UV and blue parts of the spectrum. The reduced decrease of the upwelling irradiance at 559 nm towards larger depths is likely caused by reflection of downwelling irradiance from the ocean bottom, while chlorophyll fluorescence is likely the reason for the relatively enhanced (change of slope) upwelling irradiance at 683 nm for depths below ca. 7 m.

Note that extrapolated surface values and depth-merged profiles for a specific site are not necessarily derived from the same observations. This is especially true for helicopter-based operations typically comprising several up- and downcast close to the water surface to increase the likelihood of having sufficient observations for the surface extrapolation, while usually only one or two deep-water casts were made to obtain observations at greater~~larger~~ depth. The best-suited observations for surface extrapolation, therefore, do ~~do therefore~~ not necessarily correspond to the cast used for depth merging.

2.3~~4~~.4 Profiles of the diffuse attenuation coefficient

Downward (K_{Ed}, m^{-1}) and upward (K_{Eu}, m^{-1}) spectral diffuse attenuation coefficients have been derived from the depth-merged irradiance data by fitting an exponential function through the irradiance observations within a 2 m depth window, sliding down in steps of 0.2 m. The calculation was only performed if at least five observations were available for fitting within the depth window~~within each bin~~. The choice of the ~~binning~~ window width was based on a compromise between two conflicting requirements: on~~. On~~ the one hand, it ~~the window width~~ should be small enough to allow resolution of ~~resolving~~ relatively narrow vertical features, such as ~~for example~~ a deep chlorophyll ~~a~~ maximum. On the other hand, it should be large~~wide~~ enough to allow for an accurate estimation of K. The ~~This relatively narrow~~ window width

of 2 m chosen herein favors vertical resolution over accuracy: while even relatively narrow vertical features remain distinguishablevisible in the derived K profiles, these are on the other hand affected by remaining outliers, instrument tilt variations, or instrumental noise. Obviously, users may use the depth-merged irradiance profiles to derive their own K profiles or depth-integrated K values better meetingto better meet their individual requirements.

2.34.5 Chromophoric dissolved organic matter (laboratory measurements)

Measurements of chromophoric dissolved organic matter (CDOM) absorption (a_{CDOM}, m^{-1}) were performed in the laboratory using a spectrophotometer (Perkin Elmer, Lambda 12) on water samples (10 cm cuvette path length) filtered on a 0.22 μm Millipore membrane pre-rinsed with 50 ml of Milli-Q water (Babin2003). a_{CDOM} spectra were measured between either 300-750 nm or 350-750 nm at 1-nm increment.

2.34.6 Particulate, non-algal and phytoplankton absorption (laboratory measurements)

Water samples were filtered onto 25 mm glass fiber filters (Whatman, GF/F) at a low vacuum before absorption measurement (Babin2003). Total particulate absorption (a_P, m^{-1}) was measured on particles retained on the filters between 380 and 750 nm at 1-nm increment using a spectrophotometer equipped with a 60 mm integrating sphere (Perkin Elmer, Lambda 19), and following the transmittance-reflectance (TR) method (Tassan1995, Tassan1998). Afterward, pigments were removed from the particles with sodium-hypochloride to measure non-algal (or non-pigmented) absorption (a_{NAP}, m^{-1}). Finally, phytoplankton absorption (a_φ, m^{-1}) was retrieved by subtracting a_{NAP} from a_P. a_{CDOM} and a_{NAP} absorption spectra were fitted according to the following equation (Jerlov1968, Bricaud1981):

$$a_\dagger(\lambda) = a_\dagger(\lambda 0)e^{-S_\dagger(\lambda-\lambda 0)} + k$$

Where $a_\dagger$ is the absorption coefficient, either a_{CDOM} or a_{NAP} (m^{-1}), λ is the wavelength (nm), $\lambda 0$ is a reference wavelength (443 nm), $S_\dagger$ is the spectral slope (nm^{-1}) that describes the approximate exponential decrease in absorption with increasing wavelength and k a background constant (m^{-1}) accounting for scatter in the cuvette and drift of the instrument. a_{CDOM} spectra were fitted using observations below 700 nm, whereas a_{NAP} fits were performed between 380 and 730 nm, excluding spectral regions between 400-480 nm and 620-710 nm to avoid possible residual pigment absorption (Babin2003). a_{CDOM} and a_{NAP} were baseline corrected by subtracting the background parameter (k) derived from the following equation:

$$\hat{a}_\dagger(\lambda) = a_\dagger(\lambda) - k$$

Finally, the spectral background of a_P spectra (i.e., the average absorption measured between 746 and 750 nm) were added to a_{NAP}. The underlying rationale is that absorption by phytoplankton in the near-infrared (NIR) is null, and that absorption measured in the NIR for total particles using the TR method is real (Tassan2003) and belongs exclusively to non-algal particles. The absorption signal measured for total particles retained on untreated filters is assumed more reliable than that measurement on bleached filters.

2.~~3~~4.7 Total absorption and scattering (in situ measurements)

In situ vertical profiles of absorption (a, m^{-1}) and beam attenuation (c, m^{-1}) were acquired at nine wavelengths (412, 440, 488, 510, 555, 630, 532 or 650 depending on the instrument configuration, 676, and 715 nm) using a flow-through in situ absorbance-attenuance meter (AC9, Wetlabs). As the measurements are referenced to pure (Milli-Q) water, the obtained absorption and attenuation coefficients exclude the contribution of water. To correct absorption measurements for incomplete recovery of scattered light,

$a(715)$ was subtracted from $a(\lambda<715)$. Scattering coefficients (b, m^{-1}), were calculated by subtracting absorption from attenuation. As the contribution of molecular (water) scattering is excluded (see above), this coefficient essentially corresponds to particle scattering and is hereafter denoted b_p (m^{-1}).

The such derived values for a, b_p, and c were subsequently averaged over the first attenuation length (1 / K_{Ed}) to provide values representative of the surface layer. Due to practical difficulties when operating the AC-9 from a helicopter (e.g., trapped air in the instrument), the availability of a, b_p, and c values is limited for helicopter-based campaigns, especially for COASTlOOC 3 (Adriatic Sea). Surface-to-bottom depth merging of the full AC-9 profiles has not been attempted so far, but this is envisaged for a potential future reprocessing of the COASTlOOC dataset (see Section 5).

3. Data quality control and data processing

Different general quality control procedures were adopted to ensure the integrity of the data. First, the raw data were visualized and screened to eliminate errors both originating from the measurement devices, including sensors (systematic or random), and errors inherent to measurement procedures and methods. Statistical metrics such as average, standard deviation, and range were computed to detect and remove anomalous values in the data. Then, data were checked for duplicates and remaining outliers. The complete list of variables is presented in Table 1.

4. Data description (an overview)

4.1 Spatial variability along the coastal-ocean gradient

The COASTLOOC sampling strategy was primarily designed to capture the bio-optical gradient across the sampled ecosystems and along transects from the coast towards the open ocean within each area (Fig. 1). The following

sections present an overview of a few selected variables measured in the different areas. An extensive and detailed explanatory visualization analysis is presented in the final report of the COASTLOOC project (Coastlooc2000).

4.1.1 Chlorophyll *a* and particulate organic carbon

As observed in Fig. 4, both total chlorophyll *a* and particulate organic carbon (POC) concentrations varied markedly across the sampled areas, reflecting the natural gradient captured by the sampling strategy. Across all ~~the~~ stations, total chlorophyll *a* ranged between 0.05 and 29 mg m^{-3}. The median chlorophyll *a* ranged between ca. 0.15~~0.1~~ mg m^{-3} in the Atlantic Ocean and 5.9 mg m^{-3} in the Baltic Sea (Fig. 4A). Based on the definition proposed by Antoine1996, the sampled stations were representative of a wide range of trophic status: oligotrophic (n = 17, chlorophyll $a \leq 0.1$ mg m^{-3}), mesotrophic (n = 106, $0.1 \leq$ chlorophyll $a \leq 1$ mg m^{-3}) and eutrophic (n = 245, chlorophyll $a > 1$ mg m^{-3}). A similar shore-to-sea gradient pattern could be observed for POC, with median values varying between ca. 0.07~~0.1 g m^{-3}~~ and 0.8 g m^{-3} in the Mediterranean Sea (case 2) and Baltic Sea, respectively (Fig. 4B, range between 0.02 and 2.5 g m^{-3}).

4.1.2 Inherent optical properties (IOPs)

Measured~~Averaged~~ absorption spectra in each ~~by~~ area for total particulate (a_p), non-algal particles (a_{NAP}), phytoplankton (a_φ), and chromophoric dissolved organic matter (a_{CDOM}) are presented in Fig. 5. Phytoplankton maximum absorption peaks~~,~~ at around 440 nm and 675 nm~~,~~ are easily distinguishable in Fig. 5C and were found to be highly correlated with the concentration of total chlorophyll *a* (Pearson's $r > 0.90$). Across all the areas, a$_{CDOM}$(350), a proxy for ~~watercolor and~~ dissolved carbon concentration, varied between 0.03 and 3.66 m^{-1}. These values fall within the ranges of the mean values reported globally in the ocean (0.14 m^{-1}), coastal (1.82 m^{-1}), and estuarine~~estuaries~~ (4.11 m^{-1}) ecosystems (Massicotte2017). Overall, the highest

absorption by dissolved organic matter was observed in the Baltic Sea (Fig. 5D). Sampled stations in this area were located West of the Oder River plume, which drains important quantities of humic substance from its catchment area. In contrast, the lowest CDOM absorption was observed in the Atlantic Ocean, where stations were located away from land and thus less influenced by terrestrial inputs (Fig. 1). The spatial variability of CDOM is further emphasized ~~strengthened~~ in Fig. 5E, where $a_{CDOM}(350)$ decreased by a factor of four between the two most distant~~distanced~~ points (approximately 40 km) of the westernmost transect sampled in the North Sea (see Fig. 1B).

Light scattering by suspended particles in the water column is a driver of reflectance variability and is often used by remote sensing applications to discriminate between Case 1 and Case 2 waters (Sathyendranath2000, Morel2006). The particulate scattering coefficient at 440 nm, $b_p(440)$ ranged between 0.05 m^{-1} and 35.8 m^{-1} (Fig. 6A). The median values varied by almost two orders of magnitude between the Atlantic Ocean and the Adriatic Sea (Fig. 6A). Furthermore, $b_p(440)$ showed more or less the same spatial pattern as the downward irradiance attenuation coefficient at 443 nm, $k_{Ed}(443)$, with median b$_p$(440) values varying~~which median values varied~~ between 0.05 m-1 and 1.21 m-1 in the Atlantic Ocean and the North Sea respectively (Fig. 6B, Pearson's r = 0.76).

4.1.3 Apparent optical properties (AOPs)

Subsurface reflectance $R(0^-)$ for all COASTlOOC sites is shown in Fig. 7, reflecting the fundamental differences in the bio-optical properties of the different areas visited. For example, due to enhanced CDOM absorption, reflectance in the blue part of the spectrum is generally low at the Baltic Sea sites, while it is higher in the Case-1 waters encountered in the Atlantic Ocean or the Mediterranean Sea. In the red and NIR, very high reflectance values are observed in the sediment-rich Case-2 ~~rich~~ waters of the North Sea and~~or~~ the

English Channel.

Sample profiles of the diffuse attenuation coefficient k_{Ed}(490) at a vertical resolution between 2.0 m close to the surface and 5.0 m underneath are shown in Fig. 8 for two transition zones ~~, each~~ ranging from turbid waters close to the coast to clearer waters ~~further~~ offshore. Fig. 8A shows k_{Ed}(490) values along a transect taken in the Gulf of Lion ~~Lion's Gulf~~ on 30. Sept. 1997 from site C4014 located ca. 5 km to the west of the mouth of the main Rhône branch (Grand Rhône) to site C4019 some 20 km further south in clear Mediterranean waters. While k_{Ed}(490) at the offshore sites C4017 to C4019 adopts low values between ca. 0.02 and 0.05 m^{-1} reflecting clear-water conditions, sites C4014 and C4015 are strongly influenced by the Rhône river plume, resulting in k_{Ed}(490) reaching values of up to 0.8 m^{-1} in the surface layer. The turbid Rhône freshwater is floating on top of ~~much~~ denser and clearer Mediterranean seawater, resulting in a strong stratification. Fig. 8B shows a similar transect taken in the North Sea on 12. Sept. 1998 extending from site C6063 close to the western shore of Texel island ~~the island of Texel~~ to site C6069 located ca. 22 km further west in the open North Sea. Again, there is a distance-to-coast-dependent decrease of k_{Ed}(490) values likely related to reduced turbidity, from ca. 0.8 to 1.0 m^{-1} close to the coast down to values of ca. 0.2 to 0.3 m^{-1} offshore. In contrast to the transect in the Gulf of Lion, ~~Lion's Gulf, there is~~ no obvious stratification is observed; ~~.~~ k_{Ed}(490) values are rather homogeneously distributed ~~homogeneous~~ over the entire water column, pointing to effective mixing.

4.2 Co-variability across bio-optical measurements

Few research papers have previously shown how optically significant components and AOPs/IOPs measured during the COASTLOOC missions co-varied (Ferrari2000, Ferrari2003, Babin2003, Babin2003a). In this section, a brief overview of selected pairwise relationships is presented. Total ~~Figure 4~~

~~shows that total~~ chlorophyll *a* co-~~varies~~varied with POC (Fig. 4) and the relationship was~~, and Fig. 9A shows that this relationship is~~ highly variable (Fig. 9A). ~~across the areas.~~ Whereas the global Pearson's correlation is 0.79 (*n* = 299), it ranged between 0.36 in the North Sea (*n* = 88) to 0.78 in the Adriatic Sea (*n* = 37) showing that optically significant components do not necessarily covary altogether. Likewise, the global Pearson's correlation between a_φ and total chlorophyll *a* (Fig. 9B) was relatively high (*r* = 0.909, *n* = 338). The lowest and highest correlations were observed in the North Sea (*r* = 0.69, *n* = 88) and the Adriatic Sea (*r* = 0.91, *n* = 38), respectively. POC concentration is well known to be an important driver of both IOPs and AOPs in aquatic ecosystems (Stramski2008, Cetinic2012). Unsurprisingly, positive correlations were observed between POC and $k_{Ed}(443)$ (Fig. 9C, *r* = 0.81, *n* = 204) and $b_p(440)$ (Fig. 9D, *r* = 0.67, *n* = 165).

Based on the size-reactivity continuum model proposed by Benner2015, physicochemical and photochemical processes are shaping the size distribution of organic matter along the aquatic continuum, and determine the contrasted intrinsic nature of the particles for each type of ecosystem. This is likely one key reason explaining why the observed relationships are quite variable across the different sampled ecosystems (Fig. 9). For example, in coastal areas (e.g., Baltic Sea, Fig. 1), the POC content is generally influenced by large humic organic particles drained by rivers from the surrounding watershed (Babin2003a). Consequently, in these areas, POC concentration (Fig. 4B), absorption (Fig. 5), and downward attenuation coefficients (Fig. 9C) were higher than in other sampled areas. In contrast, the stations that were farther away from the coast (such as the Atlantic Sea, Fig. 1) had lower POC concentrations (Fig. 4B) and absorption (Fig. 5). ~~This is suggesting that the particulate pool was typically made up of smaller, less colorful molecules that were mostly made by primary production in situ~~

5. Code and data availability

The COASTLOOC data is provided as a collection of comma-separated values (CSV) files that are regrouping measurements associated with each measurement. The processed and tidied version of the data is hosted at SEANOE (SEA scieNtific Open data Edition) under the CC-BY license (https://www.seanoe.org/data/00824/93570/).~~a particular type of measure.~~ To aid the user to merge these files, there is a lookup table file designated *station list* (https://www.seanoe.org/data/00824/93570/data/100311.csv)~~called stations.csv~~ that can serve as a table to join the data together based on the station's unique identifier. Table 1 shows the complete list of available measurements. Please note that the actual filenames (e.g., /100311.csv) are assigned by SEANOE during the upload process and cannot be altered. ~~and their associated principal investigators. The processed and tidied version of the data is hosted at SEANOE (SEA scieNtific Open data Edition) under the CC-BY license (https://www.seanoe.org/data/00824/93570/).~~ All statistical analyses were performed in R version 4.3.0~~2.2~~ (RCoreTeam2022). The code used to produce the figures and the analysis presented in this paper is available under the GNU GPLv3 license (10.5281/zenodo.8091717). Irradiance depth merging was done with an IDL code available under the Creative Commons Attribution 3.0 Germany license (10.5281/zenodo.8096682). Note that there are plans to update and further improve the current data set. More data will likely be recovered from the backup archives, most prominently AC-9 measurements to derive in situ vertical profiles of a, b_p and c. Intended improvements to radiometer processing comprise (not exhaustively) a better dark current correction, the application of a self-shading correction, a more accurate treatment of the air/sea transmission of downward irradiance (Ed), and a closer look at uncertainties. Any update of the COASTlOOC dataset will be made publicly available through SEANOE~~7708653}~~.

6. Conclusions

The consolidated, quality-controlled, data collected during the COASTLOOC oceanographic expeditions offer, even if collected more than twenty years ago, many possibilities to better understand the bio-optical dynamics along the land-to-sea gradient where optically-complex water drained from watersheds mix with seawater. To this date, nearly 40 peer-reviewed studies have made use of the COASTlOOC data set to develop remote sensing algorithms and to study the optical properties of seawater across the coastal-to-open ocean gradient (see Table A1). This now consolidated and easily accessible data set will facilitate future development and evaluation of new bio-optical models adapted for optically-complex waters. In this paper, only a subset of variables has been presented. The reader is referred to Table 1 for a complete list and description of variables collected during the COASTLOOC expeditions.

Author contribution

Acknowledgments

RV Poseidon, RV Victor Hensen) and helicopter (Commerc'Air SA) we used for collecting the field data.

Financial support