



**CNR
ISMAR**
ISTITUTO
DI SCIENZE
MARINE



Aequora
consulting

serco



CDOM-proxy retrieval from aeOLus ObseRvations



Deliverable Report D3-AUX

Auxiliary/ancillary data set description

**Dionisi¹ D., Bucci² S., Cesarini¹ C., Colella¹ S., D'Alimonte³ D., Di Ciolo² L., Di
Girolamo⁴ P., Di Paolantonio¹ M., Franco⁴ N., Kajiyama³ T., Liberti¹ G. L.,
Organelli¹ E., Santoleri¹ R.**

1: Consiglio Nazionale delle Ricerche, Istituto di Scienze Marine (CNR-ISMAR), Rome, Italy

2: Serco Italy S.p.a, Rome, Italy

3: AEQUORA, Lisbon, Portugal

4: School of Engineering, University of Basilicata, viale dell'Ateneo Lucano 10, 85100 Potenza, Italy

Date: 11 October 2021

Version: 2.0

Acceptance

Accepted on date:

By (for ESA):

TABLE OF CONTENTS

1	DOCUMENT INTRODUCTION	5
1.1	VERSION CONTROL	5
1.2	APPLICABLE DOCUMENTS	5
2	AUXILIARY/ANCILLARY DATA SET DESCRIPTION	6
2.1	INTRODUCTION	6
2.2	CAL/VAL PRODUCTS	6
2.2.1	BGC-ARGO	6
2.2.2	SATELLITE PRODUCTS	12
2.2.3	TEST DATASET IDENTIFICATION AND GENERATION	12
2.3	AEOLUS PRODUCTS	12
2.3.1	DESCRIPTION OF DATASET USED	13
2.3.2	DATA PRE-PROCESSING ACTIVITIES	14
2.4	ANCILLARY DATA SETS	16
2.5	SIMULATION DATA SETS	16
2.6	REFERENCES	16
3	APPENDIX 1	17

LIST OF FIGURES

Figure 1: The free-drifting profiling float (type PROVOR CTS-4) used for acquired 0-250 m radiometric measurements in open-ocean waters across the globe. The floats are equipped with: (a) Iridium antenna; (b) oxygen sensor; (c) sensor for chlorophyll fluorescence, chromophoric dissolved organic matter (CDOM) fluorescence, and particle light backscattering; (d) nitrate sensor; (e) conductivity–temperature–depth sensor; (f) radiometer [three $E_d(\lambda)$ + PAR]. (Organelli et al., 2016). 10

Figure 2: The quality-control procedure developed by (Organelli et al., 2016) for BGC-Argo measurements of $E_d(380)$, and used in the frame of COLOR. 11

Figure 3: Pre-processing flow diagram..... 15

LIST OF TABLES

Table 1: Overview table of auxiliary BGC-Argo data files..... 7

Table 2: Details of the data variables inside the AUX files corresponding to BGC-Argo $E_d(380)$ profiles..... 8

Table 3: Details of the data variables inside the AUX files corresponding to BGC-Argo $K_d(380)$ coefficients within the first optical depth (Z_{pd}). Overview table of auxiliary BGC-Argo data files. 8

Table 4: Version 1 of the released database for K_d dataset..... 12

Table 5: Aeolus satellite dataset metadata..... 13

Table 6: List of selected variables..... 13

1 Document introduction

1.1 Version Control

Version	Reason	Date of Release
1.0	First version delivered	13 09 2021

1.2 Applicable Documents

AD-1	Statement of Work: Statement of Work ESA Express Procurement Plus - [EXPRO+] - Aeolus+ Innovation (Aeolus+I)
AD-2	CDOM-proxy retrieval from aeOLus ObseRvations Detailed Proposal: COLOR
AD-3	Contract: ESA Contract 4000133933/21/I-BG Aeolus+ Innovation (Aeolus+I) COLOR – CDOM-proxy retrievals from aeOLus ObseRvations
AD-4	Deliverable Report D4-ALGO: Algorithm Development

2 Auxiliary/ancillary data set description

2.1 Introduction

The objective of this document is to provide a description of the suitable auxiliary/ancillary products that have been created for the areas of interest identified in COLOR (see section 2.6 of the deliverable report D1-RB). These individual datasets are used as input data for the data processor and later for verification, and for calibration and validation of the output product.

The repository hosting the data pool is available at:

<ftp://aeolus@51.68.87.22>

Port:21

Password: provided separately.

2.2 CAL/VAL Products

The datasets containing the retrieved variables, used for the development, testing and validation of the retrieval technique are described in this section.

2.2.1 BGC-ARGO

RATIONALE

To develop and validate the new AEOLUS's CDOM products, vertical profiles of downwelling irradiance (E_d) in the UV light at 380 nm, collected by autonomous robotic Biogeochemical–Argo (BGC-Argo) floats, have been primary selected and used to derive the diffuse attenuation coefficients (K_d) in the first optical depth (the depth that is seen by Ocean Colour satellites). The BGC-Argo database is currently the only database including E_d measurements in the UV at high temporal, horizontal and vertical resolutions acquired at a global scale. The BGC-Argo database here used was also published by (Organelli et al., 2017), and it's freely distributed by the SEANO (SEA scientific Open data Edition) publisher at <https://doi.org/10.17882/49388> and <https://doi.org/10.17882/47142> for vertical profiles and products within the first optical depth, respectively. Original data files can also be freely downloaded from the Coriolis Global Data Assembly Centre (GDAC) at <ftp://ftp.ifremer.fr/ifremer/argo>.

DESCRIPTION/COMMENTS

- Autonomous robotic measurements of downwelling irradiance at 380 nm acquired in the upper 250 m of the water column by an array of BGC-Argo floats deployed in COLOR's Test Areas. These data sets are most preferred as in situ data sources because of their high temporal, horizontal and vertical resolutions;
- The data sets consist in quality-controlled vertical profiles of $E_d(380)$ and derived K_d coefficients within the first optical depth (Z_{pd}) for selected Test Areas;
- Gaps within the profile due to detected cloud occurrences and wave focusing/defocusing are not completed; Before K_d computation, measured profiles are completed with mathematical extrapolation of the value just below the sea surface;

- Version: 1.0. Original data files distributed in Real-Time mode by Coriolis GDAC have been quality checked with Near-Real-Time protocols specifically designed for OC and space-based applications. NRT QCed data are not distributed by Coriolis GDAC. Details in Organelli et al. (2017).

- Region of Interest (RoI): North and South Atlantic subtropical gyres; North Atlantic subpolar gyre; Levantine Sea (Eastern Mediterranean Sea). Other RoIs identified in Deliverable 1 (Southern Ocean – Indian Sector, Black Sea, North Western Mediterranean Sea) will be delivered at KO+12.

- Period of the analysed data: November 2012-January 2016

- Spatial resolution: profiles

- Temporal sampling: every 1 to 10 days

- File extension: .txt (no specific routines are required to read it)

- Nomenclature of each file is on the format of:

BGC_Argo_RoI_Variable_CoriolisQCtype_NRT_dateofcreation(ddmmyyyy).txt

An overview of the main characteristics of the files is provided by Table 1.

Table 1: Overview table of auxiliary BGC-Argo data files.

Name	Short Description	Format	Size	Static/Dynamic	Spatial Coverage	Spatial Representation Type	Generation Frequency	Validity	Source
BGC_Argo_NASTG_Ed380_RT_NRT_02082021.txt	BGC-Argo Quality-controlled E _d (380) profiles	txt	3 MB	S	North Atlantic subtropical gyre	profile	N/A	N/A	SEANOE, https://doi.org/10.17882/49388
BGC_Argo_SASTG_Ed380_RT_NRT_02082021.txt	BGC-Argo Quality-controlled E _d (380) profiles	txt	4.8 MB	S	South Atlantic subtropical gyre	profile	N/A	N/A	SEANOE, https://doi.org/10.17882/49388
BGC_Argo_NASPG_Ed380_RT_NRT_02082021.txt	BGC-Argo Quality-controlled E _d (380) profiles	txt	12.2 MB	S	North Atlantic subpolar gyre	profile	N/A	N/A	SEANOE, https://doi.org/10.17882/49388
BGC_Argo_SEMED_Ed380_RT_NRT_02082021.txt	BGC-Argo Quality-controlled E _d (380) profiles	txt	3.9 MB	S	Levantine Sea (Eastern Mediterranean Sea)	profile	N/A	N/A	SEANOE, https://doi.org/10.17882/49388
BGC_Argo_NASTG_Kd380_RT_NRT_02082021.txt	BGC-Argo Quality-controlled K _d (380) coefficients	txt	10 KB	S	North Atlantic subtropical gyre	First optical depth	N/A	N/A	SEANOE, https://doi.org/10.17882/47142

D3-AUX Auxiliary/ancillary data set description

BGC_Argo_S ASTG_Kd380 _RT_NRT_02 082021.txt	BGC-Argo Quality- controlled K _d (380) coefficients	txt	16 KB	S	South Atlantic subtropical gyre	First optical depth	N/A	N/A	SEANOE, https://doi.org/10.17882/47142
BGC_Argo_N ASPG_Kd380 _RT_NRT_02 082021.txt	BGC-Argo Quality- controlled K _d (380) coefficients	txt	130 KB	S	North Atlantic subpolar gyre	First optical depth	N/A	N/A	SEANOE, https://doi.org/10.17882/47142
BGC_Argo_S EMED_Kd380 _RT_NRT_02 082021.txt	BGC-Argo Quality- controlled K _d (380) coefficients	txt	18 KB	S	Levantine Sea (Eastern Mediterranean Sea)	First optical depth	N/A	N/A	SEANOE, https://doi.org/10.17882/47142

Specifically, the data variables that are used by the data processor or used in the verification and validation exercise shall be described as outlined in Table 2 for E_d(380) profiles, and in Table 3 for K_d(380) coefficients.

Table 2: Details of the data variables inside the AUX files corresponding to BGC-Argo E_d(380) profiles.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
FLOAT_WMO	World Meteorological Organization float identification number	-	-
CYCLE	Float cycle	A cycle corresponds to the period between two float's stays at a 1000 m parking depth	-
PROFILE	Cast number	Vertical profile from 1000 m (or other given depth) up to the surface	-
DATE	Date of the cast	Format yyyy-mm-dd	-
LATITUDE	Latitude	-	[°N]
LONGITUDE	Longitude	-	[°E]
FLOAT_PI	Project's Principal Investigator	-	-
PROJECT	Funding	-	-
PRESSURE	Pressure data at <i>n</i> levels	-	dbar
ED_380_NM	Downwelling irradiance at 380 nm	Quality-controlled profiles with gaps due to clouds and wave focusing/defocusing occurrences; Profiles completed with values extrapolated just below the sea surface	μW cm ⁻² nm ⁻¹

Table 3: Details of the data variables inside the AUX files corresponding to BGC-Argo K_d(380) coefficients within the first optical depth (Z_{pd}). Overview table of auxiliary BGC-Argo data files.

Variable Names	Variable Description	Source/Reference/Citation	Variable Units
FLOAT_WMO	World Meteorological Organization float identification number	-	-

D3-AUX Auxiliary/ancillary data set description

CYCLE	Float cycle	A cycle corresponds to the period between two float's stays at a 1000 m parking depth	-
PROFILE	Cast number	Vertical profile from 1000 m (or other given depth) up to the surface	-
DATE	Date of the cast	Format yyyy-mm-dd	-
LATITUDE	Latitude	-	[°N]
LONGITUDE	Longitude	-	[°E]
FLOAT_PI	Project's Principal Investigator	-	-
PROJECT	Funding	-	-
Zpd	First Optical Depth	Computed from quality-controlled PAR profiles	m
Kd.380.	Diffuse attenuation coefficient at 380 nm within the first optical depth	Quality-controlled $K_d(380)$ coefficients computed from QC profiles described in Tables 1 and 2	m^{-1}
Serr_Kd.380.	Standard error of $K_d(380)$ coefficients	Computed from the linear fit between $\ln(E_d(380))$ and depth	m^{-1}

BGC-Argo programme, platform, sensor description and measurement protocol

The BGC-Argo programme is an on-going international effort (<https://biogeochemical-argo.org>) that started in 2012 through deployment of an array of more than 100 floats equipped with bio-optical and biogeochemical sensors, besides a CTD probe for acquiring simultaneous profiles of temperature and salinity (Roemmich et al., 2019). The array was initially funded by a few scientific research programmes that deployed floats in several areas around the globe, including those areas of interest for the COLOR project. In Figure 1, we report the example of a PROVOR CTS4 (NKE Marine Electronics – France) platform fully equipped with instruments to acquire physical, bio-geochemical and bio-optical variables.

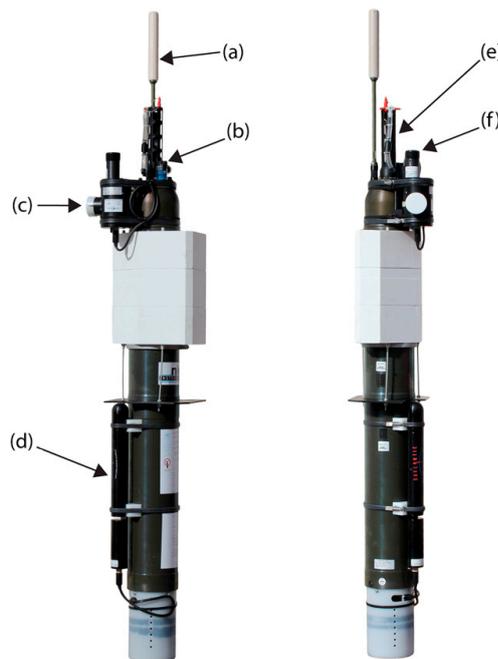


Figure 1: The free-drifting profiling float (type PROVOR CTS-4) used for acquired 0-250 m radiometric measurements in open-ocean waters across the globe. The floats are equipped with: (a) Iridium antenna; (b) oxygen sensor; (c) sensor for chlorophyll fluorescence, chromophoric dissolved organic matter (CDOM) fluorescence, and particle light backscattering; (d) nitrate sensor; (e) conductivity–temperature–depth sensor; (f) radiometer [three $E_d(\lambda)$ + PAR]. (Organelli et al., 2016).

A Seabird SATLANTIC OCR-504 radiometer acquired 0-250 m vertical profiles of E_d at three wavelengths (380, 412 and 490 nm; Units of $\mu\text{W cm}^{-2} \text{ nm}^{-1}$) and photosynthetically available radiation (PAR) integrated between 400 and 700 nm (Units of $\mu\text{mol quanta m}^{-2} \text{ s}^{-1}$). Each float was deployed according to published best practices and the radiometer was installed on the top of the platform to minimize platform's self-shading (Bittig et al., 2019). Radiometry measurements were programmed to be acquired every 1 to 10 days and surface around local noon when solar elevation is at its maximum, following international standard procedures (Mueller et al., 2003; IOCCG Protocol Series, 2019). E_d measurements were acquired with at least 1-m resolution, which increased approaching the ocean surface (every 0.2 m).

All BGC-Argo radiometric data distributed in Real-Time (RT) were gathered from the Coriolis Global Data Assembling Centre (GDAC; <ftp://ftp.ifremer.fr/ifremer/argo/dac/coriolis/>). Then Near-Real-Time (NRT) protocols specifically designed for OC applications have been applied to RT-QC $E_d(380)$ profiles. NRT QCed data are not distributed by Coriolis GDAC. In Version 1 of this deliverable, released data cover the temporal window from November 2012 to January 2016.

Data processing and quality control procedures

NRT Quality-Control of E_d profiles was performed by applying the protocol developed and fully detailed by (Organelli et al., 2016). This protocol consists in a series of successive mathematical and statistical tests (Figure 2), and accepts measurements acquired both under clear and cloudy sky conditions as good as these remain stable during the cast. Harsh sea conditions may deviate radiometers (and floats) from the vertical with respect to the sunlight which strongly reduce sensor performances. A first step of the quality control thus consists of identifying and discarding each profile acquired under very unstable sky and sea conditions (quantitative metrics are detailed in (Organelli et al., 2016)). The remaining profiles are quality controlled to identify and remove (1) dark measurements at depth, (2) sporadic atmospheric clouds, and (3) wave focusing/defocusing in the upper part of the profile. The (Organelli et al., 2016) protocol has been specifically developed for in situ bio-optical and remote sensing applications. It returns radiometric profiles of the highest quality to derive biogeochemical quantities (such as K_d coefficients), which are mandatory to achieve COLOR's objectives.

When NRT quality control procedures have been applied, each radiometric profile is associated with quality flags. A first set of quality flags returns the overall quality of the profile (i.e., profile type): Type 1 for good profiles, Type 2 for probably good profiles, and Type 3 for probably bad profiles. A second set of flags, still from 1 to 3, is assigned to each record within a profile and these flags are used to mark clouds, wave focusing/defocusing and dark counts within the profile. In the frame of COLOR, only flags "1" and "2" of profile "Type 1" are retained for further analysis.

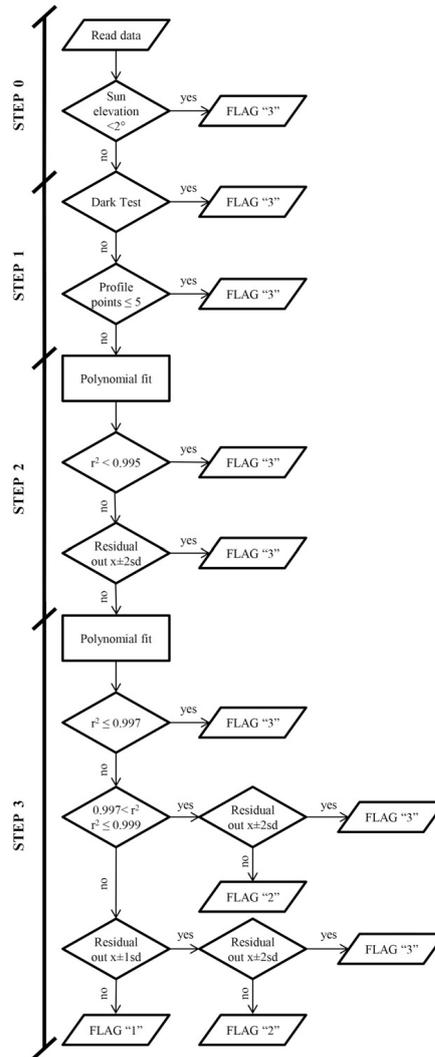


Figure 2: The NRT quality-control procedure developed by (Organelli et al., 2016) for BGC-Argo measurements of $E_d(380)$, and used in the frame of COLOR.

Radiometric profiles were also quality-checked for sensor drifts including biofouling and other sensors issues. The procedure we followed is the one published by (Organelli et al., 2017) which checks E_d measurements at the surface with those expected by atmospheric clear-sky models (Gregg and Carder, 1990), and cross-checks multiple variables acquired by the same float (e.g., radiometry, chlorophyll concentration, fluorescent dissolved organic matter and salinity) as detailed by (Organelli et al., 2017).

Because BGC-Argo E_d and PAR measurements are collected up to a few centimeters from the sea surface, NRT quality-controlled vertical profiles were completed by values just below it ($E_d(0^-)$). The $E_d(0^-)$ value was calculated by extrapolation within the first optical depth (Z_{pd}) using a second-degree polynomial fit according to (Organelli et al., 2016). The value of Z_{pd} was computed as $Z_{eu}/4.6$ (Morel, 1988), where Z_{eu} is the depth at which PAR is reduced to 1% of its value just below the surface and was calculated using measured PAR profiles. To achieve robust $E_d(0^-)$ values, the initial values of Z_{eu} and Z_{pd} were first estimated from the shallowest PAR measurement and subsequently from that corresponding to 0^- .

Then, the vertical diffuse attenuation coefficient for downward irradiance at 380 nm ($K_d(380)$) within Z_{pd} was computed for each measured $E_d(380)$ profile that passed NRT QC procedures described above. $E_d(380)$ profiles were first binned in 1-m intervals. $K_d(380)$ values then correspond to the slope of the linear fit, after removal of outliers, between the natural logarithm of $E_d(380)$ and depth (in units of pressure) following

(Mueller et al., 2003). Each slope of a linear fit is associated with the standard error of the computation, that thus corresponds to the error associated to obtained $K_d(380)$ coefficients. All the $K_d(380)$ coefficients based on less than three points or with a determination coefficient (r^2) lower than 0.90 were discarded (Organelli et al., 2017). The units of $K_d(380)$ are m^{-1} .

Expected upgrades for Version 2 of the database

Next release of the database will see the following upgrades:

- Source radiometric profiles from Coriolis GDAC quality controlled in Delayed Mode (DM) instead of RT. DM-QC implies correction of dark counts of irradiance measurements for sensor's dependence on water temperature changes and aging (protocols published in 2021 by Jutard et al., 2021);
- The database will be extended up to 2021 for all selected test areas according to BGC-Argo float availability (final temporal coverage 2012-2021);

2.2.2 Satellite products

Satellite dataset will be provided in the next phase of this activity.

2.2.3 Test dataset identification and generation

Based upon analyses and considerations duly reported in **Deliverable 1 – Science Requirements Baseline**, seven test areas have been selected to represent the expected global variability of $K_d(380)$ coefficients. Currently, preliminary analyses to verify the overall response of the algorithm in the developing and testing phases have been conducted on four ROIs: North and South Atlantic subtropical gyres, North Atlantic subpolar gyre, and Levantine Sea (Eastern Mediterranean Sea). Other ROIs identified in Deliverable 1 (Southern Ocean – Indian Sector, Black Sea, North Western Mediterranean Sea) will be delivered at KO+12. The Version 1 of the database released contains (Table 4):

Table 4: Version 1 of the released database for K_d dataset.

Test Area	Number of quality-controlled $E_d(380)$ profiles/ total number of n levels	Number of quality-controlled $K_d(380)$ coefficients within the first optical depth	Average $K_d(380) \pm$ standard deviation (units of m^{-1})
North Atlantic subtropical gyre	154	95	0.03 ± 0.006
South Atlantic subtropical gyre	234	156	0.03 ± 0.006
North Atlantic subpolar gyre	1711	1273	0.15 ± 0.05
Levantine Sea (Eastern Mediterranean Sea)	299	176	0.05 ± 0.02

2.3 Aeolus products

2.3.1 Description of Dataset used

The study uses L1B Aeolus products, containing data at measurement and observation resolution. Measurement ground linear resolution is approximately 3 km, observation resolution is about 90 km. Complete description of the selected variables is included in the IODD [AD-3]. In the following table is listed relevant information about the data used in the study:

Table 5: Aeolus satellite dataset metadata.

Baseline	10	
IODD document	521666_IODD_4_12	
Time range	From 2020-06 to 2020-08	
Data gap analysis document	Aeolus_data_documentV12.pdf	
Selected locations	North Atlantic Subtropical gyre (NASTG)	upper/lower latitude: 26N/16N western/eastern longitude: 55W/30W
	South Atlantic Subtropical gyre (SASTG)	upper/lower latitude: 14S/22S western/eastern longitude: 33W/19W
	Levantine Sea(LEV)	upper/lower latitude: 38N/30N western/eastern longitude: 22E/35E
Quality selection criteria	Multiple thresholds on Mie SNR (see Deliverable Report D4-ALGO: Algorithm Development)	
Selected variables	Listed in Table 6	

In the following table we list the variables actually used. For details about their description and their precision, refer to [AD-3].

Table 6: List of selected variables.

Dataset	Resolution	Variable name
geolocation/measurement_geolocation/geolocation_of_dem_intersection/	Measurement	longitude_of_dem_intersection
geolocation/measurement_geolocation/geolocation_of_dem_intersection/	Measurement	latitude_of_dem_intersection
geolocation/measurement_geolocation/measurement_mie_geolocation/	Measurement	altitude_of_height_bin
geolocation/measurement_geolocation/measurement_mie_geolocation/	Measurement	sattelite_range_of_height_bin
geolocation/measurement_geolocation/measurement_rayleigh_geolocation/	Measurement	altitude_of_height_bin
geolocation/measurement_geolocation/measurement_rayleigh_geolocation/	Measurement	sattelite_range_of_height_bin
geolocation/measurement_geolocation/geolocation_of_dem_intersection/	Measurement	altitude_of_dem_intersection
ground_wind_detection/measurement_ground_wind_detection/mie_measurement_ground_wind_bin/	Measurement	ground_bin_thickness_above_dem
ground_wind_detection/measurement_ground_wind_detection/mie_measurement_ground_wind_bin/	Measurement	surface
ground_wind_detection/measurement_ground_wind_detection/mie_measurement_ground_wind_bin/ground_bin_property/	Measurement	ground_bin_num
ground_wind_detection/measurement_ground_wind_detection/mie_measurement_ground_wind_bin/ground_bin_property/	Measurement	offset_dem_bin
ground_wind_detection/measurement_ground_wind_detection/mie_measurement_ground_wind_bin/ground_bin_property/	Measurement	dem_weight
ground_wind_detection/measurement_ground_wind_detection/mie_measurement_ground_wind_bin/ground_bin_property/	Measurement	snr_weight

D3-AUX Auxiliary/ancillary data set description

Dataset	Resolution	Variable name
ground_wind_detection/measurement_ground_wind_detection/mie_measurement_ground_wind_bin/ground_bin_property/	Measurement	fwhm_weight
ground_wind_detection/measurement_ground_wind_detection/rayleigh_measurement_ground_wind_bin/	Measurement	ground_bin_thickness_above_dem
ground_wind_detection/measurement_ground_wind_detection/rayleigh_measurement_ground_wind_bin/	Measurement	surface
ground_wind_detection/measurement_ground_wind_detection/rayleigh_measurement_ground_wind_bin/ground_bin_property/	Measurement	ground_bin_num
ground_wind_detection/measurement_ground_wind_detection/rayleigh_measurement_ground_wind_bin/ground_bin_property/	Measurement	offset_dem_bin
ground_wind_detection/measurement_ground_wind_detection/rayleigh_measurement_ground_wind_bin/ground_bin_property/	Measurement	dem_weight
ground_wind_detection/measurement_ground_wind_detection/rayleigh_measurement_ground_wind_bin/ground_bin_property/	Measurement	snr_weight
ground_wind_detection/measurement_ground_wind_detection/rayleigh_measurement_ground_wind_bin/ground_bin_property/	Measurement	fwhm_weight
useful_signal/measurement_useful_signal/mie_altitude_bin_useful_signal_info/	Measurement	useful_signal
useful_signal/measurement_useful_signal/rayleigh_altitude_bin_useful_signal_info/	Measurement	useful_signal_channel_a
useful_signal/measurement_useful_signal/rayleigh_altitude_bin_useful_signal_info/	Measurement	useful_signal_channel_b
product_confidence_data/measurement_pcd/meas_alt_bin_pcd/	Measurement	mie_signal_to_noise_ratio
product_confidence_data/measurement_pcd/meas_alt_bin_pcd/	Measurement	rayleigh_signal_to_noise_ratio_channel_a
product_confidence_data/measurement_pcd/meas_alt_bin_pcd/	Measurement	rayleigh_signal_to_noise_ratio_channel_b

Those variables have been used globally and on the selected locations specified in Table 5.

2.3.2 Data pre-processing activities

Rationale

Data pre-processing is an activity performed for two main reasons:

- data quality and availability assessment
- cross-compatibility of the output format with scientists' platforms and systems

The first activity is of great importance, because it guarantees the creation of a data pool with a quality level and availability that is always transparent to the teams that use it. It allows a faster and more conscious selection of the study period.

The second point is vital to speed-up research activities among different teams and multiple data processing platforms.

Pre-processing chain overview

The pre-processing chain is described by the diagram below:

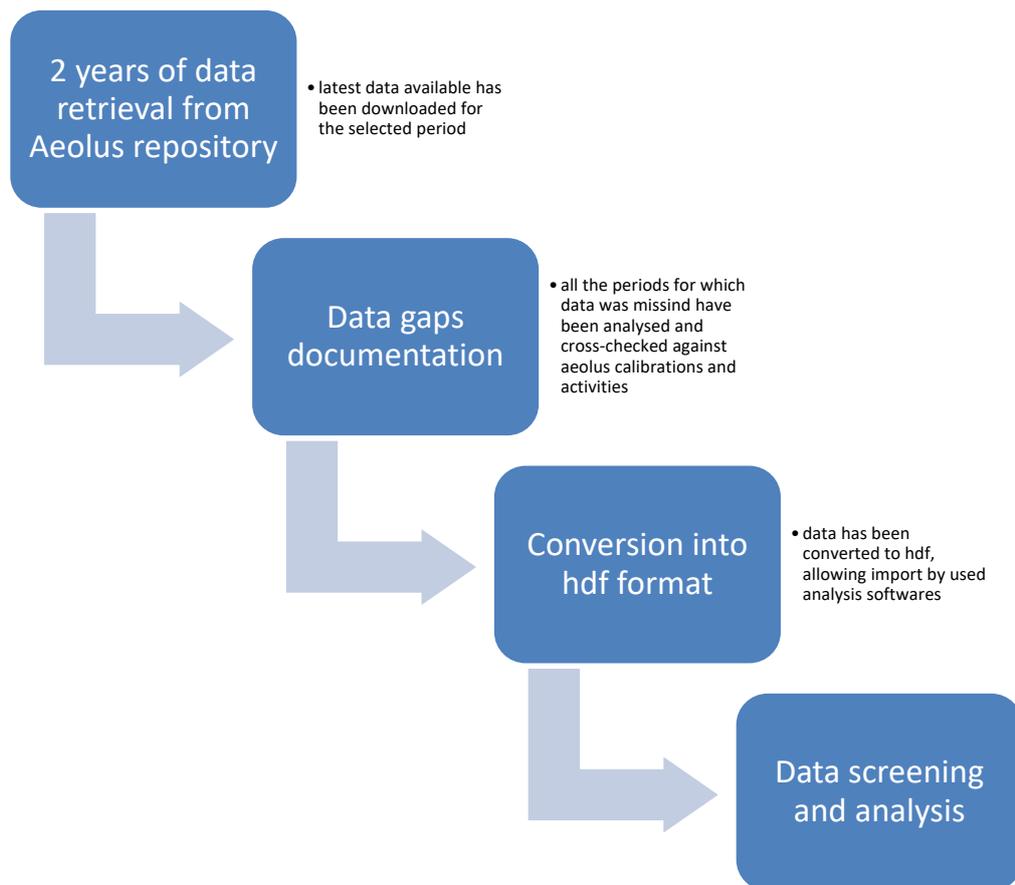


Figure 3: Pre-processing flow diagram

Let's see in detail all the blocks:

1. Aeolus satellite data are available on the distribution facility of the mission. Multiple baselines have been saved on the repository during real time operations and reprocessing activities. The first step is to select and download the latest baselines available for each period included in the study. Latest baselines are supposed to have better quality than older ones, due to evolutions in processing algorithms and software.
2. then the analysis of gaps has been performed, to know if:
 - a. discrepancies observed before and after the gap were due to what happened in between
 - b. a calibration was available at a certain point, in case the team would decide to use it. ("Data gap analysis document" is specified in Table 5)
3. An automatic script for data conversion has been developed and put in place. This was run and the complete dataset has been converted into HDF format to make it readable from the most common data science software (go to Appendix 1 for format conversion information).
4. Data screening and analysis step is based on variables that represent data quality. In this study the SNR of the Mie signal has been chosen, as specified in Table 5.

L1.5 AEOLUS COLOR dataset

A preliminary dataset for the last five AEOLUS acquisition bins (21-25), containing the selected variables listed in Table 6, was produced seasonally for each ROI.

Nomenclature of each file is on the format of:

AEOLUS_L1.5COLOR_ROI_season_year(yyyy)_dateofcreation(ddmmyyyy).txt

2.4 Ancillary data sets

Ancillary data sets will be provided in the next phase of this activity.

2.5 Simulation data sets

Simulation data sets will be provided in the next phase of this activity.

2.6 References

- Bittig, H.C., Maurer, T.L., Plant, J.N., Schmechtig, C., Wong, A.P.S., Claustre, H., Trull, T.W., Udaya Bhaskar, T.V.S., Boss, E., Dall’Olmo, G., Organelli, E., Poteau, A., Johnson, K.S., Hanstein, C., Leymarie, E., Le Reste, S., Riser, S.C., Rupan, A.R., Taillandier, V., Thierry, V., Xing, X., 2019. A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage. *Front. Mar. Sci.* 6, 502. <https://doi.org/10.3389/fmars.2019.00502>
- Gregg, W.W., Carder, K.L., 1990. A simple spectral solar irradiance model for cloudless maritime atmospheres. *Limnol. Oceanogr.* 35, 1657–1675. <https://doi.org/10.4319/lo.1990.35.8.1657>
- IOCCG Protocol Series, 2019. Protocols for Satellite Ocean Colour Data Validation: In Situ Optical Radiometry. Vol. 3.0. International Ocean Colour Coordinating Group (IOCCG). <https://doi.org/10.25607/OBP-691>
- Jutard et al. - 2021 - Correction of Biogeochemical-Argo Radiometry for S.pdf, n.d.
- Morel, A., 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (Case I Waters). *Journal of Geophysical Research* 93, 10749–10768.
- Mueller, James L, Fargion, G.S., McClain, C.R., Mueller, J L, Morel, A., Frouin, R., Davis, C., Arnone, R., Carder, K., Steward, R.G., Hooker, S., Mobley, C.D., McLean, S., Holben, B., Pietras, C., Knobelspiesse, K.D., Fargion, G.S., Porter, J., 2003. Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 4, Volume III: Radiometric Measurements and Data Analysis Protocols 84.
- Organelli, E., Barbieux, M., Claustre, H., Schmechtig, C., Poteau, A., Bricaud, A., Boss, E., Briggs, N., Dall’Olmo, G., D’Ortenzio, F., Leymarie, E., Mangin, A., Obolensky, G., Penkerch, C., Prieur, L., Roesler, C., Serra, R., Uitz, J., Xing, X., 2017. Two databases derived from BGC-Argo float measurements for marine biogeochemical and bio-optical applications. *Earth Syst. Sci. Data* 9, 861–880. <https://doi.org/10.5194/essd-9-861-2017>
- Organelli, E., Claustre, H., Bricaud, A., Schmechtig, C., Poteau, A., Xing, X., Prieur, L., D’Ortenzio, F., Dall’Olmo, G., Vellucci, V., 2016. A Novel Near-Real-Time Quality-Control Procedure for Radiometric Profiles Measured by Bio-Argo Floats: Protocols and Performances. *Journal of Atmospheric and Oceanic Technology* 33, 937–951. <https://doi.org/10.1175/JTECH-D-15-0193.1>
- Roemmich, D., Alford, M.H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., Owens, W.B., Pouliquen, S., Purkey, S., Scanderbeg, M., Suga, T., Wijffels, S., Zilberman, N., Bakker, D., Baringer, M., Belbeoch, M., Bittig, H.C., Boss, E., Calil, P., Carse, F., Carval, T., Chai, F., Conchubhair, D.Ó., d’Ortenzio, F., Dall’Olmo, G., Desbruyeres, D., Fennel, K., Fer, I., Ferrari, R., Forget, G., Freeland, H., Fujiki, T., Gehlen, M., Greenan, B., Hallberg, R., Hibiya, T., Hosoda, S., Jayne, S., Jochum, M., Johnson, G.C., Kang, K., Kolodziejczyk, N., Körtzinger, A., Traon, P.-Y.L., Lenn, Y.-D., Maze, G., Mork, K.A., Morris, T., Nagai, T., Nash, J., Garabato, A.N., Olsen, A., Pattabhi, R.R., Prakash, S., Riser, S., Schmechtig, C., Schmid, C., Shroyer, E., Sterl, A., Sutton, P., Talley, L., Tanhua, T., Thierry, V., Thomalla, S., Toole, J., Troisi, A., Trull, T.W., Turton, J., Velez-Belchi, P.J., Walczowski, W., Wang, H., Wanninkhof, R., Waterhouse, A.F., Waterman, S., Watson, A., Wilson, C., Wong, A.P.S., Xu, J., Yasuda, I., 2019. On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array. *Front. Mar. Sci.* 6, 439. <https://doi.org/10.3389/fmars.2019.00439>

3 Appendix 1

Original mission files are available at:

https://aeolus-ds.eo.esa.int/oads/access/collection/L1B_Wind_Products/tree

From here you can select ALD_U_N_1B, the desired baseline and time period.

Files are in TGZ format.

The extraction of the file results in a .tar archive that contains a .DBL and a .HDR file.

The first one can be converted into HDF format using CODA (more information available at <https://stcorp.github.io/coda/doc/html/index.html>).

The command used for the conversion is described in the following link:

<https://stcorp.github.io/coda/doc/html/codadump/index.html>

An example of the command usage is reported below:

```
codadump hdf4  
AE_OPER_ALD_U_N_1B_20210526T105856026_003371995_015977_0001.DBL
```

WHERE IS SUFFICIENT TO SPECIFY THE FORMAT AND THE NAME OF THE FILE TO BE CONVERTED.