Framework Service Contract EEA/DIS/R0/20/001 Lot 1 for Services supporting the European Environment Agency's (EEA) cross-cutting coordination of the Copernicus programme's in situ data activities – Observational data

C-RAID– Phase 2

Spring delivery - Final Activity Report

Patricia ZUNINO, Jean-Philippe RANNOU and Alain MAZURIER

Issue: 2.0
Date: 28/06/2021

[Document Release]

	Name(s)	**Affiliation**
Coordinated by:	Thierry Carval	IFREMER
Contributions:	Patricia Zunino-Rodriguez Jean-Philippe Rannou Alain Mazurier	CAP GEMINI ENGINEERING CAP GEMINI ENGINEERING CAP GEMINI ENGINEERING
Approval:	Henrik Steen Andersen	European Environment Agency

Prepared for:	European Environment Agency (EEA)
Represented by: (Project Manager)	Henrik Steen Andersen
Contract No.	EEA/DIS/R0/20.001

[Change Record]

Version	**Date**	**Changes**
1.0	17/10/2020	Creation
2.0	28/06/2021	Update

Table of Contents

Figures

Tables

Acronyms

Acronyms	Definition
AOML	Atlantic Oceanographic and Meteorological Laboratory
ATMS	Parameter name of sea level atmospheric pressure
CLS	Collecte Localisation Satellite
DMQC	Delayed Mode Quality Control
GDP	Global Drifter Program
RTQC	Real Time Quality Control
SST	Sea Surface Temperature

Reference documents

Reference N°	Title	Link
#RD1	C-RAID drifters Quality Control Manual	
#RD2	C-RAID drifters NetCDF format reference manual	

1. INTRODUCTION

This activity rapport summarizes the actions carried out for the second phase of the C-RAID project (EEA work on Specific Contract 4 n° 3436_R0-COPERNICUS_EEA.57652).

The objects of the C-RAID project is to gather, treat, decode and make available all the historical data measured by drifters starting in 1979.
Since 2016, most drifters are using Iridium for data transmission and Global Navigation Satellite System (GNSS) for positioning, whereas most earlier platforms used Argos for both functions.
The second part of the project focuses on years 1979-2010.

The main steps of the work done during this project are:

- To gather all available data: meta-data, original data transmitted by the drifters and already decoded data available at data centers;
- To decode anew original data when available;
- To create the C-RAID drifter data set by merging newly decoded and already decoded data;
- To quality control the obtained drifter data set with real time tests followed by a delayed mode phase (including data visualization).

We decided to focus on the AOML drifters. The concerned drifter deployed during the 1979-2010 period were equipped with Argos transmission device. CLS, as our partner in the project, was responsible for inventorying and extracting all available drifter Argos original data from their databases. The extraction of drifter Argos original data was achieved during this second phase of the project.

2. METADATA PROCESSING

Most of the metadata used in the project have been downloaded from the AOML GDP website (https://www.aoml.noaa.gov/phod/gdp/).

2.1 GDP website metadata

The available metadata are:

- Drifter individual metadata files (dirfl_1_5000.dat, dirfl_5001_10000.dat, dirfl_10001_15000.dat and dirfl_15001_mar19.dat);
- List and details of all drifters (https://www.aoml.noaa.gov/phod/dac/dirall.html);
- Drifter deployment log (https://www.aoml.noaa.gov/phod/dac/deployed.html);
- Quality controlled drifter metadata (https://www.aoml.noaa.gov/phod/dac/dirall.html);
- Drifter death/aground/picked-up probabilities (https://www.aoml.noaa.gov/phod/dac/drifter_deaths.html);
- Drifter specification metadata (https://www.aoml.noaa.gov/phod/dac/Drifter_Specifications.html);
- Barometer specification metadata (https://www.aoml.noaa.gov/phod/dac/Barometer_Metadata.html);
- Drogue specification metadata (https://www.aoml.noaa.gov/phod/dac/Drogue_Specifications.html);
- Drifter Id vs WMO number (https://www.aoml.noaa.gov/phod/dac/wmoid.html)

2.2 Additional metadata

An additional Excel file ("97-10_CLS_Meta_Request.xlsx") was kindly received from CLS (fblanc@groupcls.com), it contains AOML drifter decoding templates and sensor calibrations.

2.3 Metadata processing

The metadata processed correspond to all the drifters deployed before January 1st 2019, that is to say 22 473 drifters.
The processing consisted of merging the available metadata from all inputs mentioned above.
Some information are present in more than one input, we then had to check their consistencies.
Inconsistencies were found for 156 drifters, we then asked AOML ('Shaun.Dolk@noaa.gov', 'mayra.pazos@noaa.gov', 'Erik.Valdes@noaa.gov' and 'rick.lumpkin@noaa.gov') to fix the identified issues.
As we never received any answer from them, we decided to exclude these drifters from our lists.

We finally stored the processed metadata of the 22 317 remaining drifters in three Excel files named:

- "finalize_aoml_meta_ALL_1_20191017T141257.xlsx";
- "finalize_aoml_meta_ALL_2_20191017T141257.xlsx";
- "finalize_aoml_meta_ALL_3_20191017T141257.xlsx";

3. INPUT DATA

Two types of input data have been used to generate the C-RAID drifter data set:

- The original drifter Argos messages transmitted by the drifters. They are collected, processed and distributed to users by CLS;
- The archive of already decoded data available at the GDP ftp site (ftp://ftp.aoml.noaa.gov/pub/phod/buoydata/unkriged/);

3.1 Drifter Argos messages

The drifter Argos messages are stocked in the PVR/DS format (see https://www.argos-system.org/manual/6-data/632_data_formats.htm). The files contain all the information transmitted by the drifters; see an example in Figure 1. Each DS file contains satellite messages emitted by different drifters concerning location and geophysical parameter.

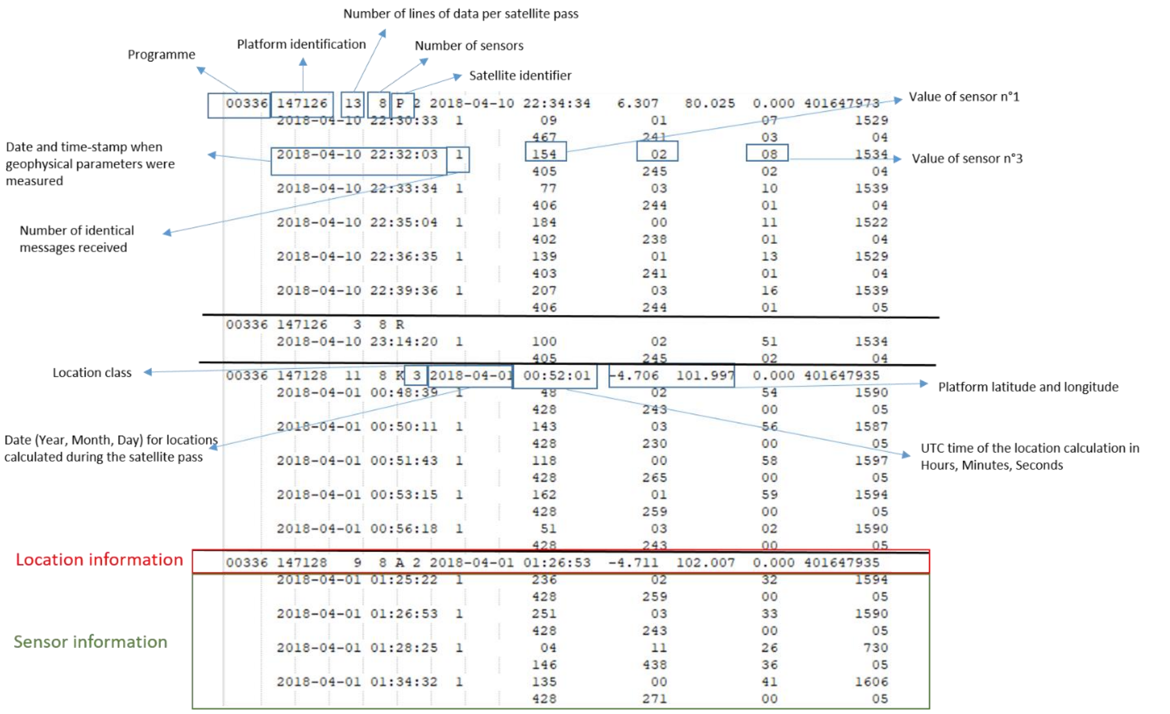

Figure 1: Example of a file in DS format with the information of four satellite passes. Horizontal black lines separate different satellite passes.

During this phase of the project, CLS finished the collect of all the Argos messages, data from 1997 to present. Unfortunately, Argos messages emitted by drifters before 1997 are not available anymore, they were lost. Part of these Argos messages, from 2002 to 2019, are also available in the AOML website (https://www.argos-system.org/manual/6-data/632_data_formats.htm).

We collected, treated and compared the PVR/DS files coming from both CLS and AOML. We found 16 471 Argos Ids in both CLS and AOML (doubled), 10 586 Argos Ids present only in CLS, and 1 094 Argos Ids present only in AOML dataset. Once doubles removed, we got a total of 28 151 Argos Id, from 1997 to 2019. This is our main source for decoding the C-RAID drifter data transmitted by Argos.

3.2 Already decoded data

The already decoded data downloaded from AOML are available from February 14th 1979 (deployment date of the GDP oldest drifter) to October 2019. This is our only source of C-RAID drifter data deployed between 1979 and 1996 (because we have not be able to retrieve Argos original drifter messages prior to 1997). Furthermore, these data were also used to fill time-gaps of decoded data (see section 4.3).

Data were stored in three ASCII columns format files containing the data of all the archived drifters. These files contain the following information:

- AOML drifter ID;
- Time of the location;
- Latitude of the drifter;
- Longitude of the drifter;
- Time of the measurement;
- Drogue information;
- SST measurement;
- Voltage information;
- Sensor4 to sensor6 (used to store additional sensor measurements);
- Location quality Index.

4. PROCEDURE

In this section, we describe the different steps performed to create the C-RAID drifter data set (formatting, decoding and merging data) and to quality control it (Real Time and Delayed Mode quality control phases) for generating the final C-RAID product.

In this phase, we focused on drifters with SST sensor or with SST and ATMS sensors, deployed between 1979 and 2010. Note however that data of drifters still active after 2010 were also processed in order to provide all the data measured by a same drifter.

4.1 Formatting data

4.1.1 Argos satellite messages

1) The Argos message files were first split by Argos identifier so that all the data received from a given Argos Id are gathered in the same file. Note that one Argos Id can be subsequently used by multiple drifters.
2) The second processing step consists in correcting the satellite pass headers from possible anomalies.
3) The resulting satellite passes are then cleaned from any duplicates.

After processing, we obtain one file per emitter identifier (ArgosId), containing cleaned and unique satellite passes.

4.1.2 Already decoded data

Already decoded data are available in ASCII columns format, multiple drifter data are present in the same file. The process consists to split these files by drifter Id.

4.2 Decoding data

The Argos satellite message (see Figure 1) contains location and sensor information. The first line of each satellite pass indicates location information: the date, time and location (longitude and latitude) calculated for each satellite pass. The following lines indicate the time of message reception together with the values of the geophysical parameters measured by the drifter.

For drifters using Argos transmission, the sensor information contained in the satellite message is only associated with the time-stamp of the Argos message reception, the position of each measurement should be inferred by time-position linear interpolation of the drifter trajectory.
Before interpolation, the drifter trajectory is subjected to RTQC tests (TEST14, TEST02, TEST03 and TEST16, detailed in section 4.4 of this document); only locations with QC equal to 1 and 2 were used to locate the sensor data.

Note that the measurement of geophysical parameters given in the satellite message is the original “raw” sensor output (generally expressed as “counts”), without physical meaning.

A calibration equation is necessary to obtain real value of the geophysical parameter.
For example, the received “raw” sea surface temperature is SST_COUNT and SST (in °C) is obtained as
SST = SST_COUNT * SST_SLOPE + SST_OFFSET
where calibration coefficients SST_SLOPE and SST_OFFSET should be provided in the metadata.

In order to decode the sensor data, the following information, normally available in metadata, is required:

a) Number of sensor mounted on the drifter;
b) Binary code pattern to decode the original “raw” sensor output;
c) Calibration equations for each geophysical parameter.

For the C-RAID project, we will implement decoders for each type of drifter, which depends on the number of sensors mounted in the drifter and the binary code pattern.
At this state of C-RAID project, we have implemented the decoder for two type of drifters:

a) Drifters with 3 sensors: with a SST sensor;
b) Drifters with 8 sensors: with SST and ATMS sensors.

4.3 Merging decoded data

By decoding the Argos satellite messages, we generate time series of location and geophysical parameters by drifter.
However, we have noticed that, in occasions, there are time periods without data in the Argos satellite message but data are present in the AOML already decoded data (i.e. original Argos data were not correctly archived).

In the merging process, we use AOML already decoded data to fill the time-gaps in the C-RAID drifter data set.

4.4 Real-Time Quality Control

The first step of the drifter quality control is to run automatic RTQC tests. We have adapted (from RTQC tests of Argo float data) or developed 16 RTQC tests.
These RTQC tests are listed in Table 1 with a brief description; they are detailed in the C-RAID drifters Quality Control Manual [RD1].

The RTQC tests are applied one after the other. When a data fails a test, the data is not further tested. Consequently, it is important to previously define the order the tests should be applied (see section 2.2 of [RD1]). After the run of all the RTQC tests, two new variables are created for each parameter (PARAM) indicating the QC given to the data and the RTQC test that first set the data its final QC: <PARAM>_QC and <PARAM>_QC_FAILED, respectively.

Table 1: List of RTQC tests automatically applied to drifters data

TEST NAME	TYPE OF DATA TESTED	BRIEF DESCRITION
TEST01: Platform identification	Identification	The drifter identification should have the identification format defined in C-RAID project.
TEST02: Impossible date	Time	The Julian day (JULD) of drifter data should be later than 1st January 1979 and earlier than the current date of the check (in UTC time).
TEST03: Impossible location	Location	The observed latitude and longitude should be sensible.
TEST04: Position on land	Location	The latitude and longitude from a drifter should be located in an ocean.
TEST06: Global range test	Geophysical parameters	The parameter measured by the drifter should be inside the limits defined for each parameter.
TEST07: Regional range	Geophysical parameters	This test affects data in specific regional seas. The parameter measured by the drifter should be inside the limits defined for each parameter.
TEST08: Time-Continuity	Geophysical parameter	The ratio $\frac{\Delta PARAMETER}{\Delta TIME}$ should be lower than the indicated threshold.
TEST09: Spike	Geophysical parameters	The absolute difference of the value tested with the precedent and following ones should be lower than a defined threshold.
TEST10: Digit rollover	Geophysical parameters	It checks a sudden increase or decrease in the parameter value due to a data storage overflow.
TEST11: Stuck value	Geophysical parameter	It checks whether a parameter remains constant during 5 days.
TEST12: Grey list	Geophysical parameters and location	The grey list is updated during the DMQC by the DMQC-operator, who visualizes the data and QC given by automatic RTQC tests. The operator indicates in the grey list periods of sensor malfunction, drifter on land, or data that were inadequately flagged by the RTQC.

TEST13: Argos redundancy	Geophysical parameters	It is applicable only to data of drifters with Argos transmission. It checks the possible corruption of the received message to provide a reliable information.
TEST14: Inside of Mission	Geophysical parameters and location	All the data transmitted by the drifter emitter are available in the C-RAID NetCDF file even before or after the drifter mission. However, this test flags to QC = 4 all the data timely outside the drifter mission.
TEST16: Questionable Argos position	Location	It identifies questionable Argos position data collected, considering the drifter speed at the sea surface and Argos position accuracy.
TEST19: Spike two points	Geophysical parameters	It tests high variability of a given parameter by comparison of the tested data with the data two time-steps before and after. The absolute difference should be lower than a defined threshold.

4.5 Delayed-Mode Quality Control

This section details the steps carried out after the run of the automatized RTQC tests.

4.5.1 Comparison with ERA5 data

First, we implemented a comparison test for SST and ATMS drifter data with co-localized, in time and space, ERA5 data. ERA5 is a product of reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). It provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. The data cover the Earth on a 30 km grid. For details of ERA5 and data access visit https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.
Because ERA5 is a product of reanalysis, quality-assured data are available within 3 months of real time. Consequently, this test is not directly in the RTQC procedure. However, when dealing with historical data, as in C-RAID, this test should be directly applied in the RTQC procedure.

The methodology for preparing the ERA5 comparison test is the following:

- First, the RTQC tests affecting location data (TEST02, TEST03 and TEST16) were run. Only position with QC equal to 1 or 2 were selected to do the interpolation of sensor data (see section 4.2), and so, the co-localization with ERA5 data.
- Then, the ERA5 data were spatially interpolated to the drifter positions, using the ERA5 time step the nearest to the time-span of the drifter position. The Python function “interp” of “xarray” package (http://xarray.pydata.org/en/stable/generated/xarray.Dataset.interp.html) was used for the 2D – linear interpolation. If only 2 or 3 positions around the drifter positions were in the ocean (not on land), the meridional and/or zonal gradient of the available data were used to interpolate the parameter to the drifter position. The recovery of ERA5 data was carried out in DATARMOR calculator hosted at Ifremer (Brest, France).

Moreover, ERA5 wind data, specifically wind stress and wind velocity at 10 m, were also co-localized to the drifter positions (for further studies). All the co-localized ERA5 data are provided together with the drifter data.

4.5.2 On land DMQC test

Whether a drifter position is on land is checked in the RT on land test. This test estimates the mean values of the 1, 2 or 4 elevations, depending on the drifter position, on the GEBCO_2020 grid. The test fails if the mean elevation is ≥ 0. However, this test can fail when the drifter is in shallow waters, and so, in the ocean, e.g. near an island, over a bank, inside an atoll. Therefore, we propose a DMQC on land test as follow.
The trajectories of drifters failing the RT on land test are generated in KML format. The DMQC-operator visualizes the trajectory of the drifters using Google Earth (https://www.google.com/earth/) in order to verify whether the drifter was on land (bathymetry >= 0 m). When a drifter is on land, all the technical parameters, locations and geophysical parameters, should be flagged as bad data (4). The operator should indicate it in the grey list, indicating PARAMETER = MISSION, and the period of time the drifter was on land. Additionally, drifter positions inside an atoll, on the beach affected by the tide, or blocked in the ice, should be also be indicated in the grey list, without QC. This action is just to indicate that, despite the drifter was in the ocean, it was in shallow waters and/or blocked.
Once the DMQC on land test was carried out, the RTQC procedure should be ran again with the RT on land test (TEST04) deactivated.
By the process of visualization of drifter trajectories in Google Earth we can also identify:

a) Drifters whose trajectories were affected by human actions.

For example, a drifter was recovered in a boat, brought to port and deployed again later, this situation is clearly detectable (see example in Figure 2). In situation like this, we consider that the drifter trajectory was modified by human actions, and two missions with two different C-RAID identifiers are defined.

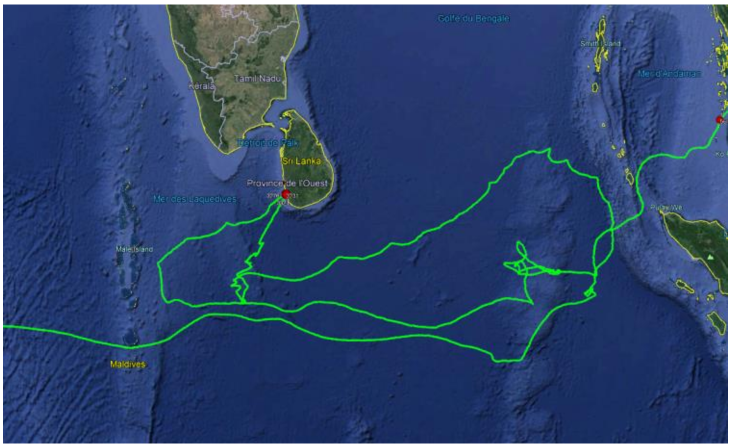

Figure 2: Trajectory of C-RAID drifter 1270087; the drifter was recovered, brought to port and deployed again later

b) Drifters with deployment data later than the metadata one.

This case is identified when i) the first positions of the drifter are located on land, or ii) the drifter trajectory is a straight line and its surface velocity is greater than 3 m s-1, which indicates that the drifter was still on board the deployment boat.

c) Drifters with end mission data earlier than the metadata one.

This case is identified when i) the last positions of the drifter are located on land, or ii) the drifter trajectory is a straight line and its surface velocity is greater than 3 m s-1, which indicates that the drifter has been recovered.

4.5.3 Visualization of time series of drifter data and the QC given during the RT tests

The time series of geophysical parameters (SST and ATMS) with their QC are visualized drifter by drifter. This step allows:

a) The identification of biased or wrong data not detected by the RTQC tests
b) The identification of potentially reliable data flagged as “bad data” during the RTQC process.

For this step, a specific scientist skill is required.
For this propose, several discussion sessions about SST were carried out between the DMQC operator (Patricia Zunino) and the CLS colleagues (Hélène Etienne, Stéphanie Guinehut and Christine Boone). These discussions were very fruitful for the DMQC operator, who proposed a protocol for the SST data. The discussions about ATMS were less fluent; they were limited to the exchange of several emails directly with Gilles Reverdin (LOCEAN) and with Christophe Billon (METEOFRANCE) who transferred the questions to meteorologist colleagues. With this low exchange, the DMQC operator could however propose a protocol to the DMQC of ATMS data.

In order to facilitate the DMQC operator decisions, ERA5 data and AOML already decoded data are visualized together with our decoded data. It is interesting to note that the ATMS data provided by the AOML are very rough; it is probably because they have not been quality controlled yet. Unfortunately, there is no documentation about that, and AOML colleagues have never answered to our questions.

During the visualization of the time series of geophysical parameters, we also detected some problems in the sensor calibration. The calibration equations initially given in the metadata were used during the decoding step (see section 4.2). However, we found that for some time periods, or even for the whole drifter mission, data presented the same time variability than ERA5 or AOML already decoded data, but they were different in magnitude. Because of that observation, we considered the data suspicious and new calibration equations were estimated by comparison with the AOML or ERA5 data. The data resulting of the new calibration are coherent with the AOML already decoded data. The new calibration equations estimated in C-RAID are available in the provided NetCDF files.

Besides, the visualization of the time series of geophysical parameters are also useful to have clues or insights that:

a) The drifter was on air, which is clearly observable by the large SST daily variability. It is indicative that the drifters was recovered or on land. By the visualization of the time series of geophysical parameter, the DMQC operator can suspect a drifter finished its mission, and so use another tool, for example the visualization of drifter trajectory (see section 4.5.4), to validate whether her/his suspicion was true.
b) The drifter started/continued measuring geophysical parameters before/after the deployment/end mission dates. Our tool is able to display all the data transmitted by the drifter Argos emitter, even before and/or after the indicated mission dates. The DMQC operator can suspect the drifters was naturally drifting in the ocean before or after the indicated mission dates, and so use another tool, for example the visualization of drifter trajectory (see section 4.5.4) to validate whether her/his suspicion was true.

4.5.4 Visualization of drifter trajectory and a first estimator of drifter velocity

We also developed another tool to visualize the drifter trajectory. At the same time, we calculate an estimator of the surface velocity of the drifter, simply by using the location positions considered as good by the RTQC tests. Despite this estimator is not precise, it gives insights whether the drifter was naturally drifting in the ocean, on board à vessel, or blocked against a bank.
The visualization of the drifter trajectory and the time series of drifter velocity is carried out simultaneously with the visualization of time series of geophysical parameters.

The tool represent:

a) All the positions transmitted by the drifter, even before/after the deployment/end mission dates indicated in the metadata,
b) The QC of positions in the time period deployment date - end mission date,
c) The time series of surface velocity of the drifter.

It is very useful:

a) To validate the deployment date indicated in the metadata. Sometimes we find many positions before the deployment date showing a natural drifter trajectory with velocities lower than 3 m s^{-1}. In this case, the deployment date should be set earlier. Differently, we can detect straight trajectory with velocities greater than 3 m s^{-1}, which is indicative the drifter was still on board the deployment boat. In this case, the deployment date should be set later.
b) To validate the end mission date indicated in the metadata. Sometimes we find many positions after the deployment date showing a natural drifting trajectory with velocities lower than 3 m s^{-1}. In this case, the end mission date should be set later.
c) To validate the drifter was recovered before the indicated end mission date. We see that the drifter shows a straight trajectory with velocities greater than 3 m s^{-1}. In this case, the end mission date should be set earlier.
d) To identify two mission for the same drifter.Two cases are possible:
- First, we see a straight trajectory and velocities greater than 3 m s^{-1} indicating the drifter was recovered, and later we see the drifter was drifting naturally in the ocean.
- Second, we see a straight trajectory to a port, and later, we see a straight trajectory to the ocean and then the drifter drifting naturally.

  In both cases, the trajectory of the drifter was affected by human actions and two missions, with two C-RAID identifier, should be set for that drifter.

4.5.5 Changes in the deployment date or end mission date initially indicated in the metadata or identification of two missions

During the visualization of time series of geophysical parameters (section 4.5.3) and the drifter trajectories
(section 4.5.4) the operator can decide to change the deployment date or end mission date initially given in the metadata. These modifications are indicated in the grey list.

Considered the original set of 10 039 drifters available in this phase of the C-RAID project, the number of drifters whose deployment /end mission dates were modified, and the drifters with two missions, are indicated in the Table 2.

Table 2: Information about the number of drifters in C-RAID Phase 2 Spring Delivery, for which the metadata have been partially modified

	TOTAL C-RAID 1
INITIAL NUMBER OF DRIFTERS PROCESSED IN DMQC	9 993
NUMBER OF INITIAL DRIFTERS WITH TWO MISSIONS	46
FINAL NUMBER OF DRIFTERS PROCESSED IN DMQC	1 039
N° DRIFTER DEPLOYMENT DATE CHANGED	284
N° DRIFTER END MISSION DATE CHANGED	786

By changing the deployment date and the end mission date, and considering the new mission of the drifters with two missions, we recuperate 102.5 years of data.
The user can find in the grey list the drifters for which the metadata have been partially modified. For example, filter:

- PARAMETER_NAME = TWO_MISSIONS to find drifters with two missions;
- PARAMETER_NAME = CHANGE_DEPLOY_DATE to find drifters whose deployment date has been modified;
- PARAMETER_NAME = CHANGE_END_MISSION_DATE to find drifters whose end mission date has been modified.

4.6 NetCDF files generation

The C-RAID drifter data set is transmitted in NetCDF files compliant with the dedicated format (see [RD2]).

The mission of a drifter is the time period which starts at drifter deployment date and ends when the drifter is recovered (or considered as lost). Each drifter mission has a unique C-RAID drifter identifier (stored in DRIFTER_NUMBER NetCDF variable).
If a drifter is recovered, reconditioned and deployed again. It is a new mission (with a new C-RAID drifter identifier).
One NetCDF file is generated for each drifter mission. It contains the meta-data of the drifter and all data sampled during the mission. When available, additional data (received from the drifter transmission identifier but sampled outside the mission time period) are also stored in the NetCDF file (and affected with a QC flag 4).

The file naming convention of C-RAID NetCDF files is *TYYNNNN*.nc.
Where *TYYNNNN* is the C-RAID drifter identifier for which:

- *T*: refers to the drifter transmission type (1: for Argos, 2: for Iridium);
- *YY*: refers to the drifter deployment date (00: if deployed before 1979, YY+1978: deployed year otherwise. E.g. *YY* = 22 for a drifter deployed in 22+1978 = 2002);
- *NNNN*: is a number designed so that the C-RAID drifter identifier is unique.

5. UPDATE DATA OF DRIFTERS ALREADY IN C-RAID PHASE 1 DATABASE

The C-RAID Phase 1 database was produced from an incomplete and temporary dataset of Argos messages. The collect of all the Argos messages was finally achieved by our partner CLS during this second phase of the project, exactly on April 2021. The new dataset of Argos messages is richer in data than the previous one. Therefore, all the decoded data of drifters in the C-RAID Phase 1 database have been re-decoded. The differences between both repetitions are mainly due to the absence of Argos messages for specific periods in the first Argos messages dataset that are now available in the new Argos messages dataset. The DMQC process has been performed anew when drifter data in both repetitions differed.

6. DELIVERY

The product generated during the second phase of the C-RAID project contains documents and data.

The documents delivered are:

a) The activity report of C-RAID project phase 2 (this document);
b) The C-RAID drifters Quality Control Manual [RD1];
c) The C-RAID drifters NetCDF format reference manual [RD2].

In the C-RAID drifter data set, we find, for each C-RAID drifter identifier:

- The NetCDF file containing the drifter data. It is a NetCDF file compliant with the C-RAID dedicated format, named “C-RAID_ID.nc” (where C-RAID_ID stands for C-RAID drifter identifier).
- Complementary data files used during the process (or to be used during further studies) which are:
    - The original Argos satellite messages. It is an ASCII file at the PRV/DS format, named “ArgosId_YYYY-MM-DD-hh-mm-ss_YYYY-MM-DD-hh-mm-ss.txt” (where ArgosId stands for drifter emitter identifier). These data have been used during the decoding process.
    - The SST and ATMS data extracted from ERA5 and co-localized to the drifter positions. It is an ASCII file named “C-RAID_ID_sst_msl_era5_extracted.txt”. These data have been used during the quality control process.
    - The wind stress data extracted from ERA5 and co-localized to the drifter positions. It is an ASCII file named “C-RAID_ID_ewss_nsss_era5_extracted.txt”. These data have not been used during the C-RAID quality control process, they are provided for future studies.
    - The wind velocity at 10 m data extracted from ERA5 and co-localized to the drifter positions. It is an ASCII file named “C-RAID_ID_u10_v10_extracted.txt”. These data have not been used during the C-RAID quality control process, they are provided for future studies.
    - A README file, with the headers of the four ASCII complementary data files.

Complementary data files are provided in a zipped archive named "CRAID_ID_complementary_data.zip".

Additional files are also provided, they concern:

- The whole data processing. We provide the following lists (of AOML drifter identifiers)
  - " C-RAID-phase2_driftersId_ALL.txt" contains the list of all the drifters of the data set delivered in C-RAID-Phase2-SpringDelivery;
  - " C-RAID-phase2-springDelivery_driftersId_SST_1979-1996.txt" contains the list of drifters that measure only SST, deployed between 1979 and 1996;
  - " C-RAID-phase2-springDelivery_driftersId_SST_1997-2001.txt" contains the list of drifters that measure only SST, deployed between 1997 and 2001;
  - " C-RAID-phase2-springDelivery_driftersId_SST_2002-2010.txt" contains the list of drifters that measure only SST, deployed between 2002 and 2010;
  - " C-RAID-phase2-springDelivery_driftersId_SST-ATMS_2002-2010.txt" contains the list of drifters that measure SST and ATMS, they were deployed between 2002 and 2010;
  - "C-RAID_ID_vs_AOML_ID_Phase2.xlsx" contains the AOML drifter identifier VS C-RAID drifter Identifier list.

- The metadata processing phase. The "finalize_aoml_meta_ALL_1_20191017T141257.xlsx", "finalize_aoml_meta_ALL_2_20191017T141257.xlsx" and "finalize_aoml_meta_ALL_3_20191017T141257.xlsx" excel files contain the result of the metadata processing phase for the 22 317 AOML drifters deployed before January 1st 2019 (excluding the 156 ones with inconsistencies in their metadata).
- The DMQC phase. The "GREY_LIST_C-RAID_PHASE2_20210616.xlsx" excel file is the grey list completed during the delayed mode process of drifters treated in C-RAID Phase 2.

Different to the C-RAID Phase 1 database, in C-RAID Phase 2 – Spring Delivery the whole DMQC process is finished for all the drifters. Details about both deliveries are provided in table 3.

Table 3: Details about the dataset in C-RAID Phase 2 Spring Delivery and in C-RAID Phase 1

	C-RAID Phase 2 - Spring Delivery	C-RAID Phase 1
Number of Drifters	10 039	7 493
Number of years	10 391	8 650
Period	1979 - 2010	1997 - 2010
Quality Control Level	all in DMQC	5 154 drifters in DMQC / 2 339 drifters in RTQC
Parameters	8 498 drifters SST & 1 541 drifters SST-ATMS	5 992 drifters SST & 1 501 drifters SST-ATMS