Framework Service Contract EEA/DIS/R0/20/001 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
Autumn Delivery 2022 – Activity Report
Patricia ZUNINO and Jean-Philippe RANNOU

Issue: 5.0
Date: 07/11/2022

[Document Release]

	Name(s)	Affiliation
Coordinated by:	Thierry Carval	Ifremer
Contributions:	Patricia ZUNINO RODRIGUEZ Jean-Philippe RANNOU	Capgemini Engineering Capgemini 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 SC2

[Change Record]

Version	Date	Changes
1.0		First draft release

REFERENCES

Reference N°	Title	Link
#RD1	C-RAID drifters NetCDF format manual.	https://www.seanoe.org/data/00660/77184/
#RD2	C-RAID drifters quality control manual.	https://www.seanoe.org/data/00660/77184/
#RD3	Lopez, R. and Malardé, J-P, (2011). Improving Argos Doppler Location Using Kalman Filtering.	https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6648418

Table of Contents

1. Introduction

The objectives of the C-RAID project are to gather, to treat, to decode and to make available all the historical data measured by drifting buoys 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. Both groups of drifter data will be deal with in the C-RAID database. So far, only drifter data with Argos positioning have been included in the C-RAID database.

The C-RAID database has been implemented in several Phases:

1. C-RAID Phase 1, funded by EEA work on Specific Contract 4 n° 3436_R0-COPERNICUS_EEA.57652. In this Phase 1 we dealt with data of drifting buoy deployed from 1997 to 2010 and present in the AOML database. The main steps of the work done during this phase were:
    - 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).

   At the end of C-RAID Phase 1 the database contained data of 7 493 drifters, deployed between 1997 and 2010. The accumulation of these data was equivalent to 8 650 years of data.

2. C-RAID Phase 2, funded by Framework Service Contract EEA/DIS/20/001 for services supporting the EEA's cross-cutting coordination of Copernicus' in situ data activities. During this phase the goal was to include in the C-RAID database data of drifting buoy deployed between 1979 and 2010 and present in the AOML database. The main steps of the work done during this phase were:
    - 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).

   At the end of C-RAID Phase 2 the database contained data of 10 039 drifters (those of Phase 1 included), deployed between 1979 and 2010. The accumulation of these data was equivalent to 10 391 years of data. The DMQC process was fully finished for the data of the 10 039 drifting buoys.

3. C-RAID Phase 3, funded by the MC0 – CORIOLIS 2021 and 2022 contracts.
   The objectives in this phase were:
    - To introduce the Argos fixes repositioned by the Kalman filter.
    - Deal with data of drifting buoys rescue from the CLS data base and not present in the AOML database.

4. C-RAID Phase 4, funded by funded by Framework Service Contract EEA/DIS/20/001 for services supporting the EEA's cross-cutting coordination of Copernicus' in situ data activities.
   The objectives in this phase were:
   - Assessment of the incorporation of Argos fixes re-calculating by the method based on the Kalman filtering.
   - To continue the formatting, decoding and quality control of data of drifting buoy with Argos positioning.

   At the end of this phase 4, we provide data of around 14 000 drifters:
   - Drifters dealt with in Phase 1 et 2, with recalculated Argos positions
   - drifters recovered from the CLS database and not present in the AOML database
   - New drifters treated during this phase.

2. Metadata

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.

The calibration equation of the sensors mounted in the drifting buoy are listed in this file. This information was very useful for decoding our data. Unfortunately, this information was available only for drifters deployed between 1997 and 2010.

2.3 Metadata for data exclusively available in the CLS database

All the data available in the GDP website are available in CLS database but not the contrary, this point is explained in section 4.3. CLS transferred some information, mainly related to commercial purposes, about drifters Id exclusively present in their database. However, any information concerning decoder template, calibration equations or start and end mission dates were available. The metadata for these drifters were created by us by gathering, editing, and cleaning the metadata transferred by CLS and the data processing, this point is explained in section 4.3.4.

2.4 Metadata processing

2.4.1 GDP metadata processing

We processed the metadata of 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. Without further information 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”

2.4.2 CLS metadata processing

The metadata for drifters ‘only CLS’ were created by us by gathering, editing, and cleaning the metadata transferred by CLS and our data processing, this point is explained in section 4.3.4.

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/);

The main provider of original Argos data in the C-RAID project is our partner CLS. During the C-RAID Phase 1, CLS realized their tasks of data recompilation exclusively measured by drifters and all the treatment that this entails. Meanwhile, we collected drifter Argos data available in an AOML ftp site (ftp://ftp.aoml.noaa.gov/phod/pub/pazos/data/argos_cd) and we used them in order to develop our tools to generate the delivery products of the C-RAID project. In C-RAID Phase 2, CLS transferred a comprehensive delivery of Argos messages emitted by drifting buoy. From C-RAID Phase 2 hereafter, the CLS delivery is our main source of data to decode drifter data.

3.1 GDP Drifter Argos messages

The drifter Argos messages downloaded from AOML are archived monthly from 2002 to 2019 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 3.1. Each DS file contains satellite messages emitted by different drifters concerning location and geophysical parameter.

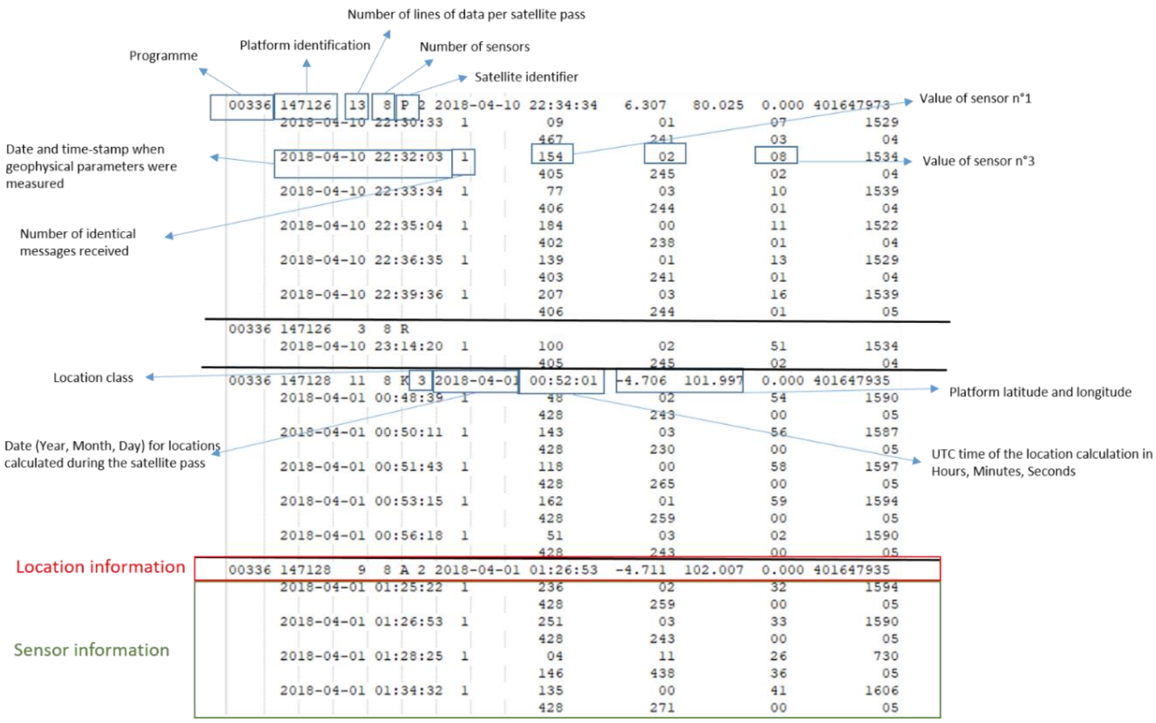

Figure 3.1 : Example of a file in DS format with the information of four satellite passes. Horizontal black lines separate different satellite passes.

3.2 CLS Drifter Argos messages

During the C-RAID Phase 2 we received a comprehensive delivery of:

- The drifting buoy Argos fixes estimated by i) the Least Square method, ii) the Kalman Filtering method, and iii) the Smothing Kalman filtering fixes.
- All the Argos messages emitted by drifting buoy existing in the CLS database. This database contains more Argos Ids than the GDP drifter Argos messages.

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.

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.

These data were used to validate our decoded data and to filling time-gaps when data are not available in the CLS database.

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.

4.1 Formatting data

4.1.1 Argos satellite messages

- 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.

- The second processing step consists in correcting the satellite pass headers from possible anomalies.
- 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 archived by months. The process consists in splitting these files by drifter Id.

4.2 Decoding data of drifters with metadata

The Argos satellite message (see Figure 3.1) 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 4.3 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 sensors 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.
In the Phases 1 and 2 of C-RAID project, we have implemented the decoder for two types of drifters:

a) Drifters with 3 sensors: with a SST sensor;

b) Drifters with 8 sensors: with SST and ATMS sensors.

4.3 Decoding data of drifters without metadata

This section is specifically for decoding data of drifting buoy without metadata, without information of deployment date and end mission date, number of sensors mounted in the drifters, calibration equations, etc … We explain the procedure carried out during the phase MCO – CORIOLIS for decoding the Argos messages of drifters found exclusively in the CLS database without metadata.

Argos data stored in CLS database and not in the AOML database concern 9 918 Argos Ids ("only CLS" database). CLS transferred some information about them, mainly related to commercial purposes. This information is: "Platform Id", "Program Number", "SUA_ID" (customer Id), "WMO", "location algorithm" (Least Square or Kalman), "start valid date", "end valid date", "family", "activity", "program name" or "operating organization".
The start and end valid dates indicate the first and last dates of the satellite messages.
The family can be "bird", "car", "container", "drifter", "fix station", "ice buoy", "land animal", "marine animal", "mooring buoy", "ship", "sub float", "underwater stat", "underwater vehicle" or "unknown".
The activity is indicative of the instrument the Argos antenna was mounted on: "acoust", "albatross", "argo_float", "bird", "car", "cargo", "caribou", "container", "deer", "drift", "eagle", "elephant", "fad", "figge", "float", "fox", "glider", "gull", "hawk", "ice buoy", "land animal", "marine animal", "mb wave", "met", "mooring buoy", "pinguin", "petrel", "poll track", "pop up", "porpoi", "profil", "river stat", "seal", "shark", "ship", "stork", "sub float", "sub surf", "SVP", "SVP_B", "SVP_BW", "swan", "turtle", "unknown", "vessel", "vos whale", "whale_sh", "yacht" or empty.
However, any information concerning decoder template, calibration equations or start and end mission dates were available.
From this information we deduced that the data of the 9 918 Argos Ids are not exclusively drifter data. We selected a theoretical list of Argos Ids by filtering family equal to "drifter" which resulted a total of 4 801 Argos Ids. These selected data from the "only CLS" database were potentially drifter data and were initially considered in this contract.

4.3.1 Identification of data decoder templates for each Argos Id

In this section we explain the methodology applied to identify the data decoder template of each Argos Id. A same Argos Id can be used on different instruments over the time and then have a number of sensors during a period, and a different number of sensors during another period. In order to decode the satellite messages, we need to know i) how many sensors were mounted on the drifter, ii) the parameters measured by the sensors and iii) the order in which the data of the measured parameters were registered in the satellite message. None of this information were known for the group of 9 918 Argos Id exclusively in the CLS database. To decode them, first, we identify the number of sensors the drifter probably had by reading the satellite messages. Second, we decode with the decoder templates already implemented. If the decoded parameters are not the parameter we look for, e.g., SST, ATMS or DROGUE, we implemented new decoder templates by changing the order of the sensors.

4.3.1.1 Dealing with Argos Id with three, four or eight sensors

We initially processed the Argos Ids and periods with three, four and eight sensors for which four decoder templates were already implemented in Phases 1 and 2 of the C-RAID project.

We found ~340 Argos Ids with 3 sensors, ~ 2 350 with 4 sensors and ~ 450 with 8 sensors.

Then, Argos Ids with hypothetically three, four or eight sensors were decoded. Next, their trajectories and time series of decoded parameters were visualized in order to verify that:

- the device was drifting in the ocean.
- the parameters were correctly decoded: for SST and ATMS by comparison with ERA5 data co-localized to the Argos fixes and for the drogue and battery voltage by operator experience.

Note that ERA5 database is a re-analysis product. It provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. The data cover the Earth on a 30km grid. For details of ERA5 and data access visit https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. The ERA5 data were spatially interpolated to the measurement drifter positions, using the ERA5 time step the nearest to the measurement time. 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.

The parameters decoded with our four initial decoder templates are: Sea Surface Temperature (SST), drogue (which indicates de submergence of the device), battery voltage, and sea level air pressure (ATMS), each of them in a specific position in the satellite message.

For Argos Ids with parameters wrongly decoded, we have implemented new decoder templates by changing the order of the sensors associated to each parameter. So, when a parameter was wrong, we re-decoded the data with other decoder template up to find SST, ATMS, drogue or battery voltage data. When the data registered in the satellite message was not identified as SST, ATMS, drogue or battery voltage, the data of the unknown parameter is stored as UNKNOWN_PARAMETER. That is, all the information, identified or not, in the satellite message are transferred in the C-RAID NetCDF files (see details in the C-RAID NetCDF format manual [R2]). For checking the decoded data of the ~ 3150 Argos Ids we have implemented a total of 27 decoder templates summarized in Table 1, all of them with less than eight sensors. The details about the nomenclature of decoder Ids are exposed in section 4.3.1.3.

Difficulties found

- Reading the satellite messages: the drifter could transfer the data in decimal format, hexadecimal format or both formats at the same time. To correctly read the satellite messages, we implemented several tools.
- Identifying the decoder template: for a same Argos Id the number of sensors can change from one period to another, with the two periods overlapped. For example, from January 1999 to March 2002 an Argos Id present 4 sensors, and from April 1999 to February 2003, the same Argos Id present 8 sensors. In this case the Argos Id was decoded twice, and two different decoder templates could be retained; we retained the decoder templates providing more correct SST data. Note that SST is our main reference parameter to validate a decoder template.

- Identifying the decoder template: for some Argos Ids, data decoded as SST presented large daily variability indicating that the sensor was measuring air temperature rather than SST. In these cases, from the information given by CLS, we discovered that these Argos Ids, initially classified as "drifter", were associated to "FAD" (Fishery Aggregated Devices), or to "ice-buoys", or to a project related with pelagic or fishery issues. We have then discarded these Argos Ids from the final drifter data set. These data, concerning to 91 Argos Id, were already decoded, so, they are also delivered in separate datasets.
- Identifying the decoder template: the variable DROGUE is usually given in the drifter data, however not all the time at the same position in the satellite message. The time series of DROGUE have been carefully visualized. It has been decoded as DROGUE only when the operator thought that it was really DROGUE values, it was decoded as UNKNOWN_PARAM otherwise. The DROGUE could be measured with a strain gauge sensor (presence or not of the drogue) or with a submergence sensor. The criterium to decide whether the variable is the DROGUE depends on the method used. In the case of strain gauge sensor, the value is constant and drop to 0 when the drogue is lost. In the case of the submergence sensor, the results are not so clear.
- A group of Argos Id with eight sensors and decoder Id 1000011 provided ATMS values with the same time variability as the ERA5 co-localised ATMS, but different magnitude. It is very probably that ATMS calibration equation different to the standard was used. The method to find a right calibration equation is exposed in section 4.3.3.3.

4.3.1.2 Dealing with Argos Id with more than eight sensors

The following step was to analyze the Argos Ids and periods the satellite messages presented more than eight sensors. At this point, there were around 3 400 Argos Ids that could be decoded. The identification of parameters in the satellite message was done by eye looking at the .DS file, and by comparison with SST, ATMS or AIR_TEMP from ERA5 database co-localized to the drifter positions. Therefore, when the satellite message has data for more than eight sensors, the identification of the measured parameters is a very tedious task. Moreover, if the values corresponding to the sensors data are in count, that is, a calibration equation is needed to decode the data, the task became even more complicated because we do not have the calibration equation. Consequently, when more than eight sensors were identified in the Argos satellite message, we studied only the Argos Ids with decimal values (with physical meaning), and temperature (SST or AIR_TEMP) or ATMS in the first five sensor values. From the initial group of 3 400 Argos Ids, only 560 Argos Ids fulfilled the condition. The decoder Id 1000112 was initially implemented to decode this group, it was successful for 385 Argos Ids of the 560 Argos Ids studied. Then, to decode the rest of the 560 Argos Ids, we implemented new decoder templates (1000113, 1000116, 1000118, 1000124, …) by changing the order of the sensor that measured SST or ATMS. Furthermore, we realized that, for a group of Argos Ids, the variable initially decoded as SST presented high daily and interannually variability and compared correctly with the air temperature at 2 m over the seal level of ERA5. Consequently, we implemented four more decoder templates for decoding Argos Ids that measured air temperature (AIR_TEMP) and ATMS. Interestingly, the Argos Id decoded with the former templates were drifting in the Arctic and Antarctic Oceans.

4.3.1.3 Decoder templates

In this section we describe the method implemented to obtain the forty decode templates indicates in table 1. Note that the satellite message can contain both count data (i.e. direct sensor output) and/or decimal data with physical meaning. The decoder templates depend thus on i) the number of sensors, ii) the order of appearance of the parameter associated to each sensor, and iii) whether the data were an integer (count) or a decimal (data with physical meaning). Between the decode templates in table 1 we can identify four different groups:

- 10000**XX**: the sensor values in the satellite messages are integer, they need a calibration equation to obtain a value with physical meaning.
- 1000**1XX**: the sensor values in the satellite messages are decimal (with physical meaning), they do not need a calibration equation to be decoded.
- 1000**2XX**: the sensor values in the satellite messages are unrecognizable as drogue, battery voltage, SST, ATMS or AIR_TEMP, but their positions shown a nice drifter trajectory.
- 1000**3XX**: the sensor values contained in the satellite messages are decimal, nevertheless, SST is not identified, and only the variables AIR_TEMP and ATMS are identified.

The index XX of each decoder template corresponds to an evolution of an initial decode template by changing the order of the parameters and the number of sensors identify in the satellite message.

The large variability of decoder templates was mainly developed during the Phase MCO – CORIOLIS. However, these decoder templates were later used during the C-RAID Phase 4. In fact, the Argos messages with sensor values in decimal is not only in the Argos messages of drifters 'only CLS' but also in the Argos messages of drifters in both CLS and AOML database.

Table 1. List of decoder templates implemented in the C-RAID project. It indicates de decoder Id. The C-RAID Phase when the decoder template was implemented. The number of sensors to be decoded. The decode type indicates whether the sensors values were in count (a calibration equation was needed), or whether the sensor values were directly read from the message satellite (values in decimal). The last column indicates the parameters to be decoded and the order they are in the satellite message. Struck out lines indicate that the decoder template has not been finally used.

Decoder Id	Implemented in C-RAID Phase	Number of sensors	Decode type	Parameters and order
1000001	1	3	calibration equation	Drogue (sensor#1), Battery Voltage (sensor#2), SST (sensor#3)
1000002	1	8	calibration equation	Format 8.0
1000003	1	8	calibration equation	Format 8.02: Rank and Air pressure tendency == 0
1000004	1	4	calibration equation	Drogue (sensor#1), Battery Voltage (sensor#2), SST (sensor#3)), with the last sensor unused (== 0)
~~1000005~~	~~MCO-COR~~	~~3~~	~~calibration equation~~	~~Drogue (sensor#1), SST (sensor#2), Battery Voltage (sensor#3)~~
1000006	MCO-COR	4	calibration equation	Drogue (sensor#1), SST (sensor#2), Battery Voltage (sensor#3), with the last sensor unknown (~= 0))
~~1000007~~	MCO-COR	4	calibration equation	SST (sensor#1), Drogue (sensor#2), Battery Voltage (sensor#3), with the last sensor unknown (~= 0)
1000008	MCO-COR	3	calibration equation	Battery Voltage (sensor#1), SST (sensor#2), Drogue (sensor#3)

1000009	MCO-COR	4	calibration equation	Drogue (sensor#1), Battery Voltage (sensor#2), SST (sensor#4), with the sensor #3 unknown (~= 0)
1000010	MCO-COR	4	calibration equation	Drogue (sensor#1), Battery Voltage (sensor#1), SST (sensor#1), PARAM_UNKNOWN (sensor#4)
1000011	MCO-COR	8	calibration equation	Drogue (sensor#7), Battery Voltage (sensor#8), ATMS (sensor#4) and SST (sensor#5), 4 PARAM_UNKNOWN
~~1000012~~	MCO-COR	8	calibration equation	Drogue (sensor#7), Battery Voltage (sensor#8), ATMS (sensor#3) and SST (sensor#4), 4 PARAM_UNKNOWN
1000101	MCO-COR	3	direct	Drogue (sensor#1), Battery Voltage (sensor#2), SST (sensor#3)
1000102	MCO-COR	4	direct	Drogue (sensor#1), Battery Voltage (sensor#2), SST (sensor#3), sensor#4 unused (==0)
1000103	MCO-COR	3	direct	Drogue (sensor#1), SST (sensor#2), Battery Voltage (sensor#3)
1000104	MCO-COR	8	direct	Drogue (sensor#6), Battery Voltage (sensor#2), ATMS (sensor#1), SST (sensor #3), 4 PARAM_UNKNOWN
1000105	MCO-COR	8	direct	Drogue (sensor#7), ATMS (sensor #4), SST (sensor #5), 5 PARAM_UNKNOWN
1000106	MCO-COR	8	direct	Drogue (sensor #7), ATMS (sensor#1), SST (sensor#3), 5 PARAM_UNKNOWN
1000107	MCO-COR	4	direct	Battery Voltage (sensor#2), ATMS (sensor#1), SST (sensor#3), Drogue (sensor#4),
1000108	MCO-COR	4	direct	Batt (sensor #3), SST (sensor #4), PARAM_UNKNOWN (sensor#1), PARAM_UNKNOWN (sensor#2)
~~1000109~~	MCO-COR	4	direct	Drogue (sensor#1), SST (sensor#2), Battery Voltage (sensor#3), sensor#4 unused (==0)
1000110	MCO-COR	4	direct	SST (sensor#2), Drogue (sensor#3), Battery Voltage (sensor#1), sensor#4 unused (==0)
1000111	MCO-COR	4	direct	Drogue (sensor#1), Battery Voltage (sensor#2), SST (sensor#3), PARAM_UNKNOWN (sensor#4)
1000112	MCO-COR	> 8	direct	ATMS (sensor#1), SST (sensor #3), Unknown parameters not indicated
1000113	MCO-COR	> 8	direct	ATMS (sensor#4), SST (sensor #5), Unknown parameters not indicated
1000116	MCO-COR	> 8	direct	ATMS (sensor#2), SST (sensor #3), Unknown parameters not indicated
1000118	MCO-COR	> 8	direct	ATMS (sensor#1), SST (sensor #5), unknown parameters not indicated
1000124	MCO-COR	24	direct	Drogue (sensor#6), Battery (sensor #2), ATMS (sensor#1), SST (sensor #3), 20 PARAM_UNKNOWN
1000201	MCO-COR	3	only trajectory	3 PARAM_UNKNOWN
1000202	MCO-COR	4	only trajectory – with last sensor = 00	3 PARAM_UNKNOWN
1000203	MCO-COR	4	only trajectory – with last sensor ~= 00	4 PARAM_UNKNOWN

1000204	MCO-COR	8	only trajectory	8 PARAM_UNKNOWN
1000301	MCO-COR	3	direct	Battery (sensor#2), ATMS (sensor#1), AIR-TEMP (sensor#3), unknown parameters not indicated
1000302	MCO-COR	4	direct	Battery (sensor#2), ATMS (sensor#1), AIR-TEMP (sensor#3), sensor#4 unused (= 0), unknown parameters not indicated
1000303	MCO-COR	4	direct	Battery (sensor#2), ATMS (sensor#1), AIR-TEMP (sensor#3), PARAM_UNKNOWN (sensor#4)
1000304	MCO-COR	8	direct	Drogue (sensor# 7), ATMS (sensor#1), AIR-TEMP (sensor#4), 5 PARAM_UNKNOWN
1000305	MCO-COR	> 8	direct	ATMS (sensor#1), AIR-TEMP (sensor#4), unknown parameters not indicated
1000306	MCO-COR	> 8	direct	ATMS (sensor#2), AIR-TEMP (sensor#4), unknown parameters not indicated
1000307	MCO-COR	> 8	direct	ATMS (sensor#3), AIR-TEMP (sensor#5), unknown parameters not indicated
1000308	MCO-COR	> 8	direct	ATMS (sensor#2), AIR-TEMP (sensor#1), unknown parameters not indicated

4.3.2 Identification of start and end mission dates

The start and end mission dates delimit the period from the moment the drifter was deployed up to the moment the drifter was recovered, dead or beached. This information is available for drifters Ids in the AOML database, but not for drifters Ids in the “only CLS” database. We have then implemented a tool that considers:

- whether the drifter is in the ocean and not on land (see Real Time Quality Control test 4 [RD1])
- the reliability of the Argos positions (see Real Time Quality Control test 16 [RD1])
- the velocity of the drifter (smaller than 3 m s^{-1} to be considered in mission).

This tool also identifies Argos Ids with more than one mission when two measurements are timely separated by more than two months. When more than one mission is found for one Argos Id, we verified whether the same decoder template is valid for all the missions. The tool estimates approximated start and end mission dates. All the dates were checked later during the whole process of RTQC and DMQC detailed in the C-RAID Quality Control Manual [RD1].

4.3.3 Calibration equations

In previous Phases 1 and 2 of the C-RAID project, the calibration equations for decoding drifter data were indicated in metadata. When the decoded data seemed to be biased in relation to the ERA5 data, the calibration equation were re-calculated by fitting the raw data with the AOML decoded data. The challenge in this contract was that we had either the calibration information in the metadata, nor the AOML decoded data to compare with. Our first approximation was to use the “SST standard calibration equations” used in previous Phases (equation 1). When the decoded data were systematically biased in relation with ERA5 data, we re-decoded the data using calibration equations used in previous Phases. The SST calibration equations used for decoding SST data in this contract are:

- SST = 0.05*count – 5 equation (1)
- SST = 0.04*count – 2 equation (2)
- SST = 0.17*count – 2 equation (3)
- SST = 0.16*count – 5 equation (4)

In all the cases, the used SST calibration equation is indicated in the C-RAID NetCDF file.
When a ATMS calibration equation is needed, we used the ATMS standard equation used in previous C-RAID phases (eq. 5). However, for a group of Argos Id with decoder Id 1000011, the decoded ATMS with eq. 5 showed the same time variability as the ERA5 co-localised ATMS, but different magnitude. In this case, we estimated new calibration equations by linear regression of ATMS_COUNT with the ERA5 co-localized ATMS. The ERA5 data are not exactly the reality, and the spikes of the data measured by the sensor can result in different calibration equations for each Argos Id. In order to be homogeneous, we considered a mean value of the slopes and offsets obtained by linear regression for each of the Argos Id presenting the problem (~20 Argos Id). The specific ATMS calibration equation for this group of Argos Id is shown in eq. 6

- $\mathrm{ATMS} = 0.1 * COUNT + 850$ equation (5)
- $\mathrm{ATMS} = 0.02 * COUNT + 900$ equation (6)

Difficulties found

- Manipulation of SST_COUNT

We found some Argos Ids with continuous drifter trajectory in the ocean indicating that the drifter was not recovered. However, the time series of SST_COUNT presented an important discontinuity before and after 16th January 2001. This problem has been observed on 8 different Argos Ids (5178, 5288, 5566, 5877, 7857, 8154, 8044, 21866). A change in the order of magnitude of the SST_COUNT during the period the drifter was in the water is not understandable, it is thus the processing performed at CLS that changed at this specific date. Moreover, as the SST_COUNT data need a calibration equation different to the SST standard calibration equation from the beginning of the mission up to 16/01/2001, and the SST standard calibration equation after that date, we decided to flag bad (QC = 4) the SST_COUNT and SST values before 16/01/2001. This information is indicated in the grey list.

4.3.4 Metadata

Once the decoder template identified and the start and end mission dates defined, we could cross deeper the information transferred by CLS. We assigned then additional information to each of the mission of the Argos Id found. This information concerns: “transmission system id”, “WMO number”, “drifter maker”, “project name”, “operating institution”, and “operating country”.

Difficulties found

A same Argos Id can change part or all of the information enumerated above from one period to other.

4.3 Merging decoded data

By decoding the Argos satellite messages, we generate time series of location and geophysical parameters by drifter.
In C-RAID Phase 1, as the available Argos data set was not necessarily complete, we were aware that for some months (December 2008 for example) we did not have any Argos satellite message thus any decoded data. These missing files cause time-gaps in our time series.
Moreover, we have noticed that, in occasions, there were 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 used AOML already decoded data to fill the time-gaps in the C-RAID drifter data set.
Note that during C-RAID Phase 2 we received the comprehensive delivery of Argos messages from CLS. Hereafter, the time-gaps in our decoded time series was unusual.

4.3 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 2 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 C-RAID drifters Quality Control Manual [#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 2: List of RTQC tests automatically applied to drifting buoy 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.
TEST44: Position on land Q3	Location	The latitude and longitude from a drifter that was previously flagged QC= 3 with test16 and which is on land should be flagged QC = 4.
TEST06: Global range test	Geophysical parameters	The parameter measured by the drifter should be inside the limits defined for each parameter. The latitude and longitude from a drifter that was previously flagged QC= 3 with test16 and which is on land should be flagged QC = 4.
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.

4.4 Delayed-Mode Quality Control

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

4.4.1 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 quality control process.

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.

4.4.2 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_2021 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). 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.

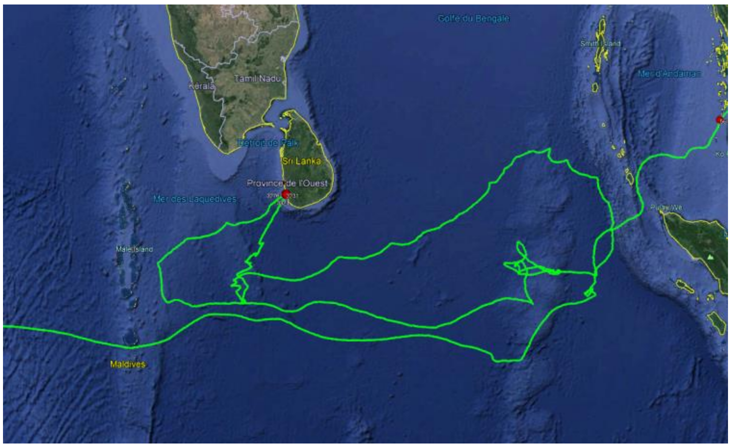

Figure 4.1: Trajectory of C-RAID drifter 1270087; the drifter was recovered, brought to port and deployed again later

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.

4.4.3 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:

(i) The identification of biased or wrong data not detected by the RT tests
(ii) 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 0), 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 0) to validate whether her/his suspicion was true.

4.4.4 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 missions 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 identifiers, should be set for that drifter.

4.4.5 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 4.4.3) and the drifter trajectories (section 0) the operator can decide to change the deployment date or end mission date initially given in the metadata. These modifications are directly indicated in the drifter information used for decoding the data.

As an example, we show here the changes of dates effectuated during the C-RAID Phase 1. Considered the original set of 5 136 drifters studied in delayed mode in the Phase 1 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 3.

Table 3: Information about the number of drifters studied in DM during the C-RAID Phase 1, for which the metadata have been partially modified

	TOTAL C-RAID 1
ORIGINAL NUMBER OF DRIFTERS PROCESSED IN DMQC	5 136
NUMBER OF ORIGINAL **DRIFTERS WITH TWO MISSIONS**	18
FINAL NUMBER OF DRIFTERS PROCESSED IN DMQC	5 154
N° DRIFTER DEPLOYMENT DATE CHANGED	9
N° DRIFTER END MISSION DATE ANTICIPATED	75
N° DRIFTER END MISSION DATE DELAYED	35

By delaying the end mission date of the 35 drifters over the total of 5 136 drifters, we recuperate 18.8 years of data.
Differently, when the end mission date has been anticipated, it affects a very small time period, from some hours to some days. Therefore, the final number of years of data of drifters whose mission date have been anticipated is not significantly different.
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.

4.4 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. Recalculated Argos fixes

5.1 Introduction

CLS is responsible for Argos fixes estimation. The fixes, originally estimated from Least Square algorithm are now processed by improved methods based on Kalman filtering. CLS processed anew all the drifter fixes collected since October 2007 with this improved method and provided us with resulting CSV data. The new algorithm significantly improves the positioning accuracy. Improvements are most significant for locations obtained in difficult conditions (class A and B locations). See #RD3 for more details.
The improved Argos positions are not directly included in the Argos message distributed by CLS. In the framework of C-RAID project, the improved fixes were recovered, formatted and introduced in the C-RAID database.
Some examples of the drifter trajectory obtained with the Least Square method and with the Kalman method are shown in figures 5.1 – 5.3. The trajectory of the fixes obtained with the Least Square method (in blue) is more irregular than the trajectory of the fixes obtained with the Kalman method (in green).

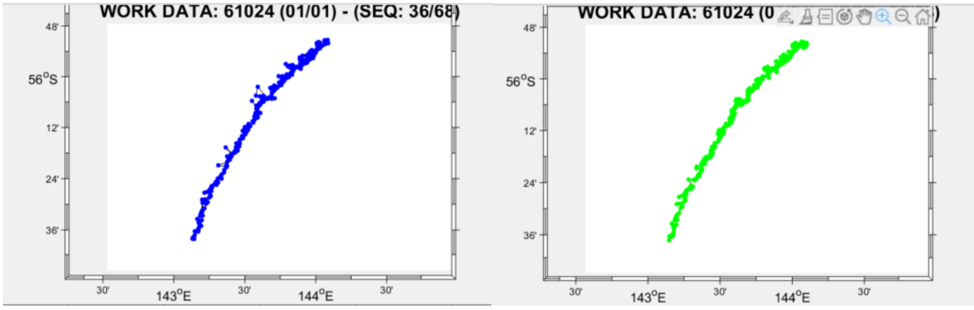

Figure 5.1: Part #36 (over 68) of the Argos Id 61024 trajectory; on the left in blue, fixes obtained with the Least Square method and, on the right in green, fixes obtained with the Kalman method.

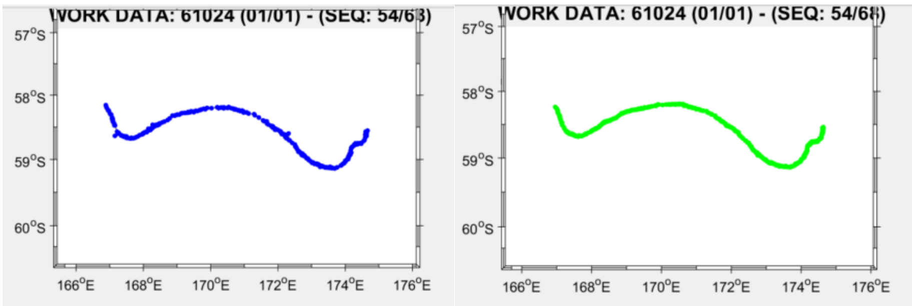

Figure 5.2: Part #54 (over 68) of the Argos Id 61024 trajectory; on the left in blue, fixes obtained with the Least Square method and, on the right in green, fixes obtained with the Kalman method.

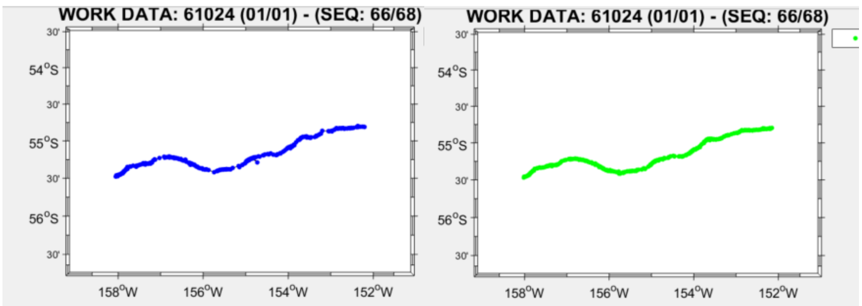

Figure 5.3: Part #66 (over 68) of the Argos Id 61024 trajectory; on the left in blue, fixes obtained with the Least Square method and, on the right in green, fixes obtained with the Kalman method.

In the following, when fixes obtained with the Kalman method are available, they replace the Least Square ones in the drifter trajectory of C-RAID database.
Additionally, when available, the .csv files with the fixes of the Argos Id trajectories obtained with i) the Least Square method, ii) the Kalman method and iii) the smoothed version of the Kalman method are provided as complementary data with the decoded data of each Argos Id.

5.2 Improvement of the incorporation of Argos fixed recalculated by Kalman method

The incorporation of the recalculated Argos fixes was done during de MCO – CORIOLIS Phase. At that point, we had dealt with the data and their Quality Control of 10 039 drifters deployed from 1979 to 2010. The recalculated fixes by the Kalman method are available from October 2007 hereafter, consequently, the data of only 2 778 drifters from the 10 039 drifters delivered previously were affected by the re-calculation of fixes. In this section we show the results of the study carried out to identify the differences caused by the incorporation of Argos fixes recalculated by the Kalman method.

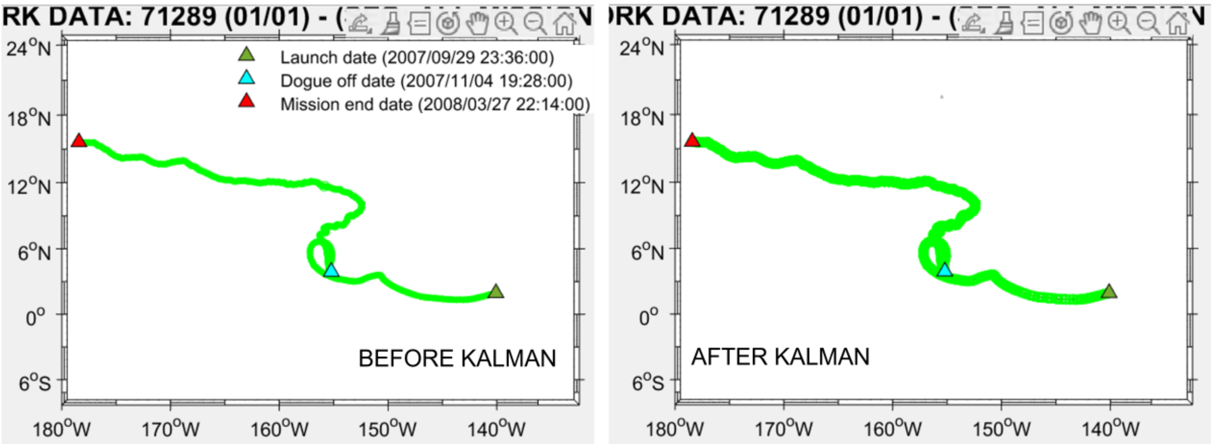

Figure 5.4: Trajectory of drifter id 71289 before (on the left) and after (on the right) incorporating the recalculated Argos fixes.

The first interesting point when comparing data before and after Kalman is that the drifter trajectories are enriched in positions. Figure 5.4 shows the trajectory of drifter Id 71289 before and after incorporating Kalman fixes. For this drifter there is a small gain in Kalman fixes: the number of positions before/after Kalman are 3 528/3 730, a gain of 202 positions (a gain of 5% of positions of the trajectory).

Another example is shown in figure 5.5, it displays the trajectory of drifter Id 71451 before and after incorporating Kalman fixes. There are 83/2 815 positions before/after incorporating Kalman fixes, a gain of 2 732 (a gain of 97% of positions of the trajectory). Moreover, the Argos Class of the gained positions after incorporating Kalman are mainly Argos Class A, B, Z and 0 (see figure 5.6 that shows drifter Id 71451 trajectory after Kalman: all the Argos Classes on the right and only Argos Classes = 1, 2 and 3 on the left).

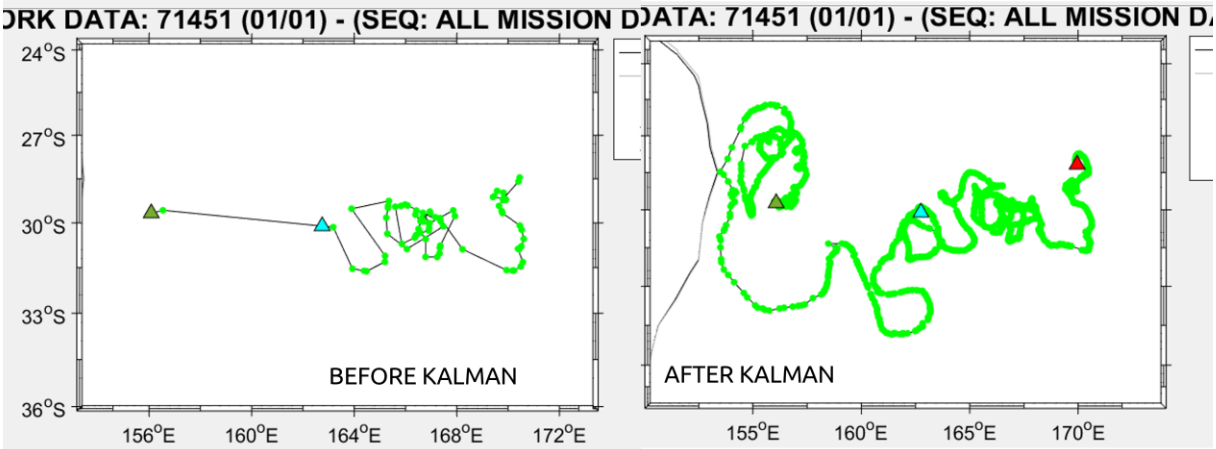

Figure 5.5: Trajectory of drifter id 71451 before (on the left) and after (on the right) incorporating the recalculated Argos fixes.

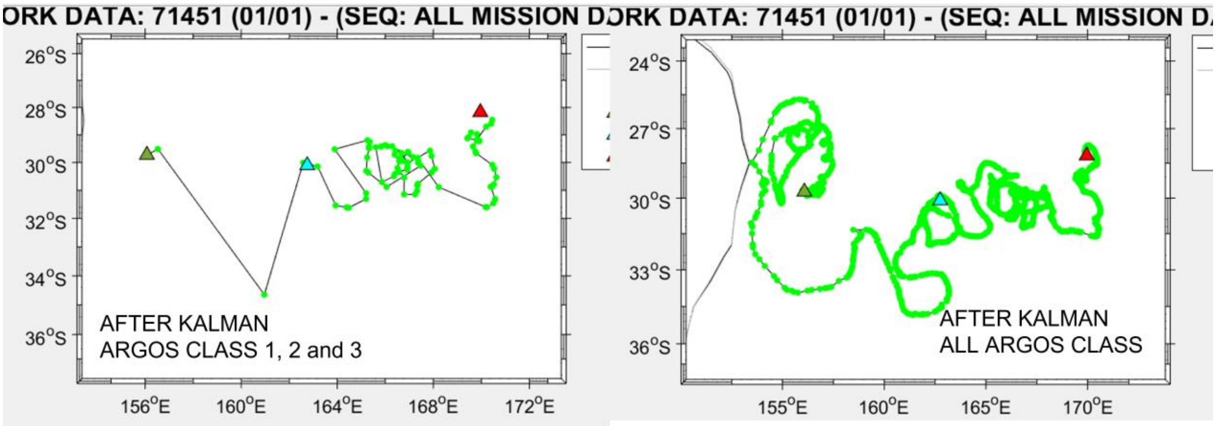

Figure 5.6: Trajectory of drifter id 71451 after incorporating the recalculated Argos fixes: Argos Class 1, 2, 3 positions on the left, and all the positions independently of the Argos Class on the right.

We have then studied the gain of positions after incorporation of Kalman fixes considering the ensemble of 2 778 drifters which positions were recalculated by the Kalman method. In the following we will refer to delivery 1 to the positions before Kalman fixes incorporation and delivery 2 to the positions after Kalman fixes incorporation. Table 4 summarizes de number of fixes in both deliveries.

Table 4. Number of positions in the 2 778 drifters for which positions were recalculated by the Kalman method: Delivery 1 refers to the delivery before incorporating Kalman fixes and Delivery 2 refers to the delivery after incorporating Kalman fixes

	Number of fixes (all Classes)	number of fixes Class 0, A, B, Z	number of fixes Class 1, 2, 3	% of fixes Class 0, A, B, Z	% of fixes Class 1, 2, 3
Delivery 1	39 274 229	910575	38363654	2.32	97.68
Delivery 2	46 876 227	6862790	40013437	14.64	85.36

Delivery 1 contains 39 274 229 positions: 2.32 % of these positions are classes 0, A, B, Z and

97.68 % are classes 1,2,3.
Delivery 2 contains 46 876 227 positions: 14.64 % of these positions are classes 0, A, B, Z and 85.36 % are classes 1,2,3.
Delivery 2 has 7 601 998 additional positions than delivery 1: 5 952 215 of classes 0, A, B, Z and 1 649 783 of classes 1,2,3.
Figure 5.7 shows histograms of % of gain of positions from delivery 1 to delivery 2. Considering positions of all the Argos classes, most of the drifters have increased the number of positions in 15-25%. When only classes 0, A, B, Z positions are considered, most of the drifters increased the number of positions in 96-99%. Finally, when only classes 1, 2, 3 positions are considered, most of the drifters increased the number of positions in 5 - 8%

Figure 5.7: Histograms of % of gain of positions: all the Argos classes on the left, positions Argos classes 0, A, B, Z in the center, and positions Argos Class 1, 2, 3 on the right.

The main gain of positions in delivery 2 is because classes 0, A, B, Z positions, which were not all available for all the drifter in delivery 1. Furthermore, more marginally, class 1,2,3 positions were also added in Delivery 2. Delivery 1 was therefore probably incomplete in this respect.

The quality control process was carried out on the data of both deliveries, we have then studied the quality given to the positions in Delivery 1 and 2. Globally, the positions classes 0, A, B, Z in delivery 1 and 2 have the same percentage of “good data” (QC = 1, 2) and “bad data” (QC = 3,4). The quality of the positions does not depend on the method of estimation of positions.
For example, for drifter Id 46067, for which the number of classes 0, A, B, Z positions in delivery 1 and 2 are the same order of magnitude (3 265 and 3 312, respectively). The % of data QC = 1,2 are similar in the two deliveries (95,4% and 95,9%, respectively). Most of the drifters show this behavior.

If we look for a relative significant change in the quality of the positions in the delivery 1 and 2, we found drifter Id 70917, which has approximately the same number of classes 0, A, B, Z positions in delivery 1 and 2 (10 379 and 10 362, respectively). The % of data QC = 1,2 slightly increased from delivery 1 to delivery 2: 93% to 99%, respectively. In this case we could say that the Kalman filtering improved the quality of the positions making exploitable (QC = 1,2) data previously unexploitable (QC = 3,4). Nevertheless, this is not the general rule found between our drifters.

5.3 Conclusions

The incorporation of fixes recalculated by the method based on Kalman filtering has brought to the light:

1. The increase of positions in delivery 2 (after Kalman) in relation to delivery 1 (before Kalman) are mainly classes 0, A, B, Z positions. There is also an increase in the number of class 1, 2, 3 positions. Consequently, this is a problem of an incomplete delivery 1 rather than an improvement related to the position estimation algorithm.
2. The classes 0, A, B, Z positions do not change in quality (QC = 3, 4 vs. QC = 1,2) depending on the position estimation algorithm used.

6. Autumn Delivery (November 2022)

The C-RAID Autumn delivery **replaces any previous C-RAID database**. It contains the data of the 13 964 drifters (around 13 657 years of data) deployed between 1979 and 2016, with the RTQC and DMQC process completed. Note that most of these drifters are also in the AOML database, however, about 2 000 drifters were directly provided by CLS and are exclusively in the C-RAID database.

Together with drifter data, we provided **complementary data** for these drifters:
The Argos satellite messages in the PVR/DS format;
The ERA5 data co-localized to the drifter positions (sea surface temperature, mean sea level, wind at 10 m, wind stress and temperature 2 meter over the sea level (when this variable is decoded));
When available, Argos fixes estimated by CLS by (i) the Least Square method, (ii) the Kalman method and (iii) the smoothed version of the Kalman positions.

They **grey list** resulting from the DMQC process, and used in the final decoding data, is also provided.

The final Excel file (7_drifter_information.xlsx) **specifically used for decoding the drifters** Ids of the delivery.

The **metadata** are also provided in this delivery:
The Excel files of metadata obtained from AOML (1_finalize_aoml_meta_ALL_1_20191017T141257.xlsx; 2_finalize_aoml_meta_ALL_2_20191017T141257.xlsx; 3_finalize_aoml_meta_ALL_3_20191017T141257.xlsx)
The additional metadata provided by CLS of drifters in the AOML
(4_Copie de 97-10_CLS_Meta_Request.xlsx)
The Excel file of metadata provided by CLS (5_Donnees_declaratives_from_CLS.xlsx);

The final Excel file of metadata for drifters "only CLS" where CLS metadata have been gathered, edited and cleaned (6_Metadata_organized_in_Capgemini.xlsx);

The decoded data measured by Pelagic – Larval - Fishery devices, a total of 91 Argos Ids (30 years of data), are also delivered in RTQC and with complementary data except ERA5 data in the file /DATA/data_no_drifter/. They are named by their Argos Ids and not with C-RAID Ids as it is the case for the drifter data.