Drifter Quality control

A focus on C-RAID ocean drifters reprocessing: Improve the access to historical drifter data

In situ

Ifremer, Copernicus Marine & Copernicus in situ

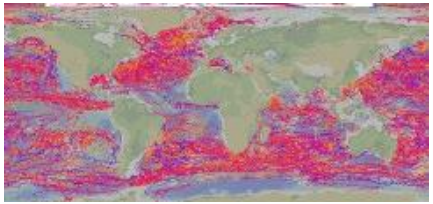
Copernicus
Europe's eyes on Earth

European Commission

In situ

C-RAID DRIFTING BUOYS PROJECT

- C-RAID project is a global rescue and reprocessing of drifting buoys data. The C-RAID dataset contains the metadata of 20.000 drifting buoys, deployed between 1979 and 2018.
- During “C-RAID phase 1”, the data of 10.035 drifting buoys deployed between 1997 and 2010 have been delayed mode processed (including comparison with Copernicus ERA5 reanalysis).
- The project is continuing in 2022-2023 as “C-RAID phase 2” to reprocess drifting buoys data deployed before 1997 and after 2010.



Copernicus Europe's eyes on Earth

European Commission

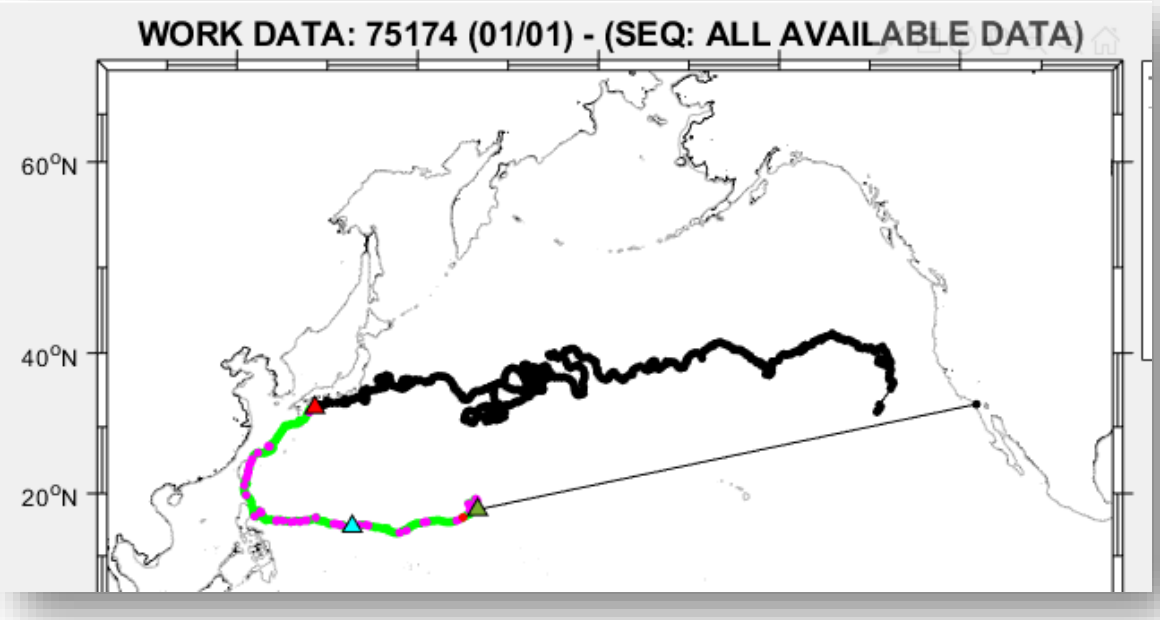
C-RAID deliverables

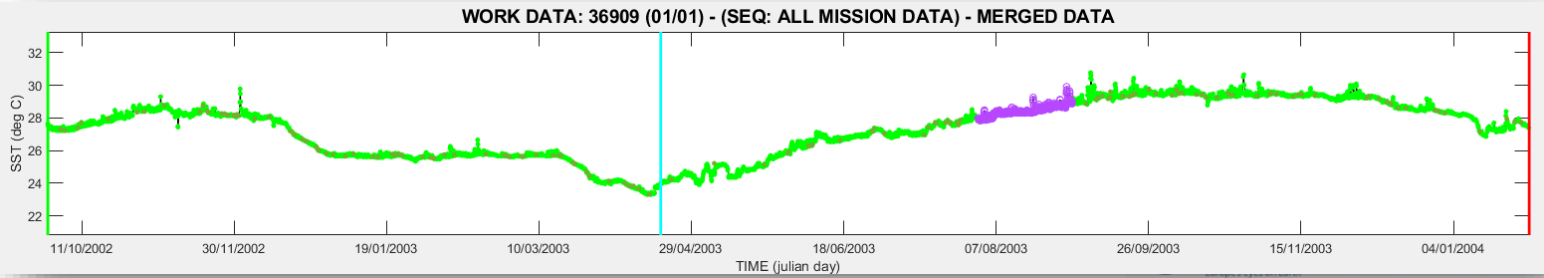
- C-RAID deliverables
  - 1) An improved drifting buoys data archive
  - 2) FAIR interfaces to drifting buoys data :
    - Web data discovery for human users
    - API for data discovery/subsetting/download services (machine-to-machine data access)

  - “Improved drifting buoy data record”
    - Missing datasets and parts of datasets recovered -> data rescue activity
    - Homogeneous and rich metadata and data NetCDF format
    - Homogeneous QC and assessment on marine and atmospheric data
    - Matchup with Copernicus ERA5 reanalysis (temperatures, atmospheric pressure, wind)

Two examples of improved data record

In situ





C-RAID data distribution

- C-RAID data, metadata and documentation are published on https://doi.org/10.17882/77184
  - Complies with FAIR data rules
    - Findable — DOI published on DataCite, google indexed (Schema.org)
      Link with bibliography and authors bibliography (ORCID)
    - Accessible — One click download, anonymous access
      Links to big data services
    - Interoperable — CF and SeaDataNet standards, QC documented, rich metadata
    - Re-usable — CC-BY license

- C-RAID metadata are transmitted to OceanOPS team (IODE-WMO)

C-RAID FAIR data Human and machine readable DOI

C-RAID buoys DOI https://doi.org/10.17882/77184

One click download of data (https)

Authors ORCID

Data licence

Documentation (NetCDF format, QC manual)

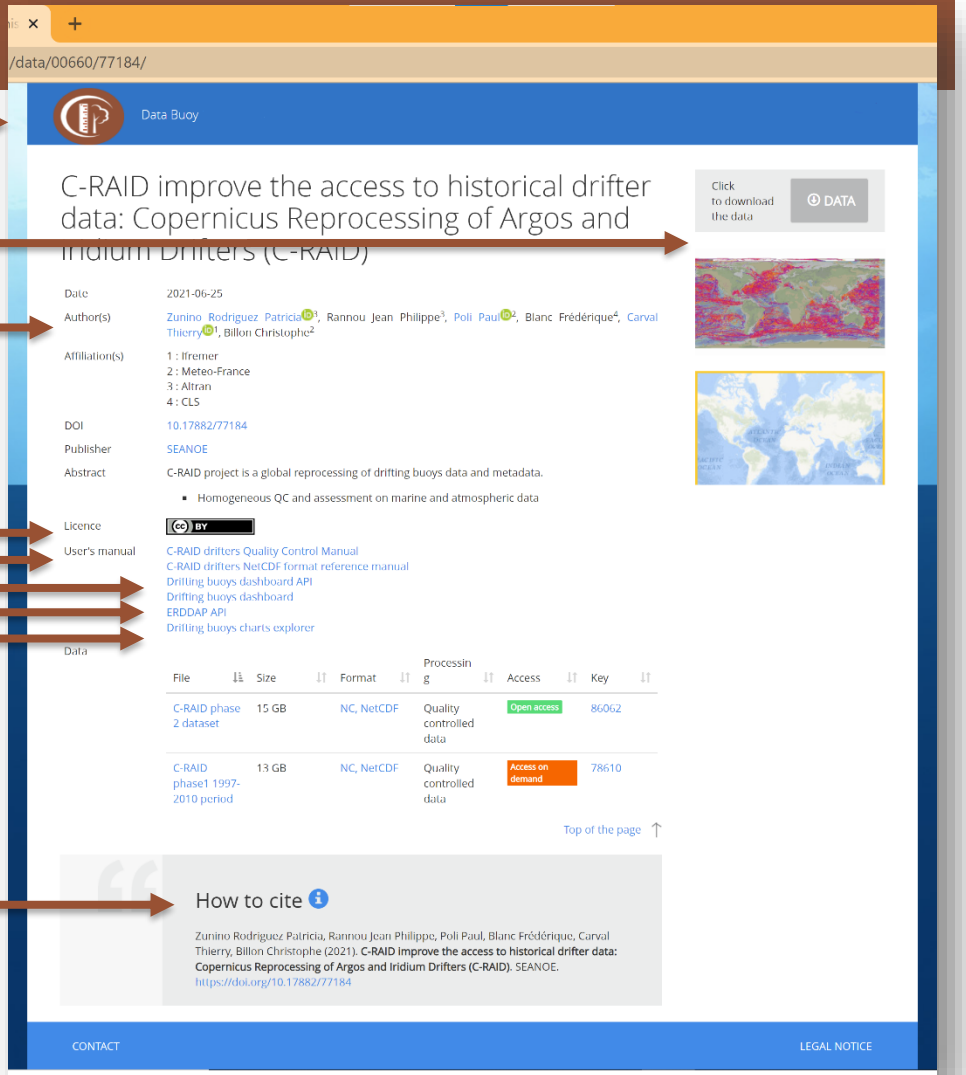
Dashboard and its API

ERDDAP server

Data charts explorer

How to cite statement

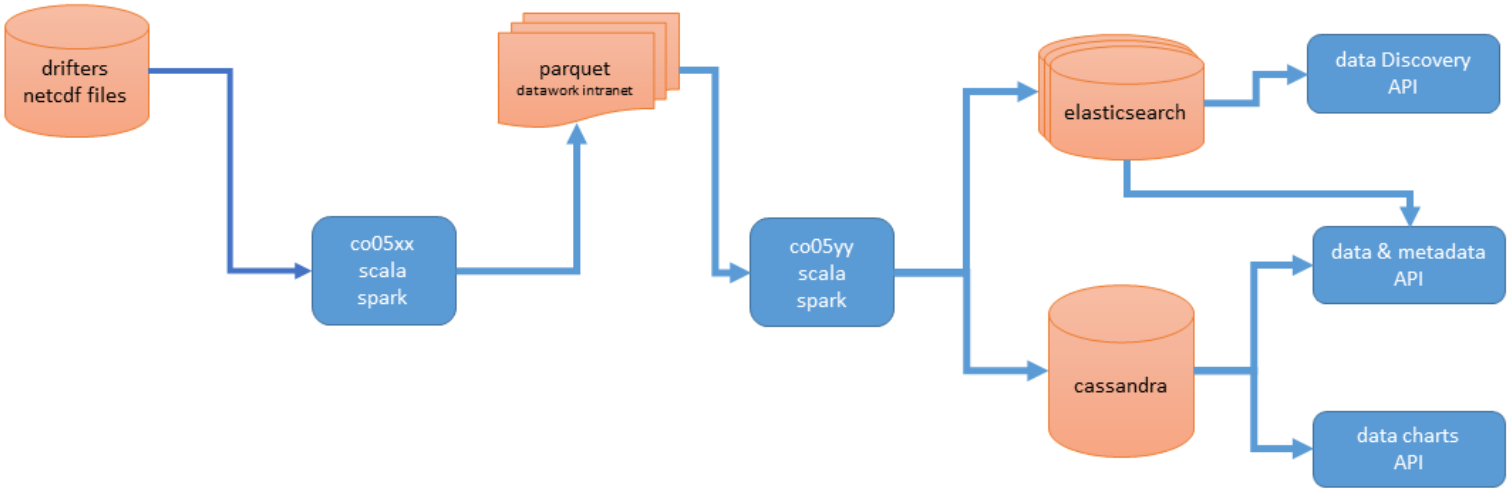
And more…



C-RAID Drifters bigdata infrastructure

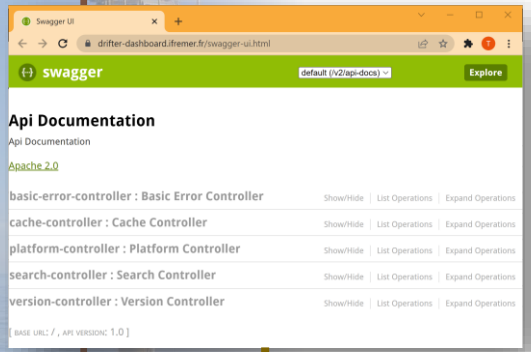
- Data discovery and subsetting API on Ifremer big data infrastructure
  NetCDF - Parquet - Cassandra NoSql – Elastisearch - API

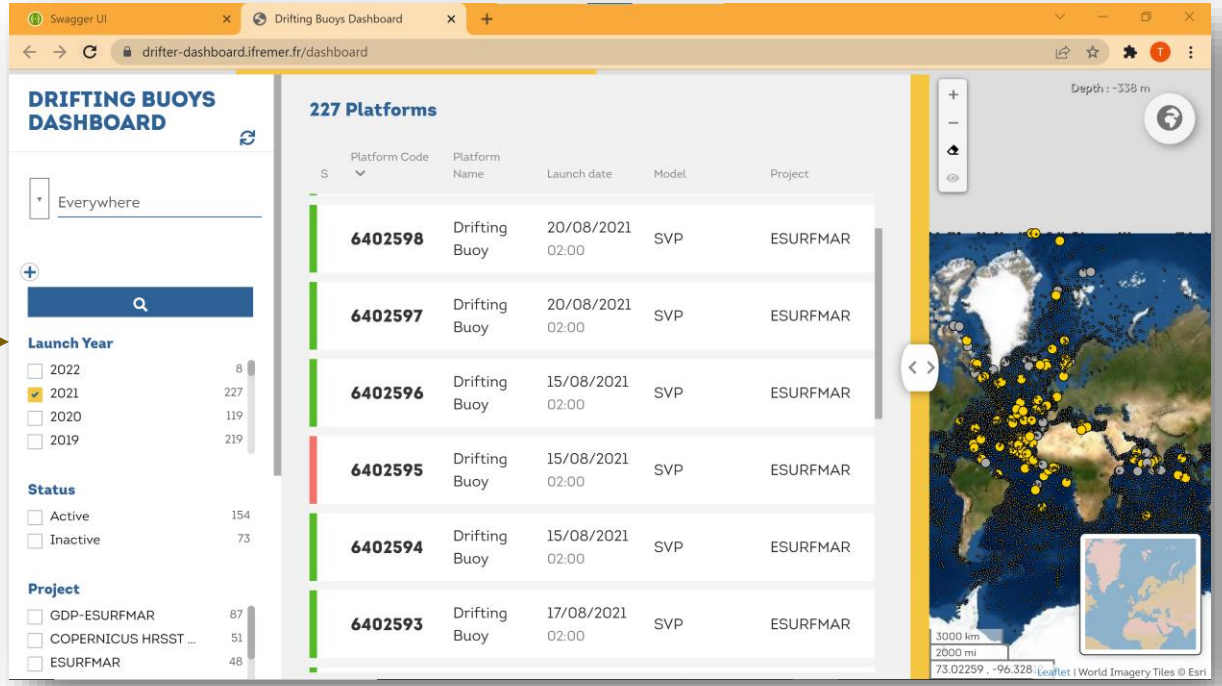
Drifting buoys bigdata architecture



C-RAID Drifters big data infrastructure

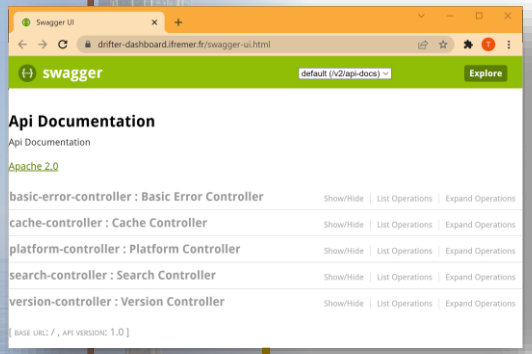
- The drifting buoys API and its dashboard are online
  https://drifter-dashboard.ifremer.fr
  https://drifter-dashboard.ifremer.fr/swagger-ui.html

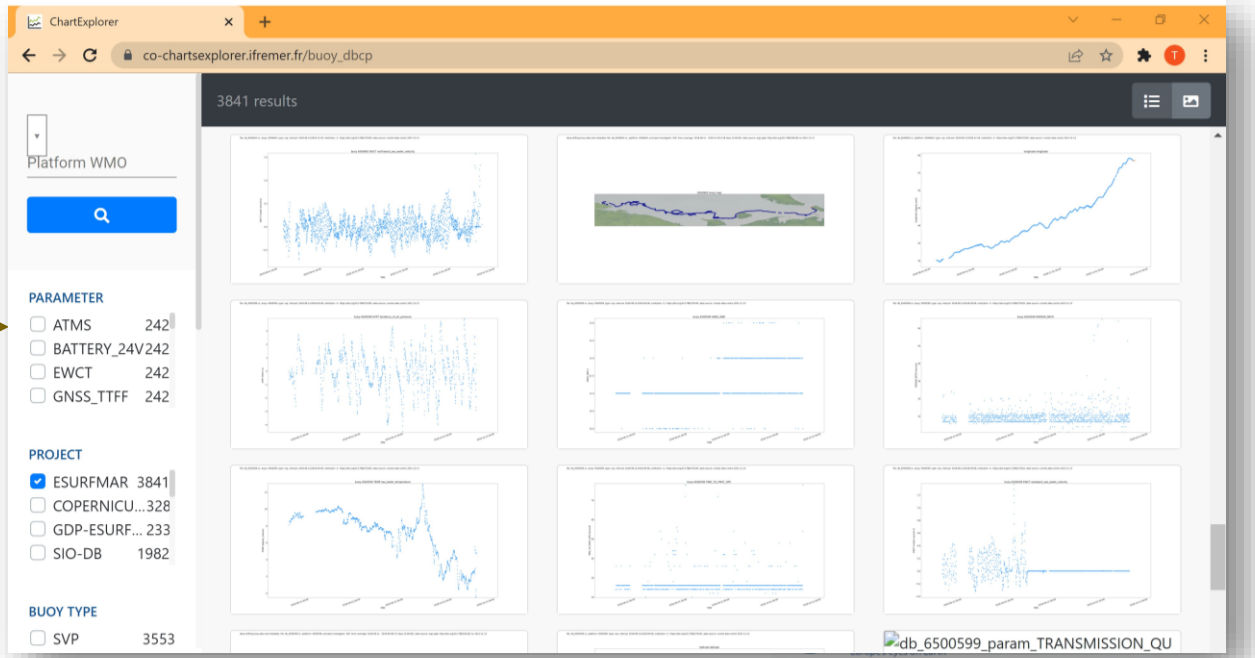




C-RAID Drifters big data infrastructure

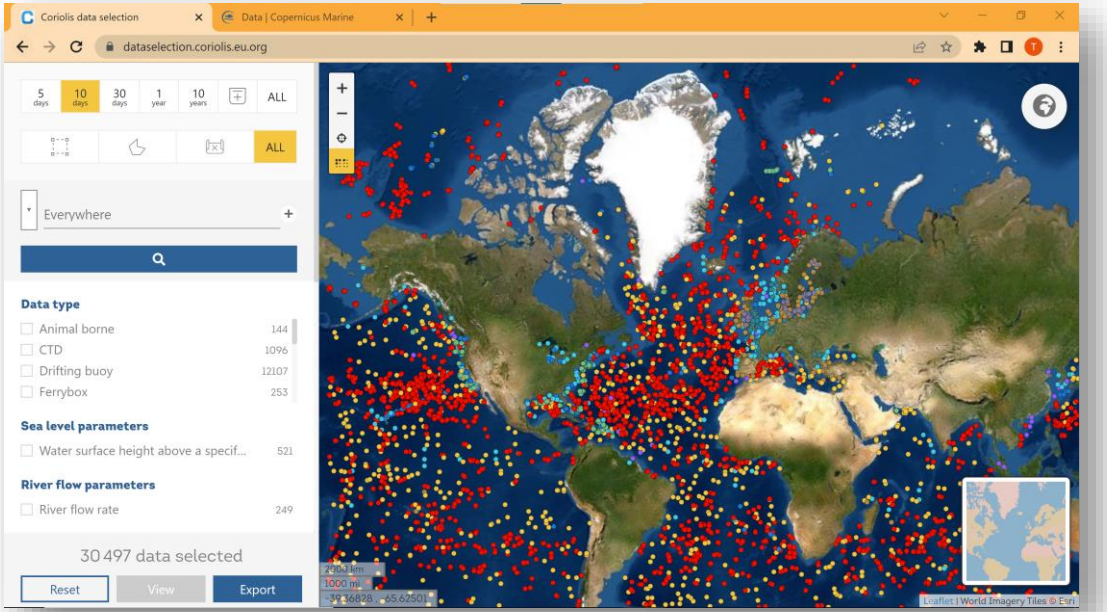
- The drifting buoys data charts explorer
  https://co-chartsexplorer.ifremer.fr/buoy_dbcp





C-RAID activity

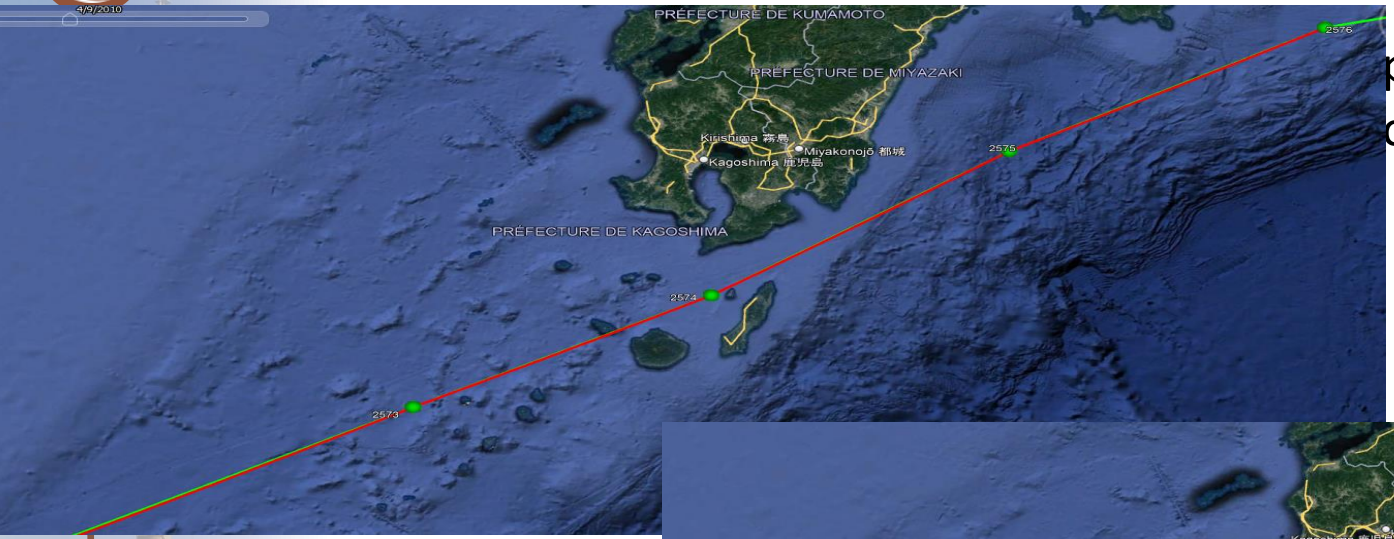
- The C-RAID dataset delivery is ingested in:
  - OceanOPS metadata dashboard https://www.ocean-ops.org/board
  - Copernicus Marine in situ TAC https://doi.org/10.48670/moi-00036
  - CIS2: Copernicus In Situ Information System https://cis2.eea.europa.eu



Succes story: Kalman improved trajectories

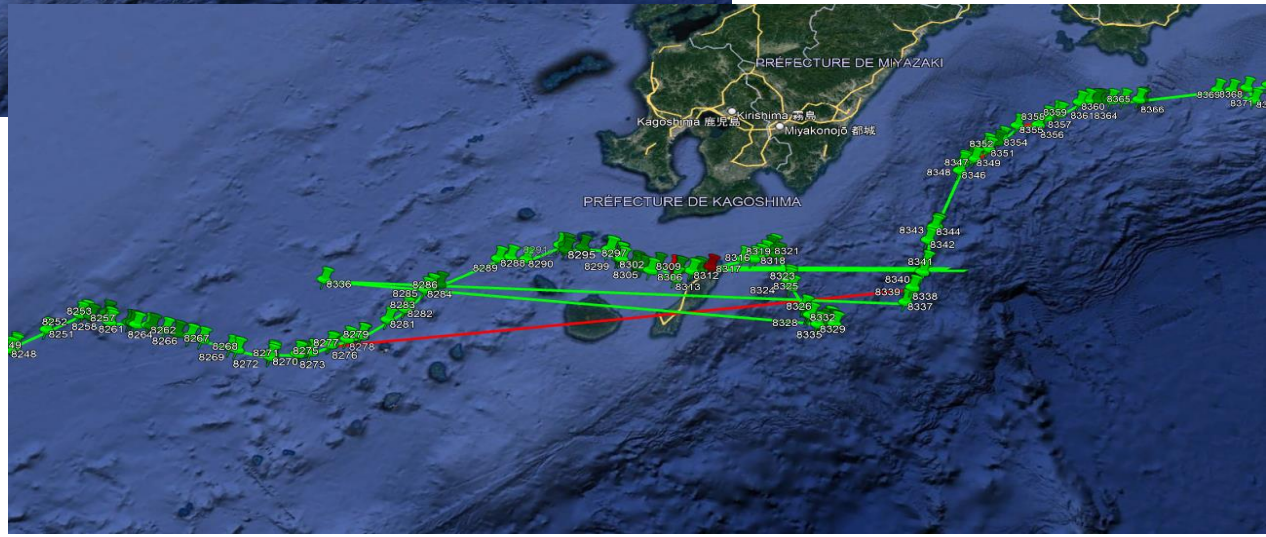
- CLS reprocessed the trajectories of their recent database
- We received in 2020 the Kalman reprocessed trajectories of 3000 drifters
- On a majority of drifters, it provides a huge increase in the quality of trajectories, on particular on poor satellite coverage positions

WP8 Highlights



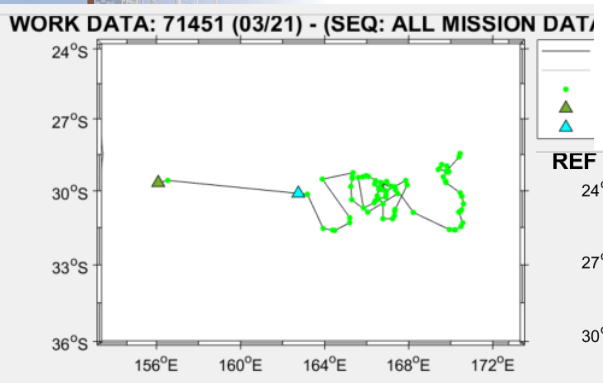
provided a **huge**

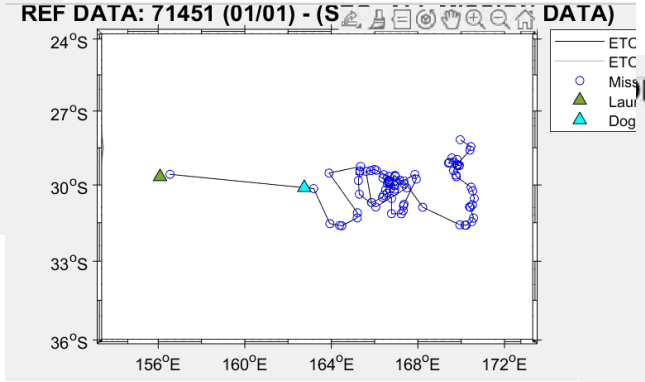
or poor satellite

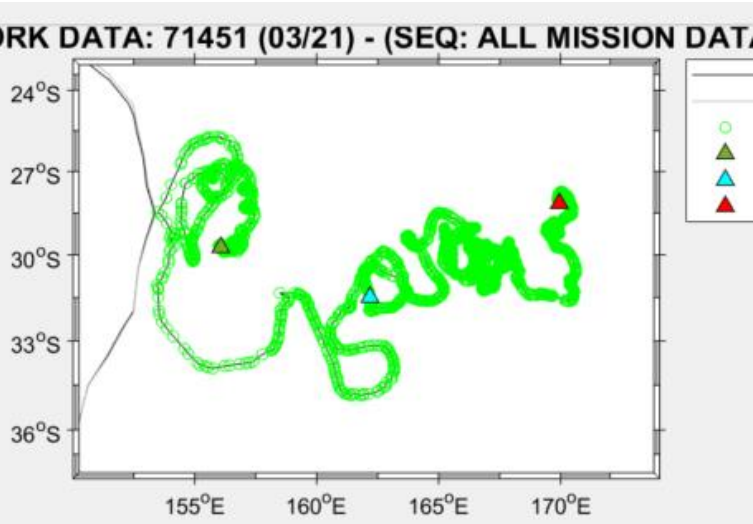


WP8 Highlights

- A success story : the Kalman filter Argos positioning provided a **huge increase in the quality of trajectories**, in particular for poor satellite coverage positions







In situ

Drifters quality control procedure

The drifters quality control procedure are documented on:

- C-RAID drifters Quality Control Manual
  https://doi.org/10.13155/81639
  - The QC manual is regularly improved, last release v6: December 2022)
  - Major chapters
    - RTQC : real-time quality controls : 15 automated tests
    - DMQC : delayed mode quality controls : 4 procedures performed by a scientific expert
  - Quality Control (QC) flags
    - Each date, position, parameters has a QC flag
      - 0 : no QC performed
      - 1 : good data
      - 4 : bad data

Reference document

- Argo Quality Control Manual for CTD and Trajectory Data https://dx.doi.org/10.13155/33951
- Guide to Drifting Data Buoys, UNESCO, 1988 https://www.oceandocs.org/handle/1834/2734

Real-time QC tests

In situ

The data processing chain decodes Argos or Iridium messages and applies a list of 15 automated QC tests

- Impossible date test (TEST02)
- Impossible location test (TEST03)
- Position on land test (TEST04)
- *Impossible speed test (to be implemented)*
- Global range test (TEST06)
- Regional range test (TEST07)
- Time-continuity test (TEST08)
- Spike test (TEST09)
- Digit rollover test (TEST10)
- Stuck value test (TEST11)
- Grey list test (TEST12)
- Argos Redundancy test (TEST13)
- Inside of mission test (TEST14)
- Questionable Argos position test (TEST16)
- Spike two points test (TEST19)

Real-time QC tests

The RTQC tests are applied in this well defined order

The last test (grey list) is the last applied, it contains human decisions that revisit some of the previous tests

Order	Test number	Test name
1	1	Platform identification test
2	14	Inside of mission test
3	13	Argos redundancy test
4	2	Impossible date test
5	3	Impossible Location test
6	16	Questionable Argos position test
7	44	Position on land QC = 3 flagged by TEST16
8	6	Global range test
9	7	Regional range test
10	9	Spike test
11	19	Spike two points test
	8	~~Time-continuity test~~ (UNUSED)
12	10	Digit rollover test
13	11	Stuck value test
	4	~~on land test~~
14	12	Grey list test

In situ

Delayed mode QC tests

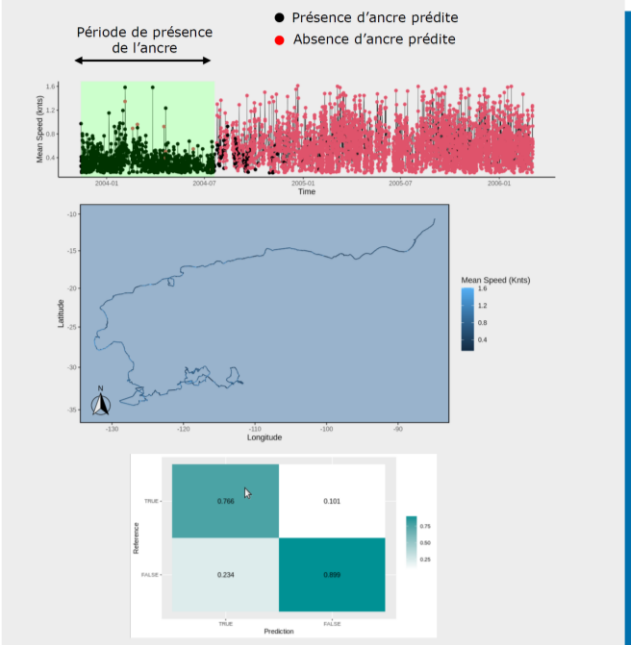
The delayed-mode quality control is performed by an expert who revisits the real-time QCs and take decisions on a visual inspection of data and metadata.

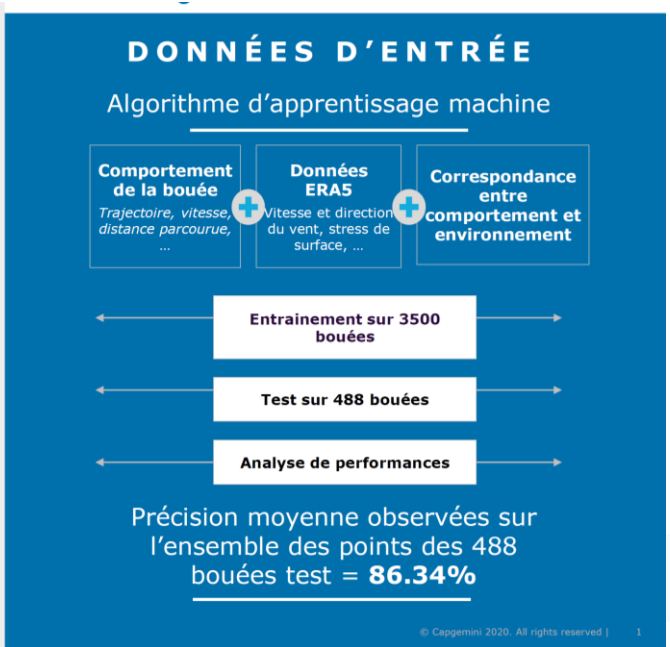
- The very fine resolution bathymetry used (GEBCO 2020), in occasion, is not enough to provide the exact depth at the position of the drifter.
  The natural processes in both the ocean and the atmosphere can result in values of geophysical parameters that are real but flagged automatically by the RTQC tests.
- During the DMQC process, originally set QCs can be modified **from good** data **to bad** data for erroneous measurements that have been “missed” by RTQC tests but also **from bad data to good data** for erroneously flagged measurements.
- The flags given during the DMQC steps revoke the flags given by specific RTQC tests. Once the different steps of DMQC finished, the RTQC tests are ran again with the corresponding revoked RTQC test inactive, or with the new test order indicated in the DMQC test.
  All the decision taken in the DMQC tests are indicated in the grey list, which is the last test to run during the RTQC procedure. Consequently, the operator decisions can be reproduced in a further run of the RTQC procedure.
- **For QC traceability, we indicate in the C-RAID drifter grey list each modification of QC values performed during this phase.**

C-RAID Machine Learning activity- xgboost

Machine Learning activity on C-RAID data to detect the drifting buoys loss of floating anchor

- Question: is a buoy drifting with the nominal 15 meter deep current ?
- Solution
  - train a model with 3500 buoys (with anchor loss detected by an expert) and ERA5 (wind, temperature,…)
  - Run the model on 500 buoys





European Commission

In situ

Thank you for your attention, any question ?

Copernicus Europe's eyes on Earth

European Commission

Implemented by

European Environment Agency