Step by step guide to the DMQC procedure using reference data, scripts and data needed

Birgit Klein, Simo-Matti Siiriä, Małgorzata Merchel

Marginal Seas Argo DMQC workshop, Sopot, Poland 18.04.2023-19.04.2023

Step by step guide based on the floats from Poland

In all scripts operator needs to select the appropriate WMO number of the float.

Download *.nc file with Argo profiles, e.g. 3902104_prof.nc from the GDAC

1. Create *.mat file from *.nc file using the **IOPANreadQCprof.m** script
2. Following steps need reference database as *.mat file e.g. IOPAN_database.mat. But you can also create your own database or use ices database from the ctd_xxxx.mat files
3. Run **IOPANplotsDMQC.m** script that searches for reference data at a distance of 30 km and 30 days
4. Run **IOPANmeanDMQC.m** script that calculates, for each Argo profile and the closest CTD profiles, the mean salinity in the selected layer (10-30 m the most stable) and difference between them.
5. Run **IOPANplotsDIFF.m** (Matlab R2018b needed) script that plots salinity difference between CTD and Argo.
6. Finally, create D files and send to Coriolis

Step by step guide based on the floats from Poland

Step 2: Create *.mat file from *.nc file using the **IOPANreadQCprof.m** script

Caveats and points for discussion:

- The script uses flags from RTQC procedures and additionally removes profiles whose average salinity in the 0-10 m layer >= 8.1.
- Problems: RTQC procedures do not quite work for the Baltic Sea data. Often correct profiles are flagged as 4, while wrong profiles are flagged as 1.
- Solution: Adaption of RTQC procedures for data from the Baltic Sea area.
- Workaround: Check each profile in the scoop program and change the QC, but there are often doubts about whether the profile is correct or not. It is also very time consuming.
- What if we are not sure if the data is good. Should we flag the data as 4 or 3?

Input is the *_prof.nc file downloaded from the GDAC
Output is a mat file with a structure array for the argo data

Step by step guide based on the floats from Poland, Step 2

Problems with wrongly assigned QC flags:

A set of prescribed real time tests is performed on all floats as described in the Argo manuals. But these test have all been devised for the open ocean and were never meant for the Baltic.

A check of the tests failed by float data were inspected during the MOCCA project EuroArgo Rise project and summarized in MOCCA deliverable D4.4.7 'Data management for floats in the Baltic' https://www.euro-argo.eu/EU-Projects/Completed-projects/MOCCA-2015-2020/Deliverables and EuroArgo Rise deliverable D2.7 'A report on the adaptation of existing DMQC methods to marginal seas' https://www.euro-argo.eu/EU-Projects/Euro-Argo-RISE-2019-2022/Deliverables

Tests concerned are: Density inversion test, Digit rollover test, Stuck value test, Gradient test and a new test proposed as ‚Incorrect near-surface salinity'

Each RTQC test has a unique Test Number (n). The binary ID (2^n) of the unique test number (n) is used to record rtqc tests performed and failed in the history variable HISTORY_QCTEST.

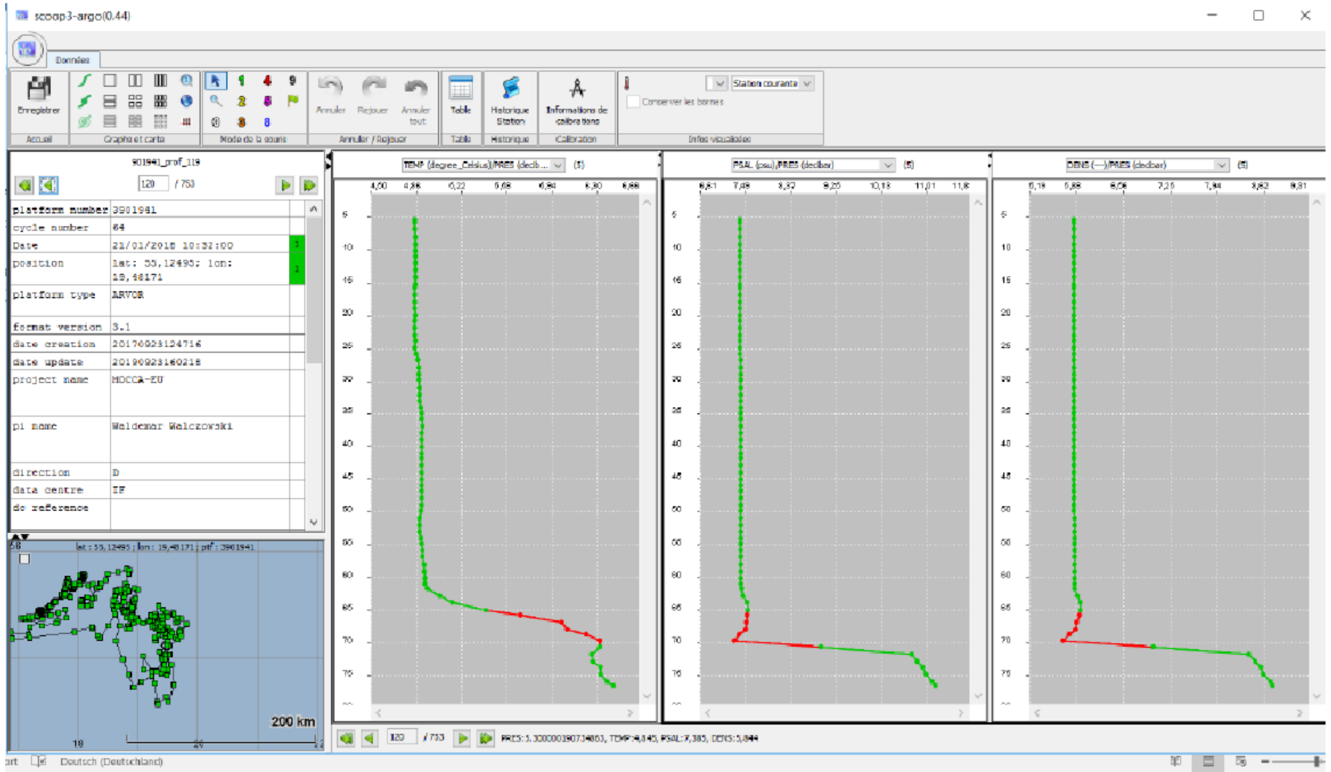
Test number (n)	Binary ID (2^n)	Test name
1	2	Platform Identification test
2	4	Impossible Date test
3	8	Impossible Location test
4	16	Position on Land test
5	32	Impossible Speed test
6	64	Global Range test
7	128	Regional Global Parameter test
8	256	Pressure Increasing test
9	512	Spike test
~~10~~	~~1024~~	~~*Top and Bottom Spike test (obsolete)*~~
~~11~~	~~2048~~	~~*Gradient test (obsolete)*~~
12	4096	Digit Rollover test
13	8192	Stuck Value test
14	16384	Density Inversion test
15	32768	Grey List test
16	65536	Gross Salinity or Temperature Sensor Drift test
17	131072	Visual QC test
18	261144	Frozen profile test
19	524288	Deepest pressure test
20	1048576	Questionable Argos position test
21	2097152	Near-surface unpumped CTD salinity test
22	4194304	Near-surface mixed air/water test
23	8388608	RTQC flag scheme for Deep SBE CTD data > 2000 dbar
24	16777216	RTQC flag scheme for RBR*Argo*[3]|2K CTD data (< 2000 dbar)
25	33554432	MEDD test

Argo Quality Control Manual for CTD and Trajectory Data
https://archimer.ifremer.fr/doc/00228/33951/

Step by step guide based on the floats from Poland, Step 2

Density inversion test:
The real-time density inversion test uses a threshold of 0.03 kg/m-3 and catches some small hooks at the base of the mixed layer of the Baltic.

Based on the experience of FMI and IO-PAN the threshold used to flag the data as bad appears reasonable and it is suggested that DMQC operators use thermal lag corrections to check if this reduces the salinity hook.

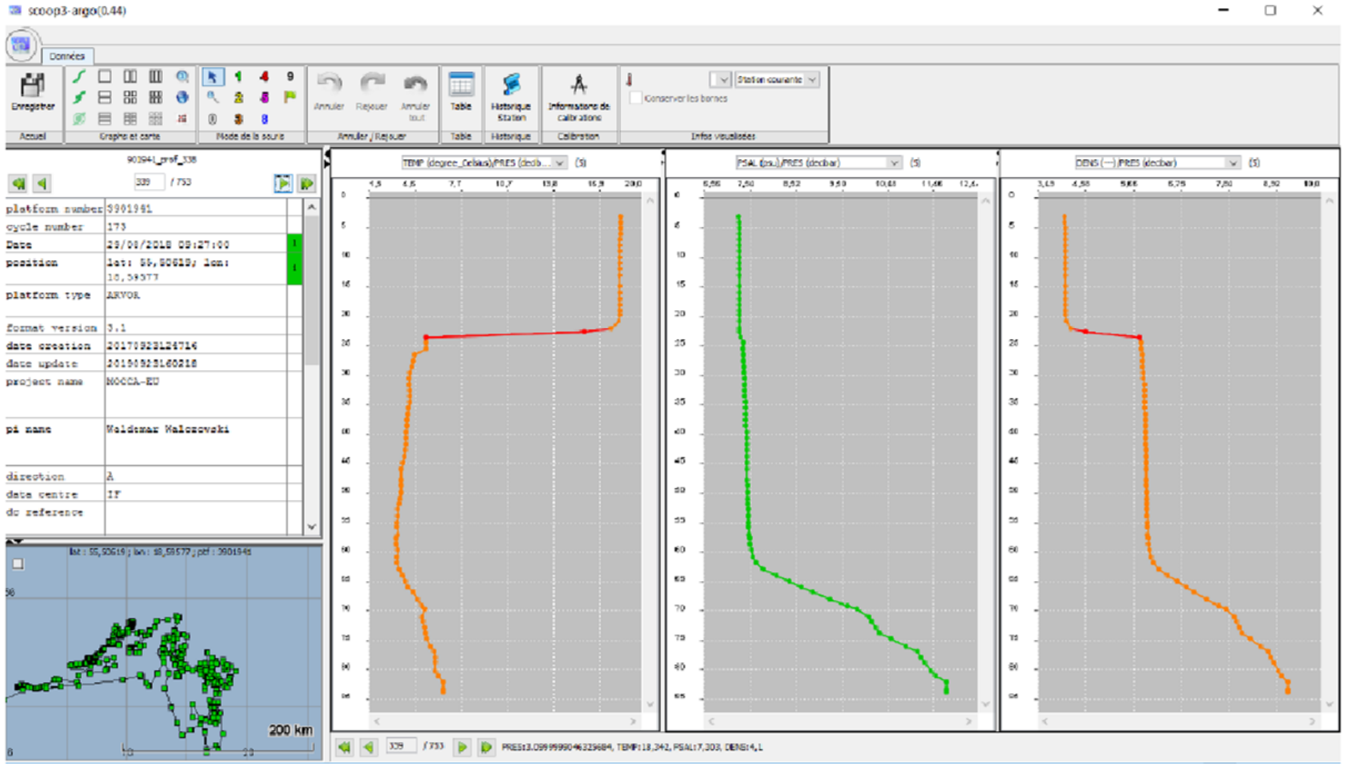


Source: D2.7 ‘A report on the adaptation of existing DMQC methods to marginal seas’ https://www.euro-argo.eu/EU-Projects/Euro-Argo-RISE-2019-2022/Deliverables

Step by step guide based on the floats from Poland, step 2

Digit rollover test: The digit rollover test is a remnant from early Argo days when only a limited amount of bits were available for transmission in the Argos satellite system. The range of encountered temperature and salinity data however was not always large enough to accommodate them and when the range was exceeded stored values rolled over to the lower end of the range.

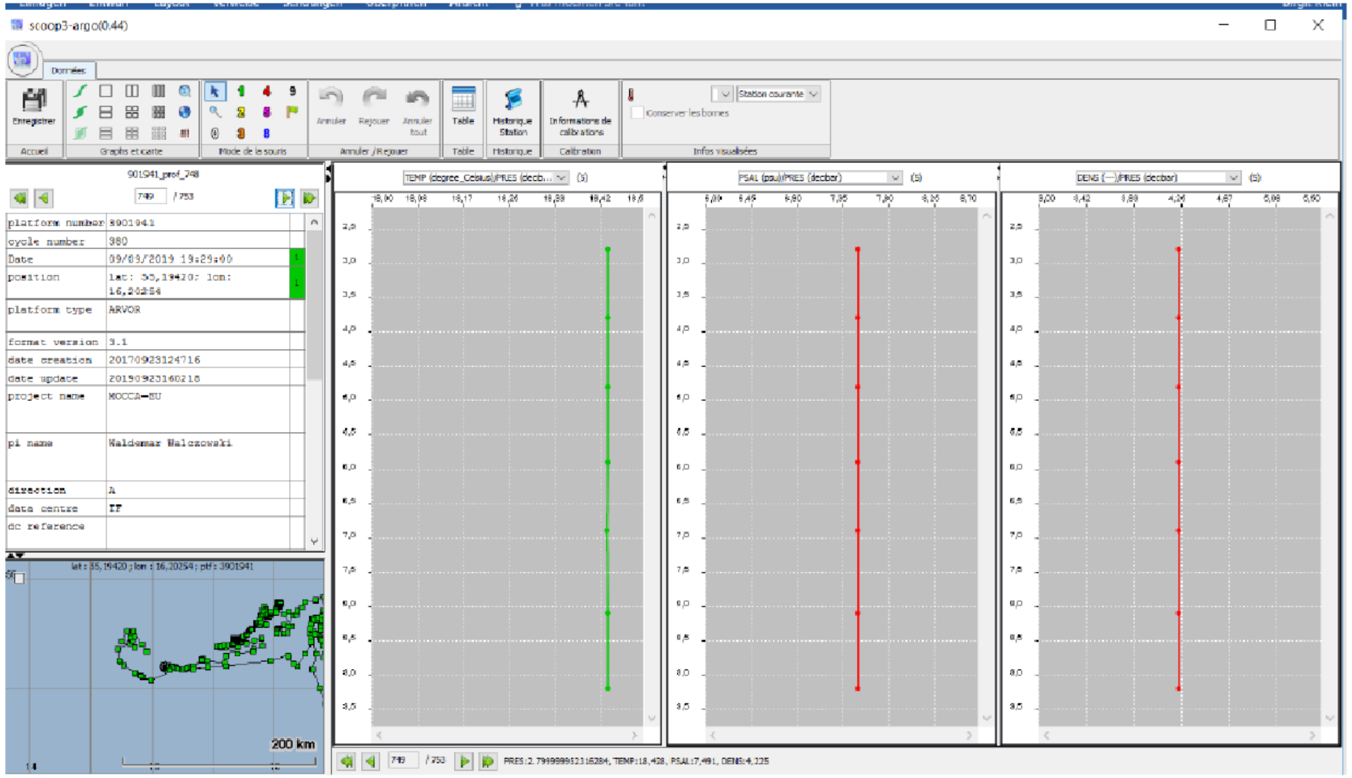
To detect the rollover the test considers temperature differences between adjacent pressures > 10°C as a sign of rollover and salinity differences of >5 psu. It was also never designed for strong, shallow thermo- and haloclines as encountered in the Baltic. In this example the test fails and wrongly flags cycle 173A as bad because of the strong thermocline gradient of > 10 °C . This demonstrates that this test should definitely be disabled for the Baltic.



Source: D2.7 ‘A report on the adaptation of existing DMQC methods to marginal seas’ https://www.euro-argo.eu/EU-Projects/Euro-Argo-RISE-2019-2022/Deliverables

Step by step guide based on the floats from Poland, step2

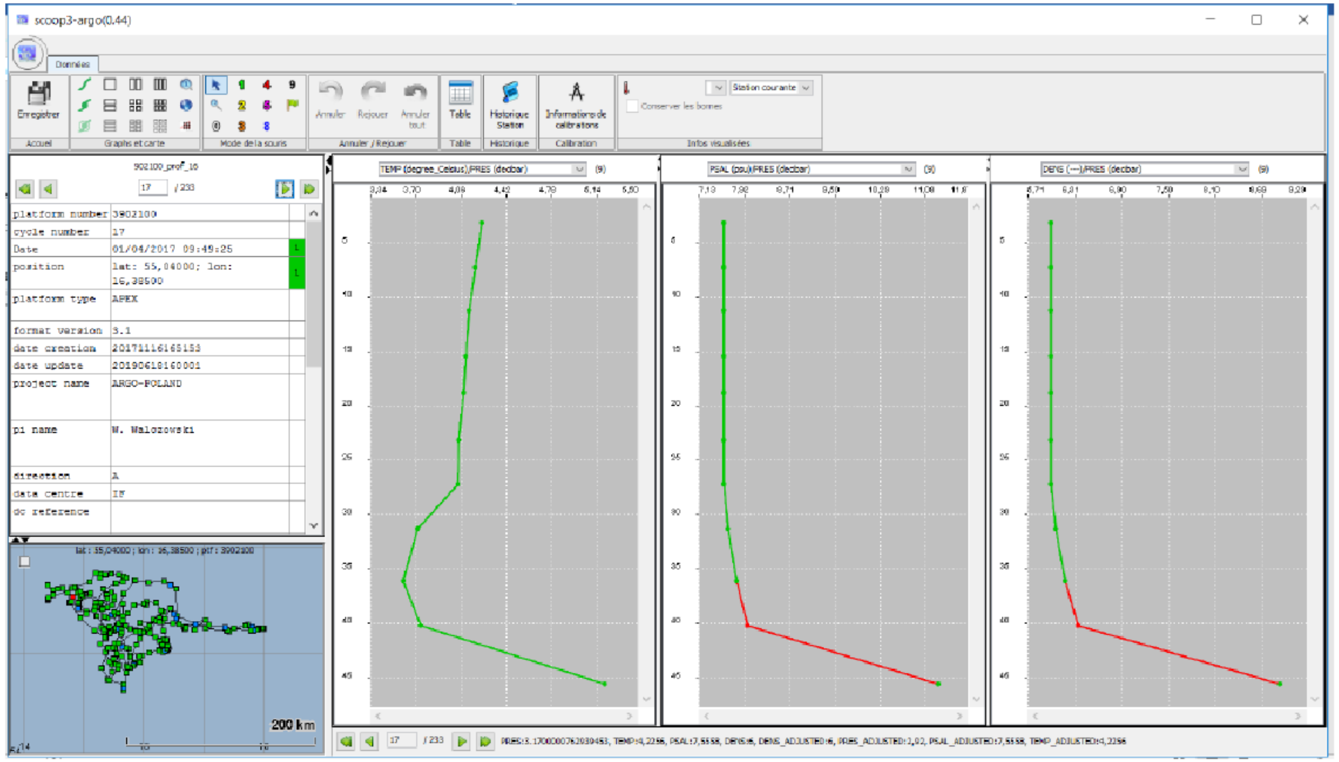
Stuck value test: The stuck value test looks for measurements of temperature and salinity in a profile being identical.
In the example the profiles of temperature and salinity are nearly constant. All salinities are exactly the same and thus are flagged as bad, while temperatures at least show a 0.6 mK standard deviation and thus escaped a degradation in flagging. This test was never intended to work on short profiles and has the potential to catch homogenous winter profiles. It should be disabled for the Baltic.



Source: D2.7 ‘A report on the adaptation of existing DMQC methods to marginal seas’ https://www.euro-argo.eu/EU-Projects/Euro-Argo-RISE-2019-2022/Deliverables

Step by step guide based on the floats from Poland, step 2

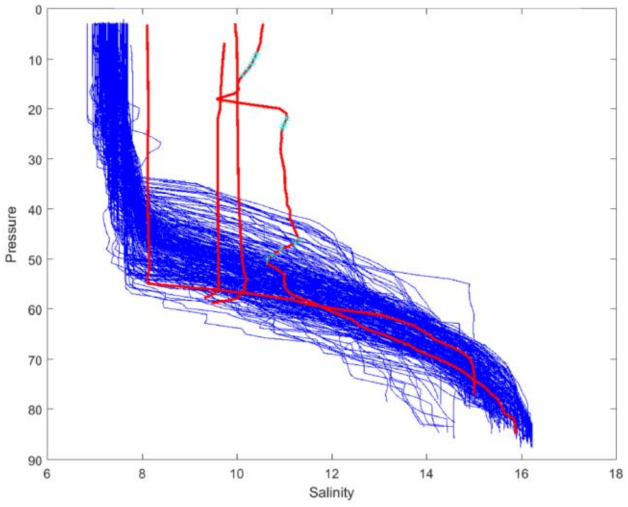
Gradient test: This test had been designed to assign a ‘bad data’ quality flag when the difference between vertically adjacent measurements is too steep and has the potential to catch strong gradients in deep layers of the Baltic caused by the inflowing North Sea waters. It was declared obsolete in October 2019 at ADMT20 but it seems that the data from the Baltic floats need to be reprocessed to accommodate this decision.

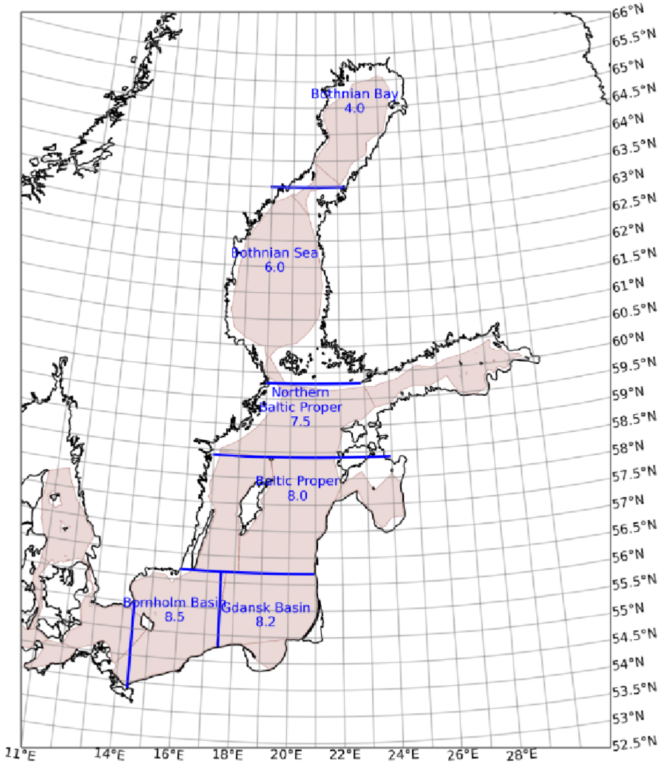


Source: D2.7 ‘A report on the adaptation of existing DMQC methods to marginal seas’ https://www.euro-argo.eu/EU-Projects/Euro-Argo-RISE-2019-2022/Deliverables

Step by step guide based on the floats from Poland, step 2

A new test ‚Incorrect near-surface salinity‘:
The current real-time tests in the Baltic failed to detect unreasonable high surface salinities. Surface salinities in the Baltic are quite low because of the large fresh water supply to the Baltic. Sometimes, however, recorded float profiles exhibit surface salinities in excess of 8 salinity units. The issue was initially named clogging and was associated with a insufficient flushing of the conductivity cell with salty water from the deep layer





Source: D2.7 ‘A report on the adaptation of existing DMQC methods to marginal seas’ https://www.euro-argo.eu/EU-Projects/Euro-Argo-RISE-2019-2022/Deliverables

Step by step guide based on the floats from Poland

Step 3: Following steps need reference database as *.mat file e.g. IOPAN_database.mat. Has be created by Gosia for the present reference data. But you can also create your own database or use ices database from the ctd_xxxx.mat files

Caveats and points for discussion:

- How much quality control of the reference data is required
- How frequent do we update the reference data, who is performing the updates and sorting into wmo-boxes
- What consequences does this have for the scheduling of dmqc

The mat file contains the structure variable CTD from all the IOPAN cruises
With the attributes Latitude, Longitude, Pressure, Temperature, Salinity, Time

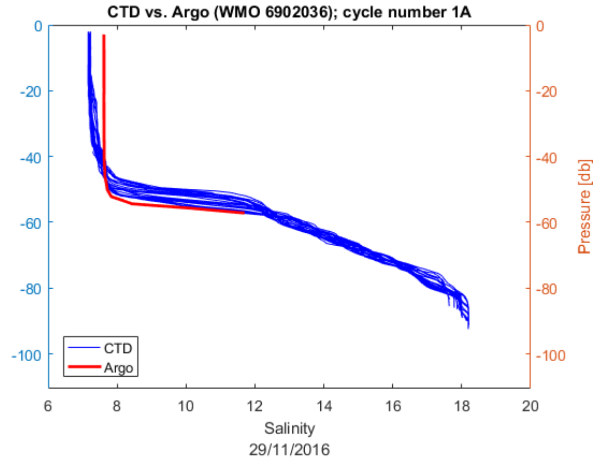
Step by step guide based on the floats from Poland

Step 4: Run **IOPANplotsDMQC.m** script that searches for reference data (IOPAN_database.mat) at a distance of 30 km and 30 days for the float selected. It creates figures for each float cycle with the float data compared to the reference data.

Caveats and points for discussion:

- The distance in time and space can be changed in the script depending on the needs of the region.
- Which rules should we apply, limitations from physical constraints or select until enough data are found
- What is enough data?

A plot of vertical salinity profiles is created for each Argo cycle, do we need temperature as well? What needs to be checked from the plots.



Step by step guide based on the floats from Poland

Step 5: Run **IOPANmeanDMQC.m** script that calculates, for each Argo profile and the closest CTD profiles, the mean salinity in the selected layer (S=10-30 m, T=70-90m) and difference between them.

Caveats and points for discussion:

- Profiles that have salinity difference greater than 0.2 should be flagged as 4 (bad data). Is that consensus or are the possibilities for such large differences?
- Other rules to apply?
- How to determine drift of the salinity sensor?
- Other quality checks needed?

Input is mat file for the selected float and the reference data base
Output is a mat file xxxxxx_meanRTQC with a structure varaiable srednie
Inside script the conditions for selection are specified as search radius and selection of depth layers.
No checks on number of data in the mean, should there be a limit?
No checks on variabiliy, should we do another check if reference data have reasonable variability?

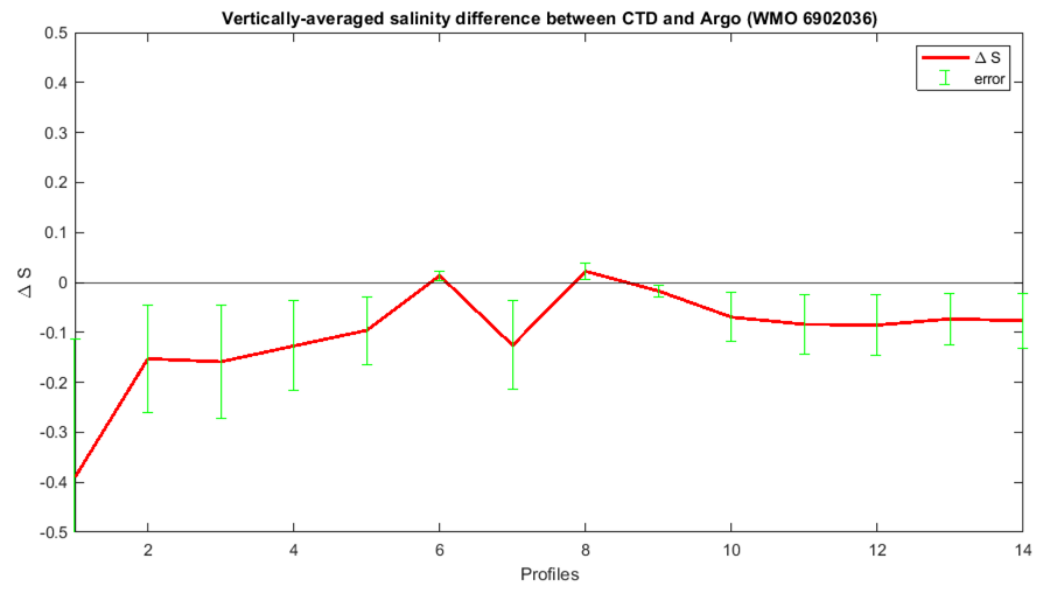
Step by step guide based on the floats from Poland

Step 6: Run **IOPANplotsDIFF.m** (Matlab R2018b needed for rmmissing) script that plots salinity difference between CTD and Argo.

Caveats and points for discussion:

- Is that sufficient to visualize the success of the dmqc?
- How to determine drift of the salinity sensor?

Input is the matfile with the mean values of the argo and reference data,
! Y-axis plotted here is just the index of the differences and not cycle number



Step by step guide based on the floats from Poland

Step 7: Create D-files and submit them to Coriolis

Caveats and points for discussion:

- First needs formal approval of ADMT
- Problem: No guidelines for D files with data from the Baltic Sea area and no scripts to execute them
- What to write into the SCIENTIFIC_CALIB_FIELDS
- What to used for PSAL_ADJUSTED_ERROR
- What is our expected accuracy for these data and will users be ok with larger errors?

Checking the sensor drift with calibration data sheets, example

SENSOR SERIAL NUMBER: 4793
CALIBRATION DATE: 27-Mar-13

COEFFICIENTS:

```
g = -9.821290e-001
h =  1.423607e-001
i = -3.273814e-004
j =  4.338152e-005
```

SENSOR SERIAL NUMBER: 4793
CALIBRATION DATE: 29-Mar-15

COEFFICIENTS:

```
g = -9.821266e-001
h =  1.423310e-001
i = -3.134470e-004
j =  4.250802e-005
```

SENSOR SERIAL NUMBER: 4793
CALIBRATION DATE: 24-Jan-17

COEFFICIENTS:

```
g = -9.836474e-001
h =  1.427120e-001
i = -4.180454e-004
j =  5.025814e-005
```

WMO number	Float serial. No	CTD serial no.	Float type	Country/ Programme	Deployment date
6901901	5397	3511	APEX	Argo Finland	17.05.2012
6902013	5396	3503	APEX	Argo Finland	13.06.2013
6902014	6711	4793	APEX	Argo Finland	14.08.2013
6902017	5397	3511	APEX	Argo Finland	31.05.2014
6902018	6710	5051	APEX	Argo Finland	31.05.2014
6902019	7191	5699	APEX	Argo Finland	21.08.2014
6902020	6711	4793	APEX	Argo Finland	05.08.2015
6902021	6710	5051	APEX	Argo Finland	22.09.2015
6902022	5396	3503	APEX	Argo Finland	13.05.2016
6902023	5397	3511	APEX	Argo Finland	13.07.2016
6902024	7191	5699	APEX	Argo Finland	03.08.2016
6902036	7507	7248	APEX	Argo Poland	29.11.2016
6902025	7958	8893	APEX	Argo Finland	09.05.2017
6902026	7959	8894	APEX	Argo Finland	06.06.2017
6902027	6711	4793	APEX	Argo Finland	15.06.2017
6902028	6710	5051	APEX	Argo Finland	06.08.2017

Recovery of floats is practised routinely in the Baltic.
Floats are redeployment many times (example SN4793) and drift can be calculated from SBE‘s laboratory analysis

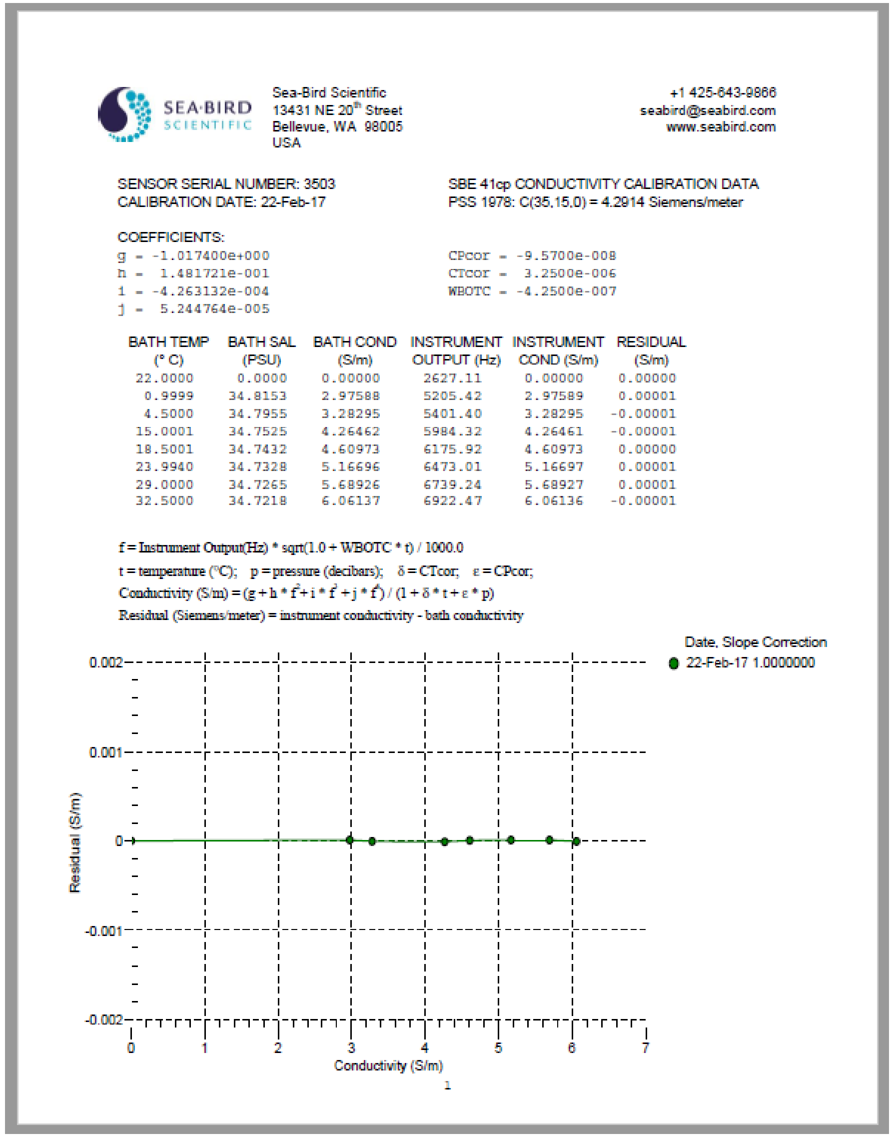
$$\text{Conductivity (S/m)} = (g + h * f^2 + i * f^3 + j * f^4) / 10 (1 + \delta * t + \varepsilon * p)$$

Procedure and tools (in development)

- So far experience only with APEX floats
  - Teledyne/Webb research
  - Provides calibration as PDF files
- So far applied on the salinity drift
  - Sensor calibrated with:

    $$\text{Conductivity (S/m)} = (g + h * f^2 + i * f^3 + j * f^4)/10\,(1 + \delta * t + \varepsilon * p)$$

  - All the parameters available
- Small (python) script to calculate difference in calibration between deployments
  - Available in googledrive, with sample sheets.
    - https://drive.google.com/drive/folders/1ITElbJGUlnPrleP8BMCm7AuTjtDup8L5?usp=share_link



SEA·BIRD SCIENTIFIC

Sea-Bird Scientific
13431 NE 20th Street
Bellevue, WA 98005
USA

+1 425-643-9866
seabird@seabird.com
www.seabird.com

SENSOR SERIAL NUMBER: 3503
CALIBRATION DATE: 22-Feb-17

SBE 41cp CONDUCTIVITY CALIBRATION DATA
PSS 1978: C(35,15,0) = 4.2914 Siemens/meter

COEFFICIENTS:

g = -1.017400e+000
h = 1.481721e-001
i = -4.263132e-004
j = 5.244764e-005

CPcor = -9.5700e-008
CTcor = 3.2500e-006
WBOTC = -4.2500e-007

BATH TEMP (° C)	BATH SAL (PSU)	BATH COND (S/m)	INSTRUMENT OUTPUT (Hz)	INSTRUMENT COND (S/m)	RESIDUAL (S/m)
22.0000	0.0000	0.00000	2627.11	0.00000	0.00000
0.9999	34.8153	2.97588	5205.42	2.97589	0.00001
4.5000	34.7955	3.28295	5401.40	3.28295	-0.00001
15.0001	34.7525	4.26462	5984.32	4.26461	-0.00001
18.5001	34.7432	4.60973	6175.92	4.60973	0.00000
23.9940	34.7328	5.16696	6473.01	5.16697	0.00001
29.0000	34.7265	5.68926	6739.24	5.68927	0.00001
32.5000	34.7218	6.06137	6922.47	6.06136	-0.00001

f = Instrument Output(Hz) * sqrt(1.0 + WBOTC * t) / 1000.0

t = temperature (°C); p = pressure (decibars); δ = CTcor; ε = CPcor;

$$\text{Conductivity (S/m)} = (g + h * f^2 + i * f^3 + j * f^4) / (1 + \delta * t + \varepsilon * p)$$

Residual (Siemens/meter) = instrument conductivity - bath conductivity

1

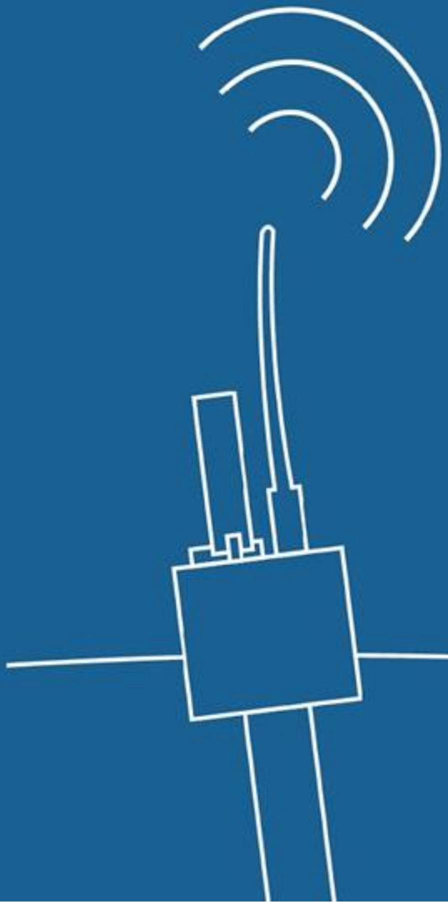
So, with recovered floats:

Recommendations for DMQC of recovered floats:

- Aim for recovery on annual or biannual basis, recalibrate the float before redeployment
- After consecutive lab calibrations only correct for significant drift (>0.1 conductivity units /12 month)
- Give DMQC operators access to the calibration sheets

Further things to do:

- Remember to store calibration sheets when available!
    - Should there be a common database?
- Modify tool to accept various kind of calibration sheets
- Need to convert to Matlab?

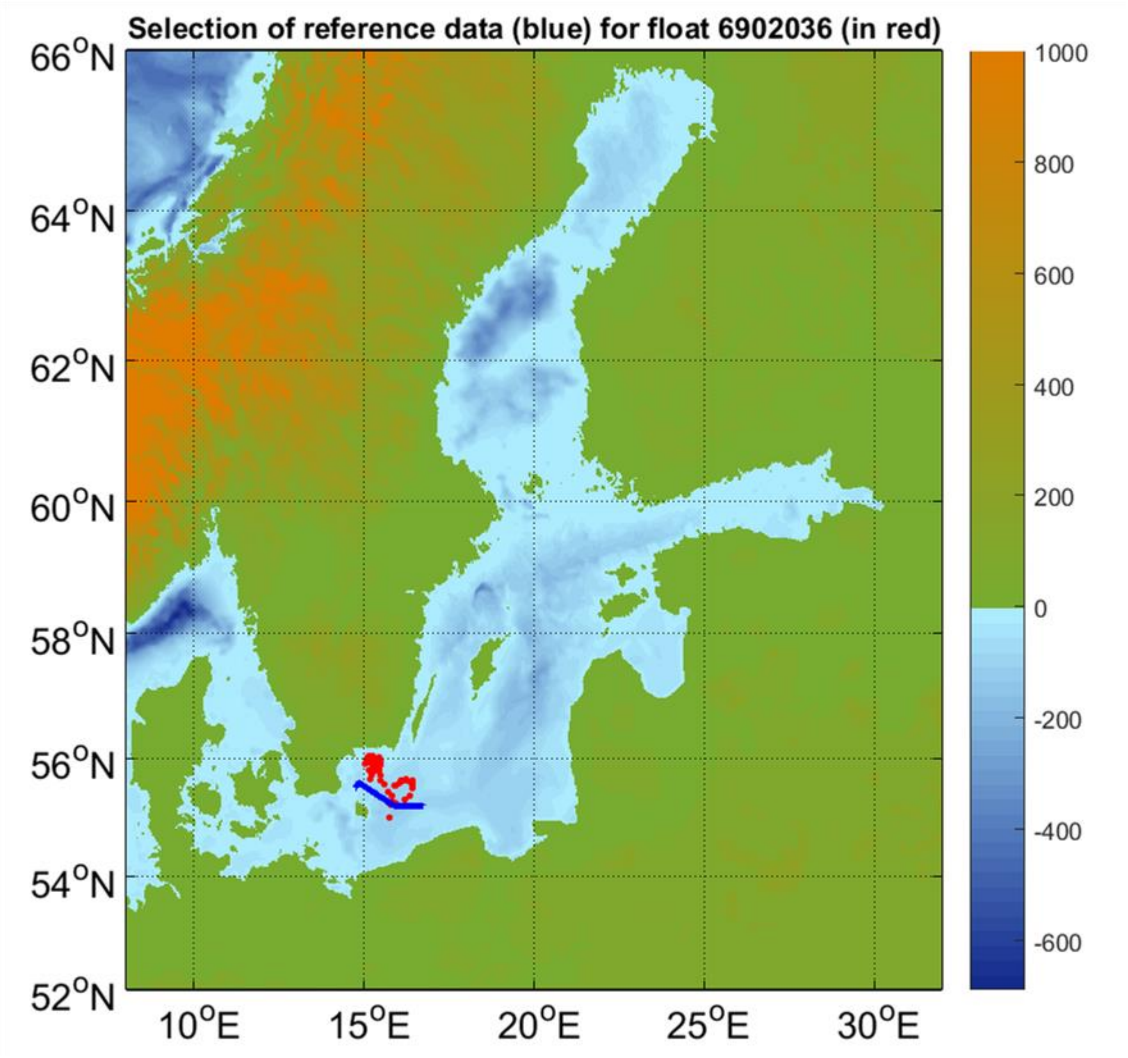


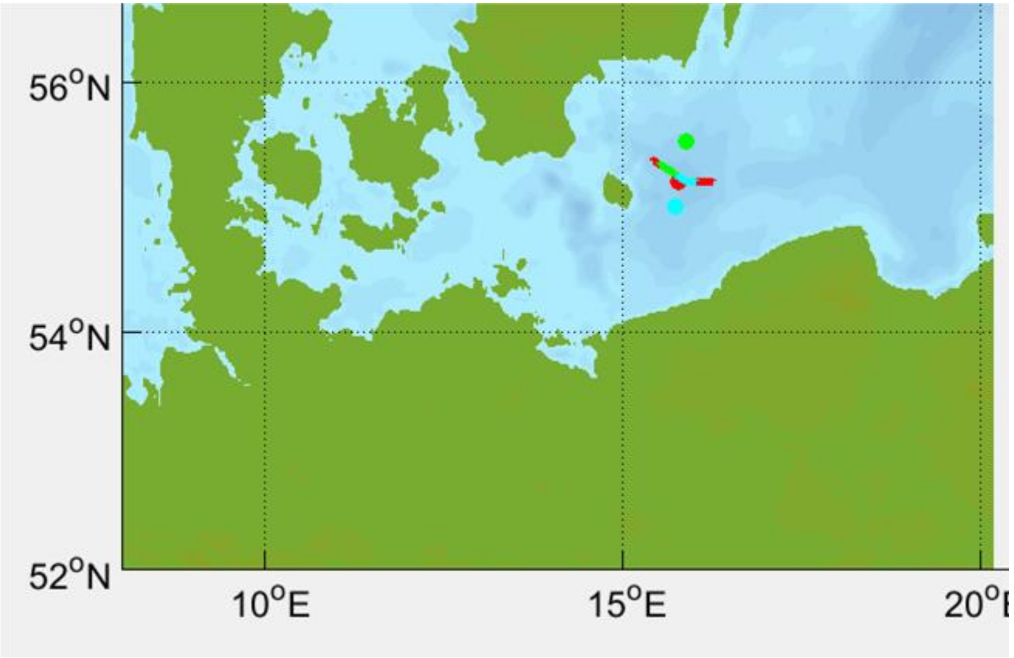
EURO-ARGO.EU

contact@euro-argo.eu

@EuroArgoERIC

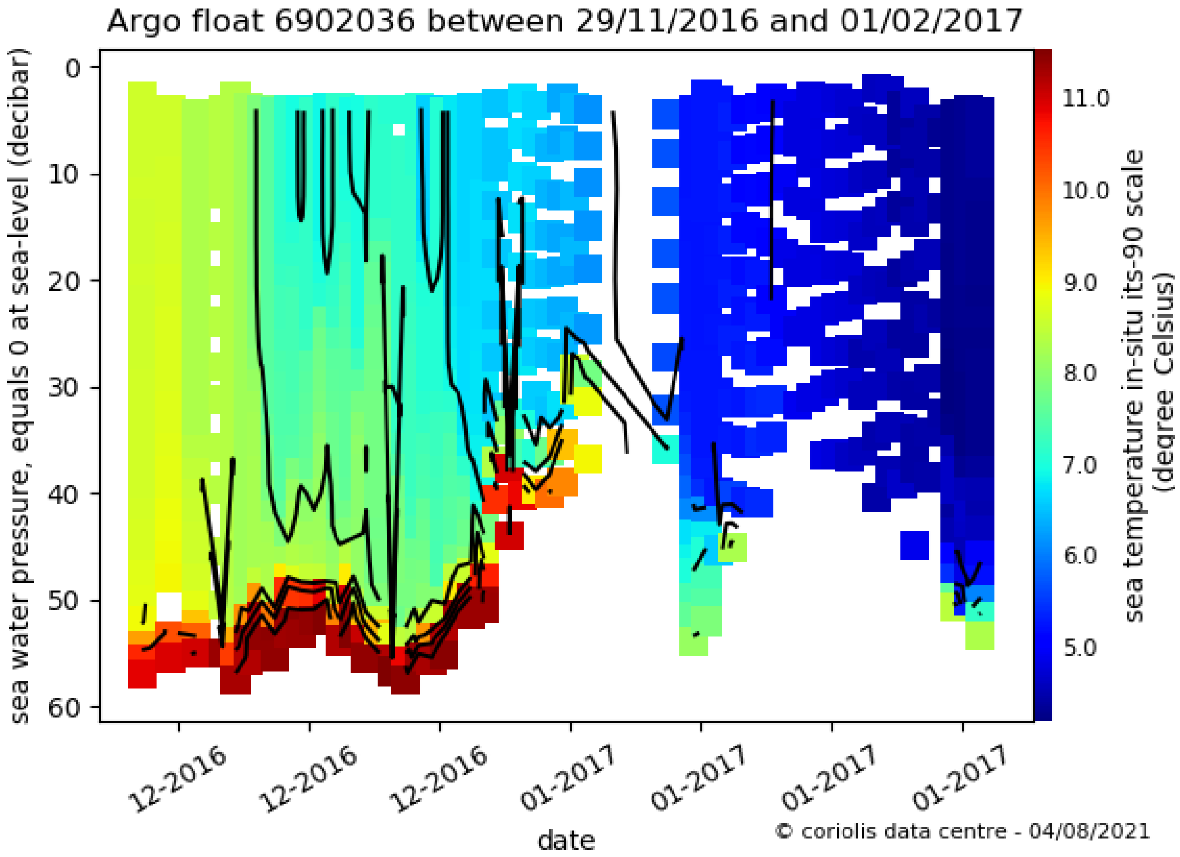
Additional plots for discussion and further steps

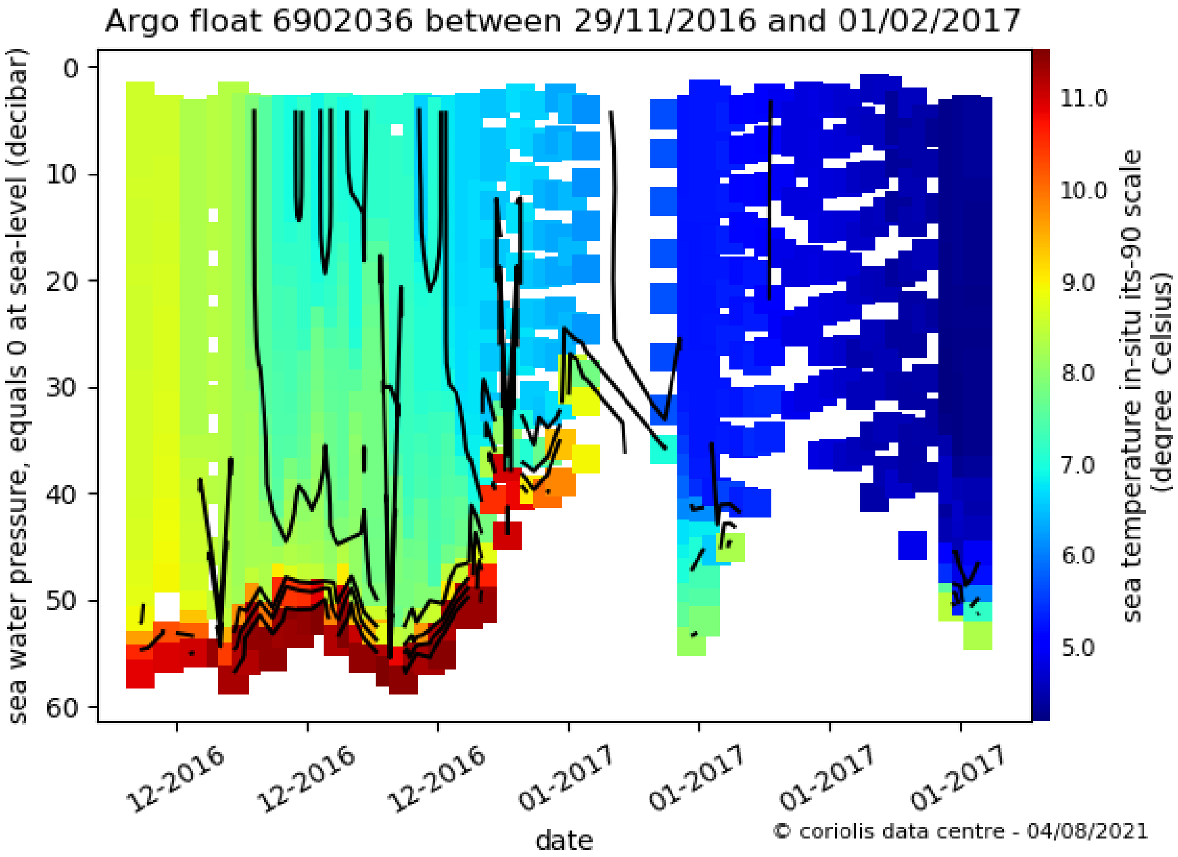




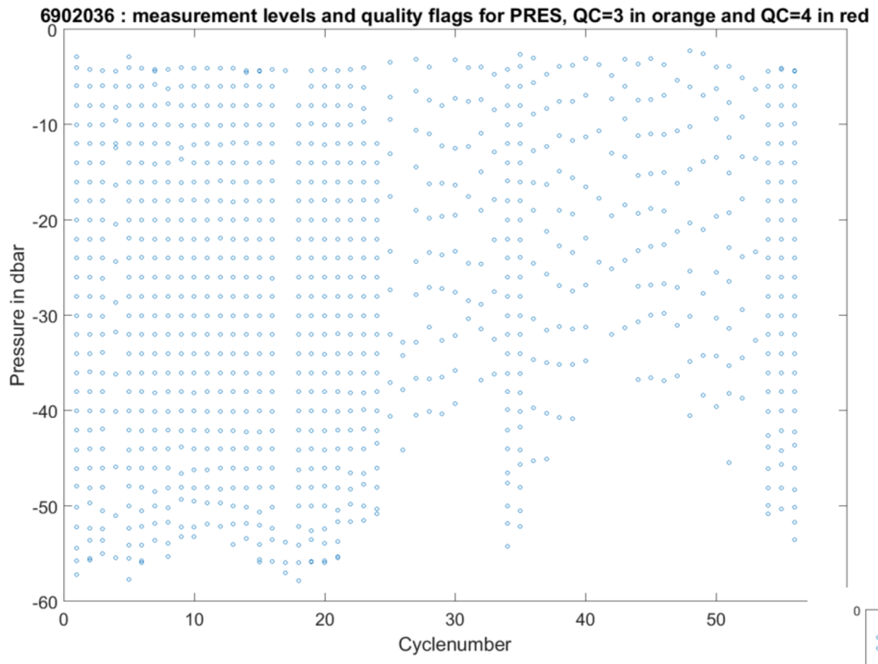
Selects different parts of the IOPAN CTD reference data, but with overlap. Argo profile 1 has 29 matches, **Argo profile 2** is nearly on section and gives 84 matches, will **Argo profile 4** has only 21 matches
Do we consider all the matches as independent?

Additional plots for discussion and further steps

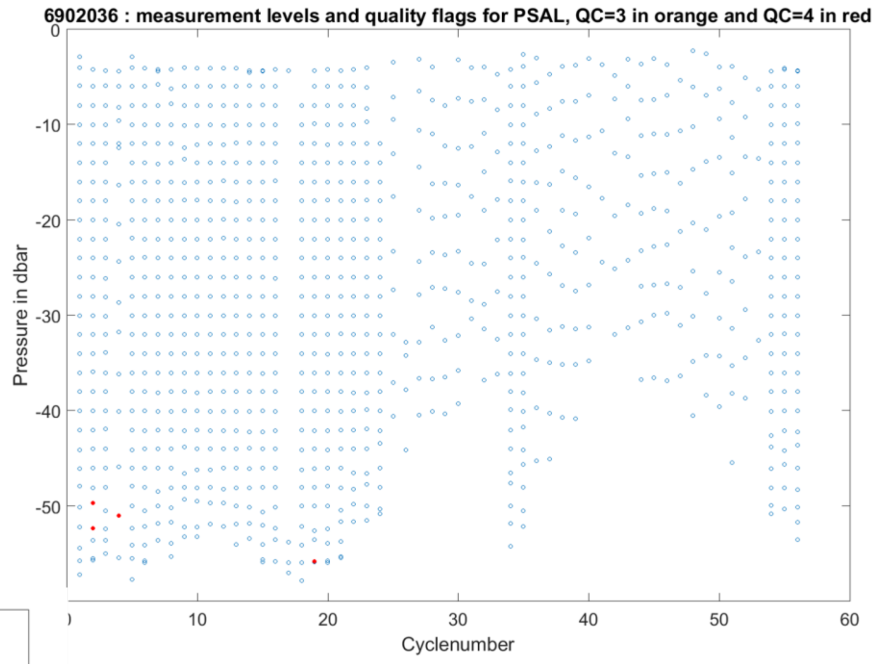


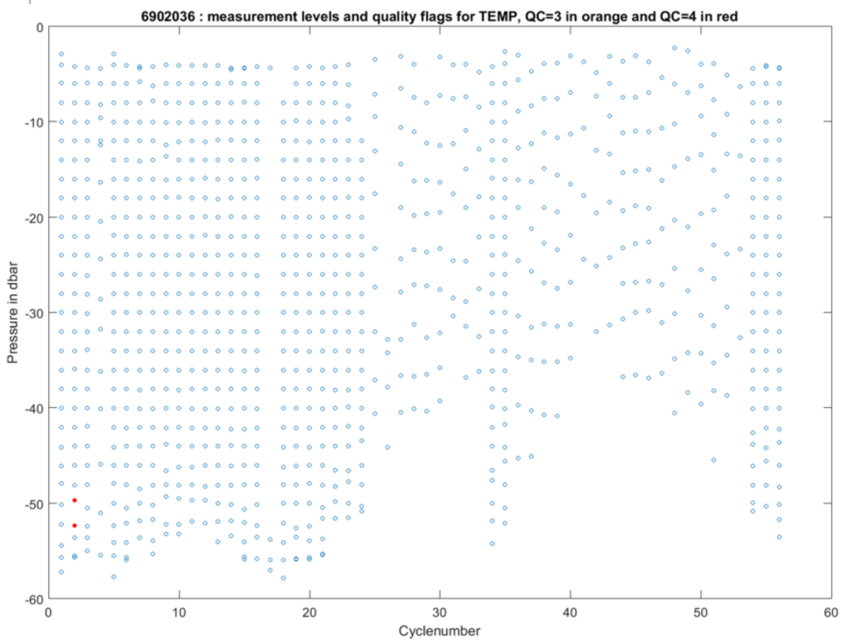


Additional plots for discussion and further steps

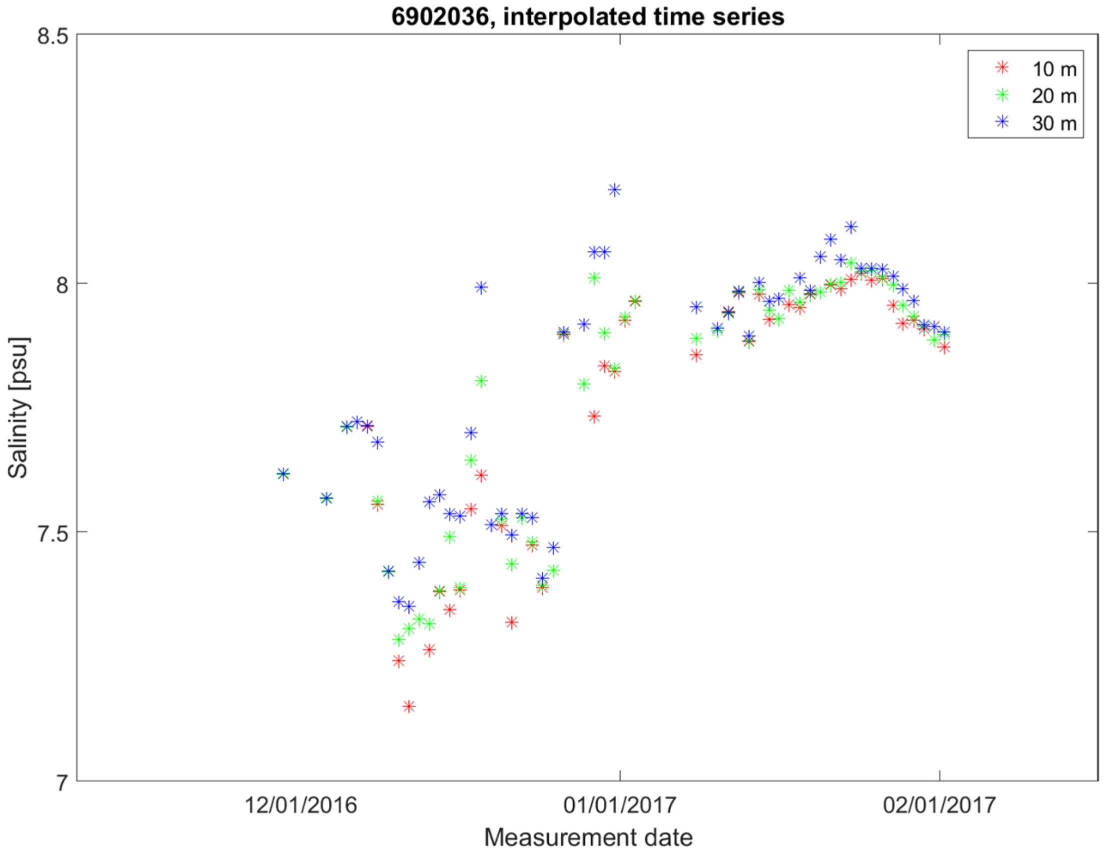


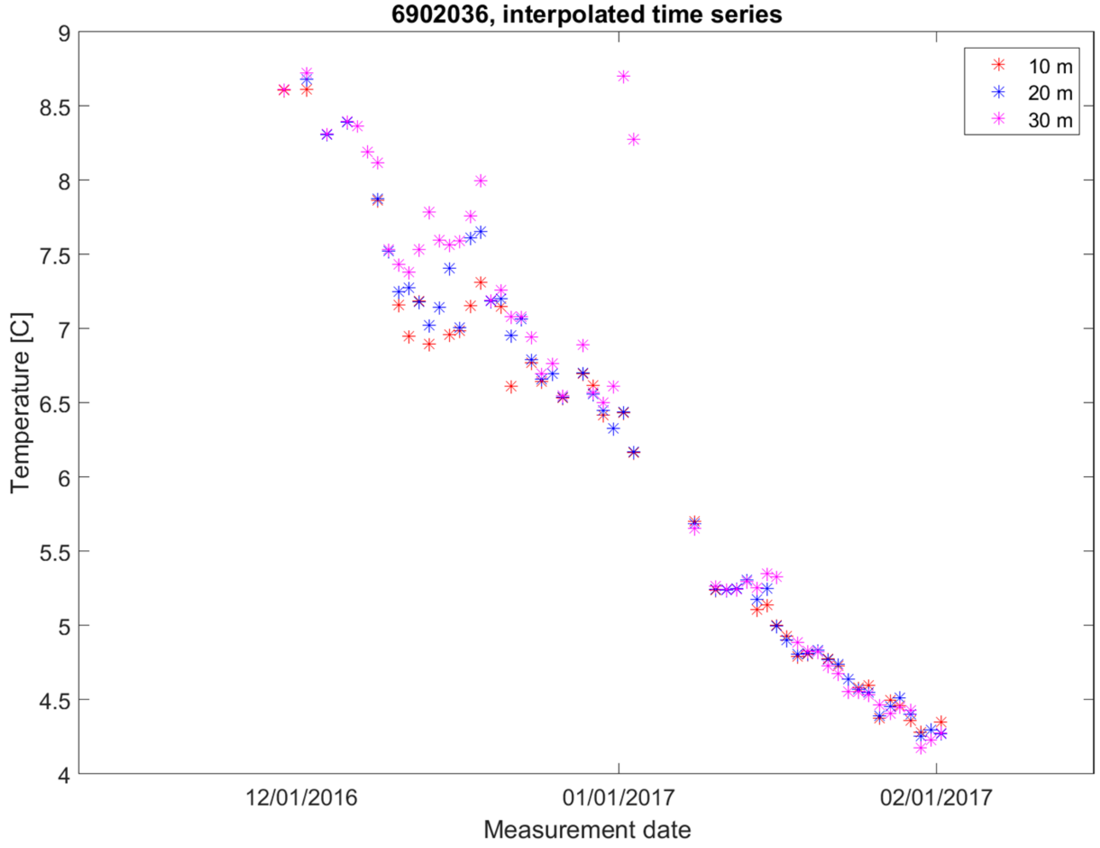
Why the irregular measurement levels? Is that related to the hydraulic capacity of the APEX buoyancy engine?



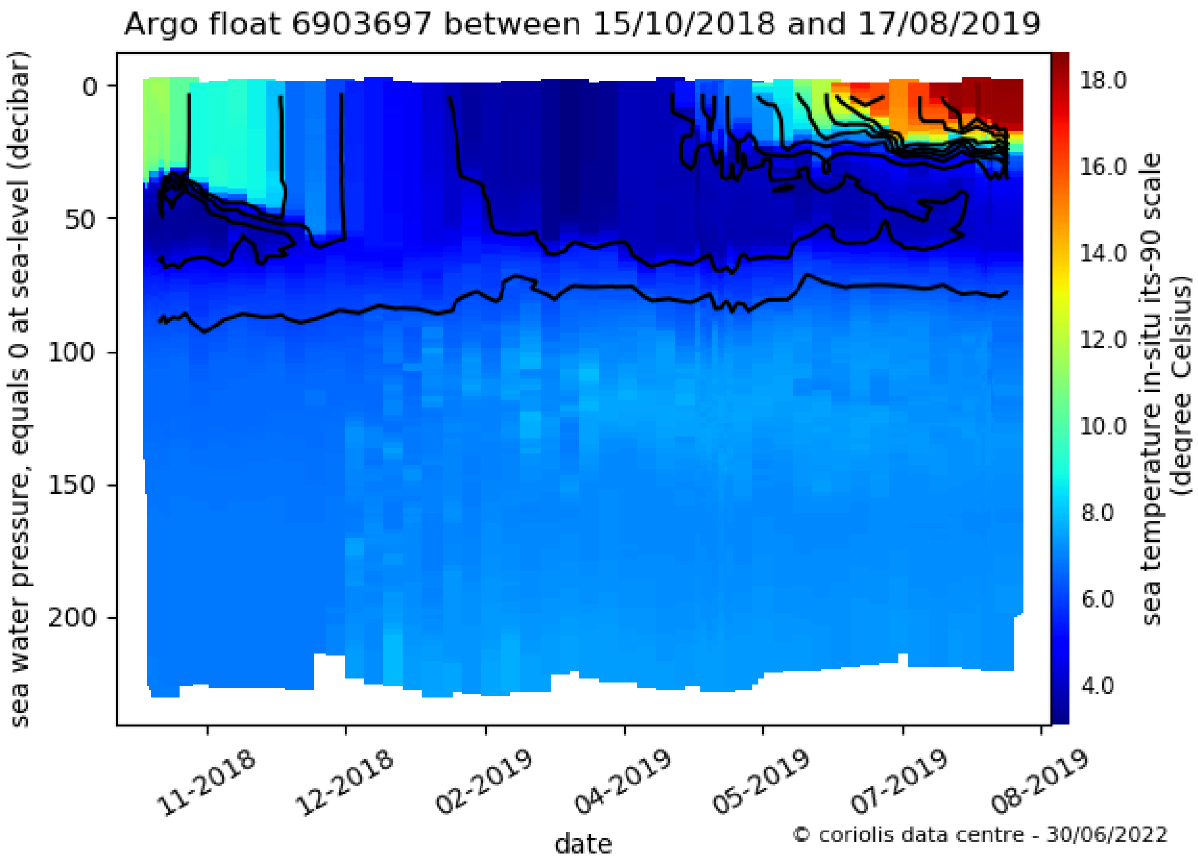


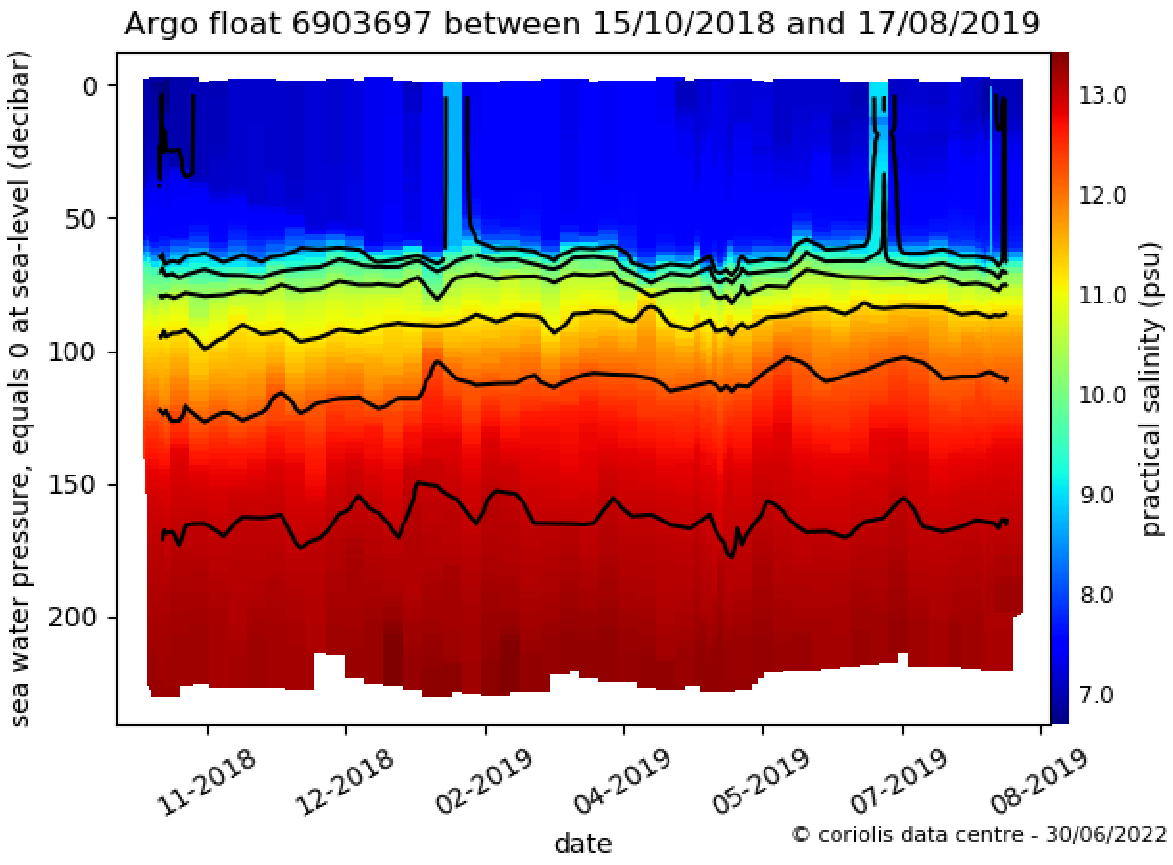
Additional plots for discussion and further steps



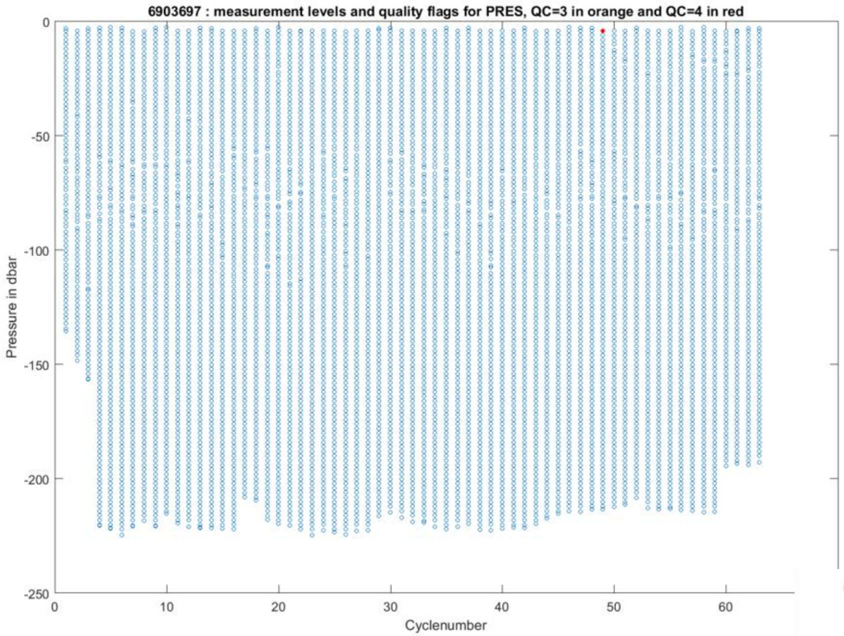


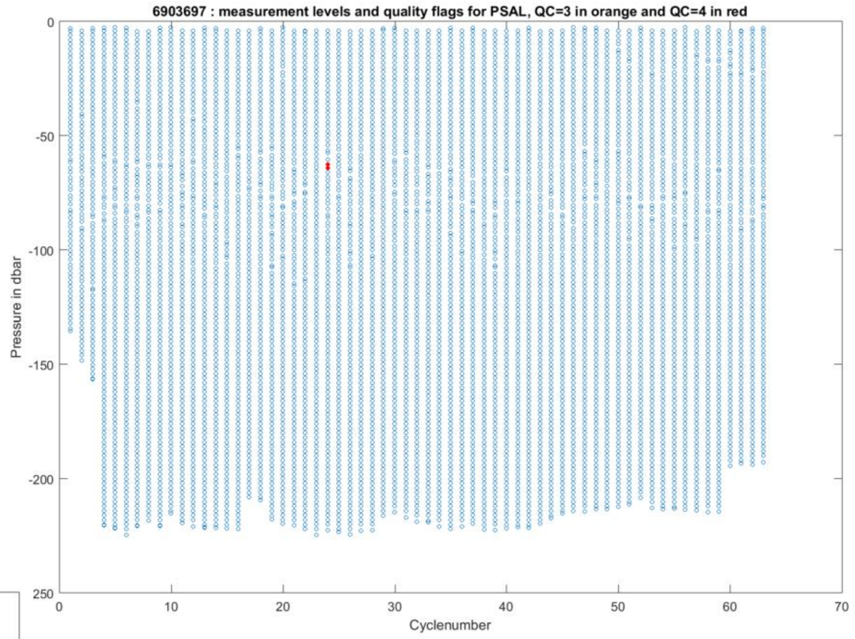
Additional plots for discussion and further steps

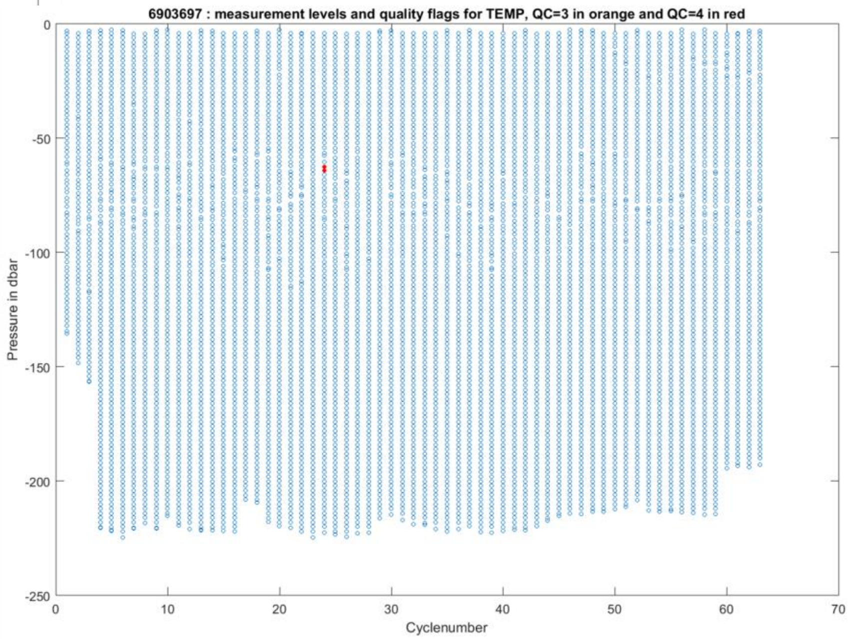




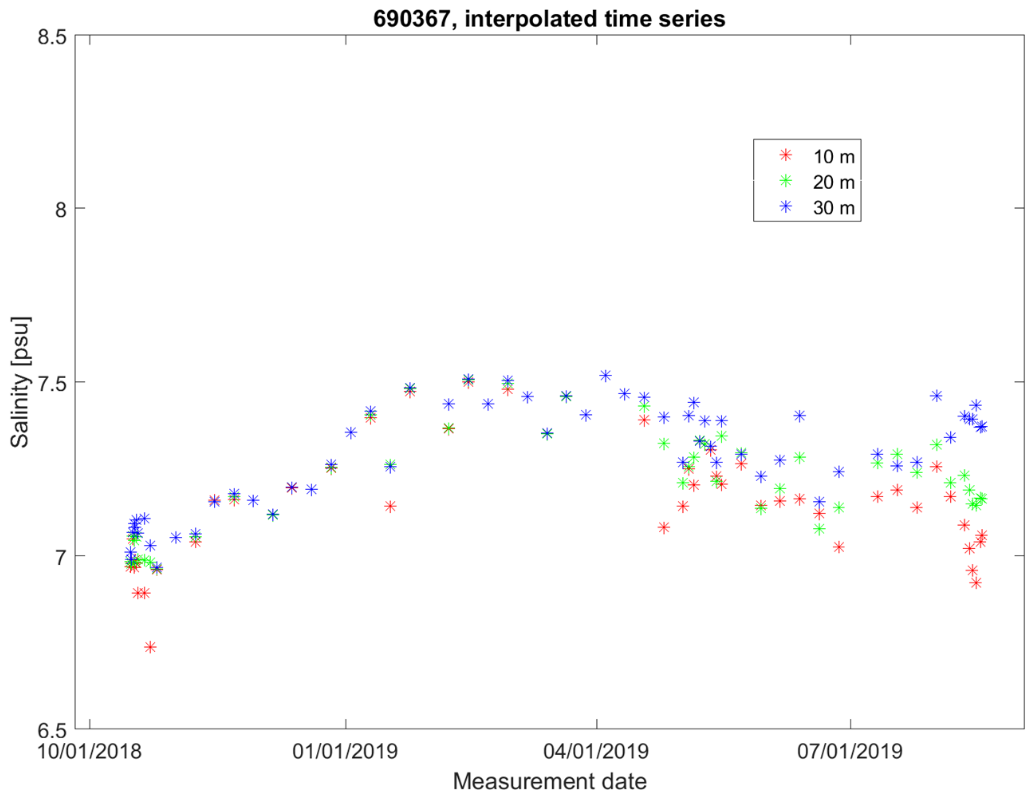
Additional plots for discussion and further steps

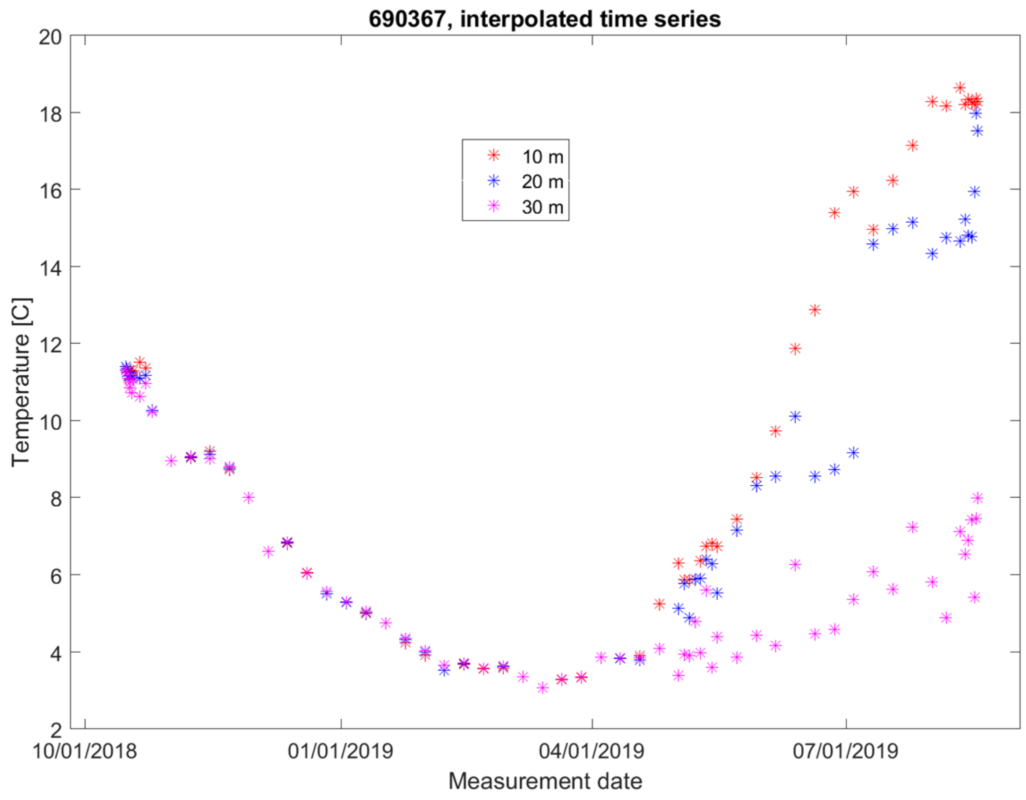




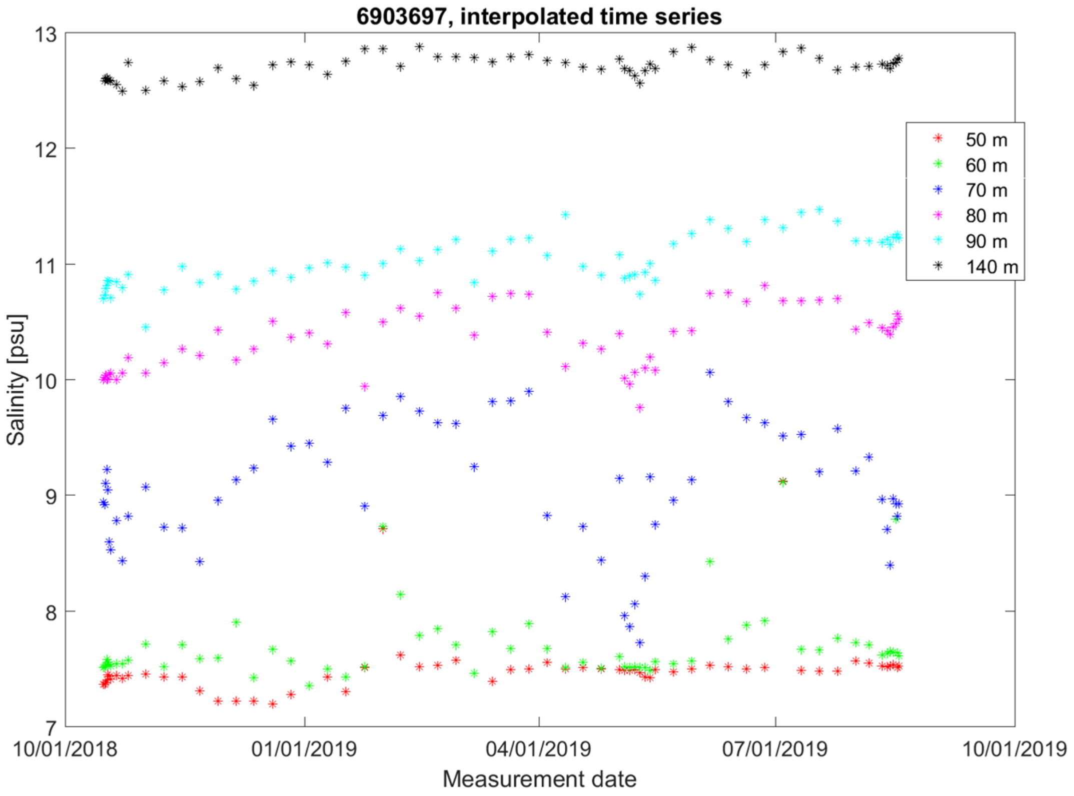


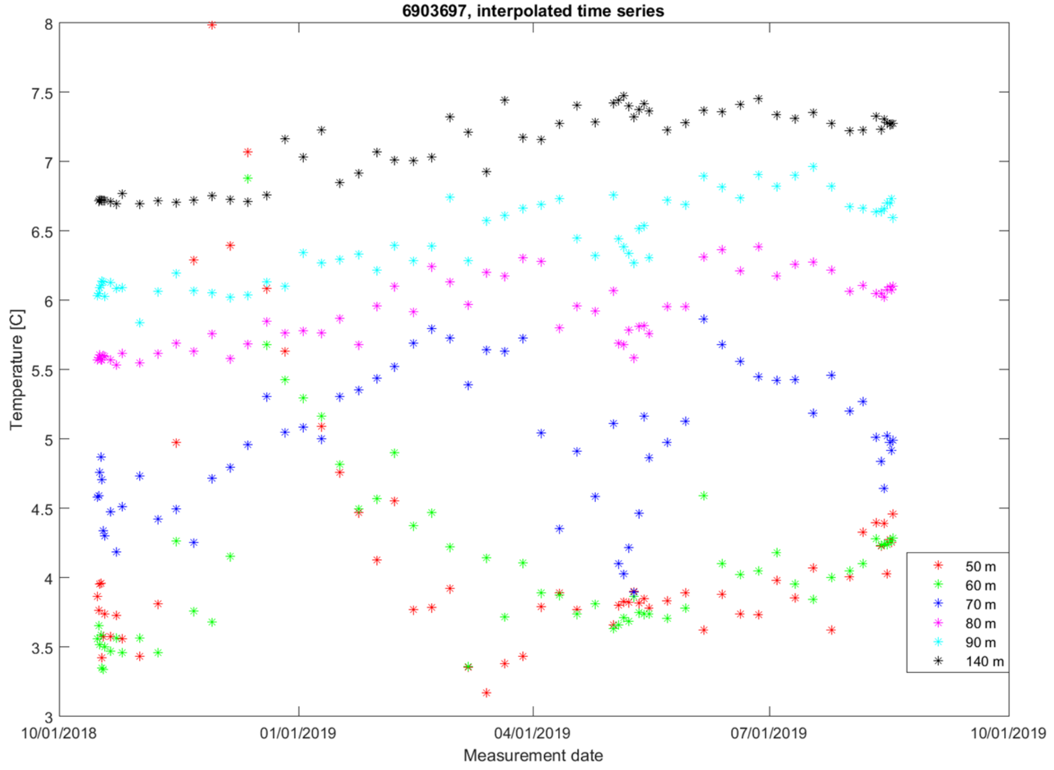
Additional plots for discussion and further steps





Additional plots for discussion and further steps





Additional plots from me to be integrated ref data fmi