EUROPEAN COMMISSION
Executive Agency for Small and Medium-sized Enterprises (EASME)

Department A - COSME, H2020 SME and EMFF
Unit A3 - EMFF

Agreement number: EASME/EMFF/2015/1.2.1.1/SI2.709624

Project Full Name: Monitoring the Ocean Climate Change with Argo

European Maritime and Fisheries Fund (EMFF)

MOCCA

D4.4.8 Review and development of DMQC training and resources

Circulation:	PU: Public
Lead partner:	British Oceanographic Data Centre
Contributing partners:	National Oceanography Centre
Authors:	Matthew Donnelly, Kamila Walicka, Clare Bellingham, Ed Small
Quality Controllers:	Romain Cancouët, Sylvie Pouliquen
Version:	1.0
Reference	D4.4.8 Review and development of DMQC training and resources
Date:	20.08.2020

Euro-Argo ERIC
European Research Infrastructure
(2014/261/EU)

Consisting of:

Organisation/Natural person	Represented by	Statute	Contributing entities[1]
Euro-Argo ERIC	N/A	Coordinator	N/A
The French Republic	Ifremer	Member	SHOM, INSU/CNRS, Meteo-France, IRD, IPEV
The Federal Republic of Germany	BSH	Member	GEOMAR, University of Hamburg, Alfred-Wegener-Institute for Polar and Marine Research (AWI)
The Hellenic Republic	HCMR	Member	N/A
The Italian Republic	OGS	Member	N/A
The Kingdom of the Netherlands	KNMI	Member	N/A
The Republic of Finland	FMI	Member	N/A
The United Kingdom of Great Britain and Northern Ireland	Met Office	Member	NOC, BODC
The Kingdom of Norway	IMR	Observer	N/A
The Republic of Poland	IOPAN	Observer	N/A

Document History

Version[2]	Issue Date	Stage	Content and Changes
0.1	10.07.2020	Draft	Initial document creation
0.2	04.08.2020	Draft	Revision
0.3	05.08.2020	Draft	Revision
0.4	11.08.2020	QC	For internal quality control
1.0	20.08.2020	Final	Final version for submission

[1] As indicated in the "Technical and Scientific description of the Euro-Argo ERIC" July 2013 attached to the Euro-Argo Statutes.

[2] Integers correspond to submitted versions.

Table of Contents

1. INTRODUCTION

This document describes the contribution that the British Oceanographic Data Centre (BODC), which is part of the National Oceanography Centre (NOC), has made to the review and development of resources for use by Argo Delayed-Mode Quality Control (DMQC) operators. DMQC operators are individuals who apply defined procedures to the unadjusted data transmitted by Argo floats through comparison with reference data. The aim of the DMQC process is to either confirm data from an Argo float is already sufficiently accurate for climate-grade research uses, to apply an at-sea calibration if the sensor has been identified as having a known and correctable issue, or to certify the data as beyond correction and unfit for scientific use.

To achieve the above, DMQC operators must be able to understand the procedures, understand float and sensor behavior, use the community DMQC software effectively, and make expert judgements. DMQC operators must also be able to identify when their knowledge has been exceeded and when to refer to Principal Investigators or regional data quality experts for further guidance.

Near the beginning of the MOCCA project, the composition of the BODC Argo Team began changing with existing members moving to other roles and new members joining. Whilst the necessary expertise and procedures were still available within BODC, it was necessary to train a new generation of DMQC operators. This included contributing to and benefitting from the 1st European Argo Delayed-mode QC Workshop in 2018, from which the need for improved knowledge sharing and tools was identified. Additionally, BODC aimed to make a broader contribution to reviewing and enhancing DMQC training and associated resources as part of a wider community effort capture knowledge and experience not already found within research papers and Argo manuals.

2. REVIEW AND DEVELOPMENT OF DMQC RESOURCES

2.1. DMQC cookbook contribution

BODC (Kamila Walicka) has contributed to the development of an Argo DMQC cookbook for core parameters (*DMQC_cookbook)* led by Ifremer. This contribution covers:

- The guidelines regarding DMQC workflow of Argo core data (pressure, temperature, salinity), providing a list of steps from getting R-files (uncalibrated real-time) from the GDAC to sending the D-files (calibrated delayed-mode) back;
- Description of examples of hydraulic or sensor problems;
- A template has also been prepared with best practice for generating a report on the DMQC of an individual float;
- Description of DMQC analysis and decision-making process of a float deployed in the Southern Ocean.

2.2. OWC Python conversion

BODC had been planning to invest in the development of the DMQC software and workflow as part of the MOCCA project. As part of the EuroArgo RISE project (WP 2: Evolution of the Core Argo Mission), a survey of the entire core Argo DMQC community was undertaken by BODC (Kamila Walicka and Matt Donnelly) in 2019 to identify the barriers and opportunities to improving the efficiency and capacity of the overall community effort. It was identified as part of this survey that the existing DMQC software – known as OWC after the primary authors, Owens, Wong and Cabanes – being written in the software language and environment Matlab was a barrier for many institutions. This is because Matlab is paid-for licensed software and many institutions either had few licenses, or did not have the necessary additional licensed toolboxes. This has resulted in the use of different versions of Matlab and consequently different versions of the OWC software amongst the Argo community, as well as proving a barrier to some institutions being involved at all. A decision was reached to assess the potential for converting the OWC Matlab code to a free software, with the widely used Python being the preferred language.

A 1-year software developer post was recruited in the middle of 2019 and a year of detailed assessment and development has followed. The original code was first mapped and reviewed, and a set of recommendations proposed to the international Argo Data Management Team in October 2019. After approval to proceed with 'OWC Python' was given, the developer (Ed Small) supported by BODC DMQC operators (Kamila Walicka and Matt Donnelly) undertook several months of conversion work leading up to several phases of testing in mid-2020. The converted code is now functional, with a final round of performance enhancements, testing and evaluation underway to ensure it is ready for operational use.

As part of the development work, there has been close collaboration with Ifremer to prepare the software package to be used as part of a Jupyter notebook, and even to do so online using the Pangeo Binder. Combined with the potential to fully parallelize the analysis code, the conversion of OWC Matlab to OWC Python marks a step-change in capability and sets a new standard in quality control software development for the Argo community.

2.3. DMQC report template

A review of the current state of DMQC tools and methods (undertaken by BODC as part EuroArgo RISE project, see section 2.2) identified a large diversity in software used to create the final DMQC reports produced by DMQC operators at various institutions. The production of comprehensive and consistent DMQC reports is useful to document the decisions made to calibrate an Argo float, and in-turn enable the reproducibility of the DMQC analysis and inform future rounds of DMQC, either by the same operator or by future operators. This becomes increasingly important as a new generation of DMQC operators who were not active in the early stages of the Argo program become increasingly responsible for data which they were not originally involved in assessing.

To address this issue, BODC (Clare Bellingham and Kamila Walicka) have drafted a template of a DMQC report implemented using the free open-source LaTex (*https://www.latex-project.org/*) typesetting system. The DMQC report generator includes the detailed description of visual inspection of the float notes, comparison with satellite altimetry provided by CLS, the OWC configurations for the specific regions, diagnostic plots generated by OWC software, and scientific justification of the decisions made to determine a high-quality at-sea calibration for a given Argo float. The DMQC report template can be found in Appendix 1. The DMQC report template will be distributed via GitHub at the public QC forum (see: *https://github.com/euroargodev/publicQCforum*) and through Argo email lists for further discussion and refinements based on feedback from the global group of Argo DMQC operators.

2.4. DMQC workshop participation and coordination

BODC contributed to the preparation and delivery of the 1st European Argo Delayed-mode Workshop for CTD data held in Brest, France on 17th-18th April 2020. This included contribution to training material by Matt Donnelly and Justin Buck, delivery of material at the event by Justin Buck, and attendance by two new members of the BODC Argo Team as part of their training.

BODC has undertaken the preparation of hosting the 2nd European Argo/7th International Argo Delayed-mode Workshop for CTD data in Liverpool, UK. This workshop aimed to include the DMQC analysis for both core (2000 m) and deep (4000 m-6000 m) Argo floats, with the latter being a focus of the EuroArgo RISE project WP3 on developing deep DMQC methods. The agenda and registration were advertised with the support of the Euro-Argo ERIC Office via *https://www.euro-argo.eu/News-Meetings/Meetings/Others/2020-DMQC-workshop*. The meeting was planned to happen from 12th May to 15th May 2020. However, due to the ongoing pandemic of COVID-19 virus, this workshop has been postponed to an as yet undefined date, but the planning effort around venue, arrangements and format remain valid.

3. FUTURE

BODC has invested in community-level resources to improve the effectiveness, efficiency and sustainability of DMQC at UK, European and international level. The contribution to the cookbook has helped to capture knowledge and expertise not otherwise documented, whilst the conversion of OWC to Python and the development of a DMQC report template has laid the foundations for future improvements to DMQC in the Argo community. Whilst this work is directly applicable to the core Argo mission, it also sets a standard in the approach for the deep and biogeochemical Argo missions as well. Whilst the Argo DMQC workshop has been postponed, BODC will offer to host it when it does take place. The lessons learnt through the MOCCA project will be applied to future challenges funded through the EU H2020 project Euro-Argo RISE as well as UK national capability funding.

4. APPENDIX

Delayed Mode Quality Control of Argo float 6901180

Kamila Walicka

British Oceanographic Data Centre (BODC), National Oceanography Centre
Joseph Proudman Building, 6, Brownlow St, Liverpool L3 5DA

31 July 2020

Float decision

Profile 1-x QC=xx, error=xx. Cell Thermal Mass corrections not applied. No corrections required.

1

Contents

2

1 Introduction

Delayed mode analysis was performed for float WMO number 6901180 (WMO: 6901180) where salinity and temperature values were separately compared to nearby historical CTD profiles and nearby Argo profiles as a reference database. The OWC (Cabanes et al., 2016) method was run to estimate a salinity offset and a salinity drift. For more information about float 6901180click on the following link: http://www.ifremer.fr/argoMonitoring/float/6901180

2 Quality Check of Argo Float Data

2.1 Visual check of float data

The visual inspection showed no further need of manual corrections of QC flags on this float.

2.2 Satellite Altimeter comparison

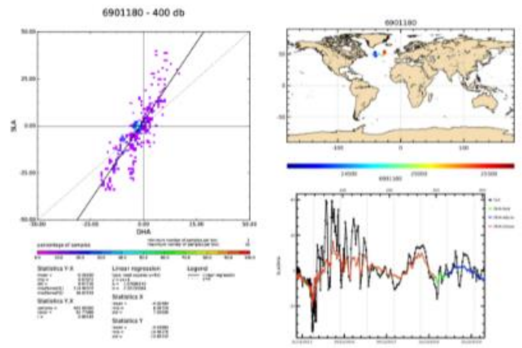

Figure 1: Float 6901180. The comparison between the Sea Surface Height (SSH) from the satellite altimetry and Dynamic Height Anomaly (DHA) extracted from the Argo float temperature and salinity data (ftp://ftp.ifremer.fr/ifremer/argo/etc/argo-ast9-item13-AltimeterComparison/figures/).

3

2.3 Time Series of Vertical Distribution of Data

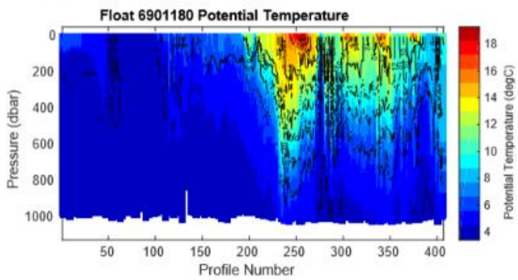

Figure 2: Float 6901180. Time series of the vertical distribution of potential temperature (°C).

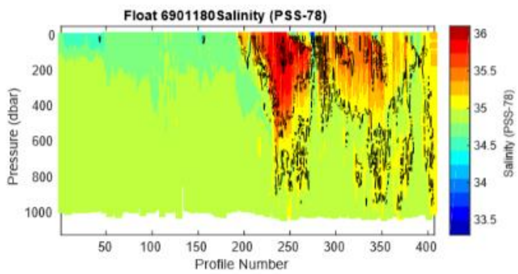

Figure 3: Float 6901180. Time series of the vertical distribution of salinity (PSS-78).

4

2.4 Comparison between Argo Float and Climatology

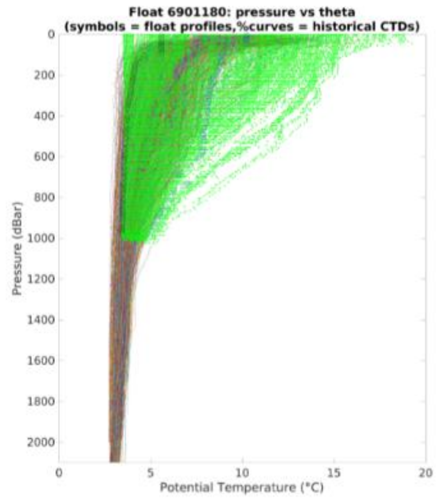

Figure 4: Float 6901180. Potential temperature (°C) plotted with pressure (dBar) and data from WMO boxes of CTD reference data (CTD for DMQC 2019V01) +/- 10 °of latitude and longitude. The black and blue cycles indicates the first and the last Argo profile, respectively. Green symbols represent other Argo profiles from this float. The thin colors lines indicate the reference data

5

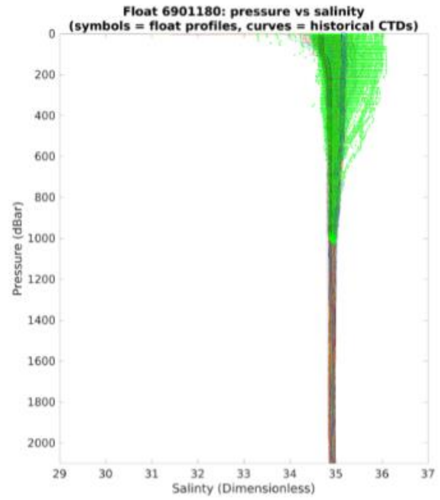

Figure 5: Float 6901180. Salinity (PSS-78) plotted with pressure (dBar) and data from WMO boxes of CTD reference data (CTD for DMQC 2019V01) +/- 10°of latitude and longitude. The black and blue cycles indicates the first and the last Argo profile, respectively. Green symbols represent other Argo profiles from this float. The thin colors lines indicate the reference data.

6

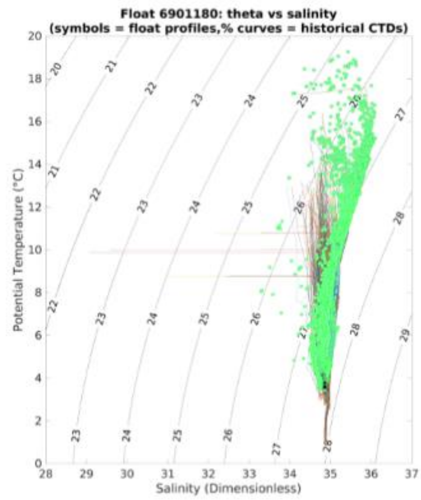

Figure 6: Float 6901180. T/S diagram plotted with and data from WMO boxes of CTD reference data (CTD for DMQC 2019V01) +/- 10°of latitude and longitude. The black and blue cycles indicates the first and the last Argo profile, respectively. Green symbols represent other Argo profiles from this float.

7

3 Correction of Salinity Data

3.1 Comparison between Argo floats and CTD Climatlogy

3.1.1 Configuration

```
% ==============================
%
%   Climatology Data Input Paths
%

HISTORICAL_DIRECTORY=/users/argo/climatology
HISTORICAL_CTD_PREFIX=/historical_ctd/CTD_for_DMQC_2019V01/ctd_
HISTORICAL_BOTTLE_PREFIX=/historical_bot/bot_
HISTORICAL_ARGO_PREFIX=/argo_profiles/ARGO_for_DMQC_2020V01/argo_

% ==============================
%
%   Float Input Path
%

FLOAT_SOURCE_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/float_source/
FLOAT_SOURCE_POSTFIX=.mat

% ==============================
%
%   Mapping Output Path
%

FLOAT_MAPPED_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/float_mapped/ctd/
FLOAT_MAPPED_PREFIX=map_
FLOAT_MAPPED_POSTFIX=.mat

% ==============================
%
%   Calibration Output Path
%

FLOAT_CALIB_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/float_calib/ctd/
FLOAT_CALIB_PREFIX=cal_
FLOAT_CALSERIES_PREFIX=calseries_
FLOAT_CALIB_POSTFIX=.mat

% ==============================
%
%   Diagnostic Plots Output Path
%

FLOAT_PLOTS_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/float_plots/ctd/

% ==============================
%
%   Constants File Path
%

CONFIG_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/constants/
CONFIG_COASTLINES=coastdat.mat
CONFIG_WMO_BOXES=wmo_boxes_ctd.mat
CONFIG_SAF=TypicalProfileAroundSAF.mat
```

8

```
% ================================
%
%    Objective Mapping Parameters
%

% max number of historical casts used in objective mapping
CONFIG_MAX_CASTS=300

% 1=use PV constraint, 0=don't use PV constraint, in objective mapping
MAP_USE_PV=1

% 1=use SAF separation criteria, 0=don't use SAF separation criteria, in objective mapping
MAP_USE_SAF=0

% spatial decorrelation scales, in degrees
MAPSCALE_LONGITUDE_LARGE=2.5
MAPSCALE_LONGITUDE_SMALL=0.8
MAPSCALE_LATITUDE_LARGE=1.5
MAPSCALE_LATITUDE_SMALL=0.5

% cross-isobath scales, dimensionless, see BS(2005)
MAPSCALE_PHI_LARGE=0.3
MAPSCALE_PHI_SMALL=0.08

% temporal decorrelation scale, in years
MAPSCALE_AGE=0.69
MAPSCALE_AGE_LARGE=5

% exclude the top xxx dbar of the water column
MAP_P_EXCLUDE=100

% only use historical data that are within +/- yyy dbar from float data
MAP_P_DELTA=250
```

9

3.1.2 Results

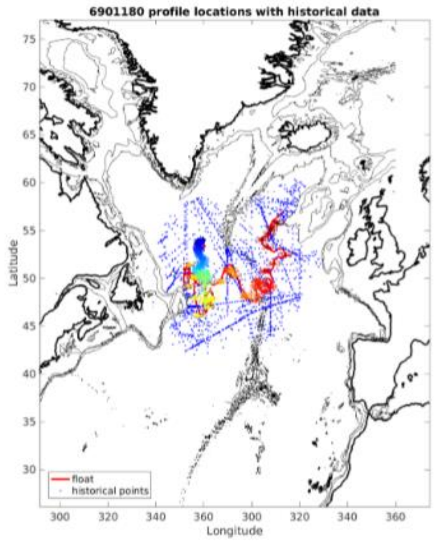

Figure 7: Float 6901180. Location of the float profiles (red line with coloured numbers) and the CTD reference data selected for mapping (blue dots). The black contours indicate the bathymetry at 0, 200, 1000 and 2000 m.

10

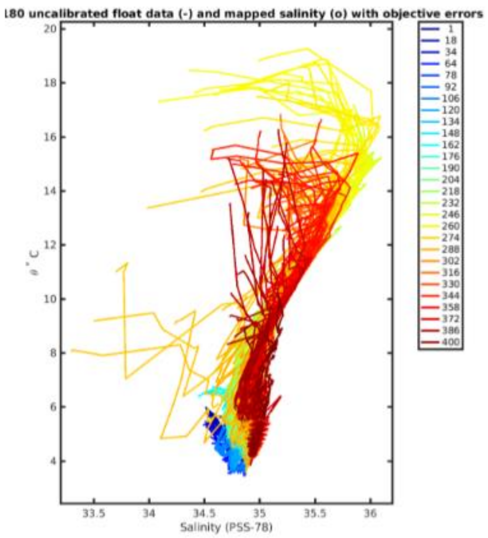

Figure 8: Float 6901180. Plots the original float salinity and the objectively estimated reference salinity at the 10 float theta levels that are used in calibration.

11

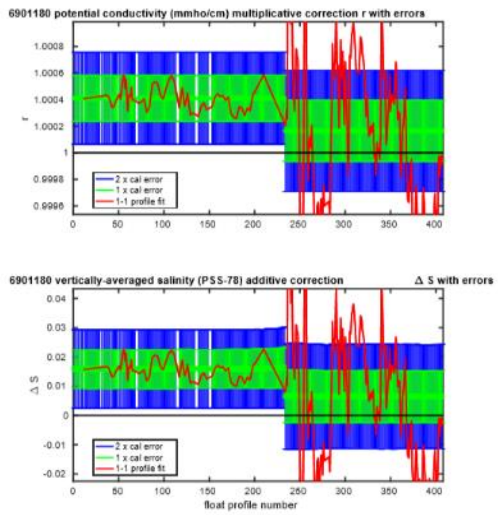

Figure 9: Float 6901180. Evolution of the suggested adjustment with time. The top panel plots the potential conductivity multiplicative adjustment. The bottom panel plots the equivalent salinity additive adjustment. The red line denotes one-to-one profile fit that uses the vertically weighted mean of each profile. The red line can be used to check for anomalous profiles relative to the optimal fit.

12

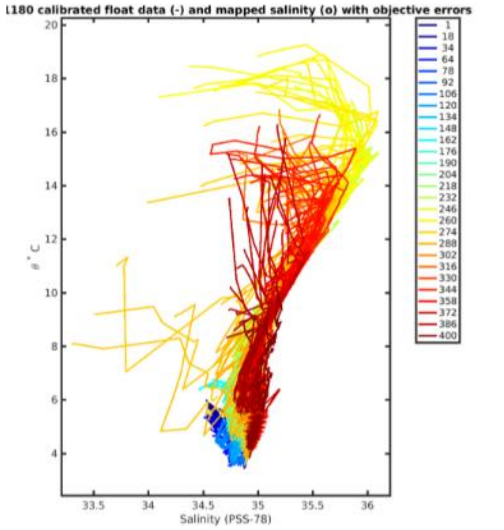

Figure 10: Float 6901180. Plots of calibrated float salinity and the objectively estimated reference salinity at the 10 float theta levels that are used in calibration.

13

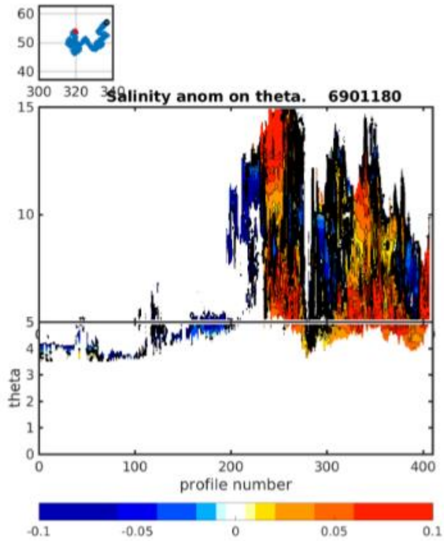

Figure 11: Float 6901180. Salinity anomaly on theta levels.

14

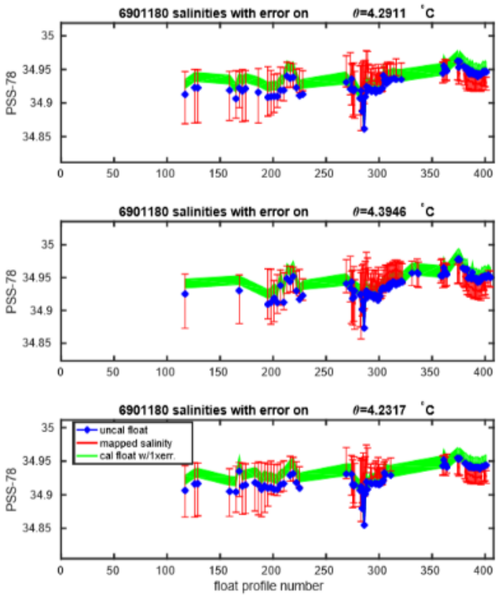

Figure 12: Float 6901180. Plots of the evolution of salinity with time along with selected theta levels with minimum salinity variance.

15

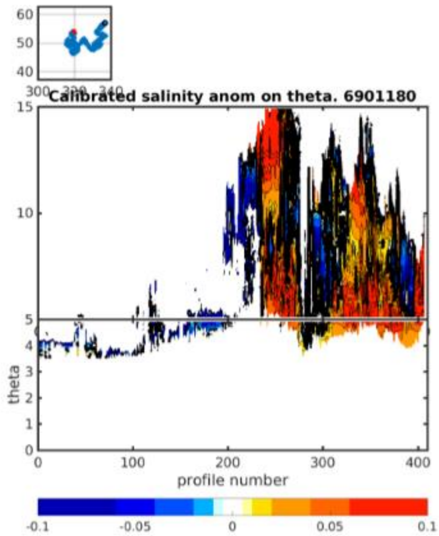

Figure 13: Float 6901180. Calibrated salinity anomaly on theta levels.

16

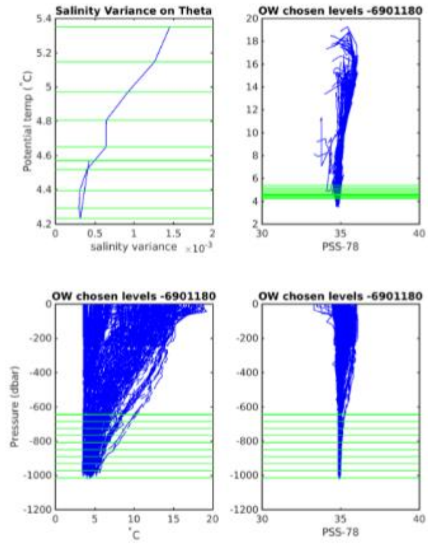

Figure 14: Float 6901180. Plots including the theta levels chosen for calibration: Top left: Salinity variance at theta levels. Top right: T/S diagram of all profiles of Argo float. Bottom left: potential temperature plotted against pressure. Bottom right: salinity plotted against pressure.

17

3.2 Comparison between Argo floats and Argo Climatlogy

3.2.1 Configuration

```
% ==============================
%
%    Climatology Data Input Paths
%

HISTORICAL_DIRECTORY=/users/argo/climatology
HISTORICAL_CTD_PREFIX=/historical_ctd/CTD_for_DMQC_2019V01/ctd_
HISTORICAL_BOTTLE_PREFIX=/historical_bot/bot_
HISTORICAL_ARGO_PREFIX=/argo_profiles/ARGO_for_DMQC_2020V01/argo_

% ==============================
%
%    Float Input Path
%

FLOAT_SOURCE_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/float_source/
FLOAT_SOURCE_POSTFIX=.mat

% ==============================
%
%    Mapping Output Path
%

FLOAT_MAPPED_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/float_mapped/argo/
FLOAT_MAPPED_PREFIX=map_
FLOAT_MAPPED_POSTFIX=.mat

% ==============================
%
%    Calibration Output Path
%

FLOAT_CALIB_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/float_calib/argo/
FLOAT_CALIB_PREFIX=cal_
FLOAT_CALSERIES_PREFIX=calseries_
FLOAT_CALIB_POSTFIX=.mat

% ==============================
%
%    Diagnostic Plots Output Path
%

FLOAT_PLOTS_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/float_plots/argo/

% ==============================
%
%    Constants File Path
%

CONFIG_DIRECTORY=/users/argo/ow/matlabow-2.0.1/data/constants/
CONFIG_COASTLINES=coastdat.mat
CONFIG_WMO_BOXES=wmo_boxes_argo.mat
CONFIG_SAF=TypicalProfileAroundSAF.mat

% ==============================
```

18

```
%
%    Objective Mapping Parameters
%

% max number of historical casts used in objective mapping
CONFIG_MAX_CASTS=300

% 1=use PV constraint, 0=don't use PV constraint, in objective mapping
MAP_USE_PV=1

% 1=use SAF separation criteria, 0=don't use SAF separation criteria, in objective mapping
MAP_USE_SAF=0

% spatial decorrelation scales, in degrees
MAPSCALE_LONGITUDE_LARGE=2.5
MAPSCALE_LONGITUDE_SMALL=0.8
MAPSCALE_LATITUDE_LARGE=1.5
MAPSCALE_LATITUDE_SMALL=0.5

% cross-isobath scales, dimensionless, see BS(2005)
MAPSCALE_PHI_LARGE=0.3
MAPSCALE_PHI_SMALL=0.08

% temporal decorrelation scale, in years
MAPSCALE_AGE=0.69
MAPSCALE_AGE_LARGE=5

% exclude the top xxx dbar of the water column
MAP_P_EXCLUDE=100

% only use historical data that are within +/- yyy dbar from float data
MAP_P_DELTA=250
```

19

3.2.2 Results

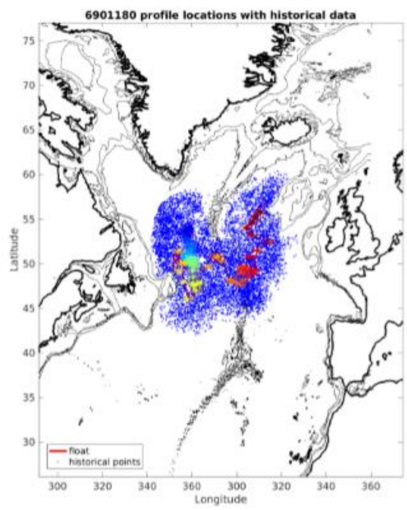

Figure 15: Float 6901180. Location of the float profiles (red line with coloured numbers) and the CTD reference data selected for mapping (blue dots). The black contours indicate the bathymetry at 0, 200, 1000 and 2000 m.

20

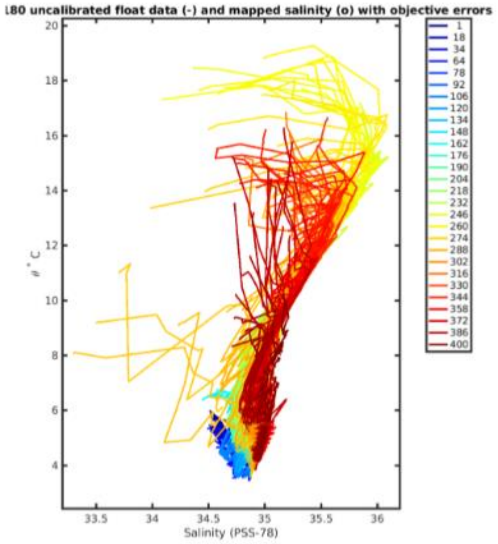

Figure 16: Float 6901180. Plots the original float salinity and the objectively estimated reference salinity at the 10 float theta levels that are used in calibration.

21

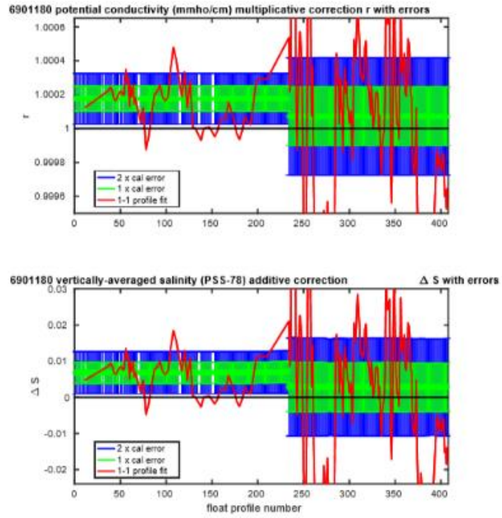

Figure 17: Float 6901180. Evolution of the suggested adjustment with time. The top panel plots the potential conductivity multiplicative adjustment. The bottom panel plots the equivalent salinity additive adjustment. The red line denotes one-to-one profile fit that uses the vertically weighted mean of each profile. The red line can be used to check for anomalous profiles relative to the optimal fit.

22

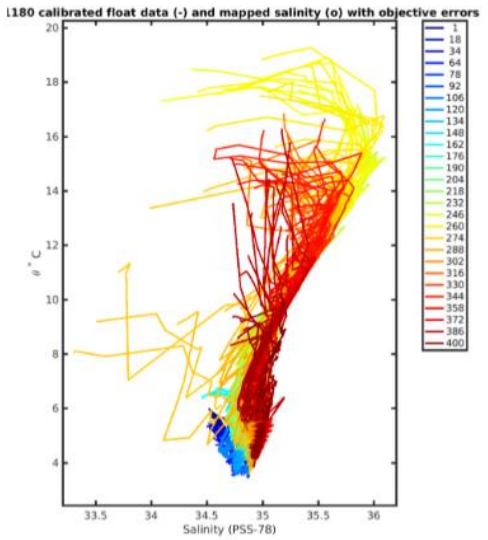

Figure 18: Float 6901180. Plots of calibrated float salinity and the objectively estimated reference salinity at the 10 float theta levels that are used in calibration.

23

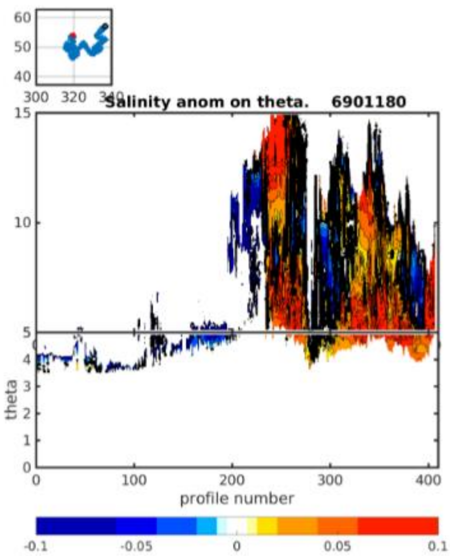

Figure 19: Float 6901180. Salinity anomaly on theta levels.

24

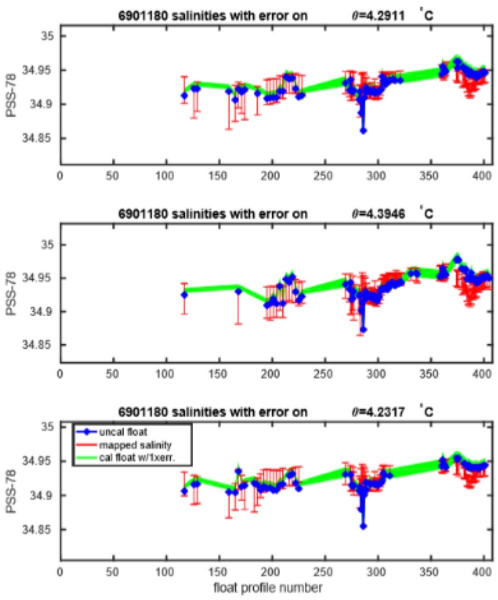

Figure 20: Float 6901180. Plots of the evolution of salinity with time along with selected theta levels with minimum salinity variance.

25

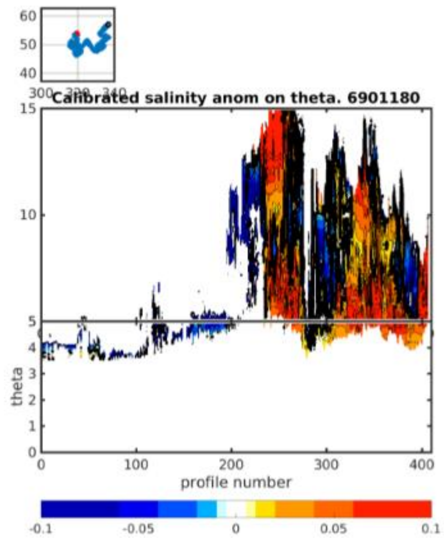

Figure 21: Float 6901180. Calibrated salinity anomaly on theta levels.

26

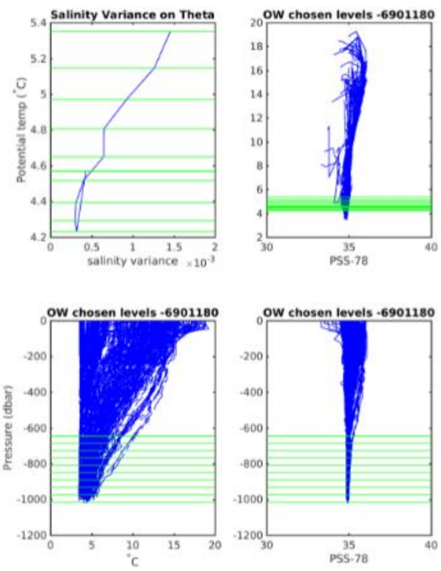

Figure 22: Float 6901180. SPlots including the theta levels chosen for calibration: Top left: Salinity variance at theta levels. Top right: T/S diagram of all profiles of Argo float. Bottom left: potential temperature plotted against pressure. Bottom right: salinity plotted against pressure.

27

3.3 Summary and Conclusions

E.g." The float was deployed in the Newfoundland Basin, where further was travelling along the Subpolar gyre rim, across the Mid-Atlantic Ridge toward the Labrador Sea.The analysis has been conducted using both CTD and Argo reference data.

The configuration of the objective mapping parameters, set for this float, were applied separately for CTD and Argo reference data. In set calseries the time series has been separated onto two periods due to different sampling: from profile 1 to 276 data were collected daily and from 277 to 407 data were samples every 10 days. Additionally, we constrained the theta levels to be selected below 600 m due to strong seasonal signal of Atlantic Water in the upper layer.

For the first 277 profiles flat was located in relatively homogenic layer of water masses, whith salinity error between the climatology and Argo float of 0.01(based on Argo ref. data). Further after crossing Mid-Atlantic Ridge float showed much stronger variability with a salinity difference exceeding 0.02. The variability is a result of relatively shallow water ¡1000 taken to analysis and strong natural variability in the subpolar north Atlantic. The salinity error for was around 0.014. This error was assigned to salinity data of the entire float."

28