Review of the current state of the delayed-mode quality control tools and methods used within the global Argo community

Ref.: Subpart_D2.5

Date: 19/11/2019

Euro-Argo Research Infrastructure Sustainability and Enhancement Project (EA RISE Project) - 824131

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 824131.
Call INFRADEV-03-2018-2019: Individual support to ESFRI and other world-class research infrastructures

Disclaimer:

This Deliverable reflects only the author's views and the European Commission is not responsible for any use that may be made of the information contained therein.

Document Reference

Project	**Euro-Argo RISE - 824131**
Deliverable number	**This report is a subpart of D2.5**
Deliverable title	**Review of the current state of delayed-mode quality control tools and methods used within the global Argo community**
Description	Review and summary of variety of tools and procedures used by the global Argo community in delayed mode quality control. Assessment of the current state and future improvements require to increase the consistency of the Argo delayed-mode dataset and enhance the efficiency and sustainability of the delayed-mode quality control procedure based on global delayed-mode operator community feedback.
Work Package number	**2**
Work Package title	**Improvement of the Core Argo Mission**
Lead Institute	**British Oceanographic Data Centre, National Oceanographic Data Centre**
Lead authors	**Kamila Walicka, Matt Donnelly**
Contributors	**DMQC operators from 19 international organisations involved in DMQC**
Submission date	**Click here to enter text.**
Due date	**19/11/2019**
Comments	**Click here to enter text.**
Accepted by	**Guillaume Maze**

Document History

Version	Issue Date	Author	Comments
v0.1	11/11/2019	Kamila Walicka	Preparation of report draft
v0.2	14/11/2019	Matt Donnely	Document review
V0.3	18/11/2019	Kamila Walicka	Further revisions

EXECUTIVE SUMMARY

This report summarises a variety of delayed mode quality control (DMQC) tools and procedures used by the global Argo community. The aspiration of this project is to improve the sustainability of DMQC within the global Argo community, through effective tools. A survey was undertaken on aspects of the DMQC processing: operators, software, the shared reference data repository, quality control procedures and generation of DMQC reports for future reference. This report highlights the main opportunities for improvements reported by the operators in processing Argo float in delayed mode, to enhance the sustainability and efficiency of Argo dataset and DMQC procedures.

Responses to the survey were received from 19 international organisations involved in DMQC processing. The survey recorded 39 DMQC operators including well-experienced Argo members and a rising number of new contributors. The survey shows a relatively low number of operators per organisation, mostly around 1-3 DMQC operators. The most common operating system was Windows and Linux. The organisations use a wide range of MATLAB versions for running the analysis code, often associated with high costs of software licences and its toolboxes. The MATLAB codes currently used in the DMQC analysis are not uniform across the entire Argo community. Most organisations have developed their own version of Graphical User Interface (GUI) to the DMQC analysis, with various improvements. Moreover, operators use a number of additional checks and databases to verify the quality of Argo float data. Only 50% of operators are using the latest version of DMQC analysis software. There is a broad range of CTD and Argo reference data versions used to salinity drift assessments, where the majority of the community use the most updated data from 2018. The is a lack of uniform software to generate DMQC reports, resulting in different details and focus in the report.

The recommendations suggested by the DMQC operators are to provide routinely trainings for current and new Argo DMQC operators, implement a free programming language for DMQC analysis, introduce an uniform and best extensive DMQC GUI and other common tools for the entire Argo community, develop a reference database for the shallow seas, improve flexibility in selecting data from the currently available reference database, include the available metadata covering the CTD reference data used, introduce a sharing repository with stored list and reports of problematic floats, consider the other available opportunities to improve methods of the DMQC process.

TABLE OF CONTENT

1 Introduction 5
2 Methods 6
3 Results 7
3.1 DMQC Operators 7
3.2 DMQC software 9
3.2.1 Operating Systems 9
3.2.2 MATLAB Versions 9
3.3 Data repository of shared codes 10
3.4 Quality Check Procedures 10
3.4.1 Software to Adjust Flags on Data 10
3.4.2 Additional Quality Checks 11
3.4.3 Initial Adjustments 12
3.4.4 Salinity drift assessment 13
3.5 Final DMQC Reports 15
4 Suggestions and Recommendations 17
5 Conclusions 19
6 References 20

1 Introduction

The aim of this report is to investigate the accessibility and usage of available DMQC tools within the Argo community from a range of organisations, to provide an understanding of the current state of development and practice. The questionnaire covers matters such as a number of DMQC operators, type of the operating software used, the version of MATLAB software, types of software used to adjust flags on data, the level of implementing the initial adjustments to float data, versions of DMQC software and referenced database used and software to generate DMQC reports. Moreover, this document highlights the issues reported by the global DMQC operators associated with using the DMQC software and also their suggestions to improve the DMQC procedures.

The report will be used as a basis for all partners to create and agree on the framework for collaborative DMQC development.

2 Methods

The survey about the currently available DMQC tools has been performed through a questionnaire (Appendix A). All data obtained from questionnaires are stored securely on BODC's server and will only be used within the framework of the Euro-Argo RISE project. Access to the answers is restricted to the responsible persons, Kamila Walicka and Matt Donnelly.

The method of answering the questionnaire was voluntary. The answers were analysed offline. The questionnaire technical results and conclusions deduced from the results can be further published within the Euro-Argo RISE project deliverables, reports and documentation, however, no personal information will be published in any form. All questionnaire answers will be deleted latest at the end of the Euro-Argo RISE project.

The questionnaire was sent to all known DMQC operators via the DMQC operator list extracted from jcommops.org website and also sent to the e-mail list (argo-dm-dm@jcommops.org). This email list is commonly used to communicate between the operators.

3 Results

Among the organisations involved in the DMQC process, 19 distinct organisations took part in the survey, however, this was reduced to 16 functioning DMQC groups. The reduction of a number of organisations involved in the survey was motivated by the presence of a few associated organisations that are working together, while the DMQC software is run through one organisation.

The result section consists of statistics constructed based on the questionnaire responses covering issues associated with the DMQC operators, DMQC software, data repository, quality control procedures and final DMQC reports.

3.1 DMQC Operators

The first parameter considered in this survey was the number of DMQC operators per organisation. Figure 1 shows a relatively low number of operators per organisation. In six organisations there was only one DMQC operator, while only one organisation has five operators. Moreover, in four organisations we found four operators, however, among these operators, there were also inactive members of the team. These results showed a relatively low number of DMQC operators per organisations that may lead to complications in terms of the succession planning and knowledge exchange as team composition changes.

Additionally, this survey showed that some operators revealed worries in terms of the experience and the level of subjective decisions made during the DMQC processing such as in applying flags to data and choosing the number of profile where the correction will be applied. This survey showed that there are 39 DMQC operators within the global Argo community. Most of the DMQC operators have a background related to physical oceanography, often including a doctorate degree. Among these 39 operators, four operators declared that they have less than one year of experience in DMQC process, eight operators had less than six years experience, 13 operators had more than six years experience and the remaining 14 operators did not provide an answer for this question (Figure 2). This result gives confidence of well-qualified operators working with Argo floats. However, this statistic also revealed a relatively large number of operators that require some support in DMQC analysis and further training. The experience and background of operators is a key element in terms of consistency of scientific decision made in analysing the Argo floats in delayed mode.

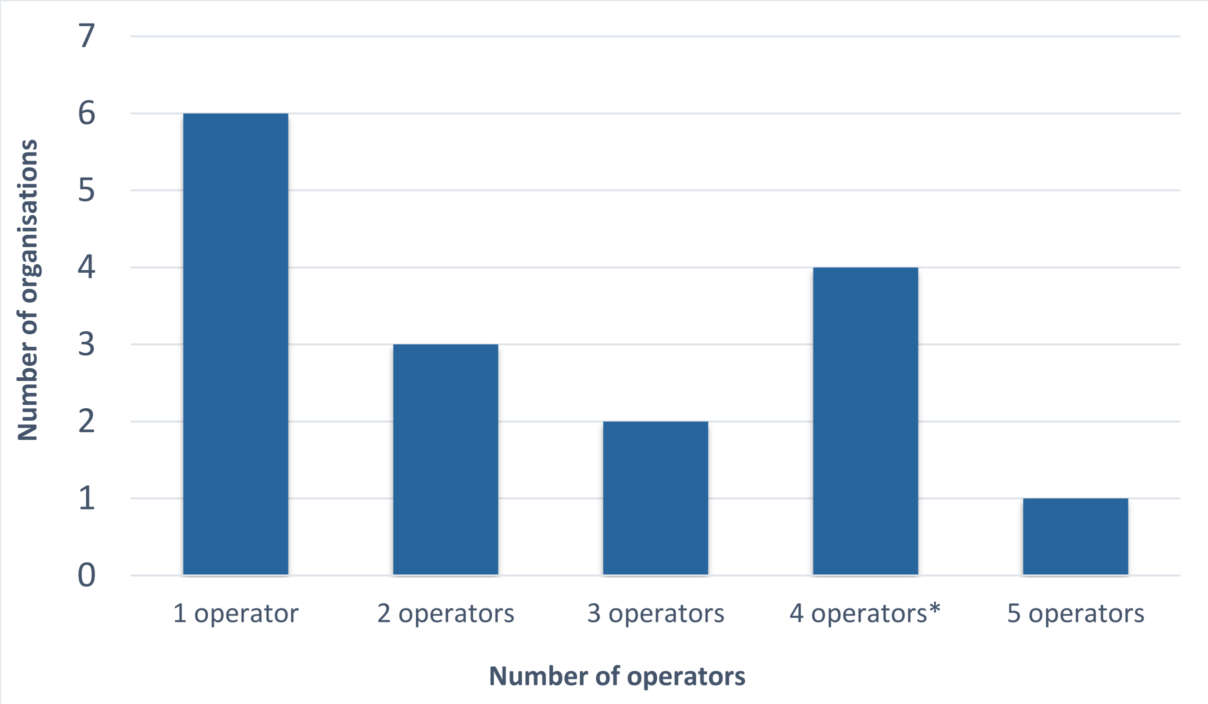

Figure 1: Number of recorded DMQC operators within an organisation. The asterix (*) informs that statistics include also former DMQC operators.

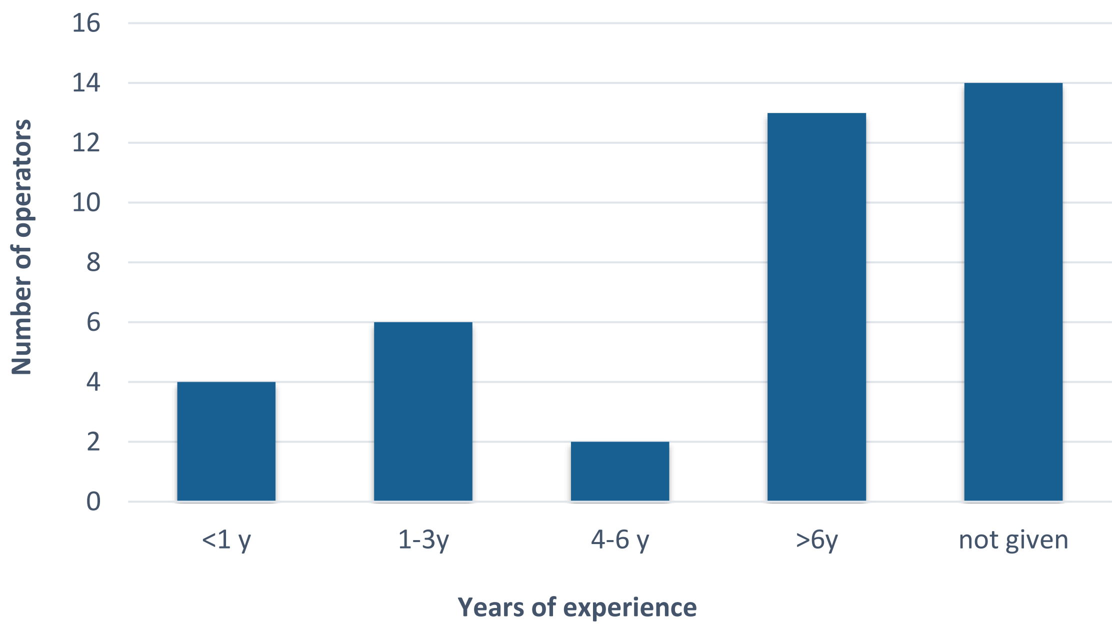

Figure 2: Number of years of experience in DMQC.

3.2 DMQC software

3.2.1 Operating Systems

The survey reveals that DMQC operators were using various operating systems and in a few cases, were using more than one operating software in their organisation. These statistics include all operating software used in organisations. Figure 3 shows that in most cases DMQC operators used Windows (seven organisations) and Linux software (six organisations). A minority of organisations worked with MacOS (three organisations) and Unix (one organisation), respectively.

The type of operating system may have a significant impact e.g. on the speed of DMQC software execution. Moreover, the use of multiple operating systems at the same time may influence the infrastructure and functionality of DMQC procedures.

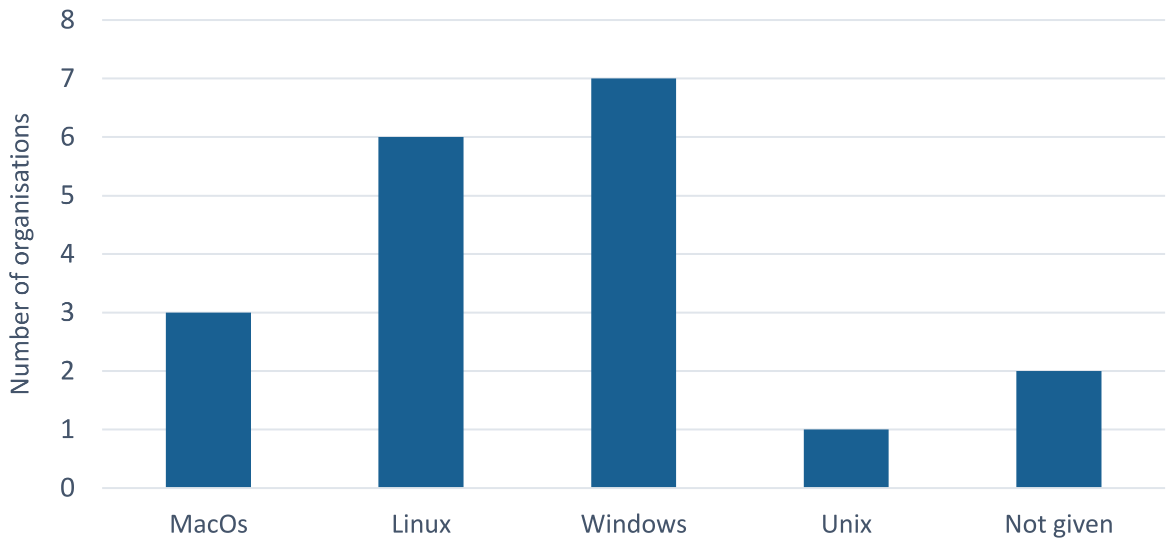

Figure 3: Types of operating systems used by organisations. There was more than one operating system used per organisation.

3.2.2 MATLAB Versions

This survey found that operators are using more that one MATLAB version in their DMQC processing. These statistics include all types of MATLAB versions used within organisations. Figure 4 shows that DMQC organisations used a few different MATLAB versions. The oldest reported version came from 2006, while the newest R2019 version was used only in three organisations. The most common MATLAB version was R2014.

Additionally, the operators reported that the use of older versions of MATLAB and their toolboxes may lead to some differences in the OW and OWC results. This is due to minor variations in previous versions of MATLAB and their toolboxes reflect different function behaviours. This finding highlights a need to improve the consistency in the approach of DMQC, through improved software sustainability.

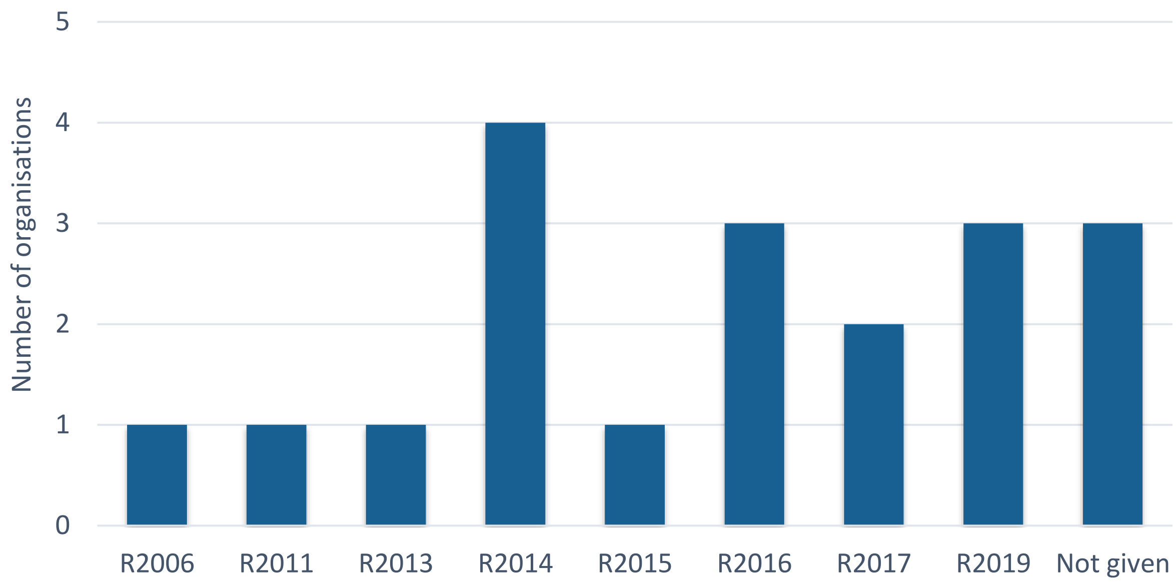

Figure 4: Types of MATLAB version used in organisations. There was more than one MATLAB version used per organisation.

3.3 Data repository of shared codes

A primary objective of the DMQC process is to assess the presence and magnitude or offset in the salinity drift in Argo float, including the estimate of the associated error. The survey found two DMQC software used within the Argo community (Section 3.4.4.1), *OW version 1.1*, created based on Owens & Wong (2009) and *OW version 2* includes further modifications proposed by Cabanes et al. (2016). The OW version 2 includes two releases 2.0 and 2.0.1. These releases include the same major code with some further minor fixes and additional expansion of toolboxes written in MATLAB. In this survey, these two released of OW version 2 were classified as one called *OWC*. Recently (October 2019), there has been released another version of OWC software- 2.1. Both OW and OWC versions of the software are freely available on a GitHub repository and can be found in https://github.com/ArgoDMQC/matlab_owc/releases.

3.4 Quality Check Procedures

3.4.1 Software to Adjust Flags on Data

Adjusting flags is one of the first steps before the DMQC process, where an operator undertakes a visual quality control of profiles and applies flags to suspicious or problematic profiles. The survey showed a broad range of tools used by operators. Figure 5 shows that the majority of organisations adjusted flags utilizing their own in-house GUIs using MATLAB software. Additionally, the survey showed that organisations are also using other tools such as :

- **Ocean Data View** (https://odv.awi.de/),
- in-house codes in the **Python** language,
- **Scoop** (https://www.seanoe.org/data/00374/48531/) and

- **Edserplo** (https://www.bodc.ac.uk/submit_data/what_do_we_do_with_your_data/software_engineering/edserplo/).

All of these tools allow to adjust flags on data, however, there are varying modes of operation and capabilities. For instance, some of them provide an easy way to look at data from the entire float dataset and improve the speed of the quality check. While other tools provide additional data from a climatology that support the decision-making process.

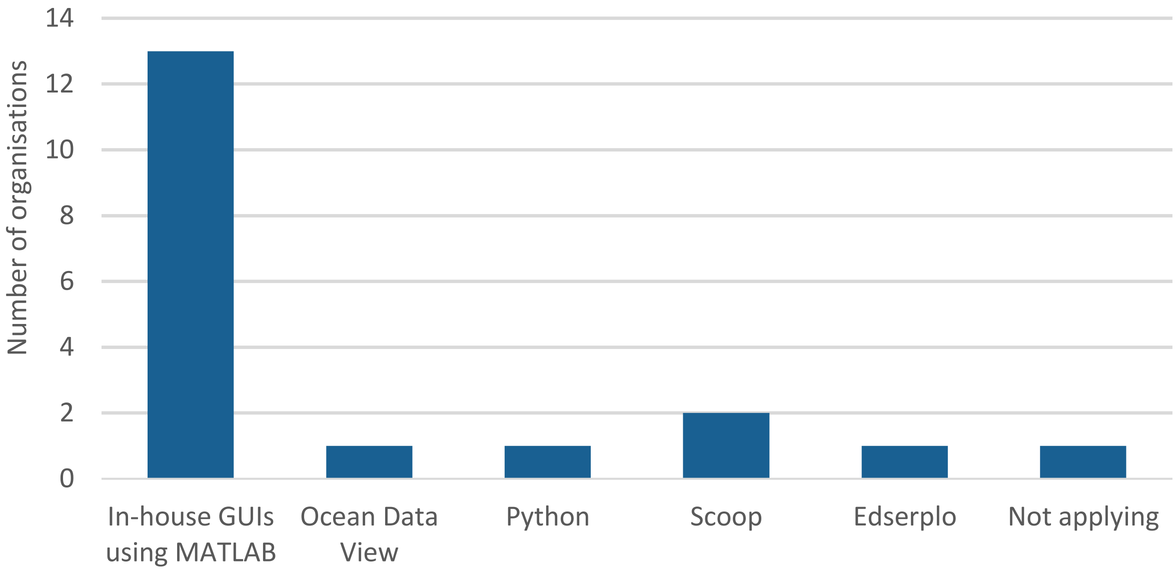

Figure 5: Types of software used to adjust flags on data in organisations. More than one type of software was used per organisation.

3.4.2 Additional Quality Checks

The analysis of the DMQC procedures shows that operators were using different approaches in checking Argo data before using the OW/OWC software. Some operators declared only visual inspection of data, to identify any abnormal values, spikes or other problems with the quality of data. Other operators developed their own GUI which encapsulates data validation routines against criteria set out in the Argo User Manual, but also a set of additional checks.

The additional quality checks often used by DMQC operators are:

- Assessing the quality of pressure, temperature and salinity profiles using the nearby Argo floats (in the real-time), comparison to ship-based CTD on deployment were available, comparing Argo float to data from different climatologies (e.g. CARS 2006, CARS 2009 and Gouretski and Koltermann, WOA2005).
- Density inversion checks.
- In one organisation the thresholds for the range check, density inversion check, and sensor drift check are set more strictly than the ADMT real-time QC. The thresholds are set with reference to the sea region in which float was deployed.

- One organisation mentioned the use of an in-house tool to manually correct salinity on a profile for floats affected by biofouling or other non-systematic salinity offsets or drifts.
- Verifying the consistency between the Argo float data and data from satellite altimetry through the altimetry warnings for float drift generated by Ifremer.
- Routine checks on general parameters of float (number of cycles, float coordinates, time), metadata, technical parameters such as battery voltage.
- One organisation is using an additional procedure to DMQC, based on the artificial neural networks, to identifying anomalous temperature and salinity from Argo floats.
- One organisation has introduced the method to improve the estimation of correlation scales.

3.4.3 Initial Adjustments

In terms of the initial adjustments, the survey investigated the source and methods of pressure corrections and cell thermal mass corrections applied to core Argo floats. The pressure adjustments have to be applied to those floats where the pressure sensor is not auto-corrected to zero while at the sea surface, such as the Teledyne Webb Research Apex floats. The pressure data from these floats then require calibrations during processing in DMQC.

Figure 6 shows that in most organisations the sea surface pressure was inspected visually and any corrections were applied using in-house GUIs. The corrections were calculated based on the methods specified in the Argo manual (https://archimer.ifremer.fr/doc/00228/33951/32470.pdf). Furthermore, there were two organisations who were not applying the corrections. That is mostly because they are using floats where the corrections are applied onboard automatically or they found some difficulties to apply these corrections.

A similar result was obtained in applying the cell thermal mass corrections, these corrections were also mostly applied using various in-house GUI's using MATLAB. In the questionnaire, most of the surveyed did not provide an answer about the source or method used to apply the corrections. Only two organisations declared that they implemented the method from Johnson et al. (2007). The study found four organisations who did not apply cell thermal mass corrections. One organisation reported that most of their fleet is composed of Apex float and they found difficulties with applying this kind of correction.

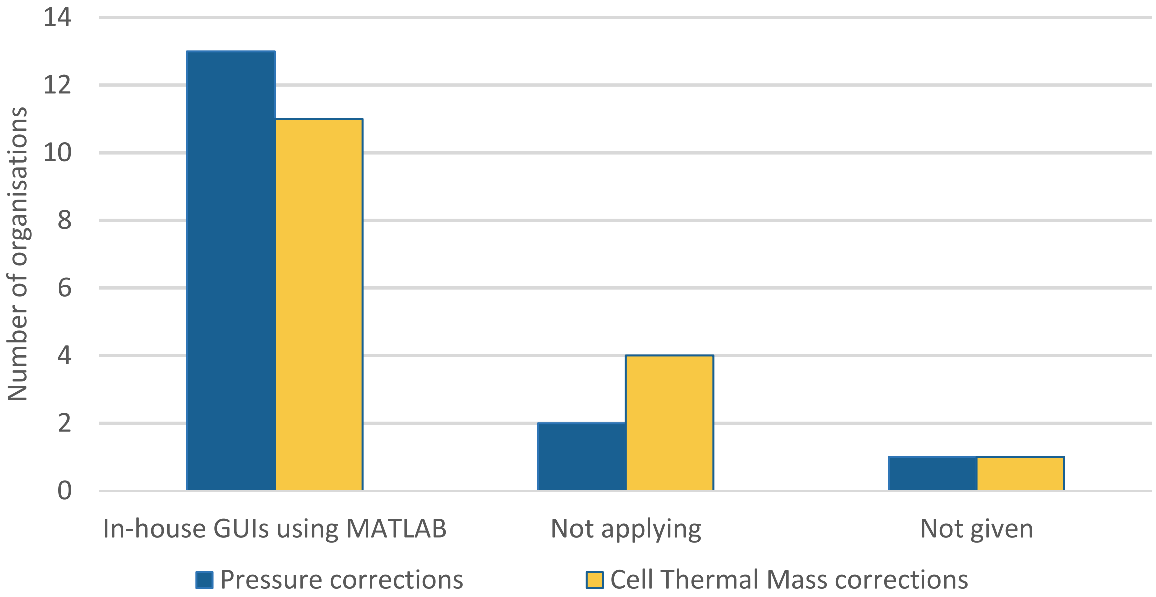

Figure 6: Initial Adjustments (sea surface pressure corrections and cell thermal mass corrections) applied to data.

3.4.4 Salinity drift assessment

3.4.4.1 Software to Salinity Drift Assesment

The survey found two major DMQC software used within the Argo community, OW and OWC version (Section 3.3). Additionally, some organisations are using both OW and OWC software. These statistics include all responses from each organisation. Figure 7 shows that in 50 % answers the organisations were using OWC version, while in 44 % of answers organisations were still working with OW software. Additionally, this group includes a few organisations who intend to implement the OWC version in the nearest future. Overall, for some of the ocean regions, the use of these two versions of DMQC software may lead to some differences in the results of processed data.

A few DMQC operators have also reported the issue in the efficiency of working with this software. They experienced a relatively long duration of OWC code execution that may lead to a decrease of a number of analysed and calibrated Argos available in delayed mode.

Additionally, one organisation reported that they are not running the OWC software for every float. Instead, the decision on using OWC is made based on the comparison of the float data from other databases. The OWC software is used only on some clear evidence of salinity drift or offset. That may lead to increase the subjective decisions DMQC processing.

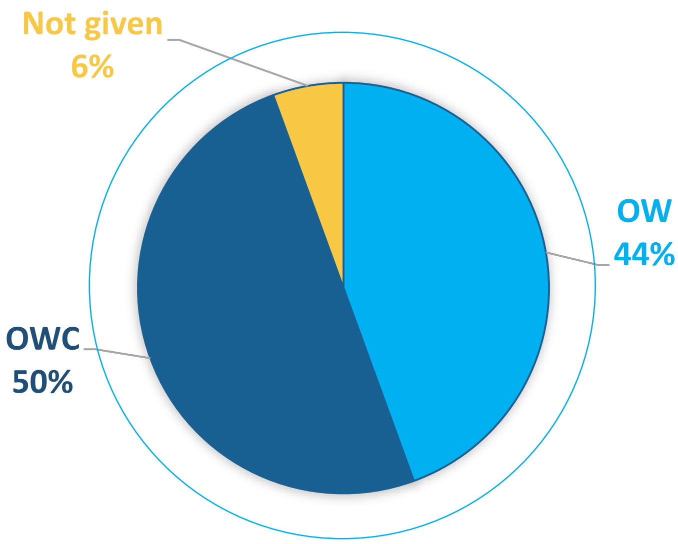

Figure 7: Types of DMQC software to salinity drift assessment. There was more than one software used per organisation.

3.4.4.2 CTD and Argo Reference Data

The major source of CTD and Argo reference data used by operators is data distributed by Ifremer via the FTP site at /coriolis/data/DMQC-ARGO/, with statistics made for CTD, Argo and other databases, respectively. In terms of the CTD data, in 2018 there was two releases of these data, version 01 and 02. For the same year, there was only one release of Argo reference data. The survey shows a few different versions of reference data used by organisations. Figure 8 shows that the most common CTD reference data used by organisations came from 2018. Where eight organisations used version 2018v02 and four organisations used 2018v01, respectively. However, there was also one organisation that used solely some other databases for the specific seas.

In terms of Argo reference data, there is a broader range of reference data used. The oldest versions of Argo reference data came from 2015 (one organisation) and 2017 (one organisation). The majority of operators, however, used data from 2018. Additionally, it has been found that one organisation is not using Argo reference data et all.

Overall, these results showed a wide range of both CTD and Argo the reference data used in the DMQC analysis. The use of the most updated version of reference data should give the most appropriate estimates of the salinity drifts and associated error.

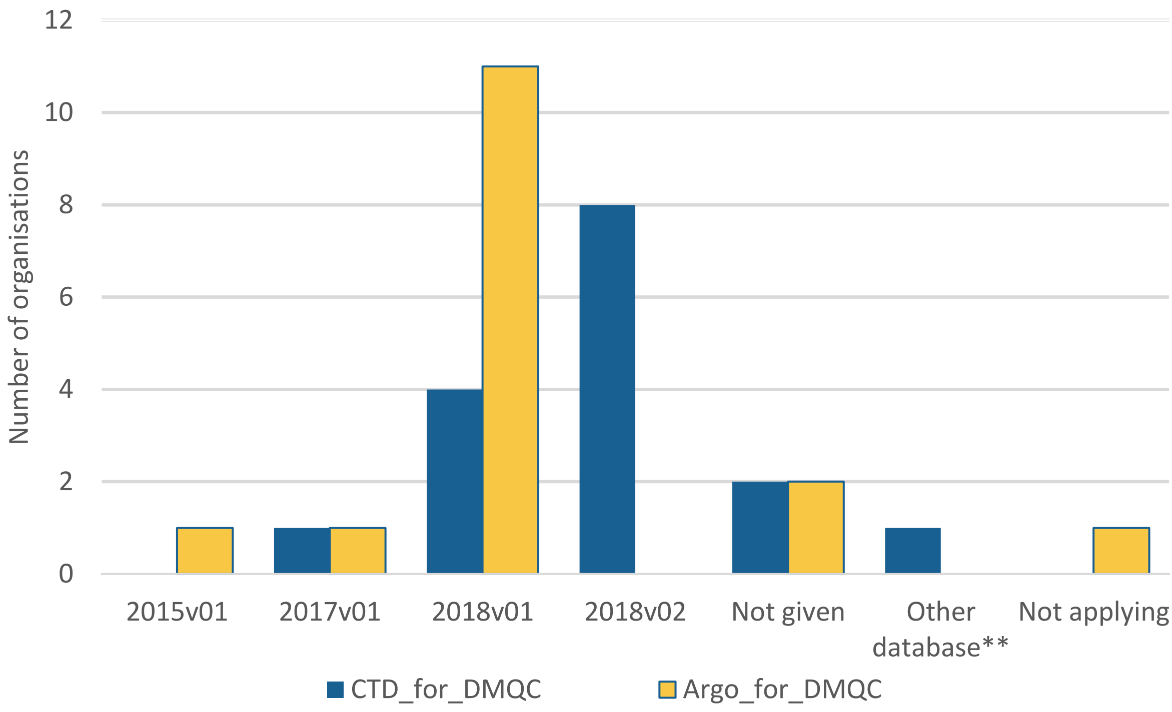

Figure 8: Versions of reference data, CTD and Argo data respectively, used by organisations. Double asterix (**) relates to other databases using a solely regional database for the specific seas.

3.5 Final DMQC Reports

The final DMQC report should include the information about the DMQC processing allowing the reproducibility of the analysis and justification of the decision made to calibrate and correct the Argo float. This survey found that among the organisations, operators used more that one type of software to generate DMQC reports. In order to investigate all types of software, these statistics include all responses. Figure 9 shows that 10 organisations used the family of MS office software (Word or Excel) or other word processors (Hangul) to generate the DMQC reports. Another five organisations kept their reports on online websites. The Latex software was used by four organisations. The survey also finds that two organisations saved the DMQC processing information in the in-house database.

Overall, these results showed a large diversity in software used to create the final DMQC reports. Moreover, we found that there is also a lack of guidelines and requirements to create that report.

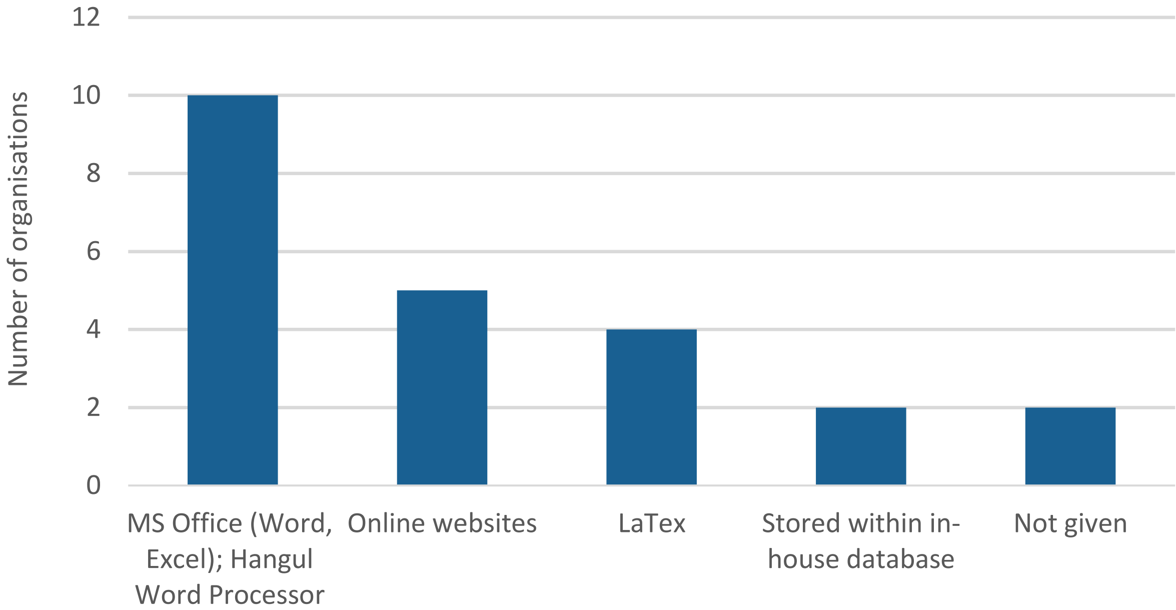

Figure 9: Types of software used to generate DMQC report. There was more than one software used per organisation.

4 Suggestions and Recommendations

This section presents a set of recommendations that were suggested by global Argo DMQC operators to improve the sustainability of DMQC tools and procedures.

DMQC training

In order to improve the consistency of the DMQC analysis, many DMQC operators suggested frequent DMQC training courses with the case studies to illustrate real scenarios in processing the problematic floats. The training courses are needed for the new members of the Argo delayed mode group but also for the current operators to improve the awareness about the newest available tools, processing methods, software and fine-tune skills.

Free programming language

One of the key responses from this survey is the relatively old versions of MATLAB used by the operators (Section 3.2.2) linked to the very high costs of MATLAB licences and its toolboxes. Many organisations noted that they are struggling with these expenses. We received many responses where operators have suggested a need to use other free open-source programming languages e.g. Python, to implement the DMQC analysis.

Unification of the graphical user interface

Another recommendation from this survey is the implementation of a uniform GUI, including all elements of DMQC procedures. Currently, many organisations have to create their own GUI, resulting in non-uniform MATLAB codes across the Argo community. Some operators already highlighted problems in their currently unstable GUI and difficulties in creating the MATLAB codes such as

- generate OW matrix from real-time netCDF format to MATLAB matrix,
- apply the initial adjustments to floats (cell thermal mass correction and sea surface pressure corrections). This resulted that some organisation was not applying e.g. cell thermal mass correction to float data (Section 3.4.3).
- generate the netCDF D-files with corrections suggested from OW/OWC method.

Overall, the lack of uniform commonly shared MATLAB codes for all DMQC processing may lead to inconsistency of DMQC procedures, which can lead to inconsistency of generated results.

Reference database

The survey showed that DMQC operators also reflect a need for more information about the metadata of CTD historical data. The currently distributed reference data does not include the information about e.g. the number of profiles used from each source, for each cycle. The suggestion is to make these data available for DMQC operators.

Another idea related to the reference data is to improve the selection of reference data from the specific regions. This will allow a more flexible selection of relevant area of analysis such as data inside region characterised and with specific water mass properties. This improvement will lead to a decrease in the variability of salinity error estimates in some region and more effective assessments of salinity drift and offsets.

Some operators who are working with floats from the shallow seas, below 800 m, recommended the works associated with the creation of the reference database specifically for these shallow region including both CTD and Argo data. These regions include a valuable source of information that is not captured by the currently available reference datasets.

Additionally, DMQC operators reported a limited capability to acquire CTD ship-based data from some organisations that collect data in specific regions. The suggestion is to verify and update the project licences to encourage PIs to share the CTD ship-based data within a certain period.

Repository for sharing float analysis

Another improvement suggested by operators is a need for a repository for sharing case studies and reports of interesting for analysis and difficult floats, with a set of guidelines or common tool for selecting the most urgent floats. This repository will allow a quick alerting and verification of problematic floats and improve the consistency of decision made within the Argo DMQC community.

Improvements in DMQC methods

Some organisations who are working with data from the specific regions often have to apply some additional methods to improve their analysis. This survey reveals two potential improvements suggested by the operators: (1) improvements in the estimation of correlation scales (Gradient-Dependent Correlation Scale Method), (Chunling Z. et al. 2012) and (2) Artificial Neural Networks for performing DMQC, to identifying anomalous temperature and salinity from Argo floats (Bhaskar T.V.S.Udaja, peer review).

5 Conclusions

There is an issue with the sustainability of core Argo DMQC, and this problem will only grow with the development of deep and biogeochemical DMQC proceedures.

To improve the overall consistency of the Argo delayed-mode dataset, there is a need to enhance the overall consistency of the DMQC tools, methods and operator skills development

There is a need to improve the efficiency of the core DMQC process, with different parts of the community facing different challenges in this respect

Do we need a technical road map for international collaboration on software development for DMQC?

The independent improvements of the DMQC methods require a revision and the most appropriate solutions needs to be uniformly implemented in the international organisations involved in the DMQC processing. Any further suggested improvements should be reported in the shared repository and frequently reviewed.

The international DMQC operators are obligated to follow the most updated DMQC procedures, implement to their internal systems the newest software updates and use the most recent release of reference data for the DMQC analysis.

6 References

Cabanes, C., Thierry, V., & Lagadec, C. (2016). Improvement of bias detection in Argo float conductivity sensors and its application in the North Atlantic. *Deep-Sea Research Part I: Oceanographic Research Papers*, *114*, 128–136. https://doi.org/10.1016/j.dsr.2016.05.007

Johnson, G. C., Toole, J. M., & Larson, N. G. (2007). Sensor corrections for Sea-Bird SBE-41CP and SBE-41 CTDs. *Journal of Atmospheric and Oceanic Technology*, *24*(6), 1117–1130. https://doi.org/10.1175/JTECH2016.1

Owens, W. B., & Wong, A. P. S. (2009). An improved calibration method for the drift of the conductivity sensor on autonomous CTD profiling floats by θ-S climatology. *Deep-Sea Research Part I: Oceanographic Research Papers*, *56*(3), 450–457. https://doi.org/10.1016/j.dsr.2008.09.008