argo data management

Argo data management

DOI: http://dx.doi.org/10.13155/84370

BGC-Argo quality control manual for nitrate concentration

Version 1.0

October 01, 2021

1

ARGO

part of the integrated global observation strategy

BGC-Argo quality control manual for nitrate concentration

Authors: Kenneth Johnson, Tanya Maurer, Joshua Plant, Henry Bittig, Christina Schallenberg, Catherine Schmechtig

How to cite this document

Kenneth Johnson, Tanya Maurer, Joshua Plant, Henry Bittig, Christina Schallenberg, Catherine Schmechtig (2021). **BGC-Argo quality control manual for nitrate concentration.** http://dx.doi.org/10.13155/84370

Table of contents

History of the document

Version	Date	Authors	Modification
1.0	October 1st 2021	Kenneth Johnson, Tanya Maurer, Joshua Plant, Henry Bittig, Christina Schallenberg, Catherine Schmechtig	Initial version

Reference Documents

Reference N°	Title	Link
#RD1	Argo Quality Control Manual for CTD and Trajectory Data	http://dx.doi.org/10.13155/33951
#RD2	Argo Quality Control Manual for Biogeochemical Data	http://dx.doi.org/10.13155/40879
#RD3	Argo user manual	http://dx.doi.org/10.13155/29825
#RD4	Processing Bio-Argo nitrate concentration at the DAC Level	http://dx.doi.org/10.13155/46121

Preamble

At the 16th Argo Data Management Team (ADMT) meeting, it was decided to split the Argo quality control manual in two manuals:

- the Argo quality control manual for CTD and trajectory data (JULD, LATITUDE, LONGITUDE, PRES, TEMP, PSAL, TEMP, CNDC, (#RD1) and,
- the Argo quality control manual for biogeochemical data (#RD2).

As there are many different groups of experts in charge of the assessment of different biogeochemical data sets, the Argo quality control manual for biogeochemical data should be considered as the cover document for all biogeochemical data quality control manuals. This document is dedicated to the description of the specific tests for the quality control of nitrate concentration and the related intermediate parameters.

WARNING:

Users should be aware that although biogeochemical data are now freely available at the Argo Global Data Assembly Centres (GDACs) along with their corresponding CTD data, the accuracy of these biogeochemical data in their raw state is generally not suitable for direct usage in scientific applications. Users are warned that the raw biogeochemical data should be treated with care and that adjustments are almost always needed before these data can be used for meaningful scientific applications. NITRATE_ADJUSTED is the optimal nitrate parameter for scientific applications. The data user should always inspect the data quality flag, NITRATE_ADJUSTED_QC, for each measurement, as well as the nitrate PARAMETER_DATA_MODE assignment.

Any user of these biogeochemical data who develops a specific and dedicated adjustment improving data accuracy is invited to contact the ADMT for potential inclusion of their method in a future edition of this document.

1. Introduction

This document is the Argo quality control (QC) manual for nitrate concentration, where the metadata parameter name for the state variable is NITRATE with units µmol kg^{-1}. The document describes two levels of quality control:

- The first level is the “real-time” (RT) quality control system, which always includes a set of agreed-upon automatic quality-control tests on each measurement. Data adjustments can also be applied within the real-time system, and quality flags assigned accordingly.
- The second level is the “delayed-mode” (DM) quality control system where data quality is assessed in detail by a delayed-mode operator and adjustments (based on comparison to high-quality reference fields) are derived. As mentioned, these adjustments can then be propagated forward and applied to incoming data in real-time until the next delayed-mode assessment is performed.

In core-Argo profile files, where <PARAM> = PRES, TEMP, PSAL (and sometimes CNDC), each <PARAM> has 5 QC and adjusted variables that are used to record real-time qc test results and delayed-mode adjustment information:

<PARAM>_QC, PROFILE_<PARAM>_QC, <PARAM>_ADJUSTED, <PARAM>_ADJUSTED_QC, and <PARAM>_ADJUSTED_ERROR.

In b-Argo profile files, <PARAM> can be classified into 3 groups:

(a). B-Argo <PARAM>: these are the ocean state biogeochemical variables that will receive real-time QC tests, adjustment in real-time, and delayed-mode adjustments. They are stored in both the b-files and the GDAC merged (Sprof) files.

(b). I-Argo <PARAM>: these are the intermediate biogeochemical variables that are only stored in the b-files. They will receive real-time QC tests and may receive adjustments.

(c). PRES: this is the stand-alone vertical index that links the core- and b-files.

B-Argo and I-Argo parameters for nitrate are identified in Section 2.1, Table 1 in #RD4.

The following are some clarifications on the QC and adjusted variables that are included in the b-files:

(a). B-Argo <PARAM>: all 5 qc and adjusted variables are mandatory for B-Argo PARAM in the b-files.

(b). I-Argo <PARAM>: <PARAM>_QC and PROFILE_<PARAM>_QC are mandatory for I-Argo <PARAM>. <PARAM>_ADJUSTED, <PARAM>_ADJUSTED_QC and <PARAM>_ADJUSTED_ERROR are optional.

(c). PRES: the b-files do not contain any QC or adjusted variables for PRES. These are in the core-file.

In b-Argo profile files, biogeochemical parameters can receive adjustments at different times. Therefore, the variable PARAMETER_DATA_MODE (N_PROF, N_PARAM) is added to b-Argo profile files to indicate the data mode of each <PARAM> in each N_PROF. The PARAMETER_DATA_MODE describes the data mode of the individual parameter:
R : real time data
D : delayed mode data

A : real time data with adjusted values

In b-Argo profile files, the variable PARAMETER_DATA_MODE associated to the variable PRES is always ‘R’, as adjusted values provided for PRES are only stored in the core profile file. Thus, to access the ‘best’ existing version of parameter (<PARAM>) data, except PRES, the user should:

1. Retrieve the data mode of the <PARAM> (from DATA_MODE(N_PROF) in a c-file and from PARAMETER_DATA_MODE(N_PROF, N_PARAM) in a b-file or an s-file),
2. Access the data:
    - If the data mode is ‘R’: In <PARAM>, <PARAM>_QC and PROFILE_<PARAM>_QC,
    - If the data mode is ‘A’ or ‘D’: In <PARAM>_ADJUSTED, <PARAM>_ADJUSTED_QC, PROFILE_<PARAM>_QC and <PARAM>_ADJUSTED_ERROR.

Note that the data mode of an I-Argo parameter may depend on the DAC’s decision of whether or not to include the adjusted fields for a particular I-Argo parameter in the b-Argo profile file:

- If <PARAM>_ADJUSTED, <PARAM>_ADJUSTED_QC and <PARAM>_ADJUSTED_ERROR are present in the file, the data mode of the I-Argo parameter can be ‘R’, ‘A’ or ‘D’,
- If not, the data mode of the I-Argo parameter should always be ‘R’.

Following the rules applied in the computation of PROFILE_PARAM_QC see #RD3 QC flag values of 5 or 8 should be considered GOOD data, while QC flag values of 9 (missing) should not be considered as an indicator of quality.

Flag	**Meaning**	**Real-time comment** *applicable to _QC in 'R' mode and _ADJUSTED_QC in 'A' mode*	**Delayed-mode comment** *applicable to _ADJUSTED_QC in 'D' mode*
0	No QC is performed	No QC is performed.	No QC is performed.
1	Good data	Good data. All Argo real-time QC tests passed. These measurements are good within the limits of the Argo real-time QC tests.	Good data. No adjustment is needed, or the adjusted value is statistically consistent with good quality reference data. An error estimate is supplied.
2	Probably good data	Probably good data. These measurements are to be used with caution.	Probably good data. Delayed mode evaluation is based on insufficient information. An error estimate is supplied.
3	Probably bad data that are	Probably bad data. These measurements are not to be used without scientific adjustment,	Probably bad data. An adjustment may (or may not) have been applied, but the

	potentially adjustable	e.g. data affected by sensor drift but may be adjusted in delayed-mode	value may still be bad. An error estimate is supplied.
4	Bad data	Bad data. These measurements are not to be used. A flag '4' indicates that a relevant real-time qc test has failed. A flag '4' may also be assigned for bad measurements that are known to be not adjustable, e.g. due to sensor failure.	Bad data. Not adjustable. Adjusted data are replaced by FillValue.
5	Value changed	Value changed	Value changed
8	Estimated value	Estimated value (interpolated, extrapolated or other estimation).	Estimated value (interpolated, extrapolated or other estimation).
9	Missing value	Missing value. Data parameter will record FillValue.	Missing value. Data parameter will record FillValue.

2. Real-time quality control for nitrate concentration and associated intermediate parameters

2.1. Introduction

Because of the requirement for delivering data to users within 24-48 hours of the float reaching the surface, the quality control procedures on the real-time data are limited and automatic.

Real-time tests are defined below for the biogeochemical parameter NITRATE (and NITRATE_ADJUSTED).

2.1.1. Correspondance between NITRATE and core parameters and QC flags

NITRATE delayed-mode quality control and adjustment may occur before or after those of core Argo PTS variables. To have a common guide, the following specifications are made:

- In 'R' Mode, NITRATE_ADJUSTED and NITRATE_ADJUSTED_QC are FillValue.
- In 'A' Mode, NITRATE_ADJUSTED(_QC) is computed from NITRATE(_QC) (from raw P/T/S).
- In 'D' Mode: An optional step in DMQC after a float has died is to recalculate NITRATE using ADJUSTED PRES, TEMP, and PSAL to give NITRATE_RECALCULATED (not stored in the Argo files!). NITRATE_ADJUSTED can be computed from NITRATE, or NITRATE_RECALCULATED. Whatever step is used is recorded in the SCIENTIFIC_CALIB section. Whatever step is not used is accounted for in NITRATE_ADJUSTED_ERROR.

In all three parameter data modes 'R', 'A', or 'D', NITRATE is the raw value, computed from the raw PTS. NITRATE_QC obeys the flag propagation policy from section §2.2.2.1 using the core QC flags.

2.2. Argo real-time quality control tests on vertical profiles of nitrate concentration

2.2.1. Common Argo real-time quality control tests on vertical profiles

This section lists the real-time tests that are common between CTD data and biogeochemical data. The same real-time test numbers for CTD data are used here. See Argo quality control manual (#RD1, #RD2)

The following tests are applied to nitrate concentration, see Argo Quality Control Manual for Biogeochemical Data (#RD2).

6. Global range test

This test applies a gross filter on observed values for NITRATE.

- NITRATE in range [−15, 65] micromole/kg
- NITRATE_ADJUSTED in range [-2 50] micromole/kg

Action: Values that fail this test should be flagged with a QC = '4' for NITRATE.

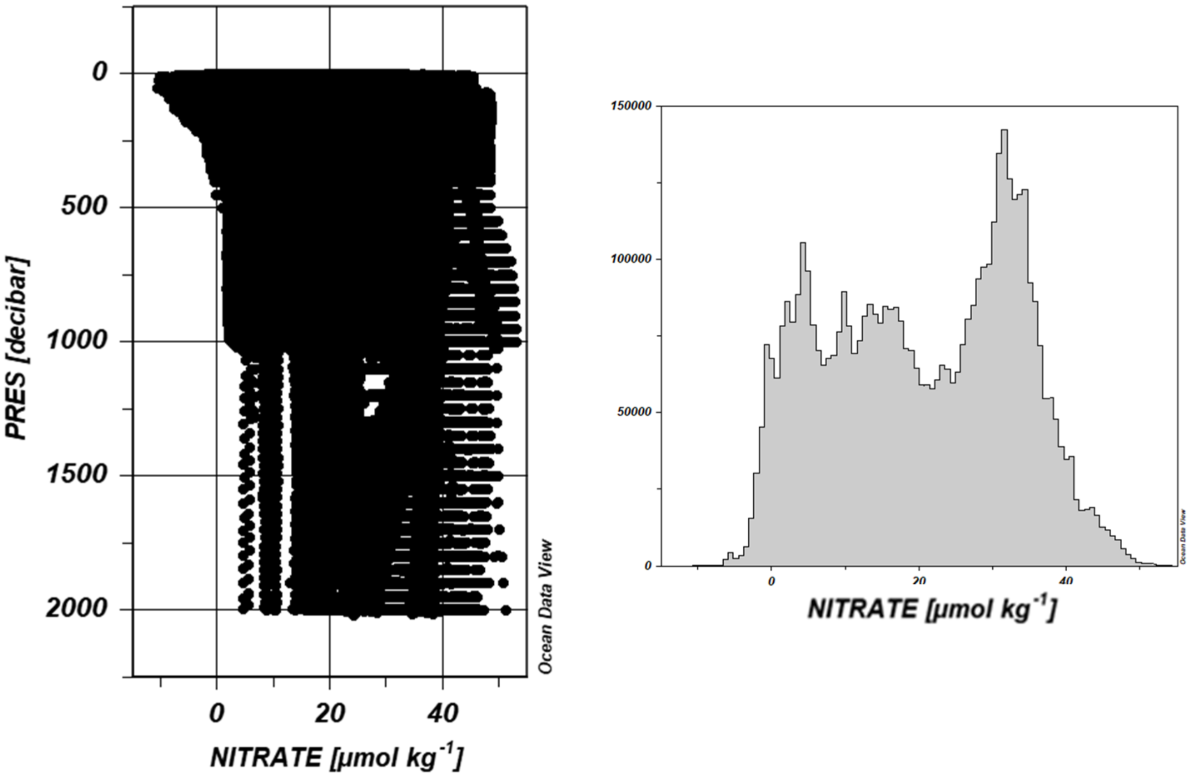

Figure 1: NITRATE (raw) profile data range from GDAC holdings (source: Sprof files).

9. Spike test

We calculate the absolute difference between the nitrate concentration at a certain depth (V2) and a running median (5 values, V0, V1, V2, V3, V4) along the whole profile:

- TestValue = ABS[V2 - MEDIAN(V0,V1,V2,V3,V4)]

The action taken on the threshold test value defined above will vary regionally. For example, specific ranges for observations from the Mediterranean Sea and the Red Sea further restrict what are considered reasonable values.

The Red Sea is defined by the region 10N, 40E; 20N, 50E; 30N, 30E; 10N, 40E.

The Mediterranean Sea is defined by the region 30N, 6W; 30N, 40E; 40N, 35E; 42N, 20E; 50N, 15E; 40N, 5W; 30N, 6W.

Action:

In Red Sea and Mediterranean Sea:

IF TestValue > 1 micromole/kg THEN the test failed and V2 should be flagged with a QC = ‘4’ for NITRATE.

Other places:

IF TestValue > 5 micromole/kg THEN the test failed and V2 should be flagged with a QC = ‘4’ for NITRATE.

13. Stuck value test

This test looks for all biogeochemical sensor outputs (i.e. ‘i’ and ‘b’ parameter measurements transmitted by the float) in a vertical profile being identical.

Action: Stuck values should be flagged as bad data (NITRATE_QC = ’4’).

15. Grey list

See Argo quality control manual (#RD1).

2.2.2. Specific Argo real-time quality control tests on vertical profiles

59. NITRATE specific Argo real-time quality control tests

2.2.2.1. Initial QC

It was decided at the BGC Argo Data Management task team meeting on May 26, 2020 that real-time unadjusted NITRATE data should receive a quality flag of ‘3’. This is because the majority of nitrate sensors deployed on BGC Argo profiling floats suffer from shifts in calibration (of varying magnitude) that often occur during the time between initial laboratory calibration and float deployment. Because this is a known bias that affects the majority of nitrate sensors within the array, and because it is something that can be corrected (see Section 4 of this manual), NITRATE_QC should be set to ‘3’. The following real-time test should be used in order to populate NITRATE_QC:

If <PARAMETER> == NITRATE

NITRATE_QC = 3

End

PRES, TEMP and PSAL are used to compute NITRATE. Considering the impact of PRES and TEMP on the NITRATE calculation, when PRES_QC=4 and/or TEMP_QC=4 then NITRATE_QC should be set to 4.

When PSAL_QC=4, NITRATE_QC should be kept to 3 in real time because in many cases PSAL is not bad enough to justify a NITRATE_QC of 4. However, further review of the impact of PSAL on NITRATE in such cases should be performed in delayed-mode, as the impact on NITRATE uncertainty is dependent on the degree of degradation in PSAL. For example, floats with CTDs identified as “rapid salty drifters” may experience significant rapid degradation in PSAL quality over a short period of time resulting in heightened impact on the quality of computed nitrate.

Action:

If TEMP_QC=4 and/or PRES_QC=4, then NITRATE_QC=4

If PSAL_QC = 4, then NITRATE_QC=3 (to be further reviewed in delayed-mode)

2.2.2.2. Saturation of spectrophotometer

The method used to measure dissolved nitrate concentration (NITRATE) with sensors mounted on profiling floats is based on the absorption of light at ultraviolet wavelengths by dissolved sea salt and nitrate ions (#RD4). These measurements are performed with a spectrophotometer. Occasionally, the I-Argo parameter UV_INTENSITY_NITRATE may reach the saturation value of the detector (this SATURATION_VALUE depends on the output value type, e.g., for the SUNA V2 on CTS4 floats, this value is coded on two bytes so the SATURATION_VALUE = 2^{16} -1=65535). If saturation occurs, the retrieval of the nitrate concentration may not be correct and should be flagged with a '3', "probably bad".

If R is a pixel count running from PIXEL_FIT_START to PIXEL_FIT_END, and one saturation value occurs, then the nitrate concentration is probably bad.

Action: IF ANY UV_INTENSITY_NITRATE(R) = SATURATION_VALUE THEN NITRATE_QC = '3'

A skilled operator may recover these values in delayed mode with custom code that ignores pixels that are saturated.

2.2.2.3. Absorbance at 240 nm

The seawater absorbance spectrum can be used to diagnose the sensor's optical performance. Neither nitrate nor bromide absorbs light strongly at 240nm. Therefore, high absorption at 240 nm suggests the optical window hass fouled or the optical path has been compromised. High absorbance means a low amount of light is reaching the detector and analytical sensitivity can become an issue. Calculate the seawater absorbance at the maximum pixel <= 240nm (ABSORBANCE_SW($i_{\sim 240nm}$), see Equation 1 in #RD4)

Action: IF 0.8 ≤ ABSORBANCE_SW(i_{240nm}) < 1.1 THEN NITRATE_QC = '3'
Action: IF ABSORBANCE_SW(i_{240nm}) ≥ 1.1 THEN NITRATE_QC = '4'

2.2.2.4. RMS Error of nitrate fit residuals

Nitrate is calculated by assuming a model seawater UV light absorbance spectrum composed of nitrate, bromide and a linear baseline. If the measured data fit the model poorly, this suggests the calibration has changed or that there are other constituents in seawater absorbing UV light which are not accounted for in the model. To evaluate the fit between predicted and observed absorbances, we calculate the root mean square error of the residuals over the wavelength range used to fit nitrate (all terms in this equation are described in #RD4):

FIT_ERROR_NITRATE = SQRT{ sumR (ABSORBANCE_COR_NITRATE(R) –
[BASELINE_INTERCEPT + BASELINE_SLOPE*OPTICAL_WAVELENGTH_UV(R) +
MOLAR_NITRATE * E_NITRATE(R)]) ^2 / (PIXEL_FIT_END - PIXEL_FIT_START) }

where the sumR runs from PIXEL_FIT_START to PIXEL_FIT_END

Action: IF FIT_ERROR_NITRATE ≥ 0.003 THEN NITRATE_QC = ‘4’

2.2.2.5. Presence of high bisulfide concentrations.

The bisulfide ion (HS-) also absorbs light in a UV region that overlaps with the nitrate spectrum. Elevated levels of bisulfide in anoxic basins, such as the Black Sea, will interfere with the computation of nitrate. If bisulfide concentration exceeds 5 µmol kg^{-1}, then nitrate should be flagged as bad. Note that computation of bisulfide is not a routine procedure and will likely only be done by skilled operators that are processing data from floats in regions where oxygen concentration is less than 2 µmol kg^{-1}.

Action: IF BISULFIDE ≥ 5 THEN NITRATE_QC = ‘4’

2.2.3. Test application order on vertical profiles

The Argo real time QC tests on vertical profiles are applied in the order described in the following table. See Argo quality control manual (#RD1, #RD2) for tests not listed in §**Erreur ! Source du renvoi introuvable.**.

Order	Test number	Test name
1	1	Platform Identification test
2	2	Impossible Date test
3	3	Impossible Location test
4	4	Position on Land test
5	5	Impossible Speed test
6	15	Grey List Test
7	19	Deepest pressure test
8	6	Global Range test
9	9	Spike test
10	13	Stuck Value test
11	59	Nitrate concentration specific tests

2.2.4. Scientific calibration information for each profile

If PARAMETER_DATA_MODE is ‘R’, there is no reason to fill the scientific calibration information, thus:

For PARAMs (B-Argo PARAMs and I-Argo PARAMs) in ‘R’-mode	
SCIENTIFIC_CALIB_COMMENT	FillValue
SCIENTIFIC_CALIB_EQUATION	FillValue
SCIENTIFIC_CALIB_COEFFICIENT	FillValue
SCIENTIFIC_CALIB_DATE	FillValue

A specific comment should however be set for PRES parameter

For PRES	
SCIENTIFIC_CALIB_COMMENT	'Adjusted values are provided in the core profile file'
SCIENTIFIC_CALIB_EQUATION	FillValue
SCIENTIFIC_CALIB_COEFFICIENT	FillValue
SCIENTIFIC_CALIB_DATE	FillValue

(see Chapters **Erreur ! Source du renvoi introuvable.** and 3.4.1 how to fill scientific calibration information when PARAMETER_DATA_MODE is 'A' or 'D' respectively).

2.3. Argo real-time quality control tests on trajectories

Nitrate concentration trajectory data are sometimes duplicates of vertical profile data, e.g., dated levels of PROVOR/ARVOR profiles that are present in the profile file (without their time stamps) and duplicated in the trajectory file (with their associated time stamps). These data should be duplicated with their associated QC values, which were set during the real-time quality control tests performed on the vertical profiles.

2.4. Argo real-time quality control tests on near-surface data

No near-surface data related to nitrate concentration are acquired therefore no tests are performed on near-surface data.

2.5. Argo real-time quality control tests for deep float data

No specific tests are defined for deep float data.

2.6. Quality control flag application policy

The QC flag value assigned by a test cannot override a higher value from a previous test. Example: a QC flag '4' (bad data) set by Test 6 (range test) cannot be decreased to QC flag '3' (bad data that are potentially correctable) by Test 59 (Nitrate specific test).

A value with NITRATE_QC flag '4' (bad data), or with a NITRATE_ADJUSTED_QC flag '4' (bad data) or '3' (bad data that are potentially correctable) is ignored by the quality control tests.

Note that flag values of 5 or 8 should be considered GOOD data, while QC flag values of 9 (missing) should not be considered as an indicator of quality.

3. Adjustments of nitrate in Real Time and Delayed Mode

Data from Argo floats that pass through automatic quality control procedures and are delivered to the Argo GDAC, typically within 24 hours, are referred to as Real-Time (RT) data. If the float Principal Investigator applies further corrections to the data, which usually involves visual inspection of data relative to a reference data set, then the data are referred to as Delayed Mode (DM) data. DM corrections for core Argo pressure and salinity data are normally made within 6 to 12 months of collection. The adjustments that are applied to DM data after visual inspection may also be applied to RT data, as they are received, without visual inspection. These data are referred to as Adjusted RT or Adjusted Mode (A Mode) data (Argo User's Manual, 2017).

As noted by Johnson et al. (2017) and Maurer et al. (2021), most BGC-Argo nitrate data must receive real-time adjustments to meet the Argo goals of delivering research-quality observations. Applications of using uncorrected data are relatively limited in comparison to the utility of corrected values. An Argo quality flag of 3 (questionable, probably bad) should be assigned to unadjusted RT nitrate data that passes the real-time tests outlined in Section 2.2. The Argo goals for research-quality data then require that the RT sensor data be adjusted in real-time, as noted above, to receive a quality flag of 1 (good data). To accomplish this goal, best practice is to perform the first delayed mode correction after only a few (~5) cycles (see Section 3.3 below). Automatic real time procedures can then carry these initial corrections forward to produce more accurate Adjusted Mode RT data. Note that if an initial DM assessment is unavailable at the DAC, real-time adjustments can be based purely on automatic procedures, and further refined once DM corrections are performed. The procedures to accomplish each of these tasks are outlined in the text below.

3.1. General data adjustment process

Chemical sensors typically suffer from two problems: inaccurate initial calibrations, which result from sensor instability during storage and transport before deployment, and subsequent drift or offsets that occur post-deployment. Addressing such issues through the data adjustment process is essential. Similar to Argo salinity adjustments (Owens and Wong, 2009), the nitrate adjustment process depends on having an accurate model for nitrate concentrations in waters below 1000 m depth, where temporal and spatial variability is minimal over decadal time scales. This reference data set will be used in the following adjustment procedures to populate the NITRATE_ADJUSTED variable. The basic approach is to calculate the correction between the reference data and the measured data at operator selected depths between 1000 and 2000 m, and then apply this correction to the entire profile of raw measured data to yield the adjusted profile data. It has been shown that the correction value determined at depth can be applied to the entire profile (Johnson et al., 2017).

3.2. Model Reference Datasets

There are multiple methods available to estimate the model reference data for nitrate at a global scale (see table below). These include Multiple Linear Regression (MLR) equations (e.g., Locally Interpolated Nitrate Regression (LINR) method, Carter et al., 2018), a neural network prediction system known as CANYON-B (Bittig et al., 2018) and the Mediterranean version, CANYON-MED (Fourrier et al., 2020), as well as the World Ocean Atlas (WOA) a gridded monthly climatology (Garcia et al., 2019). More recently available are the Empirical Seawater Property Estimation Routines (ESPERs), as described in Carter et al (2021), which are now available for use. Many of these methods utilize adjusted oxygen (DOXY_ADJUSTED) as an input variable for computing predicted nitrate concentrations. Therefore, it is strongly recommended to always deploy both nitrate and oxygen sensors together and to perform any needed corrections to raw oxygen values prior to making any adjustments to nitrate. Without an oxygen measurement, the correction schemes for NITRATE are much less robust. In that case, they rely on an estimation of reference nitrate based on a spatial and temporal interpolation of the WOA data set at depths between 1000 and 2000 m. Note that presently the LINR function is only coded as a Matlab function and the CANYON-B estimation routine is only available as a Matlab or R function. Please see links and references in the table below for detailed information. If the user decides to use the WOA database it is easiest to download all nitrate data files to a local computer first (1.5Gb).

Reference	Method	Source for equation or Matlab functions or data
Carter et al. (2018)	LIR	https://github.com/BRCScienceProducts/LIRs
Bittig et al. (2018)	CANYON_B	https://github.com/HCBScienceProducts/CANYON-B
Fourrier et al.(2020)	CANYON-MED	https://github.com/MarineFou/CANYON-MED/tree/master/v2
Carter et al. (2021)	ESPER	https://github.com/BRCScienceProducts/ESPER
Garcia et al. (2019)	WOA	ftp://ftp.nodc.noaa.gov/pub/data.nodc/woa/WOA18/DATA/nitrate/netcdf/all/1.00/

3.3. Real-Time adjustment and quality control options for nitrate

3.3.1. Automatic Real-Time adjustment procedure (no previous DM assessment)

There is one suggested method currently in practice for the automatic adjustment of NITRATE when no previous delayed-mode assessment has been performed. This method utilizes World Ocean Atlas nitrate fields as the reference dataset because this reference option allows for the greatest flexibility in DAC automatic RT operations (for example, it eliminates the reliance on other parameters in the adjustment process and does not require specific software to access/run). The procedure is as follows: The NITRATE(PRES_WOA) concentration at a PRES_WOA pressure which is max(Profile Pressures) – 100 dbar is calculated. The nitrate concentration (n_woa) at PRES_WOA dbar is then retrieved from WOA at the closest neighbour of LATITUDE and LONGITUDE of the profile. The CORRECTION is the median of NITRATE(PRES_WOA)-n_woa(PRES_WOA) cumulated over two months after the deployment.

Then, NITRATE_ADJUSTED = NITRATE – CORRECTION over the complete profile.

When NITRATE_QC(PRES_WOA) = ‘4’ (RMS error test, see 2.2.2.3), the associated NITRATE(PRES_WOA) should not be used to compute the CORRECTION.

If, after the **two months** period after deployment, less than **5** NITRATE(PRES_WOA) values are available to compute the CORRECTION, the period is extended so that a minimum of **5** values is used in the calculation of the median.

If NITRATE_ADJUSTED is calculated in this way, then:

- NITRATE_ADJUSTED_QC is initially set to ‘2’ and then the NITRATE_ADJUSTED field should go through standard RT QC tests, similar to the raw parameters (tests described in 2.2.1)
- NITRATE_ADJUSTED_ERROR = 5 micromol kg-1

3.3.2. Real-Time adjustment procedure based on previous DM assessment

As mentioned previously, ideally a nitrate record on a biogeochemical float should receive a delayed-mode adjustment within the first two months of life. The delayed-mode adjustment procedures are described below in Section 3.4 and roughly follow the method outlined in Maurer et al. (2021). Once a DM adjustment has been performed, the adjustment applied to the last profile evaluated within the DM assessment ($CORRECTION_n$) can be applied automatically to incoming profiles in real-time. Thus, for profile n+i, the real-time adjustment becomes

$$NITRATE\_ADJUSTED_{n+i} = NITRATE_{n+i} - CORRECTION_n$$

until a subsequent DM assessment is performed. While this method cannot explicitly account for any future sensor drift, it serves as a more accurate first-guess than the method described in section 3.3.1 as the correction applied is closer in time to incoming cycles and thus more highly correlated.

If NITRATE_ADJUSTED is calculated in this way, then:

- NITRATE_ADJUSTED_QC is initially set to '1' and then the NITRATE_ADJUSTED field should go through standard RT QC tests, similar to the raw parameters (tests described in 2.2.1)
- NITRATE_ADJUSTED_ERROR = Elast + 1 umol/kg year-1 · (JULD-JULDlast)/365 where Elast = NITRATE_ADJUSTED_ERROR at last point of DMQC, JULDlast = JULD at last point of DMQC, and JULD = current date.

3.3.3. Parameter data mode and scientific calibration information for each profile

When a biogeochemical parameter ('b' parameter) has been through an adjustment procedure, its PARAMETER_DATA_MODE is set to 'A' which means "adjusted in real-time". The PARAMETER_DATA_MODE of all intermediate parameters ('i' parameters) associated to this adjusted biogeochemical parameter are also set to 'A' when they have an "_ADJUSTED" field (but left as 'R' if not).

If PARAMETER_DATA_MODE is 'A', none of the scientific calibration information should be set to FillValue and every field should be filled.

As mentioned in §1, for I-Argo <PARAM>, while <PARAM>_QC and PROFILE_<PARAM>_QC are mandatory, <PARAM>_ADJUSTED, <PARAM>_ADJUSTED_QC and <PARAM>_ADJUSTED_ERROR are optional.

The three fields SCIENTIFIC_CALIB_COMMENT, _EQUATION, and _COEFFICIENT have netCDF dimensions (N_PROF, N_CALIB, N_PARAM, STRING256). This means that for each N_CALIB, each field is a 256-length character string. If character strings longer than 256-length are needed, the procedure should be separated and stored as multiple N_CALIB.

For a single calibration that needs multiple N_CALIB:

- the SCIENTIFIC_CALIB_DATE should be identical for all N_CALIB,
- once the different fields are correctly filled, the remaining empty fields (unused) should be filled as follows:
  - ✓ SCIENTIFIC_CALIB_COMMENT: 'No additional comment',
  - ✓ SCIENTIFIC_CALIB_EQUATION: 'No additional equation',
  - ✓ SCIENTIFIC_CALIB_COEFFICIENT:'No additional coefficient'.

3.3.3.1. Sample fields for RT adjustment based on automatic comparison to WOA nitrate concentration

For I-Argo PARAMs with no corresponding _ADJUSTED field and for which the associated B-Argo PARAMs have been through adjustment in real-time	
SCIENTIFIC_CALIB_COMMENT	'not applicable'
SCIENTIFIC_CALIB_EQUATION	'not applicable'
SCIENTIFIC_CALIB_COEFFICIENT	'not applicable'
SCIENTIFIC_CALIB_DATE	YYYYMMDDHHMISS(*)

For B-Argo NITRATE that has been through adjustment in Real-Time	
SCIENTIFIC_CALIB_COMMENT	CORRECTION is the median of NITRATE(PRES_WOA)-n_woa(PRES_WOA) cumulated over two months after the deployment;PRES_WOA=Profile Pressure-100;n_woa (LATITUDE,LONGITUDE) (closest neighbour) from WOA annual file
SCIENTIFIC_CALIB_EQUATION	NITRATE_ADJUSTED = NITRATE - CORRECTION ; CORRECTION=med[NITRATE(PRES_WOA)-n_woa(PRES_WOA)cumulated over two months after the deployment]
SCIENTIFIC_CALIB_COEFFICIENT	CORRECTION=2.71504,NITRATE(PRES_WOA)= 36.5545,n_an(PRES_WOA)=33.7413(*)
SCIENTIFIC_CALIB_DATE	20190930072909 (**)

(*): Example values are extracted from the float 3902120, profile 4 (Coriolis DAC). Values are given as example and are subject to change depending on reprocessing.

(**): for a given calibration, the SCIENTIFIC_CALIB_DATE of an adjusted B-Argo parameter and of its associated I-Argo parameters should be identical.

3.3.3.2. Sample fields for RT adjustments based on previous DM assessment

For I-Argo PARAMs with no corresponding _ADJUSTED field and for which the associated B-Argo PARAMs have been through adjustment in real-time

SCIENTIFIC_CALIB_COMMENT	'not applicable'
SCIENTIFIC_CALIB_EQUATION	'not applicable'
SCIENTIFIC_CALIB_COEFFICIENT	'not applicable'
SCIENTIFIC_CALIB_DATE	YYYYMMDDHHMISS(*)

For B-Argo NITRATE that has been through adjustment in Real-Time	
SCIENTIFIC_CALIB_COMMENT	CORRECTION(n) is the value of the adjustment applied to the last NITRATE profile assessed in delayed mode. n is the cycle number of the last NITRATE profile assessed in delayed mode.
SCIENTIFIC_CALIB_EQUATION	NITRATE_ADJUSTED = NITRATE – CORRECTION(n);
SCIENTIFIC_CALIB_COEFFICIENT	CORRECTION(n)=1.3530(*)
SCIENTIFIC_CALIB_DATE	YYYYMMDDHHMISS(**)

(*): The example correction coefficient displayed here would be used in the real-time adjustment of incoming nitrate data beyond cycle 299 (the last cycle used in DMQC assessment) for AOML/UW/MBARI float 5906039 (see Section 3.4).

(**): for a given calibration, the SCIENTIFIC_CALIB_DATE of an adjusted B-Argo parameter and of its associated I-Argo parameters should be identical.

3.4. Delayed-mode adjustment and quality control options for nitrate

3.4.1. Calculation of the CORRECTION

The delayed mode adjustment process should follow the general approach recommended in section 3.1. Calculation of the CORRECTION time series is the first step. The CORRECTION is defined as the difference between the raw measured nitrate and the model reference nitrate at the reference depth for a given profile (reference options were described in section 3.2):

$$\mathrm{CORRECTION}_{(I,P)} = \mathrm{NITRATE}_{(I,P)} - \mathrm{REF}_{(I,\,P)}$$

REF is the reference value at cycle I and pressure P. P should equal the sample pressure chosen by the operator between 1000 and 2000 dbar. The chosen pressure is likely to represent the depth with the least variability in nitrate. The float should reach the chosen pressure consistently. The CORRECTION should be calculated for all cycles at this reference pressure for NITRATE_QC not equal to 4. If the CORRECTION is calculated using the LINR or CANYON-B reference models, follow the documentation instructions found in the links in the table within section 3.2. If WOA is used, a reference profile must be created through spatial interpolation of the gridded annual data file (woa18_all_n00_01.nc, variable "n_woa"). It is often useful to calculate model reference data for the whole profile and compare this to the measured profile data as an additional visual quality control check. In this case, a spatial & temporal interpolation of the monthly and annual files is required. Matlab code to accomplish this task is freely available at the following GitHub repository maintained by MBARI:https://github.com/SOCCOM-BGCArgo/ARGO_PROCESSING/tree/master/MFILES/WOA.

Once a CORRECTION time series is calculated for the life of the float, all NITRATE measurements within a profile, *I*, can be adjusted as,

$$\text{NITRATE\_ADJUSTED}_{(I)} = [\text{NITRATE}_{(I)} - \text{CORRECTION}_{(I,P)}]/\text{GAIN}$$

For the vast majority of case the GAIN should = 1 because a correction at depth should be valid at the surface. Under certain circumstances, such as for very warm tropical surface waters where the nitrate temperature correction may be imperfect, the GAIN can be used as a tuning factor to bring the adjusted data into better alignment with the expected true values. GAIN adjustments should be used with caution by expert DM operators and only after careful inspection to ensure that proper calibration data are being used for the nitrate calculation and that there is not a sensor malfunction. The derivation of a nitrate sensor gain is a manual, iterative process and is not required for the majority of the floats within the array. GAIN values within the range of [0.95 1.05] are considered acceptable, and should be applied after the CORRECTION has been incorporated.

It is important to note that due to the presence of noise within the sensor time series, it is best to first model the CORRECTION time series prior to subtracting it from the original data series. The choice of model should be compatible with the known behavior of the sensor. For nitrate sensors deployed on biogeochemical Argo profiling floats, it is not uncommon for jumps in the data series to occur due to dirty optics. These jumps are typically periodic and followed by longer periods of steady drift. We thus recommend modelling the CORRECTION series through a segmented set of piecewise discontinuous linear fits, with each segment defined by a set of breakpoints, or nodes, corresponding to the cycle (time) at which a change in an offset or drift occurs (as described in Maurer et al., 2021). These nodes can be chosen manually by the delayed mode operator, although a more objective method, such as automated change-point detection in conjunction with the Bayesian Information Criterion (or alternative statistical model selection tool), is advised. If modelling the CORRECTION series in this way, it is convenient to store the model coefficients from each linear fit in a « correction matrix » to be accessed and applied during processing. An example CORRECTION time series for float 5906039 is shown in the figure below in green. Float data at P = 1500 dbar is shown in blue and LINR reference model (at same pressure level) is shown in red.

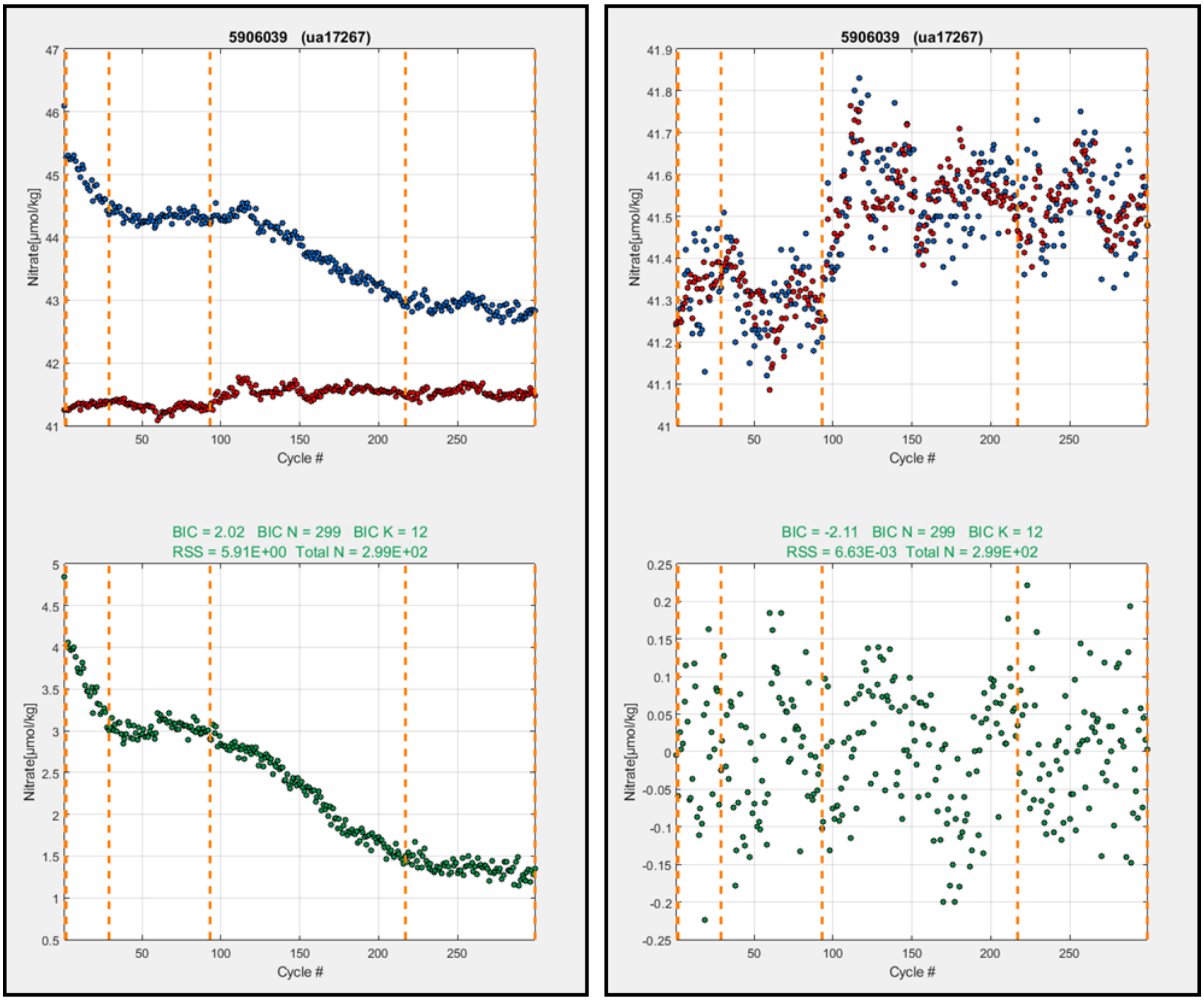

Figure 2: Example delayed-mode nitrate CORRECTION calculation for float 5906039 using SAGE software. Left panel (top) shows raw NITRATE data from the float (blue) and LINR reference (red) data at 1500 dbar and the resulting correction required in green (bottom). Right panel (top) shows NITRATE_ADJUSTED data (blue) with the LINR reference (red) and post-correction residuals in green (bottom). Orange dashed lines depict the cycle nodes (breakpoints) detected by the software using automated change-point detection.

The software used in this assessment (the SOCCOM Assessment and Graphical Evaluation, or, SAGE) was created at MBARI and is freely available at "https://github.com/SOCCOM-BGCArgo/ARGO_PROCESSING/tree/master/MFILES/GUIS/SAGE" (Maurer et al, 2021). Note that for this float, the software has automatically identified four breakpoints between the two bounding nodes at the start and end of the time series, resulting in five segments for which a least-squares fit was performed on the CORRECTION series. The resulting correction matrix for this example is as follows :

Node (*i*)	Gain	Offset (*O*)	Drift (*D*)
1	1	4.8450	0
2	1	4.0789	-13.4462
29	1	3.0304	0
93	1	2.9998	-4.2745

217	1	1.4600	-0.6592
299	1	1.3530	0

and the modelled CORRECTION at each node, *i*, becomes :

CORRECTION$_{(i,1500dbar)}$ = $O_{(i)}$

and the modelled correction at each subsequent cycle, *j*, between segment nodes becomes :

CORRECTION$_{(j,1500dbar)}$ = $O_{(i)} + D_{(i)}(T_{(j)}-T_{(i)})/365$

Where *T* represents the time (in days).

The CORRECTION then gets subtracted from the NITRATE series as described above. As mentioned previously, the number and location of the breakpoints could also be determined manually by the DM operator. However, automated methods such as the one used in this example are more objective and prevent over-correction of the data. Note that the post-correction residuals (right bottom panel in Figure 1) using this method remain well within the accuracy of the sensor (+/-0.5 micromol/kg).

If NITRATE_ADJUSTED is calculated in this way in delayed-mode, then:

- NITRATE_ADJUSTED_ERROR = 1 umol kg^{-1} + DOXY_ADJUSTED_ERROR/10

3.4.2. Editing QC flags in delayed-mode

In addition to the data adjustment assessment, delayed-mode operators should examine profile data for pointwise errors such as missed spikes and jumps, and edit and check the qc flags in <PARAM>_QC and <PARAM>_ADJUSTED_QC (when the adjustment was performed in Real Time). Here, <PARAM> refers to the biogeochemical parameters that have been through the delayed-mode process.

Examples where <PARAM>_QC and <PARAM>_ADJUSTED_QC should be edited in delayed-mode include:

- <PARAM>_QC/<PARAM>_ADJUSTED_QC should be changed to ‘4’ for bad and uncorrectable data that are not detected by the real-time tests; and
- <PARAM>_QC/<PARAM>_ADJUSTED_QC should be changed to ‘1’ or ‘2’ for good data that are wrongly identified as probably bad by the real-time tests.

3.4.3. Compulsory variables to be filled in a BD profile file

This section lists the compulsory variables that must be filled in an Argo netCDF b-profile file that has been through the delayed-mode process.

3.4.3.1. QC and ADJUSTED variables

Each B-Argo <PARAM> has 5 mandatory qc and adjusted variables in the B- profile file:

- <PARAM>_QC
- PROFILE_<PARAM>_QC
- <PARAM>_ADJUSTED
- <PARAM>_ADJUSTED_QC
- <PARAM>_ADJUSTED_ERROR

When a B-Argo <PARAM> has been through the delayed-mode process, the above 5 mandatory qc and adjusted variables must be filled in the BD profile file. PROFILE_<PARAM>_QC should be re-computed when <PARAM>_ADJUSTED_QC becomes available.

For I-Argo <PARAM>, <PARAM>_QC and PROFILE_<PARAM>_QC are mandatory, but the 3 adjusted variables are optional in the B- profile file:
<PARAM>_ADJUSTED, <PARAM>_ADJUSTED_QC, <PARAM>_ADJUSTED_ERROR.

If a data centre chooses to include these 3 adjusted variables for I-Argo <PARAM> in the B-profile file, then these 3 adjusted variables must be filled when the I-Argo <PARAM> has been through the delayed-mode process, and PROFILE_<PARAM>_QC should be re-computed with <PARAM>_ADJUSTED_QC.

Note that PRES in the B- profile file does not carry any qc or adjusted variables. It is used as a stand-alone vertical index that links the core- and b-files. Users who want delayed-mode adjusted pressure values (PRES_ADJUSTED) should obtain them from the core- files.

3.4.3.2. Scientific calibration information for each profile

It is compulsory to fill the scientific calibration section of a BD- profile file.

PARAMETER should contain every parameter recorded in STATION_PARAMETER (including PRES), even though not all STATION_PARAMETER have delayed-mode qc.

When a biogeochemical parameter (‘b’ parameter) has been through a delayed-mode procedure its PARAMETER_DATA_MODE is set to ‘D’. The PARAMETER_DATA_MODE of all intermediate parameters (‘i’ parameters) associated to this adjusted biogeochemical parameter are also set to ‘D’ when they have an “_ADJUSTED” field (but left to ‘R’ if not).

If PARAMETER_DATA_MODE is ‘D’, none of the scientific calibration information should be set to FillValue and every information should be filled.

Here are the indications on how to fill the scientific calibration section of a BD profile file.

For I-Argo PARAMs related to NITRATE with no corresponding _ADJUSTED field and for which the associated B-Argo NITRATE has been through delayed-mode qc	
SCIENTIFIC_CALIB_COMMENT	‘not applicable’
SCIENTIFIC_CALIB_EQUATION	‘not applicable’
SCIENTIFIC_CALIB_COEFFICIENT	‘not applicable’
SCIENTIFIC_CALIB_DATE	YYYYMMDDHHMISS(*)

For I-Argo PARAMs related to NITRATE with corresponding _ADJUSTED fields and for which the associated B-Argo NITRATE has been through delayed-mode qc	
SCIENTIFIC_CALIB_COMMENT	Content depends on <PARAM> and method
SCIENTIFIC_CALIB_EQUATION	Content depends on <PARAM> and method

SCIENTIFIC_CALIB_COEFFICIENT	Content depends on <PARAM> and method
SCIENTIFIC_CALIB_DATE	YYYYMMDDHHMISS(*)

For NITRATE that has been through delayed-mode qc	
SCIENTIFIC_CALIB_COMMENT	JULD_PIVOT is the time at cycle X, the pivot cycle (or time at which drift assessment began). JULD is the time at the cycle for which the correction is being applied. OFFSET [umol kg-1], DRIFT [umol kg-1 yr-1], GAIN and pivot cycle determined by climatology comparisons at depth*.
SCIENTIFIC_CALIB_EQUATION	NITRATE_ADJUSTED = [NITRATE – CORRECTION]/GAIN; CORRECTION = OFFSET + DRIFT*(JULD – JULD_PIVOT)/365
SCIENTIFIC_CALIB_COEFFICIENT**	OFFSET = 4.0789; DRIFT = 13.4462; GAIN = 1.0000; JULD_PIVOT = 25388.1680; JULD = 25389.3995
SCIENTIFIC_CALIB_DATE	20211025142319 (***)

(*): the specific climatology or reference model (and depth range) used in the assessment can be listed explicitly here, should the DMQC operator choose to do so.

(**): coefficients listed are for cycle 3 from MBARI-processed float 5906039 (see Figure 2).

(***): for a given calibration, the SCIENTIFIC_CALIB_DATE of an adjusted B-Argo parameter and of its associated I-Argo parameters should be identical.

The three fields SCIENTIFIC_CALIB_COMMENT, _EQUATION, and _COEFFICIENT have netCDF dimensions (N_PROF, N_CALIB, N_PARAM, STRING256). This means that for each N_CALIB, each field is a 256-length character string. If character strings longer than 256-length are needed, the procedure should be separated and stored as multiple N_CALIB.

For a single calibration that needs multiple N_CALIB:

- the SCIENTIFIC_CALIB_DATE should be identical for all N_CALIB,
- once the different fields are correctly filled, the remaining empty fields (unused) should be filled as follows:
  - ✓ SCIENTIFIC_CALIB_COMMENT: ‘No additional comment’,
  - ✓ SCIENTIFIC_CALIB_EQUATION: ‘No additional equation’,
  - ✓ SCIENTIFIC_CALIB_COEFFICIENT:’No additional coefficient’.

3.4.3.3. Other variables in a BD profile file

Here are other variables in a B- profile file that need to be updated after delayed-mode qc.

- The variable DATA_STATE_INDICATOR should record '2C' or '2C+'.
- The variable DATE_UPDATE should record the date of last update of the netCDF file, in the format YYYYMMDDHHMISS.
- In both the core- and b- profile files, the variable DATA_MODE(N_PROF) is not related to a specific parameter. The value of DATA_MODE(N_PROF) is set to ‘D’ when adjusted values for one or more <PARAM> in each N_PROF become available. In b-Argo profile files, there are additional biogeochemical parameters which can receive delayed-mode adjustments at different times. Therefore the variable PARAMETER_DATA_MODE(N_PROF, N_PARAM) is added to b-Argo profile files to indicate the data mode of each <PARAM> in each N_PROF.

  The adjusted section (<PARAM>_ADJUSTED, <PARAM>_ADJUSTED_QC and <PARAM>_ADJUSTED_ERROR) for each <PARAM> in each N_PROF should then be filled independently according to its PARAMETER_DATA_MODE.

  For example, in a b-Argo profile file with DOXY and NITRATE, it is possible that

  PARAMETER_DATA_MODE = ’D’ for DOXY, and

  PARAMETER_DATA_MODE = ’R’ for NITRATE.

  In this case:

  -the adjusted section for DOXY should be filled with their adjusted values;

  -the adjusted section for NITRATE should be filled with FillValues.
- A history record should be appended to the HISTORY section of the netCDF file to indicate that the netCDF file has been through the delayed-mode process. Please refer to the Argo User's Manual (§5 "Using the History section of the Argo netCDF Structure") on usage of the History section.

3.4.3.4. Profile files naming convention

When one or more <PARAM> in a single-cycle core- profile file has received delayed-mode adjusted values, the file name changes from R<WMO_ID>_xxx.nc to D<WMO_ID>_xxx.nc.

When one or more <PARAM> in a single-cycle B-profile file has received delayed-mode adjusted values, the file name changes from BR<WMO_ID>_xxx.nc to BD<WMO_ID>_xxx.nc.

When one or more <PARAM> in a single-cycle M- profile file receive delayed-mode adjusted values, the file name changes from MR<WMO_ID>_xxx.nc to MD<WMO_ID>_xxx.nc.

When one or more <PARAM> in a single-cycle S- profile file receive delayed-mode adjusted values, the file name changes from SR<WMO_ID>_xxx.nc to SD<WMO_ID>_xxx.nc.

References

Bittig HC, Steinhoff T, Claustre H, Fiedler B, Williams NL, Sauzède R, Körtzinger A and Gattuso J-P (2018). An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO2 Variables and Nutrient Concentrations From T, S, and O2 Data Using Bayesian Neural Networks. Front. Mar. Sci. 5:328. doi: 10.3389/fmars.2018.00328

Carter, B. R., Feely, R. A., Williams, N. L., Dickson, A. G., Fong, M. B. and Takeshita, Y. (2018), Updated methods for global locally interpolated estimation of alkalinity, pH, and nitrate. Limnol. Oceanogr. Methods, 16: 119-131. doi: 10.1002/lom3.10232

Carter, B.R., Bittig, H.C., Fassbender, A.J., Sharp, J.D., Takeshita, Y., Xu, Y.-Y., Álvarez, M., Wanninkhof, R., Feely, R.A. and Barbero, L. (2021), New and updated global empirical seawater property estimation routines. Limnol Oceanogr Methods. doi:10.1002/lom3.10461

Fourrier, M., Coppola, L., Claustre, H., D'Ortenzio, F., Sauzède, R., and Gattuso, J.-P. (2020), A Regional Neural Network Approach to Estimate Water-Column Nutrient Concentrations and Carbonate System Variables in the Mediterranean Sea: CANYON-MED Front. Mar. Sci. 7:620. doi: 10.3389/fmars.2020.00620

Garcia H.E., T.P. Boyer, O.K. Baranova, R.A. Locarnini, A.V. Mishonov, A. Grodsky, C.R. Paver, K.W. Weathers, I.V. Smolyar, J.R. Reagan, D. Seidov, M.M. Zweng (2019). World Ocean Atlas 2018: Product Documentation. A. Mishonov, Technical Editor.

Johnson, Kenneth S., Hans. W. Jannasch, Luke J. Coletti, Virginia A. Elrod, Todd R. Martz, Yuichiro Takeshita, Robert J. Carlson, James G. Connery. 2016. Deep-Sea DuraFET: A pressure tolerant pH sensor designed for global sensor networks. Analytical Chemistry. 88, 3249-3256, DOI: 10.1021/acs.analchem.5b04653.

Johnson, K. S., J. N. Plant, L. J. Coletti, H. W. Jannasch, C. M. Sakamoto, S. C. Riser, D. D. Swift, N. L. Williams, E. Boss, N. Haëntjens, L. D. Talley, J. L Sarmiento. 2017. Biogeochemical sensor performance in the SOCCOM profiling float array. Journal of Geophysical Research Oceans, 122, doi:10.1002/2017JC012838.

Maurer, T. L., Plant, J. N., Johnson, K. S. (2021). Delayed-Mode Quality Control of Oxygen, Nitrate, and pH Data on SOCCOM Biogeochemical Profiling Floats. Frontiers in Marine Science 8. doi: 10.3389/fmars.2021.683207

Sauzede, R., Bittig, H., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J., Legendre, L. and K.S. Johnson. 2017. Estimates of water-column nutrients and carbonate system parameters in the global ocean: A novel approach based on neural networks. Frontier in Marine Science, 4:128. doi: 10.3389/fmars.2017.00128