argo data management

Argo data management

DOI: http://dx.doi.org/10.13155/97828

BGC-Argo quality control manual for pH

Version 1.0

December 11, 2023

ARGO

part of the integrated global observation strategy

BGC-Argo quality control manual for pH

Authors: Kenneth S. Johnson, Tanya L. Maurer, Joshua N. Plant, Yuichiro Takeshita

How to cite this document

Kenneth S. Johnson, Tanya L. Maurer, Joshua N. Plant, Yuichiro Takeshita (2023).
BGC-Argo quality control manual for pH.

http://dx.doi.org/10.13155/97828

Table of contents

History of the document

Version	Date	Authors	Modification
1.0	December 11th 2023	Kenneth S. Johnson, Tanya L. Maurer, Joshua N. Plant, Yuichiro Takeshita	Initial version as a stand-alone document. Previous QC instruction was incorporated into the pH processing document v1.0 (initially a merged processing & QC document)

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 pH at the DAC Level	http://dx.doi.org/10.13155/57195

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 pH concentration and the related intermediate parameters.

Note that this document is still in progress. As such, there are highlighted sections of text throughout that need to be addressed. Yellow highlighting means this is an open topic - some things are known about this topic, but agreement needs to be reached on how best to officially address this in the data system.

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. PH_IN_SITU_TOTAL_ADJUSTED is the optimal pH parameter for scientific applications. The data user should always inspect the data quality flag, PH_IN_SITU_TOTAL_ADJUSTED_QC, for each measurement, as well as the pH 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 pH, where the metadata parameter name for the state variable is PH_IN_SITU_TOTAL (units dimensionless). 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 pH 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 with 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	Probably bad data. An adjustment may (or may not)

	potentially adjustable	without scientific adjustment, e.g. data affected by sensor drift but may be adjusted in delayed-mode	have been applied, but the 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 pH 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 PH_IN_SITU_TOTAL (and PH_IN_SITU_TOTAL_ADJUSTED).

2.1.1. Correspondence between PH_IN_SITU_TOTAL and core parameters and QC flags

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

In all three parameter data modes 'R', 'A', or 'D', PH_IN_SITU_TOTAL is the raw value, computed from the raw PTS. PH_IN_SITU_TOTAL_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 pH

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 pH concentration, see Argo Quality Control Manual for Biogeochemical Data (#RD2).

6. Global range test

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

- PH_IN_SITU_TOTAL in range [7.0 8.8]

- PH_IN_SITU_TOTAL_ADJUSTED in range [7.3 8.5]

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

9. Spike test

The pH sensor can generate occasional spikes due to electrical noise and despiking is appropriate. The default Argo spike tests in core variables (Wong et al., 2018) is

$$\text{Test value 1} = |\, V2 - (V3 + V1)/2 \,| - |\, (V3 - V1) / 2 \,| \qquad (1)$$

where V2 is the measurement being tested as a spike, and V1 and V3 are the values above and below. This test does not work well for pH. The criteria depends strongly on the vertical gradient, making it regionally dependent. The test is also sensitive to the vertical resolution of the sensor data, which is platform dependent. We do not recommend this approach.

The spike test recommended for chlorophyll (Schmechtig et al., 2014) is more appropriate for pH. The value tested is

$$\text{Test value 2} = |V2 - \text{median}(V0, V1, V2, V3, V4)| \qquad (2)$$

where the test value represents the anomaly of the observed pH from the median of the surrounding data. Schmechtig et al. (2014) recommend that the criteria used to determine if a data value is a spike be based on the statistical distribution of the 5 data points. Perhaps because pH has a smaller dynamic range than chlorophyll, we find that a constant value of 0.04 pH works well. A data point is considered a spike and marked with quality flag 4 (data bad) if Test value 2 > 0.04 pH.

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 (PH_IN_SITU_TOTAL_QC = '4').

15. Grey list

See Argo quality control manual (#RD1).

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

56. PH_IN_SITU_TOTAL 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 PH_IN_SITU_TOTAL data should receive a quality flag of ‘3’. This is because the majority of pH 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 pH sensors within the array, and because it is something that can be corrected (see Section 4 of this manual), PH_IN_SITU_TOTAL_QC should be set to ‘3’. The following real-time test should be used in order to populate PH_IN_SITU_TOTAL_QC:

If <PARAMETER> == PH_IN_SITU_TOTAL

PH_IN_SITU_TOTAL_QC = 3

End

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

When PSAL_QC=4, PH_IN_SITU_TOTAL_QC should be kept to 3 in real time because in many cases PSAL is not bad enough to justify a PH_IN_SITU_TOTAL_QC of 4. However, further review of the impact of PSAL on PH_IN_SITU_TOTAL in such cases should be performed in delayed-mode, as the impact on PH_IN_SITU_TOTAL uncertainty is dependent on the degree of degradation in PSAL. For example, floats with CTDs identified as “abrupt 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 pH as well as the reference algorithm estimate used to adjust the data. If the DMQC operator so chooses, there is an accepted protocol for restoring PH_IN_SITU data quality by using a PSAL estimate from high-quality salinity product, thus enabling of BGC parameters and adjustments in delayed mode. This method is described in #RD2, section 5.3 (Recovering BGC data when float salinity is bad).

Action:

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

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

2.2.2.2. Sensor diagnostic checks

There are a number of pH sensor diagnostics that may be returned from the pH sensor and some of these can be used to assess the health of the sensor and associated impact on PH_IN_SITU_TOTAL. These include intermediate parameters IB_PH and IK_PH. Currently not all sensors are configured to report these intermediate parameters, so they have not been incorporated into any official Real-Time test. However, certain data centers have found the following range checks on pH sensor diagnostics to be particularly useful.

- IB_PH in range [-100 100] nano amps
- IK_PH in range [-100 100] nano amps

Action (optional): Values that fail this test can be flagged with a QC = ‘3’ for PH_IN_SITU_TOTAL, for further review in delayed-mode.

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).

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	pH 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 **3.3.3.2** and **3.4.4.2** on 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

pH 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 pH 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 (pH specific test).

A value with PH_IN_SITU_TOTAL_QC flag ‘4’ (bad data), or with a PH_IN_SITU_TOTAL_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 pH 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 pH 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 pH 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 for pH should not be made (this is not the case for nitrate, for which preliminary automated corrections can be made using WOA, and further refined once DM corrections are performed.)

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 pH adjustment process depends on having an accurate model for pH 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 PH_IN_SITU_TOTAL_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 (see Section 3.4.1 for details). It has been shown for Deep-Sea DuraFET sensors that the correction value determined at depth can be applied to the entire profile (Johnson et al., 2017). In essence, we assume that sensor drift only affects the constant term (k_0) of the reference potential, and the temperature (k_2(PRES)) and pressure (f(PRES)) coefficients remain unchanged. This is a key assumption in the adjustment process, and is corroborated through field and laboratory observations (Johnson et al. 2016, Maurer et al. 2021)

3.2. Model Reference Datasets

There are multiple methods available to estimate the model reference data for pH at a global scale (see table below). These include Multiple Linear Regression (MLR) equations (e.g., Locally Interpolated pH Regression (LIPHR) 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). More recently available are the Empirical Seawater Property Estimation Routines (ESPERs), as described in Carter et al (2021), which are now available for use. Note that the ESPER routines are the most recently available option, trained on the most up-to-date quality-controlled data. Additionally, there are three versions of ESPER available: an MLR-based version, a neural-network-based version, and a version that represents average estimates from the two (ESPERmix). Note that ESPERmix is recommended for most applications (Carter et al 2021). Many of these methods utilize adjusted oxygen (DOXY_ADJUSTED) as an input variable for computing predicted pH concentrations. Therefore, it is strongly recommended to always deploy both pH and oxygen sensors together and to perform any needed corrections to raw oxygen values prior to making any adjustments to pH due to the use of this variable in the calculation of the reference data. Without an oxygen measurement, the correction schemes for PH_IN_SITU_TOTAL are much less robust. A LIPHR equation without oxygen as an input is available for use if oxygen is bad or unavailable, but note that this version of the algorithm is less skilled. If attempting to adjust PH_IN_SITU_TOTAL using LIPHR equation without oxygen (ie Eq 8 as described in Carter et al., 2018), then care must be taken to assess algorithm performance at the float's location, and PH_IN_SITU_TOTAL_ADJUSTED_ERROR should be inflated accordingly (a conservative approach would be to double the error in such cases when comprehensive error analysis cannot be performed).

Note that presently the LIPHR and ESPER function is only coded as a Matlab function (python version in development) 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.

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

It is also important to note that the current version of CANYON-B (v2) does not attempt to reconcile the various pH measurements (i.e. pH directly measured using spectrophotometry with purified meta-cresol purple and pH calculated from dissolved inorganic carbon and total alkalinity measurements), while the LIR and ESPER methods do. As a result, predictions from CANYON-B at deep reference depths (~1500m) tend to be biased high on the order of a few millipH relative to LIR predictions at the same locations.

The use of different reference models can lead to different adjustments. In some regions, e.g. the Southern Ocean, differences are close to negligible (mean differences CANYON-B pH minus LIR pH data of 0.001 pH units (Maurer et al., 2021)). Additionally, pH from model reference datasets at a reference depth of 1500 m agree excellently on a global average (mean difference ≤0.001 pH units). However, float trajectories cover only subregions, and accuracy at the regional level may vary. For example, in the Northern hemisphere or the Indian Ocean, differences between reference datasets are regionally coherent and are significantly higher (up to order 0.020 pH units). Comparing the mismatch between the various reference datasets (LIR, CANYON-B, ESPER), offers insight into regional uncertainties, yet care must be taken with this approach, as differences stemming from more recently trained algorithms may actually represent improvements to regional estimates. This is an active area of research, and continuous model validation and improvement will be a critical component for reducing these regional biases.

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

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

There is no suggested method currently in practice for the automatic adjustment of PH_IN_SITU_TOTAL when no previous delayed-mode assessment has been performed. This may be updated in the future.

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

As mentioned previously, ideally a pH 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 follows equation (3) below until a subsequent DM assessment is performed.

$$PH\_IN\_SITU\_TOTAL\_ADJUSTED_{n+i} = PH\_IN\_SITU\_TOTAL_{n+i} - CORRECTION_n * TCOR \quad (3)$$

where

$$TCOR = (TEMP_{QCdepth} + 273.15)/(TEMP + 273.15) \quad (4)$$

Note that the change in pH that is computed at depth (the CORRECTION) gets adjusted at each sample by the ratio of the absolute temperature of the sample (TEMP) to the absolute temperature at reference depth ($TEMP_{QCdepth}$ in equation (4)). This is equivalent to a change in k0, the assumed source of the shift in sensor performance and the reason behind the need for pH adjustment. 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 PH_IN_SITU_TOTAL_ADJUSTED is calculated in this way, then:

- PH_IN_SITU_TOTAL_ADJUSTED_QC is initially set to ‘1’ and then the PH_IN_SITU_TOTAL_ADJUSTED field should go through standard RT QC tests, similar to the raw parameters (tests described in 2.2.1)
- PH_IN_SITU_TOTAL_ADJUSTED_ERROR = Elast + 0.03 · (JULD-JULDlast)/365 where Elast = PH_IN_SITU_TOTAL_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 of pH

Not applicable at this time.

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 PH_IN_SITU_TOTAL that has been through adjustment in Real-Time	
SCIENTIFIC_CALIB_COMMENT	CORRECTION(n) is the value of the adjustment applied to the last PH_IN_SITU_TOTAL profile assessed in delayed mode. n is the cycle number of the last

	PH_IN_SITU_TOTAL profile assessed in delayed mode.
SCIENTIFIC_CALIB_EQUATION	PH_IN_SITU_TOTAL _ADJUSTED = PH_IN_SITU_TOTAL – CORRECTION(n);
SCIENTIFIC_CALIB_COEFFICIEN T	CORRECTION(n)=0.0210(*)
SCIENTIFIC_CALIB_DATE	YYYYMMDDHHMISS(**)

(*): The example correction coefficient displayed here would be used in the real-time adjustment of incoming pH data beyond cycle 162 (the last cycle used in DMQC assessment) for AOML/UW/MBARI float 5904395 (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 pH

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 pH and the model reference pH at the reference depth for a given profile (reference options were described in section 3.2):

$$\text{CORRECTION}_{(I,P)} = \text{PH\_IN\_SITU\_TOTAL}_{(I,P)} - \text{REF}_{(I,\,P)} \qquad (5)$$

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 pH. The float should reach the chosen pressure consistently. The CORRECTION should be calculated for all cycles at this reference pressure for PH_IN_SITU_TOTAL_QC not equal to 4. If the CORRECTION is calculated using the LIPHR or CANYON-B reference models, follow the documentation instructions found in the links in the table within section 3.2. It is often useful to view model reference data for the whole profile and compare this to the measured profile data as an additional visual quality control check.

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

$$\text{PH\_IN\_SITU\_TOTAL\_ADJUSTED}_{(I)} = [\text{PH\_IN\_SITU\_TOTAL}_{(I)} - \text{CORRECTION}_{(I,P)}*\text{TCOR}]/\text{GAIN} \qquad (6)$$

where

$$\text{TCOR} = (\text{TEMP}_{\text{QCdepth}} + 273.15)/(\text{TEMP} + 273.15) \qquad (7)$$

(As described in Section 3.3.2.) Note that currently for the vast majority of cases the GAIN should = 1 because a correction at depth should be valid at the surface. However, this term remains in the correctin equation to date in case it is required for future floats.

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 pH sensors deployed on biogeochemical Argo profiling floats, it is not uncommon for jumps in the data series to occur due to small shifts in the sensor reference potential (k_0) over time. 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 5904395 is shown in the figure below in green. Float data at P = 1500 dbar is shown in blue and LIPHR reference model (at same pressure level) is shown in red.

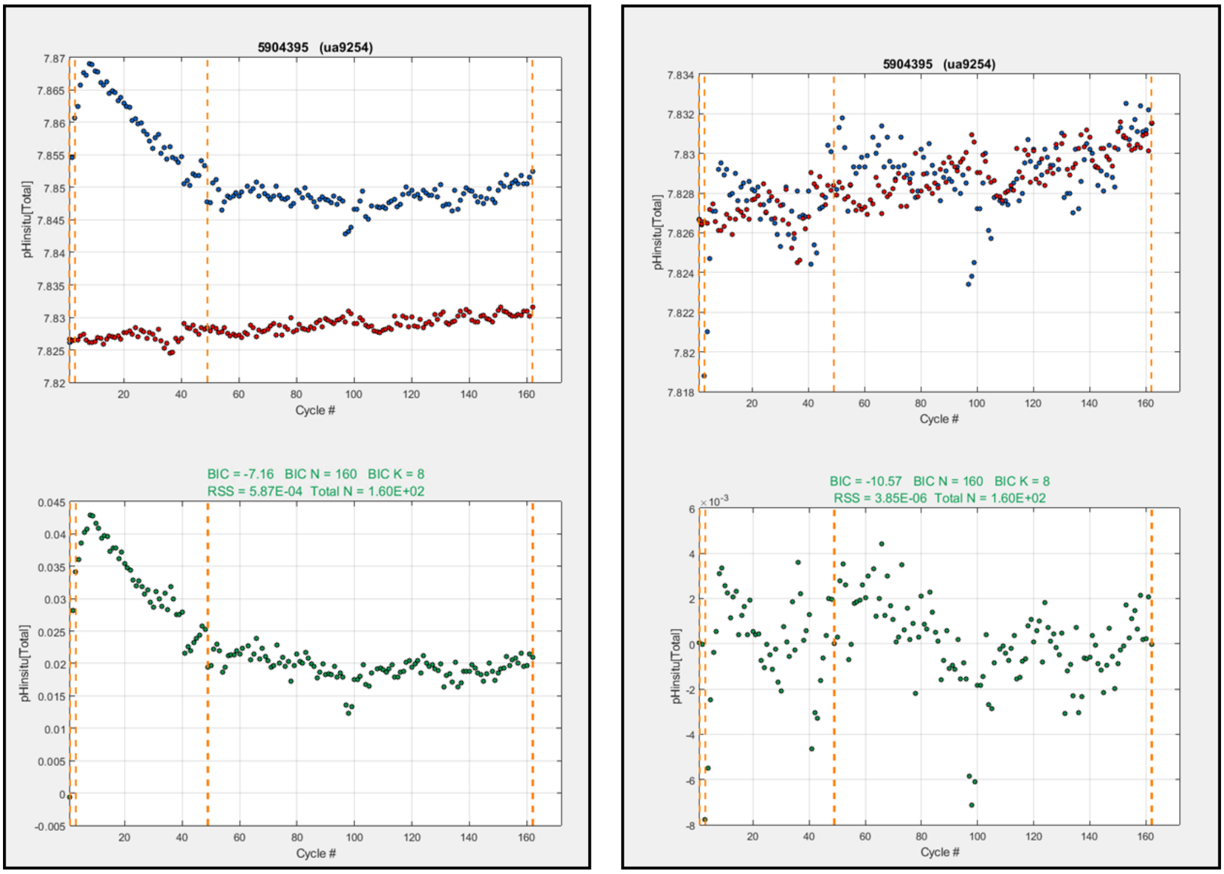

Figure 1: Example delayed-mode pH CORRECTION calculation for float 5904395 using SAGE software. Left panel (top) shows raw PH_IN_SITU_TOTAL data from the float (blue) and LIPHR reference (red) data at 1500 dbar and the resulting correction required in green (bottom). Right panel (top) shows PH_IN_SITU_TOTAL_ADJUSTED 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 two breakpoints between the two bounding nodes at the start and end of the time series, resulting in three 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	-0.0006	1.9717
3	1	0.0419	-0.0285
49	1	0.0194	0
162	1	0.0210	0

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

$$\text{CORRECTION}_{(i,1500dbar)} = O_{(i)} \quad (8)$$

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

$$\text{CORRECTION}_{(j,1500dbar)} = O_{(i)} + D_{(i)}(T_{(j)}-T_{(i)})/365 \quad (9)$$

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

The CORRECTION then gets subtracted from the PH_IN_SITU_TOTAL 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 (+/-0.005 pH units).

The expected error in float pH measurements is derived from the uncertainty in the reference data as well as sensor uncertainties. The expected error reported by Carter et al. (2018) for pH values predicted with LIR from depth, temperature, salinity and oxygen (Apparent Oxygen Utilization) at depths near 2000 m is 0.001±0.006, where the error terms are a bias and the root mean square error. Additional errors in float sensor pH may arise (Williams et al., 2017) because of uncertainty in the sensor pressure and temperature coefficients (0.0025 pH), and uncertainty in the carbon system thermodynamics used to convert laboratory pH values to in situ values (0.005). The pH sensor precision is near 0.003 (Williams et al., 2017). A combined error budget that recognizes both systematic and random errors suggests that sensor pH values corrected with the LIR equations alone have an accuracy near 0.009. However, as previously noted, the absolute value of PH_IN_SITU_TOTAL_ADJUSTED will vary depending on the choice of reference algorithm used in the adjustment process, and these discrepancies in reference algorithms vary regionally. Float data adjusted to the more recently trained ESPER routines, for example, may reflect improved uncertainties In addition, because float oxygen data is routinely recommended for use as an input to the various reference algorithms, any error from DOXY impacts the uncertainty in PH_IN_SITU_TOTAL_ADJUSTED and should be accounted for as well.

Until the uncertainty budget for PH_IN_SITU_TOTAL _ADJUSTED is better constrained (this is currently an active area of research), a reasonable approach to estimating pH adjusted error would be to calculate it roughly as the square root of the sum of squared error from the primary identified sources of uncertainty: a conservative 0.01 combined error from the sensor and reference (to be modified regionally, should the operator deem it necessary), plus 0.0016 for every umol/kg-error in oxygen (Williams et al 2016; S1). For example, based on geography alone, uncertainties in the reference might be significantly lower or higher (such as in the Southern Ocean or North Atlantic, respectively). However, it is important that any uncertainty estimates lower than the global scale recommendation be substantiated by a comparison of profiling float pH sensor data with shipboard measurements made near the time of float deployment and then converted to in situ values (Johnson et al., 2017; Maurer et al., 2021).

For PH_IN_SITU_TOTAL _ADJUSTED calculated following the method outlined in 3.4.1:

PH_IN_SITU_TOTAL_ADJUSTED_ERROR = sqrt[0.01^2 + $(\text{DOXY\_ADJUSTED\_ERROR}*0.0016)^2$] (10)

unless otherwise specified by the operator.

3.4.2. Addressing pH profiles that exhibit a “pump-offset”

For many float platforms, pH profiles may occasionally exhibit a step change at the pressure level where the CTD shifts from spot sampling to continuous profiling mode (CTD pump on; for many floats this is triggered at 985db, but can vary with float configuration). The magnitude of this offset in pH can vary from float to float, as well as through time, and although its root cause is not yet fully characterized, it is assumed to be partially related to environmental factors and flow over the reference electrode or FET. A correction protocol is currently being researched. Until a standardized correction scheme is agreed upon by the ADMT community, it is important to inspect each pH profile in delayed mode for the presence of this offset. If present (visible offset in pH of 0.01 or greater at the “pump-on” pressure level), the general recommendation is to perform the delayed mode assessment described in section 3.4 using a reference depth shallower than the pump-on pressure level, and subsequently flag data below the pump-on pressure level as a ‘3’ (which would remain biased, but could be recovered at a later date). An alternative option if the DM operator feels there is too much uncertainty in using reference depths less than 1000 m for floats consistently presenting this issue would be to operate the float with the pump on at all depths, with the understanding that this would have a modest effect on lifetime of the float (due to increased power demand from continuous pump operation). This option would eliminate the discontinuity throughout the vertical profile, and standard data adjustment, following Section 3.4, could then be performed.

3.4.3. Editing QC flags in delayed-mode

In addition to the data adjustment assessment and inspection of profiles for the presence of a pump-offset, 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.

3.4.4. 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.

3.4.4.1. 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.

3.4.4.2. 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 PH_IN_SITU_TOTAL with no corresponding _ADJUSTED field and for which the associated B-Argo PH_IN_SITU_TOTAL has not 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 PH_IN_SITU_TOTAL with corresponding _ADJUSTED fields and for which the associated B-Argo PH_IN_SITU_TOTAL 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 PH_IN_SITU_TOTAL 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 [pH], DRIFT [pH yr-1], GAIN and pivot cycle determined by climatology comparisons at depth*.
SCIENTIFIC_CALIB_EQUATION	PH_IN_SITU_TOTAL_ADJUSTED = [PH_IN_SITU_TOTAL-CORRECTION]/GAIN; CORRECTION = OFFSET + DRIFT*(JULD – JULD_PIVOT)/365
SCIENTIFIC_CALIB_COEFFICIENT**	OFFSET = 0.0419; DRIFT = -0.0285; GAIN = 1.0000; JULD_PIVOT = 23497.4056; JULD = 23502.8632

SCIENTIFIC_CALIB_DATE	20220524100121 (***)

(*): 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 4 from MBARI-processed float 5904395 (see Figure 1).

(***): 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’.

3.4.4.3. 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 PH_IN_SITU_TOTAL, it is possible that

  PARAMETER_DATA_MODE = ’D’ for DOXY, and

  PARAMETER_DATA_MODE = ’R’ for PH_IN_SITU_TOTAL.

  In this case:

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

-the adjusted section for PH_IN_SITU_TOTAL 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.

3.4.4.4. 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

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

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.

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

Wong, A., R. Keeley, T. Carval and the Argo Data Management Team. 2018. Argo Quality Control Manual For CTD and Trajectory Data. : http://dx.doi.org/10.13155/33951