Meeting agenda:

1. Discussion: standardization of QFs for RT unadjusted BGC data
2. Progress on parameter 'decision trees'
3. Formalizing a BGC reference database
4. Propagating more adjusted data to GDAC
5. Additional ADMT topics
   • Optode time response
   • Combined traj files
   • Fine-tuning Sprof files
   • Meta data

BGC-Argo ADMT 20 Summary of Action-Items:

- Organize the transfer from publication to BD files
  - BBP, CHLA: Herve, Uday, Marin, Catherine, Josh, Christina
  - DOXY: (Catherine, Henry)
- Gather the information of other data that have been published
- Document how to fill the metadata in the processing documentation
- Fill the metadata according to the documentation (all DACS)
- Organize and link all the open source software repositories on the BGC Argo Website (Catherine)
- Final decision tree
  - DOXY, pH, NITRATE (Josh, Henry, Virginie)
  - CHLA (Xiaogang, Herve, Catherine, Josh, Emmanuel)
  - BBP (Emmanuel, Giorgio)
  - Radiometry (Emanuele, Emmanuel, Antoine)
1. Discussion: standardization of QFs for RT unadjusted BGC data
Standardizing the use of BGC
Argo Quality Flags

Helping users find good data

Tanya Maurer, Josh Plant, Ken Johnson
MBARI
May 21, 2020
Overarching goals:

- Make BGC Argo data easy to use!
- Prevent accidental data misuse by novice users
- Enhance confidence and trust in the program
Problem:
• Different DACs are interpreting and applying Argo quality flags differently for BGC
• This presents a difficulty for users when trying to extract all “good” data from the global dataset
• What is the intended purpose of the quality flags for the data user population? Is there a relationship between data quality and data accuracy?

Example for DOXY:
MBARI flags RT unadjusted data a “3”, intent on conveying to users “Caution: do not use data without correction.”

Coriolis flags RT unadjusted data a “1”, intent on conveying to users: “This data has passed all real-time tests. Sensor is working.”
Problem:

• Different DACs are interpreting and applying Argo quality flags differently for BGC
• This presents a difficulty for users when trying to extract all “good” data from the global dataset
• What is the intended purpose of the quality flags for the data user population? Is there a relationship between data quality and data accuracy?

Example for DOXY:

MBARI flags RT unadjusted data a “3”, intent on conveying to users: “Caution: do not use data.

Coriolis flags RT unadjusted data a “1”, intent on conveying to users: “This data has passed all real-time tests. Sensor is working.”
Accuracy depends primarily on the individual calibration of the sensors, and on the proximity of calibration or reference data to the deployment. To allow the scientific use of DOXY data, an in-situ adjustment of DOXY (in real time or delayed mode) is crucial (see O2 QC manual; [RD18]) in order to correct O2 sensitivity drift (order several % per year when not deployed, Bittig et al. 2018, [RD17]).

Users should be aware that although biogeochemical data are now freely available at the Argo Global Data Assembly Centres (GDACs) along with their CTD data, the accuracy of these biogeochemical data at their raw state is not suitable for direct usage in scientific applications. Users are warned that the raw biogeochemical data should be treated with care, and that often, adjustments are needed before these data can be used for meaningful scientific applications.
MBARI flags RT unadjusted DOXY data a “3”. Does this make sense based on current Argo QF definitions? Yes!

<table>
<thead>
<tr>
<th>n</th>
<th>Meaning</th>
<th>Real-time comment</th>
<th>Delayed-mode comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No QC was performed</td>
<td>No QC was performed.</td>
<td>No QC was performed.</td>
</tr>
<tr>
<td>1</td>
<td>Good data</td>
<td>All Argo real-time QC tests passed.</td>
<td>The adjusted value is statistically consistent and a statistical error estimate is supplied.</td>
</tr>
<tr>
<td>2</td>
<td>Probably good data</td>
<td></td>
<td>Probably good data.</td>
</tr>
<tr>
<td>3</td>
<td>Bad data that are potentially correctable</td>
<td>Test 15 or Test 16 or Test 17 failed and all other real-time QC tests passed. These data are not to be used without scientific correction. A flag ‘3’ may be assigned by an operator during additional visual QC for bad data that may be corrected in delayed mode.</td>
<td>An adjustment has been applied, but the value may still be bad.</td>
</tr>
<tr>
<td>4</td>
<td>Bad data</td>
<td>Data have failed one or more of the real-time QC tests, excluding Test 16. A flag ‘4’ may be assigned by an operator during additional visual QC for bad data that are not correctable.</td>
<td>Bad data. Not adjustable.</td>
</tr>
<tr>
<td>5</td>
<td>Value changed</td>
<td>Value changed</td>
<td>Value changed</td>
</tr>
<tr>
<td>6</td>
<td>Not used</td>
<td>Not used</td>
<td>Not used</td>
</tr>
<tr>
<td>7</td>
<td>Not used</td>
<td>Not used</td>
<td>Not used</td>
</tr>
<tr>
<td>8</td>
<td>Estimated value</td>
<td>Estimated value (interpolated, extrapolated or other estimation).</td>
<td>Estimated value (interpolated, extrapolated or other estimation).</td>
</tr>
<tr>
<td>9</td>
<td>Missing value</td>
<td>Missing value</td>
<td>Missing value</td>
</tr>
</tbody>
</table>
15. Grey list

This test is implemented as a mechanism for data assembly centres (DACs) to flag, in real-time, sensors that are potentially not working correctly.

The decision to insert a float parameter in the grey list comes from the PI or the delayed-mode operator. A float parameter should be put in the grey list when sensor drift is too big to be corrected adequately in real-time, or when the sensor is judged to be potentially not working correctly.

16. Gross salinity or temperature sensor drift

This test is implemented to detect a sudden and significant sensor drift. It calculates the average temperature and salinity from the deepest 100 dbar of a profile and the previous good profile. Only measurements with good QC are used.

Action: For salinity, if the difference between the two average values is more than 0.5 PSU, then all salinity values from the profile are flagged as probably bad data ('3'). For temperature, if the difference between the two average values is more than 1°C, then all temperature values from the profile are flagged as probably bad data ('3').

17. Visual QC

This is subjective visual inspection of float measurements by an operator.
One of the main purposes for Argo RT tests is to screen data submitted to GTS for use in real-time operations.

“A beginner’s guide to accessing Argo data”,
http://www.argo.ucsd.edu/Argo_data_guide.pdf
states in reference to data that has passed the RT tests:

“In general these data should be consistent with ocean climatologies even though no climatology tests have been performed.“
⭐ RT unadjusted DOXY data should not go to GTS!
⭐ RT unadjusted DOXY data should not be used in scientific analysis!
⭐ Optode ‘storage drift’ a known sensor issue, well documented (Inaccuracy can be >20%, sensor dependent)

Median sensitivity loss of 7.66% (gain of 1.0766)  

BGC Argo DOXY Profile Gains (computed using WOA)

SOCCOM APEX floats
SOCCOM NAVIS floats
SOCCOM median = 5.62%
For PSAL data, this implies that data is good enough for government work (i.e., use in modeling, numerical weather prediction).

How good is “good enough” for DOXY?

If a float was deployed with optode reading low by 5, 10, 20%, is this “good enough”?

Might depend on the user’s application and scientific goals, but Argo policies should be developed to support the program’s mission and prevent misuse of the data.

QF = 1, “good data”, should imply that a certain level of accuracy has been met (in both RT and DM).
Proposed Action:

**Suggested QF=3 for RT unadjusted DOXY!**

This can easily be formulated within the context of a real-time test, if desired, in order to fit within the existing constructs of the Argo system & quality flag definitions.

**RT test:**

\[ \langle \text{PARAMETER} \rangle \sim \text{‘DOXY’} \]

- True
  - Data passed this RT test
    - \( \langle \text{PARAMETER} \rangle \_QC = 1 \)
- False
  - Data failed this RT test
    - \( \langle \text{PARAMETER} \rangle \_QC = 3 \)

Example is given for DOXY, although similar solution could (and should) be implemented for NITRATE and PH_IN_SITU_TOTAL
Summary

• Data system should be optimized for the user!
• Average users rely heavily on quality flags for screening data
• QC flag of “1” should be reserved for “Good data” (consistent with ocean climatologies, of accuracy acceptable for propagation to GTS)
• Raw, unadjusted data from sensors known to be out of calibration should not be identified as “good”. This confuses users.
• Easily prevent misuse of data and set QF=3 for RT unadjusted DOXY and other unadjusted BGC parameters

**SOCCOM-Argo user community agrees!!**
In discussing (real-time) data flagging, we need to acknowledge different uses of BGC-Argo data first of all.

Different uses like:
- Air sea flux and budgets (Gray et al. 2018, GRL; Bushinsky et al. 2017, JGR-Oceans)
- Diapycnal flux/supply of nutrients to the surface layer.
- Seasonal drawdown of nitrate (Plant et al. 2018, GBC)
- Sub-mixed layer accumulation of oxygen (Riser and Johnson 2008, Nature)
- Subduction events (Llort et al. 2018, JGR-Oceans)

More uses than we originally thought of.

While all of them benefit from more accurate data, not all of them require the same level of accuracy, i.e., accuracy is subjective.

Real-time data flagging must permit such different uses, by separating subjective from objective qualification as much as possible.

Subjective qualification like: very accurate, less accurate, or quite inaccurate
Objective qualification like: failure/malfunction or not

For the subjective qualification, we’d need a continuous qualifier (to accommodate the wide range of expectations). For the objective qualification, we’d need a discrete qualifier (it worked or not).
<table>
<thead>
<tr>
<th>Flag</th>
<th>Meaning (measurement quality)*</th>
<th>Real-time-comment</th>
<th>Delayed-mode comment (following data analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No QC was performed</td>
<td>No QC was performed</td>
<td>No QC was performed</td>
</tr>
<tr>
<td>1</td>
<td>Good measurement</td>
<td>All Argo real-time QC tests passed, i.e., sensor and processing worked as intended. Inaccuracies can still be present in BGC data.*</td>
<td>The sensor measurement and processing worked as intended and the adjusted value is statistically consistent. A statistical error estimate is supplied.</td>
</tr>
<tr>
<td>2</td>
<td>Probably good measurement</td>
<td>Probably good measurement, i.e., sensor and processing worked probably as intended.*</td>
<td>Probably good measurement. An adjustment has been applied, but &lt;???&gt;. A statistical error estimate is supplied.</td>
</tr>
<tr>
<td>3</td>
<td>Probably bad measurement that is potentially correctable</td>
<td>A flag ‘3’ may be assigned by an operator during additional visual QC for probably bad measurements that may be corrected in delayed-mode, e.g., severe sensor drift.</td>
<td>The sensor measurement is probably bad. An adjustment has been applied, but the value may still be bad. For BGC-Argo, this flag should not be used in delayed-mode with preference for flags ‘4’ or ‘1/2’.</td>
</tr>
<tr>
<td>4</td>
<td>Bad measurement</td>
<td>A flag ‘4’ may be assigned by an operator during additional visual QC for bad measurements that are uncorrectable, e.g., sensor failure or severe biofouling</td>
<td>Bad measurement. Not adjustable. Data replaced by FillValue.</td>
</tr>
<tr>
<td>5</td>
<td>Value changed</td>
<td>Value changed</td>
<td>Value changed</td>
</tr>
<tr>
<td>6</td>
<td>Not used</td>
<td>Not used</td>
<td>Not used</td>
</tr>
<tr>
<td>7</td>
<td>Not used</td>
<td>Not used</td>
<td>Not used</td>
</tr>
<tr>
<td>8</td>
<td>Estimated value</td>
<td>Estimated value (interpolated, extrapolated, or other estimation)</td>
<td>Estimated value (interpolated, extrapolated, or other estimation)</td>
</tr>
<tr>
<td>9</td>
<td>Missing value</td>
<td>Missing value</td>
<td>Missing value</td>
</tr>
</tbody>
</table>

**Suggested enhancements to QC flag definitions**
Argo quality control: back to basics

Thoughts on squaring the circle for BGC QC

Matt Donnelly (BODC, NOC)
A clean slate...

Imagine we were starting with no decisions already implemented...

We want to implement what makes long-term sense...

Ignore (most of) what already exists!
Assumptions: User Motivations

• There are different types of users, broadly:
  • **Operational** – who largely need data quickly for short-term, ephemeral purposes
  • **Research** – who largely need accurate data for long-term, policy-informing purposes

• Users will have different focuses:
  • Some will work **within broad Argo science mission** objectives
  • Some will work **outside broad Argo science mission** objectives
  • Increasingly engaged in **inter-disciplinary** activity requiring multiple types of data

• Irrespective of motivation, users need to be able to:
  • **understand** both the data and quality control
  • **use** the data and quality control quickly and efficiently
Assumptions: Operational users

• Require data that is ‘good enough’ – within climatological expectations

• Need it quickly (<<12 hours) from as many platforms as possible

• Increasingly want more than physical data

• Need to be able to filter out definitely ‘bad’ data

• Need as much useable data after filtering – otherwise nothing to use!
Assumptions: Research users

• Require data that is ‘science ready’ – to observe changes in climate

• Need a sustainable supply from the overall array

• Increasingly using many types of data for inter-disciplinary work

• Want maximum amount of accurate and precise data

• Need confidence in quality – otherwise will do their own QC!
Assumptions: User Behaviour

• We cannot and do not want to control who uses Argo data

• We cannot prevent data misuse

• We can try to minimise the risk of misuse by being:
  • clear about the meaning of Argo quality control
  • consistent in how we use that meaning throughout Argo

• Users need a data system that is:
  • maximally comprehensive – contains necessary and sufficient information
  • minimally complex – as simple as possible to understand

• Users would prefer uniformity across Argo – core, deep and BGC
Assumptions: Sensors

- **Core** Argo: CTD *likely good* at $P < 2000$ dbar and good for operational purposes unless proven otherwise

- **Deep** Argo: CTD *likely needs correction* at $P >> 2000$ dbar and real-time data not suitable for operational purposes without correction

- **BGC** Argo: all sensors *likely need correction* and real-time data not suitable for operational purposes without correction

- All sensors *need delayed-mode QC* before being ‘research-ready’
Assumptions: Data System

• We will continue to have:
  • **R-files** - real-time
  • **D-files** - delayed-mode

• We will continue to have:
  • **R-mode** data - real-time unadjusted, or ‘raw’ – present in R-files and D-files
  • **A-mode** data - real-time adjusted – present in R-files only
  • **D-mode** data - delayed-mode adjusted – present in D-file only

We will continue to have a data flag scheme as now:
  • with the **same basic flag meanings**
  • but scope to **adjust** their definitions? (no promises!)

There are other QC indicators:
  • RP2 – profile quality flag, R6 data state indicators, R11 QC test binary IDs
Question: What is ‘good enough’ in R-mode?

Good data needs to be ‘good enough’ for operational use

EITHER: we need to **universally flag** sensors known to likely be bad by default

OR: we need RT-tests that are **robust and sufficient** to allow a **good flag**

If the latter is worthwhile, can we implement tests that are quick to perform and easy to implement at DACs?

- e.g. brutal min/max climatology comparison?
- e.g. initial gain test?
- Either way, no correction for R-mode data anyway
Question: What is ‘good enough’ in A-mode?

Still R-files - good data needs to be ‘good enough’ for operational use

EITHER: we remain dependent on undertaking DMQC to make a future A-mode adjustment

OR: we introduce a mechanism to make rapid non-DMQC adjustments possible for BGC

How do we sustainably scale BGC assessments and corrections for every parameter across the entire fleet?
Question: What is ‘science-ready’ in D-mode?

Good data need to be ‘science ready’ for research

EITHER: it is our current best assessment subject, potentially subject to major review?

OR: it is our quantified assessment with associated error that can be used with confidence, but subject to further improvement?
Question: Are the current QC flag definitions sufficient?

<table>
<thead>
<tr>
<th>Flag</th>
<th>Meaning</th>
<th>R-mode</th>
<th>D-mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No QC was performed</td>
<td>No QC was performed</td>
<td>No QC was performed</td>
</tr>
<tr>
<td>1</td>
<td>Good data</td>
<td>All Argo real-time QC tests passed</td>
<td>The adjusted value is statistically consistent and a statistical error estimate is supplied.</td>
</tr>
<tr>
<td>2</td>
<td>Probably good data</td>
<td>&lt;empty&gt;</td>
<td>Probably good data.</td>
</tr>
<tr>
<td>3</td>
<td>Bad data that are potentially correctable</td>
<td>Test 15 or Test 16 or Test 17 failed and all other real-time QC tests passed. These data are not to be used without scientific correction. A flag ‘3’ may be assigned by an operator during additional visual QC for bad data that may be corrected in delayed mode.</td>
<td>An adjustment has been applied, but the value may still be bad.</td>
</tr>
<tr>
<td>4</td>
<td>Bad data</td>
<td>Data have failed one or more of the real-time QC tests, excluding Test 16. A flag ‘4’ may be assigned by an operator during additional visual QC for bad data that are not correctable.</td>
<td>Bad data. Not adjustable.</td>
</tr>
<tr>
<td>5</td>
<td>Value changed</td>
<td>Value changed</td>
<td>Value changed</td>
</tr>
<tr>
<td>6</td>
<td>Not used</td>
<td>Value changed</td>
<td>Value changed</td>
</tr>
<tr>
<td>7</td>
<td>Not used</td>
<td>Not used</td>
<td>Not used</td>
</tr>
<tr>
<td>8</td>
<td>Estimated value</td>
<td>Estimated value (interpolated, extrapolated or other estimation).</td>
<td>Estimated value (interpolated, extrapolated or other estimation).</td>
</tr>
<tr>
<td>9</td>
<td>Missing value</td>
<td>Missing value</td>
<td>Missing value</td>
</tr>
</tbody>
</table>
Challenges: Achieving the goal

• Each sensor and parameter has **unique features** but we need a **consistent data system**

• **No one** has the resource to do everything

• Definitely no one has the resource to **re-do** everything!

• How do we **prepare** for the next ten years – e.g. new sensor models?

• How do we **work together**? How do we support DACs?

• How do we **share knowledge**, without being complete experts in everything?
Finally...

- Assume we can agree a way forward

- Do we need better comms with users?

- Could a video on quality control - R/D files, R/A/D mode, and QC flags - cover in 5 minutes what it might take a users hours to fully grasp from the documentation?
2. Progress on parameter 'decision trees'

- A RT ‘roadmap’ for DACs
- Lays out the processing tests and related flagging protocols for automated RT processing and adjustment of profile data
- Should be outlined in detail within documentation
- Progress made on CHLA, NITRATE, DOXY (pH, BBP, Radiometry?)
Proposed work flow for RTQC of CHLA variable

May 2020
Christina Schallenberg & Tom Trull
in consultation with Bozena Wojtasiewicz, Pete Strutton, Catherine Schmechtig, Raphaëlle Sauzede, Josh Plant
Overview

This slide show contains the following:

1. The glossary of terms used in the flow charts
2. The proposed flow charts along with outstanding questions that need to be discussed/decided
3. A discussion of spikes in CHLA and a proposal for how to deal with them
4. Additional, more general thoughts/comments/questions
5. Previous flow charts for reference
6. Argo QC flags
GLOSSARY

- **CHLA**: Chlorophyll-a (mg m⁻³) calculated with the factory calibration
- **CHLA_QC**: Quality flag for CHLA
- **CHLA_ADJUSTED**: Adjusted chlorophyll-a (mg m⁻³)
- **CHLA_ADJUSTED_QC**: Quality flag for CHLA_ADJUSTED
- **DARK_10**: Median (or minimum) of the minimum FLUORESCENCE_CHLA from 10 initial profiles
- **DARK_CHLA**: Manufacturer-supplied dark (counts) for the calibration equation transforming FLUORESCENCE_CHLA into CHLA
- **DARK'_CHLA**: Adjusted dark (counts) used to calculate CHLA_ADJUSTED; uses DARK_10 if it exists
- **FLUORESCENCE_CHLA**: Fluorescence signal (counts), uncalibrated
- **iProf**: iᵗʰ profile measured
- **LAST_DARK_CHLA**: Stored value of the last dark (counts) that was used to calculate CHLA_ADJUSTED *(may be obsolete with new DARK_10 approach?)*
- **MLD**: Mixed Layer Depth
- **SCALE_CHLA**: Manufacturer-supplied slope for the calibration equation transforming CHLA; also used for calculation of CHLA_ADJUSTED
- **ZMaxFluo**: Depth of the fluorescence maximum
Global range test should be done early on (i.e. before dark determination) to exclude outrageous negative values.

New limits reflect the Roesler factor.

Should the lower level (-0.2) be set even more generously, considering that the factory dark that was used to calculate CHLA might skew values towards the negative?
Questions/comments:

• Define "sufficiently deep" profile depth: 900 m?
• Define MLD criteria: MLD + 200 m?

• **Note that finding the dark of the first 10 profiles involves first doing a running median on each of these profiles, and then finding the minimum on that**

• Should DARK_10 (suggested by Bozena to be based on 10 profiles rather than 5; see figures on slide 11) be ultimately the minimum or the median of the 10 darks?

• How should we assess the quality of DARK_10? Based on the standard deviations of the 10 individual darks, or based on whether it stays close enough to the factory dark, or both?

• For all the cases where DARK_10 can’t be determined, should the QC remain at 3, or should it be elevated to 2?
Questions/comments:

- How should we define the “substantial change” between DARK_10 and the dark of iProf? What should the threshold be?
- Should CHLA_ADJUSTED_QC be set to 4 because of the failed dark test? Or should it stay at 3?
- See next 3 slides for a discussion of spikes and spike removal.
- The slope adjustment represents a change in all values. Should all CHLA_ADJUSTED be flagged 5, i.e. “changed value”? What is done for other variables such as NO₃ when offsets are corrected?
Spike issues...

- Traditionally, negative spikes were flagged in both CHLA and CHLA_ADJUSTED. However, the first round of reviews of the updated CHLA QC document in March/April indicated issues with the spike test:
  - Difficulty to reliably distinguish noise from negative spikes
  - Difficulty defining a reliable reference window for the running median given varying sampling resolutions between floats and even within profiles (e.g. more coarse sampling at depth)

- Furthermore, the presence of spikes (or rather: how to deal with them) presents a conundrum when dealing with the NPQ correction:
  - Should spikes be “added back on” after NPQ correction, and if so, should an NPQ correction be attempted on them? How could that be achieved?
  - If spikes are not put back after NPQ correction, we’ll end up with a very skewed product where parts of profiles contain spikes (i.e. any non-NPQ parts), and other parts do not (the NPQ-corrected parts). In addition, there would be an artificial difference introduced between daytime (some spikes removed due to NPQ correction) and night-time (no spikes removed) profiles.
What exactly is behind spikes in the CHLA variable? Are they even Chl-a/phytoplankton?

Like Tanaka et al. (2019), who used cameras looking at an fchl and $b_{bp}$ sensor, we confirmed that spikes in FDOM, $b_{bp}$, and fchl could originate from zooplankton. Spikes in backscattering have long been associated with the presence of large particles. Spikes in FDOM can be explained by the fact that fluorescent proteins are possessed by a range of marine organisms and that both humic- and amino acid-like fluorescence are produced by zooplankton grazing and excretion (Stedmon & Cory, 2014). Spikes observed in fchl co-occurring occasionally with the layers of spikes observed in FDOM and $b_{bp}$ channels are harder to explain. A possible explanation could be the fluorescence from the gut of the zooplankton. However, it would likely be observed only for a short period (several hours) after the time of grazing which isn’t the case in our data set. Spikes in fchl could also be caused by the same fluorescent compounds seen by the FDOM sensor (Xing et al., 2017) or by an out of band response given the near saturation of the backscattering channel when measuring spikes which uses the same detector but a different light source.

Resulting issues with CHLA spikes:
If they’re not Chl-a in the sense of live phytoplankton biomass, then

- The slope factor is unknown
- Validity of NPQ correction is unknown

Proposed solution:
- If you are interested in spikes, use the CHLA variable. Spikes there are untouched.
- Get rid of all spikes in CHLA_ADJUSTED, both day and night (see next slide for more)

From Haëntjens et al. 2020: Detecting mesopelagic organisms using biogeochemical-Argo floats

None of the discussed phenomena that might be behind spikes in CHLA are actually living phytoplankton.
Advantages of RUN_MED(CHLA_ADJUSTED)

• Treats all CHLA_ADJUSTED the same, i.e. no skewing of spike counts from NPQ correction

• Preserves data (rather than losing a data point due to a spike flag, you end up with an ”interpolated” point) → less holes in the data, which may be significant for low-res profiles
  • Does that mean CHLA_ADJUSTED_QC should be 8 for “interpolated”? QC=5 probably covers all the bases, no?

• Consistent with what we know/don’t know about how spikes should be treated with respect to calibrations and adjustments (NPQ, slope factor)

• Consistent with the notion that spikes likely (or most often) do not represent living phytoplankton

• No need to improve the (difficult) spike test, i.e. less room for error there
More comments and questions

• In the proposed, CHLA will mostly be QC = 3, reflecting the fact that the factory dark is seldom the same as the in situ dark. Since CHLA will always be calculated with the factory calibration, it will mostly be QC’ed as “suspect” simply because of the dark issue.

• In the proposed, CHLA_ADJUSTED_QC only gets QC=2 if the dark in iProf isn’t substantially different from DARK_10 – is that what we want?

• Quite a few tests and corrections depend on the ability to calculate MLD; what should we do in cases where S or T are flagged bad and thus can’t be used to calculate MLD? What QC flags should be used in CHLA_ADJUSTED_QC for those cases?

• Are we sure we don’t want to do an in situ dark correction on CHLA? Really sure?

• The proposed is all meant for RTQC (and adjustments) only. There are obviously special cases arising from some floats that will need to be reviewed in DMQC. But these are beyond the scope of RTQC.
About using the first 10 profiles for the dark estimation...

Some figures from Bozena that illustrate an exponential decline in the dark at the beginning of a float’s life.

Top figure: Fluorescence minima from a float that cycled several times per day.

Bottom: A similar exponential decline was observed in the bbp minima at two wavelengths. Colours indicate profile depths.

Sensor: MCOMS
Float ID: 5904923
On-float darks over time, examples from MCOMS sensors on NAVIS floats:

On-float darks over time, examples from FLBB sensors on APEX floats:
Flow charts from version 1.1 of the CHLA QC doc

1. CHLA = (RAW_CHLA - DARK_CHLA) * SCALE_CHLA (Eq 1)
   LAST DARK CHLA = DARK_CHLA (Eq 2)
   only for the first profile

2. CHLA_ADJUSTED = (FLUORESCENCE_CHLA - DARK_CHLA) / SCALE_CHLA (Eq 6)
   Range Test (Test 5)
   -0.1 mg/m³ – 50 mg/m³

3. Negative Spike Test
   RES = 2 – median(V0, V1, V2, V3, V4) (Eq 8)
   RES > 2 * percentile10 (RES) (Test 6)

4. FLAG_MLD Test 7
   No NPQ correction

5. NPQ Correction
   depthNPQ = depth[max(median(CHLA_ADJUSTED), V0, V1, V2, V3, V4)] (Eq 9)
   depthNPQ is within 0.9 MLD and depths where RES > 2 * percentile90 (RES) are ignored in the estimation of depthNPQ (test 8)
   CHLA_ADJUSTED[0, depthNPQ] = CHLA_ADJUSTED(depthNPQ) (Eq 10)

6. CHLA QC (0, depth NPQ) = 1
   CHLA_ADJUSTED QC (0, depth NPQ) = 5
## Argo QC flags

Table 1: BGC Argo QC flags and their meaning

<table>
<thead>
<tr>
<th>Flag</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Not evaluated, raw data</td>
</tr>
<tr>
<td>1</td>
<td>Good data; data have passed the highest level of quality control</td>
</tr>
<tr>
<td>2</td>
<td>Probably good data</td>
</tr>
<tr>
<td>3</td>
<td>Data suspect or of high interest, potentially correctable</td>
</tr>
<tr>
<td>4</td>
<td>Bad data</td>
</tr>
<tr>
<td>5</td>
<td>changed value (NPQ flag)</td>
</tr>
</tbody>
</table>
QC and Adjustment
Decision Tree for CHLA
(not finalized...)

Initialisation
- initial ADJUSTED_QC = 3
- initial QC = 3
- if iprof = 1 => deepdelta Offset = 0

Range
- QC = 4

Test profile depth
- If shallow profile (if profile_depth < 500)
  - CHLA_ADJUSTED = CHLA - deepdelta offset
  - if iprof > 5 ADJUSTED_QC = 1
  - if iprof < 5 ADJUSTED_QC = 3
- If deep profile
  - deeply mixed (if MLD + 200 > profile_depth)
    - CHLA_ADJUSTED = CHLA - deepdelta offset
    - if iprof > 5 ADJUSTED_QC = 1
    - if iprof < 5 ADJUSTED_QC = 3
  - not deeply mixed
  - negative despiking
    - Minimum estimation (median filter)
      - Median comparison
        - if abs(DARK - median(DARKs)) >= 0.2 * median(DARKs) => sensor drift?
        - CHLA_ADJUSTED = CHLA - deepdelta offset
        - ADJUSTED_QC = 2
      - Factory comparison
        - if DARK = DARK_FACTORY QC = 2
        - if DARK = DARK_FACTORY QC = 3
      - Median calculation
        - => deepdelta offset
        - CHLA_ADJUSTED = CHLA - deepdelta offset
        - if nprofiles < 5 ADJUSTED_QC = 3
        - if nprofiles >= 5 ADJUSTED_QC = 1

Range
- ADJUSTED_QC = 4

QUENCHING correction
- If NIGHT => STOP
- IF DAY
  - positive despiking
    - CHLA_ADJUSTED = max(CHLA_ADJUSTED(0.0 * MLD))
    - ADJUSTED_QC(0.0 * MLD) = 5
QC and Adjustment Decision Tree for NITRATE

- Work initiated in 2013 by K. Johnson
- Several iterations
  - In place in Coriolis, Aoml
- Documentation almost finished (minor edits)
  - There should be an official release of the documentation (S. Pouliquen)
- DM part to be completed
  - NITRATE_ADJUSTED_ERROR
  - Examples of different Ways to populate SCIENTIFIC_CALIB_xxx

RT ADJUSTED_QC = ‘2’, ‘1’?

To be discussed here
 QC and Adjustment Decision Tree for DOXY (work in progress...)

DOXY: **real-time** (same as in O₂ QC manual v2.0)

More work required:
- Include RT tests for O₂ & flagging scheme
- Include RT adjustment gain estimation: 1) WOA, or 2) last DM gain assessment, ie SAGE-O₂
- Organize into single diagram
- Decision tree for DM has also been started
QC and Adjustment Decision Tree for BBP, Radiometry
3. Formalizing a BGC reference database
BGC Argo needs a formalized reference database for standardized handling/storage of shipboard data used in calibration/validation

Questions initially posed by Lynne Talley (Scripps):
• Is there a part of the ADMT that deals with the shipboard data sets that are ingested for calibration?
• Is there a group in place that takes care of detailed oversight (compilation, formatting, QC, etc) for expanding (ie BGC) Argo?

Yes, this exists for Core-Argo, Deep-Argo, but not for BGC!
BGC Argo needs a formalized reference database for standardized handling/storage of shipboard data used in calibration/validation

Requirements:
• Centralized group to receive data from PIs post-cruise
• Stringent QC in line with needs of BGC Argo
• Formatting should be in support of reference tools used in calibration (and a plan for periodic updating of reference tools should be included)
BGC Argo needs a formalized reference database for standardized handling/storage of shipboard data used in calibration/validation

**GLODAP a logical partnership** (notes from Henry Bittig):
• Already an established pathway from CCHDO (and other centers) for BGC data
• Capacity for secondary quality control, standardized formatting
• Provides convenient, easy-to-use global products
• Actively improving their data handling for better automation, timely updates/releases of merged datasets
Additional details re GLODAP submissions (notes from Bob Key):

• Data centers (ie CCHDO) primary sources of data but many datasets also come direct from PI
• WOCE, CLIVAR, GO-SHIPS cruises considered the most important included in GLODAP, but are only a small percentage of the total cruises.
• Many cruises that go into GLODAP are very small (3-20 stations).
• GLODAP focuses on BGC + Tracer parameters. We let WOA deals with the T/S/O2-only data sets
• GLODAP will gather CTD data, but we have no CTD experts and do no formal processing

Note, still need support for bio-optics (hplc, poc) ...formalize a separate pathway?
What we used to do at the LOV to keep records of the data at deployment

CTD = x
CDOM = C
Fluorescence = F
Oxygen with optode = O
PAR = L
Transmissiometry = T

Then you click on the link and you’ll be asked for a password because of the cruise embargo (as many password as cruises) then you’ll get the data but not always in the same units, formats ....
On going work with ERDDAP (WMS, CSW) at the LOV

- All pigments (more than 25 types) analysed by the french service national of analysis of pigments (SAPIGH)
- All the relative CTD, Fluorescence, Characteristics of fluorometers...

⇒ Pertinent as the LOV got some expertise

Cruise with BGC-Argo Floats deployment or cruises of BGC interest
⇒ Time consuming (to get the data, to get the data QCced, to format the data, to maintain the server)
⇒ Embargo gestion?

⇒ Maybe not so pertinent to store at the LOV
Where do we go from here?

Required Actions:
1. Decide on tentative pathway(s)
2. Further discussion/agenda at ADMT21 (invite key players)

Other questions?
4. Propagating more adjusted data to the GDAC
Propagate more adjusted data to the GDAC’s

From Henry’s status directories (lots of useful info here):
https://biogeochemical-argo.org/cloud/document/implementati...en-status/

Great progress in 2020 but a ways to go!

★ Only ~60% of DOXY data is adjusted
Flag bad data

**DOXY audit** – A way to help data managers (DMs) do this

- Compare all BGC Argo surface DOXY to Word Ocean Atlas O2 % sat
- Identify statistically anomalous profiles (MAD & Z score)
- Report anomaly list
- Requires DM participation (Agree / disagree profile with assessment)
List example

<table>
<thead>
<tr>
<th>DAC &amp; PI (float owners)</th>
<th>Profile Gain (WOA % sat/ DOXY % sat)</th>
<th>Z scores (# of std dev from mean)</th>
</tr>
</thead>
</table>

New lists published (grey list check, added data mode to list) at:
High rates of false positives in S. Ocean, Med. Sea, upwelling zones
Have only heard back from a couple folks, is this a useful product?
What to do about older floats?
Correction factor (G) generated for most floats
DOXY_ADJUSTED = DOXY * G (if no other correction scheme available)
Pushing published data into the data system

Why?
Ken’s introduction of the BGC-Argo workshop in Villefranche
⇒ Prevent multiplication of offline databases of corrected data
⇒ Put back into the data system published data that are
  - unbiased
  - visually inspected
  - grey listed (spikes, drift, ...)

We organize an OFFLINE MEETING at Villefranche (Herve, Josh, Uday, Christina)
⇒ Produce DM Files For CHLA and BBP (without waiting for official documentation, but referring to publications)

- Work begin in December 2019 for coriolis
- Some issues with the NPQ correction (sometimes worse than the « A » mode)
- Produce only BBP in « D » mode
Pushing published data into the system

For Floats dead before the 1st January 2019 (Bellacicco et al., 2019) ~ 94 floats (CTS4 and CTS5)
=> production of BD files (routines available on github: https://github.com/qjutard/chl_bbp_ttt)

- From greylist BBP700_ADJUSTED_QC => 3,4
- Not in range test [-0.00005 ; 0.1] => BBP700_ADJUSTED_QC = 4
- Report BBP700 into the BBP700_ADJUSTED (except for 1 with an OFFSET, 4 more with an offset but not in BD)
- BBP700_ADJUSTED_ERROR = MAX(3 counts * SCALE, 20% (BBP700)) => Will be improved with giorgio’s work
- Final check of the time series (routines available on github: https://github.com/qjutard/time_series_plot)

Example for 6901485, profile 171
- SCIENTIFIC_CALIB_COMMENT : BBP700 delayed mode adjustment following the work done by M. Cornec in Bellacicco et al. 2019 (http://dx.doi.org/10.1029/2019GL084078)
- SCIENTIFIC_CALIB_COEFFICIENT : BBP700_OFFSET = -0.001026
- SCIENTIFIC_CALIB_EQUATION : BBP700_ADJUSTED = BBP700 - BBP700_OFFSET

Courtesy of henry, BBP at Coriolis February 2020
5. Additional ADMT topics

- Optode time-response
- Combined traj files
- Fine-tuning Sprof files
- Meta data
An update on
Oxygen optode time-response correction

Tanya Maurer & Yui Takeshita
MBARI
May, 2020
Problem:

- Optodes are slow!
- Response time of oxygen optode is not immediate
- Leads to data inaccuracy in high $O_2$ gradient regions (errors > 10 umol/kg)
- Response time varies among sensors types (pumped SBE63 faster than Aanderaa)
Response time (tau) can be explained by 1D diffusion model (Bittig et al. 2014, 2017)

- $O_2$ diffusion into the sensing foil controls response time
- **Diffusion Controlled by:**
  - Temperature
  - Boundary layer thickness (BLT) in front of the sensing foil at each sample
- BLT is determined by the flow regime
  - Pumped optode $\rightarrow$ higher flow, more rapid boundary layer renewal $\rightarrow$ shorter response time

- If you know flow rate and temperature, you can model overall tau
- Algorithm can then be applied to inverse filter the observational profile, producing time-corrected profile (see Bittig et al. 2014, 2017)
- Time stamps required!
Method tested on 182 MBARI APEX floats

- Sample time from Nitrate data file
- Plots below show time-corrected O2 data minus original
- Largest corrections occur in regions of high gradient (first panel)
Can compare first float oxygen profile (before and after time-correction) to shipboard cast taken at time of deployment.

Mean bias reduced by 1 µmol/kg across array

Original
Mean = 2.437
Std = 8.186

Time-corrected
Mean = 1.452
Std = 8.079
But...remaining problems:

1. For platforms with high-frequency sampling, noise can get amplified, esp in regions of low gradient (Bittig et al, 2014; Takeshita et al, 2020, in prep)
2. Errors in reconstruction increase with low-frequency sampling intervals

Both of these points are known, mentioned in Bittig et al, 2014. But, how to adequately address / account for them within global BGC Argo?
Problems:

1. For platforms with high-frequency sampling, noise can get amplified, esp in regions of low gradient (Bittig et al, 2014; Takeshita et al, 2020, in prep)

2. Errors in reconstruction increase with low-frequency sampling intervals

Filtering the data prior to correction a potential solution. Although, not as much of an issue for profiling floats.

5-point gives same Stdev as raw data

Image from Yui Takeshita, time response correction work with optode on spray glider
Problems:

1. For platforms with high-frequency sampling, noise can get amplified, esp in regions of low gradient (Bittig et al, 2014; Takeshita et al, 2020, in prep)

2. Errors in reconstruction increase with low-frequency sampling intervals

What does this mean for profiling floats?
Test how vertical resolution affects reconstruction error by applying time-correction to subsampled profiles

Black = spray glider O2 profile
Colors = modeled O2 profiles for Aanderaa and sBE63 at various taus

Work done by Yui Takeshita
Test how vertical resolution affects reconstruction error by applying time-correction to subsampled profiles.

Residuals of reconstructed profiles show high error in oxycline.

Work done by Yui Takeshita.
- Similar test done using Provor float
- See how result of Bittig time response correction changes with reduced vertical resolution.
- Down-sample a Provor profile and apply correction.
Subplot 2 (light blue) is similar to sample resolution of MBARI APEX floats.
With reduced sample resolution there is increased error apparent in the correction.
• For current BGC float configurations, resolution likely high enough to warrant correction (minimal bias in low gradient regions, and improved accuracy in high gradients)
For current BGC float configurations, resolution likely high enough to warrant correction (minimal bias in low gradient regions, and improved accuracy in high gradients)??

However, with reduced sample resolution there is increased error apparent in the correction

- For current BGC float configurations, resolution likely high enough to warrant correction (minimal bias in low gradient regions, and improved accuracy in high gradients)??
SUMMARY

• Implementation of time-response correction possible for most SOCCOM floats
• Test on SOCCOM array showed improvement to bottle-data validation statistics

**However, correction uncertainties depend on sample resolution**
• Details to resolve
  • How should error fields be updated to include uncertainties in this correction? Esp, with vertical resolution varying among platform type?
  • DM correction only or RT adjusted as well?
  • How to show when this correction has been applied → document in SCIENTIFIC_CALIB_*
  • What to do about profile end-points beyond where correction is mathematically possible? (insert uncorrected data?)
  • What to do for portions of a profile where data is bad/missing, thus sample resolution is modified? (ie should only correct full profiles?)
Combined (Core and BGC) Trajectory Files

- Approved by AST
- Sample v3.2 files available at Coriolis
  [ftp://ftp.ifremer.fr/ifremer/argo/etc/trajectoryBcCombined_SampleFiles/](ftp://ftp.ifremer.fr/ifremer/argo/etc/trajectoryBcCombined_SampleFiles/)
- File size of combined traj is ~25-50% bigger than stand-alone btraj
- Proposed format not final
- Production will require close coordination between core/bgc groups
- In-air optode data (used in deriving gain adjustment) stored in traj file
- Timeing info stored in traj file

Some formatting questions:
1. As there is no PARAMETER_DATA_MODE in the core PROF files, we should consider to duplicate DATA_MODE of the core PROF file to set the TRAJECTORY_PARAMETER_DATA_MODE of PRES, TEMP and PSAL parameters (for relevant MCs, i.e. those which refer to profile measurements (i.e. 190, 301, 503 and 590 in Coriolis Traj files)).

2. The JULD data mode should be provided explicitly. This could be done by:
   a. Adding JULD to TRAJECTORY_PARAMETERS (so that JULD data mode can be stored in TRAJECTORY_PARAMETER_DATA_MODE); or
   b. Adding a JULD_DATA_MODE (this is the best choice for me).

3. What should be the JULD data mode of an estimated time? (for which JULD = FillValue and JULD_ADJUSTED ~= FillValue).
The S-profile files

* Many thanks to Coriolis for making the S-profile files become a regular feature on the GDACs. They have improved BGC data usability tremendously!

Discussions are ongoing on how to improve some minor aspects of the S-profile files:

• Create a S-profile file even when the BGC sensor malfunctions and no valid data is transmitted for that cycle. This will allow the multi-cycle Sprof file to have a complete record of the float cycles, and not artificially create missing cycles.

• Replace qc=‘8’ with ‘4’ or ‘3’ if they are bounded by ‘4’ or ‘3’.

• Revisit how to select what data will enter into the construction of the synthetic pressure axis when there are CTD profiles with no accompanying BGC data, and how to interleave those P/T/S data into the synthetic profile.

~ from Annie Wong, Henry Bittig
Improving Meta Data

• Many inconsistencies exist in population of BGC metadata fields across the Argo data system
• Filling of PREDEPLOYMENT_CALIB* fields should be a priority
• Specification for the PREDEPLOYMENT_CALIB* fields still needed in documentation for NITRATE, PH

• pH sensor version numbers should be included in meta file
  • Store within SENSOR_MODEL (ie shift from DURA to DURA_V1, DURA_V1.1, DURA_V2 etc?)
  • Sensor manufacturer already described in SENSOR_MAKER field
  • Where to store description of what versions mean – likely in documentation