Summary of 2014 IMOS Quality Control Summit

Marty Hidas and Guillaume Galibert (eMII)

The QC Summit took place at CSIRO Hobart, on 5–7 November 2014. This year the aim was to combine QC efforts across IMOS facilities, focused on particular instrument types or parameters. However, most of the attendees were associated with ANMN.

This summary was written from an eMII perspective, but aims to record all key points discussed.

Current profiles (ADCP)

- Alessandra Mantovanelli (AIMS), Bec Cowley (CSIRO) and others have produced a report on details of processing and QC for data Acoustic Doppler Current Profilers (ADCPs), with recommendations for improving the final product.

- Ming Feng (CSIRO) talked about ADCP operations in WA, including comparisons between currents measured by ADCPs (at 10m depth) and by radar. Generally the agreement is reasonably good, despite the fact that ADCP data quality is usually worst near the surface.

Action items

- **eMII** to publish the final Mantovanelli et al. 2014 report on ADCP QC.

- **Guillaume** to work with the **ADCP QC working group** to decide which of the recommendations in the report should be implemented in the IMOS Matlab Toolbox.

- **Guillaume** to work with the **ADCP QC working group** to identify vertically bin-mapped variables from along-beam ones (not tilt corrected) and extend the QC and processing efforts to Nortek ADCPs.

Dissolved oxygen measurements

- Bronte Tilbrook leads a group at CSIRO who have set up a multi-point calibration system for dissolved oxygen (DO) instruments.

- The Seabird SBE43 sensors used on most ANMN moorings (part of the WQM package) are of poor quality and making the data usable would require much more work.

- Optode sensors are more reliable. They do drift a bit, but can be corrected.

- Bronte stressed the importance of performing pre-deployment checks.

Argo delayed-mode QC and data formats

- Esmee Van Wijk (CSIRO) described the tests applied in delayed-mode QC of Argo profiles. Many of these are performed manually by trained operators.

- "Metafiles" containing metadata about each float are essential for the QC process. They contain calibration, configuration and deployment information.

- Argo has a controlled vocabulary for instruments and sensors used on their floats.
- Argo has recently adopted a new NetCDF format for profiles, with separate files for physical and biogeochemical variables.
- Delayed-mode files include *ADJUSTEDERROR* variables to report estimated uncertainty.
- For delayed-mode data, all QC flags are either 'Good' or 'Bad'. Anything in between should be dealt with in processing.

Bio-optics, fluorescence

- Converting fluorescence (FL) counts to Chlorophyll a (Chl a) concentration is problematic. The relationship between the two depends on the plankton species present.
- Suggestion to develop regional calibration equations, along with uncertainty estimates for Chl a. Lesley Clementson and Tim Lynch (CSIRO) have done some initial work. We now have lots more data, just need someone to do the work.
- Lesley Clementson mentioned that it might be worth providing WQM products that would be averaged over two hours rather than one in order to match some satellites products.
- Suggestion to combine bio-optical QC effort between all IMOS Facilities that use such instruments (ANMN, ANFOG, Argo, SRS Ocean Colour)
- It was pointed out that biogeochemical variables in models can be out by factors of 10, so even observations with relatively high uncertainties provide useful constraints.
- A lot of discussion about the choice of sensors, their location on the moorings and distribution among the NRS stations.
  - Multi-sensor WQMs vs. independent sensors?
  - Fluorometers vs. radiation sensors?
  - Are the upper WQMs in the mixed layer so measurements can be compared to satellite observations?
  - Focusing bio-optical sensors on just 3 NRS sites along the east coast?

Data quality from Seabird profiling CTDs

- Val Latham (CSIRO) has been comparing salinity measured by CTD to the bottle samples and finding offsets up to ~0.06
- The comparison may not be entirely valid because the cast and water sample are not taken at the same time.
- Some sites have started using a rosette to sample.
- The switch to rosette needs to be captured in the metadata!

Action items:

- **Claire Davies** (CSIRO) to ensure switch to rosette sampling is recorded in the CSIRO BGC database.
- **Marty** to talk to Claire and Margaret Miller about harvesting metadata from the CSIRO BGC database, including collection methods and field trip details.

Re-processing data with the IMOS Matlab Toolbox

About once a year eMII ask the ANMN data processors to re-process all existing data with a new version of the Toolbox in order to improve the quality and interoperability of the entire collection. The sub-facilities find this to be an overly time-consuming task and have asked us to consider implementing a centralised process at eMII to perform the re-processing.

Marty and Guillaume's presentation focused mainly on making re-processing by the sub-facilities easier. We mentioned centralised re-processing as something we can work towards in long term.

Once properly configured, the Toolbox can easily re-process a large volume of data in batch mode with minimal intervention (e.g. Mark Snell in WA has demonstrated this). The main reasons this tends to become a more time-consuming task are:

- Missing or erroneous metadata in the deployment database (DDB)

- Incorrect Toolbox configuration

- Semi-automated and manual QC can't be automatically re-applied during re-processing

- Errors are often only picked up by eMII or data users, leading to further re-processing requests

A more fundamental issue is that each sub-facility has its own way of populating the DDB, so that each requires a different Toolbox configuration. Therefore our main suggestions were to standardise the DDB content and Toolbox configuration across all sub-facilities, and put effort into fixing up the metadata for all deployments to date.

Guillaume also pointed out that the sub-facilities have the oceanographic expertise, so they are ultimately responsible for validating the data (independent of where the processing happens).

The option of re-processing from the existing NetCDF files was suggested. This is a possibility, but only if the required raw data and metadata in those files is complete and correct.

Action items:

- **Deployment DB working group** formed, including person in charge of deployment DB at each sub-facility and Guillaume, led by Felicity McAllister (AIMS).

- **Deployment DB working group** to produce guidelines for populating DDB.

- **IMOS/eMII** to publish guidelines as an IMOS document.

- **Sub-facility operators** to fix and maintain accuracy & completeness of metadata in the deployment database

- (Suggested) Publish guidelines for IMOS pre-deployment procedures

- (Suggested) Publish guidelines for IMOS Post-deployment processing

- (Suggested) Create IMOS Log sheet templates, to be converted to digital documents

- (Suggested) Secure archive of log sheets, raw files and deployment databases

- **Guillaume** to define re-processing required for each instrument type (depending on Toolbox changes)

- **Guillaume** to enable re-processing of historical semi-automated and manual QC in the Toolbox

- **Guillaume** to provide visual validation of metadata and data content in the Toolbox (for

data processors)

- **eMII** to be at sub-facility's service to help with Toolbox

- **eMII** to implement IMOS NetCDF checker for content compliance

Toolbox & deployment DB updates

- AIMS have updated their deployment DB

  - add IMOS processed file name

  - links to sensor retrieval images

  - sensor IDs

- Guillaume summarised the Toolbox developments since the last Summit

Processing CTD casts

Guillaume described the whole process of processing CTD cast data from raw instrument files to NetCDF files ready to publish. He also ran a live demo of running the Toolbox in profile mode. He pointed out existing documentation for all stages of the process:

- Standard pre-run check, field sampling and post-deployment procedures described in two Ingleton et al. documents (both linked from the ANMN Documents webpage)

- AIMS has produced some Java code and a document to streamline post-deployment processing of CTD profiles, including

  - Batch processing with Seabird software (.hex to .cnv)

  - How to populate deployment DB

  - How to use the Toolbox

- Instructions for installing, configuring and running the Toolbox in profile mode are on the Toolbox wiki

Charles James (SARDI) told Guillaume and Marty about an issue with SARDI's CTD casts, many of which were done without a sufficiently long surface soak. He has identified pre-deployment best practices to avoid this in the future, and a new workflow to assess the impact of the surface soak (if any) and flag the data accordingly if necessary.

Action items:

- **Shaun Byrnes (AIMS)** to finalise post-deployment batch processing document and Java code

- **eMII** to publish AIMS document and make Java code available

- **Tim Ingleton (OEH, NSW)** to update surface soak procedure in the pre-deployment and field sampling document according to Charles James's suggestions.

- **Guillaume** to implement Charles' surface-soak QC checks in the post-deployment procedures.

- (Suggested) Look at CSIRO in-house post-processing workflow in Matlab (contact Lindsey Pender?)

Facilitating user uptake

Issues

- Binary files (some can't be converted to CSV)

- Large range of parameters

- Higher level products very limited

- Variable data quality & coverage

- Data streams can be complex, even when measuring simple things (e.g. ADCP for currents). User needs some minimal understanding to make sense of the data.

- Current facets not very useful for finding some data collections (e.g. People want to search for "The NRS biogeochemical data", not "self-propelled boat")

- When users report errors in data, fixes take a long time (still have to implement workarounds in code)

- A generic user (who has no knowledge of IMOS Facilities and just wants to download some data) and an IMOS data provider (who knows exactly which datasets from which deployments he wants to access to) don't have the same needs when it comes to interacting with the IMOS portal. At the moment, the portal is not catering for IMOS data providers.

Suggested solutions (Craig Steinberg, AIMS)

- Improve links from portal to documentation, including:

  - Deployment information (e.g. [AIMS deployment web pages](#)), cruise reports

  - Coverage and extent of data available

  - How & why measurements are made in a certain way (e.g. bursts, averages)

  - Details of processing

- Provide summary plots on portal that can be viewed before data download – "try before you buy"

- Provide higher-level products for timeseries (like OceanCurrent)

- Manage user expectations

- Provide access to raw data for experienced users

Action items:

- **Logs & reports working group** (newly formed) will determine how to standardise deployment/recovery cruise reports (including data summaries) and link them from the portal.

To make IMOS data ready for assimilation (David Griffin)

- Much higher level of QC

  - Compare to climatology

  - Compare to tides to check times

- Better consistency in file structure and metadata

- More complete and correct metadata

- Produce velocity products gridded in constant depth/height coordinates

Managing expectations

- Some users (e.g. modellers) want big chunks of data from different sources, perfectly quality-controlled and standardised so they can slurp it all in and crunch numbers without having to look at individual files.

- Many experienced researchers just want the raw data from the instrument (that's what they're used to handling), but they'll need some metadata to be able to make use of it.

- Facility operators want to know the gory details of each deployment and prefer to access it on a per-deployment basis.

Instrument recovery images

- Field operators have started taking photos of instruments at mooring recovery to record the amount of fouling on sensors.

- AIMS have added a new field in the deployment DB to contain a URL link to a recovery image for an instrument.

- AIMS convert the metadata in the DB into deployment summary web pages that now include the recovery image. These could be linked to from the portal or the metadata record. Other operators could create their own deployment summary pages.

- Alternatively, everyone could upload the images to eMII & include a link in each NetCDF file.

Manual QC

- Tim Austin (UNSW) showed some examples of QC issues that can only be picked up by comparing data from multiple deployments (i.e. couldn't easily be done in an automated way).

- It would be useful to be able to add a comment when manually flagging points in the Toolbox.

- It would be useful to have some geographic (isobaths) information on top of velocity rose-type plots.

Visualisation tools

- Need to visualise data at various stages from pre-deployment checks to validation of processed data

- Need standardised plots for IMOS reporting (if a data report template is implemented)

- AIMS has a Matlab plotting library to produce publication quality plots from NetCDF files. Other groups have their own tools. It would be useful to incorporate some of these in the Toolbox, or add them to the IMOS User Code Library.

- Roger Proctor (eMII) pointed out that eMII are working on a Data Tools page on the IMOS website, with links to such tools.

Time zone issues

- David Griffin's tidal analysis has revealed data sets in which the timestamps appear to be offset by up to 10 hours. These are likely due to
    - The instrument clock being set to local time, but timestamps being interpreted as UTC; or
    - A local time to UTC conversion being applied incorrectly or more than once.
- It is the operator's responsibility to ensure that
    - The instrument clock time zone (preferably UTC) is recorded correctly on the field logsheet and in the deployment database; and
    - If necessary, a conversion to UTC is performed correctly in the Toolbox.
- A QC procedure using tidal analysis to check timestamps could be implemented in a future version of the Toolbox (after 2.4).

Appendix: Working Group members

ADCP QC

Craig Steinberg (AIMS), Alessandra Mantovanelli (AIMS), Bec Cowley (CSIRO), Guillaume Galibert (eMII), Ian Darby (CSIRO) + one operator from NSW, SA & Tas.

Profiling CTD

Tim Inglegton (OEH), Val Latham (CSIRO), Ian Darby (CSIRO), Shaun Byrnes (AIMS), Ryan Crossing (CSIRO), Charles James (SARDI).

Deployment DB

Felicity McAllister (AIMS), Darren Moore (CSIRO), Ian Darby (CSIRO), Tim Austin (UNSW), Paul Malthouse (SA), Guillaume Galibert (eMII)

Logs & reports

Craig Steinberg (AIMS), Bec Cowley (CSIRO), + ??