IML QUALITY CONTROL

# Troubleshooting Log

### E. Chisholm, January 2019

In BCD2IML\_format.R line 20-24 were altered to read data files offline from BioChem after getting error prompt for BioChem Username and password.

File B\_batch\_BIO.m found in BIO QC toolbox updated from IML toolbox by Gordana Lazin.

Function ‘textcell.m’ found in CTD toolbox from Gordana Lazin

Date/time error, BCD2IML\_format.R producing NaN dates, edited line 105 to produce correct format, changed POSIXct format to ‘%d-%b-%Y %H%M’ from ‘%Y-%m-%d %H%M’

Function strmatchi.m missing, supplied by Gordana Lazin

Toolbox ‘SeawaterCsiro\_ver2\_0\_2’ for MATLAB required, available on shared drive. B\_batch\_BIO.m requires sw\_dpth.m and sw\_dens.m

Toolbox MLI\_CTD required for function ‘remove\_doubtful.m’ in B\_batch\_BIO.m. Available via SVN (from Gordana Lazin)

Error loading ‘east\_mask’ file in B\_batch\_BIO.m resolved by commenting out visual inspection (line 53-57) in ‘B\_control\_Q.m’

B\_batch\_BIO.m is producing blank files ‘QC\_tests\_performed’ and ‘QC\_data…’, resolved when visual inspection was commented out and tests were able to finish without errors.

Get\_flags\_IML2BCD.r needs to be modified to run offline from BioChem

AZMP\_mission\_name\_descriptor.csv file relates mission descriptors and cruise names but only for AZMP cruises, would be very difficult to replicate for entire dataset. Modified script to pull Mission descriptor for file names.

Cruise 189099001 script could not map temperature data type, added ‘Temp\_CTD\_1968’ to BCD\_IML\_MAP.csv

Added ‘O2\_Winkler’ data type to BCD\_IML\_MAP.csv, otherwise data from BCD does not get mapped to IML file

Some loops in BCD2IML\_format.r are generating warnings about NA (invalid factors) being generated, does not seem to critically damage script but does produce empty files. Had to initiate new variables as factors, resolved.

DIS\_SAMPLE\_KEY\_VALUE is formatted differently for various cruises. Gordana’s script assumed it would be consistent (‘missiondescriptor\_event\_sampleID’) in order to run through script G and Emily decided to create sample key value from pieces at beginning of formatting from other data columns.

Read\_btl\_txtfile.m is creating an empty dataframe, tried many different solutions… Some files were affected and would not run but some files were running through just fine, trying to determine the cause.

Looked at line 85-112 which matches CTD data to lab data, this is where data frame was being overwritten with empty columns

Tried commenting out line 112 but this created an error in B\_create\_btl.m

Tried mimicking other data files which were running fine, matched data dimensions and column names

Tried multiple sample key value formats including starting all with a letter, some started with an underscore and when this was rectified a few more files did run through

Thought it might be something involving the data types not mapping to IML

**Files without pressure data are unable to map onto correct depth bin, causing an error in IML script**

**Added a column zbouteille, following IML standards and matching it to BCD column, ‘START\_DEPTH’, ran files successfully through BCD2IML\_format.r and read\_btl\_txtfile.m**

Need to write process to check that all data columns from BCD are properly mapped to IML

Mapped all outstanding DATA\_TYPE\_METHOD fields which can be QC’d (list from Gordana), used reference guide to ensure proper matching units, check by GL before final version

Issue where multiple BCD data types map to same IML code, causing errors when R script attempts to map back to BCD after QC.

Solution was to add line of code which isolates BCD\_IML\_map to only the data types present in that particular mission (assuming there will be no duplicate data types within mission)

Will need to convert mmol/m to ml/l to compare against IML QC standards

Email contact with Caroline LaFleur from IML

You can keep your oxygen data in mmol/m\*\*3. You just have to create a new variable in the data\_btl.m file with the correct conversion factor. The quality control procedure will then be done in ml/l but the data will remain mmol/m\*\*3 in your BTL\_\* file.

For example: xx is the btlLIST length +1

btlLIST(xx).name = 'bottle sample oxygen';

btlLIST(xx).units = 'mmol/m\*\*3';  %original units

btlLIST(xx).type = 'labo';

btlLIST(xx).code = 'OXYM\_';   %new variable code in the BTL\_\* file

btlLIST(xx).gf3 = 'DOXY';     %gf3 code stays the same

btlLIST(xx).btl2gf3 = 1/44.66;   %Conversion factor from mmol/m\*\*3 to ml/l

btlLIST(xx).decimal = 3;

btlLIST(xx).method = {'OXYM\_01';'OXYM\_02';'OXYM\_XX';'OXYM\_03'};

btlLIST(xx).desc = {'Winkler dissolved oxygen titration method: Carpenter (1965) and Carrit and Carpenter (1966)';...

      'Winkler automated dissolved oxygen titration method: Jones, Zemlyak and Stewart (1992)';...

      'dissolved oxygen: unknown method';...

      'dissolved oxygen measured by laboratory electrode'};

Added the DATA\_TYPE\_METHOD O2\_Winkler\_Molar to BCD files, changed O2\_Winkler data to O2\_Winkler\_molar if data maximum per cruise was greater than 14 (this could likely be refined using a visual inspection process)

O2\_Winkler\_Molar maps to IML code OXYM\_01 which is converted using factor of 44.66 to ml/l before being run through QC.

~~Is currently not running properly~~

O2 molar variable and flag had to be added into B\_addQ2btl.m in order for OXYM flag to be created, once this issue was solved it became apparent that data was not being converted to ml/l before QC and was producing globally impossible value flags. UNRESOLVED

56 missions with OXYM but missing Q\_OXYM, QC is still processing OXYM variable, flags are being stored in QCFF.

Error when running through MATLAB QC where S/T dimensions were not matching.

Removed duplicated PSAL codes from BCD\_IML\_map, so only one salinity value per IML file, solved problem!

Error in IML QC MatLab related to dimensions of data

Index exceeding matrix dimension at end of quality control when attempting to merge two data frames (one with data, one with flags). Thought it might have to do with adding the new OXYM variable but some files with OXYM processed fine, some threw errors. UNRESOLVED

When running IML QC scripts, there was a recurring error coming up related to datestr (MATLAB was unable to convert date values into characters) discovered this was due to NaN values in IML format file dates. This was traced back to NA values in original data set time values (when combing date and times in BCD2IML scripts, NA was propagating). Total 5 cruises affected.

Resolved by changing all NA time values in original data set to zero

If R not reading IML format tables correctly (error about number of elements per row, ensure using comment.char = ‘’ in read.table() to disable interpretation of comments, also ensure stringsAsFactors = FALSE to ensure character variables can be read.

To resolve lack of conversion between mmol/m\*\*3 and ml/l in IML script, conversion was done manually before QC. Entered into R script as function convertOxy(), has the ability to temporarily convert data points from mmol/m\*\*3 to ml/l , using standard conversion of 44.66.

Caroline had also suggested a solution to the conversion, see files B\_addQ2btl\_CL.m and B\_create\_btl\_CL.m

Total 38 cruises could not run though processing scripts (20 errors in IML matlab scripts, 18 errors in getting flags through R scripts after QC).

Issue where some cruises did not appear to contain oxygen data was found to be an issue in the SQL code where if a cruise measured oxygen the entire mission was included but then data points were excluded based on geographic boundaries so some cruises were included who did not collect oxygen data within geographic boundaries. These cruises were excluded from QC

Issue in get flags R script where data frame dimensions were not matching isolated to script not being able to find QC flags.

Edit to line 81 to generalize formula for excluding lat, lon, date, time and not exclude any data or QC flag columns.

Matlab error in QC,

```

Test 3.6 -> phos in box(es) 8 is out of Nutrient Climatology at 400.0 m

Index exceeds matrix dimensions.

Error in B\_stage3\_Q (line 207)

fprintf(fid\_data,form,T.filename,Tn.pres(i),T.uniqueno(i),test36.name{j},par(i,J),['Out of Nutrient

Climatology in box(es) ' P.box ' (Min=' num2str(par\_climato\_min,'%.2f') ' Max='

num2str(par\_climato\_max,'%.2f') ' n=' num2str(par\_climato\_no,'%.0f') ')']);

Error in B\_control\_Q\_GL (line 47)

S(i)=B\_stage3\_Q(S(i),36); %Test 3.6: Brickman Monthly Climatology (NTRZ, PHOS, SLCA)

```

Resolved by removing nutrient columns from data frame (unnecessary for processing when focused on oxygen)

Successfully ran all cruises producing errors in R except 18PZ96027 which does not appear to contain oxygen data

**Parsed errors down to 9 cruises:**

* Not containing oxygen
  + 18HU94008
  + 18HU95011
  + 80029
  + 18PZ96027
* “Qiden” error putting matlab btl and QC together
  + 18NE95008
  + 18VA69002 (nutrient dimension error resolved)
  + 18VA70002
* Other
  + 18PZ94029 (non-unique samples)
  + 18PZ97042 (CTD doxy in mmol/m\*\*3)