**WORKFLOW – from Peter Yates**

This document outlines the workflow of these analyses:

Last updates on 5 June 2019

All code are located in C:\Users\Peter\OneDrive - Sydney Institute of Marine Science\OPCmodelling\2code

All code are run from a single R Studio project: C:\Users\Peter\OneDrive - Sydney Institute of Marine Science\OPCmodelling\OPCmodelling

To replicate these analyses, the code needs to be run in a particular order.

This is because the data.frame output from some are then the input of another.

Folders called 'old' are superceded copies of code, as indicated by the six-digit date in the file name.

Code runs in this order:

(1) CTD overlap: this takes the OPC data CTD data from MNF\_SS200408\_ctd\_trawler.csv,

does some processing (creating new variables etc), and outputs a new data set (that actually isnt used again),

secondly it goes on investigate the spatial and temporal overlap between OPC and CTD

Finally it creates OPC\_CTD discrepency 120419.csv for merging with the outputs later on. This is the distance and sal/temp differences btween OPC and CTD data streams.

(2) HYD processing and overlap: this is the equivalent of CTD overlap but for the hydrology data. This saves 'HYD cleaned 080419' that is used in HYD interpolation.

(3) CTD interpolation: This takes the output OPC data from from 'CTD overlap' (does it? actually it reads the raw data from JE but follows the same steps as CTD overlap.), along with the CTD netcdf files.

It generates and appends interpolated CTD data to the OPC data.frame, and saves it.

(4) HYD interpolation: This takes output from 'HYD processing and overlap' and 'CTD interpolation', and the CTD data (as csv)

It generates and appends the interpolated HYD data to the dataframe that contains OPC and CTD data.

The output is a dataframe that has integrated OPC, interpolated CTD and interpolated HYD data.

(5) ADCP processing: this takes the raw ADCP csv file from MNF and does some pre-processing.

It filteres the data collected within the duration of OPC tows, and assigns a location.

(6) ADCP interpolation:

This takes the output from 'ADCP processing' (ADP\_tows\_final) and HYD/CTD interpolation.

It does the ADCP interpolation and adds values of U and V to the amalgamated dataset.

(7) MISC variables intergration:

Adds the following:

OPC\_CTD\_HYD\_ADP, discrepancy from 'OPC\_CTD discrepency 120419.csv', slope, distance from shore, GA depth :)

(8) Exploration:

This is code for data exploration and cleaning.

It is located in the OPCmodelling folder (1 level up, its a Rmd thing workspace thing...)

This is for all explanatory and response variables, all sites.

Outputs CleanedData\_xxxx19.csv

Also produces a html of plots etc

(9) Biomass\_model selection

inputs CleanedData\_xxxx19.csv

This is where Peter left off

You will see there are lots of similar files here.

Peter was cross-roads with how to proceed with these analyses (i.e. ommit rows with NA and proceed with AIC-based model selection)

..or alternativley to keep all the observations (including with NAs) and just present the full model.

..how to deal with site CB which has no CTD data (i.e. no fluorescence)

..which variables to include (the team requested re-instating oxygen/nitrate despite data exploration processes.

..so as you see there are lots of decions that have not been resolved. These are up for discussion with the team.

..the best course of action would be to take re-assess with the team and start a new script (all the code you need should be in the various versions of CleanedData\_xxxx19.csv)

..also see 'Statistics discussion with Ben Maslen' in folder '7stats consults and meetings' for notes about confounding, collinearity and interactions.

(10) 'LM and GLM 210519'

Also unresolved is whether to use LM, GLM, GAM, GLMM, GAMM. Peter's recomendation was GAM.

Upon request from the team, code for additional work using LM and GLM can be found in 'LM and GLM 210519'

It is located in the OPCmodelling folder

...Then the process will need to be repeated for the other dependent variables.

Peter is happy to assist if you have questions.

peter.yates2@my.jcu.edu.au