

Guide to process and convert Particulate, Detritus and Phytoplankton Absorbance bottle data to Absorption using R

By Peter Toma, July 2021

**Department of Fisheries and Oceans
Bedford Institute of Oceanography
Dartmouth, Nova Scotia, CANADA**

**Ocean and Ecosystem Science Division
Supervisor: Dr. Emmanuel Devred
Science Information Officer: Diana Cardoso**

Updated July 20, 2021

For R version 4.0.4 (2021-02-15)

TABLE OF CONTENTS

1. PROJECT DESCRIPTION AND OBJECTIVE	3
2. METHOD AND CODE	3
3. DESCRIPTION OF EACH SCRIPT	6
4. SESSION INFO ("C:\R_WD\SESSION_INFO.TXT")	21
5. USER INTERFACE DIALOGS	23

1. Project Description and objective

The R script reads the Particulate(123456.txt) and the Detritus(123456p.txt) absorbance raw data files from spectrophotometer analysis of water samples from Niskin bottles on a CTD rosette, reads and matches metadata, converts to absorption, performs QC and writes a final format that includes metadata and QC flags with all sample ID's in one csv file for Particulate, one for Detritus and one for Phytoplankton, per mission.

The code is written as a series functions called by a main script. There are User Interface Dialogs that help read the required files and help with the identification of metadata incompatibilities. All known possible errors have been added to the User Interface Dialog.

The working directory is C:\R_WD\ and each "C:\R_WD\WORKS(1-7)" folder contains multiple report files (their file names describe the content) to facilitate problem solving if needed. All report files show the number of rows in the file name to facilitate reading. The report files listed and described in this manual are only the ones that are relevant for the most common issues that were encountered so far. This R code was used on historical data with differing data and metadata formats it is anticipated that recent and present data will have less issues in processing.

Objective: Mining and making available and useful BIO historical biological and optical data to bring new information on the state of the ecosystem on the Scotian Shelf and Labrador Sea

2. Method and Code

Note: Minimum 'need to know' to run the script it is highlighted in yellow.

This R codepack is a directory named "R_WD" that can be found within "Phytoplankton_Absorbance_to_Absorption_R_Work_Directory.zip". The "R_WD" directory must be extracted and copied in "C:\\" drive root; C:\R_WD before running.

The R script writes all files and reports in "C:\R_WD\WRITE\". It is recommended that the "WRITE" folder to be deleted or renamed if it exists each time before running the scripts to avoid file overwriting.

The User Dialog window can popup behind R Studio and R will appear frozen. If fatal errors occur R will close without saving the workspace; please save your work before proceeding.

Input Table: The file "C:\R_WD\Absorbance_to_Absorption_R_scripts_input_table_info.xlsx" contains the input table below:

Variable_name	Example
Cruise_folder_with_raw_data_txt_absorbance_files_particulate_123456.txt_and_detritus_123456p.txt	AMU2019-001 Lab sea
coded_cruise_name (known as 'cruise name' or 'cruise number')	AMU2019001
the HPLC file (must have HPLC in the file name) (xls, xlsx or xlsx, NOT CSV)	AMU2019-001HPLC - final.xlsx
the QAT file (must have _QAT in the file name) (xls, xlsx or csv, NOT xlsx)	AMU2019001_QAT.xlsx
the ODF file (must have .ODF file extension)	CTD_AMU2019001_001_01_DN.ODF
filter_diameter in millimeters (mm)	16.5
Broadband Internet Connection	Wi-Fi

Output Table:

The file "C:\R_WD\Absorbance_to_Absorption_R_scripts_output_table_info.xlsx" contains the output table below (only one column is shown):

Output table – RESULT_name:

RESULT_name
Preliminary statistics before processing that may require attention
The Absorbance new raw files 123456Particulate.txt, 123456Detritus.txt and 123456Phytoplankton.txt
The Particulate raw data files that DO NOT have 801 rows of data will not be processed; all values in all files will be set to -999
The Detritus raw data files that DO NOT have 801 rows of data will not be processed; all values in all files will be set to -999
All new Absorbance files list created including the ones with -999 for missing Detritus/Particulate pair and/or not 801 rows
Absorbance_raw_data_files_analysis – Phyto<0, Par<Det, Par>0&Det<0, Par<0, Par<0&Det>0, Par=Det, Par<0&Det<0, Det<0, Par=0, Det=0, Par=0&Det=0
The Particulate, Detritus and Phytoplankton Absorbance tables without metadata (HPLC and QAT needed columns, ODF header)
Absorbance table plots from WORKS2, without accounting for Volume(HPLC) and filter_diameter
The Particulate, Detritus and Phytoplankton Absorbance tables with metadata (HPLC and QAT needed columns, ODF header)
All Absorbance Sample ID's that were not found in HPLC and QAT in 2 separate files
The Particulate, Detritus and Phytoplankton Absorption tables with metadata (HPLC and QAT needed columns, ODF header)
Absorption table plots from WORKS5 accounting for Volume(HPLC) and filter_diameter
Absorbance vs Absorption 3D plots 2 png files in one folder for easy view before / after
QC Plots 443_to_490, CHLA_to_443, CHLA_to_550 and CHLA_to_670, 18 png image files - and QC log file
Phytoplankton Absorption table with metadata and QC columns
Phytoplankton Absorption table with metadata and QC columns and Comments column
Phytoplankton Absorption table from WORKS7 plots by QC flag color (png legend path = C:\R_WD\QC_plots_Colors_Legend.png)
Phytoplankton Absorbance and Absorption plot - Phytoplankton Absorption table from WORKS7 plots by QC flag color and Phytoplankton Absorbance from WORKS\RESULTS
Phytoplankton Absorption table from WORKS7 plots on leaflet map png, interactive html with metadata, cruise track animation
Provides details about the metadata file names used in 1 row

Output table - path:

C:\R_WD\WRITE\
\METADATA_TEST_BEFORE_PROCEED\AMU2019001_Absorbance_Particate_plus_Detritus_file_list_txt_d6_and_d6p_264_rows.csv
\WORKS\RESULTS\
\WORKS\final_d6_txt_nrow_filepath_NROW_ERR_df_0_rows.csv
\WORKS\final_d6p_txt_nrow_filepath_NROW_ERR_df_0_rows.csv
\WORKS\GENERATE_UNPAIRED_P_D_PHY\all_Absorbance_files_including_the_ones_999_created_should_be_396_and_they_are_396_rows.csv
\WORKS_Absorbance_raw_data_files_analysis\Phytoplankton_negative_values_124_rows_stopped_at_file_132_of_132.csv
\WORKS2\RESULTS\
\WORKS_PLOT_Absorbance_no_VOL\RESULTS
\WORKS4\RESULTS\Absorbance_3tables_with_metadata\
\WORKS4\ALL_samples_ID_not_found_in_HPLC_and_QAT\
\WORKS5\RESULTS\
\WORKS_PLOT_Absorption_final_tables\RESULTS
\Absorbance_vs_Absorption_3dplots_copy
\WORKS_QC\QC_plots
\WORKS6
\WORKS7\RESULTS
\WORKS_PLOT_Phytoplankton_Absorption_by_QC_flags_Comments\RESULTS
\WORKS_PLOT_Absorbance_and_Absorption\RESULTS
\WORKS_PLOT_ON_MAP_Phytoplankton_Absorption_QC_Comments\RESULTS
\WORKS_Metadata_per_cruise_stats\

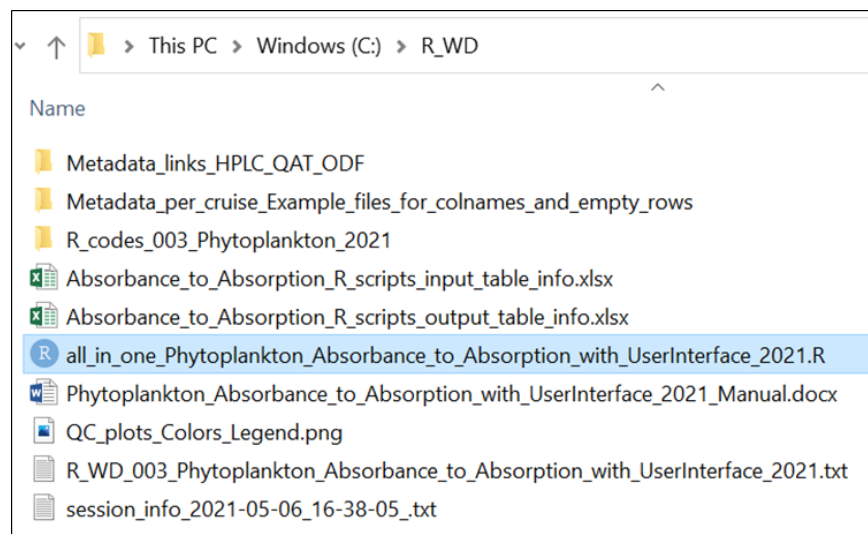
R main script name and path to run which calls all the functions:

“C:\R_WD

all_in_one_Phytoplankton_Absorbance_to_Absorption_with_UserInterface_2021.R”

Line 1 to line 90 – User input and testing, line 90 to 245 – processing.

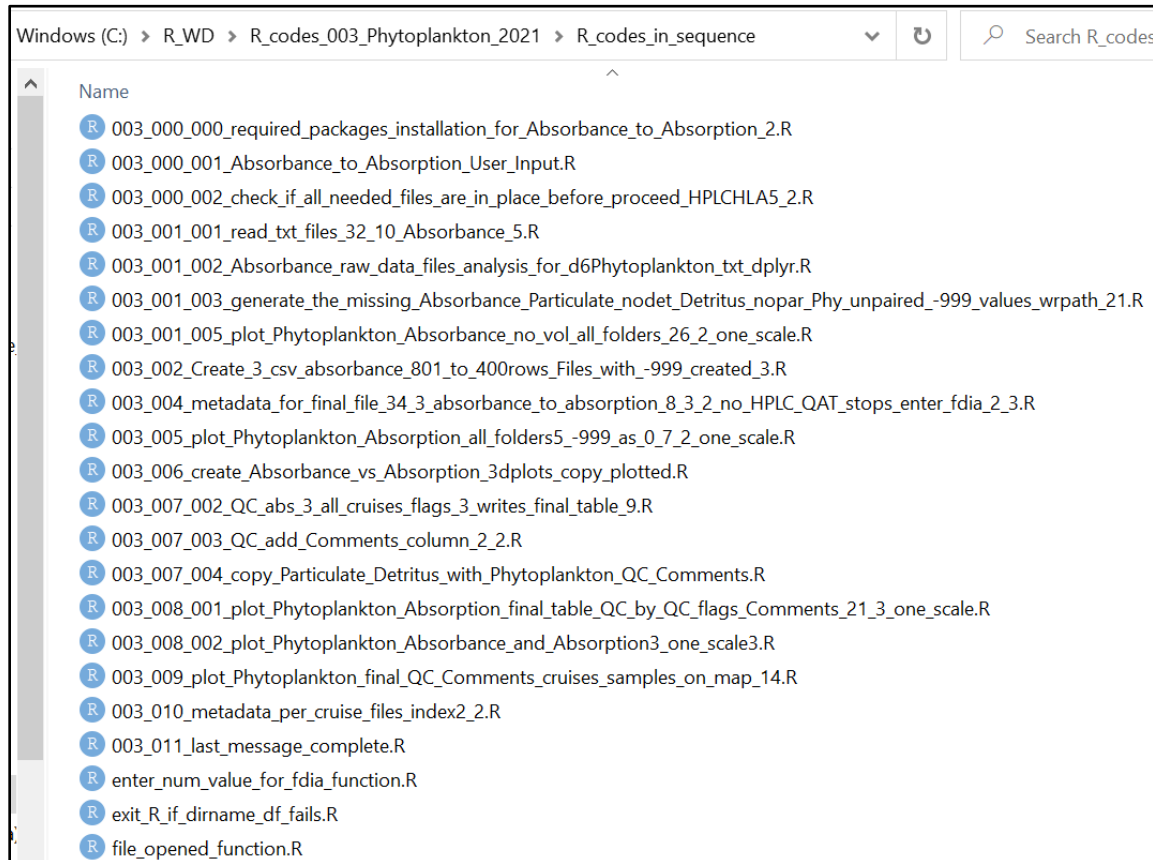
Note: The scale for the plotted values for absorbance and absorption it is determined by the minimum and the maximum values within the cruise folder selected for processing.



3. Description of each script

The R script above will call the other 20 R scripts that are located in:

“C:\R_WD\R_codes_003_Phytoplankton_2021\R_codes_in_sequence\”



Description of each R scripts (1 to 20):

1. “003_000_000_required_packages_installation_for_Absorbance_to_Absorption_2.R”

Checks if the R version is 4.0.4. The R script contains package detection and installation with dependencies with User Dialog requiring confirmation.

Packages required: "tcltk2", "DescTools", "reader", "dplyr", "tidyr", "stringr", "R.utils", "readxl", "fs", "qdapRegex", "gdata", "lmodel2", "robustbase", "leaflet", "mapview", "leaflet.providers", "leaflet.extras2", "sf", "rgl", "xfun", "webshot", "webdriver", "phantomjs"(webdriver::install_phantomjs(version = "2.1.1", baseURL = "https://github.com/wch/webshot/releases/download/v0.3.1/"))

IF PHANTOMJS INSTALLATION FAILS within the R script 'required_packages_installation_for_Absorbance_to_Absorption.R', PLEASE INSTALL IT MANUALLY WITH THE COMMAND BELOW:

```
webdriver::install_phantomjs(version = "2.1.1", baseURL =  
"https://github.com/wch/webshot/releases/download/v0.3.1/")
```

2. “003_000_001_Absorbance_to_Absorption_User_Input.R”

Contains all User Input interactive dialog for the input and files needed.

The script will store the cruise folder selection and the coded cruise name in

“C:\R_WD\ folder_name_to_coded_cruise_name_table.csv” to be called by the other scripts. If exists it will be overwritten; If the file is open, “try again” window will pop up before R will exit.

folder_name	coded_cruise_name
AMU2019-001 Lab sea	AMU2019001

The R script will store a copy of the metadata files selected (HPLC, QAT and ODF) in “C:\R_WD\Metadata_per_cruise\cruise_folder_name\”; it will call these files when needed. If exists, deletion is prompted with user dialog.

For the record,

"C:\R_WD\WRITE\METADATA_TEST_BEFORE_PROCEED\user_input_last_session.csv" is written.

The metadata files can be found in:

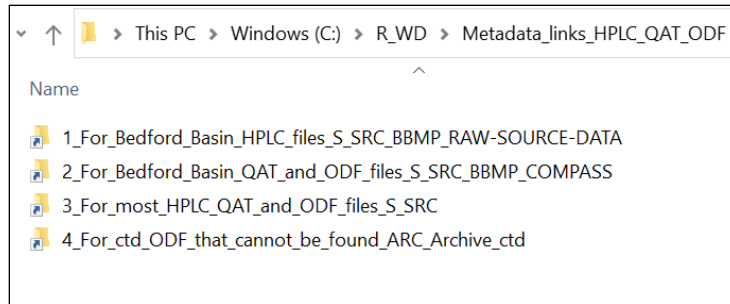
For BEDFORD BASIN - QAT and ODF files the link is: “S:\SRC\BBMP\COMPASS\”

For BEDFORD BASIN - HPLC files the link is: “S:\SRC\BBMP\RAW SOURCE DATA\”

For other missions HPLC, QAT and ODF files the link is: “S:\SRC\”

For ctd ODF files the link is: “S:\ARC\Archive\ctd\”

The folder “C:\R_WD\Metadata_links_HPLC_QAT_ODF\” contains shortcuts for fast access:



HPLC file - must contain "HPLC" in the file name, the file extension must be .xls, .xlsx or .xslm, **not .csv**

QAT file - must contain "_QAT" in the file name, the file extension must be .xls, .xlsx or .csv, **not .xslm**

ODF file - must have the file extension ".ODF"

The HPLC file must have (27) columns and QAT (40) columns; the columns are selected by the column number and then checked by name; The ODF file must have at least the first 12 rows(3 to 12 are needed). Sample files for metadata column names are in the folder below:

"C:\R_WD\Metadata_per_cruise_Example_files_for_colnames_and_empty_rows\Cruise_folder_name\"

The columns needed from the HPLC file are:

ID	DEPTH	HPLCHLA	ABS. VOL(L)
----	-------	---------	-------------

The columns needed from the _QAT file are:

Filename	Cruise_number	Event_number	Latitude	Longitude	Sample_id	Date	Pressure
----------	---------------	--------------	----------	-----------	-----------	------	----------

The rows needed from the .ODF file – the metadata header that will be used for Absorption tables (WORKS7) – are from row 3 to row 12:

```
CRUISE_HEADER
COUNTRY_INSTITUTE_CODE=data
CRUISE_NUMBER='data'
ORGANIZATION='data'
CHIEF_SCIENTIST='data'
START_DATE='dd-Mmm-YYYY 00:00:00.00'
END_DATE=' dd-Mmm-YYYY 00:00:00.00'
PLATFORM='data'
CRUISE_NAME='data'
CRUISE_DESCRIPTION='data'
```


No commas(,) should exist or be added to the ODF file if requires editing for correction – otherwise the script will fail on QC: “more columns than column names”. The script will inform the user if additional commas(,) are detected.

The data contained within the ODF header will not be processed, only verified if line 12 is “CRUISE_DESCRIPTION” – which will validate the ODF file.

3. “003_000_002_check_if_all_needed_files_are_in_place_before_proceed_HPLCHLA_5_2.R”

Runs the data and metadata tests before processing.

Writes in "C:\R_WD\WRITE\METADATA_TEST_BEFORE_PROCEED\"

Note: Sometimes absorbance raw data files containing the letter “r” or other characters in the name (123456r.txt, 123456pr.txt) instead of just the sample ID number; the script will look for these but will not process them; to be processed the “r” letter must be removed from the file name and the R script run again; the file below will be written to record the “r” files found:

"C:\R_WD\WRITE\METADATA_TEST_BEFORE_PROCEED\
cruise_name_files_with_r_in_filename_file_list_txt_d6rS.csv"

The tests that check the metadata and raw files are listed below:

1. Checks if there is at least one Absorbance Sample ID Particulate = 123456.txt and one Absorbance Sample ID Detritus = 123456p.txt in the cruise folder selected
2. Checks if raw Particulate txt files with the letter 'r' (123456r.txt, 123456rp.txt) have been found in the folder selected
3. Checks if folder “C:\R_WD\Metadata_per_cruise\” already contains a copy of the metadata files (HPLC, QAT, ODF) and prompts for deletion if true before a new selection is made
4. Checks if the HPLC file selected have exactly 27 columns and the columns 1(A),2(B),19(S),27(AA) are named 'ID', 'DEPTH', 'HPLCHLA', 'ABS. VOL(L)'. If not, HPLC column name error is displayed and the file “C:\R_WD\WRITE\METADATA_TEST_BEFORE_PROCEED\HPLC_column_names_error_table_for_file_.csv” is written
5. Checks if the first line in the second column [1,2] from the QAT selected file has the same cruise coded name as the input
6. Checks if the _QAT file selected has the right column names for the 8 needed columns from 40. The needed columns 1(A),2(B),3(C),4(D),5(E),7(F),8(G),16(P) must be named 'filename', 'cruise_number', 'station', 'eventLatitude', 'eventLongitude', 'sample_id', 'date', 'PrDM' if QAT column name error is returned. the file “C:\R_WD\WRITE\METADATA_TEST_BEFORE_PROCEED\QAT_column_names_error_table_for_file_.csv” is written
7. Checks if in the ODF file line 12 is named “CRUISE_DESCRIPTION”

8. Checks if the ODF header has the right number of commas (,) – (9 commas from line 3 to 11, no comma on line 12)
9. Checks if there are missing Absorbance Sample ID's from HPLC and QAT, shows the results on screen, writes files with the missing sample ID's
10. Shows a preview of the ODF data header for review
11. Checks if both HPLC and QAT files contain at least one Sample ID from the Absorbance raw data files selected for processing

4. “exit_R_if_dirname_df_fails.R”

R will exit without saving the workspace if “dirname(data.frame)” fails in 5 (003_001_001) which happened with older versions of R.

5. “003_001_001_read_txt_files_32_10_Absorbance_5.R”

Reads Absorbance raw data files Particulate (123456.txt) and Detritus (123456p.txt) from the selected cruise folder.

Writes the new Absorbance raw data file formats “123456Particulate.txt”, “123456Detritus.txt” and “123456Phytoplankton.txt”(contains Wavelength, Particulate, Detritus and Particulate minus Detritus – used as data input for other R scripts) in "C:\R_WD\WRITE\WORKS\RESULTS"

Writes reports in "C:\R_WD\WRITE\WORKS\" listed below:

- Shows the Particulate – Detritus – only the valid raw data files that are paired:

“final_d6_d6p_txt_related_without_NA_paired_without_nrow_errors.csv”

- Shows the Particulate – Detritus – only the valid raw data files that are paired and unpaired; this list will be used to create the new format: “123456Particulate.txt”, “123456Detritus.txt” and “123456Phytoplankton.txt”:

“d6_txt_that_ARE_in_d6p_txt_full_join.csv”

- The new format of Absorbance raw data files list that were written:

“search_result_d6Detritus_txt_files_processed_df.csv”

“search_result_d6Particulate_txt_files_processed_df.csv”

“search_result_d6Phytoplankton_txt_files_processed_df.csv”

Screen shot – the first 5 lines of the file

“search_result_d6Detritus_txt_files_processed_df.csv “

search_result_d6Detritus_txt_files_processed	filename_ext_d6p	filename_d6p	file
C:/R_WD/WRITE/WORKS/RESULTS/AMU2019-001 Lab sea/476288Detritus.txt	476288Detritus.txt	476288	476288
C:/R_WD/WRITE/WORKS/RESULTS/AMU2019-001 Lab sea/476314Detritus.txt	476314Detritus.txt	476314	476314
C:/R_WD/WRITE/WORKS/RESULTS/AMU2019-001 Lab sea/476318Detritus.txt	476318Detritus.txt	476318	476318
C:/R_WD/WRITE/WORKS/RESULTS/AMU2019-001 Lab sea/476321Detritus.txt	476321Detritus.txt	476321	476321
C:/R_WD/WRITE/WORKS/RESULTS/AMU2019-001 Lab sea/476324Detritus.txt	476324Detritus.txt	476324	476324

6. “003_001_002_Absorbance_raw_data_files_analysis_for_d6Phytoplankton_txt_dpl yr.R”

Reads new Particulate Absorbance raw data files (123456Phytoplankton.txt) from the folder
"C:\R_WD\WRITE\WORKS\RESULTS"

Checks each absorbance Sample ID on each Wavelength(801) for instances: Par<0, Par<Det,
Par>0&Det<0, Par<0, Par<0&Det>0, Par=Det(Phyto=0), Par<0&Det<0, Det<0, Par=0, Det=0,
Par=0&Det=0, Par>Det, sorted by the number of occurrences. (Par = Particulate, Det = Detritus,
Phyto = Phytoplankton)

Writes reports for each instance in

"C:\R_WD\WRITE\WORKS_Absorbance_raw_data_files_analysis\"

7. “003_001_003_generate_the_missing_Absorbance_Part particulate_nodet_Detritus_nop ar_Phy_unpaired_-999_values_wrpth_21.R”

Reads the new Absorbance raw data files formats 123456Particulate.txt, 123456Detritus.txt in
“C:\R_WD\WRITE\WORKS\RESULTS”

Creates the Absorbance missing pairs in the same folder; the names of the files are changed as
listed below to indicate missing pairs “C:\R_WD\WRITE\WORKS\RESULTS”:

- for Particulate without a Detritus pair:

123456_-999_Detritus.txt (all values = -999);
123456_nodet_Part particulate.txt;
123456_nodet_Phytoplankton.txt;

- for Detritus without a particulate pair:

123456_-999_Part particulate.txt (all values = -999);
123456_nopar_Detritus.txt;
123456_nopar_Phytoplankton.txt;

Writes reports in “C:\R_WD\WRITE\WORKS\” in folders
“GENERATE_UNPAIRED_P_D_PHY”, “FILE_LENGTH_ERROR_THE_PAIR”,
“MISSING_THE_PAIR”.

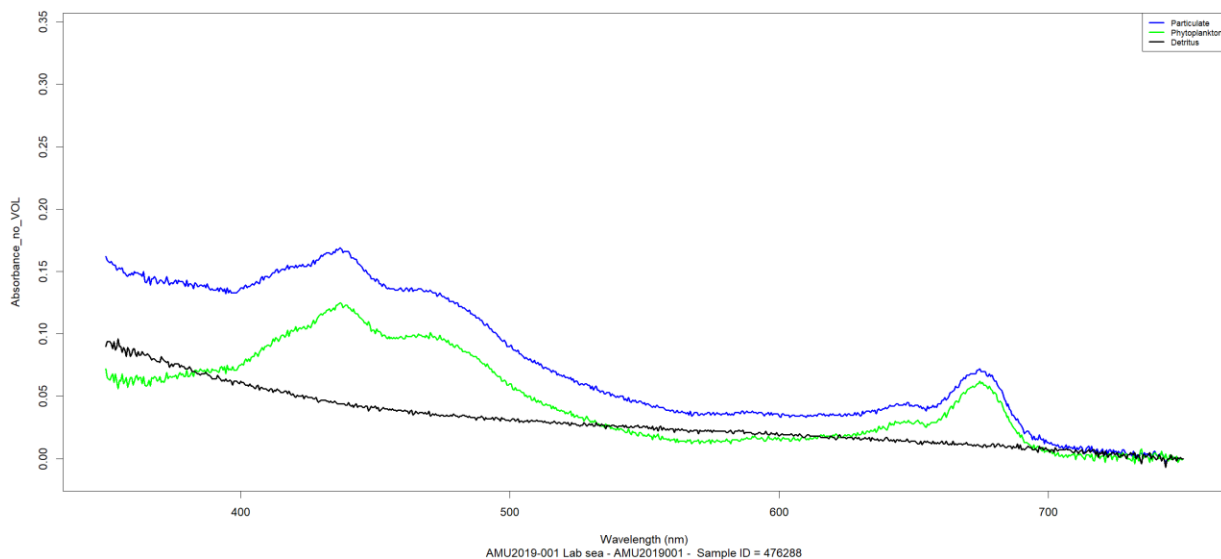
If a Particulate or a Detritus file does not have 801 rows of data, it will be considered as missing and is created with all values = -999.

Two report files are written for the nrow != 801 error to check in “C:\R_WD\WRITE\WORKS\”:

- For Particulate: “final_d6_txt_nrow_filepath_NROW_ERR_df.csv”
- For Detritus: “final_d6p_txt_nrow_filepath_NROW_ERR_df.csv”

8. “003_001_005_plot_Phytoplankton_Absorbance_no_vol_all_folders_26_2_one_scale.R”

Plots 2D(one plot per Sample ID) and 3D(one plot with all Sample ID's) of the Absorbance from tables in “C:\R_WD\WRITE\WORKS2\RESULTS” Phytoplankton, Particulate and Detritus and saves plots in “C:\R_WD\WRITE\WORKS_PLOT_Absorbance_no_VOL\RESULTS”;



9. "003_002_Create_3_csv_absorbance_801_to_400rows_Files_with_-999_created_3.R"

Combines the raw data (1 file per sample ID) into 1 table with all sample IDs for each and writes 3 Absorbance csv tables:

- “d6Particulate_project_table.csv”;
- “d6Detritus_project_table.csv”;
- “d6Phytoplankton_project_table.csv”;

in “C:\R_WD\WRITE\WORKS2\RESULTS”.

For files containing 801 wavelengths (350 to 750 step 0.5) these tables keep only 401 wavelengths (350 to 749 step 1) - this matches the historical data wavelengths used.

10. “003_004_metadata_for_final_file_34_3_absorbance_to_absorption_8_3_2_no_HPLC_QAT_stops_enter_fdia_2_3.R”

The "REFERENCE line 13 written in the Absorption table header is in this script (line 1770) ('Reference to method of absorption calculation'...)

“Reference to method of absorption calculation: 'Stramski D., R.A. Reynolds, S. Kaczmarek, J. Uitz and G. Zheng, 2015: Correction of pathlength amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region. Appl. Opt. 54: 6763?6782.”

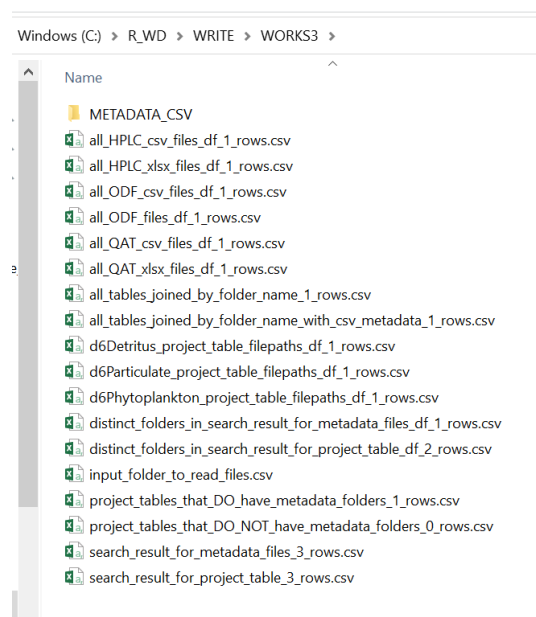
For the Absorption csv tables all commas(,) have been replaced with semicolon(;).

No commas(,) should be added to this REFERENCE line if edited – otherwise the script will fail on QC: “more columns than column names”. There is no comma(,) detection in this script for this section to inform the user about comma removal.

This script reads the metadata from the HPLC, QAT and ODF files from “C:\R_WD\Metadata_per_cruise\” keeping only the needed columns for each file; writes them as csv in “C:\R_WD\WRITE\WORKS3\METADATA_CSV\”.

Writes reports in “C:\R_WD\WRITE\WORKS3\”.

Screenshot with the folder content:



Writes the Absorbance missing sample ID's from HPLC and _QAT reports in
“C:\R_WD\WRITE\WORKS4\ALL_samples_ID_not_found_in_HPLC_and_QAT\”; if there are
no missing samples a file without data will not be written in here, only in
“C:\R_WD\WRITE\WORKS4\RESULTS\Absorbance_3tables_with_metadata\”.

Joins the Absorbance data (from “C:\R_WD\WRITE\WORKS2\”) with the metadata (csv from
“C:\R_WD\WRITE\WORKS3\METADATA_CSV\”) writing the Absorbance tables with
metadata in “C:\R_WD\WRITE\WORKS4\RESULTS\Absorbance_3tables_with_metadata\”.

Converts Absorbance to Absorption writing the Absorption tables for Particulate, Detritus and
Phytoplankton in

“C:\R_WD\WRITE\WORKS5\RESULTS\Absorption_3tables_with_metadata\”; also creates a
copy in “C:\R_WD\WRITE\WORKS5\ALL_samples_ID_not_found_in_HPLC_and_QAT\” for
the missing Absorbance sample ID's from HPLC and _QAT previous reports.

Absorption calculation used:

$\text{vol} = \text{volume} / 1000 \text{ \# in m}^3$

$\text{fdia} = \text{filter_diameter} / 1000 \text{ \# in m}$

$\text{farea} = \pi * (\text{fdia} / 2)^2 \text{ \# in m}^2$

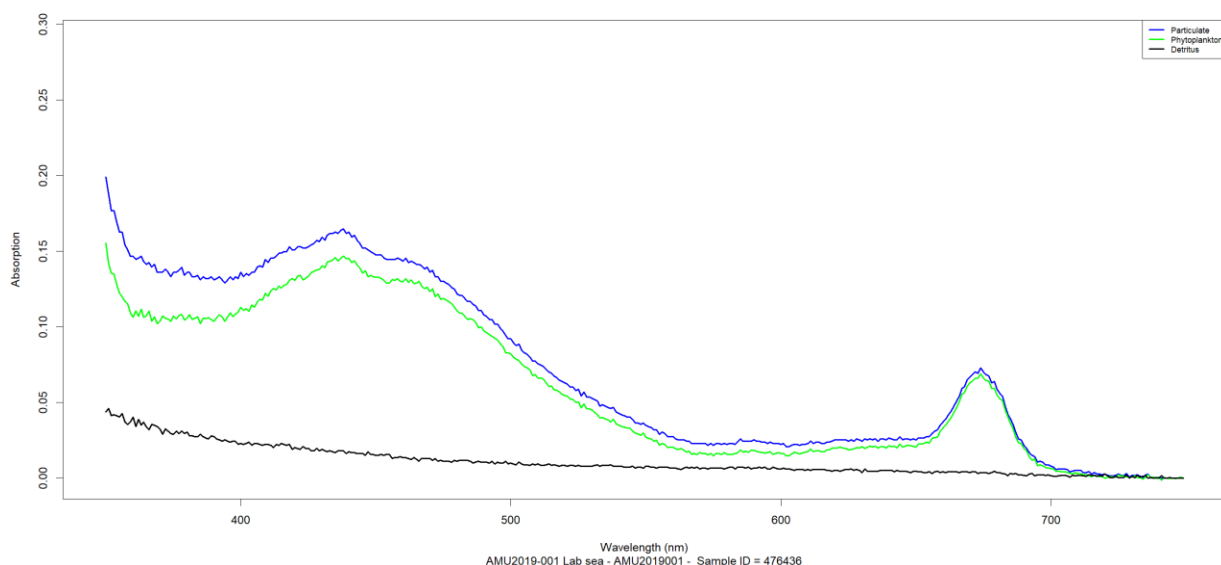
$\text{ods} = 0.679 * \text{absorbance}^{1.2804}$

$\text{absorption} = \log(10) * \text{ods} * \text{farea} / \text{vol}$

11. “003_005_plot_Phytoplankton_Absorption_all_folders5_- 999_as_0_7_2_one_scale.R”

Plots 2D(one plot per Sample ID) and 3D(one plot with all Sample ID's) Absorption from tables
in “C:\R_WD\WRITE\WORKS5\RESULTS\Absorption_3tables_with_metadata\” and saves
plots in “C:\R_WD\WRITE\WORKS_PLOT_Absorption_final_tables\RESULTS\”.

All missing values set as -999 will be plotted as 0 (zero) for scale relevance.



12. “003_006_create_Absorbance_vs_Absorption_3dplots_copy_plotted.R”

Copies the 3D Absorbance and the 3D Absorption plots in one folder in “C:\R_WD\WRITE\Absorbance_vs_Absorption_3dplots_copy\” to make them more accessible.

13. “003_007_002_QC_abs_3_all_cruises_flags_3_writes_final_table_9.R”

Performs the QC tests reading the Absorption table with metadata from “C:\R_WD\WRITE\WORKS5\RESULTS\” and the “HPLCHLA” column from the HPLC file from “C:\R_WD\Metadata_per_cruise\”.

Writes the QC plots in “C:\R_WD\WRITE\WORKS_QC\QC_plots\”.

Writes reports and the QC log file in “C:\R_WD\WRITE\WORKS_QC\ QC_log_file.txt”.

Writes the Phytoplankton Absorption table with the QC columns in “C:\R_WD\WRITE\WORKS6\”.

The QC Tests (19) are performed and saved as .png images except “TEST_0” (410nm_smaller_than_440nm) that is not a plot – the plot for the Sample ID’s that failed on TEST_0 but passed all the other QC tests are plotted blue in :

“C:\R_WD\WRITE\WORKS_PLOT_Phytoplankton_Absorption_by_QC_flags_Comments\RESULTS\”.

The QC plots legend: “C:\R_WD\ QC_plots_Colors_Legend.png”.

The 19 QC Tests performed and plots saved (18) are:

- TEST_0_410nm_smaller_than_440nm (not saved as png image);
- TEST_1_1_1_Absorption_443_TO_490_REL;
- TEST_1_2_1_hist_linear_Absolute_Residuals_443_TO_490;
- TEST_1_2_2_plot_linear_Absolute_Residuals_443_TO_490;
- TEST_1_3_1_hist_linear_Percent_Residuals_443_TO_490;
- TEST_1_3_2_plot_linear_Percent_Residuals_443_TO_490;
- TEST_1_4_1_hist_log10_Absolute_Residuals_443_TO_490;
- TEST_1_4_2_plot_log10_Absolute_Residuals_443_TO_490;
- TEST_1_5_1_hist_log10_Percent_Residuals_443_TO_490;
- TEST_1_5_2_plot_log10_Percent_Residuals_443_TO_490;
- TEST_2_1_1_CHLA_Absorption_TO_443_REL;
- TEST_2_2_1_log10_CHLA_Absorption_TO_443_REL;
- TEST_2_2_2_EXCLUDE_log10_CHLA_Absorption_TO_443_REL;
- TEST_3_1_1_CHLA_Absorption_TO_550_REL;
- TEST_3_2_1_log10_CHLA_Absorption_TO_550_REL;
- TEST_3_2_2_EXCLUDE_log10_CHLA_Absorption_TO_550_REL;
- TEST_4_1_1_CHLA_Absorption_TO_670_REL;
- TEST_4_2_1_log10_CHLA_Absorption_TO_670_REL;
- TEST_4_2_2_EXCLUDE_log10_CHLA_Absorption_TO_670_REL.

log10 CHLA Absorption Relationship to Wavelengths 443, 550 and 670 QC test on the residuals
- checks if points are greater than 3 standard-deviations.

The QC test adds 3 columns to the Phytoplankton Absorption table:

QC_flag_410_smaller_than_440	QC_flag_443_to_490	QC_flag_ChLA_to_443_550_670
------------------------------	--------------------	-----------------------------

containing one of the numbers **0**, **1**, **2** or **4** representing flags as follows:

- **0**: No quality control;
The reason is described in the “Comments” column which is the last column added post QC. Possible reasons:
 - a. The Sample ID is missing from the HPLC file (all values are set to -999) (no VOLUME to calculate absorption, no HPLCHLA for QC);
 - b. The Particulate Absorbance raw file does not have a Detritus pair (all Phytoplankton Absorption values are set to -999 for that Sample ID);
 - c. The Detritus Absorbance raw file does not have a Particulate pair (all Phytoplankton Absorption values are set to -999 for that Sample ID);
 - d. Phytoplankton Absorption contains negative values (to avoid log(-n) error in QC);
 - e. Phytoplankton Absorption value is 0 (zero) on one of the Wavelengths: 410, 440, 443, 490, 550, 670 (to avoid log0 error in QC).

- **1:** QC test passed (for all 3 tests: QC_flag_410_smaller_than_440, QC_flag_443_to_490, QC_flag_ChIA_to_443_550_670);
- **2:** QC test failed for tests “QC_flag_443_to_490” and “QC_flag_ChIA_to_443_550_670”
- **4:** QC test failed for test “QC_flag_410_smaller_than_440”

During the QC tests, on failure, the flag numerical values are replaced in sequence as follows:

- For QC_flag_410_smaller_than_440: 1, 0, 4;
- For QC_flag_ChIA_to_443_550_670: 1, 0, 2;
- For QC_flag_ChIA_to_443_550_670: 1, 0, 2.

14. “R_codes_folder, "003_007_003_QC_add_Comments_column_2_2.R”

Reads the Phytoplankton Absorption table with the QC columns from C:\R_WD\WRITE\WORKS6\, adds the comments column and writes the FINAL ABSORPTION TABLE WITH QC COMMENTS named "cruise_coded_name_Absorption_Phytoplankton_QC_Comments.csv" in “C:\R_WD\WRITE\WORKS7\RESULTS”.

The “Comments” column provides information about the filter diameter used or if the Sample ID is missing from the HPLC file (all values will be set to -999). If the Sample ID is missing from the QAT file it will not be mentioned in comments because the Absorption can be calculated without the QAT file, QAT files provide only metadata. The Sample ID’s missing from QAT will be excluded from map plots because there is no Longitude/Latitude assigned to Sample ID.

15. “003_007_004_copy_Part particulate_Detritus_with_Phytoplankton_QC_Comments.R”

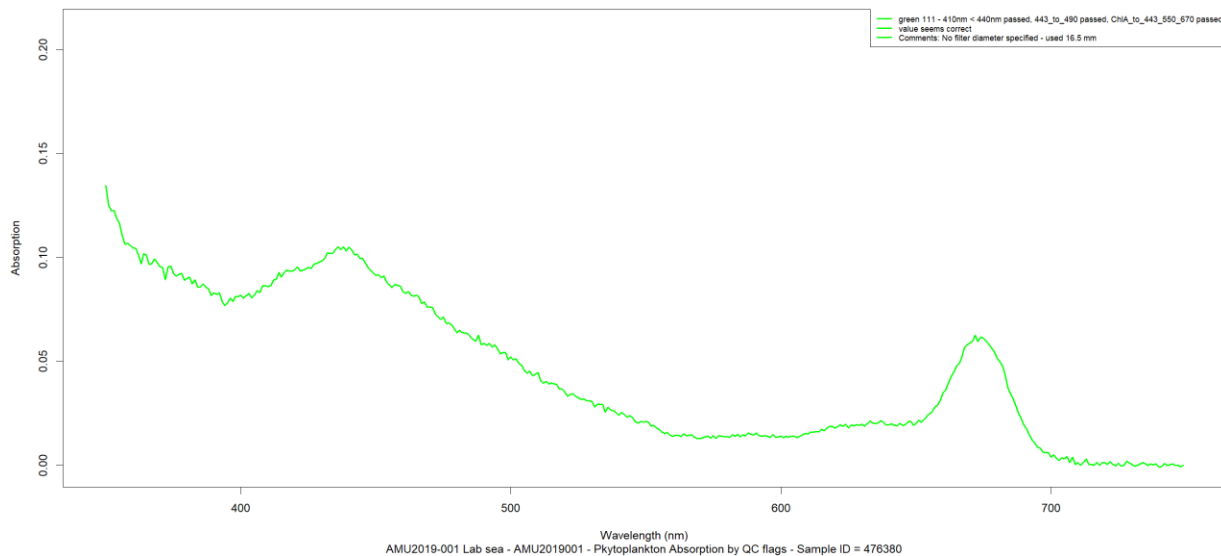
Copies from “C:\R_WD\WRITE\WORKS5\RESULTS\” the Particulate and Detritus Absorption tables to where the Phytoplankton Absorption with QC Comments is saved “C:\R_WD\WRITE\WORKS7\RESULTS\”

16. “003_008_plot_Phytoplankton_Absorption_final_table_QC_by_QC_flags_Comments_21_3_one_scale.R”

Plots 2D(one plot per Sample ID) and 3D(one plot with all Sample ID’s) Phytoplankton Absorption from table with QC Comments in “C:\R_WD\WRITE\WORKS7\RESULTS\” and saved them in “C:\R_WD\WRITE\WORKS_PLOT_Phytoplankton_Absorption_by_QC_flags_Comments\RESULTS\”.

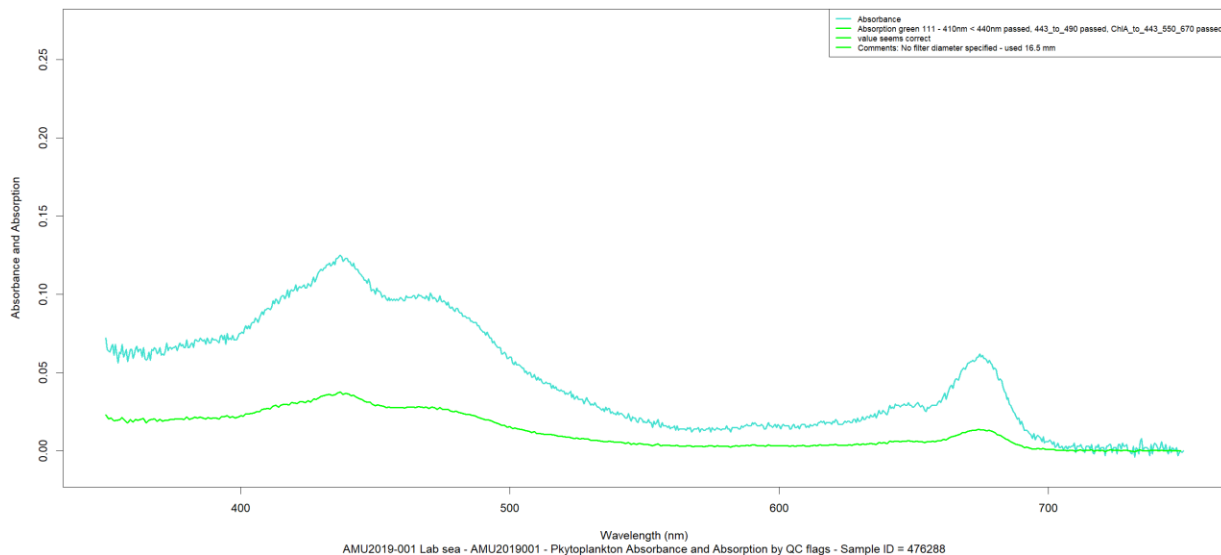
The QC plots legend: “C:\R_WD\ QC_plots_Colors_Legend.png”.

- QC_Flags = 000 color = black - no quality control
- QC_Flags = 111 color = green
 - 410nm < 440nm passed,
 - 443_to_490 passed,
 - ChlA_to_443_550_670 passed,
 - value seems correct
- QC_Flags = 411 color = blue
 - 410nm < 440nm failed,
 - 443_to_490 passed,
 - ChlA_to_443_550_670 passed
- QC_Flags = x21 color = orange
 - 443_to_490 failed,
 - ChlA_to_443_550_670 passed,
 - value appears inconsistent with other values
- QC_Flags = x12 color = purple
 - 443_to_490 passed,
 - ChlA_to_443_550_670 failed,
 - value appears inconsistent with other values
- QC_Flags = x22 color = red
 - 443_to_490 failed,
 - ChlA_to_443_550_670 failed,
 - value seems doubtful



17. "003_008_002_plot_Phytoplankton_Absorbance_and_Absorption3_one_scale3.R"

Plots each sample ID for Phytoplankton Absorption from
"C:\R_WD\WRITE\WORKS7\RESULTS\" and Phytoplankton Absorbance from
"C:\R_WD\WRITE\WORKS\RESULTS\" on the same image, writing in
"C:\R_WD\WRITE\WORKS_PLOT_Absorbance_and_Absorption\RESULTS"



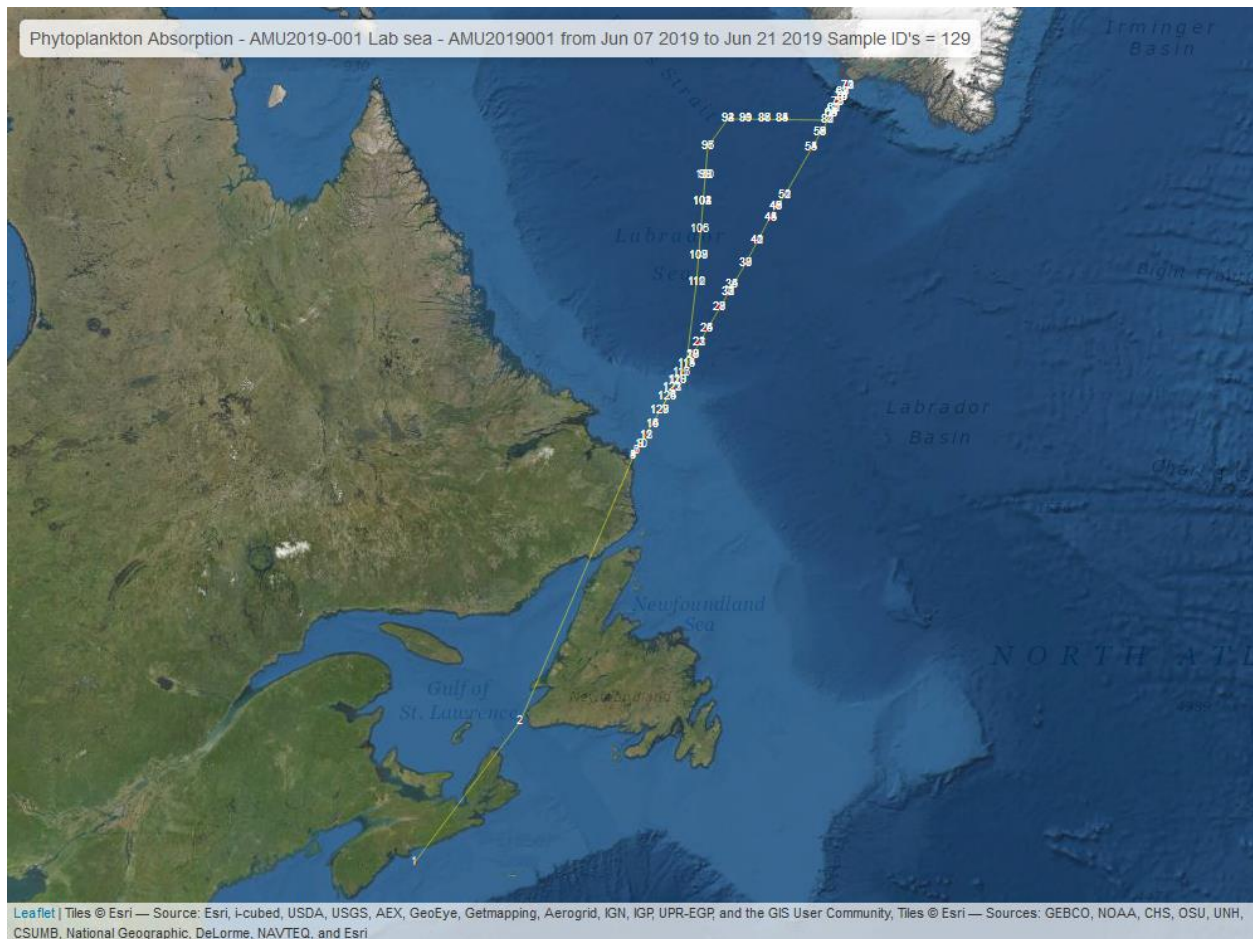
18. "003_009_plot_Phytoplankton_final_QC_Comments_cruises_samples_on_map_14.R"

Plots all Sample ID's within the cruise reading from "C:\R_WD\WRITE\WORKS7\RESULTS\" the Phytoplankton Absorption table with QC Comments on a leaflet map as .png, on an interactive html map (with metadata) and cruise track animation saved in "C:\R_WD\WRITE\WORKS_PLOT_ON_MAP_Phytoplankton_Absorption_QC_Comments\RESULTS\".

It is recommended that the folder
"C:\R_WD\WRITE\WORKS_PLOT_ON_MAP_Phytoplankton_Absorption_QC_Comments"
be zipped if a copy is needed to avoid the error "file names too long" when it is copied.

Leaflet providers for map details:

- Esri.WorldImagery;
- Esri.OceanBasemap.



19. "003_010_metadata_per_cruise_files_index2_2.R"

Creates reports in "C:\R_WD\WRITE\WORKS_Metadata_per_cruise_stats\" about the cruise folder name, coded cruise name, metadata file names used or missing in\from the process.

20. "003_011_last_message_complete.R"

Print "all went well" message box as the last script in sequence runs successfully and opens "C:\R_WD\Absorbance_to_Absorption_R_scripts_output_table_info.xlsx"

4. Session info ("C:\R_WD\session_info.txt")

R version 4.0.4 (2021-02-15)

Platform: x86_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 18363)

Matrix products: default

locale:

[1] LC_COLLATE=English_United States.1252, LC_CTYPE=English_United States.1252,
LC_MONETARY=English_United States.1252

[4] LC_NUMERIC=C, LC_TIME=English_United States.1252

attached base packages:

[1] tcltk, stats, graphics, grDevices, utils, datasets, methods, base

other attached packages:

[1] webdriver_1.0.6, webshot_0.5.2, xfun_0.22, sf_0.9-8, leaflet.extras2_1.1.0,
leaflet.providers_1.9.0

[7] mapview_2.9.0, leaflet_2.0.4.1, robustbase_0.93-7, lmodel2_1.7-3, gdata_2.18.0,
qdapRegex_0.7.2

[13] fs_1.5.0, readxl_1.3.1, R.utils_2.10.1, R.oo_1.24.0, R.methodsS3_1.8.1, stringr_1.4.0

[19] tidyr_1.1.3, dplyr_1.0.5, reader_1.0.6, NCmisc_1.1.6, DescTools_0.99.41, tcltk2_1.2-11

loaded via a namespace (and not attached):

[1] satellite_1.0.2, showimage_1.0.0, httr_1.4.2, tools_4.0.4, utf8_1.2.1, R6_2.5.0

[7] KernSmooth_2.23-18, DBI_1.1.1, colorspace_2.0-0, manipulateWidget_0.10.1, raster_3.4-5,
sp_1.4-5

[13] processx_3.5.1, tidyselect_1.1.0, Exact_2.1, curl_4.3, compiler_4.0.4, leafem_0.1.3

[19] expm_0.999-6, scales_1.1.1, DEoptimR_1.0-8, classInt_0.4-3, mvtnorm_1.1-1, callr_3.7.0

[25] proxy_0.4-25, digest_0.6.27, rmarkdown_2.7, base64enc_0.1-3, pkgconfig_2.0.3,
htmltools_0.5.1.1

[31] fastmap_1.1.0, htmlwidgets_1.5.3, rlang_0.4.10, rstudioapi_0.13, shiny_1.6.0, generics_0.1.0

[37] jsonlite_1.7.2, crosstalk_1.1.1, gtools_3.8.2, magrittr_2.0.1, Matrix_1.2-18, Rcpp_1.0.6

[43] munsell_0.5.0, fansi_0.4.2, lifecycle_1.0.0, yaml_2.2.1, stringi_1.5.3, debugme_1.1.0

[49] MASS_7.3-53.1, rootSolve_1.8.2.1, grid_4.0.4, promises_1.2.0.1, crayon_1.4.1, lmom_2.8

[55] miniUI_0.1.1.1, lattice_0.20-41, ps_1.6.0, knitr_1.32, pillar_1.6.0, boot_1.3-27

[61] gld_2.6.2, markdown_1.1, codetools_0.2-18, stats4_4.0.4, glue_1.4.2, evaluate_0.14

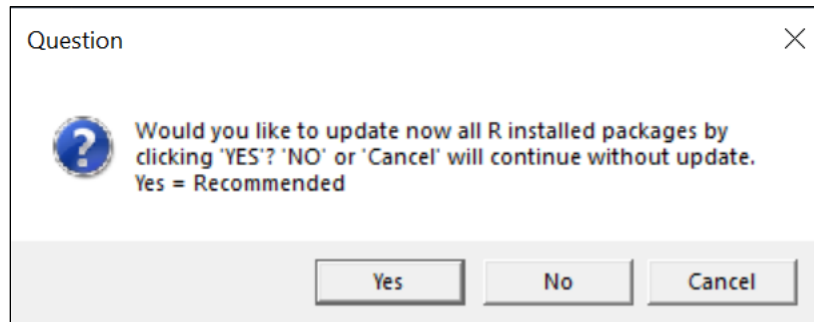
[67] data.table_1.14.0, png_0.1-7, vctrs_0.3.7, httpuv_1.5.5, cellranger_1.1.0, purrr_0.3.4

[73] assertthat_0.2.1, mime_0.10, xtable_1.8-4, e1071_1.7-6, later_1.1.0.1, class_7.3-18

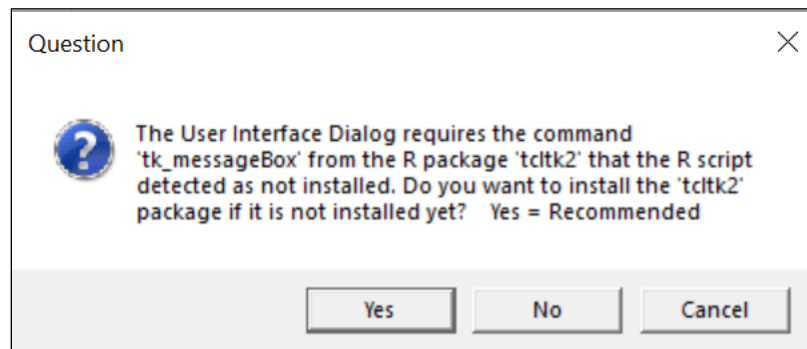
[79] tibble_3.1.1, tinytex_0.31, proftools_0.99-3, units_0.7-1, ellipsis_0.3.1

5. User Interface Dialogs

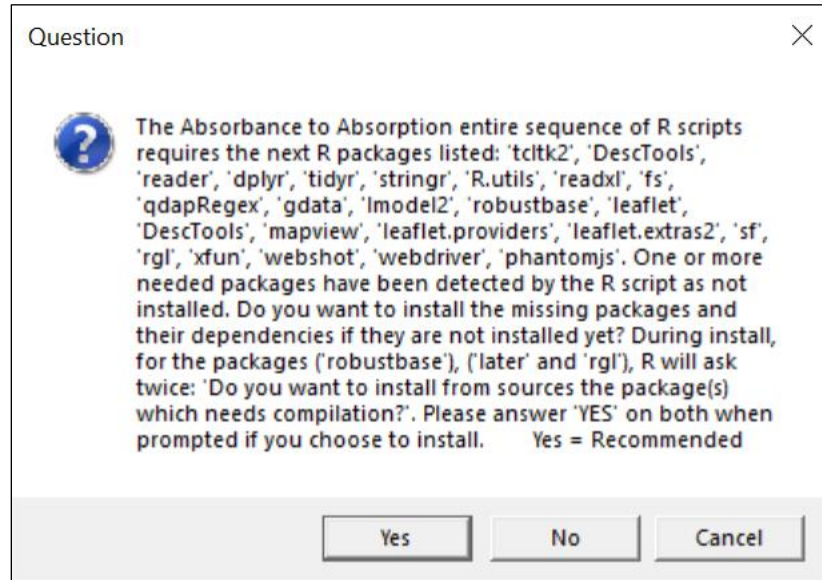
1. Every time the R script runs, from the beginning, it will open the dialog below offering to update all the packages that are already installed; it may take few minutes if packages were not updated in a while. If the R version is not 4.0.4, additional messages will appear to inform the user to choose “Continue” or “Cancel”; a newer version may run successfully, while an older version (Ex: R 3.6.3) will encounter a fatal error (on `dirname(data.frame)`) and R will close without saving the workspace. It is strongly recommended to use the R Version 4.0.4 (2021-02-15).



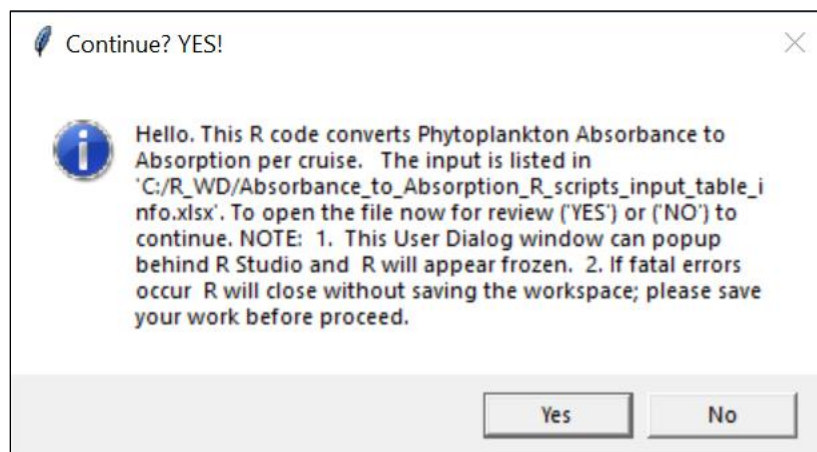
2. For the second dialog the script will check if one of the required packages named “tcltk2” is installed; if it is not, the dialog below is shown:



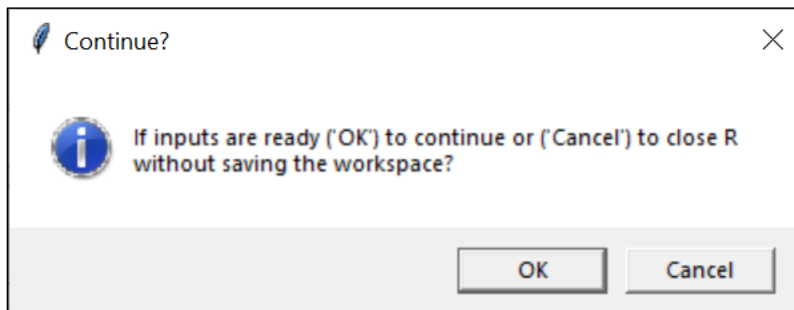
3. The script will check if all the needed packages are installed; If at least one is missing, this dialog below pops up. Choosing “Yes” will install only the missing packages listed and their dependencies. This step may take a few minutes to complete.



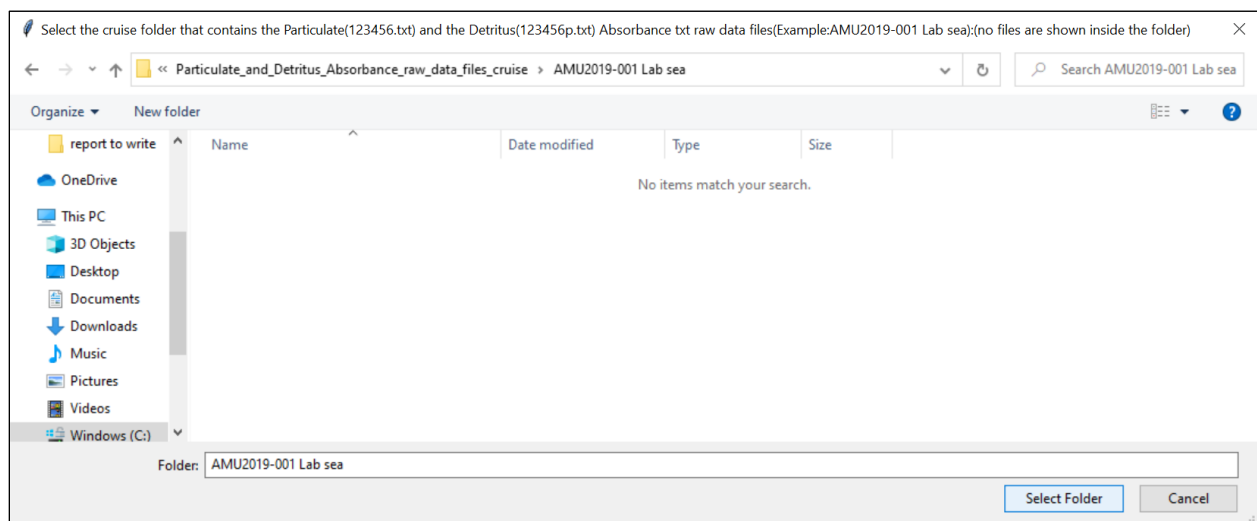
4. The second and the third dialog windows above will only pop up on the first run. If all needed packages are installed, the script will show only this fourth dialog. Choosing “Yes” will open the input table requirements for review; “No” will just move forward without opening the input table:



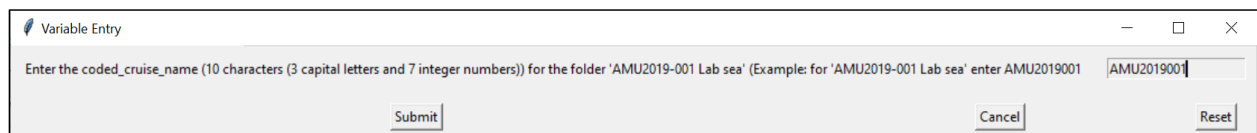
5. “Continue” confirmation (“OK”) if all the needed variables listed in the input table are ready



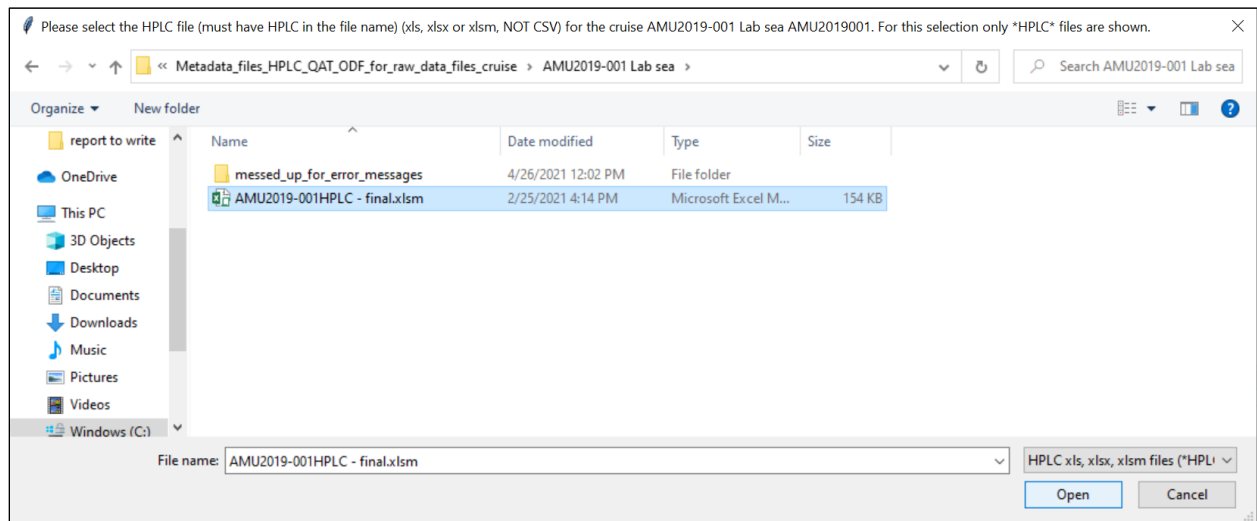
6. This window will ask to “Select the cruise folder that contains the Particulate(123456.txt) and the Detritus(123456p.txt) Absorbance txt raw data files (Example:AMU2019-001 Lab sea); no files are shown inside the folder – navigate to the folder and click “Select Folder”:



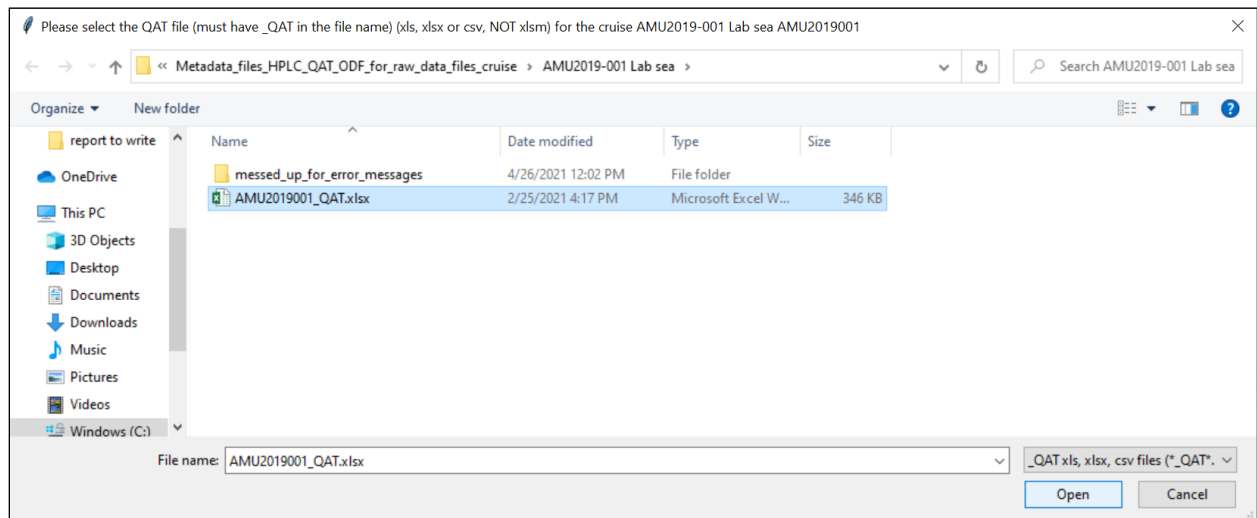
7. Enter the coded_cruise_name (10 characters (3 capital letters and 7 integer numbers)), (also known as cruise name or cruise number) for the folder selected above. (Example: for “AMU2019-001 Lab sea” enter “AMU2019001”):



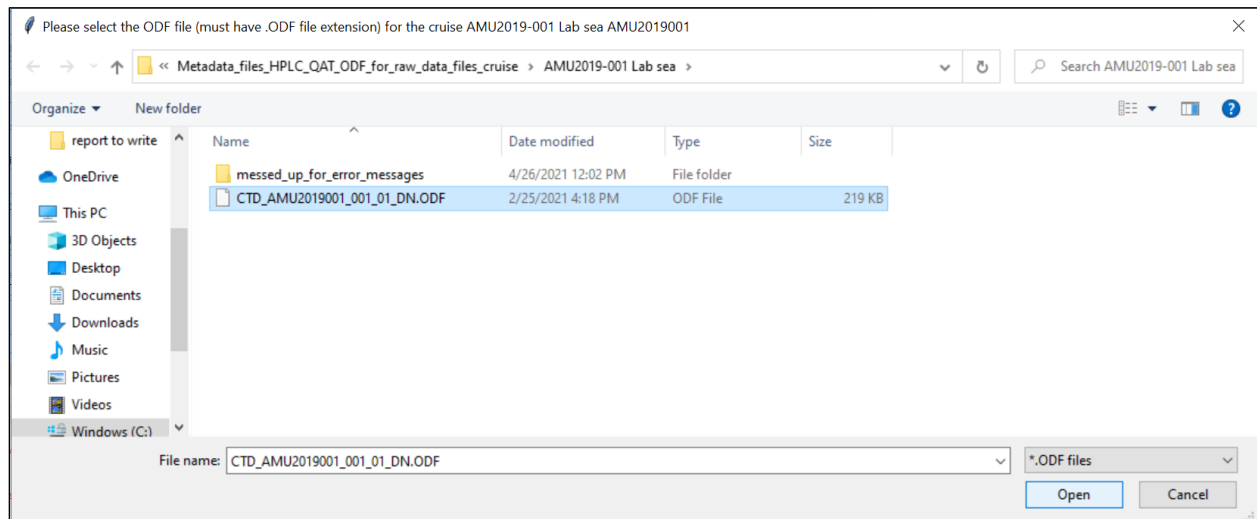
8. This window will ask to select the HPLC file (must have HPLC in the file name) (xls, xlsx or xlsx, NOT CSV) for the cruise selected for processing and click “Open”.



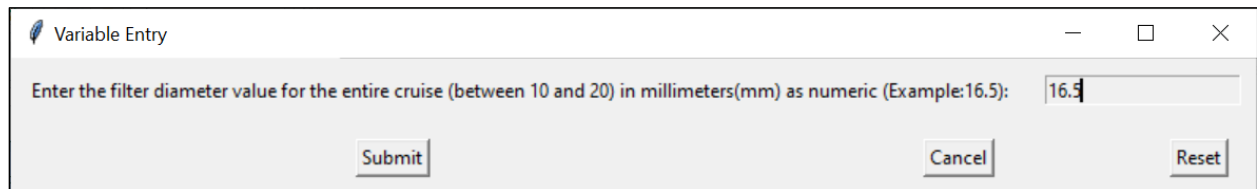
9. This window will ask to select the QAT file (must have _QAT in the file name) (xls, xlsx or csv, NOT xlsx) for the cruise selected for processing and click “Open”.



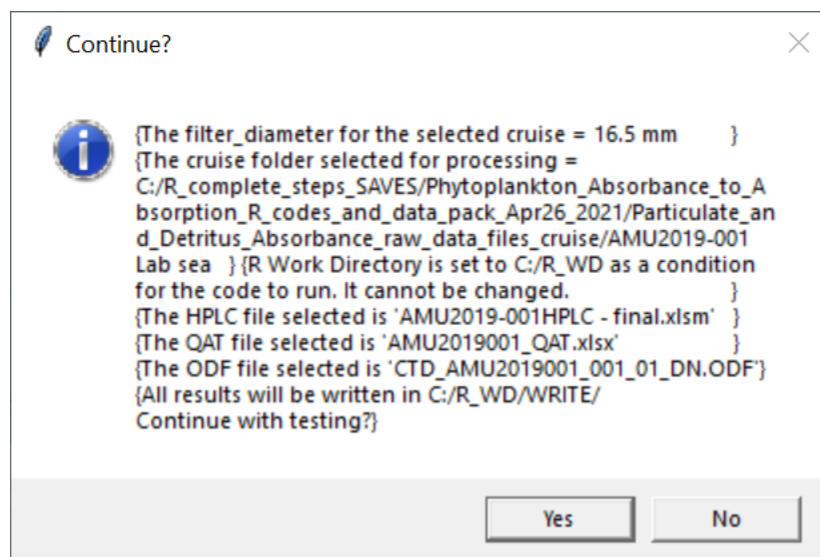
10. This window will ask to select the ODF file (must have .ODF file extension) for the cruise selected for processing and click “Open”. For this selection only *.ODF files are shown:



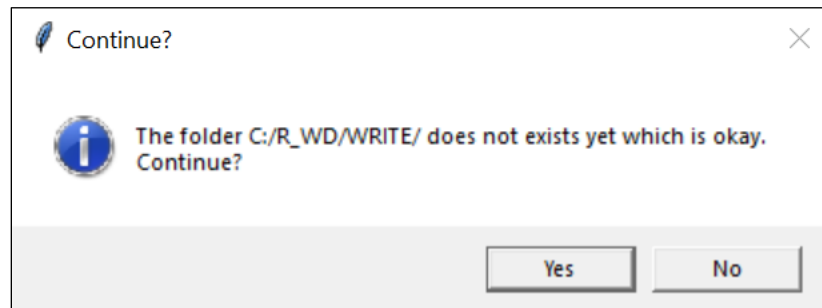
11. Enter the filter diameter value for the entire cruise (between 10 and 20) in millimeters(mm) as numeric (Example:16.5) and click “Submit”:



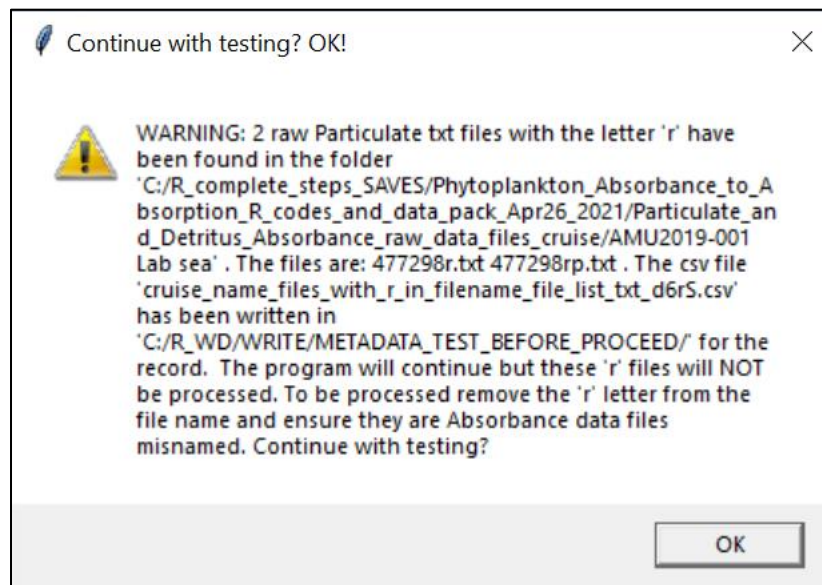
12. “Continue testing” confirmation dialog which summarises all the user inputs, click Yes to continue:



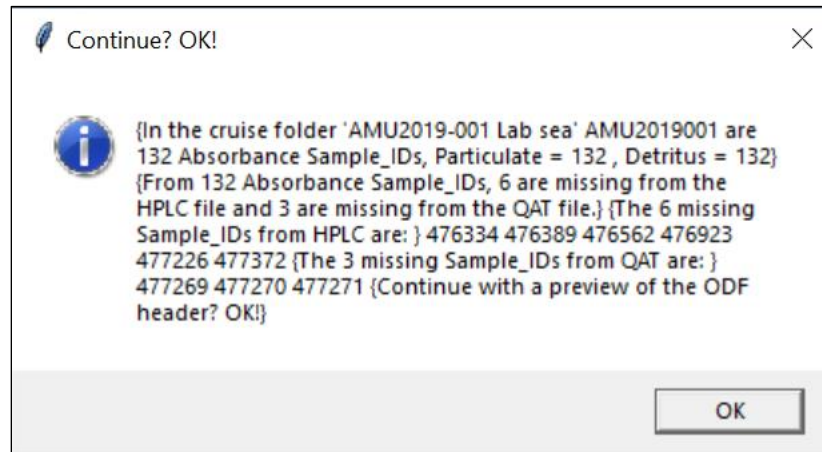
13. The script writes all results and reports in “WRITE” folder; Checking if the folder exists before writing to avoid overwriting, click Yes to continue, it will create the folder:



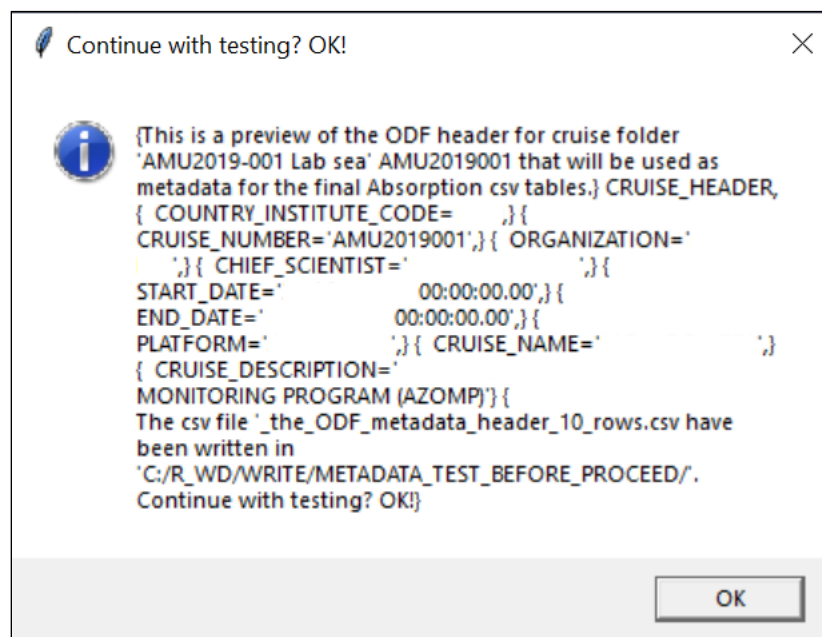
14. The scripts checks if there are raw data files with the letter “r” in the filename within the cruise folder selected for processing. The program will continue but these “r” files will NOT be processed. The files with “r” in the name are listed. To be processed remove the “r” letter from the file name and ensure they are Absorbance data files misnamed. To Continue with testing click Yes.



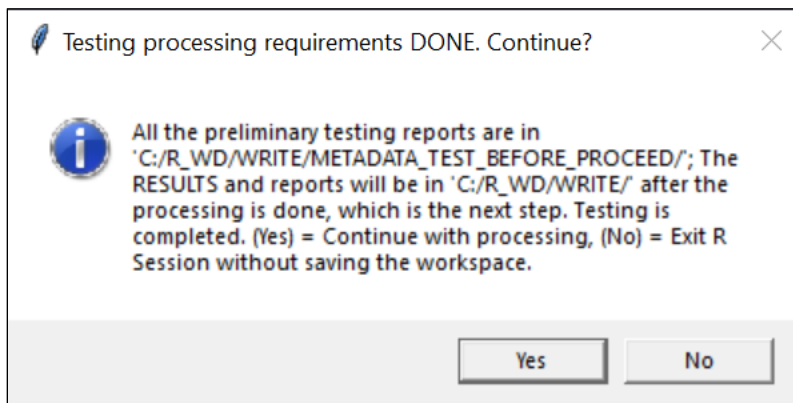
15. The next window shows a summary report of the number of Absorbance Sample ID's found within the folder selected for processing, how many are Particulate and how many are Detritus; It verifies if there are Sample ID's missing from HPLC and QAT printing the missing number from each one; review the missing IDs, click OK to continue;



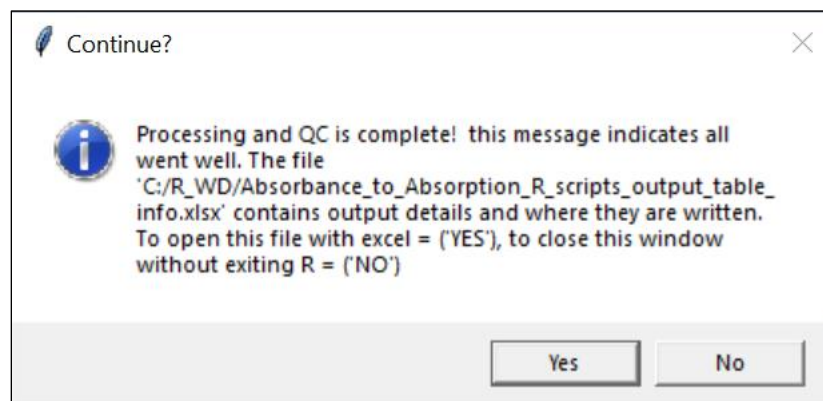
16. Displays a preview of the ODF header used as metadata for the final Absorption csv table file (data has been removed for this preview). Review the metadata, click OK to continue:



17. Next window pops up when all the data and metadata has been checked and accepted as valid and is ready to be processed; All reports generated so far are in “C:/R_WD/WRITE/METADATA_TEST_BEFORE_PROCEED/”; by clicking:
- “Yes”- the R script begins to process step by step writing results and reports as described in the output table; This step may take about 10 minutes;
 - “No” – will exit the R Session without saving the workspace.



18. The final message: “Processing and QC is complete! this message indicates all went well. The file “C:/R_WD/Absorbance_to_Absorption_R_scripts_output_table_info.xlsx” contains output details and where outputs are written. To open this file with excel = (“Yes”), to close this window without exiting R = (“No”):



Processing and QC is complete!