**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](#_Toc71715185)

[2. Method and Code 3](#_Toc71715186)

[3. Description of each script 6](#_Toc71715187)

[4. Session info ("C:\R\_WD\session\_info.txt”) 21](#_Toc71715188)

[5. User Interface Dialogs 23](#_Toc71715189)

# 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

# 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\_WDbefore 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 xlsm, NOT CSV) | AMU2019-001HPLC - final.xlsm |
| the QAT file (must have \_QAT in the file name) (xls, xlsx or csv, NOT xlsm) | 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:

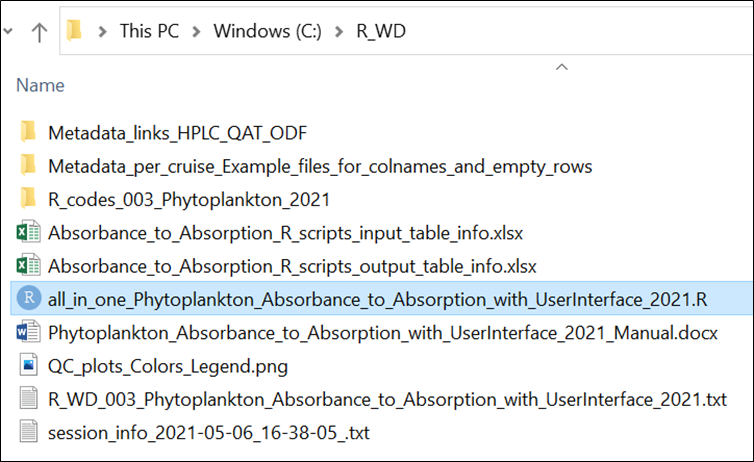


R main script name and path to run which calls all the functions:

**“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.

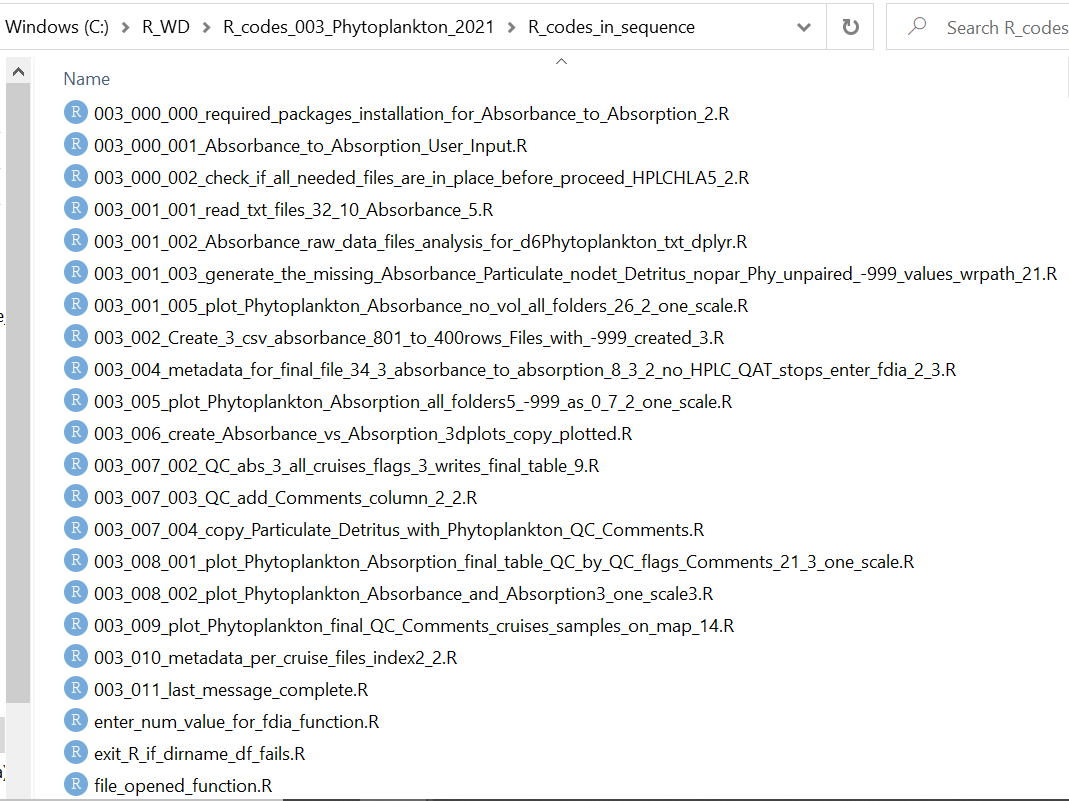
Note: The scale for the plotted values for absorbance and absorption it is determined by the minimum and the maximum values within the cruise folder selected for processing.



# 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/")

1. **“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:**

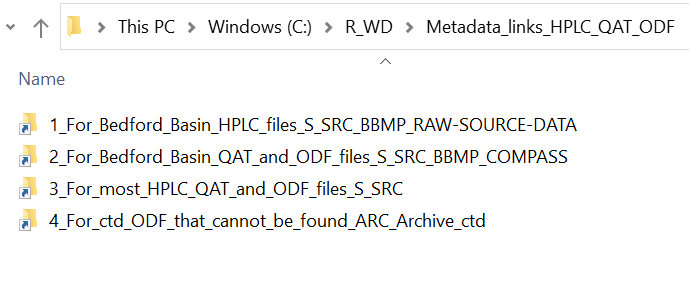
For BEDFORD BASIN - QAT and ODF files the link is: “S:\SRC\BBMP\COMPASS\”

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 .xlsm, **not .csv**

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

**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.

1. **“003\_000\_002\_check\_if\_all\_needed\_files\_are\_in\_place\_before\_proceed\_HPLCHLA5\_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
12. **“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.

1. **“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 “

1. **“003\_001\_002\_Absorbance\_raw\_data\_files\_analysis\_for\_d6Phytoplankton\_txt\_dplyr.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\"

1. **“003\_001\_003\_generate\_the\_missing\_Absorbance\_Particulate\_nodet\_Detritus\_nopar\_Phy\_unpaired\_-999\_values\_wrpath\_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\_Particulate.txt;

123456\_nodet\_Phytoplankton.txt;

* for Detritus without a particulate pair:

123456\_-999\_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”.

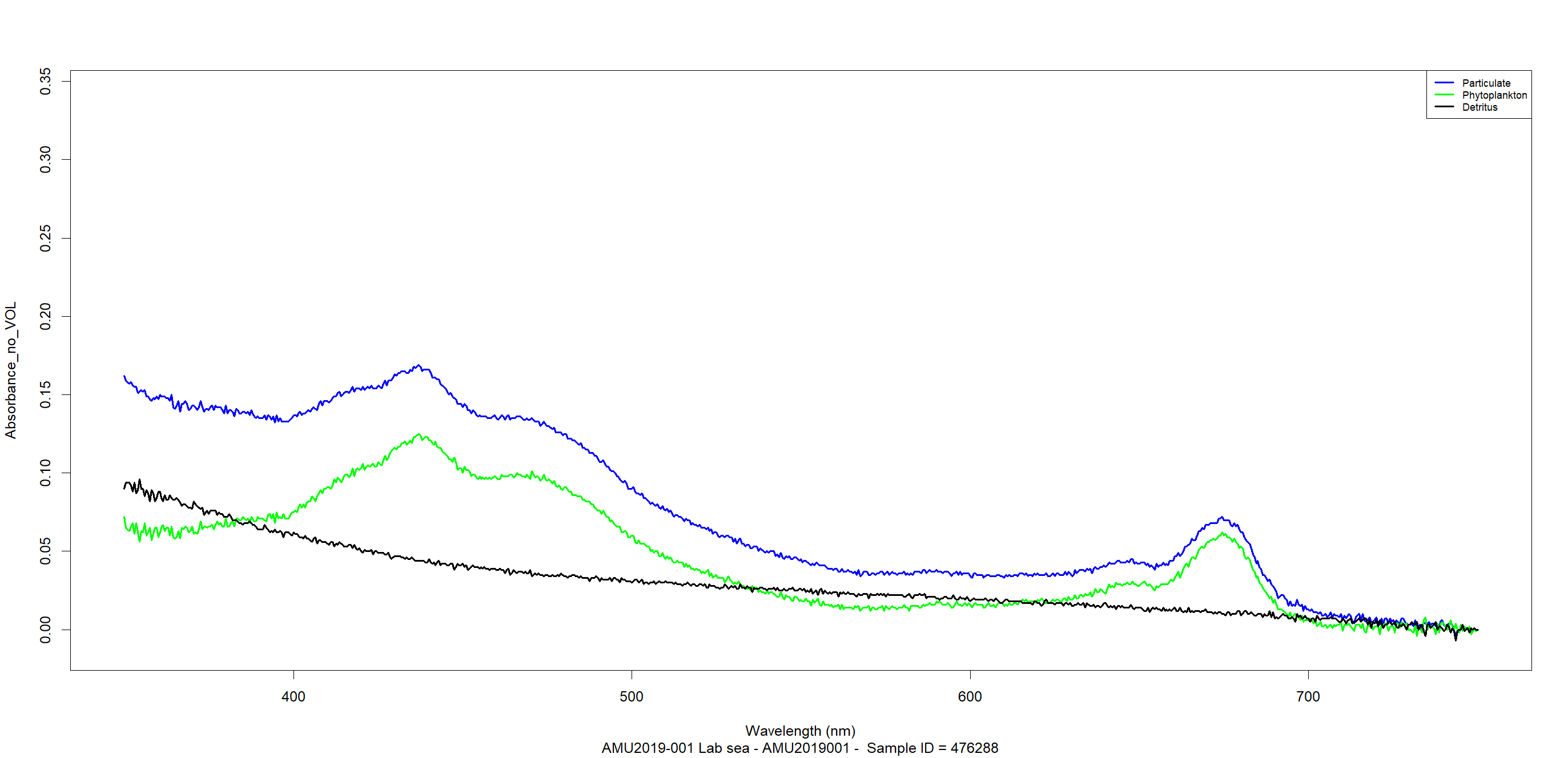
**If a Particulate or a Detritus file does not have 801 rows of data, it will be considered as missing and is created with all values = -999.**

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”

1. **“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”;



1. **"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.

1. **“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.*'”

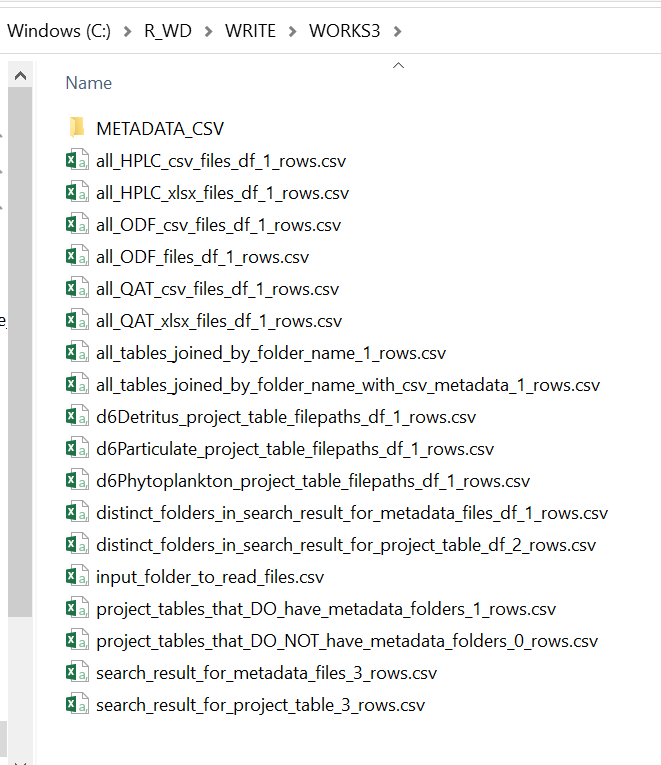
For the Absorption csv tables all commas(,) have been replaced with semicolon(;).

**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:**

vol = volume / 1000 # in m3

fdia = filter\_diameter/1000 # in m

farea = pi\* (fdia/2)^2 # in m2

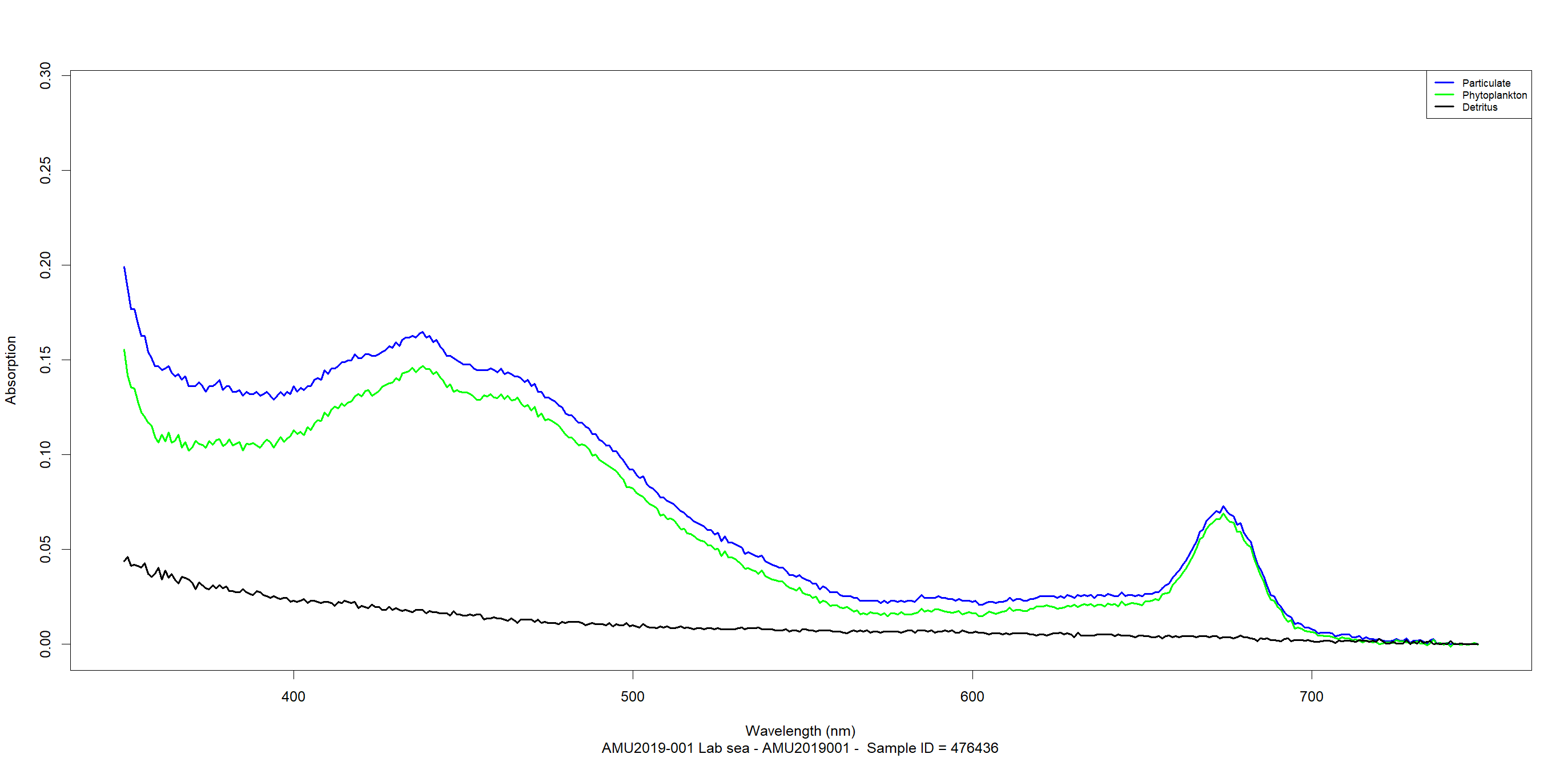
ods = 0.679 \* absorbance^1.2804

absorption = log(10) \* ods \* farea/vol

1. **“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.



1. **“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.

1. **“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:

1. The Sample ID is missing from the HPLC file (all values are set to -999) (no VOLUME to calculate absorption, no HPLCHLA for QC;
2. The Particulate Absorbance raw file does not have a Detritus pair (all Phytoplankton Absorption values are set to -999 for that Sample ID);
3. The Detritus Absorbance raw file does not have a Particulate pair (all Phytoplankton Absorption values are set to -999 for that Sample ID);
4. Phytoplankton Absorption contains negative values (to avoid log(-n) error in QC);
5. 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\_ChlA\_to\_443\_550\_670);
* **2**: QC test failed for tests “QC\_flag\_443\_to\_490” and “QC\_flag\_ChlA\_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\_ChlA\_to\_443\_550\_670: 1, 0, 2;
* For QC\_flag\_ChlA\_to\_443\_550\_670: 1, 0, 2.

1. **“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.

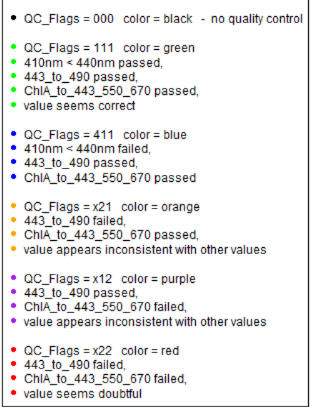
1. **“003\_007\_004\_copy\_Particulate\_Detritus\_with\_Phytoplankton\_QC\_Comments.R”**

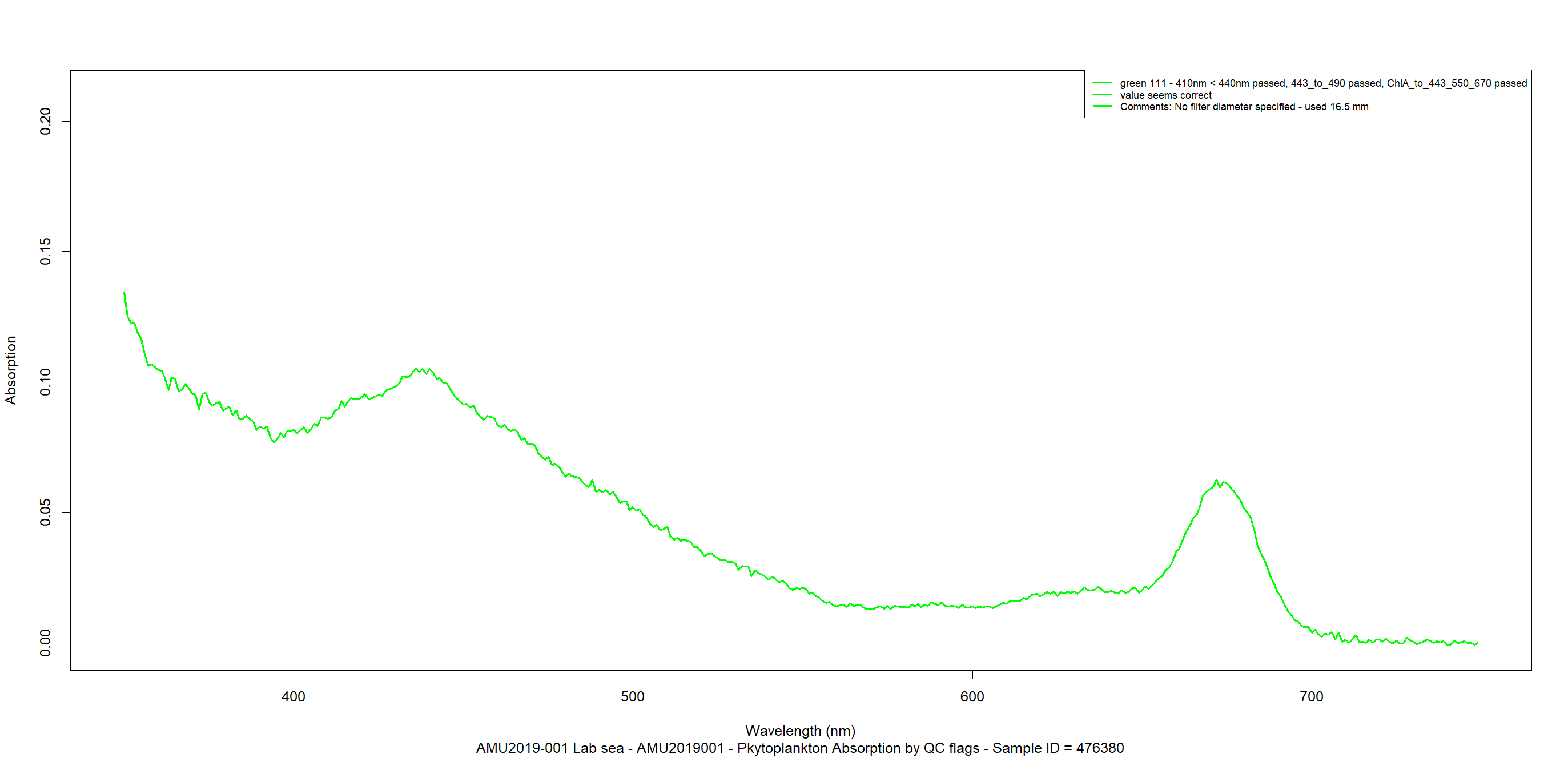
Copies from “C:\R\_WD\WRITE\WORKS5\RESULTS\” the Particulate and Detritus Absorption tables to where the Phytoplankton Absorption with QC Comments is saved “C:\R\_WD\WRITE\WORKS7\RESULTS\”

1. **“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\”.

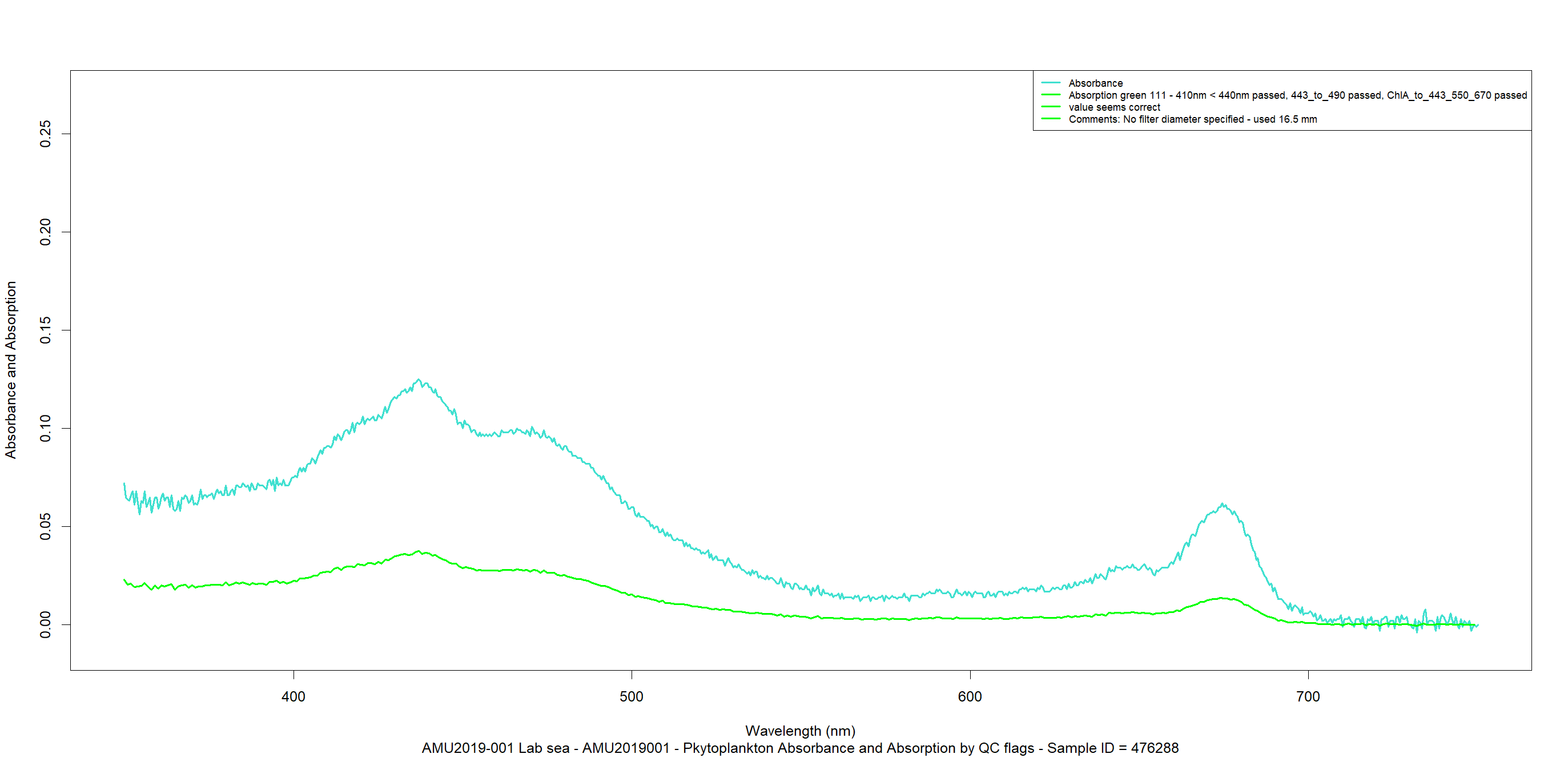
The QC plots legend: “C:\R\_WD\ QC\_plots\_Colors\_Legend.png”.





1. **"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”



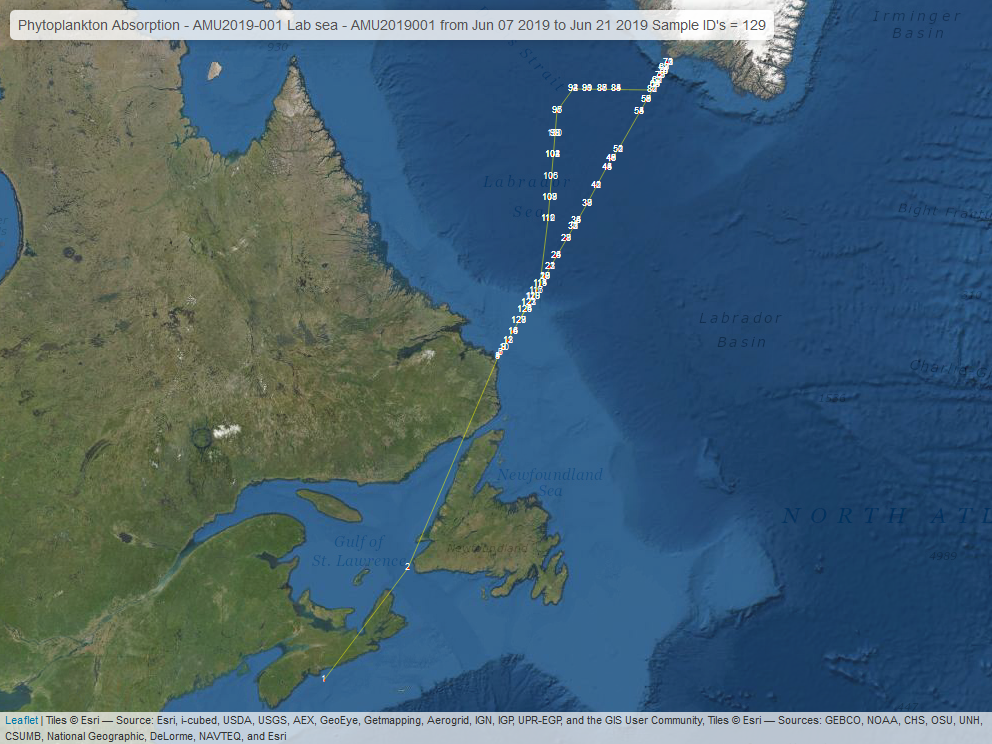
1. **“003\_009\_plot\_Phytoplankton\_final\_QC\_Comments\_cruises\_samples\_on\_map\_14.R”**

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.



1. **"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.

1. **“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"

# 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

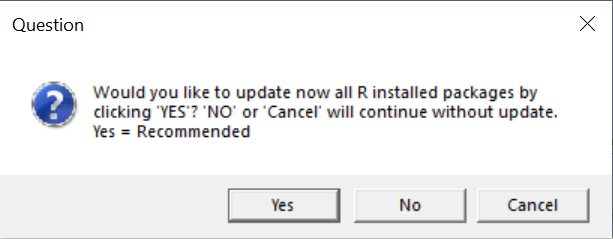
[67] data.table\_1.14.0, png\_0.1-7, vctrs\_0.3.7, httpuv\_1.5.5, cellranger\_1.1.0, purrr\_0.3.4

[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

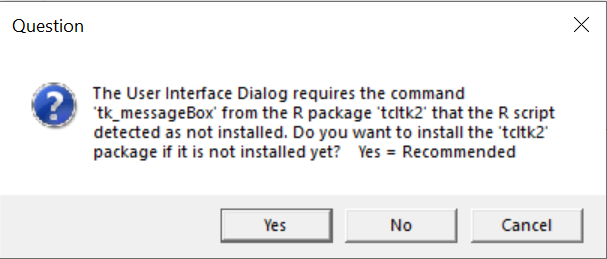
[79] tibble\_3.1.1, tinytex\_0.31, proftools\_0.99-3, units\_0.7-1, ellipsis\_0.3.1

# User Interface Dialogs

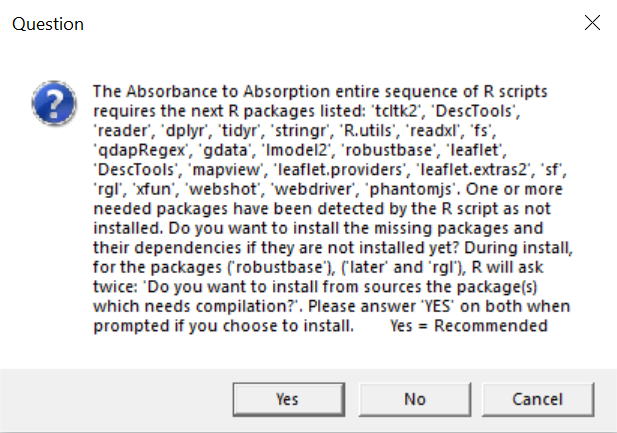
1. Every time the R script runs, from the beginning, it will open the dialog below offering to update all the packages that are already installed; it may take few minutes if packages were not updated in a while. If the R version is not 4.0.4, additional messages will appear to inform the user to choose “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).



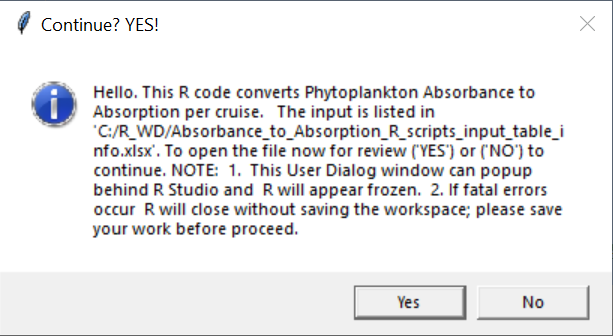
1. For the second dialog the script will check if one of the required packages named “tcltk2” is installed; if it is not, the dialog below is shown:



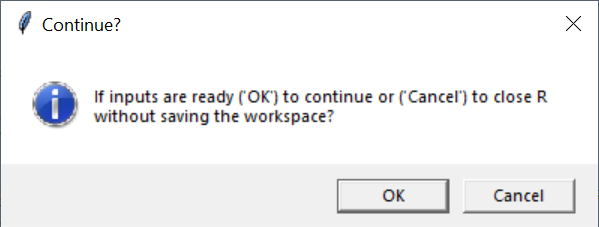
1. The script will check if all the needed packages are installed; If at least one is missing, this dialog below pops up. Choosing “Yes” will install only the missing packages listed and their dependencies. This step may take a few minutes to complete.



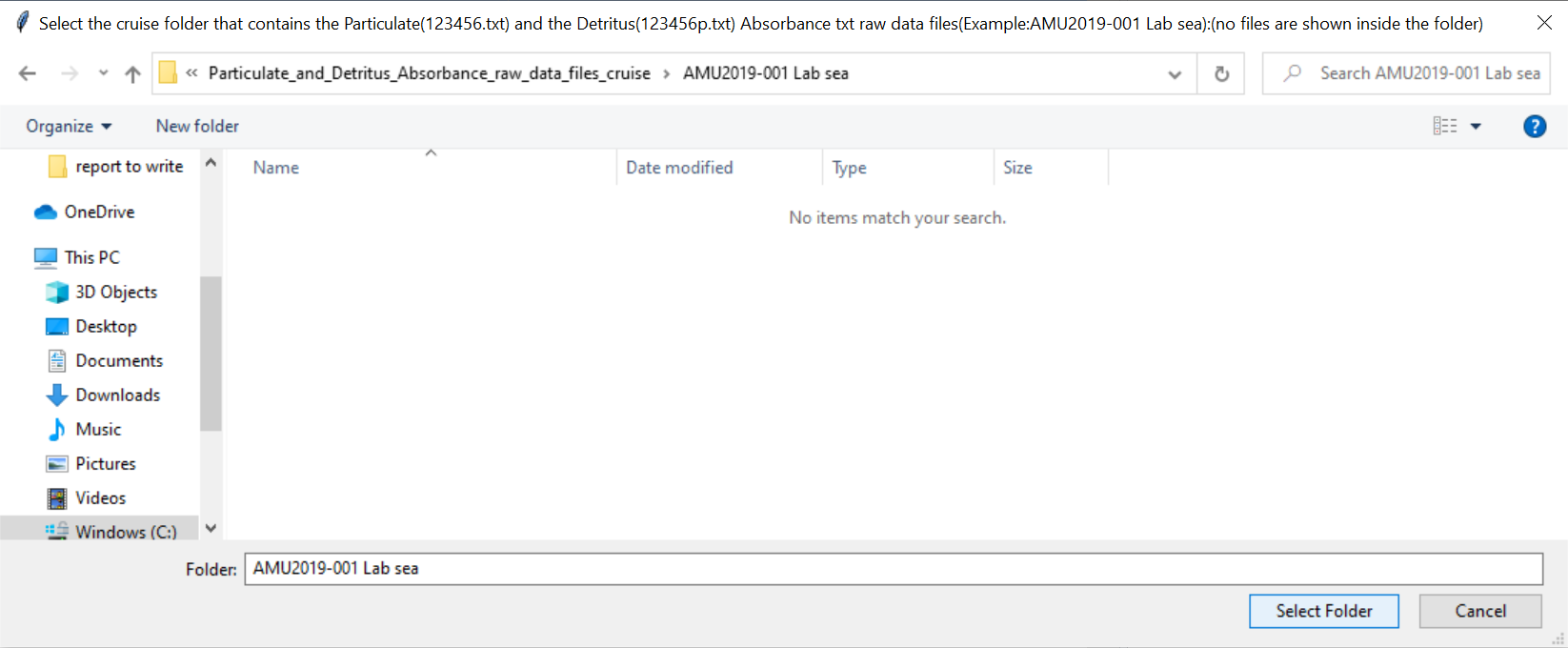
1. The second and the third dialog windows above will only pop up on the first run. If all needed packages are installed, the script will show only this fourth dialog. Choosing “Yes” will open the input table requirements for review; “No” will just move forward without opening the input table:



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

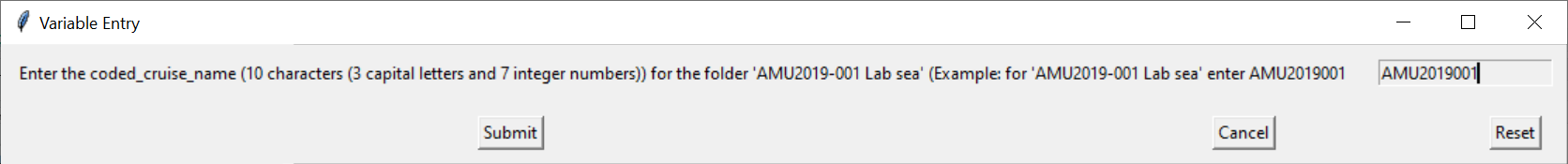


1. This window will ask to “Select the cruise folder that contains the Particulate(123456.txt) and the Detritus(123456p.txt) Absorbance txt raw data files (Example:AMU2019-001 Lab sea); no files are shown inside the folder – navigate to the folder and click “Select Folder”:

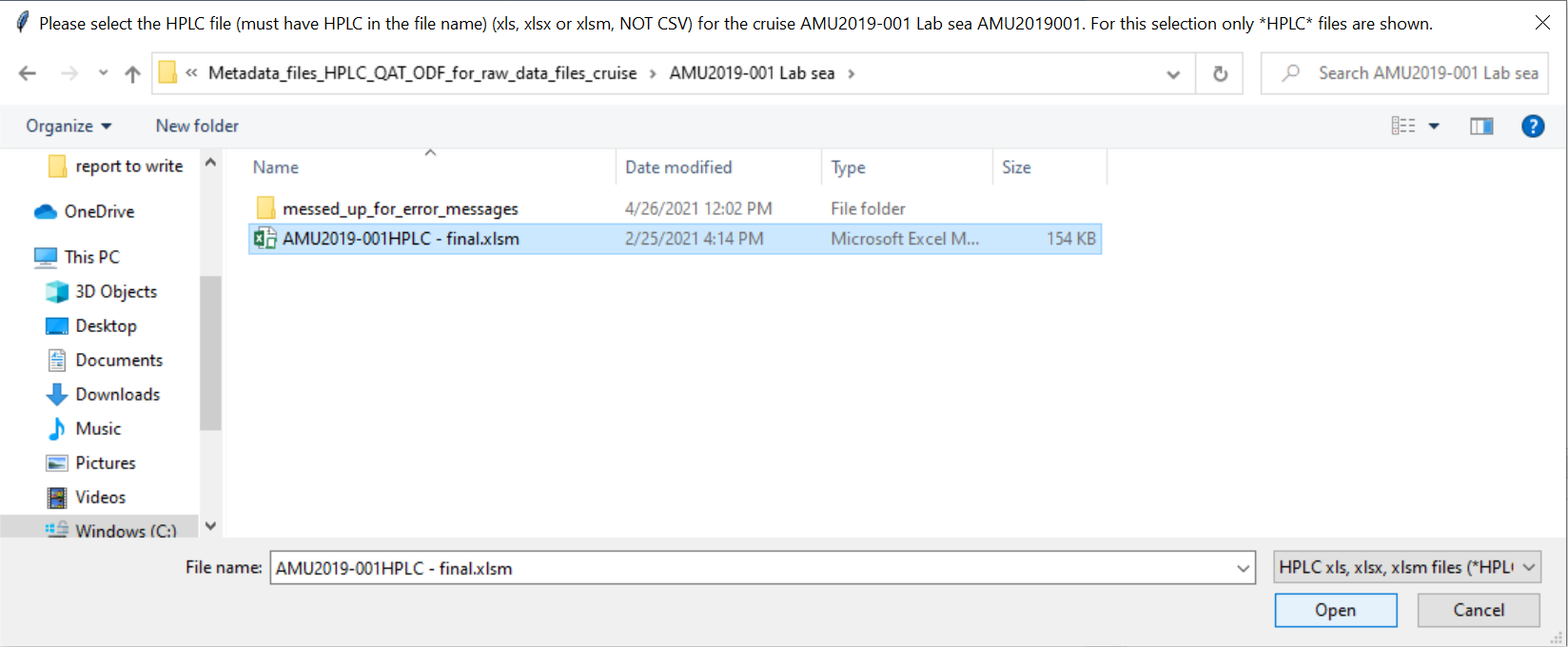


1. Enter the coded\_cruise\_name (10 characters (3 capital letters and 7 integer numbers)), (also known as cruise name or cruise number) for the folder selected above.

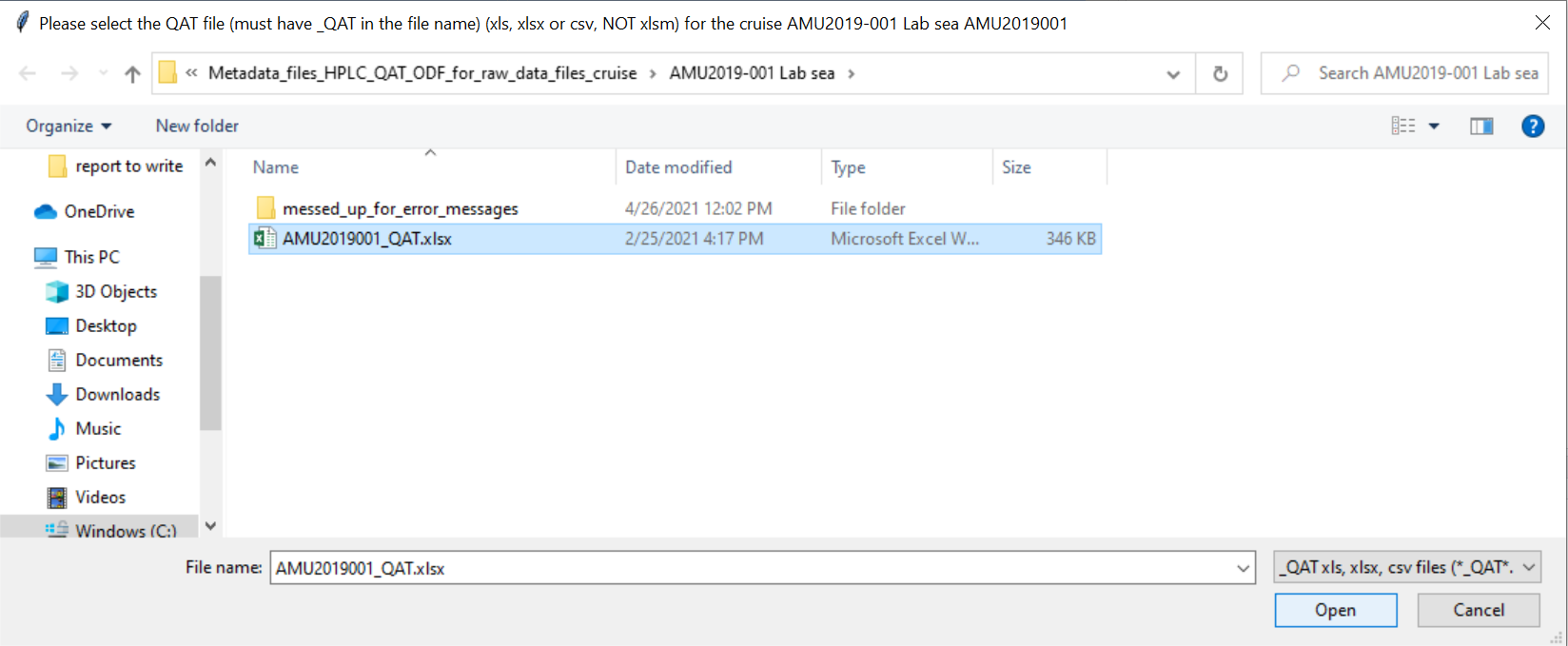
(Example: for “AMU2019-001 Lab sea” enter “AMU2019001"):



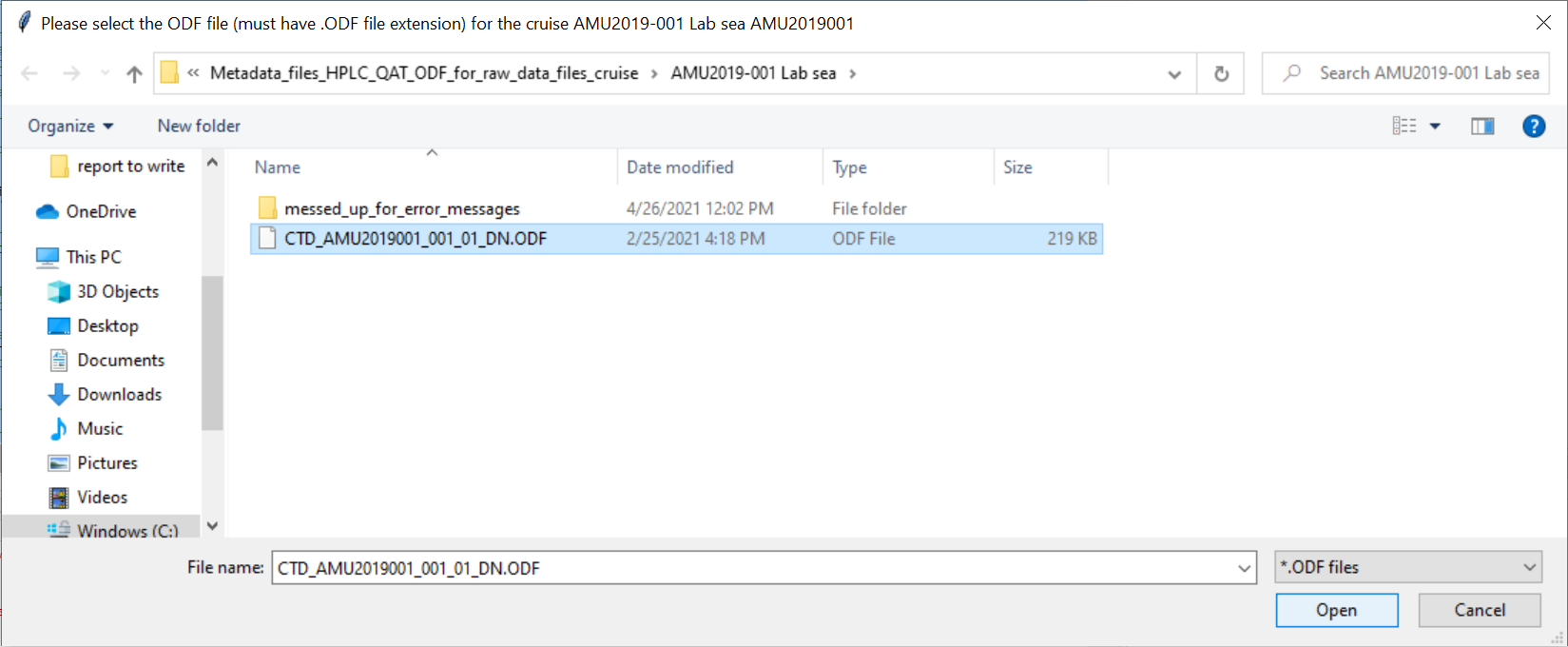
1. This window will ask to select the HPLC file (must have HPLC in the file name) (xls, xlsx or xlsm, NOT CSV) for the cruise selected for processing and click “Open”.



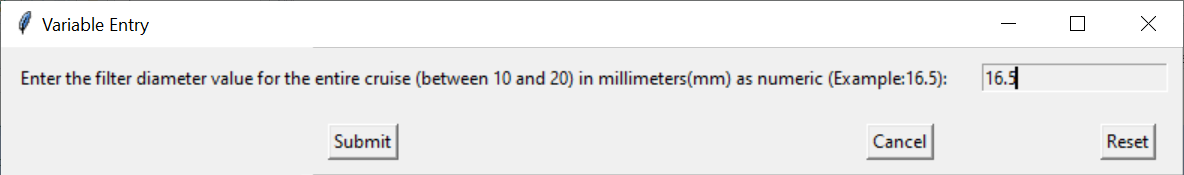
1. This window will ask to select the QAT file (must have \_QAT in the file name) (xls, xlsx or csv, NOT xlsm) for the cruise selected for processing and click “Open”.



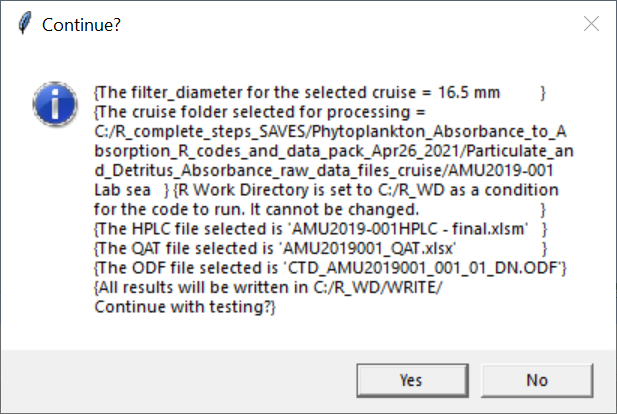
1. This window will ask to select the ODF file (must have .ODF file extension) for the cruise selected for processing and click “Open”. For this selection only \*.ODF files are shown:



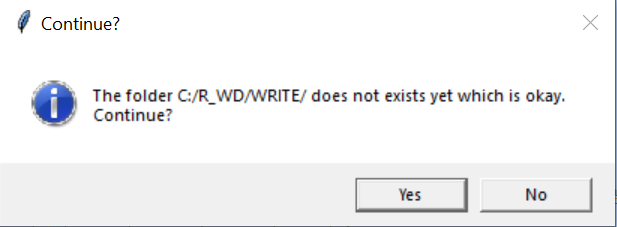
1. Enter the filter diameter value for the entire cruise (between 10 and 20) in millimeters(mm) as numeric (Example:16.5) and click “Submit”:



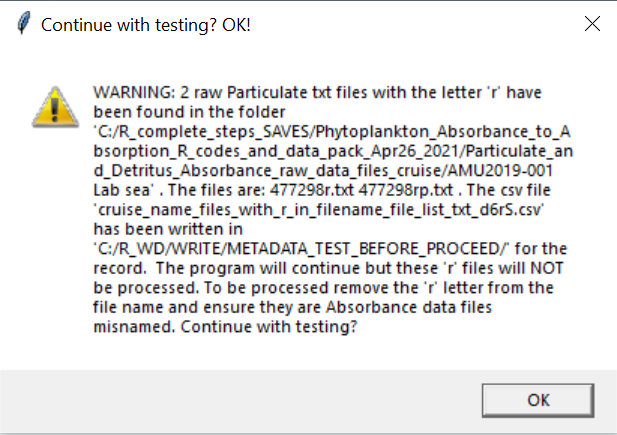
1. “Continue testing” confirmation dialog which summarises all the user inputs, click Yes to continue:



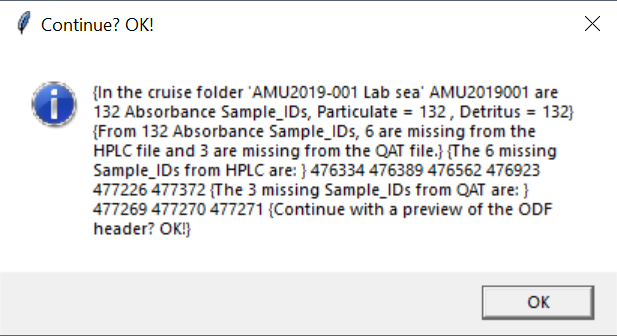
1. The script writes all results and reports in “WRITE” folder; Checking if the folder exists before writing to avoid overwriting, click Yes to continue, it will create the folder:



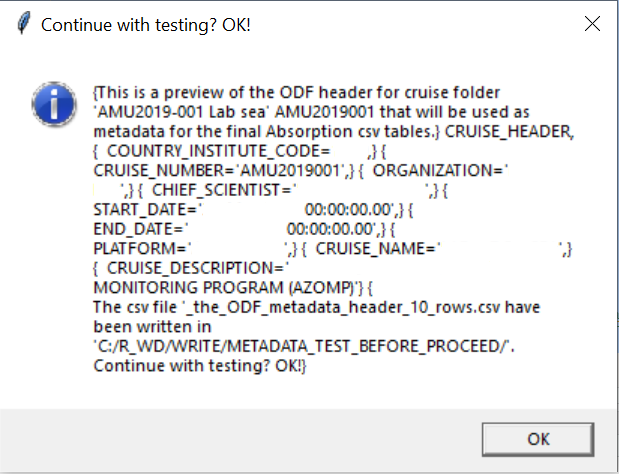
1. The scripts checks if there are raw data files with the letter “r” in the filename within the cruise folder selected for processing. The program will continue but these “r” files will NOT be processed. The files with “r” in the name are listed. To be processed remove the “r” letter from the file name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



1. 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;

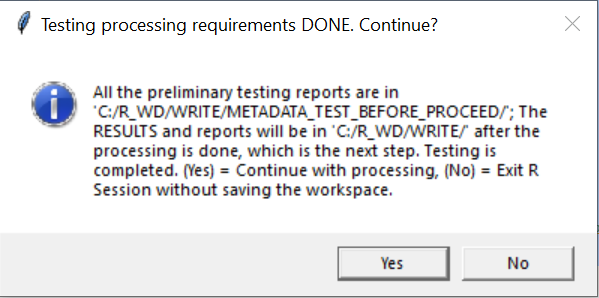


1. Displays a preview of the ODF header used as metadata for the final Absorption csv table file (data has been removed for this preview). Review the metadata, click OK to continue:

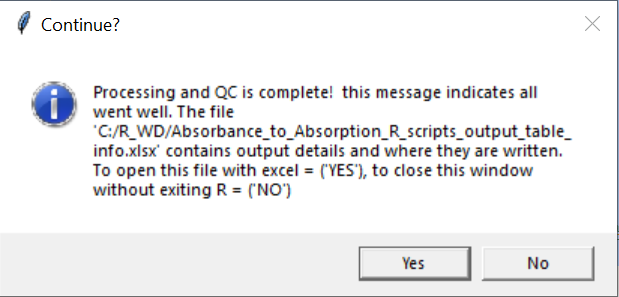


1. Next window pops up when all the data and metadata has been checked and accepted as valid and is ready to be processed; All reports generated so far are in “C:/R\_WD/WRITE/METADATA\_TEST\_BEFORE\_PROCEED/”; by clicking:

* “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.



1. 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!