SPATIAL Lab R scripts

General R tips

* Download R
  + <http://www.r-project.org/>
  + Click on download R
  + Choose a geographic location close to you (for the western US, I usually use the one at Berkeley)
  + Click on the link under your preferred location
  + Choose the correct download for your system
* Download RStudio
  + <http://www.rstudio.com/products/rstudio/>
  + Download RStudio desktop for your system
* Install the required packages to run these scripts: “RODBC” & “xlsx”
* To install packages, open RStudio and type in install.packages(“name.of.package”), eg. install.packages(“ROBDC”)
* When entering filepaths into R, R likes forward slashes. If you copy and paste from the Windows file browser, it will have back slashes. You will need to either change all the back slashes to forward slashes or add an additional back slash to each back slash (eg. C:\Users\kblevins\Desktop becomes either C:/Users/kblevins/Desktop or C:\\Users\\kblevins\\Desktop)
* If you get error messages, it can sometimes be helpful to copy those into Google and see how other people have dealt with them. Some helpful websites for R is http://stackoverflow.com/

Processing CRDS liquid data

General workflow

Three R files are used in processing liquid water data from the Picarro:

* 1. CRDS\_liquid\_1.R – contains algorithms that correct the data, shouldn’t need to be opened on a regular basis
  2. CRDS\_liquid\_2.R – contains functions that organize the data, shouldn’t need to be opened on a regular basis
  3. CRDS\_liquid\_3.R – should be opened each time data is processed, this file is meant to document how each dataset is processed

After running liquid water samples on the Picarro, open the CRDS\_liquid\_3.R file. Each entry in this file contains the code used to process the data from one machine and one run. Follow the instructions in the R file to process the data.

A set of example inputs, R script and outputs can be found at

Dropbox\ Bowen\_Lab\R\_scripts\_active\CRDS\_liquid\_example\_files

If the machine stops picking peaks during a run, the data may still be available. If this happens, use the peak\_picking.R script before processing the data as usual.

Instructions for using peak\_picking.R

1. On the Picarro computer save the folders from C:\UserData\DataLog\_User for the date(s) the samples were run into a date-specific folder in the machine-specific Dropbox folder (eg. Dropbox\hids2046\150226).
2. Create a folder at Dropbox\ Bowen\_Lab\CRDS\_liquidH2O\Peak\_Picking\_Files named with the date & machine (eg.

Dropbox\ Bowen\_Lab\CRDS\_liquidH2O\Peak\_Picking\_Files\150226\_HIDS2046).

1. Open the peak\_picking.R file and save it as a new file with the date appended to the end of the filename (eg. peak\_picking\_150226.R).
2. Follow the instructions in the R script to process the data.
3. If there is a partial file from the run that the Picarro did produce, compare the results in the csv file produced by the R script to the values produced by the Picarro to ensure they are very close.
4. Copy the data out of the csv file into the partial Picarro file in the appropriate lines, updating the Port & Injection Number columns manually. The Line number column in the Picarro file should match the Peak Number column in the output from the R script.

A set of example inputs, R file and outputs can be found at

Dropbox\ Bowen\_Lab\R\_scripts\_active\peak\_picking\_example\_files

NEON Receiving

When NEON samples are received, the spreadsheet they email should be updated with receiving information and then uploaded to the database using the R script NEON\_shipping\_processing.R.

1. In the perSample tab of the spreadsheet, add the following columns & populate for each row
   1. receivedDate (eg. 20141219)
   2. receivedBy (eg. Kali Blevins)
   3. shipmentCondition (eg. Good)
   4. receivedRemarks (eg. Sample vial broken during shipment)
2. Open the NEON\_shipping\_processing.R file and follow the instructions in this file to upload the data from the spreadsheet to the database.

Reporting CRDS liquid data

Three different types of reports can be generated using the R script report\_process.R. Spatial lab reports, SIRFER lab reports and NEON lab reports.

1. Open the report\_process.R file.
2. Follow the instructions in this file to generate the desired report.
3. If it is a NEON report, the R script will produce a final data report in the folder

Dropbox\Bowen\_Lab\Data\_reports\NEON

* 1. NEON data reports are to be uploaded by the SPATIAL lead technician to the NEON Box share using the account access provided by NEON

1. If it is a Spatial lab report, you will need to modify the excel file output by the R script.
   1. Open the output file, there will be two tabs, one with a table for data and one with a table for the qaqc parameters.
   2. Open the ClientCRDS\_template.xlsx file
   3. Delete the header row from the output file
   4. Copy and paste the two header rows from the template file into the first tab of the output file
   5. Copy and paste the qaqc parameters table from the second tab of the output file into the first tab in the appropriate place
2. If it is a SIRFER lab report, you will need to modify the excel file output by the R script.
   1. Open the output file, there will be one table in each tab and one tab per run
   2. Open the SIRFER template file
   3. Paste in the header rows from SIRFER template into each of the tables in the output file
   4. Update the header with the missing information. The date is labeled on the tab name.
   5. Paste in the missing formula rows from the SIRFER template into each of the tables in the output file
   6. If a full run of 67 vials was not run, you will need to update the formulas calculating averages & standard deviations for PT because it will be using too many cells
   7. Copy & paste the formatting for the remaining rows from the SIRFER template to the output file
   8. Cut and paste all of the tables from the different tabs in the output file into the first tab (there is one tab per run)

A set of example inputs, R script and outputs can be found at

Dropbox\ Bowen\_Lab\R\_scripts\_active\peak\_picking\_example\_files

Contents of R scripts

1. **CRDS\_liquid\_1.R** -- contains functions with algorithms used to process data, including memory correction, calibration to internal standards, drift correction and quality assessment
   1. data.mod function -- reads in the relevant files, modifies and merges them so they can be used in the subsequent functions
   2. mc.terms function -- generates memory-correction terms for d18O & d2H separately using the specified subset of ports and the specified subset of injections
   3. data.mc function -- generates memory-corrected data for the specified element using specified list of d18O and d2H memory correction
   4. cal.reg function -- calculates a regression for d18O and d2H between the known for plrm1 and plrm2 and returns the and intercept to be used in calibrating measured values in the
   5. data.cal function -- generates calibrated data the specified element using the specified list of d18O and d2H calibration slopes & intercepts
   6. drift.reg function -- calculates a regression line for d18O and d2H between the mean value each port with the slrm reference water the sequence of that port in the run
   7. data.dc function -- generates drift-corrected for the mean d18O and d2H values for each port
   8. qa.flag function -- evaluates the mean and sd for slrm to determine if they are within acceptable limits to determine the overall quality of the run, as well as the sd for each port to the data quality for each
   9. qa.summary function -- summarizes the qa metrics for the run
2. **CRDS\_liquid\_2.R** -- contains functions to look up the names needed and run the functions from CRDS\_liquid\_1.R, format the data so it can be printed, plot the data, upload the data to the database and write the data to an excel file
   1. file.lookup function -- retrieves file names for a given machine and date using regular expressions
   2. check.files function -- checks the data & ids files and returns warnings if they are not formatted correctly
   3. process.data function This function runs the functions in ‘CRDS\_liquid\_correct\_function.R' and returns a list of 6 dataframes:
      1. samples.summary - averaged and drift-corrected data with quality flags non-reference samples
      2. slrm.summary - summary data for the ports with the slrm reference
      3. ref.all - all data for references, including raw, memory-corrected, and calibrated values
      4. data.all - all data for samples, including raw, memory-corrected, and calibrated values
      5. qa.report - summary of qa parameters for the run
      6. ref.compare - comparison of the mean of the 1st 4 & last 4 injections for the references in ports 2-4
3. **CRDS\_liquid\_3.R** -- this is a file to use for running the functions to process data for CRDS liquid water runs. Each entry processes the data from one machine and a run started on one date.
4. **NEON\_shipping\_function.R** -- contains a function that can be used to upload data from a modified NEON spreadsheet to the WaterDB database in the NEON\_shipping table
   1. neon\_shipment – reads in data from a modified NEON spreadsheet and uploads it to the NEON\_shipping table in the WaterDB Access database
5. **NEON\_shipping\_processing.R** -- this is a file to use for running the neon\_shipment function each time a shipment is received from NEON
6. **Peak picking.R** -- the code in this file can be used to analyze a folder with dat files generated by the Picarro. First it reads in all of the files in a folder (including subfolders). Next, it calculates the slope for h2o vs. time. Then it picks all of the peaks. It saves a plot showing the picked peaks. Finally, it averages the isotope values for each peak. This script is meant to be run
7. **report\_functions.R** -- contains a function that is used to generate SIRFER, SPATIAL, and NEON data reports and plot data for a group of samples
8. **report\_process.R** -- this is a file to use for running the reports function each time a report needs to be generated