



Folder structure

Name

- Data0_Original
- Data1_RAW
- Data2_QC
- Data3_Aggregated
- Data4_Stats
- Scripts
- 000a_InstallLibraries.R
- 000b_InstallPandox.R
- 00_MasterScript.R
- Site_QClogs.xlsx

- **Original**
 - Master files. Do not change these! If you make a mistake and need to start over, you'll need these (unaltered) files.
- **1_RAW**
 - Files are copied from the Original folder and pasted here. These are your working files. Files need to be edited so that they have a standardized format and file naming scheme, and are .csv.
- **2_QC**
 - When the QC portion of the MasterScript is run, the QC data outputs will go here (2 outputs: Word & .csv).
- **3_Aggregate**
 - When the Aggregate portion of the MasterScript is run, the outputs go here (2 outputs: Word & .csv).
- **4_Stats**
 - When the SummaryStats portion of the MasterScript is run, the outputs go here (3 outputs: 2 .csv, 1 PDF)
- **Scripts**
 - R files that you will rarely need to work with, with the possible exception of the fun.UserDefinedValues.R file

There are 5 places in the MasterScript where you'll need to make entries - operation, SiteID, data type and data range. After making the entries, highlight the script and hit Run.

```
11 # QCReportAggregate - Performs QC on Aggregated data. Import Dir = Data3Aggregated. Export Dir = Data3Aggregated.
12 # SummaryStats - Generates summary stats on Aggregated data. Import Dir = Data3Aggregated. Export Dir = Data4Stats.
13 #####
14 # clear the workspace
15 rm(list=ls())
16 # define working directory
17 # if specify directory use "/" not "\" (as used in windows) and leave off final "/" (example below).
18 #myDir.BASE <- "C:/Users/Erik.Leppo/Documents/NCEA_dataInfrastructure/Erik"
19 myDir.BASE <- getwd()
20 setwd(myDir.BASE)
21 # library (load any required helper functions)
22 source(paste(myDir.BASE,"scripts","fun.Master.R",sep="/"))
23 #####
24 # USER input in this section (see end of script for explanations)
25 #####
26 #
27 # PROMPT; Operation
28 Selection.operation <- c("GetGageData","QCRaw","Aggregate","SummaryStats")
29 myData.operation <- Selection.operation[3] #number corresponds to intended operation in the line above
30 #
31 # PROMPT; Site ID
32 # single site;
33 # group of sites;
34 myData.SiteID <- "test2"
35 #
36 # PROMPT; Data Type
37 # type of data file
38 Selection.type <- c("Air","water","AW","Gage","AWG","AG","WG") # only one at a time
39 myData.Type <- Selection.type[3] #number corresponds to intended operation in the line above
40 #
41 # PROMPT; Start Date
42 # YYYY-MM-DD ("-" delimiter) leave blank for all data ("1900-01-01")
43 myData.DateRange.Start <- "2013-01-01"
44 #
45 # PROMPT; End Date
46 # YYYY-MM-DD ("-" delimiter) leave blank for all data (today)
47 myData.DateRange.End <- "2013-12-31"
48 #
49 # PROMPT; subdirectory, input file location. Leave blank for defaults
50 Selection.SUB <- c("Data1_RAW","Data2_QC","Data3_Aggregated","Data4_Stats")
51 myDir.SUB.import <- "" #Selection.SUB[2]
52 #
53 # PROMPT; Subdirectory, output file location. Leave blank for default.
54 myDir.SUB.export <- "" #Selection.SUB[3]
55 #
56 #####
57 # Run the script with the above user defined values
58 fun.Master(myData.operation
59           ,myData.SiteID
60           ,myData.Type
61           ,myData.DateRange.Start
62           ,myData.DateRange.End
63           ,myDir.BASE
64           ,myDir.SUB.import
65           ,myDir.SUB.export)
66
```

Selection.Operation
[1] = GetGageData
[2] = QCRaw
[3] = Aggregate
[4] = SummaryStats

Enter the SiteID

Selection.Type
[1] = Air
[2] = water
[3] = AW (air & water)
[4] = Gage
[5] = AWG (air, water, gage)
[6] = AG (air & gage)
[7] = WG (water & gage)

Enter start and end date (need to use this format - e.g., "2013-01-01" (start) and "2013-12-31" (end). To pull all available data for a site, delete the date so it's blank, except for the start and end quotes ""

Formatting - the MasterScript reads through the files and pulls out data based on file names (siteID, dates, data type). This allows for multiple files to be kept in the same directory. In order for the R script to work, the **files need to be named and formatted a certain way.**

Column headings (do not have to be in any specific order)

Required

- **SiteID**
- **“Date Time”** (one field) OR **“Date” AND “Time”** (in 2 separate fields) OR **all 3** (in 3 fields)
 - Prefer 24H time (i.e., military). R scripts will assume any time without AM/PM is 24 hr time.
- **Parameters**
 - Water Temp C
 - Air Temp C
 - Water BP psi
 - Air BP psi
 - Water Level ft

Not Required

- Discrete measurements - column heading needs to have the prefix "Discrete" – e.g., “Discrete Water Temp C”
- Air LoggerID
- Water LoggerID
- Air RowID or Water RowID (this shows up as # in the HoboWare output; you cannot leave it as “#” because R will treat this as a comment and ignore the entire line)
- Notes

Examples of column headings on properly formatted files

Non-vented pressure transducer (after water and air sensor data were combined in HoboWare)

A	B	C	D	E	F	G	H	I	J
AW RowID	SiteID	Date Time	Water Temp C	Air Temp C	Water Level ft	Water BP psi	Air BP psi	Water Logger ID	Air Logger ID
1	ECO66G12	4/26/2013 10:30	10.063	15.282	0.985	14.8024	14.3638	10229557	10229571
2	ECO66G12	4/26/2013 11:00	10.357	14.134	0.854	14.7461	14.3644	10229557	10229571
3	ECO66G12	4/26/2013 11:30	10.748	14.23	0.855	14.7428	14.3603	10229557	10229571
4	ECO66G12	4/26/2013 12:00	11.139	15.473	0.853	14.7441	14.3627	10229557	10229571
5	ECO66G12	4/26/2013 12:30	11.528	17.284	0.854	14.7431	14.3612	10229557	10229571
6	ECO66G12	4/26/2013 13:00	11.916	18.806	0.848	14.7421	14.3628	10229557	10229571
7	ECO66G12	4/26/2013 13:30	12.207	19.662	0.855	14.7453	14.3629	10229557	10229571
8	ECO66G12	4/26/2013 14:00	12.594	20.424	0.853	14.7351	14.3537	10229557	10229571

Water temperature only

A	B	C	D	E	F	G	H
Water RowID	SiteID	Date	Time	Date Time	Water Temp C	Water LoggerID	Discrete Water Temp C
1	HRCC	19-Jun-13	18:00	6/19/2013 18:00	17.819	9734130	17.7
2	HRCC	19-Jun-13	18:30	6/19/2013 18:30	17.843	9734130	
3	HRCC	19-Jun-13	19:00	6/19/2013 19:00	17.796	9734130	
4	HRCC	19-Jun-13	19:30	6/19/2013 19:30	17.748	9734130	
5	HRCC	19-Jun-13	20:00	6/19/2013 20:00	17.629	9734130	
6	HRCC	19-Jun-13	20:30	6/19/2013 20:30	17.51	9734130	

File names in the Data1_RAW folder

Example (non-vented pressure transducer, with all parameters in one file):

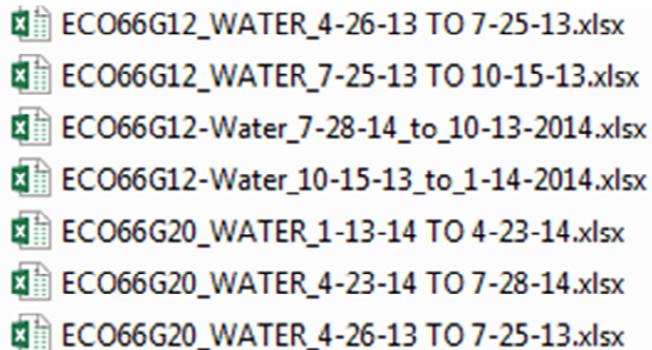
- **test2_AW_20130426_20130725.csv**
 - Site ID (no spaces or underscores) = test2
 - Data Type (Water/Air/AW/Gage) = AW (air and water)
 - Date, Start (YYYYMMDD) = April 26, 2013
 - Date, End (YYYYMMDD) = July 25, 2013
 - Each element separated by underscore (“_”).
 - CSV format (.csv), comma delimited.

What if it was just water sensor data? *Data type would be ‘_Water’ instead of ‘_AW’* (test2_Water_20130426_20130725.csv)

What if this was just air sensor data? *Data type would be ‘_Air’ instead of ‘_AW’* (test2_Air_20130426_20130725.csv)

What if this was USGS gage data? *Data type would be ‘_Gage’ instead of ‘_AW’* (test2_Gage_20130426_20130725.csv)

BEFORE



AFTER

