

Guide to website for continuous stream data QC and management

*David Gibbs (former EPA ORISE fellow), Britta Bierwagen (EPA NCEA/ORD)
Jen Stamp & Erik W. Leppo (Tetra Tech, Inc.)*

Version: 4/27/18

Guide outline

- Purpose
- Background
- Four main processes
- Instructions
 - Data preparation
 - Uploading data
 - Running QC/Aggregate/Summarize processes
 - Retrieving USGS gage data
 - Custom thresholds
- Troubleshooting
- Additional notes and resources

Purpose of website

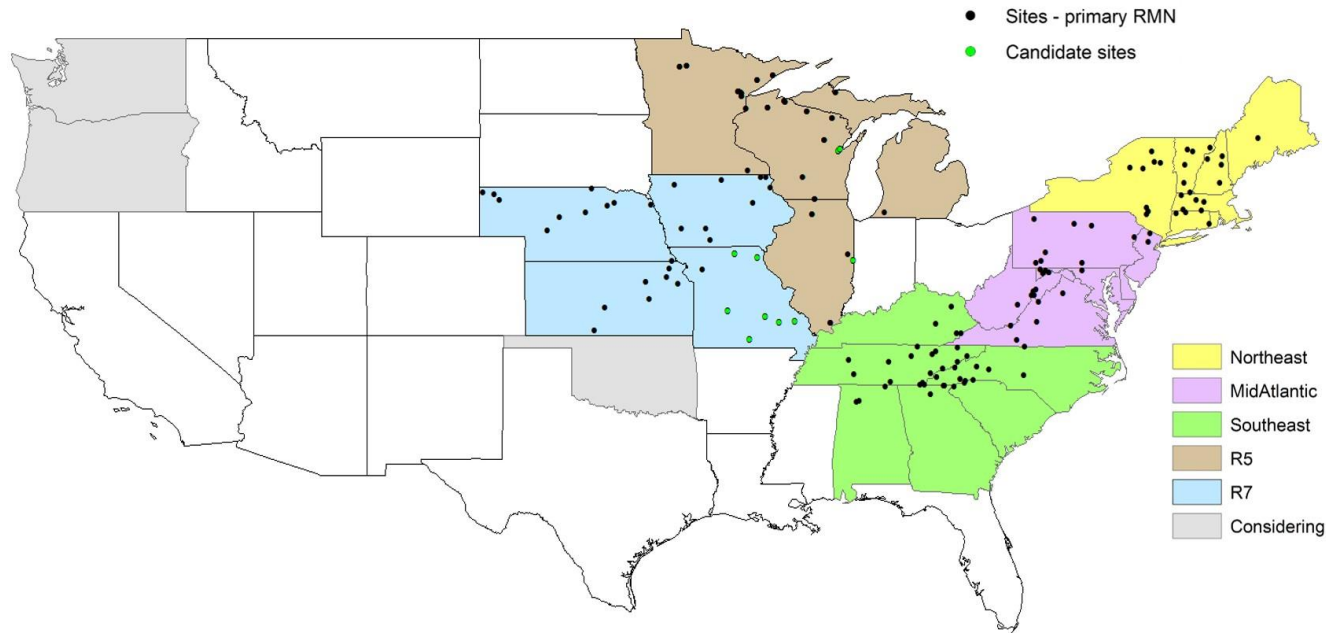
- Make QCing and getting summary statistics for continuous water sensor data as easy and consistent as possible
- Ensure that a certain (minimum) level of QC is performed on the continuous thermal and hydrologic RMN data
- Standardize and speed up the QC process
- Reduce errors in the QC process
- Standardize data for further analysis
- Facilitate data sharing

Background

Background:

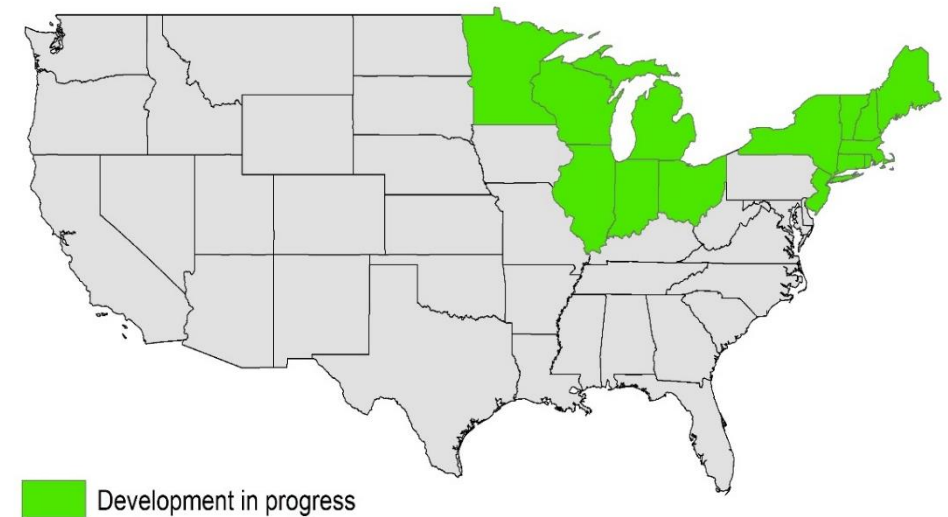
Regional Monitoring Networks (RMNs)

Streams



Sampling efforts began in the Northeast in 2012, followed by the Southeast in 2013, the Mid-Atlantic in 2014 and the Midwest in 2016-2017.

Lakes



Discussions in the Midwest and Northeast began in 2015. Some partners are planning to start piloting some protocols this year, with the goal of starting implementation in 2018.

RMN data challenges

- Lots of different data sources
- Different types of equipment, sometimes at different recording intervals, sometimes in different formats
 - **Standalone temperature** sensors (e.g. water and air temperature come in separate files)
 - **Non-vented pressure transducers** (e.g., temperature and water level data together in one output; sometimes air temperature in a separate file)
 - **Vented pressure transducers**
 - **USGS discharge data** (needs to be pulled from USGS website and combined with temperature data)



Stream RMNs – Data infrastructure

Biological data

- Macroinvertebrates
- Fish
- Periphyton

Continuous sensor data

- Temperature (air and water)
- Water level or sensor depth
- Discharge

Data source: RMN partners

Included in website

QC/processing prior to upload

QC: ContDataQC R package (4 tests) or other options that meet minimum RMN QC requirements

Upload to data storage system

Data management/storage system
(each entity needs its own system)

Query the data system
& download data to your computer

Request data from source*



Run data visualization and analysis tools**

*If the system meets certain standards and is made 'discoverable' to EPA OW's Central Catalog server, the user can view metadata for those entities (and then request desired data from the source; the EPA server will not house actual data)

**The web version of ContDataQC is under development. Also, we have not yet found a R Shiny server that can host it that is accessible to all RMN partners

ContDataQC R package

- Summary statistics and time series plots for continuous data
- Random subsampling of biological data
- PeriodStats – function that generates time series plots & stats for desired time periods preceding biological sampling events
- Outputs that are ready to run through several other relevant R scripts (CDFs, IHA metrics, thermal stats, etc.)

Included in website

QC (in accordance with RMN regional protocols)

+WQX/STORET
Water Quality Portal (EPA/USGS)

Query – 2 options:
1. WQ portal online interface
2. Data Discovery Tool (DDT)

+Temporary system:
MS Access database with RMN macroinvertebrate data; has queries that generate recommended RMN biological metrics; database will be posted on RMN Sharepoint site and Tt FTP site

Where does this website fit into your program?

It depends. Each RMN partner and region has its own unique situation; there is no 'one size fits all' solution. Maybe you'll use all the functions, or only one or two, or none.

Some of you already have a system, an SOP with QC procedures & tools for summarizing your data. **Others have nothing.**

The continuous data management website, our **data management tips**, the **RMN generic QAPP** and **training materials ensure everyone has at least something!**

This website is based on an R package called ContDataQC, written by Erik Leppo at Tetra Tech for the U.S. EPA Office of Research & Development. <https://github.com/leppott/ContDataQC>

Four main processes

Current website layout

Continuous data QC, summary, and statistics- PRELIMINARY WEBSITE

[Site introduction](#)

[Instructions & interface](#)

[Download USGS data](#)

[Advanced features](#)

[FAQ](#)

Background and features

Purpose of this site

One challenge of using continuous water monitoring devices is managing the tremendous amount of data they generate. One sensor recording one parameter every 30 minutes will produce over 17,000 records in one year. All of those records must be checked for errors before they can be used for further analyses. Following completion of quality control checks, creating summaries of the very detailed continuous records is important for conveying its key features. Unfortunately, the sheer number of records can make manual performance of QC and data summarization time-consuming and error-ridden and lead to backlogs in the analysis of valuable data.

This website helps expedite continuous data processing by performing several operations on such data. They allow users of continuous monitoring data to QC, combine, and summarize their continuous data files in a standardized way without having to download any programs to their computer. This website also allows users to download U.S. Geological Survey gage data at sites and over periods of their choosing. Collectively, these should reduce errors due to and time needed for manual processing of continuous data.

NOTE: This website is under development. New versions will be released periodically. E-mail the contacts at the bottom of this page for project updates. This website was last updated on 3/26/18.

Features of this website

Each feature below is described in more detail in the presentation found on the 'Instructions & interface' tab.

1. QC raw data:

Using this website, you can perform quality control checks on continuous stream data in a standardized way without having to download any programs to your computer. This website was developed for air and water temperature and pressure, sensor depth, and stream flow measurements. It has not been tested for other water parameters, such as conductivity, dissolved oxygen, or nutrients. You can process files from multiple sites at the same time but the more records you submit, the longer it will take for the website to process them.

This website performs four QC checks on each input parameter: unrealistic high/low values, spikes, fast rates of change, and values staying constant (not changing). Each value can pass (P), be flagged as suspect (S), or be flagged as failing (F). Whether each value is marked as P, S, or F (or X if the test is not applicable to that record) depends on the input threshold values for the QC tests. A file with default threshold values can be found in the 'Advanced features' tab. You can also upload your own custom threshold spreadsheet on that tab. Although this website performs QC checks on the data you input, it is up to you to decide how to respond to any erroneous or suspect values. The website does not change your values for you. For each input file, you receive two output files: a spreadsheet with QC flags for each record, and a summary report (html document).

2. Aggregate QC'ed data:

This website can combine spreadsheets that have been through the QC process in two different ways: by date or by data type. By date: This website can combine multiple QCed spreadsheets from the same site with the same parameters covering different time periods (e.g., combine 2/8/14-4/15/14 and 4/16/14-7/17/14 into a single spreadsheet covering 2/8/14-7/17/14). In this case, the files being aggregated should not have overlapping records (i.e. the later input should start after the end of the first input file). By parameter: This website can combine multiple QCed spreadsheets with different parameters from the same time period at the same site into a multi-parameter spreadsheet (e.g., separate air and water temperature spreadsheets from 7/1/15 to 9/30/15 into an air-water temperature spreadsheet over that same time period).

3. Produce summary statistics and plots of QCed data:

Each parameter input to this operation produces three summary output files. 1. A spreadsheet with daily average values. 2. A spreadsheet with annual, seasonal, monthly, and daily averages, medians, minima, maxima, ranges, standard deviations, and more. 3. A pdf with graphs of summary statistics by day, month, season, and year.

4. Download USGS gage data:

You can input USGS gage IDs and a date range and the website will download a separate csv for each gage over that time period. See the 'Download USGS gage data' tab for more information.

Further information

Current website layout

Continuous data QC, summary, and statistics- PRELIMINARY WEBSITE

Site introduction

Instructions & interface

Download USGS data

Advanced features

FAQ

Choose files

Browse...

No file selected

Choose operation to perform

Instructions

Summary tables

Instructions

Below are abbreviated instructions for the three operations on this tab: QCing raw data, aggregating QC'ed data, and summarizing QC'ed data. For more complete information on managing continuous data, preparing data for this website, understanding outputs, and troubleshooting, please refer to [this presentation](#). If you want to try using this website on previously tested files, you can use these [test files](#).

QC raw data:

1. Convert all the spreadsheets you will upload to this website into csvs.
2. Name your input files as follows: SITENAME_DATATYPE_STARTDATE_ENDDATE.csv. The site name should match the site name in the input files. Data types are as follows: A (air), W (water), G (flow), AW, AG, WG, and AWG. Start and end dates should match the dates in the input files and have the format YYYYMMDD (e.g., 20151203). Some example input file names are: 097_A_20150305_20150630.csv, GOG12F_AW_20130426_20130725.csv, and BE92_AWG_20150304_20151231.csv.
3. Download the continuous data [template](#). In order for this website to correctly process your continuous data, you need to format your data and column names following this template.
4. Copy the appropriate column headers from the template into the spreadsheet(s) you want the website to process. The only required fields to run the QC process are 'Date Time', 'SiteID', and at least one measurement column (air, water, sensor depth, or flow). It does not matter what order the columns are in within your spreadsheet(s).
5. Verify that the data in your spreadsheets are in the same formats as the data in the template spreadsheet (e.g., that the values in 'Date Time' are formatted the same as in the template). The website will not work on your data if the formats are incorrect.
6. Delete any extraneous columns from your spreadsheets.
7. Delete any extra header rows besides the ones with the field names. Delete any rows at the bottom of the spreadsheets that show termination of the sensor log.
8. Upload your files to the website using the 'Browse' button to the left.
9. Verify that the files are being interpreted correctly in the tables in the 'Summary tables' tab. If they are not showing as expected or if the table is replaced by error messages, it means that something is wrong with your input file(s), e.g., a column heading is incorrect.
10. Select 'QC raw data' using the drop-down menu to the left. A 'Run operation' button should appear below the operation selection drop-down menu.
11. Click the 'Run operation' button. A progress bar will appear in the bottom-right of the browser tab. It will advance after each file is processed. Thus, if you upload three files, it will wait at 0%, jump to 33%, jump to 66%, and then jump to 100%. It will not advance while each file is being processed.
12. Once the process is completed, a 'Download' button will appear below the 'Run operation' button. Click the button to download a zip file of all output files (spreadsheets and QC reports). Where the files will download on your computer depends on the configuration of your internet browser.

Aggregate QC'ed data:

1. Upload your QC'ed files to the website using the 'Browse' button to the left. NOTE: Files must have gone through the QC operation on this website before the 'Aggregate' operation can be used on them.

What does the website do?

Performs four main continuous data management operations without needing R on your computer.

The first operation is found on the third tab of the website.

The second, third, and fourth operations are found on the second tab of the website.

Operation	Description
GetGageData	Quick download of USGS data
QCRaw	Generates QC reports for temperature or hydrologic data
Aggregate	Merges files from different time periods (for a single site); for example, if you download your data quarterly, you could use the Aggregate function to join the four files together. Can also be used to subset files by date.
	Merges files that contain different parameters for overlapping different time periods (for a single site); for example, if you deploy proV2 temperature sensors at a site to measure air and water temperature, you could use the Aggregate function to bring the parameters together into one file (assuming time periods overlap), and could also potentially bring in USGS gage data as well
SummaryStats	Generates summary statistics, time series plots and other visualization tools

Website instructions

Data preparation

Before you get to the website...

Time saving tips

- For **configuration/launch** (see 'ConfigLaunch_HOBO_20170803' ppt in the 'DataProcessing_PreQC' folder of the RMN ftp site) –
 - For the **initial launch**, consider using the '**Delayed start**' option so that the sensor has time to calibrate before it starts collecting data (otherwise you will need to 'trim' those initial data points when you process the data later on)
 - If you are deploying air and water level loggers at a site, or multiple temperature sensors at the same site (e.g., on a fixed chain in a lake), **please make sure they are recording at the same time**. This will make data processing easier and will improve the quality of your data.

Before you get to the website...

Time saving tips

After downloading sensor data -

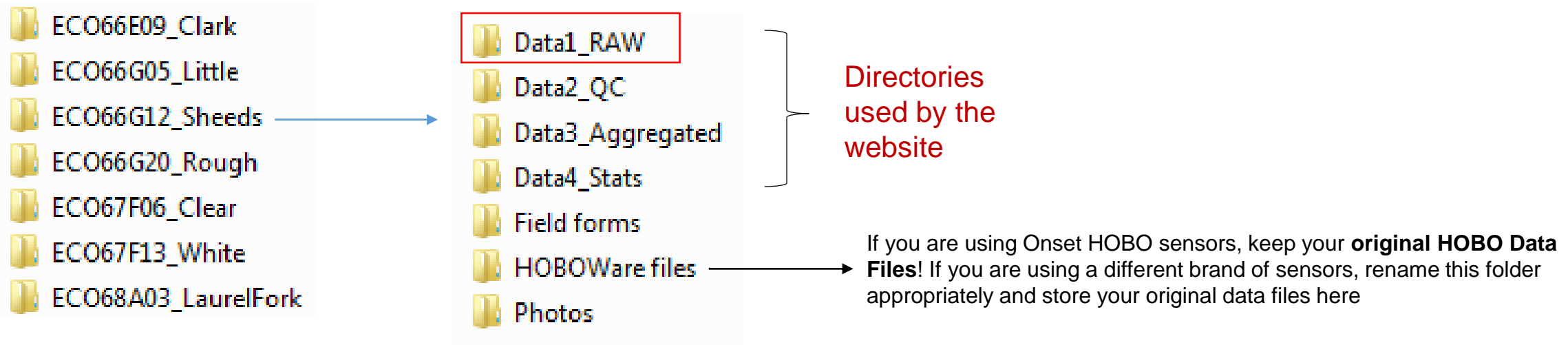
- If you are processing the data in HOBOWare, and are planning to use the website, consider using the recommended **default settings** (see appropriate ppts in the 'DataProcessing_PreQC' folder). The .csv file that you export will require fewer manual formatting edits; once you set defaults on your computer, HOBOWare will retain them.
- If you're planning to use the website, **do not put your air and water temperature files into separate worksheets in the same Excel file. Keep them as separate files.**
- **Better yet, bring the parameters together into one file before exporting the .csv file from HOBOWare** (this is recommended – it will save you steps later on; see the HOBOWare_Temperature_AW_20170824 ppt in the DataProcessing_PreQC folder).

Make sure you **keep your original HOBOWare Data Files** from the air and water sensors; this is important because if you have to go back and regenerate sensor depth in the future, you can only do that with Data Files, not Onset Project Files.

Make sure your files are backed up!

Recommended folder structure

- Each site gets its own folder
- Within that site folder, a separate sub-folder for different steps in the QC and management process



Data1_RAW – Upload spreadsheets from this folder to the website for QC

Data2_QC – Download spreadsheets and QC reports from the QC process to this folder

Data3_Aggregated – Download spreadsheets from the aggregate process to this folder

Data4_Stats – Download spreadsheets, reports, and summary plots from the summarize process to this folder

File formatting & naming

- File type: must be .csv
- File naming: SiteID_SensorType_StartDate_EndDate.csv
- Column headings (case sensitive) (**download data template from website to help with this**)
 - Required: SiteID, Date Time
 - Parameters (as available – not all are required and fields do not have to be in any specific order):
 - Water Temp C
 - Air Temp C
 - Sensor Depth ft
 - Air BP psi
 - Water P psi
 - Discharge
 - GageHeight
 - Optional: RowID, discrete, logger S/Ns
- Check for 'logged' entries in last rows; delete if needed
- Resave as .csv

More info on following slides

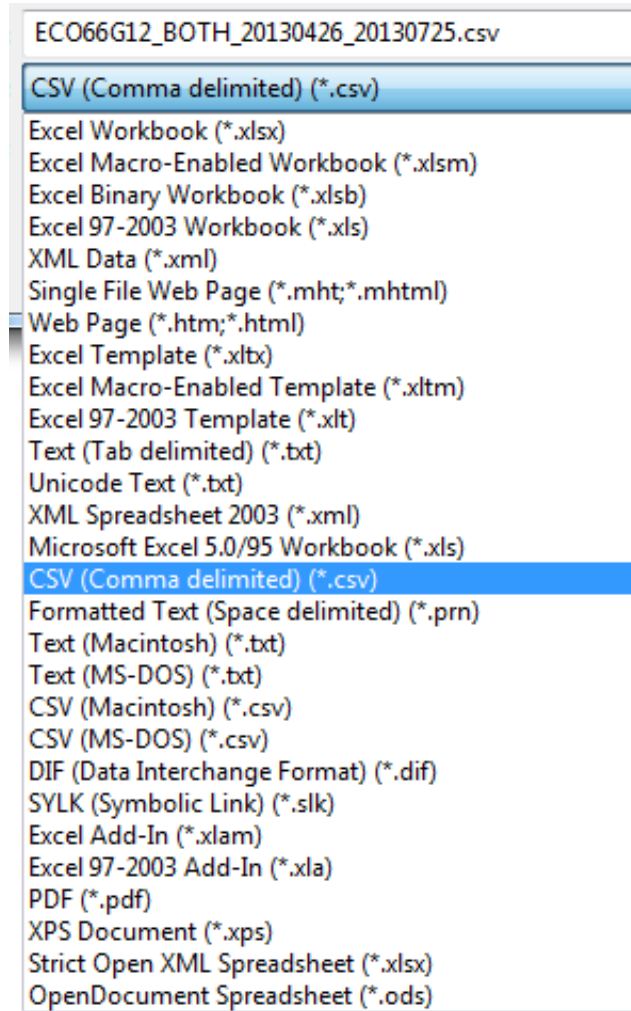
File type





.csv = comma delimited

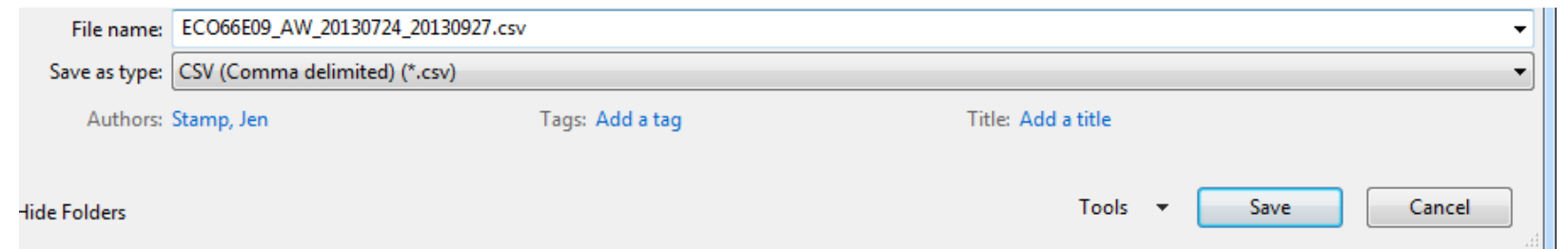
File

→ Save As

→ CSV (Comma delimited) (*.csv)



Name	Date modified	Type
 BB01CC_Air_20131022_20140428.csv	2017-08-23 20:07	Microsoft Excel Comma Separated Values File
 BB01CC_Air_20140428_20140924.csv	2017-08-23 20:12	Microsoft Excel Comma Separated Values File
 BB01CC_Water_20131022_20140428.csv	2017-08-23 20:08	Microsoft Excel Comma Separated Values File
 BB01CC_Water_20140428_20140924.csv	2017-08-23 20:15	Microsoft Excel Comma Separated Values File





File naming

Must follow a specific file naming convention.

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

- ECO66G12_AW_20130426_20130725
 - Site ID (no spaces or underscores) = ECO66G12
 - Data Type: W=water, A=air, G=gage, AW=Air/water, AWG=Air/water/gage
 - Date, Start (YYYYMMDD) = 20130426
 - Date, End (YYYYMMDD) = 20130725
 - Each element separated by underscore (“_”).

 ECO66G12_AW_20130426_20130725.csv
 ECO66G12_AW_20130725_20131015.csv

Column headings

Example: Non-vented pressure transducer (after water and air sensor data were combined in HOBOWare).
 The downloadable spreadsheet template on the website can help with this.

Before

1	Plot Title: ECO66G12					
2	Date Time, GMT-05:00	Abs Pres, psi (LGR S/N: 10229557)	Water Temp, °C (LGR S/N: 10229557)	Abs Pres Barom., psi (LGR S/N: 10229571)	Sensor Depth, feet (LGR S/N: 10229557)	Air Temp, °C (LGR S/N: 10229571)
3	2013-04-18 07:00	14.671	14.613	14.2005	1.086	14.421
4	2013-04-18 07:30	14.671	14.613	14.1998	1.088	14.709
5	2013-04-18 08:00	14.671	14.613	14.2017	1.084	15.473
6	2013-04-18 08:30	14.6733	14.613	14.2046	1.082	16.427
7	2013-04-18 09:00	14.6736	14.709	14.2051	1.082	17.379
8	2013-04-18 09:30	14.672	14.9	14.2027	1.084	17.76
9	2013-04-18 10:00	14.6704	15.091	14.2064	1.071	18.426

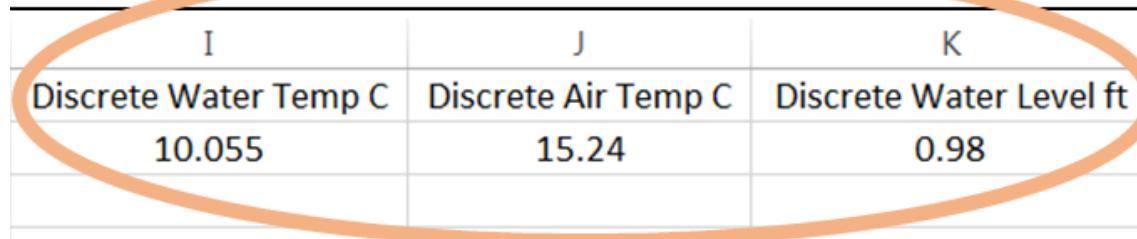


After

	A	B	C	D	E	F	G
1	SiteID	Date Time	Water P psi	Water Temp C	Air BP psi	Sensor Depth ft	Air Temp C
2	ECO66G12	2013-04-18 07:00	14.671	14.613	14.2005	1.086	14.421
3	ECO66G12	2013-04-18 07:30	14.671	14.613	14.1998	1.088	14.709
4	ECO66G12	2013-04-18 08:00	14.671	14.613	14.2017	1.084	15.473
5	ECO66G12	2013-04-18 08:30	14.6733	14.613	14.2046	1.082	16.427

Optional columns

- **Air LoggerID** – unique identifier (e.g., serial number) for land-based sensor
- **Water LoggerID** – unique identifier (e.g., serial number) for in-water sensor
- **Air RowID** or **Water RowID** (this shows up as # in the HOBOWare output; you cannot leave it as “#” because the website will treat this as a comment and ignore the entire line)
- **Discrete measurements** (must start with ‘Discrete’ as shown below)



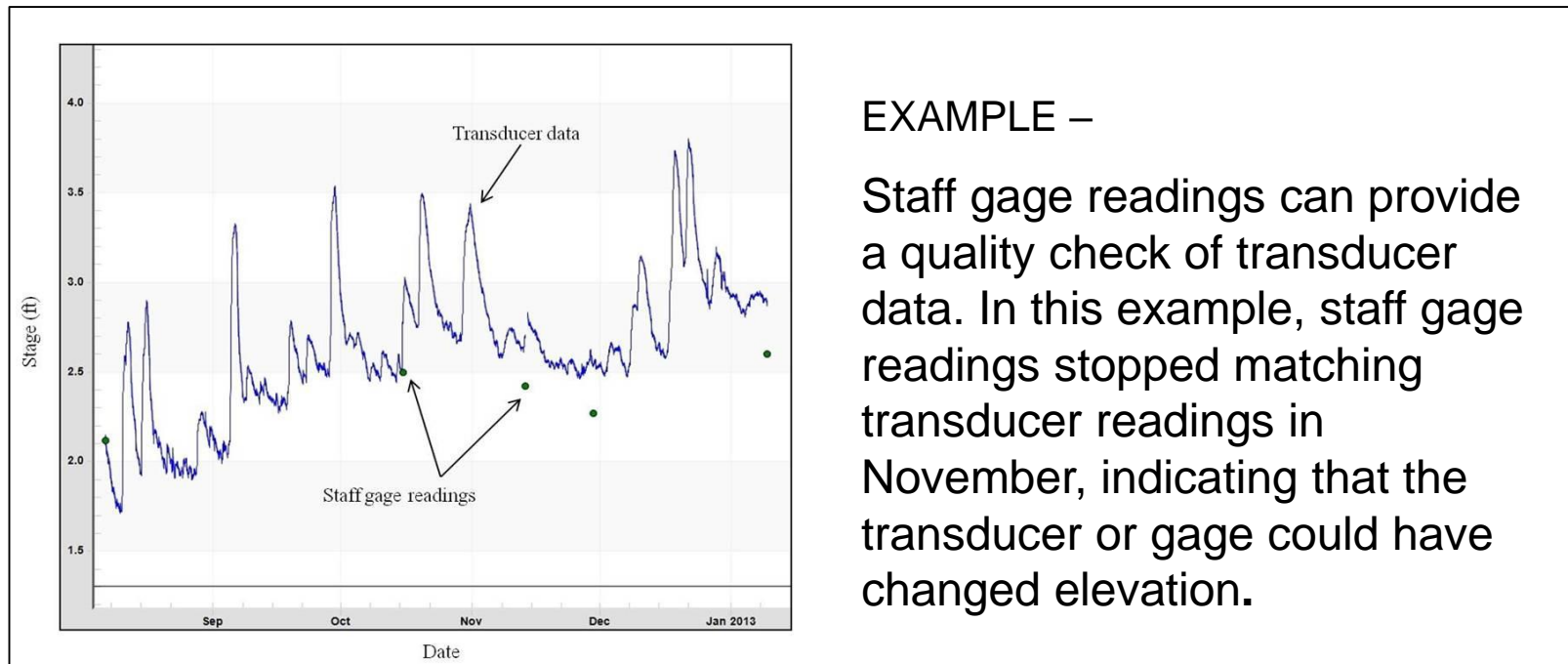
I	J	K
Discrete Water Temp C	Discrete Air Temp C	Discrete Water Level ft
10.055	15.24	0.98

Discrete measurements

Discrete measurements are instantaneous measurements taken during field visits that are compared to the closest continuous sensor measurement. They are used for:

- Accuracy checks
- Making corrections to your data (as needed)
- Validating your data (making them more defensible)

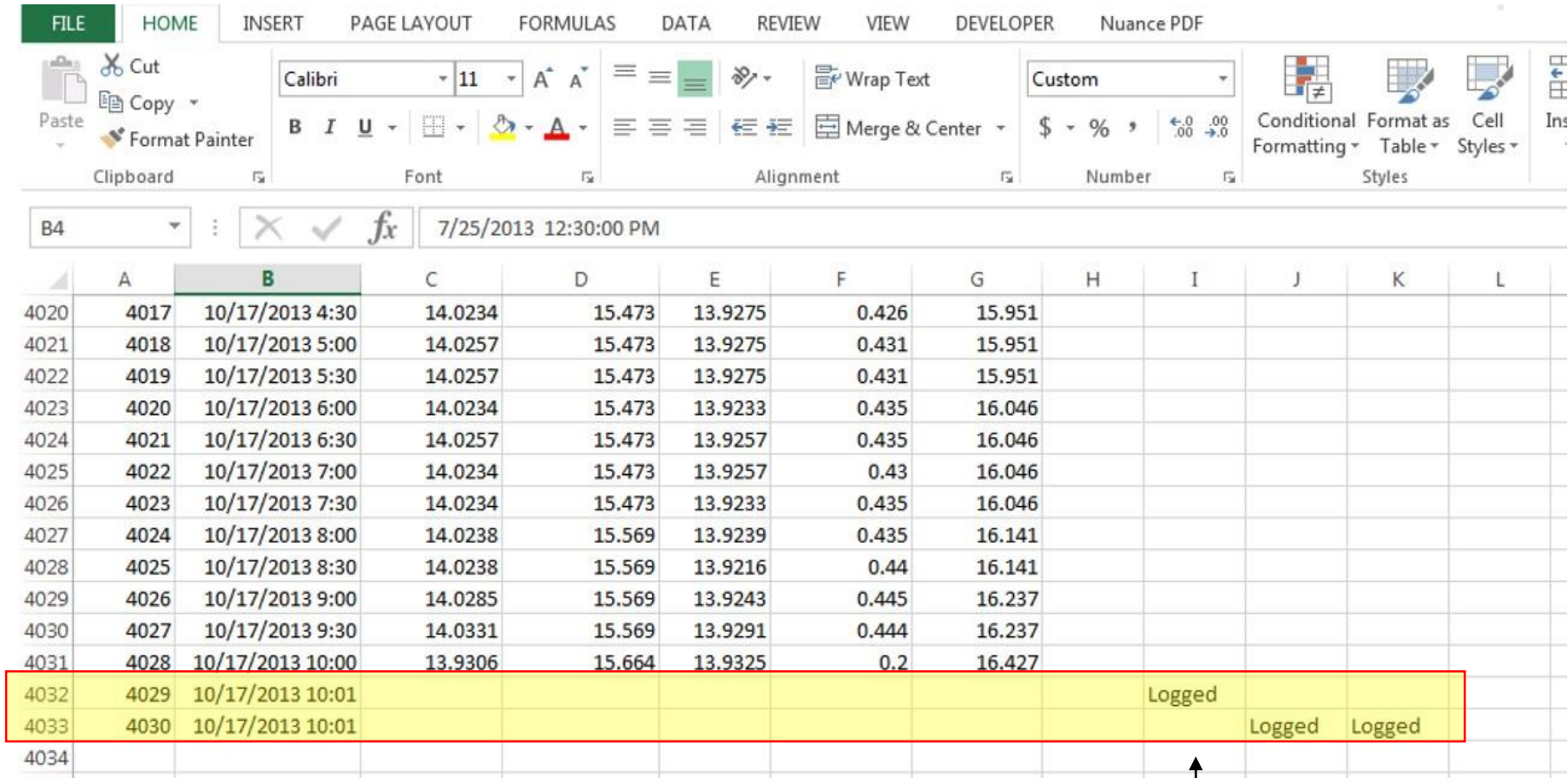
MAKE SURE YOUR DISCRETE AND MATCHING CONTINUOUS MEASUREMENT ARE BOTH IN EITHER STANDARD TIME OR DAYLIGHT SAVINGS TIME.



EXAMPLE –

Staff gage readings can provide a quality check of transducer data. In this example, staff gage readings stopped matching transducer readings in November, indicating that the transducer or gage could have changed elevation.

Check for 'logged' entries in last rows



	A	B	C	D	E	F	G	H	I	J	K	L
4020	4017	10/17/2013 4:30	14.0234	15.473	13.9275	0.426	15.951					
4021	4018	10/17/2013 5:00	14.0257	15.473	13.9275	0.431	15.951					
4022	4019	10/17/2013 5:30	14.0257	15.473	13.9275	0.431	15.951					
4023	4020	10/17/2013 6:00	14.0234	15.473	13.9233	0.435	16.046					
4024	4021	10/17/2013 6:30	14.0257	15.473	13.9257	0.435	16.046					
4025	4022	10/17/2013 7:00	14.0234	15.473	13.9257	0.43	16.046					
4026	4023	10/17/2013 7:30	14.0234	15.473	13.9233	0.435	16.046					
4027	4024	10/17/2013 8:00	14.0238	15.569	13.9239	0.435	16.141					
4028	4025	10/17/2013 8:30	14.0238	15.569	13.9216	0.44	16.141					
4029	4026	10/17/2013 9:00	14.0285	15.569	13.9243	0.445	16.237					
4030	4027	10/17/2013 9:30	14.0331	15.569	13.9291	0.444	16.237					
4031	4028	10/17/2013 10:00	13.9306	15.664	13.9325	0.2	16.427					
4032	4029	10/17/2013 10:01							Logged			
4033	4030	10/17/2013 10:01								Logged	Logged	
4034												

Remove entries like these...







Important!

- **The website is case sensitive.** For example, the website will not recognize 'sensor depth ft'; it needs to be 'Sensor Depth ft'.
- **Values must be formatted correctly**, in addition to the field names. Date formatting, in particular, is finicky. A data template is provided on the website to help with that.
- **24H time (military) is preferred.**
- **Time saving tip:** if you're using HOBOWare, it will reduce your manual reformatting time if you use the **Default settings** (see appropriate ppts in the 'DataProcessing_PreQC' folder on the FTP & Sharepoint sites). You could also develop a script of your own to automate this reformatting step.
- **Use consistent SiteIDs.** Pick one for a site and stick with it; or, if you change it, update it in all your files.
- Make sure the **start and end dates** in your file name **are correct.**

File naming & formatting: review

File naming scheme: SiteID_SensorType_StartDate_EndDate.csv

Example:

Name	Type	Size
 B0997_A_20150305_20150630.csv	Microsoft Excel Comma...	277 KB
 B0997_AW_20150305_20151231.csv	Microsoft Excel Comma...	1,209 KB
 B0997_W_20150305_20150630.csv	Microsoft Excel Comma...	263 KB
 CB192_AW_20150304_20151231.csv	Microsoft Excel Comma...	799 KB
 CB192_AW_20160101_20161231.csv	Microsoft Excel Comma...	750 KB
 ECO66G12_AW_20130426_20130725.csv	Microsoft Excel Comma...	327 KB

File formatting (field names need to be consistent; order does not)

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

	A	B	C	D	E	F	G
1	SiteID	Date Time	Water P psi	Water Temp C	Air BP psi	Sensor Depth ft	Air Temp C
2	ECO66G12	2013-04-18 07:00	14.671	14.613	14.2005	1.086	14.421
3	ECO66G12	2013-04-18 07:30	14.671	14.613	14.1998	1.088	14.709
4	ECO66G12	2013-04-18 08:00	14.671	14.613	14.2017	1.084	15.473
5	ECO66G12	2013-04-18 08:30	14.6733	14.613	14.2046	1.082	16.427

Uploading data

Upload files to website

- Click the “Browse” button on the second tab.
- Select files you want to process

Choose files

Browse...

No file selected

Choose operation to perform

[Instructions](#)

Summary tables

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values

File name	Station ID	Starting date	Ending date	Record count
Awaiting data	Awaiting data	Awaiting data	Awaiting data	Awaiting data

File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Flow
Awaiting data	Awaiting data	Awaiting data	Awaiting data	Awaiting data	Awaiting data	Awaiting data

Upload files to website

- In the two tables to the right, make sure that every field gets filled that you expect to be filled in
 - If a field is “Not found” but you believe the field is in the spreadsheet, check that the field name is formatted correctly.
 - Every spreadsheet should fully populate the top table. The bottom table will be populated depending on the fields in the input files.

[Site introduction](#) [QC tool interface](#) [Download USGS gage data](#) [Advanced features](#) [FAQ](#)

Choose files

Browse...

3 files

Upload complete

Choose operation to perform

Instructions

Summary tables

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values

File name	Station ID	Starting date	Ending date	Record count
B0997_A_20150305_20150415.csv	B0997	2015-03-05	2015-04-15	2016
ECO66G12_AW_20130621_20130725.csv	ECO66G12	2013-06-21	2013-07-25	1653
QB283_W_20150303_20151231.csv	QB283	2015-03-03	2015-12-31	14502

File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Gage height
B0997_A_20150305_20150415.csv	Not found	Found	Not found	Found	Not found	Not found
ECO66G12_AW_20130621_20130725.csv	Found	Found	Found	Found	Found	Not found
QB283_W_20150303_20151231.csv	Found	Not found	Found	Not found	Found	Not found

Running QC/Aggregate/Summarize processes

Choose an operation

- Select an operation from the drop-down menu
- NOTE: You can only aggregate and summarize output from the QC process. You do not need to aggregate spreadsheets to use the summarize process.

[Site introduction](#) [QC tool interface](#) [Download USGS gage data](#) [Advanced features](#) [FAQ](#)

Choose files

Browse...

3 files

Upload complete

Choose operation to perform

QC raw data

Aggregate QC'd data

Summary statistics

[Instructions](#) [Summary tables](#)

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values

File name	Station ID	Starting date	Ending date	Record count
B0997_A_20150305_20150415.csv	B0997	2015-03-05	2015-04-15	2016
ECO66G12_AW_20130621_20130725.csv	ECO66G12	2013-06-21	2013-07-25	1653
QB283_W_20150303_20151231.csv	QB283	2015-03-03	2015-12-31	14502

File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Gage height
B0997_A_20150305_20150415.csv	Not found	Found	Not found	Found	Not found	Not found
ECO66G12_AW_20130621_20130725.csv	Found	Found	Found	Found	Found	Not found
QB283_W_20150303_20151231.csv	Found	Not found	Found	Not found	Found	Not found

Run process

- Click the “Run process” button to run the selected process on the uploaded files
- A box in the bottom right of the screen should appear showing progress in processing uploaded files
- The progress bar will stay in place until a file is completed; then it will jump forward

Site introduction [QC tool interface](#) Download USGS gage data Advanced features FAQ

Choose files

Browse... 3 files

Upload complete

Choose operation to perform

QC raw data

Run operation

[Instructions](#) [Summary tables](#)

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values

File name	Station ID	Starting date	Ending date	Record count
B0997_A_20150305_20150415.csv	B0997	2015-03-05	2015-04-15	2016
ECO66G12_AW_20130621_20130725.csv	ECO66G12	2013-06-21	2013-07-25	1653
QB283_W_20150303_20151231.csv	QB283	2015-03-03	2015-12-31	14502

File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Gage height
B0997_A_20150305_20150415.csv	Not found	Found	Not found	Found	Not found	Not found
ECO66G12_AW_20130621_20130725.csv	Found	Found	Found	Found	Found	Not found
QB283_W_20150303_20151231.csv	Found	Not found	Found	Not found	Found	Not found

Running QCRaw Operating on B0997_A_20150305_20150415.csv

Running QCRaw Operating on
ECO66G12_AW_20130621_20130725.csv

Running QCRaw Operating on
QB283_W_20150303_20151231.csv

Running QCRaw Finished
QB283_W_20150303_20151231.csv

Download files

- After all processing is complete, a “Download” button will appear.
- Depending on your browser, you will be given the chance to open and/or save a zip file with all output files
- Transfer the output files to the appropriate folder on your computer

[Site introduction](#) [QC tool interface](#) [Download USGS gage data](#) [Advanced features](#) [FAQ](#)

Choose files

Browse...

3 files

Upload complete

Choose operation to perform

QC raw data

Run operation

Download

Instructions

Summary tables

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values

File name	Station ID	Starting date	Ending date	Record count
B0997_A_20150305_20150415.csv	B0997	2015-03-05	2015-04-15	2016
ECO66G12_AW_20130621_20130725.csv	ECO66G12	2013-06-21	2013-07-25	1653
QB283_W_20150303_20151231.csv	QB283	2015-03-03	2015-12-31	14502

File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Gage height
B0997_A_20150305_20150415.csv	Not found	Found	Not found	Found	Not found	Not found
ECO66G12_AW_20130621_20130725.csv	Found	Found	Found	Found	Found	Not found
QB283_W_20150303_20151231.csv	Found	Not found	Found	Not found	Found	Not found

QC process output

2 documents

QC report (Word or HTML document)

.csv file



Data Quality Control Report

Report Date: 2016-02-05

DATA FILE INFORMATION

Filename: QC_test1_Aw_20130725_20131015.csv

SiteID: test1

Period of Record, Requested: 2013-07-25 to 2013-10-15

Period of Record, Actual: 2013-07-25 to 2013-10-15

Recording Interval: 30 minutes

Data Type: Aw

Parameters Included: Water.Temp.C, Air.Temp.C, Water.BP.psi, Air.BP.psi, Water.Level.ft,

DATA SUMMARY, OVERALL

Month and Day versus number of records

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
8	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48
9	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48
10	48	48	48	48	48	48	48	48	48	48	48	48	49	48	21
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	31
7	0	0	0	0	0	0	0	0	0	0	27	48	48	48	48

QC_test2_Aw_20130426_20130725.csv - Excel

SiteID	Date	Time	Date.Time	Water.Temp.C	Air.Temp.C	Water.BP.psi	Air.BP.psi	Water.Level.ft	Water.LoggerID	Water.RowID	Flag.Date.Time	Flag.Water.Temp.C	Flag.Air.Temp.C	Flag.Water.BP.p
Test2	4/26/2013	10:30:00	4/26/2013 10:30	10.063	15.282	14.8024	14.3638	0.985	NA	1		P	P	P
Test2	4/26/2013	11:00:00	4/26/2013 11:00	10.357	14.134	14.7461	14.3644	0.854	NA	2		P	P	P
Test2	4/26/2013	11:30:00	4/26/2013 11:30	10.748	14.23	14.7428	14.3603	0.855	NA	3		P	P	P
Test2	4/26/2013	12:00:00	4/26/2013 12:00	11.139	15.473	14.7441	14.3627	0.853	NA	4		P	P	P
Test2	4/26/2013	12:30:00	4/26/2013 12:30	11.528	17.284	14.7431	14.3612	0.854	NA	5		P	P	P
Test2	4/26/2013	13:00:00	4/26/2013 13:00	11.916	18.806	14.7421	14.3628	0.848	NA	6		P	P	P
Test2	4/26/2013	13:30:00	4/26/2013 13:30	12.207	19.662	14.7453	14.3629	0.855	NA	7		P	P	P
Test2	4/26/2013	14:00:00	4/26/2013 14:00	12.594	20.424	14.7351	14.3537	0.853	NA	8		P	P	P
Test2	4/26/2013	14:30:00	4/26/2013 14:30	12.883	20.71	14.7315	14.3484	0.857	NA	9		P	P	P
Test2	4/26/2013	15:00:00	4/26/2013 15:00	13.173	20.901	14.7234	14.3426	0.852	NA	10		P	P	P
Test2	4/26/2013	15:30:00	4/26/2013 15:30	13.365	20.805	14.7171	14.3331	0.859	NA	11		P	P	P
Test2	4/26/2013	16:00:00	4/26/2013 16:00	13.558	21.091	14.7109	14.3302	0.852	NA	12		P	P	P
Test2	4/26/2013	16:30:00	4/26/2013 16:30	13.654	21.569	14.702	14.3215	0.851	NA	13		P	P	P
Test2	4/26/2013	17:00:00	4/26/2013 17:00	13.75	21.282	14.6978	14.3152	0.856	NA	14		P	P	P
Test2	4/26/2013	17:30:00	4/26/2013 17:30	13.846	21.473	14.6913	14.3164	0.838	NA	15		P	P	P
Test2	4/26/2013	18:00:00	4/26/2013 18:00	13.846	20.805	14.7051	14.3239	0.853	NA	16		P	P	P
Test2	4/26/2013	18:30:00	4/26/2013 18:30	13.846	19.758	14.7119	14.3294	0.856	NA	17		P	P	P
Test2	4/26/2013	19:00:00	4/26/2013 19:00	13.75	18.426	14.7368	14.3538	0.857	NA	18		P	P	P
Test2	4/26/2013	19:30:00	4/26/2013 19:30	13.75	17.665	14.7413	14.3564	0.861	NA	19		P	P	P
Test2	4/26/2013	20:00:00	4/26/2013 20:00	13.75	16.999	14.7392	14.3551	0.859	NA	20		P	P	P
Test2	4/26/2013	20:30:00	4/26/2013 20:30	13.654	15.951	14.7113	14.3289	0.856	NA	21		P	P	P
Test2	4/26/2013	21:00:00	4/26/2013 21:00	13.558	15.187	14.6812	14.2999	0.853	NA	22		P	P	P
Test2	4/26/2013	21:30:00	4/26/2013 21:30	13.558	14.325	14.6994	14.3178	0.854	NA	23		P	P	P
Test2	4/26/2013	22:00:00	4/26/2013 22:00	13.461	13.75	14.7174	14.3351	0.855	NA	24		P	P	P
Test2	4/26/2013	22:30:00	4/26/2013 22:30	13.365	13.173	14.7241	14.341	0.857	NA	25		P	P	P
Test2	4/26/2013	23:00:00	4/26/2013 23:00	13.269	12.594	14.7421	14.3606	0.853	NA	26		P	P	P
Test2	4/26/2013	23:30:00	4/26/2013 23:30	13.173	12.11	14.7416	14.3558	0.863	NA	27		P	P	P
Test2	4/27/2013	0:00:00	4/27/2013 0:00	13.173	11.819	14.7325	14.3452	0.867	NA	28		P	P	P
Test2	4/27/2013	0:30:00	4/27/2013 0:30	13.076	11.722	14.739	14.3515	0.867	NA	29		P	P	P
Test2	4/27/2013	1:00:00	4/27/2013 1:00	12.98	11.819	14.7457	14.3587	0.866	NA	30		P	P	P
Test2	4/27/2013	1:30:00	4/27/2013 1:30	12.98	11.819	14.7572	14.3656	0.877	NA	31		P	P	P
Test2	4/27/2013	2:00:00	4/27/2013 2:00	12.883	11.819	14.7637	14.37	0.881	NA	32		P	P	P
Test2	4/27/2013	2:30:00	4/27/2013 2:30	12.787	11.819	14.7541	14.3519	0.901	NA	33		S	P	P
Test2	4/27/2013	3:00:00	4/27/2013 3:00	12.787	11.722	14.7588	14.3447	0.928	NA	34		P	P	P
Test2	4/27/2013	3:30:00	4/27/2013 3:30	12.69	11.819	14.7447	14.3271	0.936	NA	35		P	P	P
Test2	4/27/2013	4:00:00	4/27/2013 4:00	12.69	11.722	14.7377	14.3175	0.943	NA	36		P	P	P
Test2	4/27/2013	4:30:00	4/27/2013 4:30	12.594	11.625	14.7718	14.3374	0.975	NA	37		P	P	P
Test2	4/27/2013	5:00:00	4/27/2013 5:00	12.594	11.625	14.7856	14.3351	1.012	NA	38		P	P	P

QC tests

- **Unrealistic values** ('Gross range')
 - Entries are flagged if values are above or below upper and lower limits
- **Spikes**
 - Entries are flagged if adjacent points change by more than 'x' amount
- **Rate of change** (RoC)
 - Entries are flagged if the RoC exceeds a given threshold
(e.g., ≥ 3 st dev within 25 hrs)
- **Flat line**
 - Entries are flagged if a certain # of consecutive measurements are within a certain amount of each other (e.g., >10 consecutive temperature measurements are within 0.01 degrees C of one another)

Flags

Flags are assigned to each data point (a single measured parameter at a unique point in time)

- **P** = Pass,
- **S** = Suspect,
- **F** = Fail,
- **X** = No Data or Not Applicable (NA).

Fail - extreme, unrealistic

- Example - water temp values flagged 'F' if ≥ 30 or ≤ -2

Suspect - highly unlikely, but not as extreme as above

- Example - water temp values flagged 'S' if ≥ 25 or ≤ -1

Thresholds

Table 1. Updated thresholds (7/31/2017) based on evaluation of continuous air temperature, water temperature, air pressure, water pressure and sensor depth data that have been collected at freshwater, high quality wadeable RMN stream sites in the eastern US.

+ Most of the data are being recorded at 30-minute intervals by Onset HOBO U20 Water Level Data Loggers.

Test	Water Temp		Air Temp		Sensor depth		Water Pressure		Air (Barometric) Pressure	
	Fail	Suspect	Fail	Suspect	Fail	Suspect	Fail	Suspect	Fail	Suspect
Gross	>30°C or <-2°C	>25°C or <-0.1°C	>38°C or <-25°C	>35°C or <-23°C	>6 <u>ft</u> or <-1 <u>ft</u>	>6 <u>ft</u> or <-1 <u>ft</u>	>17 psi or <13 psi	>16.8 psi or <13.5 psi	>15 psi or <13 psi	>14.8 psi or <13.5 psi
Spike	≥1.5°C (±)	≥1.0°C (±)	≥10°C (±)	≥8°C (±)	≥5 <u>ft</u> (±)	≥3 <u>ft</u> (±)	≥0.7 psi (±)	≥0.5 psi (±)	≥0.25 psi (±)	≥0.15 psi (±)
Rate of Change	NA	≥3 <u>stdev</u> within 25 <u>hrs</u>	NA	≥3 <u>stdev</u> within 25 <u>hrs</u>	NA	≥3 <u>stdev</u> within 25 <u>hrs</u>	NA	≥3 <u>stdev</u> within 25 <u>hrs</u>	NA	≥3 <u>stdev</u> within 25 <u>hrs</u>
Flat Line	>30 consecutive measurements within 0.01 units of one another	>20 consecutive measurements within 0.01 units of one another	>15 consecutive measurements within 0.01 units of one another	>10 consecutive measurements within 0.01 units of one another	>60 consecutive measurements within 0 units of one another	>20 consecutive measurements within 0 units of one another	>15 consecutive measurements within 0.001 units of one another	>10 consecutive measurements within 0.001 units of one another	>15 consecutive measurements within 0.001 units of one another	>10 consecutive measurements within 0.001 units of one another

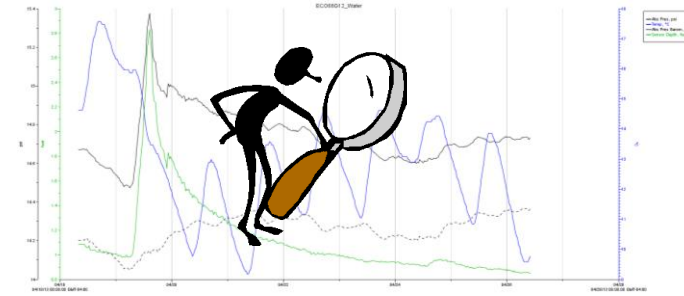
- See 'QC_ThresholdMemo_20170804' Word doc for explanation of updates
 - 'Suspect' flag for water temperature was being triggered too much (winter)
 - Spike test thresholds (water temperature) not stringent enough
- These thresholds cannot currently be changed in the website. A future version of the website will allow users to change QC thresholds.

QC report (HTML)

4 main components

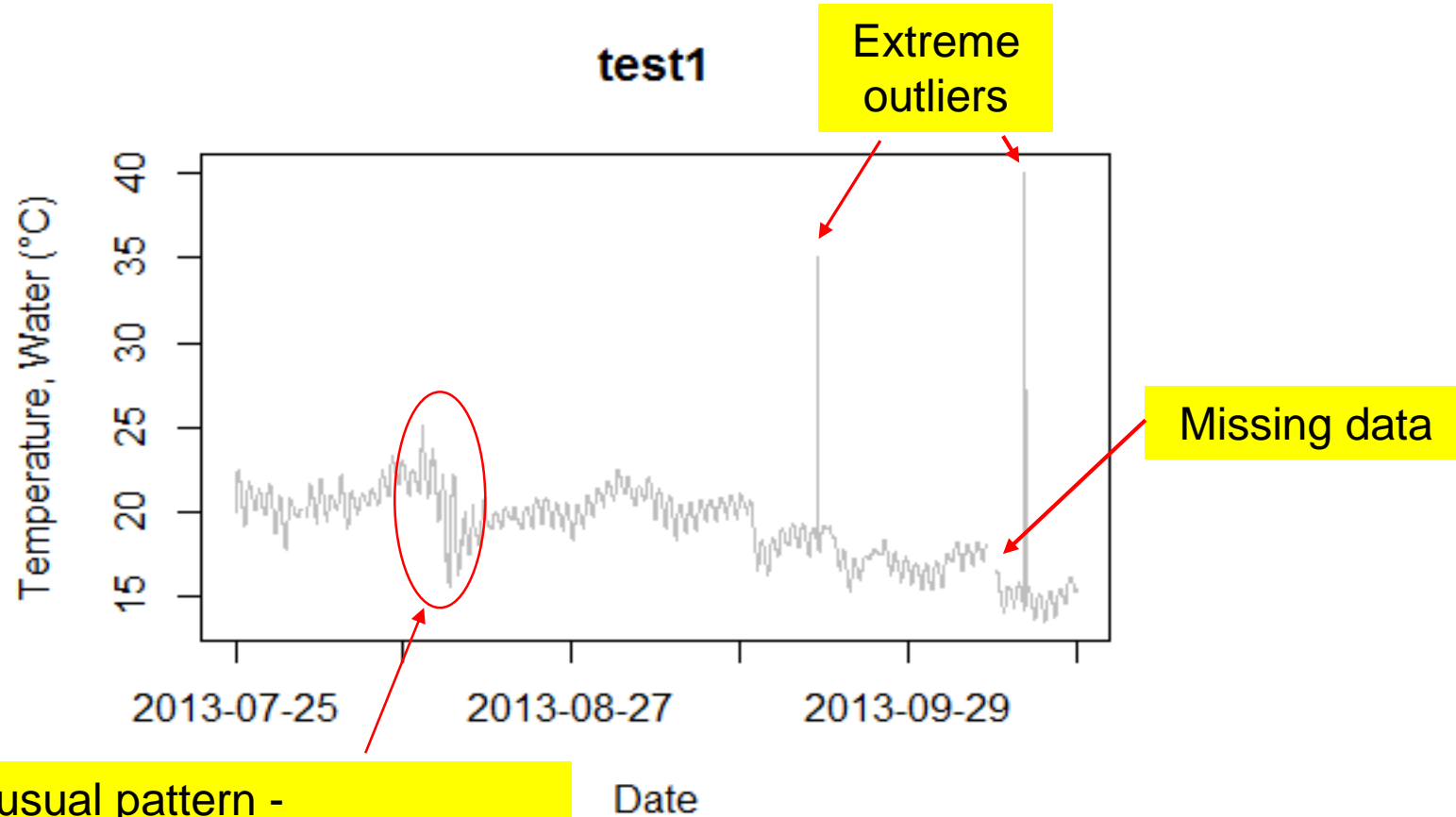
1. **Data file information** (siteID, date range, parameters, recording interval)
2. **'Count' tables (# measurements/day)** so that you can find inconsistencies (too few measurements (=missing data) or too many measurements)
3. **Results from QC tests** - # of entries marked as pass (P), suspect (S), fail (F), missing data (X) or not available (NA)
4. **Time series plots** – for each individual parameter & several combined parameters (water & air temperature; water temperature and water level)

Visual checks on time series plots



- **Missing data**
- **Closely matching water and air temperature** values (this indicates that the stream sensor may have been out of the water)
- **Diel fluxes with flat tops, or “chatter”** in the gage height record (this indicates that the sensor may have been buried in sediment)
- **Water level values of 0** (this could mean that the pressure transducer was dewatered. With vented transducers, another possibility is that moisture got into the cable and caused readings of zero water depth)
- **Water level values that are negative**
- **Outliers or rapidly fluctuating values** (it is possible that the sensor moved, e.g., due to a high flow event or vandalism)
- Values recorded before the sensor was correctly positioned & stabilized, or after the sensor was removed from its location (**“trimming”**)

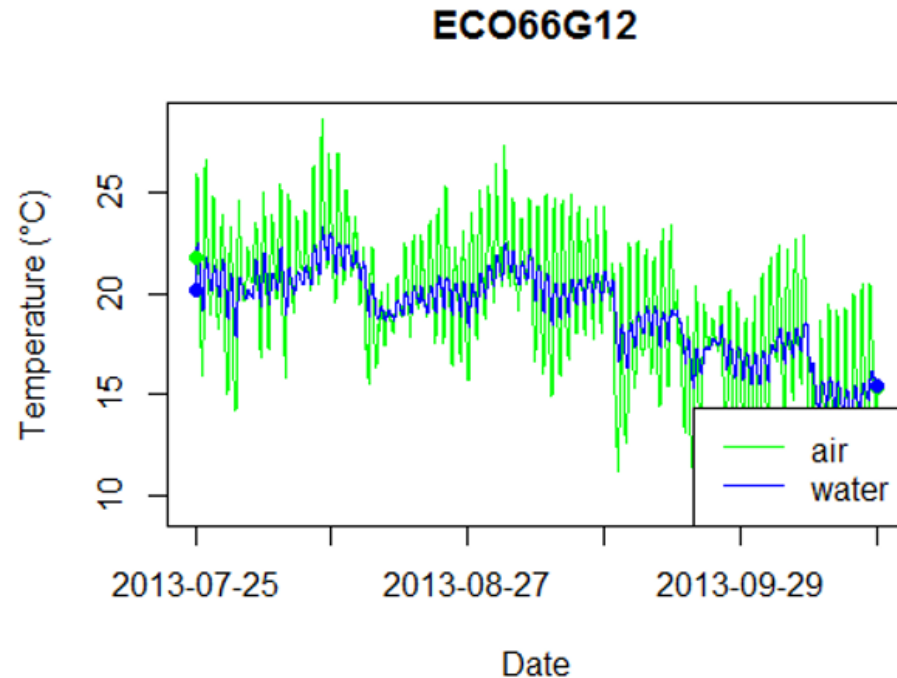
Examples of visual checks on time series plots



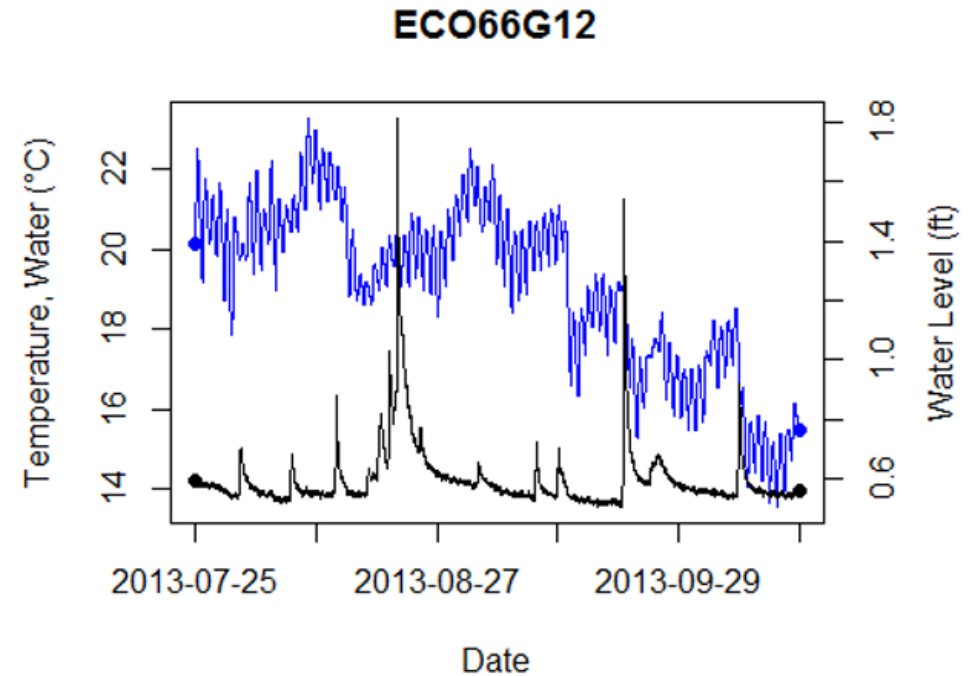
QC report – Section 4

Multi-parameter time series plots

PLOT, Temperature (Air vs. Water)



PLOT, Water Level vs. Water Temperature



Do the patterns make sense? (e.g., water temperature should fluctuate less than air temperature; increases in water level may correspond with decreases in water temperature)

.csv file

Interpretation

Values (water temp, air temp, sensor depth, etc.)

QC_test2_Aw_20130426_20130725.csv - Excel

FILE

HOME

INSERT

PAGE LAYOUT

FORMULAS

DATA

REVIEW

VIEW

DEVELOPER

Nuance PDF

Cut

Copy

Paste

Format Painter

Clipboard

Calibri

11

A

A

B

I

U

Font

Wrap Text

Merge & Center

Alignment

General

\$

%

,

.

00

.

00

Number

Conditional Formatting

Format as Table

Styles

Normal

Bad

Good

Check Cell

Explanatory...

Input

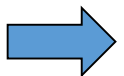
G12

fx

0.859

	A	B	C	D	E	F	G	H
1	SiteID	Date.Time	Water.Temp.C	Air.Temp.C	Water.P.psi	Air.BP.psi	Sensor.Depth.ft	Water.LoggerID
2	Test2	2013-04-26 10:30	10.063	15.282	14.8024	14.3638	0.985	NA
3	Test2	2013-04-26 11:00	10.357	14.134	14.7461	14.3644	0.854	NA
4	Test2	2013-04-26 11:30	10.748	14.23	14.7428	14.3603	0.855	NA

Scrolling to the right...



.csv file continued...

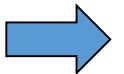
Overall flags (entries marked as pass, suspect, fail or missing data, based on the *worst* flag across all of the QC tests).

QC_test2_Aw_20130426_20130725.csv - Excel

	I	J	K	L	M	N	O
	Water.RowID	Flag.Date.Time	Flag.Water.Temp.C	Flag.Air.Temp.C	Flag.Water.P.psi	Flag.Air.BP.psi	Flag.Sensor.Depth.ft
1	1		P	P	P	P	P
2	2		P	P	P	P	P
3	3		P	P	P	P	P
4	4		P	P	P	P	P

F = fail
P = pass
S = suspect
X = Not Data or Not Applicable (NA)

Continue scrolling to the right



.csv file continued...

Flags for each parameter and each QC test (gross, spike, RoC, flat)

QC_ECO66G12_Aw_20130426_20130725 - Excel

ULAS

DATA

REVIEW

VIEW

Nuance PDF

A

≡

≡

≡

✂

▼

Wrap Text

General

▼

≡

≡

≡

↔

↔

Merge & Center

\$

▼

%

,

←.0

.00

→.0

≡

≡

≡

Conditional Formatting

≡

≡

≡

Format as Table

Normal

Bad

Check Cell

Explanatory ...

Alignment

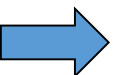
Number

Sty

X	Y	Z	AA
Flag.Gross.Water.Temp.C	Flag.Spike.Water.Temp.C	Flag.RoC.Water.Temp.C	Flag.Flat.Water.Temp.C
P	X	X	P
P	P	P	P
P	P	P	P
P	P	P	P
P	P	P	P

Example - water temperature

Continue scrolling to the right



.csv file continued...

RAW = original values; if you end up making corrections in the primary value fields, the RAW fields allow you to retain the original records and track any changes you make.

COMMENT.MOD = Explain any changes you make, including why, who and when

QC_ECO66G12_A

FILEHOMEINSERTPAGE LAYOUTFORMULASDATAREVIEWVIEWNuance PDF

Cut

Copy

Paste

Format Painter

Clipboard

Calibri

11

A

A

B

I

U

A

Wrap Text

Merge & Center

Alignment

General

\$

%

,

←.0

→.00

Number

Conditional Formatting

Format as Table

AW13

21.091

	AS	AT	AU	AV	AW
1	RAW.Date.Time	Comment.MOD.Water.Temp.C	RAW.Water.Temp.C	Comment.MOD.Air.Temp.C	RAW.Air.Temp.C
2	2013-04-26 10:30		10.063		15.282
3	2013-04-26 11:00		10.357		14.134
4	2013-04-26 11:30		10.748		14.23
5	2013-04-26 12:00		11.139		15.473
6	2013-04-26 12:30		11.528		17.284

Tips for interpreting QC output

Start by skimming through the QC Word document.

- Are there problems? If so, what kind, and what parameters do they affect?
- Then open the .csv file.
- Reformat the file so that it is easier to navigate (unfortunately we can't get .csv files to retain this formatting each time you open them)
 - **Format – Autofit Column Width**
 - **Alignment - center the column entries**
 - **Freeze top row and left two columns**
 - **Turn on Filter**
- Filter for the flagged values. Evaluate whether changes should be made.
- Document changes (if you don't already have a system in place for this, consider using the example field/data processing form on the FTP & Sharepoint sites).
- Records of any changes will also be visible in the QC .csv file, in the Comments field(s)

What the QC process doesn't do...

Does **NOT** -

- **Automatically 'fix' the data.** You still have to go through the process of interpreting the QC reports and -
 - Checking flagged data
 - Doing visual checks on the time series plots
 - Deciding if and how to make corrections
 - Document your corrections
- **Catch everything.** There may be instances where the tests miss data points that should be flagged (we would love to get your feedback when this happens! we are still learning ourselves & want to keep improving the scripts).

Making corrections

Right now we **don't have any automated corrections programed into the QC website.**

Any corrections you make need to be done **manually.**

Some types of corrections are fairly straightforward and can be done in Excel (e.g., change values by same amount, or basic drift corrections); others are more complex.

NOTE: The summarize process will not use any Failed records. If you want Failed records to be included in data summaries, delete the Fail flags.

Potential references –

- Wagner et al. 2006 (USGS) (more detailed)
- EPA 'Best Practices' report (more general - has a table with a general summary of different types of problems that can occur (e.g., missing data, failed accuracy check) and recommended actions for addressing them)

If you want to make a change in your data...

Explain changes in these

Keep these the same

Change these values

	A	B	C	D	E	G	AH	AI	AL	AM
1	SiteID	Date	Time	Date.Time	Water.Temp.C	Air.BP.psi	Comment.MOD.Water.Temp.C	RAW.Water.Temp.C	Comment.MOD.Air.BP.psi	RAW.Air.BP.psi
2	ECO66G12	4/26/2013	10:30:00	4/26/2013 10:30	10.063	14.3638		10.063		14.3638
3	ECO66G12	4/26/2013	11:00:00	4/26/2013 11:00	10.357	14.3644		10.357		14.3644
4	ECO66G12	4/26/2013	11:30:00	4/26/2013 11:30	10.748	14.3603		10.748		14.3603
5	ECO66G12	4/26/2013	12:00:00	4/26/2013 12:00	11.139	14.3627		11.139		14.3627
6	ECO66G12	4/26/2013	12:30:00	4/26/2013 12:30	11.528	14.3612		11.528		14.3612
7	ECO66G12	4/26/2013	13:00:00	4/26/2013 13:00	11.916	14.3628		11.916		14.3628
8	ECO66G12	4/26/2013	13:30:00	4/26/2013 13:30	12.207	14.3629		12.207		14.3629
9	ECO66G12	4/26/2013	14:00:00	4/26/2013 14:00	12.594	14.3537		12.594		14.3537
10	ECO66G12	4/26/2013	14:30:00	4/26/2013 14:30	12.883	14.3484		12.883		14.3484
11	ECO66G12	4/26/2013	15:00:00	4/26/2013 15:00	13.173	14.3426		13.173		14.3426
12	ECO66G12	4/26/2013	15:30:00	4/26/2013 15:30	13.365	14.3331		13.365		14.3331
13	ECO66G12	4/26/2013	16:00:00	4/26/2013 16:00	13.558	14.3302		13.558		14.3302
14	ECO66G12	4/26/2013	16:30:00	4/26/2013 16:30	13.654	14.3215		13.654		14.3215
15	ECO66G12	4/26/2013	17:00:00	4/26/2013 17:00	13.75	14.3152		13.75		14.3152

Aggregate data

- Combine multiple files from the same site and parameters that cover different time periods
- OR
- Combine multiple files from the same site and time periods with different parameters
- Outputs csv and html/Word report, just like the QC process

[Site introduction](#) [QC tool interface](#) [Download USGS gage data](#) [Advanced features](#) [FAQ](#)

Choose files

Browse...

2 files

Upload complete

Choose operation to perform

Aggregate QC'ed data

Run operation

Download

Instructions

Summary tables

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values

File name	Station ID	Starting date	Ending date	Record count
QC_QB283_A_20150303_20151231.csv	QB283	2015-03-03	2015-12-31	14502
QC_QB283_W_20150303_20151231.csv	QB283	2015-03-03	2015-12-31	14502

File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Gage height
QC_QB283_A_20150303_20151231.csv	Not found	Found	Not found	Found	Not found	Not found
QC_QB283_W_20150303_20151231.csv	Found	Not found	Found	Not found	Found	Not found

Aggregate data

- If you try to combine files from multiple sites, the website will show a warning
- If you try to aggregate files that have not gone through the QC process, the website will show a warning
- All files output from the aggregate process have “DATA_” prepended to the name

Continuous data QC, summary, and statistics- PRELIMINARY WEBSITE

Site introductionInstructions & interfaceDownload USGS dataAdvanced featuresFAQ

Choose files

Browse... 2 files

Upload complete

Choose operation to perform

Aggregate QC'ed data

Run operation

InstructionsSummary tables

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values

File name	Station ID	Starting date	Ending date	Record count
CB192_AW_20151125_20151231.csv	CB192	2015-11-25	2015-12-31	1776
ECO66G12_AW_20130621_20130725.csv	ECO66G12	2013-06-21	2013-07-25	1653

File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Flow
CB192_AW_20151125_20151231.csv	Found	Not found	Not found	Not found	Found	Found
ECO66G12_AW_20130621_20130725.csv	Found	Found	Found	Found	Found	Not found

WARNING: The spreadsheets you have selected for the aggregate process have not gone through the QC process. Please put these spreadsheets through the QC process first.

WARNING: The spreadsheets you have selected for the aggregate process include more than one site. Please select spreadsheets that include only one site.

Aggregate data

- To combine files from a single site with different parameters and different time periods, first combine files from the same time period by parameter or files of the same parameter by time period. Then combine the results of that aggregation to get a file aggregated by time period and parameter.

Example:

Step 1:

QC_QB283_A_20150303_20151231.csv + QC_QB283_A_20160101_20161231.csv =
DATA_QC_QB283_A_20150303_20161231.csv

Step 2:

QC_QB283_W_20150303_20151231.csv + QC_QB283_W_20160101_20161231.csv =
DATA_QC_QB283_W_20150303_20161231.csv

Step 3:

DATA_QC_QB283_A_20150303_20161231.csv + DATA_QC_QB283_W_20150303_20161231.csv =
DATA_DATA_QC_QB283_AW_20150303_20161231.csv

Summarize data

- Produces csvs and pdfs with summary statistics for the input parameters
- If you try to summarize files that have not gone through the QC process, the website will show a warning

Choose files

Browse... 2 files

Upload complete

Choose operation to perform

Summary statistics

Run operation

Instructions

Summary tables

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values

File name	Station ID	Starting date	Ending date	Record count
DATA_QC_ECO66G12_AW_20130621_20131015.csv	ECO66G12	2013-06-21	2013-10-15	5589
DATA_QC_QB283_A_20150303_20161231.csv	QB283	2015-03-03	2016-12-31	32070

File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Flow
DATA_QC_ECO66G12_AW_20130621_20131015.csv	Found	Found	Found	Found	Found	Not found
DATA_QC_QB283_A_20150303_20161231.csv	Not found	Found	Not found	Found	Not found	Not found

Choose files

Browse... 2 files

Upload complete

Choose operation to perform

Summary statistics

Run operation

Instructions

Summary tables

Summary tables of input files

After uploading data, confirm that the table below is showing the expected values


File name	Station ID	Starting date	Ending date	Record count
B0997_W_20150305_20150415.csv	B0997	2015-03-05	2015-04-15	2016
CB192_AW_20160101_20160212.csv	CB192	2016-01-01	2016-02-12	2064

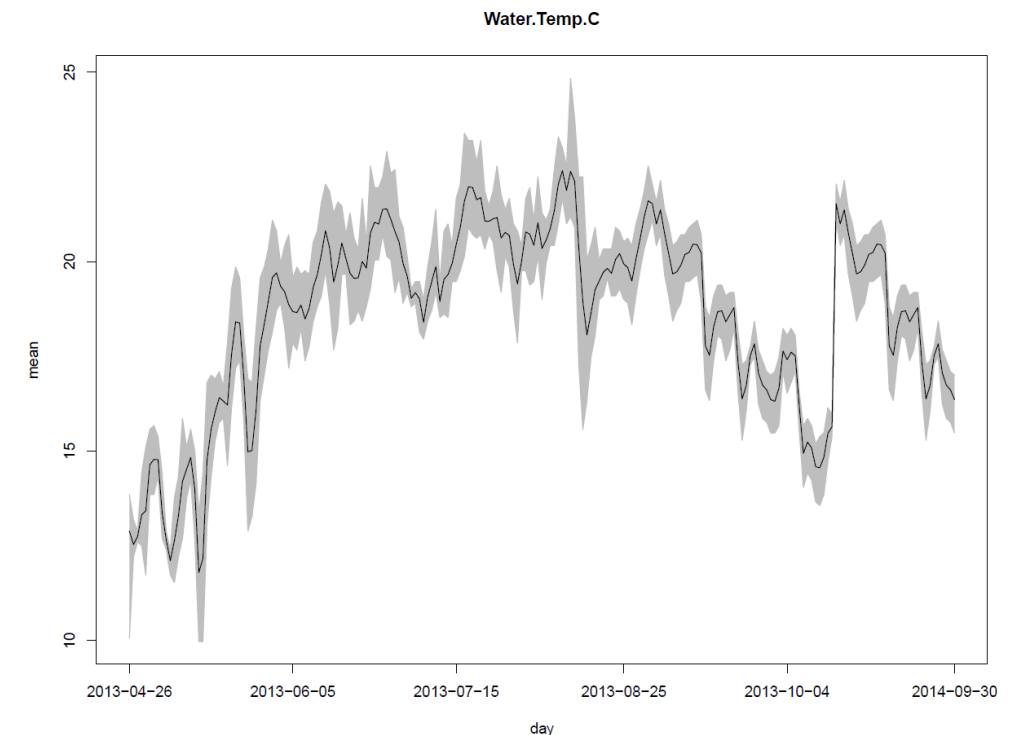
File name	Water temperature	Air temperature	Water pressure	Air pressure	Sensor depth	Flow
B0997_W_20150305_20150415.csv	Found	Not found	Found	Not found	Found	Not found
CB192_AW_20160101_20160212.csv	Found	Found	Not found	Not found	Found	Not found

WARNING: The spreadsheets you have selected for the summarize process have not gone through the QC process. Please put these spreadsheets through the QC process first.

Summarize data

- Produces 2 csvs and 1 pdf for each parameter in each input file
 - csv starting with “DV” has daily values
 - csv starting with “STATS” has mean, median, max, range, variation, percentile and more statistics by year, season, month, day, and more
 - pdf starting with “STATS” has plots of the parameter by day, month, season, and year

Name	Type	Size
 DV_DATA_QC_ECO66G12_AW_20130426_20130915_Air.Temp.C.csv	Microsoft Excel Comma...	13 KB
 STATS_DATA_QC_ECO66G12_AW_20130426_20130915_Air.Temp.C.csv	Microsoft Excel Comma...	77 KB
 STATS_DATA_QC_ECO66G12_AW_20130426_20130915_Air.Temp.C.pdf	Adobe Acrobat Docume...	14 KB
 STATS_DATA_QC_ECO66G12_AW_20130426_20130915_Water.Temp.C.csv	Microsoft Excel Comma...	77 KB
 STATS_DATA_QC_ECO66G12_AW_20130426_20130915_Water.Temp.C.pdf	Adobe Acrobat Docume...	14 KB
 DV_DATA_QC_ECO66G12_AW_20130426_20130915_Water.Temp.C.csv	Microsoft Excel Comma...	13 KB



Getting USGS gage data

Retrieving USGS gage data

- Downloads USGS gage data at the sites specified over the specified time period
- Multiple files must be separated by commas
- At this time, all input sites will have the same date range downloaded
- Each station will get its own output spreadsheet

Instructions	Download USGS gage data here
<p>You can download data from USGS gages on this tab.</p> <ol style="list-style-type: none">1. Enter as many USGS station IDs as you like separated by commas and spaces (e.g., 01187300, 01493000, 01639500).2. Enter a starting and ending date for which data will be retrieved for each station; the same date range will be used for every station.3. Click the 'Retrieve USGS data' button. A progress bar will appear in the bottom-right of the tab. It will advance as each file is completed. Thus, if you select three stations, it will wait at 0%, jump to 33%, jump to 66%, and then jump to 100%.4. After data retrieval completes, a download button will appear. Click the button to download a zip file of all station records. Where the files will download on your computer depends on the configuration of your internet browser.	<p>USGS site ID(s) (separated by commas and spaces)</p> <input type="text" value="01187300, 01493000, 01639500"/> <p>Starting date (YYYY-MM-DD)</p> <input type="text" value="2016-01-25"/> <p>Ending date (YYYY-MM-DD)</p> <input type="text" value="2016-08-17"/> <p>Retrieve USGS data</p>

Retrieving USGS gage data

- Click the “Retrieve USGS data” button to run the selected process on the uploaded files
- A box in the bottom right of the screen should appear showing progress in processing uploaded files
- The progress bar will stay in place until a station is completed; then it will jump forward

Instructions

You can download data from USGS gages on this tab.

1. Enter as many USGS station IDs as you like separated by commas and spaces (e.g., 01187300, 01493000, 01639500).
2. Enter a starting and ending date for which data will be retrieved for each station; the same date range will be used for every station.
3. Click the 'Retrieve USGS data' button. A progress bar will appear in the bottom-right of the tab. It will advance as each file is completed. Thus, if you select three stations, it will wait at 0%, jump to 33%, jump to 66%, and then jump to 100%.
4. After data retrieval completes, a download button will appear. Click the button to download a zip file of all station records. Where the files will download on your computer depends on the configuration of your internet browser.

Download USGS gage data here

USGS site ID(s) (separated by commas and spaces)

Starting date (YYYY-MM-DD)

Ending date (YYYY-MM-DD)

Retrieve USGS data

Getting USGS data Retrieving records for site 01603000

×

Download files

- After all sites' data are retrieved, a “Download USGS gage data” button will appear.
- Depending on your browser, you will be given the chance to open and/or save a zip file with all output files
- Transfer the output files to the appropriate folder on your computer

Instructions	Download USGS gage data here
<p>You can download data from USGS gages on this tab.</p> <ol style="list-style-type: none">1. Enter as many USGS station IDs as you like separated by commas and spaces (e.g., 01187300, 01493000, 01639500).2. Enter a starting and ending date for which data will be retrieved for each station; the same date range will be used for every station.3. Click the 'Retrieve USGS data' button. A progress bar will appear in the bottom-right of the tab. It will advance as each file is completed. Thus, if you select three stations, it will wait at 0%, jump to 33%, jump to 66%, and then jump to 100%.4. After data retrieval completes, a download button will appear. Click the button to download a zip file of all station records. Where the files will download on your computer depends on the configuration of your internet browser.	<p>USGS site ID(s) (separated by commas and spaces)</p> <input type="text" value="01582500, 01603000, 01488500"/> <p>Starting date (YYYY-MM-DD)</p> <input type="text" value="2016-01-25"/> <p>Ending date (YYYY-MM-DD)</p> <input type="text" value="2016-08-17"/> <p>Retrieve USGS data</p> <p>Download USGS gage data</p>

Custom thresholds

Setting custom QC thresholds

- On the “Advanced features” tab, you can upload a custom configuration R file. This allows you to change thresholds for QC.
- A configuration file is provided as a template

[Site introduction](#) [QC tool interface](#) [Download USGS gage data](#) [Advanced features](#) [FAQ](#)

R console output

This panel shows messages output by the QC, aggregating, summarizing, and USGS data retrieval processes. If there are any errors when the tool runs, please copy the messages and send them and your input files to the contacts listed on the tool background tab.

Check here for script messages after running process...

Custom QC thresholds

You can upload custom QC thresholds here. Please use [this](#) configuration document as a template.

Once you have made your changes to the configuration file, upload them below.

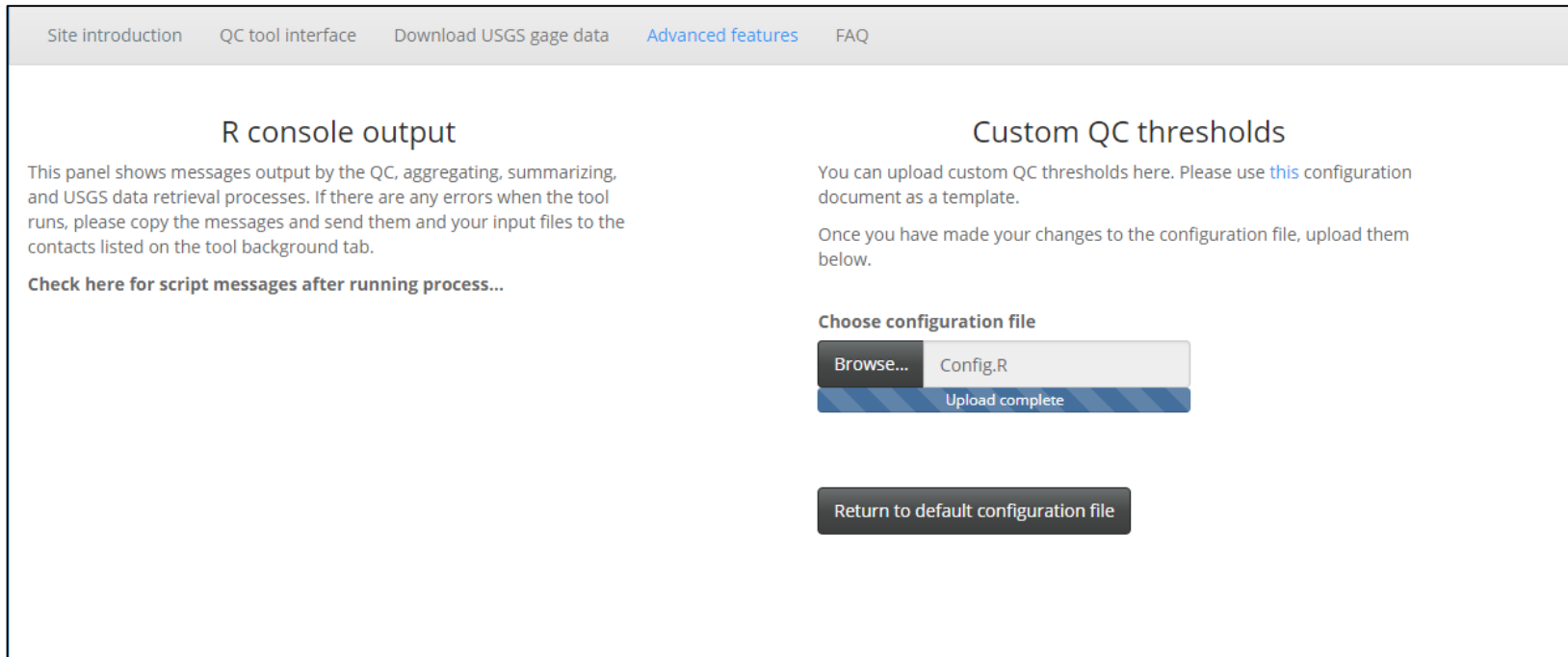
Choose configuration file

Browse...

No file selected

Reverting QC thresholds to default

- After uploading a custom configuration file, a “Return to default configuration file” button will appear.
- Click on the button to apply the default configuration file to your next QC run.



The screenshot shows the 'QC tool interface' with a navigation bar at the top containing links for 'Site introduction', 'QC tool interface', 'Download USGS gage data', 'Advanced features', and 'FAQ'. The main content area is divided into two columns. The left column, titled 'R console output', contains text about message aggregation and a link to check script messages. The right column, titled 'Custom QC thresholds', provides instructions on uploading configuration files. It includes a 'Choose configuration file' section with a 'Browse...' button, a text input field containing 'Config.R', and an 'Upload complete' status bar. Below this, a dark button labeled 'Return to default configuration file' is visible.

Troubleshooting

Use the input file summary tables

- If the summary tables don't reflect what you think is in your input spreadsheets, investigate the discrepancies.
- The summary tables report on which column names are found in the input files. If a column name is not found in an input csv, the summary table reports "Not found". If the column is found, the table reports "Found". It does not check the contents of the columns.
- Use the downloadable data template to make sure your field names are correct. Copy the field names from the template to your files to make sure file names are correct.

Use the progress bar

- If the website is crashing without completing the selected process, watch the progress bar.
- Even though it does not mark progress within a file, it marks progress between files. It can show on which file the website is crashing.
- If you are trying to process more than one file in a single upload, try processing one file at a time.
- You can also try copying your problematic input spreadsheet and deleting rows (dates), e.g., the first half of the records. Make sure to change the file name accordingly.

Use the “R console output” tab

- Every time you run a process on a file, some notes about the run will be output to the “R console output” tab.
- These can be diagnostic of issues with the process you are trying to run. Copy the output shown in this tab and send it to the contacts listed on the website.

[Site introduction](#) [QC tool interface](#) [Download USGS gage data](#) [Advanced features](#) [FAQ](#)

R console output

This panel shows messages output by the QC, aggregating, summarizing, and USGS data retrieval processes. If there are any errors when the tool runs, please copy the messages and send them and your input files to the contacts listed on the tool background tab.

R console messages for process ' QC raw data '

[1] "Total files to process = 1"

[1] "Processing item 1 of 1, WORKING (QC Tests and Flags - WaterTemp), B0997_W_20150305_20150415.csv."

[1] "Processing item 1 of 1, WORKING (QC Tests and Flags - WaterP), B0997_W_20150305_20150415.csv."

[1] "Processing item 1 of 1, WORKING (QC Tests and Flags - SensorDepth), B0997_W_20150305_20150415.csv."

[1] "Task COMPLETE. QC Report. Total time = 1.728939 secs."

[1] "User defined parameters: File Name (B0997_W_20150305_20150415.csv)."

[1] "Processing item 1 of 1, COMPLETE, B0997_W_20150305_20150415.csv."

Custom QC thresholds

You can upload custom QC thresholds here. Please use [this](#) configuration document as a template.

Once you have made your changes to the configuration file, upload them below.

Choose configuration file

Browse...

No file selected

Additional notes and resources

Additional notes

- The website will continue to be improved over time. We welcome your suggestions and comments. You can e-mail us or make comments here: https://github.com/dagibbs22/RShiny_RMN_QC_scripts
- The FAQ tab of the website has some additional information.
- The R scripts that perform the QC/Aggregate/Summarize/GetGageData processes were written by Erik Leppo (Tetra Tech) (<https://github.com/leppott/ContDataQC>). The web interface was written by David Gibbs (US EPA ORISE fellow from February 2016 to April 2018). This work was funded by Britta Bierwagen (US EPA ORD).

dagibbs22@gmail.com
bierwagen.britta@epa.gov
jen.stamp@tetrattech.com
erik.leppo@tetrattech.com

Additional training materials

- We have developed **instructional materials and the ContDataQC R package** to help improve the quality of the continuous RMN data and to make data processing and management easier for those entities that do not currently have a system

Tetra Tech FTP site - RMN_ContinuousData folder

<http://ttftp.hostedftp.com>

User = TTBaltimore

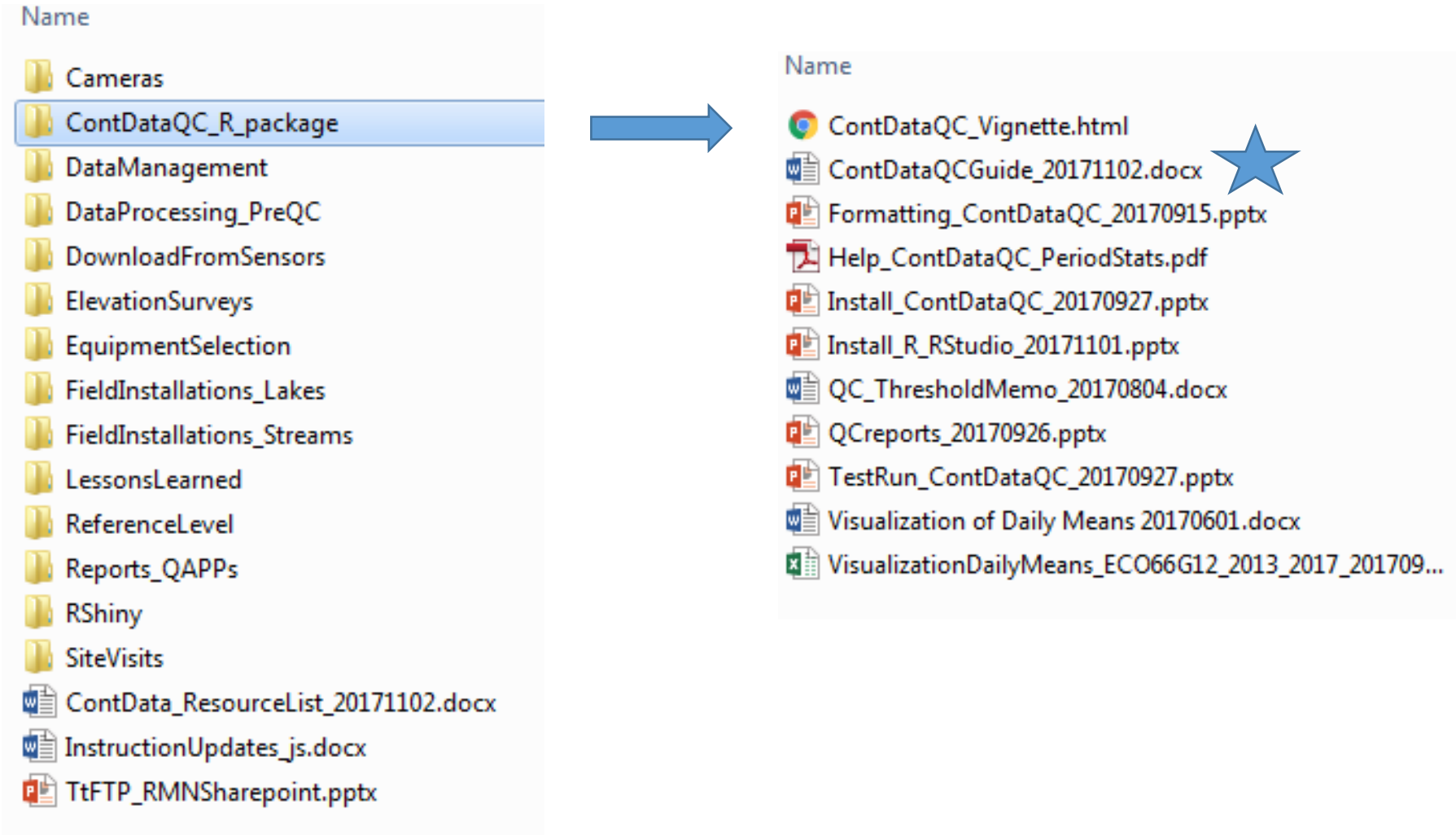
Password = Tetra2500

EPA RMN Sharepoint site – Continuous_data folder

https://usepa.sharepoint.com/sites/ORD_Community/RMN/SitePages/Home.aspx

Additional training materials

Reference the files in the '**ContDataQC_R_package**' folder



ContDataQC Guide (Word)

Good starting point!



All the instructions for running the four basic functions are in one place.

Also available in separate segments, in the PowerPoint slides of the FTP & Sharepoint sites.

For advanced R users, a vignette is posted as well (it's also on the GitHub site)

'ContDataQCGuide_20171102' Word document

Table of Contents – Introduction to [ContDataQC](#) (11/02/2017)

1	Overview	2
2	Instructional materials	3
3	Quick guide.....	4
4	Steps for first-time users	4
4.1	Install R and R Studio.....	5
4.2	Install ContDataQC R	5
4.3	Load the ContDataQC library.....	6
4.4	Format input files	6
4.5	Open the ContDataQC Help file.....	7
4.6	Set up parameters	8
4.7	Set up working directory	8
4.8	Create data directories	9
4.9	Run desired operation(s).....	10
4.9.1	QCRaw	13
4.9.2	Aggregate	13
4.9.3	SummaryStats.....	14
4.9.4	GetGageData	15
5	Advanced users	16
5.1	Changing directories.....	16
5.2	Customizing the configuration file	17