Main functions – 4a Run functions (v1; 6/17/2021)

**4a Run functions**

The three main functions are listed in the drop-down menu under ‘Choose Operation to perform’

* ‘**QC raw data’** generates QC reports.
* **‘Aggregate QC’d data’** merges the QC'd data files for the desired time periods together. The input files need to be from the same site and share common parameters.
* **‘Summary statistics’** generates two sets of CSV files and a PDF with time series plots for each parameter. One of the CSVs has daily means and the other has a wider range of summary statistics (e.g., min, median, max, quartiles).

ContDataQC works sequentially, meaning the user needs to run the QC function before running the Aggregate function, and the Aggregate function before the SummaryStats function. More detailed information on each function can be found below.

This video provides an overview of the workflow: will send link later

Click here to download a PDF with instructions on how to set up Google Chrome so that it asks you where to save each file before downloading (which can be a big time-saver).

**QC reports**

Shiny app instructions

1. Click the ‘Browse’ button and select the input file(s) from the Data1\_RAW folder. You can run multiple files at once. The status bar will tell you when the upload is complete.
2. Select ‘QC Raw Data’ from the ‘Choose Operation to perform’ drop-down menu
3. A ‘Run operation’ button will appear. Click this button. A status bar will appear in the lower right corner.
4. After the function is done running, a ‘Download’ button will appear. Click this button.
5. Save the zipped file in the Data2\_QC folder.
6. Open the QC reports. There are two types of outputs: a Data Quality Control Report that comes as either a Word document or Hypertext Markup Language (HTML) (the user specifies the desired format); and a comma-delimited (.csv) spreadsheet. The reports include results from the ‘flag’ tests (gross, spike, rate of change and flat line), missing data checks and time series plots.
7. Review each set of files, check flagged data points, do visual checks on the time series plots, and decide if and how to make corrections.

The Data Quality Control Report is organized into four main sections, starting with an overall summary followed by summaries for each parameter:

1. Data file information (siteID, date range, parameters, recording interval, QC test thresholds reference table)
2. ‘Count’ tables (# measurements/day) so that you can find inconsistencies (too few measurements (=missing data) or too many measurements)
3. Results from the four ‘flag’ tests (gross, spike, rate of change and flat line) - # 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 and two sets of combined parameters (water and air temperature; water temperature and water level). It can also plot discrete/*in situ* values taken during site visits for accuracy checks (the discrete data points are overlaid on top of the sensor data). These are in situ measurements taken when people do field work and are used for calibration/validation of these continuous sensors

The CSV file contains the time series data from the sensor plus additional data columns with flags for each test and parameter (e.g., Flag.Spike.Water.Temp.C, Flag.Flat.Water.Temp.C), comment fields for each parameter so that the user can document rationale for any changes (e.g., Comment.MOD.Water.Temp.C) and a duplicate set of the original ‘raw’ data values (e.g., RAW.Water.Temp.C). The ‘raw’ values are included in case the user decides to make corrections in the first set of data columns. It is important to maintain documentation of changes that are made when the user goes from “raw” to “quality controlled” versions. It also makes it easier for the user to go back at a later time and change how they handled a correction.

Click here to download an example output from the QC function for a HOBO temperature sensor file. It has a Word version of the Data Quality Control Report and the CSV output.

This video provides tips on interpreting QC reports and deciding whether or not to make corrections: will send link later

Currently there are no automated corrections programed into the QC function. The user must make the changes manually.

**Aggregate**

Shiny app instructions

1. Click the ‘Browse’ button and select the input files from the Data2\_QC folder that you’d like to combine together. The status bar will tell you when the upload is complete.
2. Select ‘Aggregate QC’d data’ from the ‘Choose Operation to perform’ drop-down menu
3. A ‘Run operation’ button will appear. Click this button. A status bar will appear in the lower right corner.
4. After the function is done running, a ‘Download’ button will appear. Click this button.
5. Save the zipped file in the Data3\_Aggregated folder.
6. Open and review the aggregated reports. As with the QC output, there are two types of outputs: a Data Quality Control Report that comes as either a Word document or Hypertext Markup Language (HTML) (the user specifies the desired format); and a comma-delimited (.csv) spreadsheet.

Make sure the CSV file contains data for the desired time period before running it through the Summary statistics function.

Click here to download an example output from the aggregate function in which three years of temperature data (from multiple individual data files) were combined into one file.

**Summary statistics**

Shiny app instructions

1. Click the ‘Browse’ button and select the desired input file from the Data3\_Aggregated folder. The status bar will tell you when the upload is complete.
2. Select ‘Summary statistics’ from the ‘Choose Operation to perform’ drop-down menu
3. A ‘Run operation’ button will appear. Click this button. A status bar will appear in the lower right corner.
4. After the function is done running, a ‘Download’ button will appear. Click this button.
5. Save the zipped file in the Data4\_Stats folder.
6. Open and review the summary statistics reports. There should be two sets of CSV files and a PDF with time series plots for each parameter. One of the CSVs has daily means and the other has a wider range of summary statistics (e.g., min, median, max, quartiles).

Click here to download an example output from the ‘summary statistics’ function in which three years of air and water temperature data were summarized.

**Using the R package**

If you prefer using the R package vs. the Shiny app, we recommend reviewing the material on the GitHub site (<https://github.com/leppott/ContDataQC>). The code for running the main functions can be copied and pasted from the Help file. After the user directs R to the correct directory, there are four places in the code where the user needs to make entries before running the QC function (see figure below).

Insert png ‘Rpackage\_v1\_20210617’