Approach for digging out from a big pileup of data (version 1; 8/24/2022)

Contact - Jen.Stamp@tetratech.com

Step 1 - Compile and prepare data files to run through the R tools

- Recommendation - focus on one site to start, the one that you think has the best, most complete dataset

Step 2 – Compile data into one file (and don't make any corrections yet)

- Use R tools to run your CSV files through the reformat, QC and aggregate functions. All your data will be compiled into one file.
- Evaluate QC test thresholds and fill out QC threshold check worksheet for documentation
- Customize the QC test thresholds in the configuration file if needed
- Put files from the first QC run into a folder labeled as 'old' or 'orig' (or delete them?) so that you don't accidentally use them in Step 3

Step 3 – QC the aggregated data file

- Move a copy of the aggregated CSV file into the Data1_Raw folder. Delete all the flag and comments columns from the previous run (that used the default settings). Remove the 'Data_QC_' prefix from the file name.
- Generate QC reports using the customized configuration file
- Open the Word QC report and check the 'Thresholds, Quick Reference' section to make sure the updated QC thresholds were used (vs. the defaults).
- Visually check the time series plots in the Word doc for obvious errors; keep a running list of issues/approximate timeframes to check (in case they are not picked up by the QC flag tests)
- Open the CSV QC report. Rearrange the columns so that the overall flag field for a given parameter and Comment.MOD. field are next to the column you are QCing (e.g., move Flag.Water.Temp.C and Comment.MOD.Water.Temp.C next to the Water.Temp.C column). You may want to move the columns for each individual flag test over as well.
- Update the Logger.Deployment field. Unfortunately, the start/end entries from the original input files get erased when you rerun the QC report. This is a bummer because those transition times are when errors tend to occur. To repopulate this field, under the RowID field(s), you can do a Find '1' and it should bring you to the 'start' of a new deployment period (and the previous cell would be the 'end'). If you try this, be sure to check 'match exact cell content', otherwise the Find will take you to anything with a '1' in it.
- Check each F and S flag. Verify that X flags are correct as well (=missing data). Retain or overturn flags (e.g., change 'F' to 'P'). If you are certain the data are erroneous, delete the measurement and flag as 'F'. If you are not sure, flag the data as 'S' and do not delete. Let the person using the data decide whether or not to remove questionable measurements. Document that you checked each point by adding the data qualifier to the Comment column. Download the draft Data Qualifier file. Try to use the existing qualifiers but it is ok to add new qualifiers if needed.
- Complete the accuracy check worksheet. Correct for sensor drift if needed.
- Save your QC'd CSV file regularly!

Step 4 (optional) – compare to nearby data (Daymet, weather station, USGS gage)

- If time permits, as an additional QC check, compare to nearby data. David Smith from KY Division of Water put together some instructional files for this.

Step 5 – create streamlined data file to run through summary and visualization tools and/or to upload to a data management system

- Copy and paste the columns with the QC'd data into a new file (either Excel or csv, or both). Bring in Water Temp, Air Temp, Water Level or Sensor Depth, Discharge, DO (as available); do not bring in water and air pressure. Also bring in the overall (updated) flag column for each parameter, along with the Comments field for each parameter.
- (optional) Run the streamlined file through the QC function to confirm that you didn't miss any clearly erroneous measurements
- Retain the following files for full documentation of your QC process
    - CSV data files with all the columns (including the RAW data columns, in case you want to go back and change how you handled a correction)
    - QC Word report
    - Accuracy check worksheet
    - Customized configuration file

Once you get through this, things will be easier moving ahead! You'll be familiar with the data from each site and will have a system for working through it, so you can hopefully keep up and avoid data pileups moving ahead.