Straw Man Thoughts on Conversion of Excel Splitting Records to R Script

Updated 9/13/13

**Objectives:**

1. Develop a script that will pull relevant data from NWIS Web and generate a data plot (or just .csv file to be used in for Excel), and “splitting record summary” for sites that require a flow-composite water-quality sample for runoff events.
2. For sites that require discrete samples, generate a data plot and sample summary.

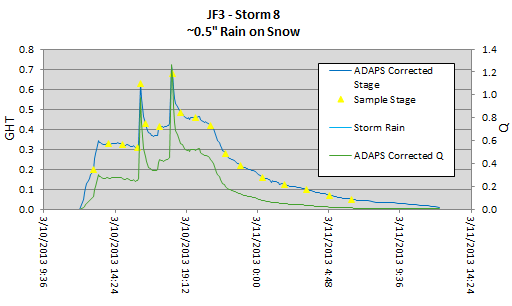
**Inputs:**

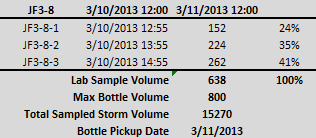
1. Station ID, Gage Height (00065), Discharge (00060), Sample Code (99909), and Precipitation (00045 - possibly from nearby station if not available?)
2. Storm Start and Storm End
3. Data Start and Data End
4. Sample Prefix
5. Initial Sample Number
6. Number of storms to split
7. Storm Number(s)
8. Max Bottle Volume
9. Max Lab Sample Volume

**Intermediate Variable Changes:** these are items that may need to be modified by the user after a preliminary run-through of the script.

Sample Dates to ignore or add back in, Max Bottle Volume, Initial Sample Number, Storm Start and Storm End, possibly a lower limit (>900, for example) for valid Sample Codes, the ability to add text (bottle pickup dates, storm descriptions, for example) to output.

**Outputs:** open for discussion, something like below perhaps.





**Variables to include:**

Storm Start and Storm End - if used to determine which data to pull from NWIS, pull data from ~3 hrs prior to Storm Start and ~3 hrs after Storm End to “bracket” the storm. Alternatively, have separate variables such as Data Start and Data End to accomplish that task.

Sample Code – typically, if this value is greater than 1000, a sample has been taken. This isn’t always the case, however, and we’ll eventually be able to have a workaround for exceptions.

Sample Date – one for each sample, typically identifiable if Sample Code > 1000.

Sample Prefix – identifier used by field crew labeling the bottles.

Sample Number – identifier used by field crew labeling the bottles. This will sometimes be tricky to determine. Some people use the actual Sample Code, but typically there is a consecutive (but sometimes resetting to zero) labeling for each site. This is probably a necessary input variable (first sample number, for example).

Sample Label – merge of Sample Prefix and Sample Number for each sample.

Storm Number(s) – This is probably a necessary input variable.

Flow Volume – computed for each data point as:

(0.5\*((TimeCur - TimePrev)\*86400))\*(0.75\*QCur + 0.25\*QPrev) + (0.5\*((TimeNext - TimeCur)\*86400)\*(0.75\*QCur + 0.25\*QNext))

Note: the 86000 constant is not necessary if Time difference is computed in seconds.

Total Sampled Storm Volume – sum of the Flow Volumes from Storm Start to Storm End. This is computed only once.

Max Bottle Volume – this is the maximum volume that can be used from any sample in the set. Typically, we default this to about 900 ml or so. Ideally, we could optimize the program so that this is set properly to give us our Sample Volume to Lab of 3.9 L? This could be tricky because getting 3.9L is not possible in a number of situations.

Sample Flow Volume – sum of the Flow Volumes from the mid-point time of previous sample to the mid-point time of the next sample. For the first and last sample of the set, flow is summed from the respective endpoint (Storm Start or Storm End) to the next (or prior) relevant midpoint. One Sample Flow Volume is generated for each sample in the composite.

Max Sample Flow Volume – this is the maximum value of Sample Flow Volume encountered for all samples in the composite.

Sample Percent – Sample Flow Volume divided by Total Sampled Storm Volume for each sample. This is computed for each sample in the composite.

Sample Aliquot – (Sample Flow Volume \* Max Bottle Volume)/Max Sample Flow Volume computed for each sample. This is how much water should be taken from each bottle to make the composite sample.

Lab Sample Volume – this is the sum of the Sample Aliquots of the samples within the composite. Currently, this value cannot exceed 4 Liters (probably make closer to 3.9L). We’ll need to do a check to make sure it is not exceeded or optimize.

Max Lab Sample Volume – typically 3900 ml or 3.9 liters.

**Code Flow:**

User inputs necessary values – consider:

1. Station ID, Gage Height (00065), Discharge (00060), Sample Code (99909), and Precipitation (00045 - possibly from nearby station if not available?)
2. Storm Start and Storm End
3. Data Start and Data End
4. Sample Prefix
5. Initial Sample Number
6. Number of storms to split
7. Storm Number(s)
8. Max Bottle Volume
9. Max Lab Sample Volume

Note: may want to consider initially hard-coding some variables for ease of use? Max Bottle Volume = 900, Max Lab Sample Volume = 3900

Pull data from NWIS web, format properly if necessary

Compute Flow Volume for each time stamp between Data Start and Data End

Compute (Total Sampled Storm Volume) by sum of Flow Volume between Storm Start and Storm End

Determine Sample Date and Sample Label for each Sample Code > 1000 for all data points between Storm Start and Storm End

Compute variables required for each Sample Date (Sample Flow Volume, Sample Percent)

Compute Max Sample Flow Volume by finding the maximum Sample Flow Volume among all Sample Dates.

Compute Sample Aliquot for each Sample Date

Compute Lab Sample Volume by summing all Sample Aliquots

Double-checks:

* Lab Sample Volume must be less than or equal to Max Sample Flow Volume. If not, error notice to user to decrease Max Bottle Volume or program optimizes.
* Sum of Sample Percent should equal 100. If not, some sort of coding issue. This will be good for initial troubleshooting
* Sum of Sample Flow Volumes should equal Total Sampled Storm Volume. If not, some sort of coding issue.

Preliminary output of plot and summary stats. Something similar to that shown above.

User option to modify items or to finalize data.

Re-run necessary functions identified in previous step

Double-checks – same as above

Secondary Output (if necessary)

Iterate as necessary to finish