Flow-Weighting and Compositing Calculator – Data Product QA Review

Team site –
https://sccwrp.sharepoint.com/:f:/s/SMCBMPProjects/EuC4Mz2OaflMq9CvMHMFIxkBe-0cwK9jR2I8FaOqsonJ-g?e=h6GK2q

Application url – https://sccwrp.shinyapps.io/FWCCalculator/

Github repo (*private) - https://github.com/SCCWRP/FWCCalculator

Version - v0.2

- Version v0.1 is a Python-based FWCCalculator found at https://github.com/SCCWRP/EMC_calculator_from_depth

Accuracy – How did we go about verifying the accuracy of the application? Who tested it and when did they test it? *Were there any outside agencies who tested the application? Which domain expert signed off on the accuracy of the application (Eddie)? If there were bugs, what process did you take to fix and document them?*

The Flow-Weighting and Compositing Calculator (FWCC) was compared with excel-calculations on data from a wetland outlet in Auckland, NZ. The Auckland Wetland Outlet data consisted of time-stamped flow rates and sample concentrations taken during a storm event in 2008. Using this data set, validation tests were created and the data set altered in order to test each element of the file validation. These include checks on items such as the file type, the number of columns and sheets in the input Excel workbook, missing values, negative values, date formatting, and other measurement criteria.

The points of comparison were flow volumes attributed to each sample as calculated by the FWCC and an excel-based method. The flow volumes become the sample compositing instructions to determine the Event Mean Concentration (EMC). Assumptions for the calculation included a left-hand Riemann sum approximation of the flowrate integration, as well as a central-attribution scheme that considered flow to be “attributed” to a sample at all times for which that sample was the closest in time.

Auckland Wetland Outlet - Hydrographs

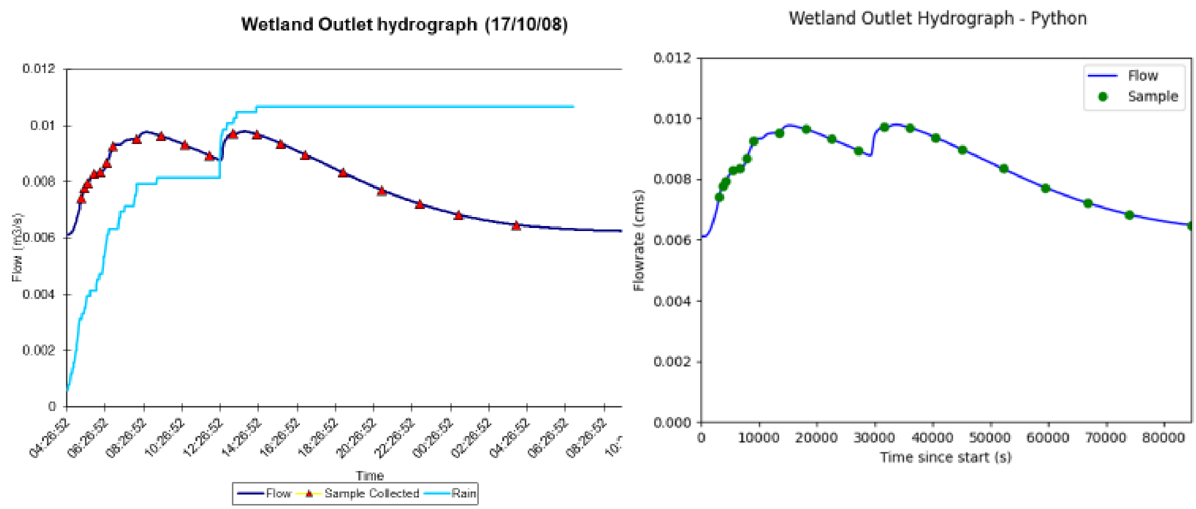
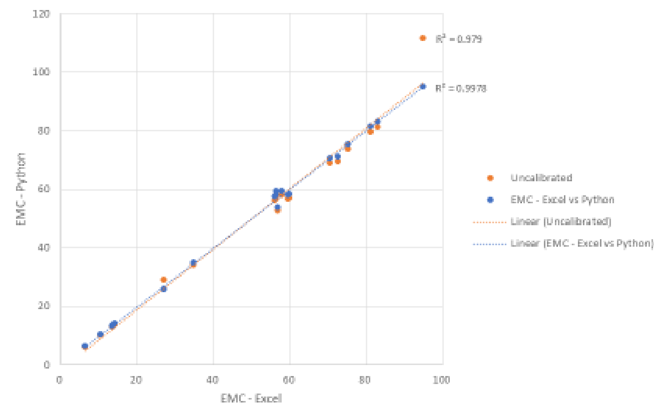


Figure 1. Accuracy data from an Auckland Wetland Outlet processed by python-version 1.0

Auckland Wetland Outlet – EMC instructions

	Bottle	Vol. (ml)		
	1	27	25.8	29.1
	2	6	6.3	6.2
	3	10	10.4	10.2
	4	13	13.4	13.1
	5	14	13.6	13.3
	6	14	14.1	13.8
	7	35	34.8	34.1
	8	58	59.4	58.1
	9	60	58.1	56.7
	10	56	57.6	56.3
	11	57	53.9	52.7
	12	56	59.4	58.1
	13	60	58.4	57
	14	56	58	56.6
	15	73	71.1	69.5
	16	81	81.4	79.5
	17	75	75.4	73.7
	18	70	70.6	69
	19	83	83.1	81.2
	20	95	95.1	111.6
Sum	20	1000	999.9	999.8
		R^2	0.998	0.97628



**Excel EMC calculated using a central Riemann sum approximation
**Python EMC calculated using a trapezoidal approximation

Figure 2. Flow-weighted sample volumes for the compositing instructions are listed in side by side comparison. Nash-Sutcliffe Efficiency (NSE) index is used to assess agreement between an excel-based method for flow-weighting and the python version of the FWCC. Note: when linearly comparing multiple models, the NSE index is functionally equivalent to the R^2 coefficient of determination.

The accuracy exercise showed excellent agreement between a calibrated FWCC and the excel-based flow-weighting method; $R^2 > 0.998$. The uncalibrated FWCC was included to illustrate the sensitivity of the FWCC to user decisions regarding processing raw data. A technical report investigating the effect of different integration methods, flow rate data resolutions, and flow-sample attribution schemes on the flow-weighted compositing instructions is being prepared.

This comparison was conducted on January 2nd by domain expert Edward Tiernan, Engineer in the Engineering Department at SCCWRP. At this time, the FWCC tool has not been reviewed by any outside agencies or beta testers.

Bugs were rooted out through extensive use of data visualization. Differential shading helped identify bugs regarding the flow rates that are attributed to a given sample. Bugs, when found, were reviewed with the domain expert and corrected by the technician. In addition to this input file validation, unit tests were created in order to test the individual functions within the application including functions for calculating the aliquot volume table, the cleaning of the input data, the volume calculation methods, and the possible linear interpolation of flow measurements. To create and run these tests, the ‘testthat’ R package was used during development. The tests were continuously run during development, and the latest run as of the creation of this document is 1/31/2023.

Stability (robustness) - What is the intended use of the application? How was the application tested for stability? What tests were done at the extremes.

The intended use of the application is to systematize and standardize the calculation procedure for generating EMC compositing instructions from field hydrograph and sample time data.

The application was tested for stability by requiring a standardized data template. An example data template, along with instructions for data requirements and uploading data to the application are included in the ShinyApp homepage.

Using this Calculator

This calculator will produce a table of aliquot volume values, a hydrograph, and, if pollutant data is provided, pollutograph(s) for the given data set of flow rate measurements and sample timestamps.

1. **Download the Excel template file here and overwrite it with your data. See the Data Requirements section below.**
   NOTE: a 'download.htm' file may be downloaded instead of the template Excel file if the link is clicked too soon after launching the application. This is a known issue with the 'shiny' R package which was used to develop this application. Please allow a few minutes for the correct download to become available.
2. **Upload your data by clicking the 'Browse' button, selecting the updated Excel spreadsheet, and clicking the 'Submit' button. The calculator will generate the aliquot volume table as well as the hydrograph and pollutograph(s), depending on the uploaded data. If pollutant data is provided, the calculator will also provide the Event Mean Concentration for each of the specified pollutants.**
3. **Use the 'Start Time' and 'End Time' inputs to filter the data to the appropriate time range. These values are measured in minutes after the start of the provided data.**
4. **After changing the 'Start Time' and 'End Time', click the 'Redraw Graph(s)' button to regenerate the aliquot volume table, hydrograph, and pollutograph(s), filtered to the provided times. The grayed-out sections of the graph will *not* be included in the aliquot volume and event mean concentration calculations.**

5. **The 'Composite Vol.' input is used in the aliquot volume calculation such that the sum of the aliquot volumes will be equal to the composite volume value entered here, measured in mL.**
6. **The 'Water Volume Units' input is used to accurately label the calculated 'Total Hydrograph Volume' output.**

Data Requirements

The uploaded Excel spreadsheet must conform to the following requirements:

- **It must contain exactly two sheets, one for the flow rate measurement data, and one for the sample timestamps/pollutant measurement data.**
- **The first sheet must have exactly two columns, one for the timestamps and one for the flow rate measurements.**
- **The first column of each sheet must be timestamps with both date and time in the 'mm/dd/yy hh:mm:ss' format. The 'Datetime' columns in the provided template file are already in the correct format.**
- **Any number of pollutant columns in the second sheet are supported. If you do not have pollutant data, delete the 'Pollutant' columns entirely before uploading the template.**
- **The column headers are required and can be renamed as needed, but cannot be exclusively numeric characters [0-9]. The flow rate and pollutant column headers will be used for axis titles and can contain the units of the measurements, for example.**
- **All flow rate and pollutant measurements must be greater than zero.**
- **There may not be any missing values in the spreadsheet.**

Data checks were included that ensured data met data requirements. Application is agnostic to the type of pollutant used, so long as the data are represented numerically.

Stability of DATETIMES is a known issue for this type of application. At present, data must be in the format of MM/DD/YY HH:MM:SS where HH is between 0 and 23. Other datetime formats may fail data checks in the ShinyApp preprocessing stage.

Consistency - Does the application work across multiple platforms? Who performed the application testing across browsers?

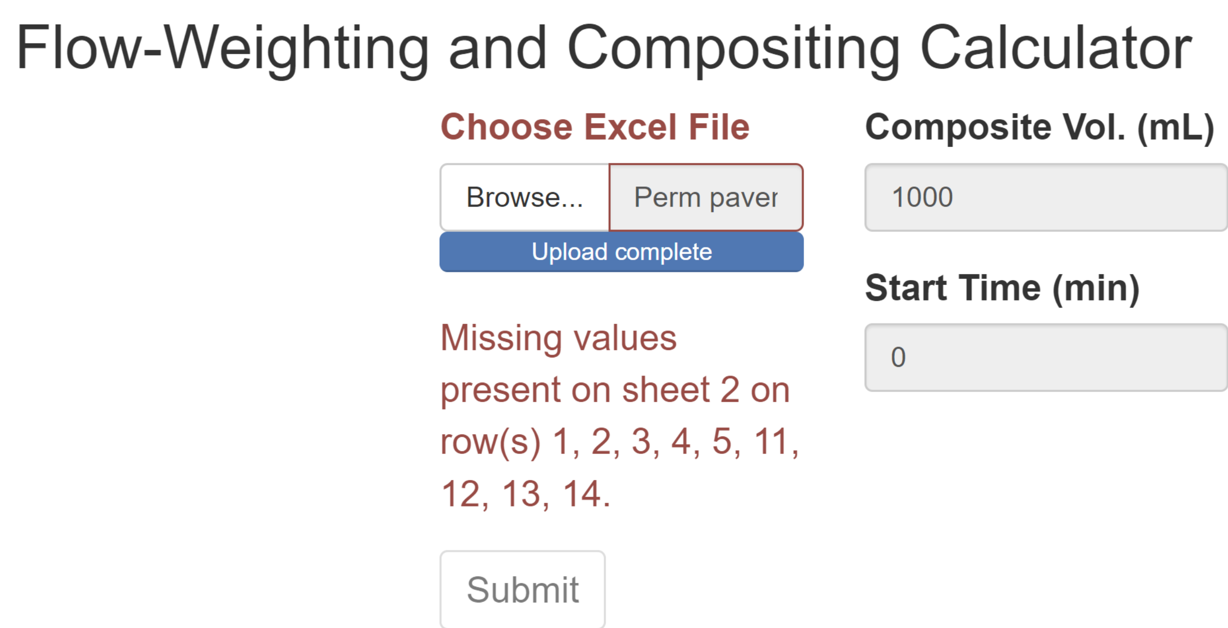
Application was built for web use and should function properly on all current browsers (circa 2023): Safari, Chrome, Firefox, and Edge. Older browsers are not supported and there is no guarantee that the application will function as expected on older computer systems.

Integrity of input data: What data did you use to test your tool and how do you know the data was “clean”? Did you check for outliers with the test data you used?

Datasets from flow and pollutant monitoring stations at an Auckland Wetland Outlet and Fashion Valley BMP were used to test the application. Raw data was manually formatted to meet the data requirements spelled out by the ShinyApp web application.

Data outliers, such as negative flow rates and non-sequential datetimes, were checked in the data formatting, upload, and preprocessing phases.

Data with missing values, incorrect data types, etc return an error message when the user attempts to upload the data to the application.



Accessibility - Survey of beta users - Which users should be able to use this tool and does the documentation provided give them guidance on how to use the tool?

Need beta users

Applicability to target audience - Survey of beta users - About how they intend to use the tool. Validate that the tool can be used to address management questions.

Applicability of application given by Readme on application Github.

The “gold standard” of stormwater control measure (SCM) performance monitoring is to collect and analyze water quality data as flow-weighted event mean concentrations (EMCs). EMCs may be generated by post-storm compositing of discrete samples as an alternative to collecting flow-weighted composite samples using automated equipment with integrated flow meters. Coincidentally, a standard procedure is not documented in industry-relevant literature to perform the requisite post-storm flow-weighting calculations. Lack of a documented procedure may deter data collection for agencies new to monitoring, or for complicated site conditions. It also leads to different approaches to post-storm flow-weighting that may influence resultant EMCs and mass loads.

This web application has been developed to enable consistent, transparent, easily applied calculations for post-storm flow-weighting and compositing and/or to generate an EMC from a pollutograph. The web app provides flow-weighted compositing instructions based on a user-uploaded hydrograph and times of sample collection, or returns an EMC based on a user-uploaded hydrograph and pollutograph. Total hydrograph volume is also returned so that users may determine a mass load from the EMC.

Version control process

All versions and development changes are documented in a github repository: https://github.com/SCCWRP/FWCCalculator

Documentation and ownership/contact

All documentation, including aspects of this QA process are included in a github repository: https://github.com/SCCWRP/FWCCalculator

Table 1. Quality control performance documentation table

Criterion	Target	Proportion Tested	Actual Performance	Date Evaluated	Stage of Process	Evaluator	Notes or Flags
Accuracy	100%	100%	99.8%	01/02/23	Dev	E.T.	
Stability and robustness	95%	100%	95%	2/14/23	Dev	N.L.	Date Time Formatting Issue
Consistency	95%	100%	100%	2/14/23	Dev	E.T./E.F.B.	
Integrity of input data	100%	100%	100%	2/14/23	Dev	E.F.B.	Missing Values Character Data Types
Accessibility - Beta	90%	—	—	—	—	—	—
Applicability to target audience - Beta	90%	—	—	—	—	—	—
Transparency of uncertainty	100%	Technical Report in preparation					
Version control process	Y/N	Github/Shiny.io					
Documentation and ownership/contact	95%	Github ReadMe / Shiny App Instructions					

X [signature]

Engineering Department Head, Dr. Elizabeth Fassman-Beck