

Metadata for the dataset associated to the manuscript:

Comparison of epifluorescence microscopy and flow cytometry in counting freshwater picophytoplankton

Pauliina Salmi^{1*}, Anita Mäki¹, Anu Mikkonen¹, Veli-Mikko Puupponen¹, Kristiina Vuorio² & Marja Tirola¹

¹Nanoscience Center, Department of Biological and Environmental Sciences, University of Jyväskylä, Surfontie 9C, P.O. Box 35, FI-40014 Jyväskylä, Finland

²Finnish Environment Institute (SYKE), P.O. Box 140, FI-00251 Helsinki, Finland

*Corresponding author: Email pauliina.u.m.salmi@jyu.fi

The dataset is divided into four subfolders:

- 1) "SEM experiment data" contains Scanning Electron Microscopy data, epifluorescence microscopy data and flow cytometry data of cultured *Synechococcus*, *Chroococcus* and *Snowella*
- 2) "raw data" contains epifluorescence microscopy and flow cytometry data of picophytoplankton from Finnish lakes. This has two sub folders "flow cytometry raw" and "microscopy raw"
- 3) "flow cytometry calibration data" contains data for cell size calibration with latex beads and volumetric calibration for the flow cytometer
- 4) "processed flow and microscopy data" contains excel workbooks for the figures shown in the manuscript

Detailed description of the subfolders:

1) SEM experiment data

- Folders named "cultures_2017_05_dd" contain flow cytometry raw data in standard flow cytometry format and analysis files in *.flow format.

- Raw data is recorded of the algae samples (e.g. chroo_1.002) and of diameters of latex beads with file name indicating the bead diameter (e.g. bead_1.005) and volumetric calibration of the flow cytometer (e.g. calib1.001).
- *.flow are analysis files created by Flowing Software (University of Turku, <http://flowingsoftware.btk.fi/index.php?page=1>) and contain information of cropped populations. These files contain references to the standard flow cytometry files, not the data itself. If you open these files, you should check the directory to the accompanying flow cytometry standard files. You can modify the directory by opening the *.flow file with e.g. Notepad++
- *.xlsx and *.txt files contain information of FSC and SSC values of each cropped algae population. These were used when calculating the average diameter of each cultured species based on flow cytometry

- Cultures_processed_data.xlsx file contain data of epifluorescence microscopy of the cultures on their own sheets. This excel file also contains abundance assessments based on microscopy and flow cytometry. This file contain also FSC (forward scattering) conversions to cell diameters on the last sheet, named fsc_diam_calibration. The formation of calibration equations can be found in the "4) "flow cytometry calibration data" subfolder.

- SEM_measurements.xlsx contain data of cell diameter measurements done using Scanning Electron Microscope (SEM). Each strain is on own sheet with an example image.

2) raw data

flow cytometry raw

- Folder names indicate the date when flow cytometry was carried out for the specific lake samples.
 - In the folders, file names indicate the lake where the sample was taken from. "LakeName_alg.00N" or "LakeName.00N" are raw data yielded by the flow cytometer and are in standard flow cytometry format.
 - Raw data of flow cytometry runs for volumetric calibrations are also included in the folders named with dates.
 - "LakeName.flow" and "flowRate.flow" are analysis files generated by Flowing Software (University of Turku, <http://flowingsoftware.btk.fi/index.php?page=1>). Note that *.flow files contain only references to the raw data files, not the data itself.

microscopy raw

- File names indicate the lake names
- *.bio files are analysis files generated by the proprietary Biomass counting programme programmed for Windows using ActiveX. These files can be opened with e.g. Notepad or MS excel, but the configuration might look obscure.
- *csv files are data files exported from the Biomass counting programme. These files show abundance, biomass and 95% confidence intervals for biomass in the microscopy samples.
- *Luotettavuudet.xlsx or *luotettavuudet.xlsx files contain calculations of 95% confidence intervals for abundances. These were calculated by opening .bio files in MS excel and using the matrix formed based on counted units per cell type per microscopic field.

3) flow cytometry calibration data

- This folder has a subfolder "calibration beads" that contain flow cytometry standard files of the beads used for calibrating the forward scattering to cell diameters in addition to the files created on flow cytometry for the SEM experiment.
 - "beads_size calibration.xlsx" file is the analysis file where the bead data is combined and the regression model established.
 - "flow_rate_calibration.xlsx" holds information of the measured flow rates on each flow cytometry analysis day
-

4) processed flow and microscopy data

-“FSC_and_SSC_data.xlsx” contain the FSC and SSC values of the PC, PE and Euk cells populations from all lakes.

-“microscopy_abundances.xlsx” holds microscopy abundances compiled in one excel file. This file contains results from years 2014-2015. Only results from year 2015 were utilized in this manuscript, because in 2014 no flow cytometry was done.

-“microscopy_biomasses.xlsx” holds microscopy biomasses compiled in one excel file. This file contains results from years 2014-2015. Again, only results from year 2015 were utilized in this manuscript.

- “comparison_data.xlsx” contains the processed flow cytometry and microscopy data that was used to create the figures in the manuscript

PE cells gave distinguished signal on FL3 channel in FacsCalibur, but their signals were mixed in PC and Euk populations that were cropped using FL2 channel. To eliminate this, very minor but existing source variation the following folders and files therein were created:

-Folder "FSC_individual_counts_exported" holds "LakeName_PC/PE/Euk.txt" or "*.csv" files that were exported from the Flowing Software. These files have information of forward scattering values (FSC) and FL3 (PC or Euk cells) or forward Scattering values and FL2 (PE cells). If PE cells were present in the sample, the PC and PE population values were differentiated as follows:

$PC_events = PC(FSC, FL3)_events - PE(FSC, FL2)_events$

$Euk_events = Euk(FSC, FL3)_events - PE(FSC, FL2)$

- The files "LakeName_PC_erotteltu.csv" holds scattering values specific for PC cells only.

- file "FSC_size_data.xlsx" holds the value of each picophytoplankton event converted to cell diameter

-Folder "FSCjaSSC_PE_erottelut" contain FSC and FL2 but also SSC data in .csv datafiles from each population and are marked "erotteltava" if this data contain mixed population", "erottelija" if this population contain solely FC cells and "valmisSSC_FSC_FL2_data"

"FSC_and_SSC_data.xlsx" holds forward and side scattering data of each separated picophytoplankton event. See the C# script in the accompanying script.cs -file that was used to separate the populations.

-The FSC and FL2 data from all the separated PE populations is gathered in folder "PE_erottelut_PCjaEuk_populaatioista"
