

# Tube Floc Data Analysis

## Data Analysis

### New Tasks for Spring 2013

#### Floc sedimentation velocity vs coagulant dose

Process the data from Karen Swetland's research to extract the average floc sedimentation velocity as a function of coagulant dosage. It should be possible to assess which coagulant, PACl or alum, produces flocs with a higher sedimentation velocity given the same aluminum concentration. This task does not require any additional laboratory work. It will require writing code to streamline the data extraction process. Think carefully about the best way to measure the time at which the turbidity is at 50% of the flocculated water turbidity. Some data smoothing or line fitting may be needed to get a reasonable estimate of the time.

Create graphs of mean floc size as a function of coagulant dose for each of the turbidities that are available. Determine if the ratio of mass of coagulant per mass of clay is a good predictor of floc sedimentation velocity. Analyze the results using the insights from our new flocculation model that emphasizes the role of fractional surface coverage. Determine if equilibrium floc size can be predicted from the fractional surface coverage or if a more complex analysis is required. Make sure to take into account that for low turbidity samples more of the coagulant ends up on the flocculator walls.

## Past Research

### Overview

During each experiment, data are automatically recorded in an Excel file (see example). The address for the excel file is recorded into the Meta file which is a spreadsheet with all the directories of the experimental data that will be/has been analyzed. Process Controller will be referring to the Meta file in order to acquire appropriate information on the experiment being analyzed; the Meta file will be an address book for Data Processor to efficiently access data.

The first graphs produced by the Mathcad file are the plots of the raw turbidity data vs. time and raw turbidity vs. settling velocity. The settling velocity ( $V_s$ ) is calculated by dividing the distance between the ball valve and the zone illuminated by the infrared LED ( $z$ ) of FRETa by the time elapsed ( $t$ ).

$$V_s = \frac{z}{t}$$

For experiments involving varying alum dosage, another graph was produced to obtain the residual turbidity vs. settling velocity. Residual turbidity is the turbidity resulting from the flocs that failed to reach the capture velocity, which is 0.12 mm/s for the sedimentation tanks of the AguaClara plants. From this, we can get the mean residual turbidity for each variation of alum dose. The mean residual turbidity is the mean of all the raw turbidity measured after a time of approximately 22 min till the end of the sedimentation state. It corresponds to a range of sedimentation velocities from 0.12 mm/s to 0.089 mm/s. See example of the residual turbidity graph: [Analysis Alum Dose Experiment](#).

Then data smoothing and normalization are performed on the raw turbidity data. The smoothing allows us to exclude outlying data points caused by really large flocs passing in front of the light sensor of FRETa, which create turbidity fluctuations. The normalization allows us to compare data sets with varying influent turbidities. The plot of normalized turbidity vs.  $V_s$  (settling velocity) can be interpreted as a cumulative distribution function (CDF) of turbidity with respect to  $V_s$ . A CDF describes the probability that a variable is less than or equal to some value.

To make the analysis more robust, the experimental data was fit to a gamma distribution. Then a derivative of the CDF of the gamma distribution gives a probability distribution of the particle population with respect to  $V_s$ . From the fitted data, the data processor retrieves the mean sedimentation velocity for each alum dose as well as the coefficient of variation of the distribution. The coefficient of variation is the standard deviation of the distribution divided by the mean. It is an indication of the width of the sedimentation velocity distribution.

For more discussion of the data analysis, see Ian Tse's [thesis](#) (Chapter I, data analysis). See an example of the gamma PDF graph: [Example of Analysis Alum Dose Experiment](#).

### Analysis Process

The data analysis process of the current project emphasizes **organization** and accuracy. The following is a step-by-step guide for the analysis process.

1. Go to the file containing the desired experimental data
2. Skim through the datalog (excel sheet with experimental data) to check for any apparent error. (Wrong influent turbidity, no variation in alum dose...)
3. If no apparent error is found, open the Meta file and record the appropriate information\*\*

#### \*\*Meta File Information

-The Meta File is like as an address book for the datalogs. It is important to update this file after an experiment for the Data Processor to access. The steps following will show how information is recorded into the Meta File:

- a. Put the appropriate ID Tag number in the A column
- b. If the duration of the experiment was three days, put the first day in the B column, the second day in the C column, and the third day in the D column; if the experiment only lasted for 2 days, C column will be left as zero, and the following will be applied for an experiment with a one day duration-only the B column will be filled.
- c. Record the appropriate directory and flocculator length in column F and G
- d. Column J will be adjusted according to the datalog; it will be the first row in the datalog where the settle state has been activated. \*\*Note, for most experiments, start row is not adjusted
- e. Comment on the experiment the datalog is referring to (eg. influent turbidity, alum dose, etc.)

**Remember to save the Meta file.** If not, the Data Processor won't be able to access the data!

4. Open Data Processor on MathCad
  5. Change the directory Data Processor is referring to for data (Enter the ID tag) \*\*
- \*\*Reference address will be located at the top of Data Processor, and only the last portion of the directory will need to be changed-which is the ID tag appearing on the Meta file. To see where to put the MetaID, see figure 1

Data Processor Functions must be referenced correctly for this Data Processor to work:	
Reference: N:\AguaClara Team Folders\Research and Development\LabFlocTeam\Spring 2010\Data Processor\Data Processor Functions v 2.xmcd(R)	
Directory of Meta File:	directory <sub>metafile</sub> := "N:\AguaClara Team Folders\Research and Development\LabFlocTeam\Spring 2010\Data Processor\"
Meta data ID of experimental run (look in Excel meta file for description of each experiment)	metaID := 1

**Figure 1:** MetaID location in Data Processor

6. Run the Data Processor; this program will automatically give graphic forms of the data and calculate needed values.
  7. In the data processor you will need to change a value. FirstFittedcolumn.
- The Data processor cannot fit the data for alum dose too low. This will enable you to eliminate the data that would not be fitted from the data vector. To see where to put the fixed FirstFittedcolumn see figure 2.

$NTU_{normed} := \text{normalize}(\text{time}_{plot}, \text{data}_{plot})$	
$V_s := \text{vel}_{logged}(\text{time}_{plot}, h_{column}, 10)$	
$\text{Firstcolumn}_{fitted} := 2$	<p><b>FirstFittedColumn edited</b></p> <p>Number of the first alum dose from which the data can be fitted</p>
$\text{FittedCDF} := \text{getfits}(\text{submatrix}(NTU_{normed}, 0, \text{rows}(NTU_{normed}) - 1, \text{Firstcolumn}_{fitted}, \text{cols}(NTU_{normed}) - 1), V_s, \text{vlog}, \text{time}_{plot})$	
$\alpha, \beta, \gamma := \text{FittedCDF}_0$ <p>array of curve fit parameters (one column for each data set)</p>	
$\text{FittedPDF} := \text{derivative}_{CDF}(\text{vlog}, \text{FittedCDF}_1)$	

**Figure 2:** FirstFittedColumn location in Data Processor