ENU 4134/6135 – Some Notes on Handling Data

Fall 2018

Correlating – No Single Algorithm

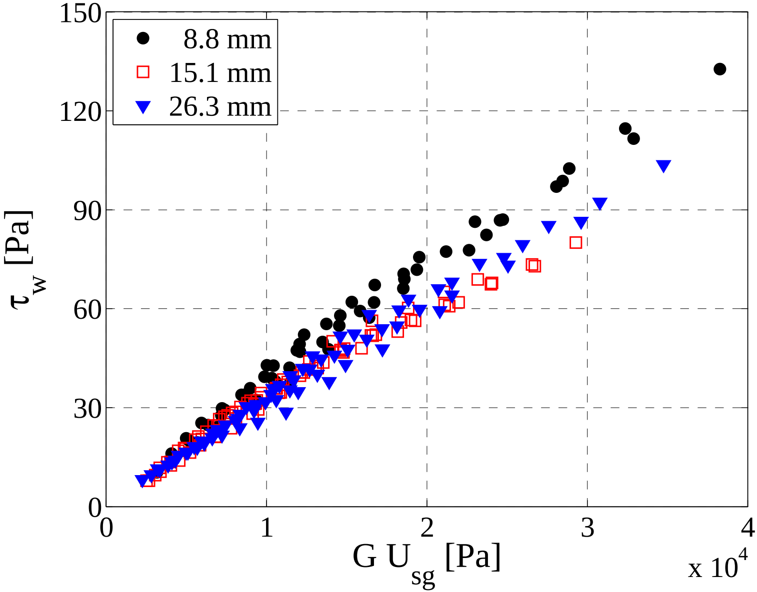
There is no single method to develop the equation to correlate experimental data, once you've determined which parameters to include.

Power laws (such as $Nu = ARe^{B}Pr^{C}$) are the customary first approach, but have no guarantee of success. They are especially poor in two-phase (in which both gas-based and liquid-based parameters have the same dimensions).

Suggestion: find a group of parameters that match dp/dz in dimension and then multiply by a dimensionless factor (friction factor, f). For example: a density times a velocity times a velocity divided by a length. Or: a mass flux times a velocity divided by a length. There should be "good-ish" correlation between your group of parameters and dp/dz, since f will not be terribly powerful.

Something to consider...

Similar, *but not identical*, data range to Mini-Project 1.



Using Flow-by-Flow Errors in Two-Phase Data

Consider making a spreadsheet grid of your data (with gas flow rising left-to-right, liquid rising bottom-to-top) and showing the error for each flow in the form.

- If left side is negative, right positive (or vice versa), gas effect mispredicted
- If top negative, bottom positive (or vice versa), liquid effect mispredicted
- If top left (high water, low gas) is positive, bottom right (low water, high gas) is negative (or vice versa), quality effect misprediced.
- If top right (high water, high gas) is positive, bottom left (low water, low gas) is negative (or vice versa), total mass flow effect misprediced.
- If two opposite corners high, the other two low... your correlation is probably non-physical. (But may give OK results for MAE/RMS and an OK-to-good grade.)

Metrics of Average Error

$$MeanError = \frac{1}{n_{FC}} \sum_{FC} \frac{XX_{corr} - XX_{exp}}{XX_{exp}} \times 100\% \quad (1)$$

$$MAE = \frac{1}{n_{FC}} \sum_{FC} \left| \frac{XX_{corr} - XX_{exp}}{XX_{exp}} \right| \times 100\% \quad (2)$$

$$RMS = \sqrt{\frac{1}{n_{FC}} \sum_{FC} \left(\frac{XX_{corr} - XX_{exp}}{XX_{exp}} \times 100\% \right)^2} \quad (3)$$

These averages are:

- Constant under multiplication of data and correlation by a constant (*e.g.*, m vs. ft).
- *Not Always Constant* if data shifted by an additive constant (*e.g.* temperature scale).
- Constant if data non-dimensionalized using dimensional variables (*e.g.*, ΔP vs friction factors).

ME, MAE, RMS Observations

- For an optimization correlation, ME should be very close to 0 (1% or less)
- MAE is the most useful metric to minimize
- The further MAE and RMS diverge, the more physical your correlation is likely to be (1.25 ratio from simulated data with random Gaussian uncertainty)

Statistical Metrics of Correlation

Correlation coefficient:

$$R \;=\; \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{s_x}\right)\left(\frac{Y_i-\overline{Y}}{s_y}\right) \tag{4}$$

... where X_i and Y_i are (in either order) correlated and experimental results.

Often reported as R^2.

R^2 is:

- Constant under multiplication of data and correlation by a constant (*e.g.*, m vs. ft).
- Constant if data shifted by an additive constant (*e.g.* temperature scale).
- *Not Always Constant* if data non-dimensionalized using dimensional variables (*e.g.*, ΔP vs friction factors).

Which to Use?

The dominant means of reporting accuracy in the literature in with an error statistic – Mean Error, MAE, or RMS. The second most popular is by citing the fraction of data within a specific band (*e.g.*,, 80% of the data within 20%).

Recommendations:

- Use MAE unless there is a compelling reason to do otherwise.
- If there is not a physically meaningful zero, consider using R^2 instead (or re-defining the zero so that it will have meaning). .
- It is often a good idea to compute both, even if you don't use them. Example: even a poor friction factor correlation (judged by R^2) can have a good MAE if the range of ρv^2 is very wide.