**Prediction of Pressures as a Result of Air & Liquid Flow**

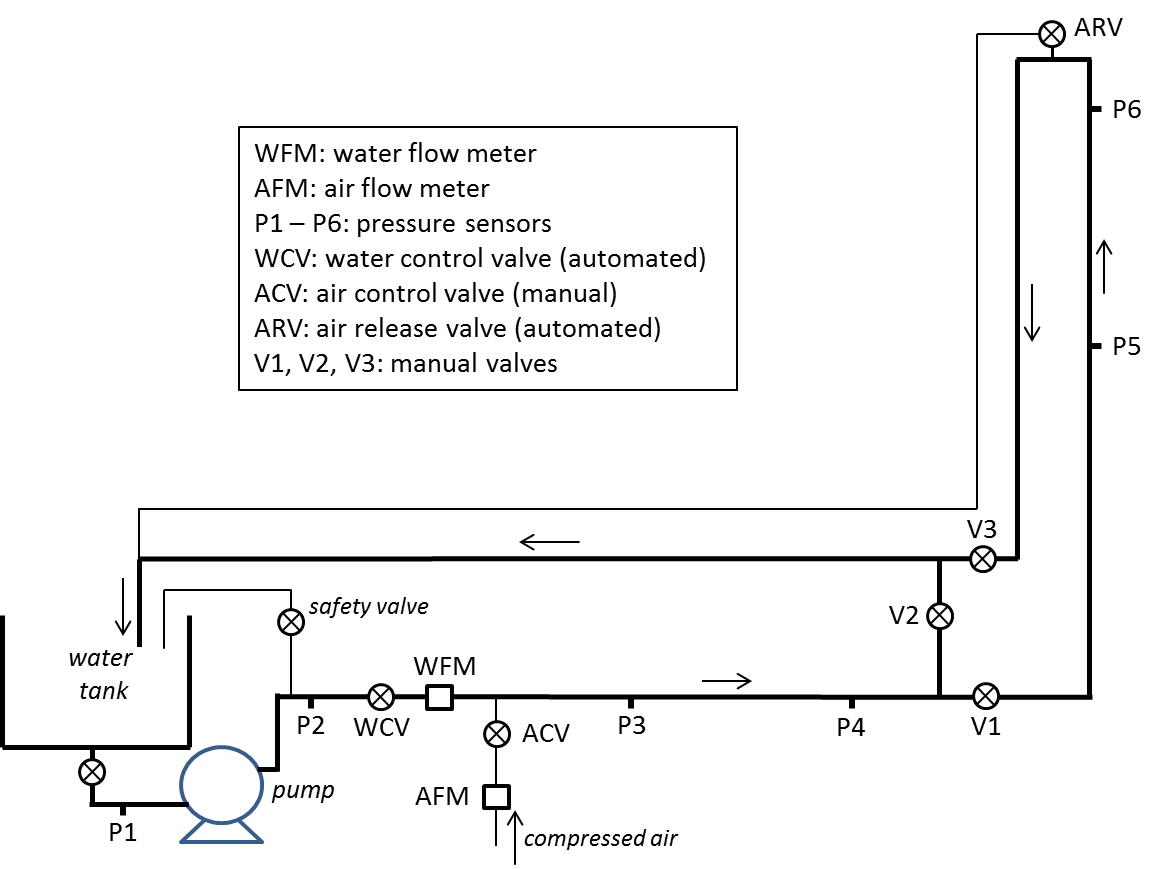
Group 2 (Lisa Reisenauer, Triton Wolfe, Joshua Randrup, Zhenzi Yu, Yuhan Yang)

**PROJECT BACKGROUND**

In a production environment, a significant amount of resources are invested in being able to anticipate phenomena, such as reliability events and process variability, before they occur. Examples would be trying to predict fouling factors and the differential pressures of a system based on process flow rates. In some processes, like the Methacrylic Acid production process patented by Rohm and Haas, where injected air causes two-phase flow to be present, this can be particularly difficult.

**PROECT OBJECTIVE**

A unit operations set-up on Georgia Tech’s campus injects air into a liquid water stream and measures downstream pressures. The schematic diagram is shown as Fig. 1. This project aims to use data from this set-up to create a regression model that can correlate the downstream pressures to the pressures in upstream and the air/water flow rates.  Namely, the single target variable is P6 (Fig. 1), features are P1-P5, air/water flow rates.



**Figure 1**. Schematic diagram of two-phase flow

**DATA DESCRIPTION**

The dataset is provided by the Georgia Tech ChBE Unit Control Lab, which is obtained during the two-phase flow correlation experiment. It consists of 8 parameters: air flow rate, water flow rate and 6 pressures along the pipeline. We will use the pressure-6 as our target variable, while the rest 7 features are the inputs. We have 8057 data points available, which is a good sample for this system, considering the number of features. As shown in Jupiter notebook, after data visualization we find that there are mainly three types of distribution: near normal, random, and damped which indicating a pulsing flow. Moreover, the orders of magnitude of these features vary a lot. Thus we will rescale the features before further processing.

**GENERAL STRATEGY**

After data exploratory analysis and cleaning, algorithms that have low complexity will be considered first to build a baseline model. The baseline model can later be optimized by utilizing time-series analysis and modeling techniques. Principle component analysis and feature engineering will also be implemented if necessary, to further improve the model.

**PROJECT RISKS & CONTINGENCY PLANS**

The main risk our project currently faces is potentially insufficient data to reach appreciable generalizability with our model. This would likely arise from lacking variation in certain input conditions such as valve position that would change the amount of fouling present in the system. If such a case did arise and the data was found to have insufficient variations from our desired inputs, we would have to obtain additional data ourselves. Presuming we can schedule a time around those in charge of the two-phase flow mechanism of Georgia Tech’s unit operations lab, obtaining additional data should not be a significant setback.

**BIBLIOGRAPHY**

[1] Curtis Ingstad Carlson, J. H., Michael Stanley DeCourcy, H. T., & Jamie Jerrick John Juliette, H. T. (2007, August 07).*United States Patent No. US 7.253,307 B1.*