Prediction of Pressures as a Result of Air & Liquid Flow

By 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.

## PROJECT OBJECTIVE

A unit operations set-up on Georgia Tech’s campus injects air into a liquid water stream

and measures downstream pressures. This project aims to use data from this set-up to create a

model that can predict pressures downstream of the air injection based on air and water flow rates.

## 

Figure 1: Pressure System Used to Collect Data

## GENERAL STRATEGY

Using various pressure measurements as well as known values of air and liquid flow,

multilinear regression can be performed on the data to obtain a model correlating pressure to two-phase flow. This baseline model can later be optimized by utilizing time-series analysis and

modeling techniques.

## 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

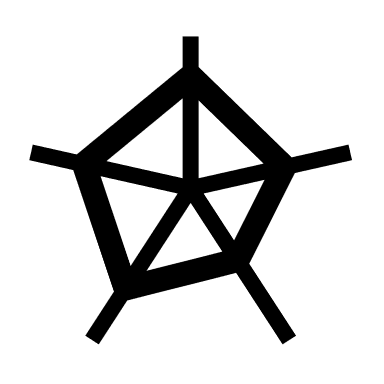
around those in charge of the two-phase flow mechanism of Georgia Tech’s unit operations lab,

obtaining additional data should not be a significant setback.

## BASELINE MODEL AND WORKFLOW

The baseline model pipeline is shown as Figure 2. As a starting point for analyzing our data and determination of likely favorable models, a correlation matrix was created for the various features of our data. Variables P3 and P4 show a relatively linear relationship; the 1-D regression analysis indicates a single feature is not enough to give accurate prediction of our target variable (P6).

Linear, non-linear, non-parametric models and time-series models were tried out at this exploratory stage. Specifically, the models implemented were multi-linear regression, Kernel Ridge Regression, Lasso, and time-series. As shown in the Jupiter notebook. the multi-linear model got comparable r2 score for training set and test set. But the model tends to underestimate the P6 in the high-pressure region. The KRR model was relatively successful, as it was able to predict the test set quite well based on the r2 scoring metric. One downside of the model was that it required significant tuning of the hyperparameters, potentially signifying limited applicability outside the range of training data. The Lasso model showed surprisingly poor performance, given how well the KRR model performed. The r2 score of its prediction was relatively constant below 60%, showing little variation with hyperparameters tended towards zero until producing errors. Given that the data is a time-series dataset, we also calculated autocorrelation using the time-series model and we found that 2 prior points having significant partial autocorrelations, but no long-term seasonal variations were visible. This make senses because we expect the system reaches equilibrium relatively quick. Also we see some scatter data points prior to 5 steps before also have high partial correlation, this probably due to the experiment design flaw. This is just a sample test for the incorporation of time-series theory while more details will be potentially included in the future work. Moreover, as a baseline model the KRR model’s success definitely bears continued attention and optimization as we progress.



Model Output

Data from Unit Operations Laboratory Pressure & Flow Set-Up

Feature Description

Correlation Matrix

1-D Regression of Each Feature

Basic Model

Multi-Linear Regression

Kernel-Ridge Regression

LASSO

Hold-Out

Training Data

Testing Data

Testing Data

Validation Data

*Figure 2. Workflow of the project*

## BIBLIOGRAPHY

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.