CS39620 Minor Project - Outline Project Specification
Smart Decisions for Welsh Water

Amy Richards (amr21@aber.ac.uk)
GG14 Computer Science/Mathematics

Supervised by Amanda Clare (afc@aber.ac.uk)

7th February 2018
Version 1.1 (Release)

Project Description

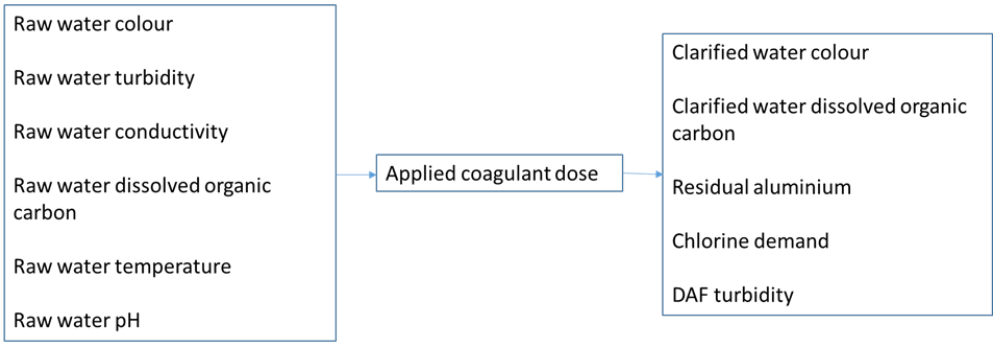
During this project I will be working with Dr Kate Martin, a process scientist for the company Welsh Water who supply drinking water to most of Wales and parts of Western England. 1.4 million people rely on Welsh Water for safe drinking water in Wales through their work of managing drinking water treatment and contamination avoidance.

An immediate reaction is required from Dr Martin to events as they occur at any time of the day in the interest of public safety. Such events include heavy rainfall changing the turbidity of the raw water entering the plant and will often present a challenge to conventional water treatment processes particularly coagulant control.

Turbidity of raw water is caused by suspended particles that are too small to settle called colloids. There are two types of colloid, organic (algae, bacteria and protozoa) and inorganic (clay, silt and mineral oxides). If the bacteria are ingested by humans, it can be fatal and there is a direct link between turbidity and risk of disease due to toxic compounds adsorbing to the surface of the suspended colloids. Turbid water is a cause of gastrointestinal infections including cryptosporidiosis.

A coagulant is a substance that causes a liquid to coagulate, Aluminium Sulphate $Al_2(SO_4)_3$ is a common choice for water treatment plants as it is effective, non-toxic and most importantly insoluble so that it can be removed after the treatment. Coagulation is the method of altering colloids so that they flocculate – the process in which colloids aggregate to form larger particles called flocs and settle to the bottom of the solution. Water isn't potable after undergoing flocculation, it must go through skimming and filtration to remove these particles.

Effective coagulation control is vital in ensuring the safety of potable water. For this project **I will be analysing historical data of coagulant doses to determine if the optimal choice was made at the time.** Coagulation choice is affected by several variables such as raw water pH, temperature, turbidity, alkalinity, conductivity and the dissolved organic carbon content. Once the chosen coagulant dose is applied, the outputs measured to determine if a dose was appropriate are the clarified water colour, the dissolved organic carbon content, residual coagulant, chlorine demand and the dissolved air floatation turbidity. All the mentioned outputs should be as low as possible for a good coagulant dose. I will be analysing 6 months of 15-minute data for all monitoring points on site and manual samples of other monitoring points are taken 2 or 3 times a week.



There are some commercially available units used to control coagulation dose including Com::pass, however a problem with this kind of unit is that they don't look at the effect of the dosing throughout the plant and have limited feedback loops. These units don't learn and are merely an instrument with a pre-set algorithm rather than an evolving one.

I am hoping that this project will give more of an insight into the factors that affect coagulation and help to optimise the process for choosing a coagulant dose in the future.

Proposed Tasks

- **Investigation into automatic coagulant control systems**
  As this is a new subject to me, I will spend some time reading around coagulant control systems that already exist and how effective they are. I will also learn about the effects of improper coagulation and different methods used to decide on a suitable coagulant dose.

- **Investigation into helpful Python libraries**
  As I have never developed a project of this size before that requires scientific algorithms I will be learning about useful Python libraries before I begin work on this project.

- **Development**
  - **I/O Functions**
    The first step of development will be to read in the large amount of data from the .csv files provided by Dr Martin into Python. I will organise this data into tables.
  - **Visualisation of Relationships between Input, Output and Dosage**
    The next step will be to plot the relationships between the inputs (raw water colour, turbidity, conductivity, dissolved organic carbon, temperature and pH) and the outputs (clarified water colour, dissolved organic carbon, residual coagulant, chlorine demand and dissolved air flotation turbidity) for the specified coagulant dose. Some relationships will be obvious to plot i.e. raw water colour vs clarified water colour but the relationship between other factors i.e. residual coagulant and the chlorine demand will be interesting. From these graphs I hope to find a correlation between factors and from that be able to find a deeper insight into the factors that affect coagulation.
  - **Exploration of Various Machine Learning Techniques to Determine Effect of Coagulants**
    Lastly, I will research into machine learning techniques and use the findings from the plots of relationships to determine the effects of the coagulant dose used and hope to use this part of the project to find useful data for Dr Martin. The data found here will hopefully be of use to Dr Martin when deciding on future coagulant dosages.

- **Project Meetings and Project Blog**
  During this project I will attend weekly supervisor meetings – one with a small group of students every Monday and a 1-1 with my supervisor every Thursday. These meetings will be a good time to ask my supervisor for guidance and to see how other students are progressing in their projects. I will be keeping a blog throughout the duration of this project where any notes from the meetings will be kept, along with a log of what I have achieved during the working week and what my goals are for the next week. The blog will be hosted on WordPress.

- **Preparation for Demonstrations**
  There will be two project demonstrations during this project, one of them will be held half-way through the project in March and the other held at the end of the project in May. The final demonstration will be held after the submission of the final report and technical work. At this point, I am not sure what functionalities will be ready by the mid-project demonstration.

Project Deliverables

- **Mid-Project Demonstration Notes**
  A set of notes will be produced that summarise the work presented at the demonstration. I will include this as an appendix in the final report.

- **Software**
  The software developed in this project will be developed in three parts as stated above. The software will read in data from .csv files, process this data and then plot the data showing the relationships between factors. It will then use machine learning to predict an optimum coagulant dose in the future.

- **Final Report**
  This document will be the major report on the project and will discuss the work completed in depth. I will discuss the findings of the data analysis in detail and discuss what went well and what could've gone better.

- **Final Demonstration**
  No formal documentation will be produced for this demonstration. However, as it is worth 20% of the final grade it should be considered when planning the work.

Bibliography

H. Ratnaweera, J. Fettig (2015) State of the Art of Online Monitoring and Control of the Coagulation Process. Available at **http://www.mdpi.com/2073-4441/7/11/6574** [Accessed: 7th February 2018]

S. K. Dentel (2009) Coagulant Control in Water Treatment. Available at http://www.tandfonline.com/doi/abs/10.1080/10643389109388409 [Accessed: 7th February 2018]

Version	Date	Change Made
1.0	06/02/2018	Initial document
1.1	07/02/2018	First draft