Investigation of Root Causes of Wastewater Pump Station Obstruction and Ragging Faults

Ellen Marinko

Berwin Turlach
Department of Mathematics and Statistics
University of Western Australia

David Burton
CEED Client: Water Corporation

Abstract

Ragging and obstruction of wastewater pumps is a significant performance issue across the Water Corporation's wastewater assets. For the Water Corporation to continue to deliver effective wastewater management to customers, there is a real need for greater understanding of the factors which influence incident rates, such as wastewater pump obstructions, enabling effective prioritisation of proactive maintenance. There is currently limited understanding of the root causes and drivers of ragging and obstruction events at wastewater pump stations. Previous investigations of root causes for wastewater pump ragging and obstruction at the Water Corporation have been limited and confined to relatively narrow focus areas. This project provides an updated approach to root cause analysis which includes exploring current preventative maintenance practices. The study comprehensively combines all available dynamic, seasonal, static, demographic and spatial data, resulting in 30 data inputs to address all potential root causes. Generalised Linear Mixed models will be developed to determine the key predictors which effect the probability of obstruction fault events, and the results obtained used to predict stations that are most at risk of ragging and obstruction events.

1. Introduction

The Water Corporation (WC) is a State Government owned entity that provides the majority of Western Australia's drinking water, wastewater and drainage services. The WC operates approximately 800 wastewater pump stations in the Perth metropolitan area. However, operation of each of these assets is intermittently affected by wastewater pump obstructions caused by rags, sand, and other material entering the sewers. From 2006 to 2017 inclusive, approximately 28,000 fault events associated with wastewater pump station obstructions and ragging were recorded by the WC. A rag is defined as a specific form of pump obstruction. Ragging is a gradual accumulation of textiles, tissues and fibers in a wastewater pump, which attach to the pump impeller vanes. Specific examples of common rags found in Western Australian wastewater systems are stringy materials such as wet wipes, hygiene products and table cloth fabrics. For example, a 2 meter obstruction weighing 50kg was flushed from plumbing in Port Headland. The massive mixture of fibrous materials, depicted in Figure 1, which contained wet wipes, sanitary items, and clothing took more than eight hours to remove after it created a blockage at the Hedditch Street pump station (Dougherty, 2018). It is an example of the costly and timely corrective maintenance needed to flush obstructions from waste water pumps and highlights the safety and hygiene considerations that mechanical tradespeople must take when performing corrective maintenance.

Figure 1 **Figure 1. Water Corporation mechanical tradesperson next to a 2m Obstruction, removed from a West Australian sewer, 2017. [Dougherty, The West Australian, 2017]**

Ragging and obstruction of wastewater pump components is driven in part by the decisions made by households and businesses regarding what they dispose of into the wastewater collection system. A lack of awareness, societal attitudes, and potentially misleading advertising from manufacturers of wipe products leads to inappropriate items entering the wastewater collection system; this includes fat, wet-wipes, nappies, and sanitary products. It therefore is plausible to investigate whether there is relationship between the populations and their assumed actions and the fault rate of pump stations. We thus explore external pump station factors, related to the pump station catchment demographics such as socio-economic status, land use and population density in relation to frequency of fault events.

The WC also wishes to evaluate preventative maintenance spending on pump stations in the Perth metropolitan area in relation to ragging and obstruction fault events at pump stations. The goal is to identify if the assets have been properly maintained in the past and whether this affects the obstruction handling ability of the station. Operation and maintenance procedures of the management authority are identified as key indicators of pump station fault in Hans Korving's work into Statistical Modeling of the Serviceability of Sewage Pumps (Korving, 2006). By comparing historical preventative maintenance patterns, we aim to determine if more frequent and greater spending is directly effecting fault rate of assets.

Modeling and predicting obstruction fault events of waste water systems have been attempted in previous external studies, with a variety of methods that incorporate a range of predictive covariates. The study of Predicting Combined Sewer Overflows Chamber Depth using rainfall radar data (Mounce, 2014) introduced the use of rainfall within a catchment to detect the occurrence of blockages. Mounce confirmed rainfall can influence obstruction faults over periods of several months. This was attributed to the soil and sand moisture, influenced by rainfall washing sediment into waste water sewage systems. These findings emphasise the need to incorporate temporal rainfall as a predictive covariate in the cause analysis. Predictive risk modeling of real-world wastewater network incidents (Bailey et al., 2015) present models for provision of a blockage likelihood level for pump stations using the asset and historical incident data from the wastewater network of Dwr Cymru Welsh Water. The findings supported the Water Corporation's preliminary study that the basic characteristics of the sewer (length, diameter, material), can provide good explanatory capability of the risk of blockage occurrence. However, the study concluded that the inclusion of temporal variables, such as rainfall and planned maintenance, would improve model accuracy, which confirms the motivation for the wider scope of this study and inclusion of such variables. We conclude that spatial, demographic and dynamic factors involving wastewater pump station catchments (with the exception of rainfall), are rarely reported in previous external studies. This analysis seeks to develop a robust understanding of root causes that lead to fault events at wastewater pump stations, and in doing so, identify wastewater pump stations which are at the greatest risk of obstruction fault in the Perth metropolitan area.

2. Methodology

The main goal of data acquisition and preparation was to create a longitudinal dataset. The dataset combined all covariates which relate to the fault event's time and location. Each fault record contained the technical and demographic information for the station and catchment in which the fault occurred, and the temporal covariates related to when the event occurred. The original dataset was compiled from 25,408 fault records which represent repeated measurements on 666 pump stations associated with obstruction and ragging faults.

2.1 Data Acquisition

The data extraction process was focused towards sourcing as much information as possible pertaining to the 25,408 fault records in the original dataset and the assets associated with them. This required data to be obtained from internal WC databases and external Australian Bureau of Statistics and Bureau of Meterology sources. All extracted data was validated, processed and collated in R to create the final dataset for exploratory analysis. The data extraction process resulted in 30 variables which represent possible causes of ragging and obstruction. We summarise the temporal, static and spatial data that was extracted using the framework that have previously been applied to pipe leakage fault causation studies (Gould et al. 2013). The potential causes of pump station fault rate variation are summarised by three distinct levels of factors:

- Asset Technical Specifications**:** such as pump station configuration, asset start up date, pump make and model, motor design, impeller specifications, and the pressure, speed & power the station was designed to function.
- Broader Environmental Conditions: such as pump station catchment observations. This includes weather and rainfall, socio-economic indexes, land use and population and population density of the waster water catchments.
- Asset Management Paradigms: such as rehabilitation practices, and risk tolerance. In particular, this is focused on preventative practices which involve the proactive maintenance spending and frequency at each pump station.

2.2 Statistical Methodology

The statistical methodology of this project is focused on implementing regression mixed models. Generalized linear mixed-effects (GLMM) models describe the relationship between a response variable and independent variables, using coefficients that can vary with respect to one or more grouping variables, for data with a response variable distribution other than normal. In this case, we fit a logistic regression, which is used to model the binary outcome fault or non-fault event, in which the log odds of the outcomes are modelled as a linear combination of the

30 predictor variables. The model includes fixed and also random effects which account for variations between each pump station that might affect the response. We aim to implement generalised linear mixed models with different frameworks and compare variable selection and thus inferences drawn under each model. A relatively new Deep Net GLMM technique proposed by M-N Tran (Tran, 2018), using a Bayesian framework will be compared to frequentist approaches to generalised linear mixed models.

3. Results

3.1 Exploratory Analysis

The initial data exploration resulted in some interesting observations regarding the frequency and spread of pump station faults across the Perth Metropolitan Area. Armadale, Kwinana, Warnbro, Gosnells and Rockingham exhibit high total fault counts which are largely disproportionate to the number of assets in the region, with the average number of faults per pump station above 50 total events and at least 40% of assets recording conspicuously high historical fault counts in each region. These results lead us to question what makes these regions different and whether regions share similar asset technical specifications, experience similar weather patterns or demographic criteria.

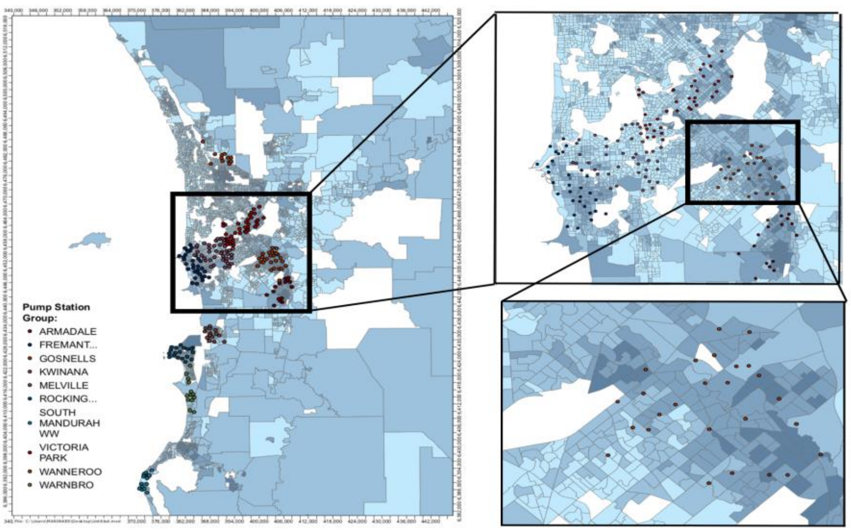

Figure 2 Pump station groups of interest in the Perth Metropolitan. The catchment variation in socio-economic disadvantage is represented by shading the ares by socio-economic disadvantage index score. (Darker shading indi cates high socio-economic disadvantage)

For the purpose of this paper, we present one key feature of the data. Figure 2 displays the distribution of pump stations, across the Perth Metropolitan area. It is not possible to identify SA1 demographic variation, from the view of the entire Perth metropolitan area, as differences in shades are not distinguishable on one map, without a closer view. The inset of the same plot, focuses on only Gosnells and Armadale groups, with the intent of examining variation in socio economic disadvantage levels in problematic areas, outlined in the exploratory analysis. Using 2016 data, we see that the Gosnells pump station group lie in an area of increased socio-economic disadvantage. This indicates that catchment demographics could effect fault rate of pump stations. We hypothesise that areas of higher disadvantage have lower education levels which might lead to a lack of awareness and societal attitudes towards flushing items into wastewater collection systems. However, these findings are limited as it only from a one year view point and focused on one particular area. We can not conclusively determine if socio-economic indexes of sewer catchments relate to pump station fault frequency during the exploratory analysis phase.

3.2 Variable Selection

The Bayesian Deep Net GLMM incorporates variable selection into the deep learning method (Tran, 2018). It may be the case that only a small fraction of the variables are substantially associated with the response, and we need to indentify which variables are key predictors of an increased probability of a fault event. As such, it is important to include only the variables which are significantly associated with the response in the final model to allow accurate model inferences to be concluded. At the time of publication, variable selection to determine the model fit in a Bayesian framework was in progress. Figure 3 is a graphical display of the shrinkage parameter for each variable plotted against the model iterations. Although, the mathematic technicalities are omitted here, the plot serves to demonstrate initial variables which are to be removed from the model. The shrinkage parameters for irrelevant variables keep increasing and are thus shown at the top right of the plot, where as the parameters with respect to the relevant variables keep decreasing towards 0 and remain in the model. Figure 3 (left) shows the initial run, where two variables (1. land use and 2. rain fall) were identified as irrelevant. Figure 3 (right) visualises the same variable selection process, once land use and rainfall variables were removed, indicating that 3. population count and 4. pump brand should also be removed.

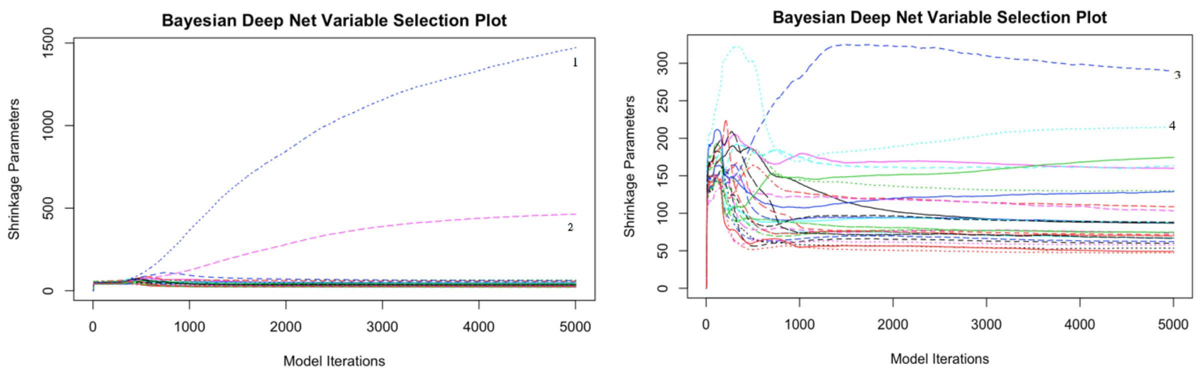

Figure 3 Plots of the shrinkage parameters over model iterations. The shrinkage parameters w.r.t the irrelevant variables keep increasing, while the relevant variables decrease.

We focus on the variables that are removed from the model. The Water Corporation accounts for areas of greater population with more pump stations in the region and thus pump stations in these areas should not be under more strain compariatively with other areas, which may explain why population is not a significant predictor of obstruction events. Interestingly, population density remains in the model which may indicate pump stations in high density areas, may not be able to process the volume or type of wastewater and have not been accounted for by the Water Corporation. The exclusion of land use variables is more surprising. As previously stated, ragging and obstruction of wastewater pump components are driven in part by the decisions made by households and businesses regarding what they dispose of into the wastewater collection system. For example the Victoria Park and Kwinana regions, contain a mixture of residential, commercial and industrial areas, and we hypothesised as a result, mixed materials may be entering the sewages and affecting pump station obstruction handling ability in these areas. However, this is not supported by the data, as land use is deemed insignificant. It is acknowledged that these findings must be compared to the frequentist framework of backwards variable selection and are not finalised.

4. Conclusions and Future Work

The initial findings presented in the exploratory analysis phase, are purely investigative and only serve as a weak indication of behaviour between variables rather than a confirmation of a relationship. Further work remains to be done in developing the two statistical models to determine how each variable effects the probability of an obstruction fault event occuring. Interaction between multiple variables and how they effect each other, is yet to be explored and the generalised mixed model will account such intereactions. This also requires a robust method for variable selection, which will be implemented using backwards selection in the generalised mixed modelling process, and compared to the variable selection of the Bayesian Deep Net method presented in this paper. Of particular focus is defining the relationship between preventative maintenance spending and obstruction and rag fault events, as exploratory examination of the data did not reveal any insight into whether prior preventative maintenance spending and frequency does affect the probability of a fault event occuring in the future.

5. Acknowledgements

This project was made possible by David Burton and Ian Gibb at the Water Corporation and by Associate Professor Berwin Turlach at the University of Western Australia. The author would like to thank all those involved, with an additional thanks to Chris Bryant and the entire Metro Asset Planning Management team and Medhi Peri and Nic Flett from the Geo-Spatial team at the John Tokin Water Centre, whom were always willing to help, their patience and guidance is much appreciated.

6. References

Bailey, J., Keedwell, E., Djordjevic, S., Kapelan, Z., Burton, C., & Harris, E. (2015) Predictive risk modelling of real-world wastewater network incidents. *Procedia Engineering*, **119**, pp. 1288–1298.

Dougherty, R. (2018) Eight hour slog to flush out 50kg blockage. Availiable from: https://thewest.com.au/news/regional/eight-hour-slog-to-flush-out-50kg-blockage-ng-b88704632z [2nd May 2018]

Korving, H. (2006) Statistical Modeling of the Serviceability of Sewage Pumps. *Journal of Hydraulic Engineering,* **132**(*10)*, pp. *1076-1085*

Mounce, S., Shepherd W. & Sailor G. (2014) Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data. *Water Sci Technol,* **69**(6), pp. 1326-1333

Gould, S., Marlow, D., and Beale, D. WSAA Fault Database Analysis General Report. Water Services Association of Australia and CSIRO, 2013.

Tran, M.-N., Nguyen, N., Kohn, R., & Nott, D. (2018) Bayesian Deep Net GLM and GLMM. *eprint (arXiv:1805.10157).*