

District Metered Area Customer Summer Night Irrigation Use Allowance Model

Kevin Kuek

Rachel Cardell-Oliver

Department of Computer Science and Software Engineering
University of Western Australia

Joe Standring

CEED Client: Water Corporation

Abstract

Estimating the amount of leakage present across areas of a water distribution network is needed to target repairs and reduce water loss in the network. Existing methods of leakage estimation do not account for seasonal variations in water use, and as a result are not suitable for use in Western Australia. The recent installation of smart meters that record water usage at hourly intervals on each service point in the suburb of Yanchep, has created the opportunity to analyse the population's water usage patterns and develop a model to predict their usage. Such a model can be applied to non-smart metered regions and subtracted from the total flow rate supplied to estimate the amount of leakage present in the distribution mains. The polynomial model developed is evaluated by its root mean squared error when predicting flows using unseen data.

1. Introduction

Estimating the amount of leakage across areas of a water distribution network is important for allowing the water service provider to target repairs. Existing methods of leakage estimation are not suitable for use in Western Australia outside of winter due to seasonal factors such as night garden irrigation. Due to this, non-visible leaks typically run continuously for extended periods and cause large water losses. This project aims to develop a method of measuring the growth in non-visible leakage in an area that is applicable to WA's conditions by modelling customer usage across the year and subtracting it from the total water supplied.

In 2017/18 the Water Corporation supplied 262GL of water to the Perth region including Mandurah, of which recoverable losses in the distribution network comprised an estimated 10GL (Standring 2018). Based on an ongoing active leak detection program, non-visible leaks contribute an estimated 85% of these losses. Visible leaks tend to be quickly reported by the public and repaired, while non-visible leaks typically run for over 2 years. Approximately 30% of non-visible leaks surface and are repaired within 12 months of occurring (Standring 2017) so the ability to identify non-visible leaks would be highly beneficial.

The Water Corporation has established five district metered areas (DMAs) in the Perth region that meter the supplied flow rate at 15-minute intervals. The Yanchep DMA is unique in that each service point (SP) has been fitted with a smart meter recording the flow rate supplied at hourly intervals. Smart metered every SP in a DMA would allow exact water usage and leakage

rates to be known but is expensive to implement and poses a large data storage challenge. Yanchep's data can be used to develop a model for predicting customer usage that can be applied to DMAs without smart meters, allowing the benefits of smart metering to be realised while reducing the associated cost.

2. Process

A model is needed to estimate the water usage in an area that can be applied to DMAs without smart metered service points. The model currently used by the Water Corporation was tested against Yanchep's smart meter data to provide a baseline prediction accuracy for evaluating models. The Yanchep data was used to develop a more robust model and methods of fitting it to DMA's without smart meters, such as training on a sample population, were investigated.

2.1 Data Preparation

The smart meters in Yanchep were installed in early 2017 and were activated in batches. Analysis was done on the period following the last batch coming online on 20/6/17, and new data is periodically analysed as it is recorded. Of the 3952 service points, 3776 were residential properties, the remaining service points were commercial properties, lots of vacant land and SPs labelled "OTHER". To account for missing data points on individual meters, the average flow across meters with a valid reading was taken for each timestamp. This is justified because the median percentage of invalid readings for each timestamp was 1.6%, and the distribution of readings during the target hours are heavily skewed towards zero. This is consistent with the power law and means that the missing readings are unlikely to impact the average flow. The daily maximum temperature was used for analysis and taken from the Bureau of Meteorology's website. This gave readings from the closest weather station to Yanchep, Gingin Aero.

2.2 Baseline Model

The current method employed by the Water Corporation for quantifying leakage is to analyse the minimum night flow (MNF) into a DMA using a method developed by the Water Services Association of Australia (WSAA). In this method, distribution leakage is estimated using:

$$MNF = \text{Minimum Night Use (MNU)} + \text{Unmetered Use} + \text{Night Leakage}$$

Unmetered use is assumed to be estimated by the Water Corporation and the MNU is broken into residential and non-residential use. The WSAA found that non-residential properties, excluding exceptionally high users, use average of 5-10L/Hr/SP, but analysis of Yanchep's data gave an average of 1L/Hr/SP and a median of 0.37L/Hr/SP. This difference is likely because the WSAA estimate comes from a study done in the UK, where water use patterns differ from Australia's. The WSAA's residential component of MNU is modelled as a single value dependant on variables such as: the expected toilet flush volume, number of residents, number of service points, average distribution pressure and an estimate of small household leaks (Wide Water Bay Corporation 2011). This model doesn't account for outdoor usage and consequently should only be used in periods without garden irrigation (Wide Water Bay Corporation 2011). In WA there is a sprinkler ban in effect through Winter during which the Water Corporation may use this model, but for the remaining 9 months it can't be used.

2.3 Polynomial Model

MNFs and MNUs exhibit two characteristics visible in figure 1: a recurring yearly pattern, and a relationship with the daily maximum temperature. By modelling these relationships, the expected MNU may be predicted and used to estimate the amount of distribution leakage in the DMA. Though this model must be trained on smart meter data for a DMA, after it is trained the smart meters may be redeployed to a different area, reducing the number that need to be purchased and the volume of data to be handled.

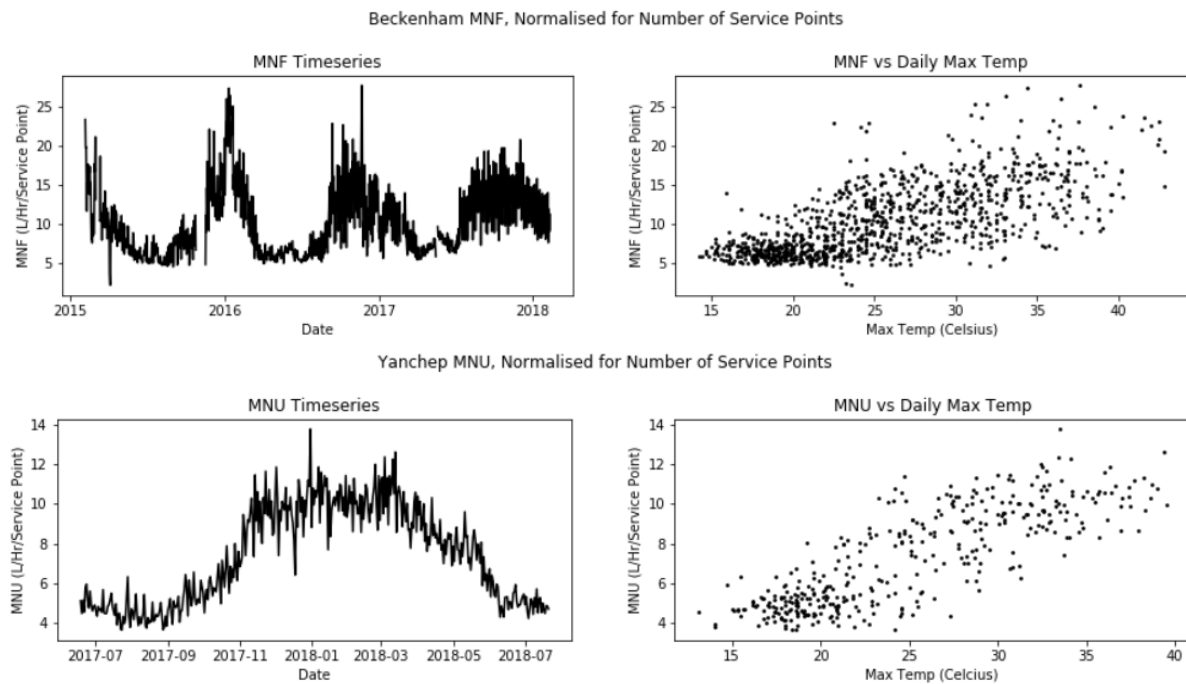


Figure 1 MNF in Beckenham and MNU in Yanchep timeseries and maximum temperature scatterplot.

2.3.1 Modelling Minimum Night Use

The timeseries suggest that a 4th order polynomial based on a reading's the day of the year (DoY) could be fit to the annual pattern and the scatterplots show a 3rd order polynomial relationship between the minimum flow and daily maximum temperature. The two explanatory variables were combined into multivariate polynomial models of various orders to determine which form best fit the data. To determine if combining the variables improved results, three sets of explanatory variables were analysed: the DoY, maximum temperature and the DoY and maximum temperature combined. Initial testing found that polynomials of order 6+ performed poorly due to being overfit to the training set, so the 3 sets of inputs were tested with polynomial of order 0-5. The timeseries was divided into training and testing sets separated at a 3:1 ratio of weeks. The model was then fit to the training set using ordinary least squares regression, and the root mean squared error (RMSE) of testing set predictions was used as a metric for performance. The RMSE of predictions is tied to the standard deviation of the residuals, and represents the uncertainty in predictions. As the models' performance varies with the composition of the training and testing sets, the mean and standard deviation of the model's RMSE was taken over 1000 different samples.

2.3.4 Limitations

This model must be trained on one year of data that doesn't contain distribution leakage, and consequently can only currently be applied to Yanchep's smart meter data. These smart meters have only been online for 15 months, so it can't yet be determined whether the yearly trend will be consistent across years. DMA level data in Yanchep is unavailable, so leakage volumes can't currently be quantified. The differences in usage patterns between regions also means that a model fit to Yanchep cannot be directly applied to another DMA without a method of retraining its parameters. The DMA level data available in other DMAs contain both MNU and distribution leakage, meaning that a model trained on it would be fit to the accumulated leaks and random bursts present in the training period.

2.4 Smart Meter Sample Population

Having smart meters on every SP in a DMAs would allow MNU to be accurately modelled but would be prohibitively expensive to deploy. Determining how the accuracy of a model trained on a sample population varies with the sample size would allow model accuracy to be balanced with cost of deployment. The polynomial model was trained on sample populations of increasing size, with each tested on data from the full population to measure performance.

3. Preliminary Results

3.1 Baseline Model

Testing the WSAA's MNF model against the Yanchep MNU data found that with the default parameters it severely underpredicts actual usage in the area (Wide Water Bay Corporation 2011). The WSAA noted that the model's parameters can be trained for specific DMAs if needed, though a training method isn't provided. The model gives a single value to estimate the MNF, so the best possible prediction is the mean MNU over the target period. Using this simplification, the best case prediction for Yanchep was found as 6.36L/Hr/SP for the whole year, and 3.98L/Hr/SP during the winter period within which the model is applicable. The models' fit can be seen in figure 2.

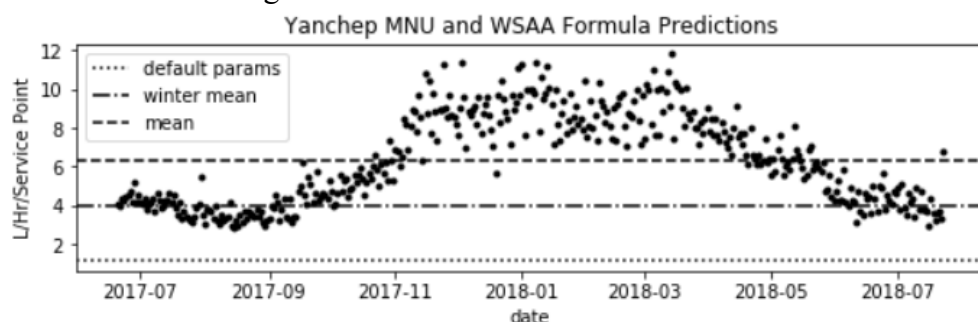


Figure 2 WSAA MNF model predictions tested on Yanchep's smart meter MNU data.

3.2 Polynomial Model

The experiments comparing different polynomial models of MNU found that a polynomial including both the DoY and maximum temperature was able to make predictions with the least uncertainty, as can be seen in the mean graph in figure 3. Though models based on the maximum temperature are stable across training populations, they have limited prediction

accuracy and are best represented as a linear function. Increasing the polynomial order exponentially increases the number of permutations of the input variables present in the formula and consequently makes it more prone to being overfit to the training population. This was observed with orders greater than five, for which the model's RMSE began to increase for both models containing the DoY. Due to this risk of overfitting, the 3rd order polynomial was determined to be the most appropriate model. When trained on a full year of data, the model was of the form:

$$\text{Daily MNU} = C_0 - C_1t - C_2d + C_3t^2 + C_4td + C_5d^2 + C_6t^3 + C_7t^2d + C_8td^2 + C_9d^3$$

Where t = maximum temperature, d = DoY and the C values, rounded to two significant figures, are the fitted coefficients: 35.13, -1.56, -2.03×10^{-1} , 2.84×10^{-2} , 7.06×10^{-3} , 2.87×10^{-4} , -1.46×10^{-4} , -2.85×10^{-5} , -1.26×10^{-5} and 3.90×10^{-7} respectively. The smaller coefficients are the result of the inputs being put to a high power and testing the ranges of the inputs found that they have a potential impact of at least 8L/Hr/SP.

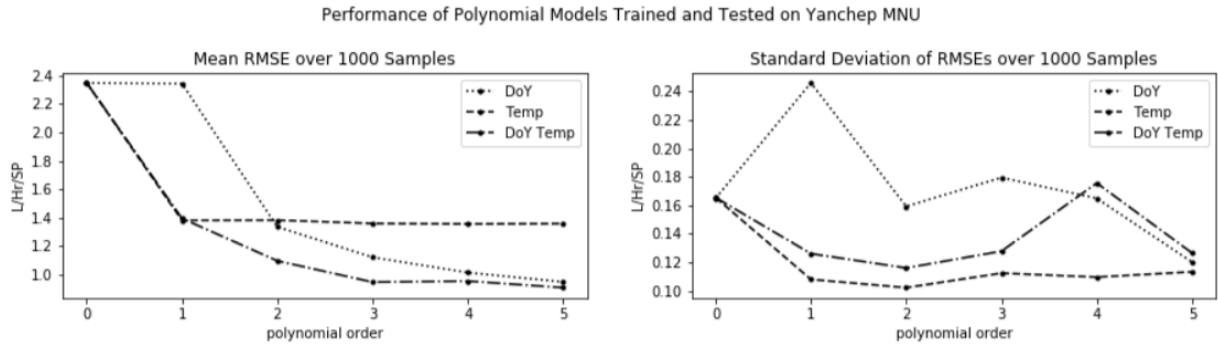


Figure 3 Average and variance of the tested polynomial models' performance.

3.3 Smart Meter Sample Population

The chosen polynomial was trained on sample sizes of 100 to 3700 in steps of 100 and tested on the full population. The mean and standard deviation of each sample size's RMSE was taken over 1000 instances of the full time series split into random sets training and testing weeks at a 3:1 ratio. These mean plot shows the average uncertainty of a model trained on a given sample size, and the standard deviation plot shows the variance between training samples. As expected, increasing the sample size gives better results with more reliability, but the derivative plots show diminishing returns at larger sample sizes. Based on this, improvements slow significantly beyond samples of 400 and 700 service points.

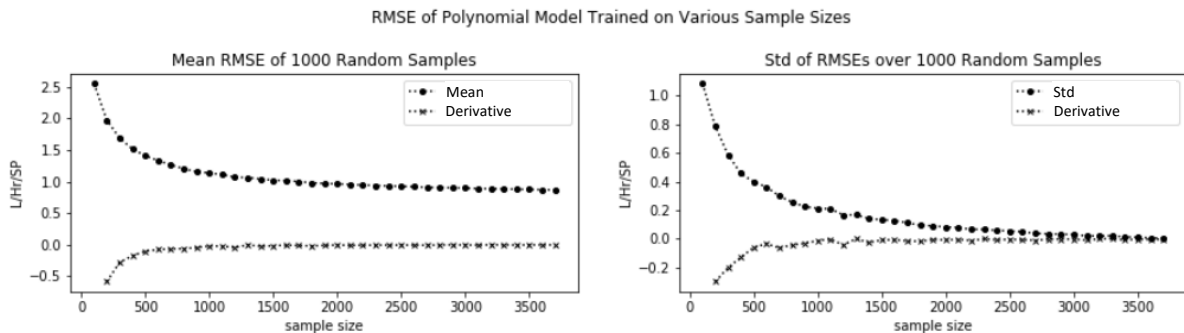


Figure 4 Average and variance of the max temp, DoY 3rd order polynomial model and their derivatives.

3.4 Uncertainty Comparison

Each model's performance can be evaluated by the uncertainty in their predictions of MNU as represented by their RMSE, which is an estimate of the models' errors' standard deviation. The size of these errors are put into perspective by comparison to figure 2, which shows that Yanchep's MNU has a range of 3.6 – 13.9L/Hr/SP across the measured period. Table 1 shows that the WSAA model performs well when restricted to winter, but not on a full year of data, as expected. The polynomial model can reduce the uncertainty to less than 1L/Hr/SP when trained on the full population and can produce better results than the baseline model for full year predictions when trained on a reduced training population.

Model	RMSE (L/Hr/SP)
WSAA Baseline fitted on full year	2.3196
WSAA Baseline fitted on Winter	0.6177
DoY, Temp Polynomial (3 rd order) (Full Population)	0.9098
DoY, Temp Polynomial (3 rd order) (Sample Size 400)	1.5137
DoY, Temp Polynomial (3 rd order) (Sample Size 700)	1.2651

Table 1 Root mean squared error of the baseline model and the polynomial model trained on different size samples of service points.

4. Conclusions and Future Work

It is possible to model the MNU of a DMA as a function of the date and maximum temperature to estimate the volume of leakage present in its distribution mains. Using smart metered data from the Yanchep DMA, a model was developed to predict MNU across the year with less uncertainty than the current method in use by the Water Corporation. This model can be trained on data from a sample of the DMA's population, reducing the potential cost of training for regions without smart meters. Future work will include using purposeful sampling of water user types, determined based on billing data, in place of random sampling as well as extending the model to account for factors like access to a private bore and evaporative air conditioning. A method of training the model in non-smart metered DMAs will be explored as well as its performance over different years, using MNF readings from non-smart metered DMAs. The leakage present in Yanchep may be estimated once a period of MNF readings are recorded.

5. Acknowledgements

I would like to thank the AIPM team from the Water Corporation for their help with the project, particularly Ian Gibb, Pete Lam and Christopher Bryant. Would like to especially thank my Water Corporation mentor, Joe Standring, for making this project possible and always making time to share his expert knowledge with me, and my supervisor Rachel Cardell-Oliver for keeping me on track and providing clarity through the sea of ideas and information.

6. References

- Standring, J. 2017, *Water Corporation Internal document #14638811*. Water Corporation.
 Standring, J. 2018, *Water Corporation Internal document #16676999*. Water Corporation.
 Wide Water Bay Corporation, Water Loss Research & Analysis Ltd 2011. Targeting Leakage Using Nightflow Measurements. Water Services Association of Australia.