

Fault Detection of Non-Residential Water Meters

S.E. Roberts^a and I.R. Monks^a

^a *City West Water, Footscray, Victoria, Australia*
Email: Stuart.Roberts@citywestwater.com.au

Abstract: City West Water (CWW) provides water services to the CBD and western suburbs of Melbourne, Australia. The water industry constantly faces the problem of non-revenue water, i.e. water lost for various reasons including bursts, leaks, and metering inaccuracy. Effective asset maintenance is a vital part of ensuring water companies can both reduce non-revenue water and continue to provide services to their customers.

To fairly bill customers for water consumption, a water meter is located at each customer's property to measure the volume of water being supplied from the water main to the property. However, these meters can read inaccurately over time and, if inaccurate, most of the time record less than the actual consumption. This results in under billing customers for the water they consume, contributing to CWW's non-revenue water loss. Despite the small number of non-residential (business) customers serviced by CWW (9% of customers), these customers account for over 50% of total water consumption each year. Therefore it is vital to ensure the meters servicing these customers are maintained correctly.

Most water meters are read only quarterly so consumption history data is quite sparse. Without removing a meter from service, or at least physically running tests at the property itself, it is uncertain whether or not these meters are performing with satisfactory accuracy. Previous replacement programs replace a meter after it has recorded up to a certain volume of water, but for meters that under record volumes, this level may never be reached and thus faulty meters remain in service indefinitely. For this reason, it is important to monitor these water meters by other means.

Unlike residential meters, there was no existing data from non-residential meter tests, so it was not possible to probabilistically determine which meter characteristics (model, age, size etc.) typically lead to faulty water meters. Instead MATLAB was used to develop an algorithm which was given the name Checkmate Lite. The purpose of Checkmate Lite is to assist CWW choose which meters servicing non-residential customers may need to be replaced. Checkmate Lite uses an algorithm that attempts to identify anomalies in a meter's consumption history. These meters are flagged by Checkmate Lite. The user can then interactively review the properties of these meters and their consumption histories to decide whether or not any of the meters should be replaced. With no non-residential meter test results, the current accuracy of the algorithm is unknown, hence it is being used to assist in decision making, and not make the final decision itself.

Keywords: *Water meter, non-revenue water, non-residential customer, sparse data*

1. INTRODUCTION

City West Water (CWW) provides water services to the CBD and western suburbs of Melbourne, Australia. CWW supplies around 400,000 customers, consisting of 35,000 non-residential (business) and 365,000 residential customers, representing approximately 1 million people. Although non-residential customers make up only 9% of CWW's customers by number, each year they account for over 50% of the total water usage.

A major problem facing the water industry is non-revenue water. Non-revenue water is water lost due to a number of causes including bursts and leaks in pipes, theft, and inaccurate water meters at customer properties. No revenue is received for water lost by these means. Customers are billed for water usage by taking readings, usually quarterly, from a water meter located at their property. From this, their consumption since the last billing period can be determined. The problem is that not all water meters record consumption accurately. Over time, meters generally begin to under record the amount of water being used meaning customers are being under charged for water consumption, and hence, these losses contribute to the non-revenue water problem.

CWW has responsibility to ensure water meters measure within $\pm 4\%$ of the actual consumption, in line with the National Metering Institute recommendation. Ideally, as a meter begins recording outside this range, it will be replaced by a new meter. However, without removing the meter from service or running tests in situ, it is not possible to determine that the water meter is performing inaccurately. Previous replacement programs replace a meter after it has recorded a certain volume of water, but for meters that under record volumes, this level may never be reached and thus faulty meters remain in service indefinitely.

There has been previous investigation at CWW by Blackhall (2013) into identifying faulty water meters for residential customers. Residential meters generally have a consistent consumption pattern. Using historical consumption data of meters that have been removed and tested it was possible to undertake statistical analysis to assist in identifying faulty meters based on meter characteristics (model, age, location etc.) and change of recorded consumption patterns.



Figure 1. A non-residential water meter

In the case of non-residential customers, there was no test data available on meters that have previously been removed from service that could be used. In addition to this, there was not enough consistency among customers nor information regarding the size of each consumer to enable the development of a benchmark for each business type or industry sector. In the absence of this data, it is not possible to build a model like the one used for residential meters and instead the meter fleet needs to be monitored by other means.

The main data available is the billing history relating to each non-residential meter which gave the consumption recorded by the meter at approximately three monthly intervals. This billing history can be used to construct time series data for each water meter's consumption. Within these series, there are some anomalies and these prompt CWW to further investigate the meter.

2. HIGH, LOW, AND SLOWING METERS

When reviewing a meter's consumption history, there are some anomalies that may indicate a problem with either the meter, or within the owner's property (e.g. a leak). The three different meter problem types to search for are referred to as 'high', 'low', and 'slowing', to reflect their respective characteristics.

A 'high' meter is defined to be a meter that has recently displayed a sharp rise in consumption. A sharp rise in a meter's consumption from one period to the next may indicate that there is a leak on the property. Conversely, a 'low' meter is defined to be a meter that has recently displayed a sharp drop in consumption. This could indicate that the meter has had a sudden internal failure, possibly by wearing out or due to an

obstruction inside the meter (such as a small stone) preventing the meter from recording correctly. Figure 2 displays the typical consumption patterns for a high meter and a low meter with quarterly reads.

A ‘slowing’ meter is a meter where the accuracy has deteriorated over time to the point where it is no longer considered accurate. This is the most difficult problem to test for given that $\pm 4\%$ is not a large error to be able to detect, especially given the amount of noise in most of the time series of consumption records.

Although a meter’s consumption pattern displays one or more of these anomalies, they may not necessarily be indicators of faults. For example, a business may not use any water for some time because the owner has taken a holiday. Such issues cannot be resolved by the analysis techniques presented, and would require follow up with the owner.

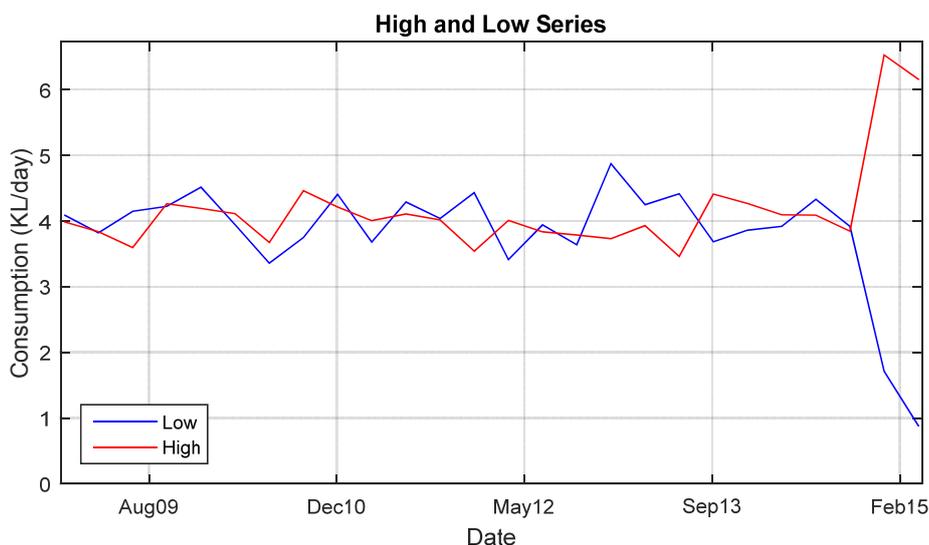


Figure 2. Time series plot of a high meter and low meter series.

3. THE ANOMALY DETECTION ALGORITHM

While the meter anomalies mentioned previously can be quite easy to spot by viewing the meter’s consumption history, with thousands of series it is impractical to do this manually. Using an automated algorithm to first shortlist meters avoids the need to review all meters manually.

3.1. Identifying High and Low Meters

High and low meters are identified by looking for significant changes from the meter’s past consumption to the final few consumption records. These can be identified by fitting a trend curve to the data and analysing residuals to determine whether or not the final observations deviate significantly from this trend curve in comparison to all other residuals. Given that the consumption data is rarely linear and often has outliers, it is important to use a method that can capture the changing trend and not be strongly influenced by outliers. The regression method chosen to do this is a locally weighted scatterplot smoothing (LOWESS) as outlined by Cleveland (2005).

The regression model \hat{X} is fitted to the consumption series X (KL/day) for each meter. The series X is found by dividing the total consumption for each reading by the number of days since the last reading. After fitting the model the residuals are calculated as $R_i = X_i - \hat{X}_i$, and then used to calculate upper and lower limits for the residuals. The bounds are calculated by

$$\text{Lower Bound} = Q1(R) - W_R * IQR(R) \text{ and, } \text{Upper Bound} = Q3(R) + W_R * IQR(R). \quad (1)$$

W_R is a constant, Q1 is the 25th percentile, Q3 is the 75th percentile, and IQR is the inter quartile range. Increasing W_R increases the allowable deviation from the fitted trend before a problem is reported. If either of the final two residuals fall above or below these bounds, the meter is reported to be either high or low. The final two observations are used so that only recent changes in consumption are reported.

3.2. Identifying Slowing Meters

Assume that upon installation a meter is recording 100% of the actual consumption (this will be referred to as accuracy (A)), and that this accuracy of the meter deteriorates over time. That is, $A(0) = 1$ and $A'(t) < 0$ for $t > 0$, where A is the accuracy of the meter and t is time in days. Consider the simple case, that a customer has some mean daily consumption, μ . The amount of this consumption measured by the meter for observation i is given by

$$x_i = \int_{T_{i-1}}^{T_i} \mu A(t) dt, \tag{2}$$

where T_i is the time the reading was taken from the meter for reading i ($T_0 = 0$) of N total readings. Converting this to an average daily consumption read gives

$$X_i = \frac{\int_{T_{i-1}}^{T_i} \mu A(t) dt}{T_i - T_{i-1}}. \tag{3}$$

Given the properties of A , this implies $X_i < X_{i-1}$. So by taking the average daily consumption of a slowing meter, a negative slope is expected in the trend of the data. Of course, this doesn't take into account the possibility that the average daily consumption is subject to change. It is possible that a change in ownership, business type, or business activity could significantly change the average consumption of the meter. If this change persists, the change in trend needs to be considered in different groups of the series, once before the change and once after the change. Figure 3 displays a pattern with this change in mean.

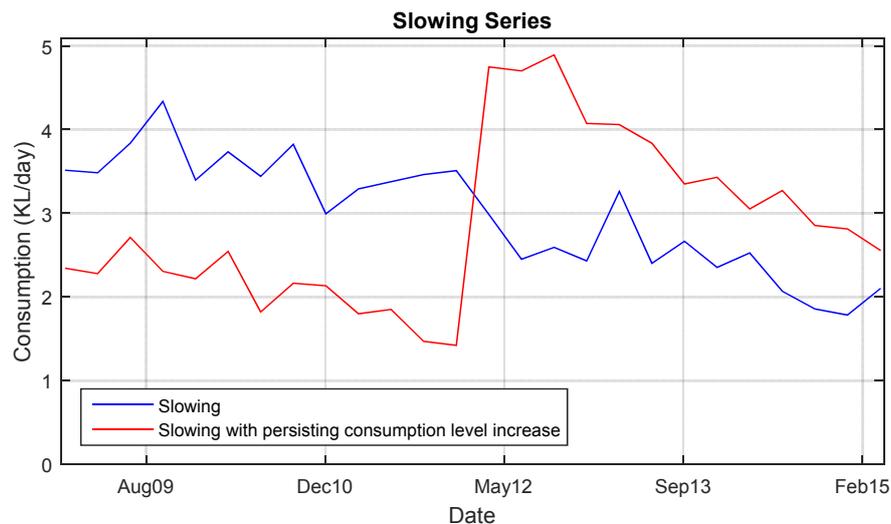


Figure 3. Time series plot of two slowing meters.

The subset, $\{X_{S_m}, X_{S_m+1}, \dots, X_{E_m}\}$, of the consumption series X will be treated as the m^{th} group when

$$\begin{aligned} |X_{S_m} - X_{S_m-1}| &> K, \\ |X_i - X_{i-1}| &< K \text{ for } i \in [S_m + 1, E_m] \text{ and,} \\ |X_{E_m+1} - X_{E_m}| &> K \end{aligned} \tag{4}$$

where K is a threshold value. S_m and E_m represent the start and end index respectively of section m . Let K depend on the distribution of the absolute differences at lag one of the series and $Y_i = |X_{i+1} - X_i|$. Then like the upper bound for the residuals in section 3.1, K will be chosen to be

$$K = Q3(Y) + W_L * IQR(Y), \tag{5}$$

where W_L is a constant, Q3 is the 75th percentile, and IQR is the inter quartile range. The greater the value of W_L the fewer sections the series is to be broken into. This process gives n groups as

$$\text{Group}_1 = \{X_{S_1}, X_{S_1+1}, \dots, X_{E_1}\}, \text{Group}_2 = \{X_{S_2}, X_{S_2+1}, \dots, X_{E_2}\}, \dots \text{Group}_n = \{X_{S_n}, X_{S_n+1}, \dots, X_{E_n}\}. \quad (6)$$

These groups will be ordered ($E_m < S_{m+1}$ etc.) with no overlapping. Now calculate the proportion of times that an observation is greater than a preceding observation within the same group. Calculate

$$\text{Drop Proportion} = \frac{\text{Count}}{\text{Total}}, \quad (7)$$

where

$$\text{Count} = \sum_{m=1}^n \sum_{i=S_m}^{E_m-1} \sum_{j=i+1}^{E_m} f(X_i, X_j) \quad (8)$$

and

$$\text{Total} = \frac{1}{2} \sum_{v=1}^n N_m(N_m - 1). \quad (9)$$

N_m is the number of observations in group m ($N_m = E_m - S_m + 1$). The function

$$f(X_i, X_j) = \begin{cases} 1 & \text{for } X_i > X_j \\ \frac{1}{2} & \text{for } X_i = X_j \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

is used to count the number of times an observation is greater than the preceding observations. The resulting test value (drop proportion) is a number between 0 and 1. The closer to 1, the more times the series decreased within each group. The meter will be flagged as a slowing meter if the test value is greater than some user defined threshold value, which should be greater than 0.5 to ensure meters with increasing trend are not reported.

4. USER INTERFACE

The algorithm has been embedded in an application, Checkmate Lite, which is used to assist the user in identifying potentially faulty water meters. Using MATLAB, a user interface was created allowing the user to; select the meter data to be analysed, set the algorithm parameters, run the algorithm, review the results, and select a shortlist of meters from the results for further investigation.

4.1. Importing Data and Setting Algorithm Parameters

Checkmate Lite's data importing function uses a Microsoft SQL database connection to import meter and consumption data filtered by the user search criteria. These criteria include the meter model, age, and industry type serviced by the meter. As an alternative the user can also import comma delimited text files containing the data.

The tests discussed in section 3 are dependent on a number of parameters. These parameters include

- the smoothing parameter used by the LOWESS regression in section 3.1,
- W_R to vary the lower and upper bound in section 3.1,
- W_L to vary the value of K in section 3.2, and
- "drop proportion" as a threshold value for the slowing meter test in section 3.2.

Currently the algorithm has not been tested in order to maximise its accuracy, and as such the parameters used by the algorithm have not been optimised. Because of this, the user has been given the option to change these parameters and modify the sensitivity of each test.

4.2. Running, Reviewing and Exporting Results

The program imports the data requested and runs it through the algorithm using the user’s parameters. Any meters that fail a test are flagged with the corresponding meter fault and returned in a list for the user to review. A boxplot displays the average daily consumption for every meter servicing the same business type as the current selected meter. For comparison a single dot representing the average daily consumption of the current meter is also plotted on the same graph. In the centre of the screen is a time series plot of the meter’s consumption history. This is the consumption data that was imported by the user. On the right hand side is an information box which displays information about the current meter including; CWW consumer id, meter name, meter model code, problem reported, average daily consumption, business type, install date, and age.

If the user identifies a meter that they want to investigate further, they can add the meter to a shortlist. Once the user has finished reviewing and selecting meters, the shortlisted meter names can be exported to an excel file together with the attributes for each meter.

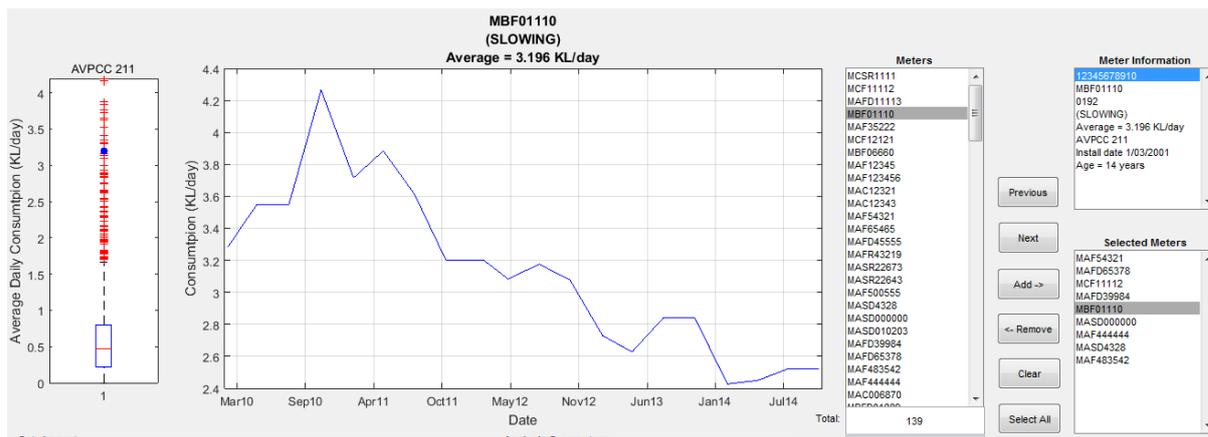


Figure 4. Checkmate Lite user interface displaying a meter found by the algorithm to be slowing.

5. OUTCOMES & LIMITATIONS

The development of Checkmate Lite has given CWW a rapid system to review the consumption history of its non-residential meters and flag any meters that may be recording consumptions outside of tolerance for further investigation. Checkmate Lite eliminates the need to manually review all non-residential meters by first shortlisting meters by identifying anomalies in their consumption history. This program is being used to assist CWW in deciding which meters to target for testing and/or replacement. Checkmate Lite is also being used by CWW as a way of identifying possible leaks at a non-residential property.

Currently the algorithm has not been calibrated or tested for accuracy using any statistical measures, since no non-residential meter test results are available. This is the reason that the program is being used to assist decision making, and not to make the final decision itself. When these test results do become available, the algorithm parameters can be calibrated or the algorithm could be used to directly calculate a probability of failure to maximise the accuracy of the program.

REFERENCES

Blackhall, K & Monks, I (2013). Checkmate: a new algorithm for detecting faulty water meters. Paper presented at Ozwater Conference, Ozwater, Perth, WA, May 24.

Cleveland, W.S. (2005). Robust locally weighted regression and smoothing scatterplot. *Journal of the American Statistical Association*. 74(368), 829- 836.