

https://dspace.org/

Decision enhancement for the management of apparent water losses

Ncube, Mthokozisi

2019

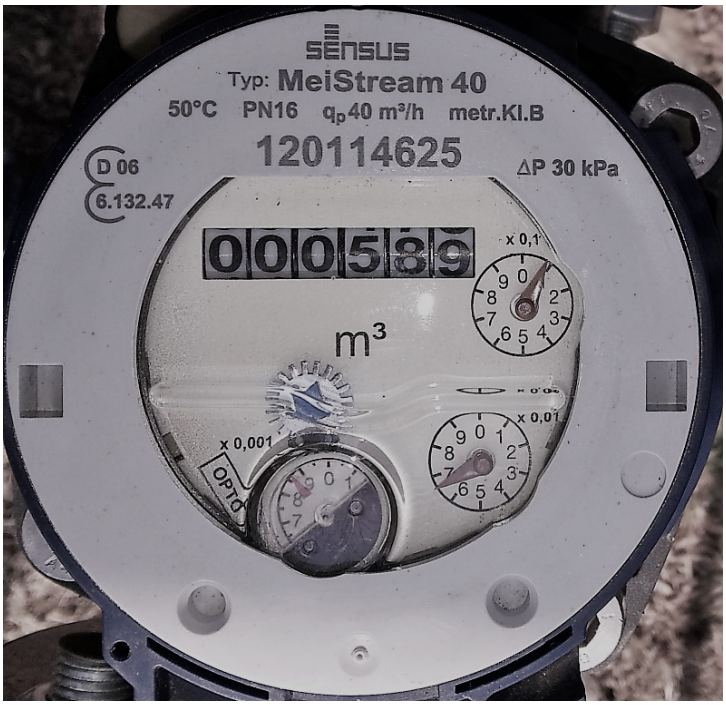
https://hdl.handle.net/10539/28791

Ncube, Mthokozisi, (2019) Decision enhancement for the management of apparent water losses, University of the Witwatersrand, Johannesburg, https://hdl.handle.net/10539/28791

Downloaded from a DSpace repository. https://dspace.org/

Decision Enhancement for the Management of Apparent Water Losses

Mthokozisi Ncube



Doctor of Philosophy:

A thesis submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg, May 2019

Dedicated to the next generation, who shall do even greater work

- Beracah Unathi and Benjamin Nkosinathi R. Ncube -

Declaration

I hereby declare that this thesis is my own unaided work. It is being submitted to the Degree of Doctor of Philosophy to the University of the Witwatersrand, Johannesburg. It has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University.



Mthokozisi Ncube

May 2019

Acknowledgements

This has been a long journey with the usual high and lows and numerous individuals have contributed toward the preservation of my sanity, perseverance and joy throughout the years I have been working on this research work. The following is by no means meant to be an exhaustive list of all individuals who have contributed but rather some of those whose support has been the most impactful and I am indebted to.

I am so grateful to my previous employer, Johannesburg Water for the opportunity, the resources and the support during my tenure in carrying out research work that gave birth to and carried a significant portion of this work. Special thanks go to the many former colleagues for the support and providing input – in particular, Mbali Matiwane, with her Physical Loss team and Fahlani Sibanda for agreeing to be sounding boards at various stages during and after my tenure with Johannesburg Water. Special mention goes to Ntshavheni Mukwevho and Enoc Mudau, among others, for their encouragement at various stages.

I am thankful for the support from my wife and kids and their patience when I had to take some of their time and devote it towards my studies. And the many hours they were my driving partners in and around Johannesburg. On a related matter, I will like to appreciate the numerous times Fahlani and family looked after the kids, giving me the much-needed time to focus on this work. Special mention goes to my mother, and late father, for the sacrifices, foundation and support that set up my academic path from early days and for always being supportive and continually lifting me up in prayer.

Prof Taigbenu, my supervisor, is also gratefully acknowledged for always being available, supportive and encouraging through the years, even before these PhD studies. His invaluable guidance has been instrumental in shaping this research output and in keeping me inspired to the finish line.

Justine Brown, previously with Sensus Metering South Africa and currently with JOAT, was very helpful and supportive in helping me find appropriate meters for flow logging and providing additional test meters, smart metering data and associated utilities. The numerous discussions and support during the early years of my studies are greatly appreciated. I will also like to appreciate the support of Mark Sheperd and the JOAT Group for the grant funding to the School of Civil and Environmental that enabled attendance and presentation of part of the work at an international conference.

Several Wits BSc Civil Engineering final year students, particularly Nobuhle Matsebula, Rui Mutemba, Muzi Nkosi and Stanley Ndlovu were also particularly helpful in meter logging of large consumers as part of their investigational projects. Thank you for lessening the load and mileage, and the company on some of the lonely days!

I will also like to appreciate the contribution of a few colleagues and friends that include Nicholas Moult for meaningful pointers and discussions on meter testing data; Arlvin Moyo for getting me started on SQL Server and the several afternoons he sacrificed in getting me up to speed on the language; Zachariah Katambara and Paul Kagoda for the many manuscripts they reviewed and for always being good encouraging friends throughout the period; Meghan Brinkey for the management of contractors for the fieldwork and associated assistance, even beyond the call of duty, and for introducing me to Pieter Crous whose valuable insights helped me refine my code, thoughts, processes and some manuscripts.

To my siblings, broader family, many friends and colleagues who have had to listen to my thinking-out-loud, my frustrations and who have provided some form of support and insight in one form or the other, I am so very grateful. God bless you!

Over all, "… to Him who is able to do exceedingly abundantly above all that we ask or think, according to the power that works in us, to Him be glory in the church by Christ Jesus to all generations, forever and ever. Amen".

Abstract

Non-revenue water (NRW) is a prevalent and pervasive problem in both the developed and developing nations of the world. Coupled with increasing water demands that are outstripping available supply, the need to significantly improve water use efficiency and water loss management has never been more urgent. Of the two main components of water losses, real (physical) losses have enjoyed greater attention in both research and practice that apparent (commercial) losses despite that latter potentially being more impactful as it is measured in terms of the retail cost of water.

Research in apparent water loss has shown that metering errors are the leading contributor but also that traditional assessment methods are costly, laborious and out of reach for many utilities. There is also limited adoption of data-driven discovery processes to glean insightful trends from existing data to inform improved apparent loss management. Existing management tools are also fragmented and do not provide for the integrating assessment of the various component of apparent water losses. There is therefore a clear need to improve the status quo in both the assessment methodologies and the management tools for the improved and integrated assessment and management of apparent water losses, particularly within the context of the developing world.

This thesis contributes towards the quantification and the management of apparent losses, with the City of Johannesburg, South Africa as the study area. Representative consumption profiles for the city were determined using results of the meter logging of 408 properties in Johannesburg and additional third-party data for 256 properties from various areas within South Africa. Residential consumers have up to 37% of flow occurring at less than 0.03m^3/hr, indicating a substantial proportion of low flow rates that might not be adequately measured by water meters. A 63% occurrence of on-site leakage with a significant monthly leakage of 11 – 41m^3/month/property was found in the Johannesburg.

Meter testing at a SANAS accredited flow laboratory was successfully done for 123 small diameter meters at 10 test flowrates to better represent the meter curve and characterise the meter accuracy of Johannesburg water meters. This was supplemented by the analysis of over 3,000 historical of meter testing records to determine the meter accuracy degradation rates at various flowrates. The starting flowrate showed the expected highest degradation rate of up to 0.25% per every 1,000kℓ and 0.32% per year of meter age, with other flow rates exhibiting lower degradation rates.

Apparent water losses within Johannesburg were evaluated using three different methodologies, two of which were alternative methods that had not been previously validated against the best practice weighted accuracy method. Using the weighted meter accuracy method, apparent losses due to meter inaccuracy estimated to range from 9.4% to 14.6%, on average, depending on the actual ratio of meter sizes used. The comparative billing analysis method estimate was closer to the worst-case scenario at 14%, while the meter change analysis method understated the losses at 4.7% for a subset of the data. Both alternative methods were found to be unreliable in their current form for the estimation of apparent losses although they do provide some insight into the meter management practices of a utility. A proposal is made on how the weighted error methodology can be also modified and enhanced to better quantify errors at low flows. The annual apparent losses were found to be location specific and reliant on prevailing meter regulation and therefore the default reference annual apparent losses of 5% will need to account for this. Proposals on how this can be done are included in the research.

An improved comparative billing methodology was developed using a combination of data preparation techniques, clustering analysis and classical regression analysis. Consumption was found to be reducing on at a rate of 0.7% per annum, similar to the weighted error degradation rate found in other studies using the weighted accuracy methodology. Apparent losses due to metering error was estimated to be 8.1% of billed consumption for the residential sector against 11.2% estimate from the weighted accuracy method for the same sector. The difference is fully accounted by meter under-registration, shown to be higher than 4% for new meters, which the developed methodology does

not account for. A general relationship is also proposed for similar utilities to estimate their apparent losses using the results of the developed method.

Based on the results of this research, a simplified meter management framework is proposed and is made up of four main components namely Systems, Audits, Analysis and Metering (SAAM). The main objective of the framework is to simplify the main requirements for achieving improved assessment and management of apparent water losses. Based on the SAAM meter management framework, an Apparent Loss Management Tool (ALMAT) was also developed to enable utilities to perform, among other things, meter audit analysis to derive consumption profiles; meter error analysis; optimum meter replacement period determination; and to assess apparent losses and meter accuracy degradation analysis using the weighted accuracy method and the improved comparative billing analysis method developed in this study.

Keywords

Apparent water loss, water metering, water meter management, non-revenue water, water demand management, decision support

Contents

Declaration ..... **v**
Acknowledgements ..... **vii**
Abstract ..... **ix**
Contents ..... **xiii**
List of Figures ..... **xviii**
List of Tables ..... **xxi**
List of Equations ..... **xxii**
Chapter 1 Introduction **1**
1.1 Apparent Water Losses 4
1.1.1 Water Meter Errors 4
1.1.2 Data Acquisition Errors 7
1.1.3 Unauthorised Consumption 7
1.1.4 Errors on Estimates of Unmetered Consumption 8
1.2 Water Supply in South Africa 8
1.2.1 Johannesburg Water 9
1.3 Need for this Research 11
1.4 Objectives of the Research 12
1.5 Scope of the Research 13
1.6 Outline of the Thesis 14
Chapter 2 Review of Methods **17**
2.1 Water Loss Management 17
2.1.1 Top-down Assessment 18
2.1.2 Bottom-Up Assessment 19
2.1.3 Hydraulic Network Modelling Assessment 20
2.1.4 Statistical Flow Analysis 21
2.2 Apparent Water Loss Assessment 21
2.2.1 Weighted Meter Accuracy 27
2.2.2 Meter Testing 29

2.2.3 Consumption Characterisation 32
2.2.4 Alternative Assessment Methods 35
2.3 Integrated Meter Management 39
2.3.1 Meter Selection and Replacement 40
2.3.2 Meter Sizing and Maintenance 42
2.3.3 Metrology Regulation and Standards 43
2.4 Data Mining 46
2.4.1 Anomaly and Fraud Detection 48
2.4.2 Time Series Similarity Search and Pattern Discovery 48
2.4.3 Clustering and Classification 50
2.5 Data Mining Techniques 51
2.5.1 Artificial Neural Networks (ANN) 51
2.5.2 Support Vector Machine (SVM) 51
2.5.3 Self-Organising Maps (SOM) 52
2.5.4 Dynamic Time Warping (DTW) 53
2.5.5 k-Nearest Neighbour (kNN) 53
2.5.6 Hidden Markov Model (HMM) 54
2.5.7 k-Shape Clustering 54
2.5.8 Fuzzy Clustering 55
2.6 Summary 55
Chapter 3 Consumption Characterisation 59
3.1 Preparatory Work 59
3.1.1 Consumer Sampling 59
3.1.2 Logging Equipment Selection 62
3.1.3 Logistics Planning 64
3.2 Site Activities 64
3.2.1 Bulk Meter Logging 65
3.2.2 Small Meter Logging 67
3.3 Smart Metering Consumption Data 68
3.4 Consumption Patterns 71
3.4.1 Bulk Consumers 71
3.4.2 Small Connections 75
3.5 Onsite Leakage 78
3.5.1 Bulk Consumers 78

3.5.2 Small Connections 78
3.6 Summary 81
Chapter 4 Water Meter Testing 85
4.1 Historical Meter Test Data 85
4.1.1 Likelihood of Meter Failure 88
4.1.2 Meter Degradation 90
4.2 Controlled Meter Testing 93
4.2.1 Preparatory Work 93
4.2.2 Results 96
4.3 Summary 101
Chapter 5 Apparent Water Loss Assessment 105
5.1 Traditional Assessment Method 105
5.1.1 Weighted Meter Accuracy 106
5.1.2 Reference Annual Apparent Losses (RAAL) 111
5.1.3 Unauthorised Consumption 113
5.1.4 Data Acquisition Errors and Unmetered Consumption Errors 114
5.2 Meter Change Method 115
5.2.1 Labelled Dataset 115
5.2.2 Complete Consumption Dataset (Unlabelled) 118
5.2.3 Key Observations 119
5.3 Comparative Billing Assessment Method 123
5.3.1 Data Preparation 123
5.3.2 Analysis and Results 124
5.4 Summary 127
Chapter 6 Comparative Billing Assessment 2.0 129
6.1 Data Preparation 131
6.1.1 Database Creation 131
6.1.2 Data Cleansing 134
6.1.3 Final Dataset 136
6.2 Clustering Analysis 140
6.2.1 Preliminary Clustering Analysis 141
6.2.2 Final Clustering Analysis 143
6.3 Accuracy Degradation Analysis 146
6.3.1 Exploration of the Dataset 146

6.3.2 Error Degradation 148
6.4 Method Validation 153
6.5 Generalised Equation for AL Estimation 155
6.6 Summary 156
Chapter 7 Decision Enhancement for the Management of Apparent Losses 159
7.1 Meter Management Framework 160
7.2 Systems for Metering and Consumption Information 163
7.2.1 Consumption Information 164
7.2.2 Meter Information 165
7.2.3 Meter Maintenance 167
7.3 Consumers and Meter Audits 167
7.3.1 Meter Audits 168
7.3.2 Consumers Audit 171
7.3.3 Meter Verification 173
7.4 Analysis and Assessments 175
7.4.1 Meter Error Analysis 176
7.4.2 Cost Benefit Analysis and Optimum Replacement Period 180
7.4.3 Apparent Loss Assessment 183
7.5 Meter Management 190
7.5.1 Meter Selection 191
7.5.2 Meter Installation 192
7.5.3 Meter Maintenance 194
7.5.4 Quality Control and Assurance 195
7.6 Summary 198
Chapter 8 Conclusion and Recommendations 201
8.1 The Context 201
8.2 Achieved Results 202
8.2.1 Consumption Characterisation 202
8.2.2 Meter Accuracy Evaluation 203
8.2.3 Apparent Loss Assessment and Method Comparison 204
8.2.4 Development of an Alternative Apparent Loss Assessment Method 205
8.2.5 Apparent Loss Management Tool 205
8.3 Contribution to Knowledge 206
8.4 Recommendations and Future Research Needs 208

References **211**

Appendix A **Meter Test Results Sample Data** **A-1**

Appendix B **Comparative Billing Analysis Code** **B-1**

List of Figures

Figure 2-1: Real Loss Components 18
Figure 2-2: Standard IWA Water Balance.....19
Figure 2-3: Apparent Loss Components 22
Figure 2-4: Methodology for determining weighted meter accuracy.....28
Figure 2-5: Meter Error Curve 31
Figure 2-6: Consumption Patterns for various areas.....34
Figure 2-7: Nine Steps towards Integrated Meter Management.....40
Figure 2-8: Illustration of Difference in Meter Classes (Q_p = 3,500ℓ/h) 46
Figure 3-1: Meter Reading Patterns 60
Figure 3-2: Location of Sample Points for Consumption Characterisation .61
Figure 3-3: iPerl Meter and its error curve 63
Figure 3-4: Meter Logging Setup for Large Meters 65
Figure 3-5: Examples of meters that could not be logged.....66
Figure 3-6: Meter Logging for Small Meters 67
Figure 3-7: Consumption Pattern: Smart Metering Data 70
Figure 3-8: Examples of Bulk Meter Logging Results 72
Figure 3-9: Consumption Distribution for Large Connections.....73
Figure 3-10: Variation of Flowrate with Units in Residential Complexes 74
Figure 3-11: Consumption Distribution for Small Connections.....76
Figure 3-12: Residential Consumers Consumption Distribution by AADD .77
Figure 3-13: Distribution of night time consumption 80
Figure 3-14: Distribution of the ratio of night flow to total flow.....81
Figure 4-1: Distribution of Meter Test Data. 87
Figure 4-2: Probability of Failure with Age - Volumetric Meters 88
Figure 4-3: Probability of failure with Meter Reading - Volumetric Meters..89
Figure 4-4: Accuracy degradation at q_{min} with meter reading 91
Figure 4-5: Accuracy degradation at q_{min} with meter age 92
Figure 4-6: Location of Sample Point for Meter Testing.....95
Figure 4-7: Meter Testing Results Extract 96
Figure 4-8: Model 2 – 15mm Test Results.....98
Figure 4-9: Q1 Error Degradation – Volumetric Meters.....98
Figure 4-10: Average Error per Meter Model and Size 100

Figure 5-1: Weighted Meter Accuracy – 15mm Meters.....109
Figure 5-2: Weighted Meter Accuracy – 20mm Meters.....109
Figure 5-3: Error Curve for a random 15mm meter.....112
Figure 5-4: Historical Record of Unauthorised Connections 114
Figure 5-5: Distribution of meter change data.....117
Figure 5-6: Error degradation with meter reading - Meter Change.....117
Figure 5-7: Distribution of meter reading data – Meter Change Data.....119
Figure 5-8: Histogram of the Impact of Meter Change 121
Figure 5-9: Days to Fix Meter Related Jobs 122
Figure 5-10: Meter Reading Data Distribution – Comparative Billing 124
Figure 5-11: Annual Consumption Variation with Meter Age.....125
Figure 5-12: Annual Consumption Variation with Meter Age.....126
Figure 6-1: Data Import within SSIS – Pre 2010 Data.....132
Figure 6-2: Data Import within SSIS – Post 2010 Data 133
Figure 6-3: Staging Database Fields 133
Figure 6-4: Consumption Time Series 133
Figure 6-5: Lookup Table Examples.....135
Figure 6-6: Average Running Annual Consumption.....137
Figure 6-7: Start Reading Distribution.....138
Figure 6-8: Average Monthly Consumption Distribution.....139
Figure 6-9: k-Shape Clusters.....144
Figure 6-10: Fuzzy Clusters 145
Figure 6-11: Running Average Annual Consumption per Category 147
Figure 6-12: Cluster Average Consumption.....148
Figure 6-13: Increasing Consumption Cluster.....149
Figure 6-14: Decreasing Consumption Cluster 149
Figure 7-1: The Meter Management Process 161
Figure 7-2: Proposed Meter Management Framework 162
Figure 7-3: Introduction to the Management Tool 163
Figure 7-4: Meter Logging Result 168
Figure 7-5: Introduction to Meter Audit 169
Figure 7-6: Consumption Profile and Meter Sizing Guideline.....169
Figure 7-7: Night flow Analysis and Meter Sizing Implications 171
Figure 7-8: Introduction to Consumers Audit 172

Figure 7-9: Consumption Characterisation 173
Figure 7-10: Meter Error Analysis Introduction 176
Figure 7-11: Meter Error Preliminary Analysis 177
Figure 7-12: Probability of Failure Analysis 178
Figure 7-13: Probability of Failure with Meter Reading 178
Figure 7-14: Degradation of Error Analysis – Meter Type 179
Figure 7-15: Degradation of Error Analysis – Meter Model 180
Figure 7-16: Cost-Benefit Analysis Input Screen 181
Figure 7-17: Optimum Replacement Period and CBA 182
Figure 7-18: Degradation Rate Sensitivity Analysis 183
Figure 7-19: Apparent Loss Assessment Introduction 184
Figure 7-20: Weighted Accuracy Estimates 185
Figure 7-21: Degradation of Accuracy - Meter Type 185
Figure 7-22: Degradation of Accuracy - Meter Model 186
Figure 7-23: Data Preparation 188
Figure 7-24: Preliminary Clustering 188
Figure 7-25: Final Clustering Analysis 189
Figure 7-26: Error Degradation Analysis 190
Figure 7-27: Error Degradation Results 190
Figure 7-28: Meter Selection Input 191
Figure 7-29: Meter Selection Recommendations 191
Figure 7-30: Meter Quality Assurance Process 197
Figure 7-31: Job Card Extract - Meter QA 198

List of Tables

Table 2.1: International Apparent Loss Estimates References 24
Table 2.2: Meter and Data Acquisition Error Estimates 25
Table 2.3: Unauthorised Consumption Estimates.....25
Table 2.4: Proposed Test Flow Rates for Different Permanent Flowrates..32
Table 2.5: Metrological Class of Meters (SANS 1529).....45
Table 2.6: Permissible Tolerance on Indication (SANS 1529) 45
Table 3.1: Water Consumption Sample Selection 60
Table 3.2: Successful Meter Logging Summary 65
Table 3.3: Smart Metering Data.....69
Table 3.4: Night flow average statistics for small connections 79
Table 3.5: Consumption Profiles.....82
Table 4.1: Summary of Meter Testing Data. 86
Table 4.2: Stratification of Meters 86
Table 4.3: Error Degradation Rates. 92
Table 4.4: Stratified Sampling for Small Meters.....94
Table 4.5: Test flow rates 96
Table 4.6: Meters Test Results.....99
Table 5.1: Average Meter Accuracy 106
Table 5.2: Consumption Profile for 15mm 107
Table 5.3: Consumption Profile for 20mm 107
Table 5.4: Weighted Meter Accuracy.....108
Table 5.5: Estimated Meter Error.....110
Table 5.6: Meter changes per property.....116
Table 5.7: Annual Volume Degradation with Age 125
Table 6.1: Surviving Dataset 139
Table 6.2: Cluster Validity Indices Output Example 143
Table 6.3: Calculation of Degradation Rate 150
Table 6.4: Degradation Rates.....151
Table 6.5: Estimated Apparent Water Losses.....153

List of Equations

Equation 2.1: Alternative Apparent Losses Estimates 26
Equation 2.2: Meter Error Degradation (Age or Volume) 27
Equation 2.3: Meter Error Degradation (Age and Volume).....27
Equation 2.4: Weighted Meter Accuracy.....28
Equation 2.5: AL from the Reverse Approach.....36
Equation 2.6: Metering Error 37
Equation 2.7: Variation of Billed Volume.....38
Equation 2.8: Meter Error Degradation 38
Equation 2.9: Meter Selection Minimisation Function 40
Equation 2.10: Real Discount Rate.....41
Equation 2.11: NPV for Meter Selection with Fixed Replacements.....41
Equation 2.12: NPVC for Meter Replacement and Selection 41
Equation 2.13: Artificial Neural Network Model Function 51
Equation 2.14: Support Vector Machine Function.....52
Equation 2.15: Self-Organising Maps Function.....52
Equation 2.16: Dynamic Time Warping Distance.....53
Equation 2.17: k-Nearest Neighbour Function 53
Equation 2.18: Shape Based Distance – k-Shape Clustering 54
Equation 2.19: Fuzzy Clustering Centroid Function 55
Equation 4.1: Probability of Failure.....88
Equation 4.2: Joint Probability of Failure 90
Equation 4.3: Modified Weighted Meter Accuracy 101
Equation 6.1: Apparent Loss Estimate.....155

Chapter 1 Introduction

The World Economic Forum considers water crises, defined as a significant decline in the available quality and quantity of fresh water resulting in negative impacts on human health and/or socio-economic activities, as one of the top 10 global risks (World Economic Forum, 2017). This is against a background of increasing freshwater withdrawals globally that have grown by about 1% per year since the 1980s, mainly due to growing demand in developing countries (WWAP, 2016). To mitigate against water risks, water-use efficiency improvements are critical, particularly in addressing the projected 40% gap between demand and supply by 2030 if no improvement in the efficiency of water use is achieved (UNEP, 2011). The management and reduction of non-revenue water (NRW) is one of the many components of improving water use efficiency and reducing current and future gaps between water supply and demand, particularly within the potable water supply environment. This becomes particularly relevant as reducing water losses costs less than three percent of adding the equivalent in new production capacity and can be accomplished significantly faster (Dobbs et al., 2013).

Despite this water use efficiency potential, water loss control remains a major challenge for water utilities worldwide, which in turn limits the sustainability and promotion of efficient use of water as a finite natural resource (Loureiro et al. 2014). Reducing municipal water losses is therefore seen as a key opportunity in many areas that can unlock both resource and financial benefits (Green Cape, 2017). Kingdom et al. (2006) estimates that water losses cost utilities a minimum of USD141 billion per year, with a third of it occurring in the developing world. The amount lost in developing countries, constitutes about 45Mm^3/d in water distribution networks – enough to serve nearly 200 million people, which is very concerning particularly in areas where populations remain unserved with safely managed water. Recent estimates show that London has an estimated 25% of non-revenue water

(NRW) while Norway has 32%, clearly showing that the challenge is not isolated to the developing world (WWAP, 2016). In South Africa, recent estimates (2015/16) puts non-revenue water (NRW) at 41%, an increase from 34.6% in the 2013/14 assessment (Department of Water and Sanitation, 2017). This increase is not only attributed to system degradation, but to improved participation and quantification of water losses by rural municipalities who previously did not fully participate in these non-revenue water assessments. Despite the increased participation, only 107 water services authorities out of a total of 152 participated. The water loss management challenges have been persistent in South Africa with the President of the Republic of South Africa declaring in the 2010 State of the Nation Address that measures had been put in place to halve water losses by 2014, which has not been achieved to date, four years later than the anticipated date. The second edition of the National Water Resources Strategy (NWRS) (Department of Water Affairs, 2013) concedes that despite water conservation and water demand management (WCDM) being a priority and critical in ensuring water security for South Africa, it has not been effectively implemented to date.

The above statistics, albeit negative, have been facilitated by the considerable progress that has been made since the mid-1990s in the assessment and benchmarking of water losses (McKenzie & Seago, 2005). Key to this progress is the adoption of the International Water Association (IWA) water balance as best practice that is largely followed in many parts of the world (Lambert & Hirner, 2000). Despite this wide acceptance of the IWA water balance, a review of its application shows that very few utilities produce complete water balances owing to the effort involved and the lack of adequate and efficient methods for determining the various components (Klingel & Knobloch, 2015). Mutikanga (2012) also notes that most of the components of the water balance are estimated using ill-defined techniques that are open to interpretation and manipulation by practitioners. Estimates of water balance components can therefore be subjective and difficult to reproduce or compare with other utilities. Coupled with only a 70% utility water loss assessment participation in South African, this demonstrates that the ability to quantify, and therefore manage, water losses remains a challenge. The solutions to such challenges could potentially include the streamlining of existing methods and the development of new methods for enhanced water loss management.

Another important factor with the use of the IWA water balance is that there is a general estimation of apparent losses (not quantification or assessment) as a percentage of non-revenue water such as 20% of water losses in South Africa (McKenzie, Siqalaba & Wegelin, 2012) and a proposed international benchmark of 5% of water sales as an acceptable level of apparent losses (Rizzo et al., 2007). McIntosh (2003) estimates that 50 – 60% of NRW in Asian water utilities are due to apparent losses while Browne et al. (2013) estimates that worldwide utilities lose USD6.5billion/annum to apparent losses. Such estimates could be based on inappropriate assumptions, particularly in a number of contexts such as developing countries (AL-Washali, Sharma & Kennedy, 2016). This exposes water utilities to choosing water loss interventions that may not be necessarily optimum as apparent losses are largely unknown and are estimated to be as high as 80% of the water loss in some areas and as low as 5% in others (McKenzie, Siqalaba & Wegelin, 2012; Mutikanga, Sharma & Vairavamoorthy, 2013). This is particularly important considering that, although apparent losses are normally reported in terms of volume, they are the most significant financially as they are calculated at the retail price charged to the consumer unlike real losses which are calculated at the variable production costs (AWWA, 2009).

Related to the general estimation of apparent losses in practice, academic research and publications in water loss management have largely focused on the real loss component of water losses (Johnson, 2007; Liemberger et al., 2007; McKenzie and Seago, 2005; Mutikanga et al., 2011). This gap is also echoed by the IWA Water Loss Task Force tasked with narrowing the gap in the assessment of apparent losses (Johnson & Vermersch, 2011) through the development of appropriate guidelines. Vermersch et al. (2016) is the proposed final guideline document which remains rather cumbersome to follow and appears to be more of a compilation of various papers by the contributing authors and team members of the task force over the years that what would be expected of an informative guideline.

1.1 Apparent Water Losses

Vermersch et al. (2016) summarises the guidelines of the Apparent Loss Initiative of the IWA Water Loss Specialist Group (WLSG) for the reduction of apparent losses. The main interacting components of apparent losses have been updated to comprise metering errors, unauthorised consumption, data acquisition errors and errors in the estimation of unmetered consumption. From a management and key performance indicator perspective, key indicators include;

- Current annual apparent losses (CAAL), the estimated value of apparent losses over a period of twelve months for the water supply system being evaluated.
- Reference annual apparent losses (RAAL), an annual estimate of a reference target of apparent losses for a water supply system. The reference annual apparent losses (RAAL) default value is assumed to be 5% of billed authorized consumption and forms the basis on evaluating the performance of any water supply system.
- Economic apparent loss level of (EALL) is the level at which the financial savings gained from implementing a specific apparent loss project equals the financial expenses incurred for the same project.
- The apparent loss index (ALI) which is a ratio of CAAL and RAAL.

The primary objective of apparent water loss management is to reduce CAAL to either the EALL or the RAAL by addressing either of the four components of metering errors, unauthorised consumption, data acquisition errors and errors unmetered consumption. Such exercises do require some rational estimation of either of the errors for any intervention to be properly designed.

1.1.1 Water Meter Errors

The majority of the work done on apparent losses has mostly focused on metering errors (Allender, 1996; Arregui et al., 2010; Mutikanga, Sharma & Vairavamoorthy, 2011b,a; Arregui, Soriano &

Gavara, 2012; Szilveszter, Beltran & Fuentes, 2015) which have been found to be the highest contributor to apparent losses (Rizzo & Cilia, 2005). This is because mechanical meters, which are predominantly and historically in use, experience reduced efficiency with age and usage due to wear and tear, incorrect installation, water quality, incorrect meter sizing, demand profile, among other factors. Most studies determined the weighted meter accuracy using the methodology outlined in Noss et al. (1987), Yee (1999) and Arregui et al. (2006) that involves the determination of metrological performance of a meter at different flow rates from meter testing and comparing it against the user demand temporal pattern from consumer audits though meter logging (Criminisi et al., 2009; Johnson, 2012). Meter testing and meter logging both involve varying levels of field and laboratory activities that are costly and require extensive amounts of time to be done properly.

The error rate decay in most studies using the weighted meter accuracy methodology employ sampling techniques and regression analysis to establish the relationship between meter accuracy degradation rate and the meter age and/or meter usage i.e. the totalised volume recorded through the meter (Mutikanga, Sharma & Vairavamoorthy, 2013). Criminisi et al. (2009) and Mutikanga et al. (2011) also looked at the influence of private water tanks on meter errors as the addition of ball valves in the tanks increases low flow rates across the meter, in which region the meters are the least accurate. Yaniv, Freni & Fantozzi (2005) Barfuss, Johnson & Neilsen (2011) and Shields, (2011) investigated the performance of meters at very low flows and high flows and the use of fittings to reduce the accuracy loss at ultra-low flows as meter performance is poor at these flow rates.

Invariably, all metering error studies consist of a combination of field and laboratory work to determine consumption profiles and actual meter errors to estimate apparent losses. This is a challenging undertaking, also noted in Arregui *et al.* (2009), for a variety of reasons that include;

- Meter testing requires that the meters be removed from site and tested at a reputable, if not accredited, flow laboratory. There are very few flow labs accredited by the South African National Accreditation System (SANAS). Meter testing involves the removal and replacement of the meter to be tested, often permanently, incurring some form of capital expenditure for meter replacement in addition to the operational expense of testing the meters. Furthermore, to be able to reproduce

the error curve adequately, the number of test flow rates required increases implying that the testing time and cost also increases. It is also not guaranteed that meter testing will be successful for all meters as they might be damaged in the process.

- Meter logging requires site activities that involve the inspection of meters and the installation of meter logging equipment to allow for consumer logging over a reasonable period such as seven days. Some consumer meters may not have the requisite pulse output, or similar form of signal, to connect the logger and therefore it might be necessary to temporary install a meter with such functionality (Ncube & Taigbenu, 2016), which further increases the cost.

- Both activities require meaningful and representative samples for the results to be applicable to the utility. And because metering databases are known to be inaccurate in some instances, there is bound to be some repetitive work in either case as documented in Ncube and Taigbenu (2016). This further drives the costs and time of implementing the activities of estimating the weighted meter error.

The methodology is therefore costly and time consuming which in most studies results in very few meter samples being taken (Couvelis, 2012) that do not yield good estimates of apparent losses. Despite these challenges, to date, field-based assessment of apparent losses remains the gold standard of estimating these losses as no credible alternative methods have been successfully developed, validated and implemented widely. The potential of alternative methodologies that can leverage on existing information without the disadvantages of the traditional laboratory and field work could therefore be invaluable to the water sector. A few attempts in this regard exist and they include Couvelis and van Zyl (2015) who used a meter replacement database to derive the meter degradation rate from the accuracy difference between the new meter and old meter, and Mbabazi *et al.* (2015) who estimated meter accuracy degradation by evaluating the evolution of the average yearly metered volume against meter age. Both methods are yet to be validated and their accuracies remain untested.

1.1.2 Data Acquisition Errors

This component of water loss is seen as directly related to water utility management and may be reduced by improving company procedures (Criminisi et al., 2009). This therefore suggests that well managed utilities will likely not have much of a problem with this component of apparent losses and in cases where it exists, the solution resides with management. Local experience in South Africa shows that this management deficiency is prevalent with several municipalities and cities being in the press in recent years for various billing related challenges. This phenomenon is described in Mazibuko (2013) for major cities in South Africa where incorrect and inaccurate municipal billing is a pervasive and serious challenge that is eroding public confidence in institutions and the payment for services. This leads to poor collection of revenue, which in turn undermines the utilities financial viability (Manyaka, 2014).

There is limited research literature that addresses and quantifies this water loss component, and it includes Mutikanga, Sharma & Vairavamoorthy (2011a) who offer a simple methodology of using sampled field and database information as a basis of estimation in a manner similar to a process modelling based solution of Thornton et al. (2008a). In cases with pervasive billing problems, as those described in Mazibuko (2013), the evaluation of this component of AL will be difficult if not impossible to implement.

1.1.3 Unauthorised Consumption

While some authors such as Criminisi et al. (2009) view this component as also directly related to water utility management, it is a problem that is prevalent in the developing world (Mutikanga, 2012). There are however nuances that the source of such challenges are neither technical nor managerial but rather social and are reflective of the prevailing socio-economic conditions of the society. Traditional approaches to estimate this component similarly include the use of field surveys of sampled data as in Thornton et al. (2008b) which provide an assessment of the prevalence of unauthorised consumption and quantify its contribution to water losses. Field surveys are however not a trivial task

and the lack of an objective and systematic methodology for estimating unauthorized consumption remains a limitation (AL-Washali, Sharma & Kennedy, 2016).

There are however fraud detection applications within the computer science domain that are relevant to this fraud detection problem. Humaid (2012) implemented a water fraud detection methodology that increased the detection hit rate of 1-10 % random manual detection to 80% intelligent detection of water theft. While this work is very relevant it is still reliant on field observations to label unauthorised consumption appropriately to be able to classify them automatically. Fraud detection is related to anomaly detection applications that are also relevant for Data Acquisition and Water Accounting Errors that have found numerous applications in the electricity sector such as in Cabral et al. (2008), Chauhan and Rajvanshi (2013), Depuru et al. (2011) and Depuru et al. (2013). Such applications are predicated on having adequately labelled and verified meter and consumer data and ongoing input from field work to incorporate into the labels of the data.

1.1.4 Errors on Estimates of Unmetered Consumption

This is a relatively new concept introduced in Vermersch et al. (2016) and it is therefore not covered in the literature. It is also doubtful if there will be consensus on this component of the water balance as it relates to estimating an error of an unmeasured quantity, which is rather confusing.

1.2 Water Supply in South Africa

South Africa's status as a water-scarce country is reflected in the formulation of the legislative framework pertaining to water which emphasises water scarcity and the effective management of national resources (Algotsson et al., 2009). Projected total water requirements in 2025 are estimated at 17,000Mm^3 per annum against a yield of 15,000Mm^3, which is further compounded by the increasing negative impacts of climate change and population growth (Department of Water Affairs, 2013). The spatial and temporal variation of water availability within the country is significant, with provinces such as Gauteng, the economic heartland of the country, importing most of its water requirements from neighbouring catchments and countries, such as Lesotho.

The responsibility and authority of delivering water and sanitation services falls within local government in terms of the Constitution of the Republic of South Africa (Act 108 of 1996). The Municipal Structures Amendment Act (33 of 2000), the Municipal Systems Act (32 of 2000) and the Water Services Act (108 of 1997), among others, define which municipalities are water services authorities, how they can perform the water services provision function and other requirements on how they are meant to function. Water loss management is therefore the direct responsibility of all duly authorised water services authorities and their water services providers, which generally tend to be the same municipality, although the functions may be accounted for separately. There are, however, exception to this – a few municipalities have created and/or outsourced the water services provision function to external entities that are accountable to them. A good example of this is the City of Johannesburg.

The City of Johannesburg is one of the largest metropolitans in South Africa and the African Continent with a population of 4.4million in 1.4million households (2011 Census Results). It is situated within the Gauteng Province and covers an area of 1,645km^2. Despite its small geographic footprint, it is estimated to contribute about 16.5% to the South African GDP in 2011 (Gauteng Provincial Treasury, 2012) and is therefore referred to as one of the key economic hubs of the country. The water service function of the metropolitan is performed by Johannesburg Water, a 100% municipal owned entity.

1.2.1 Johannesburg Water

Johannesburg Water SOC[1] Ltd (JW) is a municipal entity that commenced business on 1 January 2001 and is currently supplying about 1 572 Mℓ/day of potable drinking water through a water distribution network of 11 896 km pipelines, 122 reservoirs and water towers, and 37 pump stations (Johannesburg Water, 2016). Formal connections are estimated at 561,000 connections of which 81% are metered. These connections consist of 110,000 deemed consumption connections (flat rate

[1] An SOC is a State-Owned Company in South Africa

tariff based on 20kℓ/hh/mth consumption), 325,000 conventional metered connections (post payment) and 130,000 pre-payment connections.

Prior to June 2014, JW was reporting its water losses in terms of “unaccounted-for-water” (UFW) and not the “non-revenue water” in the best practice water balance. UFW is equivalent to water losses without the authorised unbilled consumption component. The water loss estimates before June 2013 were at 31% of the system input volume, of which 26% of the losses were estimated as real losses and the remaining 74% as apparent losses. The approach was that real losses are estimated on a bottom-up approach and the rest is assumed to be apparent losses contrary to the guidelines provided by IWA in Lambert and Hirner (2000) which recommend estimating apparent losses first.

As of June 2016, the equivalent estimate for water losses were at 22.6% while the total NRW was 35.3% (Johannesburg Water, 2016). Interestingly, the contribution and proportion of real apparent losses water losses is currently the opposite of what it was in prior years. Physical losses now make up 72% of the water losses while commercial (apparent) losses are estimated to be 28% of the losses. Apparent losses were in this case estimated using the guidelines figures in Seago, Bhagwan and McKenzie (2004), also promoted by the regulator, the Department of Water and Sanitation, while the balance was taken as physical water losses.

The difficultly of estimating the components of the water balance, referred to in Klingel and Knobloch (2015) and Mutikanga (2012), are quite clear in the progression of water loss assessment in Johannesburg. The switch-over in the make-up of the losses clearly shows that there is inadequate confidence in how water losses are currently estimated. Depending on the outcome of the assessment, water loss management interventions would differ remarkably for either real losses and for apparent losses as both require different interventions. This underscores the need of improving the tools and methods of assessing and the measurement of water losses, and in particular, apparent losses. Applying such refined tools with Johannesburg Water stands to benefit the wider water sector as JW is the largest water service provider in the country, and incidentally has one of the largest losses in the country by volume due to the high system input volumes consumed daily.

1.3 Need for this Research

Generally, water loss management remains a relevant and strategic priority for many jurisdictions of the world, with water use efficiency being one of the indicators (6.4) under Goal 6 of the Sustainable Development Goals (United Nations, 2016). In particular, apparent loss control in urban water supply systems remains in its infancy, and much work remains to be done to bring it at par with available real loss interventions (AWWA Water Loss Committee, 2003; Mutikanga, Sharma & Vairavamoorthy, 2013). Apparent losses in water distribution networks are therefore largely unquantified or crudely estimated in many areas, particularly within sub-Saharan Africa where the only known in-depth assessment was in Uganda (Mutikanga, Sharma and Vairavamoorthy, 2011). In cases where proper assessments that incorporate fieldwork and laboratory work are done, the challenges have been well described by numerous authors (Arregui et al., 2009; Couvelis, 2012; Mutikanga, 2012; Ncube & Taigbenu, 2016). These challenges and the expense involved make the adoption of such methodologies cumbersome and less attractive for many utilities, particularly in the developing world. The need for viable alternatives can therefore not be understated. While some authors have looked at alternatives (Couvelis & van Zyl, 2015; Mbabazi et al., 2015), none of these have been properly validated against the existing standard method involving field and laboratory investigations, and they have considerable subjectivity with reliability that depends on the quality of the data and billing records collected by the water utility. The development of validated apparent loss assessment methodologies that can readily be adopted by water utilities personnel remains a need.

Water loss management is currently being done mostly through the collecting field observations and generating a hypothesis, i.e. theory driven, understanding-rich processes (Babovic, 2005) and other sectors have been adopting data-driven discovery processes, such as those summarised in Morais et al. (2009) on the application of data mining techniques in power systems. Related work in water loss management has been very limited within the apparent loss assessment domain and this presents an opportunity and possibility that data mining of available meter and billing information could reveal insightful trends that could be used for apparent loss management. While there are emerging

trends that seem to suggest that smart metering could have a meaningful impact in water loss management (Kossieris et al., 2014; Loureiro et al., 2014; Cardell-Oliver, Wang & Gigney, 2016), conventional metering with coarse monthly or sub annual data will remain a reality for many utilities for many years. This therefore indicates that methods and solutions that work with currently available technology and data will remain relevant into the future.

Previous research on apparent losses has produced some tools for various purposes such as for determining the optimum replacement period of a meter (Arregui et al., 2003, 2010; Mutikanga et al., 2009; Puleo et al., 2014) and meter database for weighted error estimation (Arregui et al., 2009), but most works have focused on only specific aspect apparent water losses without integrating assessments of various components of these losses. There is therefore a current gap of tools, particularly within the developing world context, for the integrated assessment and management of apparent water losses.

The proposed research therefore seeks to bridge these identified gaps through the combination of traditional field and laboratory work to assess apparent water losses and thereafter apply data mining techniques to develop a suitable and validated alternative assessment method. Management tools that integrate the assessment and management of various aspects of apparent water losses shall also be developed.

1.4 Objectives of the Research

This thesis quantifies the apparent water losses for a South African distribution system and explores the applicability of data mining techniques on water billing datasets for the integrated assessment and enhanced management of water losses. This will be achieved through the following specific objectives;

1. Quantification of apparent losses due to metering errors using the weighted meter accuracy methodology that comprises field and laboratory investigations for a representative sample of the City of Johannesburg consumers.

2. Evaluation of viable alternative methodologies in literature to estimate apparent losses due to metering errors to gauge their applicability, shortcomings and further refinement/development requirements.

3. The development of a suitable methodology for the exploration of water billing and metering data for the assessment water meter accuracy degradation through the application of data mining techniques on monthly water metering and billing data.

4. Validation of the developed data mining methodology using the results of the weighted average field-based methodology.

5. Development of a suitable meter management framework for the study area that will enable the optimum management of apparent water losses.

6. Development of integrated decision enhancement tools for the assessment and management of apparent water losses due to metering errors based on both field survey and data mining techniques and evaluate its effectiveness by application on Johannesburg Water billing data.

1.5 Scope of the Research

The scope of the research is primarily limited to the first principles of assessment of apparent losses due to metering errors and the development and validation of a methodology for the application of data mining to water billing databases to estimate assessment apparent. Ultimately, a suite of tools for the integrated assessment of apparent losses is also developed.

While unauthorised consumption, data acquisition errors and water accounting errors are an integral part of apparent losses, the work of Mazibuko (2013), which included Johannesburg, the study area, clearly indicate underlying billing challenges that makes the assessment of these component a futile exercise. Only a cursory assessment of these components is done, and they are excluded from the research.

Field work and laboratory work is limited to the work necessary for the development and validation of the methodology and suite tools. While effort was made to broaden the meter testing flow rates to

obtain meaningful results, the evaluation of accuracy of very low flow rates is also excluded due to the cost and time involved.

With respect to the study area of Johannesburg context, only the data relating to the conventional metered areas was covered in this study while the prepayment and deemed areas and the corresponding datasets were excluded from the research and the analysis. This is due to the difficult of assessing those areas due to the consumer sensitivity and the political and physical risk involved in working on meter related issues in the respective townships. However, non-domestic consumers in these areas were included in the sample data as they are conventionally metered and do not share the same challenges.

1.6 Outline of the Thesis

This thesis comprises a total of eight chapters which are as follows;

Chapter 1, this chapter, provides an introduction into the subject of apparent water losses and the prevailing context for its management. Motivation for the need for the study, together with the research objectives and scope are also discussed.

Chapter 2 is a detailed review of the literature and application of apparent loss, together with the relevant context on how it fits within the greater water loss management paradigm. Various apparent losses methodologies are discussed together with their strengths and gaps that require attention. A review of data mining concepts and techniques are also reviewed with emphasis on their possible application in the assessment and management of apparent losses.

Chapter 3 details the activities and work done in the determination of a sample of consumers representative of the study area, the meter logging and the determination of consumption profiles and the extent of on-site leakage in the sample properties. Secondly, the analysis of third-party smart metering data to determine the consumption profiles for a few consumer categories is also included, complimenting the work of characterising consumption within a developing nation context.

Chapter 4 discusses the removal of a sample of water meters within the study area for the controlled meter testing in an accredited flow laboratory to determine the accuracy of meters at determined test flow rates and the degradation of meter accuracy. A review and analysis of historical meter testing data is also done to determine its usefulness for the proactive management of apparent losses, with the introduction of a new concept for determining the probability meter failures.

Chapter 5 consolidates the results of the previous two chapters to assess the apparent water losses for the study area using the traditional assessment method, together with a review of the applicability of key default values. Some modification of the assessment method is also proposed to make it more representative. Two alternative methods are also applied within the study area to evaluate their applicability, accuracy and the possibility of validation using the results of the traditional assessment method.

Chapter 6 introduces the development of a new methodology for the assessing apparent water losses using a combination of data preparation, clustering algorithms and regression analysis. The method is applied and validated for the data of the study area by determine the meter accuracy degradation and an estimate for apparent water losses, which is compared against the results of the traditional method. A generalised relationship for the assessment of apparent losses for similar utilities is also provided.

Chapter 7 introduces a modified meter management framework and a consolidated management tool, based on the meter management framework. The management tool provides several sub-tools that facilitate the analysis and calculation of several key metrics that will enhance the ability of decision makers in managing apparent water losses better.

Chapter 8 summarises the findings of the research, contribution to knowledge and suggest further research areas.

Chapter 2 Review of Methods

Chapter 1 has motivated why the study of apparent losses is important, including the various gaps that currently inhibit the assessment of apparent losses. This chapter reviews the literature and tools that deal with the subject and goes into detail in showing the gaps, some of which are filled in subsequent chapters. The review is extended to areas that currently might not be dealing with apparent losses but are relevant and provide an opportunity for application within this domain.

The chapter begins with the necessary broader contextual information on water loss management, together with the standard terminology, to provide the setting for the subject of this thesis. Greater detail is then placed on the various tools and methods used in literature and practice for the management of apparent water losses. While the subject of apparent losses is directly about the management of meters, a section on integrated water meter management is also included to look at existing frameworks and how all other tools can benefit meter management. Lastly, sections dealing with data mining and its applications and their relevance to apparent losses are also included.

2.1 Water Loss Management

Methods for leakage management can be broadly classified into leakage assessment methods focusing on quantifying the amount of water lost (of primary interest in this thesis); leakage detection methods and leakage control methods and models (Vassiljev, Koppel & Puust, 2010; Darsana & Varija, 2018). Leakage assessment methods can be broadly classified into either bottom-up or top-down leakage assessment. There are two indirect/alternative additional assessment methods comprising Hydraulic Network Modelling Assessments and Statistical Flow Analysis, both of which rely on the use of data about the water network and the flow to estimate real losses. The two alternative

assessment methods also fall within the leakage detection category as the exact point of leakage is also identified together with the estimated loss.

Generally, the term water loss assessments unfortunately tend to be synonymous with real loss assessment both in practice and research, without the express inclusion of apparent losses. The four main activities that real loss assessment and management are summarized in Figure 2-1.

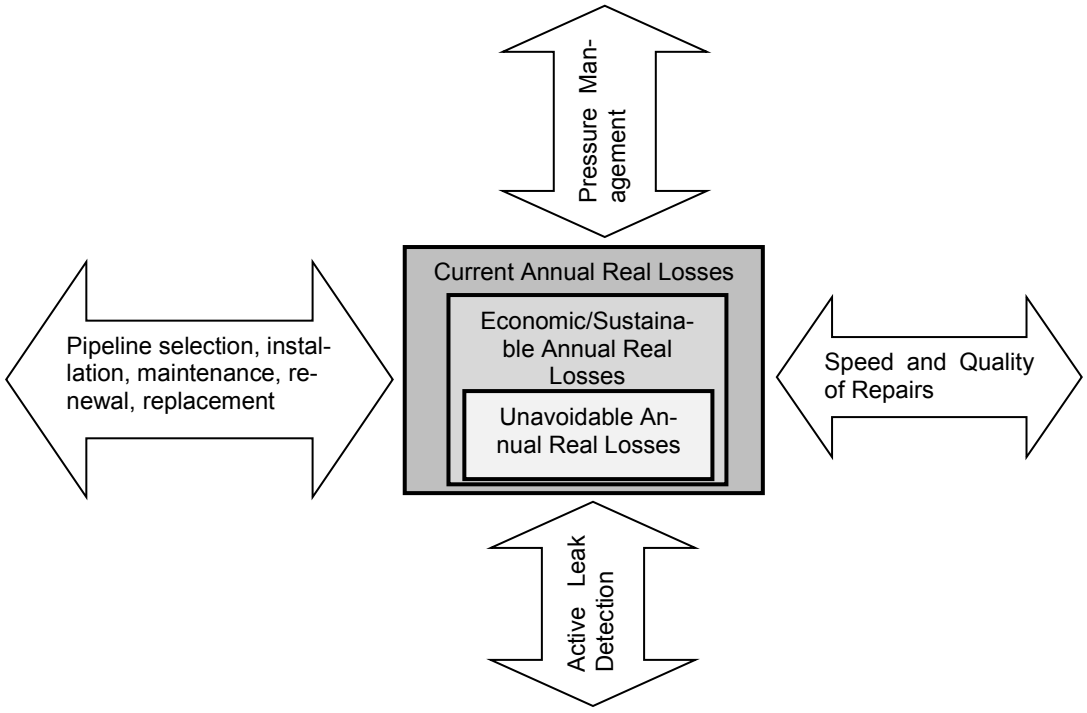


Figure 2-1: Real Loss Components

Adapted from Johnson and Vermersh (2011)

2.1.1 Top-down Assessment

The water balance, introduced in Lambert & Hirner (2000) and shown in Figure 2-2, is the most common top-down approach for assessing the components of water losses (Loureiro et al., 2014). Prior to 2000 there was no international consensus on the measurement and assessment of water losses (Pearson & Lambert, 2013). In addition to the water balance, Lambert & Hirner (2000) (commonly referred to as the Blue Pages) summarises the recommended basic standard terminology and guidelines for the calculation of losses and performance indicators for international comparisons of losses. While various adaptations and versions of the water balance exist and there might be differences of opinion in what the naming of the terms, Figure 2-2 is generally adopted as best practice.

<table>
<tr><td rowspan="8">System Input Volume,
m^3/yr</td><td rowspan="3">Authorised Consumption,
m^3/yr</td><td rowspan="2">Billed Authorised Consumption</td><td>Billed Metered Consumption (including water exported)</td><td rowspan="2">Revenue Water, m^3/yr</td></tr>
<tr><td>Billed Unmetered Consumption</td></tr>
<tr><td rowspan="2">Unbilled Authorised Consumption</td><td>Unbilled Metered Consumption</td><td rowspan="6">Non-Revenue Water,
m^3/yr</td></tr>
<tr><td>Unbilled Unmetered Consumption</td></tr>
<tr><td rowspan="5">Water Losses,
m^3/yr</td><td rowspan="2">Apparent Losses</td><td>Unauthorised Consumption</td></tr>
<tr><td>Metering Inaccuracies</td></tr>
<tr><td rowspan="3">Real Losses</td><td>Leakage on Transmission and/or Distribution Mains</td></tr>
<tr><td>Leakage and Overflow at Utility's Storage Tanks</td></tr>
<tr><td>Leakage on Service Connections up to point of Customer metering</td></tr>
</table>

Figure 2-2: Standard IWA Water Balance

The Blue Pages define real losses, also known as physical losses, as the physical water losses from the pressurised system up to the point of customer metering, including volume lost through all types of leaks, bursts and overflows. Apparent Losses, also known as commercial losses, consist of unauthorised consumption (theft or illegal use), and all types of inaccuracies associated with production metering, customer metering and related data acquisition processes. Apparent losses, therefore, comprise water that is not recorded correctly by a meter but is likely consumed – as such, the water metering device is very central to apparent losses. While the Blue Pages recommend that the components of Unauthorised Consumption and Metering Inaccuracies be assessed by "best means possible" prior to the calculation of real losses, no guidance or relevant performance indicators are provided for apparent losses but rather for the components of real losses.

Despite the popularity and perceived simplicity of the standard water balance, the estimates obtained purely from a top-down approach (mostly a desktop exercise) are only crude estimates (Puust et al., 2010). Such estimates are unfortunately the norm in many utilities and are the basis of many water loss assessments.

2.1.2 Bottom-Up Assessment

Bottom-up approaches provide more detailed information that supplements the water balance estimates by implementing component analysis and include;

- Minimum Night Flow Analysis (MNF) assessment in a discreet district metered area (DMA) (García, Cabrera & Enrique, 2006) to estimate water losses. MNF is used in conjunction with the fixed and variable area discharges (FAVAD) concept (Lambert, 2000) to factor in the variation of water pressure during the day.

- Bursts And Background Estimates (BABE) (Lambert, 1994) are premised on the volume of leakage being a function of the leakage flowrate and the duration of that leak. The BABE concept is therefore used to calculate the volumes of water lost due for avoidable leakage incidences that have been attended to and as such should only be applicable to utilities with active leak detection (AL-Washali, Sharma & Kennedy, 2016).

The Unavoidable Annual Real Losses (UARL) are estimated using a combination of the FAVAD and the BABE concepts (Lambert, 2009), the length of the water network, the average length of the service connections, the number of connections and the average operating network pressure.

2.1.3 Hydraulic Network Modelling Assessment

Hydraulic Network modelling has many applications within leakage management that include, among others, determination and optimisation of network zones, pressure management and the modelling of leakage (Wu et al., 2011). The improvements in hydraulic network modelling techniques over the last two decades have made it possible to incorporate leakage assessment within hydraulic networks (Mutikanga, 2012). Examples include Tabesh, Yekta & Burrows (2009) who proposed a novel procedure to determine the nodal and pipe leakage by using a hydraulic simulation model; Nicolini, Giacomello & Deb (2010) used multi-objective genetic algorithms to, among other objectives, minimise of the total leakage in the hydraulic system; Karadirek et al. (2012) used a model to predict water savings due to pressure reduction; Siew & Tanyimboh (2012) presented an extension to the popular hydraulic simulator EPANET 2 (Rossman, 2000) to incorporate pressure-dependent demands for better estimation of demand and water losses; and Adachi et al. (2014) who estimated area leakages of a water network to prioritize leak surveys for the areas.

The cost of developing and calibrating hydraulic network models, together with the cost of collecting and maintaining the requisite data and specific skill level, is one hindrance for the application of this techniques, especially in resource constrained utilities (Trifunovic, Sharma & Pathirana, 2009).

2.1.4 Statistical Flow Analysis

Wu & Liu (2017) included a review of statistical data-driven approaches for burst detection, a special case of leak detection. Statistical analysis is dependent on the availability of hydraulic data that is collected regularly following specific periodic patterns (hourly, daily, weekly, and yearly). Burst detection is therefore considered as a data mining task of anomaly detection (Wu & Liu, 2017). Other examples of this include Wu et al. (2010) who presented leakage analysis modelling based on pattern recognition theory; van Thienen & Vertommen (2016) described the automation of the comparison of flow pattern distributions (CFPD) to assess flow data of a supply area to pinpoint (in time, not space) and quantify changes in the amount of water supplied, and hence leakage. Similarly, Palau, Arregui & Carlos (2012) used principal-component analysis (PCA) to analyse the inflows into a DMA summarising the flow into a statistical model that explains most of the behaviour of the water distribution network.

Like the hydraulic network modelling based assessments, the statistical flow also requires continuous collection and analysis of data and as such the challenges noted by Trifunovic, Sharma & Pathirana (2009) of high development costs and high skills requirements, which are a hindrance, also apply.

2.2 Apparent Water Loss Assessment

Thornton, Sturm & Kunkel (2008a) defines apparent losses as the non-physical losses that occur when water is successfully delivered to the customer but, for several reasons, is not measured or recorded accurately, thereby inducing a degree of error in the amount of customer consumption. The lack of proper estimation of apparent losses will therefore have an impact on all parts of the water balance related to the measurement of consumption and in turn the estimation of real losses in the

top-down assessment method. The key components which make up apparent losses, and must be controlled to manage it, are illustrated in Figure 2-3.

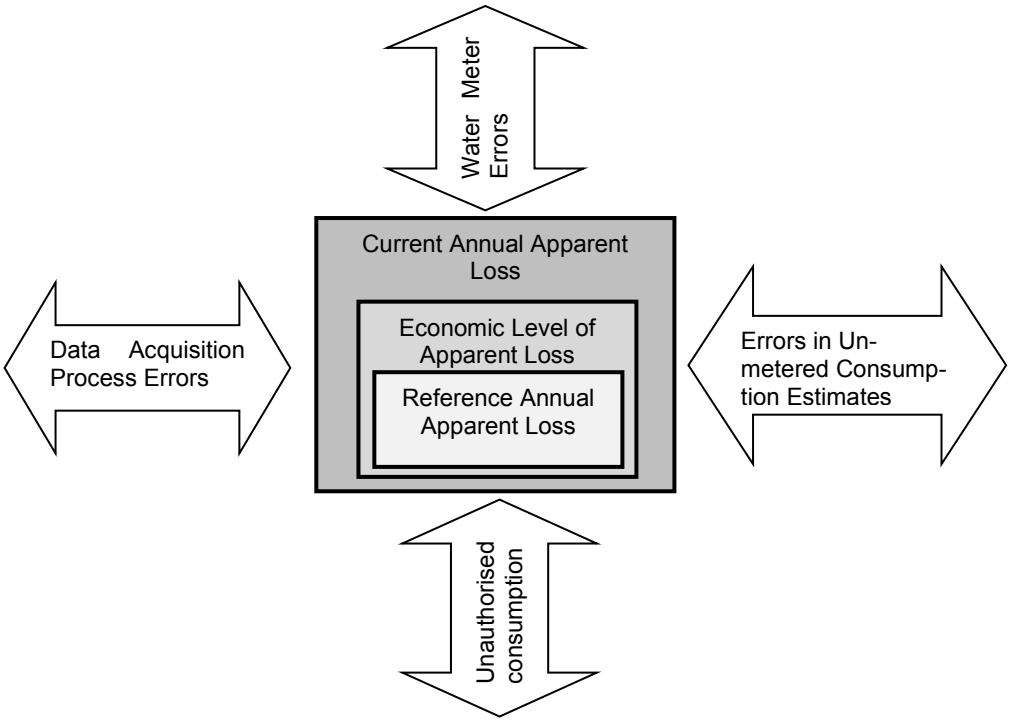


Figure 2-3: Apparent Loss Components

adapted from Johnson and Vermersch (2011) and Rizzo *et al.* (2004)

The current annual apparent loss (CAAL) is at the core of the assessment of apparent water losses and needs to be quantified and reduced appropriately as it contains portions that are recoverable. Economic level of apparent losses (EALL) represent the conceptual point at which the cost of apparent water loss control is equivalent to saving from apparent loss recovery – this is conceptual in the sense that in many jurisdictions, as in South Africa, universal consumer metering is national policy and is part of the cost of doing business. This implies that there could be cases where metering is required even when it is not necessarily economic.

Reference annual apparent losses (RAAL) have been found to be more appropriate compared to the previous Unavoidable Annual Apparent Losses (UAAL) term fashioned in line with the Unavoidable Annual Real Losses (UARL). It was found to be impractical to create an equation to estimate UAAL when there are practically many diverse parameters to be considered (Vermersch et al., 2016). Reference annual apparent losses (RAAL) represent the fact that meters are intrinsically not

100% accurate even when new – the 5% of billed authorised consumption is taken as a default value in the absence of better information. This is may be a reasonable value that is also supported by research such as Arregui, Balaguer & Soriano (2017) who found that new meters have an error as low as -1% to -6% depending on the meters used and the consumption profiles. However, it is important that the "default" value of RAAL be determined in line with local metrological standards that define what is possible and permissible from a metering perspective, and the prevalent consumption characteristics. Quantifying and understanding of a locale specific reference value, which is currently not prevalent, could also provide an impetus in the adoption of metering standard that minimises apparent losses. The final performance indicator is the apparent loss index (ALI), which is a ratio of CAAL and RAAL with values approaching 1 indicating excellent performance. The use of ALI is however not that common, with many utilities utilising the volume of the apparent losses (AL) as a percentage of the System Input Volume (SIV), which is a poor and meaningless indicator (Liemberger et al., 2007; Kanakoudis et al., 2013). AL as a percentage of the metered consumption is a much better indicator in that AL are essentially an error in the measurement of the metered volume and relating to them as such makes for better logic.

Table 2.1 summarises selected water loss studies, some of which were summarised in Mutikanga (2012), that quote AL figures as part of the water losses. The AL figures are reinterpreted for most of the studies to be in terms of the billed authorised consumption, i.e. revenue water (RW) and the value of RAAL is taken to be 5% of the RW. There is a wide variation in the values of ALI from country to country, with developing countries generally performing relatively far worse than the developed countries. This difference may also speak to the need for locale specific default values to cater for different metering capabilities and social challenges that may lead to unauthorised consumption.

It is very important to note that the method of estimation of AL in the Table 2.1 varies very widely and, in most cases, is based on a percentage of either the billed volume, the non-revenue water or any such benchmark. The unavailability of data regarding meter errors, unauthorized use and unbilled authorized use is cited as a challenge by investigators such as in Kanakoudis et al. (2013).

The task of measurement of AL is so elusive that some authors, such as Mousavi et al. (2017), omit it all together, even when the occurrences of unauthorised connections are quantified. As such, while the water balance is meant to be the means of benchmarking, its use for comparing the various water loss components can be very misleading.

Table 2.1: International Apparent Loss Estimates References (adapted from Mutikanga (2012))

References	Country	City/Utility	RW, Mm^3/yr	AL, Mm^3/yr	AL, %RW	ALI
Shin et al. (2005)[1]	South Korea	Busan	2397	116.4	48.6%	0.97
Fanner et al. (2007)	USA	Boston/BWSC	92.4	2.9	3.1%	0.6
	USA	Florida/Charlotte County Utilities	10.1	0.2	2.0%	0.4
	Canada	Halifax	166.9	2.6	1.6%	0.3
Guibentif et al. (2007)	Switzerland	Geneva/SIG	58.6	2.2	3.8%	0.8
AWWA (2009)	USA	Philadelphia	222.3	21.3	9.6%	1.9
Bidgoli (2009)	Iran	Tehran NWWEC	720.7	112.4	15.6%	3.1
Makara (2009)[2]	Papua New Guinea	Eda Ranu/Port Moresby	21.5	6.2	28.8%	6.0
Sharma & Nhemafuki (2009)	Nepal	Bhaktapur	1.5	1.6	104%	20.9
	Nepal	Dhulikhel	0.3	0.015	5%	0.9
Garzon-Contreras et al. (2009)	Colombia	Medellin EPM	185.7	29.7	16.0%	3.2
	Colombia	Cali	121.7	46.1	37.9%	7.6
	Colombia	Bogota	287.6	72.5	25.2%	5.0
Dimaano & Jamora (2010)	Philippines	Manila Maynilad	370.8	98.2	26.5%	5.3
Kanakoudis & Tsitsifli (2010)	Greece	Larisa/DEYAL	11.7	1.1	9.4%	1.9
OFWAT (2010)	England/Wales	22 Companies	4213.2	117.9	2.8%	0.6
Schouten & Halim (2010)	Indonesia	Jarkata Palyja	130	47	36.2%	7.2
Sharma & Chinokoro (2010)	Zambia	Lusaka LWSC	35.64	11.77	33.0%	6.6
Mutikanga, Sharma & Vairavamoorthy (2011a)	Uganda	Kampala NWSC	26.63	9.32	35.0%	7.0
Xin et al. (2014)	China	SA City	395.2	31.19	7.9%	1.6
Department of Water and Sanitation (2017)	South Africa	107 municipalities	2386.88	302.25	12.7%	2.5

[1] Consumer block analysis in m^3/day

[2] 2001 figures used as this is the only year that apparent losses are properly defined

In a bid to mitigate the use of percentage default values for AL, Seago, Bhagwan & McKenzie (2004) presented what they view as a more pragmatic approach, although not scientific by their own admission, based on survey responses from utilities to classify their unauthorised consumption, meter

accuracy and age, and the billing accuracy on an ordinal scale. The ratios used for AL are with reference to the water loss quantity as opposed to the volume of water recorded by meters as should be the case.

Mutikanga, Sharma & Vairavamoorthy (2011a) followed a similar approach to Seago, Bhagwan & McKenzie (2004) but rather used the weighted meter accuracy methodology to carry out a detailed assessment of AL in Kampala, Uganda. However, the main difference in the estimates provided by Mutikanga, Sharma & Vairavamoorthy (2011a) to those of Seago, Bhagwan & McKenzie (2004) is that the former is correctly based on the volume of sales, which should be the case, while the latter is a proportion of the water loss quantum. The guideline estimates for both studies are reproduced in Table 2.2 and Table 2.3 but care must be taken in interpreting them as the percentages do not refer to the same base.

Table 2.2: Meter and Data Acquisition Error Estimates

Meter Age and Accuracy					Data Transfer		
Rating	***South Africa***		***Uganda***		***Rating***	***South Africa***	***Uganda***
	Good WQ	***Poor WQ***	***Direct Supply***	***Storage Tanks***			
Poor, >10 years	8%	10%	10%	28%	Poor	8%	10%
Average, 5 - 10 years	4%	8%	8%	20%	Average	5%	6%
Good, <5 years	2%	4%	5%	15%	Good	2%	2%

Table 2.3: Unauthorised Consumption Estimates

South Africa		Uganda	
Rating	*Estimate*	*No. of connections*	*Estimate*
Very high	10%	> 100,000	10%
High	8%	50,000 – 100,000	3%
Average	6%	5,000 – 50,000	2%
Low	4%	< 5000	0.50%
Very low	2%		

For meter accuracy, Mutikanga, Sharma & Vairavamoorthy (2011a) did not differentiate by water quality, which has been known to adversely affect meter quality (van Zyl, 2011), and this could be attributed to the quality of the water supply in the study area being homogeneous. This is not the case in South Africa where different municipalities are responsible for different water supply systems.

Distinction should however be made between direct supply and that via storage tanks which was omitted in Seago, Bhagwan & McKenzie (2004),likely due to the fact that storage tanks are not prevalent in South Africa. There is a marked impact of storage tanks on the inaccuracy of the meter with up to more than twice the estimate for directly supplied meters. The use of number of connections to estimate the contribution of unauthorised connections in Uganda is not explained and may be illogical for many communities where there could be very a high proportion of unauthorised connections from a small number of connections, and vice versa. The estimation of unauthorised consumption using the ordinal scale of Seago, Bhagwan & McKenzie (2004) makes better sense in this particular case.

There are other authors who have followed a rigorous approach in estimating apparent losses, and particularly the component due to metering errors and these include Noss, Newman & Male (1987), Yee (1999), Arregui et al. (2006a, 2013), Arregui, Cabrera Jr. & Cobacho (2006), Mutikanga, Sharma & Vairavamoorthy (2011a) and Noss *et al.* (1987). All these studies use a methodology that measures both the accuracy of a meter and the effectiveness of that meter in measuring a particular consumption pattern (Ferréol, 2005). These are collectively known as determining the weighted accuracy of a meter and conversely determining the amount of water that is not registered for every 100 litres of water consumed in relation to the water consumption pattern of the user (Arregui et al., 2006a). A different approach, albeit with some commonalities, is suggested by Johnson (2012) who determines apparent losses in any year as

$$AL = n \bullet Q_{ave} \bullet \bar{\varepsilon}_w \bullet \left(\frac{P - x}{P}\right) + \delta_r \bullet (Age + x)$$

Equation 2.1: Alternative Apparent Losses Estimates

Where n is the size of the meter fleet, Q_{ave} is the average annual demand, P is the number of years in the meter replacement cycle, x is the current year, δ_r is the error decay rate and Age is the average meter age since the original weighted error of measurement was established. There is no wide usage of the formula to evaluate its usefulness. Typically, the variation of the weighted error with meter age

and/or accumulated volume is estimated through relationship of the forms in Equation 2.2 and Equation 2.3.

$$\bar{\varepsilon}_w = A(Age\ or\ Vol) + B$$

Equation 2.2: Meter Error Degradation (Age or Volume)

$$\bar{\varepsilon}_w = (A \bullet \ln(Age) + B) + (C \bullet Vol^2 + D \bullet Vol + E)$$

Equation 2.3: Meter Error Degradation (Age and Volume)

The constants A to E are context specific and must be evaluated using the homogeneous grouping of meters and consumers.

2.2.1 Weighted Meter Accuracy

The weighted meter accuracy methodology is an empirical method that is tried and tested in the estimating apparent water losses within utilities. It is currently the gold standard for the assessment of AL. The process of implementing the methodology is illustrated in Figure 2-4 and involves the evaluation of samples of meters and consumers to determine meter accuracy and consumption patterns, respectively. The approach for domestic consumers is slightly different to that for industrial consumers who are individually unique and should therefore be studied individually.

For domestic meters, it is the preferred practice to use of statistical tools and extrapolate the results obtained from a representative sample to the rest of the meters and consumers (Arregui et al., 2006a). The objective is to identify homogenous groupings of meters and consumers and to compute the weighted meter accuracy.

The concept of sampling makes it easier in that not all meters need evaluation, but representative samples still require that a meaningful number of meters and consumers be evaluated. For example, to be able obtain a sample that is representative at a 95% confidence level within ±5% margin of error, you require a sample size of about 380 for a population of 40,000 consumers/meters. This leads to the main drawback of the method as being labour-intensive and expensive to implement despite providing the best estimate of apparent water losses.

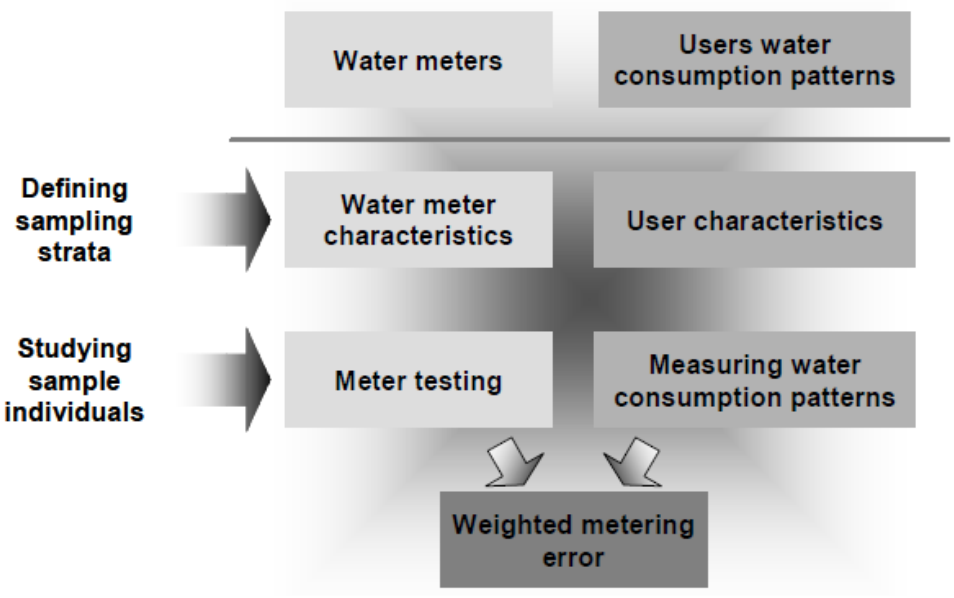


Figure 2-4: Methodology for determining weighted meter accuracy

Source: Arregui *et al* (2006)

Mathematically, the weighted meter accuracy for each homogeneous meter-consumer grouping is given by

$$\bar{\varepsilon}_{\mathrm{w}} = \sum_{q=1}^{n} \mathrm{PTC}_{\mathrm{q}} \bullet \mathrm{GAAL}_{\mathrm{q}}$$

Equation 2.4: Weighted Meter Accuracy

Where PTC is the percentage of total consumption and $GAAL$ is group average accuracy test results at the various meter test flowrates *q = 1 to n*. It is standard practice to test the meter at various flow rates to reproduce the meter error curve – this is discussed further in §2.2.2.

Arregui et al. (2014) measured the weighted error of 330 new and used meters and monitored the consumption of 200 domestic users over a week. The results show that new Class B velocity meters (single jet) have an initial error of -3.45% to -6.37 while Class C meters have a narrow range of -2.14% up to -3.5%. Positive displacement meters have an even better initial error of -0.76% up to -1.04%. For used meters, the average weighted error is about -10%, with the actual degradation best represented by both age and accumulated volume, although the accumulated volume only having an impact after 2500m^3. In a related study, Arregui, Balaguer & Soriano (2017) investigated the

performance of 11 new meters, considering different consumer patterns and metrology standards (ISO and American Water Works Association (AWWA)) and found that the ISO velocity meters under ideal conditions have an error range of -0.71% to -3.87%. On the other hand, ISO oscillating piston meters showed errors of between 0% and -1% independent of the type of user while AWWA oscillating piston meters were much worse with errors of up to -11.37%. In another study, Barfuss, Johnson & Neilsen (2011) tested AWWA single-jet meters registered an error of +4.56% while the AWWA multi-jet meters registered an error of -4.46%. As would be expected, the weighted error for users with a roof tank with ball valves are significantly worse due to the induced low flows.

Szilveszter et al. (2015) compared the weighted error for 384 meters in Ecuador using two consumption patterns, both of which were from different locales to the study area and found global errors of -9.95% and -10.9%. While the results are indicative, the use of consumption profiles that are not from Ecuador is problematic as the local consumer characteristics are likely different. Mutikanga et al. (2011a) similarly found a weighted error of -22% (due to the prevalence of storage tanks) in Uganda from the meter logging of 90 properties and the testing of 250 meters. These results show the importance of the metering standards and, in turn, the related testing, the user consumption patterns and that both the accumulated volume and age are key factors on the degradation of the weighted error. The contextual information is therefore very important in both the interpretation and comparison of different weighted error results.

2.2.2 Meter Testing

The metrological control of water meters is a foundational imperative for meter management as it forms the basis for all meter accuracy estimation through meter testing (Arregui, Cabrera Jr. & Cobacho, 2006). Meter testing in a Utility could be done for a variety of reasons that include suspicion of meter failure (e.g. meter that does not turn), customer requests/complaints, routine testing (Noss, Newman & Male, 1987), among others. Results from such meter testing programs are valuable for a variety of reasons such as determining meter failure rates and probabilities, the degradation of meter accuracy for the standard flow rates such as demonstrated in Ncube & Taigbenu (2015) and Noss et al. (1987). These may also be a means of satisfying local metrological standards with respect

to both initial and ongoing meter verification requirements, as is the case in South Africa with SANS 1529-1:2006 Standard (SANS, 2006).

Various authors have been involved in evaluating meter performance through meter testing in a laboratory environment. Bowen et al. (1991) evaluated 200 meters of twenty distinct types and found that meters which receive pulsed-flow (of which intermittent supply is a type) experienced greater meter accuracy degradation that those with constant flow, and that at flows above 3.8ℓ/m polymer body meters were less accurate that brass meters. In addition, meter chamber size and rotating element also influenced the degree of accuracy. Barfuss et al. (2011) investigated 450 new meters and 595 pulled meters and found that a higher than expected number of new meters did not meet the AWWA standards and that degradation rates for the new meters tested for endurance were higher and more apparent at low flows. There was obviously variation amongst different meter manufacturers. The pulled meters showed that meter accuracy was affected by suspended solids in the water than the other water quality factors and that the degradation rates correlated well with results from the endurance testing of new meters. These studies, both of which are from the United States of America and therefore done in line with AWWA standard, reveal some of the numerous variations that need to be considered in evaluating meter accuracy, and such variations do tend to be location specific. Also important is how the particular meter testing is done by the respective utility as failure to follow testing protocol gives rise to inaccurate results and increases uncertainty as pointed out by Shields et al. (2013).

One exacting demand of meter testing for the purposes of AL assessment is that there is a requirement to reproduce, as best as possible, the meter error curve. This requires the testing of a single meter at several selected flow rates to reconstruct a representative error curve. This requires more than the four flowrates (Q_1, Q_2, Q_3, Q_4) referred to in standard aligned with ISO 4064-1:2005, (q_{min}, q_t, q_p, and q_p) such as the South African SANS 1529-1:2006, or the three flowrates (minimum, intermediate, maximum) in the AWWA's Manual M6. This is where a few studies (Allender 1996; Yee 1999; Mutikanga et al. 2011) which evaluated AL fell short by using only three points to reconstructed the meter curve. The pitfalls in doing this are demonstrated in Arregui et al. (2009) where substantial

errors in the estimation of the metering accuracy become the inevitable result. An as example, Figure 2-5 illustrates a meter curve reproduced from ten flow rates used in this study and clearly shows that omission of certain flow rates, particularly at lower flows, will lead to a different curve thereby distorting the results

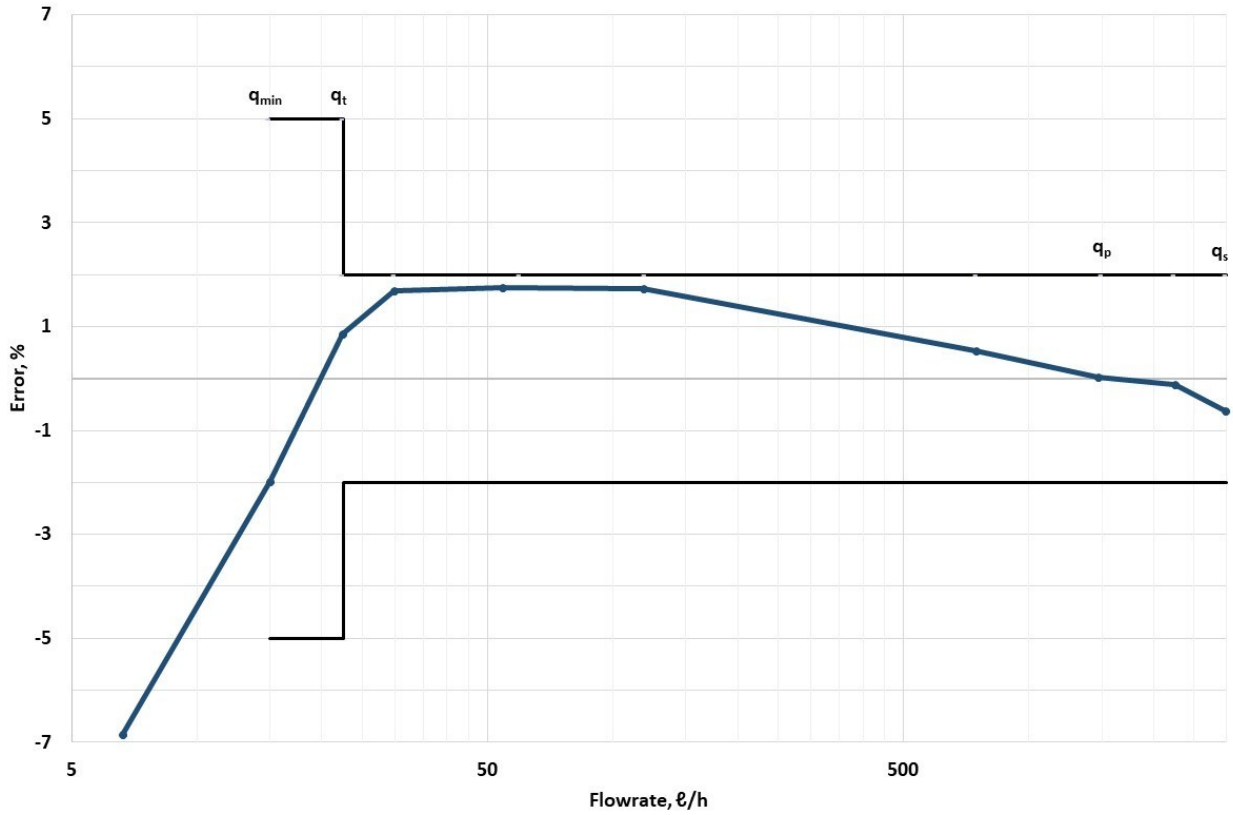


Figure 2-5: Meter Error Curve

There is therefore a need for careful consideration of the selection of the test flow rates to be utilised. This gap is covered by Arregui et al. (2013) who proposed test flow rates for meters of different permanent flow rates, based the testing of several meter models at 20 different flow rates to determine the optimum number of required test flow rate. The optimum number of test points was one that minimised the meter error and was found to be eight test points as shown in Table 2.4 for the permanent flowrates considered.

Table 2.4: Proposed Test Flow Rates for Different Permanent Flowrates

Permanent flow rate (ℓ/h)	Q_{T1}	Q_{T2}	Q_{T3}	Q_{T4}	Q_{T5}	Q_{T6}	Q_{T7}	Q_{T8}
1,600	8	16	32	64	128	800	1,600	2,000
2,500	13	25	50	100	200	1,250	2,500	3,125
4,000	20	40	80	160	320	2,000	4,000	5,000
6,300	32	63	126	252	504	3,150	6,300	7,875
10,000	50	100	200	400	800	5,000	10,000	12,500
16,000	80	160	320	640	1,280	8,000	16,000	20,000
25,000	125	250	500	1,000	2,000	12,500	25,000	31,250
40,000	200	400	800	1,600	3,200	20,000	40,000	50,000
63,000	315	630	1,260	2,520	5,040	31,500	63,000	78,750
100,000	500	1,000	2,000	4,000	8,000	50,000	100,000	125,000

The above, while minimising the error, comes at an additional cost and may lead to fewer meters being tested for any project. This additional investment has merit though, over and above the reasons in the preceding discussions, arising from the prevalence of onsite leaks in consumer properties that are characterised by low flow rates, such as the dripping taps and leaking toilet cisterns (Mayer et al., 1999; Lugoma, Van Zyl & Ilemobade, 2012; Ncube & Taigbenu, 2016) . As such, it is important to evaluate the accuracy of a meter at such low flowrates during AL assessment as such flows are experienced in real life. The only known comprehensive AL assessment study in Sub-Saharan Africa (Mutikanga, Sharma & Vairavamoorthy, 2011a; Mutikanga, 2012) achieved the tests of 250 water meters which were done only at three flowrates, the minimum of which was 15.8ℓ/h for Class C meters and 31.5ℓ/hr for Class B meters, which are much higher than the recommended 8ℓ/h in Table 2.4. While they claimed to have tested a few (number not provided) meters at lower flow rates, there is generally no evaluation of the accuracy of the meters at low flows. This means the gap of evaluating metering accuracy for the complete range of flows still exists in the region.

2.2.3 Consumption Characterisation

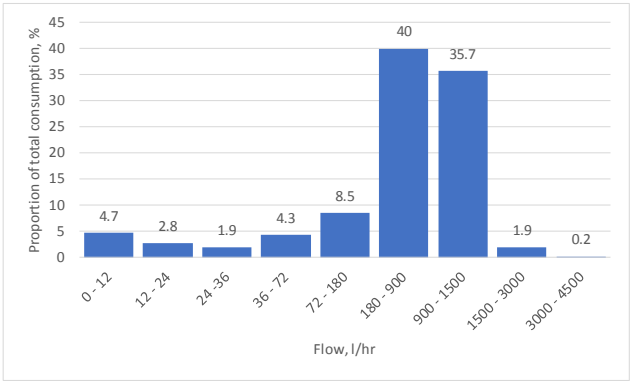
Demand profiles generated from existing consumers provide data essential for making a variety of critical decisions that can assist a utility to improve its water distribution system design, performance and management. These decisions include (AWWA, 2004);

- checking if existing meters are properly sized as part of a routine maintenance and replacement program,
- performing water use audits and leak detection programs,
- collecting peak demand information for cost-of-service studies, and
- collecting peak instantaneous demand data for creation of demand curves used for sizing new and existing service lines and meters.

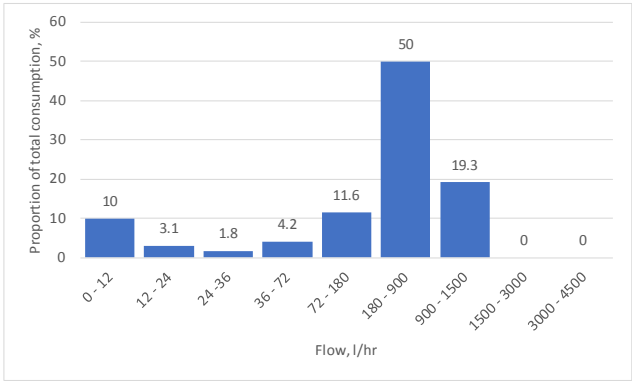
Other objectives of consumption profiling include residential end use studies and troubleshooting consumer pressure and flow problems. Residential end use studies include the works of Mayer et al. (1999), Beal & Stewart (2011) and DeOreo et al. (2016) who profiled the consumption of 1,200, 252 and 1,000 residential homes, respectively, in various cities in America and Australia. The 2016 study is essentially an update of the 1999 study. A few significant studies, all in the United States of America, have also been published on quantifying residential consumption and these include AWWA (1966) which studied 2,300 households and Bowen et al. (1993) who studied 706 households. Other studies include Kuranz (1942), Graeser (1958) and Hudson (1964). In Spain, Arregui, Cabrera Jr. & Cobacho (2006) summarised a related consumption study on 481 properties and in Uganda, Mutikanga (2012) profiled 78 properties. Arregui et al. (2017) summarised some of the popular consumption patterns internationally, and together with the Uganda profiles, they are repackaged in Figure 2-6. Pattern I (a) is associated with indoor use in 389 apartments, Pattern II (b) is indoor use of 58 residences having a roof tanks, Type I (e) are 50 single family households with elevated storage tank and Type II (f) is an apartment block of 9 units (4 floors) fed from elevated storage tanks at roof level. There is therefore a disparity in the reporting of consumption profiles, with a paucity of research, in both scope and depth, within the developing world to cover various properties and consumers.

There are several important considerations that need to be kept in mind when conducting consumption characterisation activities as discussed in Arregui, Cabrera Jr., *et al.*, (2006) and AWWA (2004). Primarily the classification of consumers by type and the selection of a representative sample for

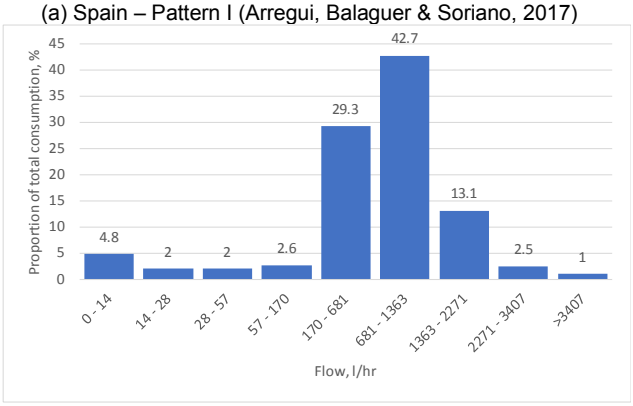
each of the categories of consumers is required. The selection of field equipment is another important consideration as equipment with the highest accuracy, particularly at low flows, the smallest possible sampling resolution (pulses or time) and adequate internal memory is required to capture detailed consumption variation to obtain a realistic consumption profile.



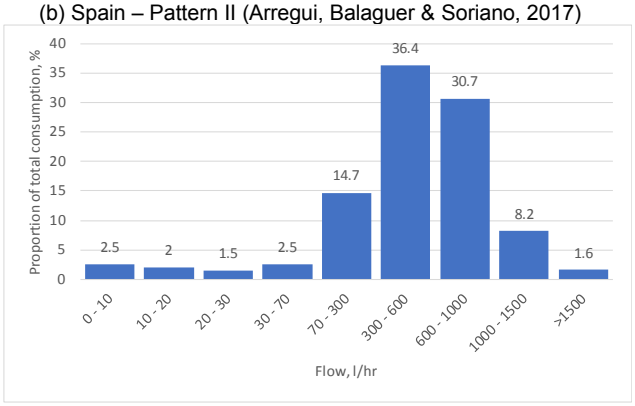
(a) Spain – Pattern I (Arregui, Balaguer & Soriano, 2017)



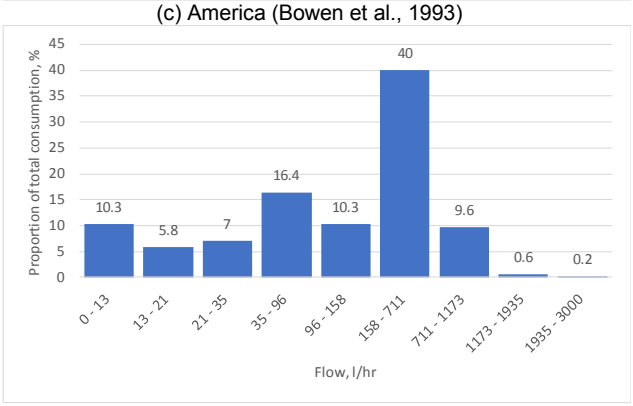
(b) Spain – Pattern II (Arregui, Balaguer & Soriano, 2017)



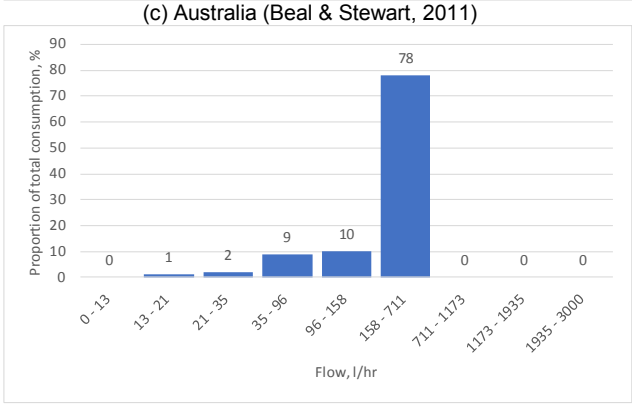
(c) America (Bowen et al., 1993)



(c) Australia (Beal & Stewart, 2011)



(e) Uganda – Type I (Mutikanga, 2012)



(f) Uganda – Type II (Mutikanga, 2012)

Figure 2-6: Consumption Patterns for various areas

This is very different from the common meter logging exercises that record the pulses at every 15 minutes interval or so. The significance of proper selection of equipment is further emphasised by activities that modify the consumption pattern such as fixtures with ball valve, roof tanks and toilets which draw water at very low flow rates, and internal plumbing leaks.

The generally accepted flow measurement period is at least one week to obtain a fair representation of typical consumption by the consumer. Lastly, verification of data accuracy is another important requirement to ensure the volume recorded by the equipment tallies with physical meter readings and the use appropriate pulse factors (AWWA, 2004). Such information is omitted by some authors, but it provides a good basis on which to evaluate the validity and applicability of the consumption profiles. Due to a variety of site conditions such as meters without any pulse output and/or meters with a sub-optimal level of accuracy, it is not uncommon to require that new meters and recording equipment be installed on site to be able to effect measurement as was the case in Mutikanga (2012). As such, the cost of consumption characterisation can be even higher than the cost of meter testing in the laboratory (A*rregui* et al., 2013), and this presents another deterrent for the proper assessment of AL.

2.2.4 Alternative Assessment Methods

Due to the difficulties and cost associated with implementing the methods of assessing AL presented in §2.2.3, a few attempts have been made by various authors to evaluate alternative methodologies.

Arregui et al. (2003) introduced "*Comparative billing analysis*"; a methodology to analyse meter performance to support decision making in water meter management. The main objective is to determine an evolution of the average yearly metered volume versus age or the total metered volume for each meter model. The multi-jet and oscillating pistol meter models compared by computing the average difference in recorded annual volumes with time and the results were also compared with the differences found from the weighted average method. The results were comparable. The authors conclude that by analysing existing data, it is possible to determine the most appropriate meter in a much more efficient method in the use of economic and human resources while still as reliable. However, no information is provided on how many annual records for each of the meters were used in the analysis, only that they were limited, nor did the analysis attempt estimating the accuracy or degradation thereof using the alternative method. Furthermore, the suggested requirements for the method are adequately labelled data with information such as meter models, diameter and installation dates which can be a problem in some utilities. Arregui, Cabrera Jr. & Cobacho (2006) concedes

that the quality and quantity of available databases invariably dictate the viability of such alternative method. There is therefore scope to broaden the method to use available data and to intuitively recognize both the various consumer and meter categories and patterns.

Al-Washali (2011) proposed a mass water and wastewater balance referred to as the "*Reverse Approach*". It is premised on the assumption that apparent losses enter the wastewater sewers while real losses do not. This assumption is invalid for many utilities with significant occurrences of on-site leakage which does not necessarily enter the sewers. The volume of AL is calculated using Equation 2.5;

$$Q_{AL} = Q_{WW} - (Q_{BW} + Q_{UAC} + Q_{OUT} + Q_{EX} + Q_{IN})$$

Equation 2.5: AL from the Reverse Approach

Where Q_{WW} is the inflow to a wastewater treatment plant or sewer outflow at the outlet of a DMA, Q_{BW} is the billed water, Q_{UAC} is the unbilled authorized consumption, Q_{OUT} is the outdoor use, Q_{EX} exfiltration of wastewater from the sewers and Q_{IN} is the infiltration entering the sewer systems. Equation 2.5 does not solve the challenge of assessing AL ; in fact, it can be argued that it compounds it. The method would require the installation of more meters to be regularly read and managed to obtain the required variables. In addition, many areas have the water and sewer management areas/district not coinciding, thereby necessitating the added responsibility of accounting for the flow differences in these non-overlapping areas. The credible assessment of outdoor use is in fact the same process as consumption characterisation which would have similar challenges as detailed in §2.2.3. Lastly, in developing countries there are high rates of sewer overflows and water ingress due to the poor state of the sewer network, and the quantification of those overflows and pollution is no small task. This alternative is therefore unnecessarily complicated and is not viewed as a viable alternative for most utilities in the developing world.

Couvelis & van Zyl (2015) used a meter change database of 142,858 15mm connections from eThekwini Municipality (Durban), South Africa, who had their meter replaced at least once between

June 2004 and February 2011. The data was cleaned and stratified into homogeneous groups according to meter age and registered volume and 65,067 records remained after the data cleansing process. The under-registration error of each replaced meter was estimated as the increase in consumption after replacing the meter with a new one. Periods of 12 months of registered consumption before and after replacement were used to account for seasonal variation in consumption, assuming consumption patterns remained the same and that the new meters have negligible errors. The metering error was thereafter calculated as;

$$\% \, Error = \frac{Q_2 - Q_1}{Q_2} \times 100$$

Equation 2.6: Metering Error

Where Q_2 and Q_1 are the average consumption after and before replacement, respectively. Deterioration in meter accuracy was found to be 0.36% per year and 0.9% per 1,000kℓ. Based on this, for the average meter age in South Africa of 10 years, the average meter under-registration error due to aging was estimated at 2.6% of consumption. Further analysis of the billing database showed that domestic consumption had a decreasing trend over the study period indicating that the error estimation is likely to be under-stated and thus conservative (Couvelis & van Zyl, 2015). The study also estimated the under registration of on-site leakage, assumed to be 15% of leakage based on an estimation from interpolation from manufacturers meter curves, to be 2.2% of consumption. The estimation of meter under-registration from manufacturers' curves is a curious feature as meter under-registration is both a function of the meter accuracy (specific to a particular meter) and the consumption profile over that meter. There are therefore better ways that have a better scientific basis for estimating meter under-registration.

The total meter under-registration error in middle to high-income areas in South Africa was thus estimated to be 4.8% of consumption. The study therefore offers a novel methodology for estimating AL due to meter errors, which by their own admission is likely a conservative estimate. The estimation of meter error from generic manufacturer's curves for new meters is grossly inadequate and does not provide meaningful results; use of consumption characterisation is considered be a better

method as all flow below the registered flow can be taken as unregistered. Validation of the methodology would therefore be critical in evaluating how conservative it is in assessing AL, together with enhancing it to cater for the reported declining trend in consumption with time.

Mbabazi et al. (2015) applied the "*Comparative Billing Analysis*" to develop the meter accuracy degradation profiles for volumetric and velocity meters (three meter models) and estimated AL in the Kampala, Uganda. The analysis started with 128,201 meters of size 15mm and 8 months of readings and these were stratified according to meter model and billing index (meter reading) to arrive at a sample of 7,769 meters. 503 meters were eventually used to model the evolutions of the average yearly metered volume (annual billed volume) versus total age for each meter model and the relationships, represented in Equation 2.7 and Equation 2.8, were found to be linear and similar to those in Arregui, Cabrera Jr. & Cobacho (2006).

$$y = \beta_0 + \beta_1 x$$

Equation 2.7: Variation of Billed Volume

$$d = \frac{\beta_1}{\beta_0}$$

Equation 2.8: Meter Error Degradation

Where y is the annual billed volume, x is the age of the meter and d is the degradation rate. The meter accuracy degradation rates were determined as 6.67%, 4.68% and 1.45% per year for Models 1, 2 and 3 respectively. It is not clear where the billing index was used in the analysis and whether the relationships were derived assuming the same index for all meters. The authors also claim that the degradation rates obtained were higher than observed for water meters (e.g. 0.3% per year in Mutikanga (2012) and 2.1% in Arregui, Cabrera Jr. & Cobacho (2006) because the methodology only considered meter age as the sole factor influencing meter degradation and did not include pressure, water quality, leaks, and users' storage tanks, and thus consideration of meter age does not explain meter degradation entirely. This is rather a very strange assertion in that the billing index, assuming the readings are factual, represents a sum total of all environmental factors experienced by the water meter in the field. Furthermore, assuming that the assertion is indeed correct, the degradation should not be lower but higher when those factors are used. However, the degradation rates are indeed very high for whcih appropiate explainations were not provided in the study. There

are also no reasons provided why only such a small number of meters (0.4%) were used from a large database when more data could have further validated the method. The use of 8 months period for the analysis, despite an unsubstantiated assertion that the study was augmented by data collected over several years, could mean different patterns were superimposed over each other in the analysis, therefore distorting the results. These, and other ambiguities in the analysis, mean the methodology in its current form will be difficult to implement and is therefore not repeatable. While the final recommendation on the best meter for the utility are likely to be correct, the discrepancies in results with laboratory- and field-based assessments in the same environment mean the method is currently not validated despite its potential for being a viable alternative.

Other than the methodology of Al-Washali (2011), the other methods could likely benefit from further refinement and development into usable methodologies by utilities in assessing AL. This will be investigated further in this thesis.

2.3 Integrated Meter Management

The assessment methods described in §2.2 provide useful information for the assessment and management of AL, with the assessment being only the first part of the proper management of apparent losses, in addition to many other key requirements. Arregui, Cabrera & Cobacho (2006) gave a comprehensive overview on the requirements for integrated meter management, with van Zyl (2011) presenting an alternative view that better reflects the South African context. The various components required to achieve integrated meter management comprise nine steps that are summarised in Figure 2-7.

Invariably, prudent meter management comprises a combination of management/back office, laboratory and field work that are once-off activities but should be ongoing and recurring. What is not immediately clear in Figure 2-7 is that these nine steps are predicated on having database systems that contain all the pertinent information on all water meters and the consumers (van Zyl, 2011). While §2.2 covers some of the aspects pertaining to Steps 1, 7 and 8 of Figure 2-7, the following are some other technical subjects covered in literature on the management of meters.

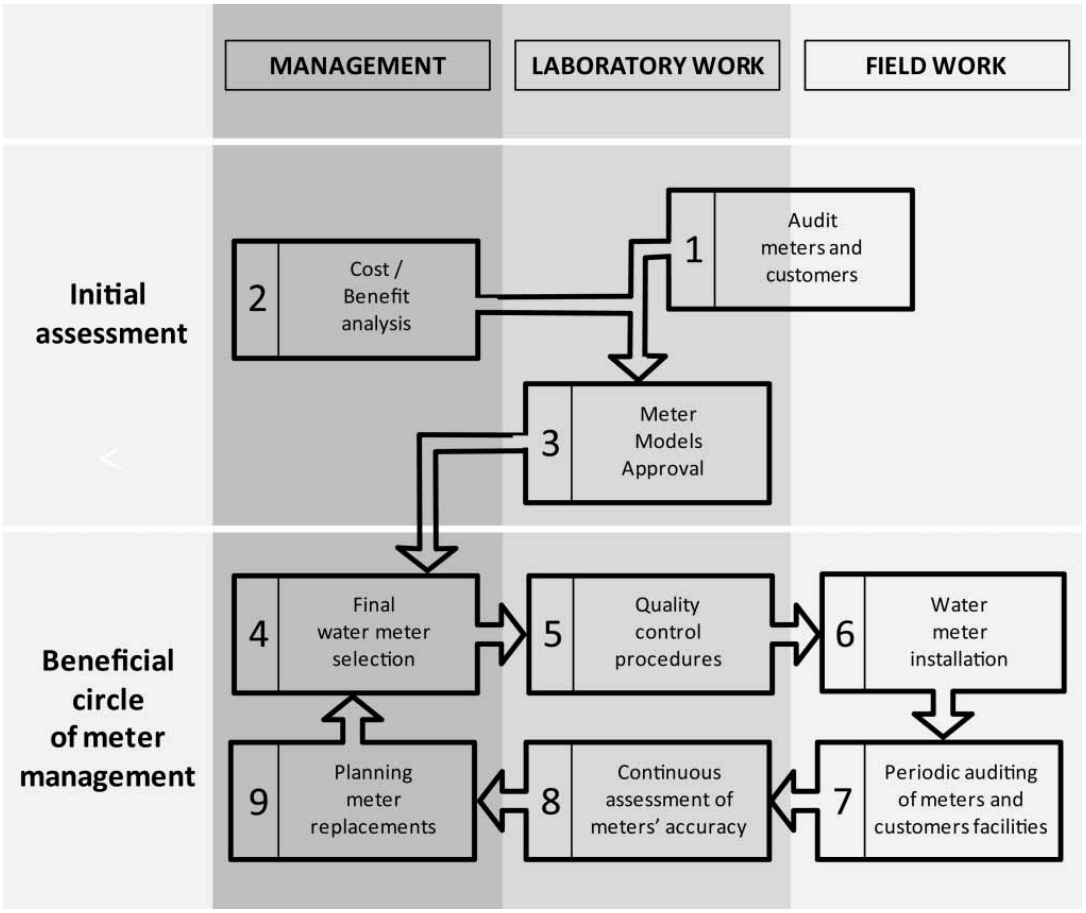


Figure 2-7: Nine Steps towards Integrated Meter Management

Source: (Arregui et al., 2012)

2.3.1 Meter Selection and Replacement

The objective of meter selection is to minimise the costs associated with the meter over its lifecycle, that is until it is due for replacement. This becomes an economic assessment which has been a subject of study by various authors including Allender (1996), Yee (1999), Johnson (2001), Arregui et al. (2003), Davis (2005), Arregui, Cabrera Jr. & Cobacho (2006), Mutikanga et al. (2009) and Arregui et al. (2010). The cost function to be minimised is given by Equation 2.9

$$\min_{costs} C_{acq} + C_{inst} + C_{adm} + C_{AL}$$

Equation 2.9: Meter Selection Minimisation Function

Where C is the cost of acquisition, installation, administration/maintenance and AL costs, respectively, over the meter's lifetime.

The inclusion of the apparent loss component implies it must be assessed, and therefore the use of the weighted error and related accuracy degradation of the meter(s) being evaluated, which makes it non-trivial. In addition, the cost of AL factors in the selling price of the water and all as such future financial losses require to be converted into their present value using the real discount rate given by (Bierman & Smidt, 2007)

$$r' = \frac{1+r}{1+i} - 1$$

Equation 2.10: Real Discount Rate

Where r is the nominal discount rate and i is the inflation. In cases where the replacement period is fixed, as is mandated by legislation in certain jurisdictions, the choice of an appropriate meter is relatively easier and guided by the net present value of current and future cost for any particular meter given by Equation 2.11 (Arregui, Cabrera Jr. & Cobacho, 2006);

$$NPV = C_{acq} + C_{inst} + C_{adm} + \sum_{t=1}^{n} Q_t \bullet \varepsilon_t \frac{C_w}{(1+r')^{t-1}}$$

Equation 2.11: NPV for Meter Selection with Fixed Replacements

Where Q_t and ε_t are the average water consumed and weighted metering error at time t, respectively, and C_w is the cost of water, which is assumed to be constant for simplicity. However, in some cases the replacement frequency is not imposed and the task changes to that of finding the most optimum replacement period and select a meter from options with different lifespans. The NPV of the costs of infinite replacements, known as the net present value of the replacement chain (NPVC) is given by Equation 2.12 (A*rregui et al.*, 2010);

$$NPVC_t = \left[C_{acq} + C_{inst} + C_{adm} + \sum_{t=1}^{n} Q_t \bullet \varepsilon_t \frac{C_w}{(1+r')^{t-1}} \right] \bullet \frac{(1+r')^t}{(1+r')^t - 1}$$

Equation 2.12: NPVC for Meter Replacement and Selection

A sensitivity analysis is normally a requirement to consider the uncertainty of the degradation rate of the weighted error of the meter. While alternative methods can assist in deriving the degradation

rate such equation demonstrates the need for some meter testing and consumption characterisation to determine the starting accuracy of meters – this however needs not be the burden of utilities as most metrology regulation already require meter manufacturers to assess the initial accuracy of meters. The discussion could be on the additional test flowrates to obtain a representative meter error curve.

While Equation 2.11 and Equation 2.12 require the degradation rate to be quantified mostly in terms of time, the average volume consumed is also included in the Q_t term. This is important as volume is, at the very least, increasingly being regarded as providing better representation of the degradation rate. Ideal representations should at least include both age and volume to describe meter degradation as demonstrated in Gavara & Arregui (2016) and most recently in (Arregui et al., 2018). Puleo et al. (2014) proposed a composite "*Replacement indicator*" that uses meter age, total volume passing through the meter and network pressure. Such composite ranking procedures, while not looking at economic evaluation and incorporating metering accuracy, are better than practices of many utilities that use meters till they failure or replace them based on some single, and at times arbitrary, indicator. They however require utilities to maintain the data of the composite indicators, which some utilities would find difficult to meet.

2.3.2 Meter Sizing and Maintenance

The problem of sizing water infrastructure such as water taps, meters and service lines is a very old one and is largely done according to the empirical methodology developed by Hunter (1940) which is the basis of all plumbing codes (Mazumdar, Jaman & Das, 2014). These plumbing codes are essential prescriptive with normative requirements on the sizing of infrastructure, and they have not changed much over years. The Hunter Curves were developed using the binomial probability function based on the 99% probability that the demand of all running water fixtures in a household will not be exceeded. Water fixtures have since changed since 1940 and are far much more efficient and various studies (AWWA, 2004; Mazumdar, Jaman & Das, 2014) have shown that the curves lead to oversized pipes, and related infrastructure. Related to this is that some utilities purely size their meters based on the size of their water connections, leading to oversized meters. Oversized meters

lead to higher AL as meters under-register the flow passing through as discussed in de Medeiros et al. (2010), a case study in Brazil.

One benefit of consumption characterisation, coupled with fixture surveys, is that it can create a wealth of information that may allow a utility to derive locally relevant curves, if required. However, a simpler approach of using the flow profile to check the adequacy of the meter is illustrated in Mutikanga, Sharma & Vairavamoorthy (2010) and should suffice for many utilities. What is lacking is a simple tool to enable users to be able to perform the analysis at a click of a button.

Van Zyl (2011) discussed meter maintenance and how it is normally neglected by utilities as long as the meters are still regarded as functional and a meter reading can be obtained. A summary of the required tasks is also included, such as the cleaning of meter strainers and proper scouring of pipe-lines after the repair of a pipe break. Similar recommendations are also included in the work of Arregui, Cabrera Jr. & Cobacho (2006).

2.3.3 Metrology Regulation and Standards

Metrology, the science of measurement, has to do with the application of legal requirements and standards to measurements and measuring instruments to protect both the buyer and the seller in a commercial transaction. Various international standards exist most of which are aligned, except for standards from the USA, and these include the following;

- *OIML R49: Water meters intended for the metering of cold potable water and hot water* from the International Organization of Legal Metrology (IOML).
- *ISO 4064 Measurement of water flow in fully charged closed conduits — Meters for cold potable water and hot water* from the International Organization for Standardization (ISO)
- *EN 14154 Water meters* from the European Union. This standard is referred to as MID after the EU directive called the Measurement Instruments Directive (MID) that gave rise to them.

The following is a summary of key South African regulations and standards, with emphasis on the management of AL and meters.

2.3.3.1 Trade Metrology Act, 1973 (Act 77 of 1973) as amended

The main requirements of the Act are that all meters used for trade purposes shall be;

a) Verified before being put into use, to ensure the accuracy levels are within the required standards
b) The subject of on-going verification to also ensure the measurements are still within the prescribed tolerances.
c) Removed from service once the meter has become defective or its accuracy is not within the prescribed tolerances.

This Act has since been repealed by the Legal Metrology Act (Act 9 of 2014) which requires that all measuring instruments are subject to initial verification and subsequent verification in accordance with the relevant legal metrology technical regulation. The relevant technical regulations are yet to be published but it is expected that they will, at the very least, maintain current practice with improvements.

2.3.3.2 South African National Standards (SANS) 1529

This suite of standards (currently parts 1 – 4 and 9), approved in 2006 and reaffirmed in 2013, determines the performance characteristics, dimensions, type approval requirements etc. for all metering devices used for trade purposes. The standard is like the outdated ISO 4064:1993, which has been revised twice. Some key requirements include;

a) All water meters used for trade purposes must be type approved.
b) The designated flow rate of a meter is its designed permanent flow rate (q_p) in cubic metres per hour. The other prescribed flow rates are the minimum (q_{min}), transitional (q_t) and maximum or overload (q_s) flow rates.
c) The metrological class of a water meter describes the capacity of the meter to measure within prescribed tolerances of accuracy at prescribed flow rates which are expressed as ratios of the permanent flow rate (q_p) as given in Table 2.5, with q_s being equal to $2q_p$.
d) Only meters with metrological classes B, C and D may be used for trade purposes
e) Meter verification can only be carried out in accredited institution by the accrediting authority (i.e. SANAS) and shall be carried out in terms of SANS 1529-1: Annex B

Table 2.5: Metrological Class of Meters (SANS 1529)

Class of Meter	*For q_p not exceeding 10 m^3/hr*		*For q_p exceeding 10 m^3/hr*	
	Minimum flow rate, q_{min}	*Transitional flow rate, q_t*	*Minimum flow rate, q_{min}*	*Transitional flow rate, q_t*
A	0.04 q_p	0.10 q_p	0.08 q_p	0.08 q_p
B	0.02 q_p	0.08 q_p	0.03 q_p	0.03 q_p
C	0.01 q_p	0.015 q_p	0.006 q_p	0.006 q_p
D	0.0075 q_p	0.0115 q_p	–	–

f) The Permissible Tolerance on indication, which is the difference between the indicated volume and actual volume, shall not exceed the values given in Table 2.6.

Table 2.6: Permissible Tolerance on Indication (SANS 1529)

Meter Status	*Flow Rates*	
	Less than q_t	*Greater than q_t*
New & Refurbished meters	5%	2%
Meters in use	8%	3.5 %

While the South African standard is outdated, some of the local meter suppliers already comply to the more recent and stringent ISO, MID and IOML standards, with relevant certification. However, such meters are still subjected to the SANS 1529 as it represents legal requirement for custody transfer within South Africa. Interestingly, manufacturers of meters with a MID approval equivalent to Class D certification in South Africa occasionally get them certified as Class C by choice simply to avoid the extended approval time for Class D. This therefore means that there are some Class C meters whose metrological performance is much superior and therefore it is important that the as-sessment and choice of meters goes beyond the class of the meter but to the meter specification. Providing utilities with tools that facilitate the selection of such meters with superior metrology but disguised as less capable meters is therefore essential.

An illustration of the meter error envelope and the minimum flowrate for the three classes A to C is presented in Figure 2-8. Regardless of meter class or nominal diameter, all meters rated by a specific permanent flow rate share the permanent and the maximum flowrate while the transitional flow rate and the minimum flow rate varies per meter class. It is clear from Figure 2-8 that the sensitivity of the meter to measure low flows gets better as you move from Class A to Class D. One disadvantage

of having outdated legislation is that no benefit is derived from improvements in technology to achieve more accurate measurements, such as the adoption of tighter accuracy envelopes in newer regulations.

In summary, the existing regulation prescribes the ongoing verification (testing of meters) to evaluate if the accuracy of meters is still within the prescribed limits. Whereas this is currently not enforced, it will be at a future time with the expectation that utilities have the capability to efficiently and effectively carry out the verification process. Furthermore, the existence of superior meters in the market offers utilities with more options, if they have the correct tools to choose the most appropriate meter for their purposes.

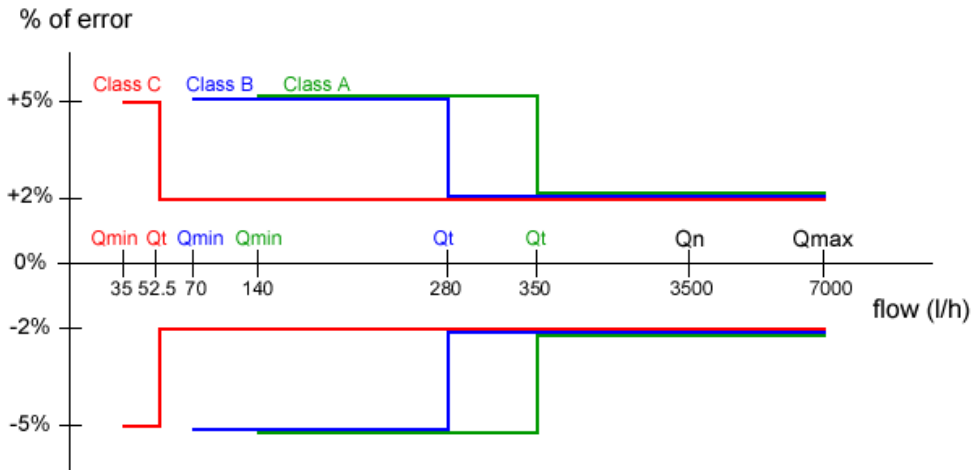


Figure 2-8: Illustration of Difference in Meter Classes (Q_p = 3,500ℓ/h)

Source: www.kwlog.com/index.asp

2.4 Data Mining

Data mining, also known as Knowledge Discovery in Databases (KDD), is the process of discovering valuable information from large amounts of data (Y*in* et al., 2011) through the automatic search and exploration of data to discover patterns, the non-trivial extraction of implicit information using computational techniques from statistics, machine learning, artificial intelligence and pattern recognition (Gorunescu, 2011). Data mining has two main goals in its application (Gorunescu, 2011):

- Using part of the variables to predict one or more other variables such as in classification, regression, anomaly detection. This is also referred to as supervised learning.

- Automatically identifying patterns that describe data and can be understood by a user such as in clustering, association rule discovery and sequential pattern discovery. This is also referred to as unsupervised learning.

The development of data mining techniques over the years has been more expertise oriented and problem-centred (Liao, Chu & Hsiao, 2012) where the users background, expertise and problem to be solved influence the direction of research in data mining. Sector applications of data mining are vast and include sectors such as banking and financial services, health care services, supply chain management, telecommunications (Gorunescu, 2011), customer relations management (Ngai, Xiu & Chau, 2009), electricity (De Silva *et al.*, 2011), water resources (Solomatine, 2003; Abrahart, See & Solomatine, 2008; Kalteh, Hjorth & Berndtsson, 2008; Dhanya & Nagesh Kumar, 2009) and water asset management (Babovic et al., 2002).

Data mining has tremendous potential of enabling rational, accurate and faster decision making processes within the local government sector (Bakırlı, Birant & Mutlu, 2012), which in South Africa includes water utilities. Boyle et al. (2011) demonstrates better classification of consumers, better tariff structures, identification of irrigators, and the evaluation of the effectiveness of water conservation programs through the data mining of widely available utility billing data. Similarly, Humaid (2012) applied data mining to identify water theft using billing information with better prediction rates. Wu & Liu (2017) shows that data-driven approaches are promising in real-life but require comprehensive performance evaluation to be validated and useable.

The future, therefore, will benefit from a combination of the best of two approaches of solving problem through the use of theory-driven, understanding-rich processes (the traditional approach) and the data-driven discovery process (data mining) (Babovic, 2005). Following are some data mining applications with relevance to water loss management;

2.4.1 Anomaly and Fraud Detection

Bolton and Hand (2002) reviews the successful application of numerous fraud detection techniques and technologies to detect activities such as money laundering, e-commerce credit card fraud, telecommunications fraud, and computer intrusion. Beyond these industries, the electricity industry, which has many similarities with the water industry, has also recorded a number of successful application of anomaly detection as summarised in Chauhan and Rajvanshi (2013). Depuru et al. (2011) uses Support Vector Machines (SVM) and later combines it with encoding and rule induction in (Depu*ru et al.*, 2013) for the detection of electricity theft. Cabral et al. (2008) uses Self-Organising Maps (SOM), a type of an unsupervised neural network, for both fraud and anomaly detection. Nagi et al. (2008) proposes a hybrid approach using Genetic Algorithm (GA) and Support Vector Machine (SVM) to identify potential fraud and in a related study Nagi et al. (2010) uses feature extraction using SVM for the same purpose. León et al. (2011) introduces statistical estimators that also use Generalised Rule Induction (GRI) to attach importance to variability and trend analysis of electric consumption in contrast to the traditional approach of comparing new consumption against a learned consumer's individual signature described in (Fawcett & Provost, 1997). It is, however, important to note that most of the electricity studies use fine resolution data such as that from smart metering which is not prevalent in certain localities in the water sector.

Humaid and Barhoum (2013) investigated customer water consumption behaviour at an area level (multiple households) using rule induction classification and were able to flag areas for inspection where fraud was suspected. In a related study, Humaid (2012) successfully applied k-Nearest Neighbour (kNN), Neural Networks (NN) and SVM classifiers in identifying water theft from monthly billing records with the SVM classifier outperforming the other methods.

2.4.2 Time Series Similarity Search and Pattern Discovery

The various data mining techniques applied on time series are summarised by Esling and Agon (2012) as falling into three main categories of representation techniques, distance measures, and

indexing methods. Most of the work done in time series within the water sector has been mostly limited to representation techniques and distance measures as summarised below.

Ouyang et al. (2010) uses the Euclidian distance and dynamic time warping (DTW) as similarities measures and for mining similar hydrological discharge processes and patterns illustrating that the procedure can objectively find out and model similar hydrological discharge processes and patterns. Nguyen et al. (2013b) and Nguyen et al. (2013) use a hybrid combination of the Hidden Markov Model (HMM), DTW and gradient vector filtering in disaggregating high resolution smart metering flow data into predefined water end use events. Kusiak et al. (2013) compares five data-mining techniques, namely, the multi-layer perceptron neural network (MLP), the boosted-tree (regression) algorithm (BT), the random-forest algorithm (RF), SVM, and kNN in modelling energy consumption by pumps in a wastewater treatment plant. MLP outperformed the other algorithms followed by SVM. Babovic et al. (2002) compares a scoring model to Bayesian network in an asset modelling exercise of prediction pipe burst and the scoring model outperforms the latter.

Räsänen et al. (2008) applied SOM together with K-means clustering and fuzzy logic to determine electricity usage profile that can be used to create personalised information to consumers. Iglesias and Kastner (2013) proposes a clustered-vector balance algorithm as a validation technique to compare the popular similarities distances of the Euclidean, Mahalanobis, Pearson's correlation distance and DTW in the clustering of building energy patterns. The Euclidean has better general performance while DTW was an improved alternative for specific applications. Tung et al. (2006) proposed a new approach to model the similarity search known as the frequent k-n-match which was more effective than existing techniques such as IGrid and Human- Computer Interactive NN search, which have been reported to be more effective than traditional kNN queries based on Euclidean distance. Chen et al. (2013) applies a CM-CoP framework that combines association rules, clustering and decision trees to predict telecommunications customer payment behaviours with an average accuracy of 79%.

Water consumption analysis, as used in the *"Comparative Billing Analysis"* is by nature a time -series problem that could be solved using these data mining applications, as has been implemented in other domains of study.

2.4.3 Clustering and Classification

Clustering is a process of grouping data items based on a measure of similarity and is useful in data mining for database segmentation, predictive modelling, and visualization of large data-bases (Jain, Murty & Flynn, 1999). The different approaches to clustering include statistical, fuzzy, neural, evolutionary, and knowledge-based methods applied over the four steps of pattern representation, similarity computation, grouping process, and cluster representation (Jain, Murty & Flynn, 1999). A review of some of the methods used in clustering are provided in García-Escudero et al. (2010) and Jain et al. (1999).

Classification is an inductive machine learning process of learning a set of rules from instances (examples in a training set) to create a classifier that can be used to generalize new instances (Kotsiantis, 2007). A large number of techniques have been developed based on Logical/Symbolic techniques (e.g. decision trees and classification rules), Perceptron-based techniques (e.g. neural networks) and Statistics (e.g. Bayesian Networks, Instance-based techniques) (Kotsiantis, 2007).

Clustering and classification are applicable for a wide range of data mining applications, including some techniques referred to under anomaly detection and pattern recognition sections. These methods are particularly relevant for AL assessment as a critical component of the evaluation is being able to derive errors for the different consumer categories and the different metering devices in use. Many utilities also have inadequate information relating to meter-type characteristics and the consumer categories and as such, unsupervised algorithms such as clustering could be well suited to the function of being able to derive distinct groupings of consumption profiles without requiring prior classification. Time-series clustering is a special type which is of particular interest due to its ubiquity in various areas and sectors (Aghabozorgi, Shirkhorshidi & Wah, 2015).

2.5 Data Mining Techniques

This succinct review of data mining techniques relates to alternative assessments methods of AL that are investigated in this work. Although these techniques have been more widely applied in the electricity sector, they have not enjoyed similar interest in AL studies.

2.5.1 Artificial Neural Networks (ANN)

ANN are mathematical models represented by a collection of computational units interlinked by a system of connections (Cheng & Titterington, 1994) producing a mathematical relationship between input variables and output variables (Ajbar & Ali, 2012). All ANN prediction models are of the form (Maier et al., 2010):

$$Y = f(X, W) + \varepsilon$$

Equation 2.13: Artificial Neural Network Model Function

Where Y is the vector of model outputs; X, the vector of model inputs; W, vector of model parameters (connection or bias weights); $f(\cdot)$ functional relationship between model outputs, inputs and parameters; ε, vector of model errors. The form of the functional relationship $f(\cdot)$ is governed by the network architecture (e.g. multi-layer perceptron) and geometry (e.g. the number of hidden layers and nodes, type of transfer function) (Maier et al., 2010).

2.5.2 Support Vector Machine (SVM)

SVM are a special type of a feed-forward neural network extensively used in pattern classification problems and non-linear regression (Gorunescu, 2011). While similar to ANNs, they have better generalization abilities and require smaller training sets (Mutikanga, Sharma & Vairavamoorthy, 2013). SVM classifies data with different class labels by determining a set of support vectors that are members of a training input set representing a hyper plane in a feature space (Chou, Hsu & Lin, 2014). The input is first mapped, using non-linear mapping, onto an *m*-dimensional feature space on which a linear model of the form below is constructed (Chou, Hsu & Lin, 2014)

$$f(X, W) = \sum_{j}^{m} w_j g_j(x) + b$$

Equation 2.14: Support Vector Machine Function

Where, $g_j(x)$ is a set of nonlinear transformations and b is a bias term and *j* is the set of neurons in the preceding layer.

2.5.3 Self-Organising Maps (SOM)

SOMs are another variant of artificial neural networks but are based on unsupervised learning of unlabelled data. Although there are many different versions of the SOM, all of them share the following two operations (Kohonen & Oja, 1998):

1. Spatial clustering of activity into a centre of high activity, the location of which is a function of the input pattern. This location is called the winner.

2. Adaptive modification of the input connections in the neighbourhood around the winner, such that all neural functions in this area become better tuned to the prevailing input activity and respond more strongly to it.

As an algorithm, each neuron *i* of the SOM can be represented as an *n*-dimensional prototype vector $\mathbf{m}_i = (m_{i1}, m_{i2}, \ldots, m_{in})$ where *n* is equal to the dimension of the input space. For each training step (the adaptive modification), a data sample **x** is selected and the unit $\mathbf{m}_c$ closest to it is located from the map and the prototype vectors of the best matching unit and its neighbours are moved towards the sample vector (Vesanto & Ahola, 1999):

$$\boldsymbol{m}_i = \boldsymbol{m}_i + \alpha(t) h_{ci}(t)(\boldsymbol{x} - \boldsymbol{m}_i)$$

Equation 2.15: Self-Organising Maps Function

Where $\alpha(t)$ is the learning rate and $h_{ci}(t)$ is a neighbourhood kernel (both decreasing monotonically with time) centred on the winner unit *c*.

2.5.4 Dynamic Time Warping (DTW)

The DTW algorithm is a popular method for measuring the similarity between two time series of different lengths by finding an optimal alignment between them with certain restrictions (Nguyen et al., 2013a). Given two time series vectors of different sizes **X** = (x_1, ..., xi, ..., x_n) and **Y** = (y_1, ..., y_j, ..., y_m), the distances between any two points x_i and y_j form an n x m cost matrix ***C*** which can be summarised by a warping path **W** = (w_1, w_2, ..., w_K) where $max(m, n) \leq K < m + n - 1$, formed by a set of matrix components subject to specific rules with respect to boundary conditions, monotonicity and the step size (Ouyang et al., 2010). The DTW distance is therefore the optimum (minimum) distance of the possible paths given by the equation:

$$d_w = min\left(\sqrt{\sum_{k=1}^{K} w_k}\right)$$

Equation 2.16: Dynamic Time Warping Distance

2.5.5 k-Nearest Neighbour (kNN)

The k-NN query is one of the most common queries in similarity search and its objective is to find the k nearest neighbours of points in horizontally partitioned data (Yin et al., 2011). The proximity of the nearest neighbours is typically defined using a distance measuring function such as the Euclidean distance. Given sample data (x_1, x_2, ..., x_n), the k-NN method searches for k point centres (m_1, m_2, ..., m_k) that minimises the problem (García-Escudero et al., 2010)

$$arg \min_{m_1,\ldots m_k} \sum_{i=1}^{n} \min_{j=1,..,k} ||x_i - m_j||^2$$

Equation 2.17: k-Nearest Neighbour Function

2.5.6 Hidden Markov Model (HMM)

HMM is a stochastic finite state automation model that is defined by the parameter λ = (π, *a, b*), where π is an initial state probability, *a* is the state transition probability and *b* is observation probability, defined by a finite multivariate Gaussian mixture (Nguyen et al., 2013b). For any observed sequence **X** = ($x_1, x_2, \ldots, x_t$), the probability of **X** given the parameter λ, $P(\mathbf{X} \mid \lambda)$, is the HMM model which can also be used to determine the corresponding state sequence that maximises the probability of **X**, $P(Q \mid \mathbf{X}, \lambda)$.

2.5.7 k-Shape Clustering

This is a form of partitional clustering algorithm created by Paparrizos & Gravano (2015) consisting of a custom centroid function (shape extraction) and a custom distance measure (shape-based distance, SBD). The function is stochastic in nature and requires *z*-normalization of the input data. SBD is based on the cross-correlation with coefficient normalization (NCCc) sequence between two series, and it is thus sensitive to scale, hence the z-normalisation requirement. SBD is given by the formula (Sard, 2015);

$$SBD(x, y) = 1 - \frac{\max\left(NCCc(x, y)\right)}{\left|\left|x\right|\right|_2 \left|\left|y\right|\right|_2}$$

Equation 2.18: Shape Based Distance – k-Shape Clustering

where $||\cdot||_2$ is the Euclidean norm of the series. Shape extraction relies on NCCc and uses it to match any two series optimally with the centroid series selected in random. Alignment can be done between series with different lengths, but the length of the resulting prototype will depend on the length of the chosen reference centroid series. K-Shape has been shown to outperforms all state-of-the-art partitional, hierarchical, and spectral clustering approaches in the computational reproducible study of Paparrizos & Gravano (2015).

2.5.8 Fuzzy Clustering

Fuzzy clustering does not create “crisp” clusters but rather fuzzy or soft partitions in which each member belongs to each cluster to a certain degree. This is a particularly useful property in water consumption analysis as it is likely that a consumer may belong to a few clusters at the same time. Common fuzzy clustering implementations utilise fuzzy c-means described in Bezdek (1981) and the Euclidean distance as a distance measure. Defining $\mu_{c,i}$ as the *i*-th element of the *c*-th centroid, and $x_{p,i}$ as the *i*-th data-point of the *p*-th object in the data, the centroid function is expressed as (Sard, 2015);

$$\mu_{c,i} = \frac{\sum_{p=1}^{N} u_{p,c}^{m} x_{p,i}}{\sum_{p=1}^{N} u_{p,c}^{m}}$$

Equation 2.19: Fuzzy Clustering Centroid Function

From Equation 2.19 the time series data will be required to have the same dimensionality, which could be a challenge for water consumption data as various time-series will have different lengths. This could be handled through the interpolation of the series to match the longest time-series assuming the individual series retain their statistical properties. The non-crisp partitions of the algorithms could also be made crisp by taking the maxima of each cluster when required.

2.6 Summary

This chapter has outlined the previous work of many authors who have been involved in apparent loss assessment and related fields, including the work of other fields which could be relevant in addressing the objectives of this research. From the literature review, it has been shown that AL assessment has enjoyed far less attention both from an application point of view and in academic literature. This is demonstrated by the prevalence of estimates for this component of water losses using a variety of usually unsubstantiated methods. There is therefore a clear case for contributing and generating data and evaluating the assessment of apparent losses using different methodologies and perspectives.

With respect to the field and laboratory-based work on apparent water loss assessment that includes the determination of meter accuracy and consumer profiles to estimate the weighted meter accuracy, it has been shown that all but one of the studies were done in the developed world that have distinct set of consumption patterns, metering standards and cultural norms which all influence AL. The one study done on the African continent, while contributing to the body of knowledge, has also been shown to have some gaps in methodology and the number of samples involved. This underscores the need of implementing field and laboratory-based assessment of apparent losses within a developing nation in Africa, to enable the generation and application of contextually relevant results within the region.

There is consensus amongst researchers that the current assessment methods of AL are both laborious and costly to implement, and there is need for equally reliable but easier and cheaper alternatives. Such alternatives are however yet to be developed. The alternative methods reviewed, while they show promise, are yet to be validated with different datasets. There is therefore a need to evaluate such proposed alternatives, as currently described, to see how well they match best practice assessment methods. It is envisaged that qualifying methods with greater potential will require further development into a fully-fledged, repeatable and validated alternative methodologies that can be used as viable alternatives.

The review of current metrology regulation highlighted the need for the ongoing verification of water meters in the field. This usually requires the ongoing testing of meters which has been shown to be expensive. The development of alternative validated methods that are cheaper and faster could be of great benefit in meeting these legal requirements that are currently not being adhered to.

Lastly, data mining applications in other sectors and subject areas have also been reviewed and they have been used to solve similar problem with respect to the problem definition. What remains is the application of such methodologies within the AL domain in developing viable alternative assessment methods from available data.

The chapters that follow detail the work implemented as part of this research work to address the highlighted gaps and in the process develop viable alternatives that can be widely applied within the water sector.

Chapter 3 Consumption Characterisation

This chapter is based, in part, on the publication *"Consumption Characterisation and On-Site Leakage in Johannesburg, South Africa"* presented at the IWA Water Loss 2016 Conference in Bangalore, India and referenced as Ncube & Taigbenu (2016).

The process of deriving the weighted meter accuracy, illustrated in Figure 2-4, involves the determination of the consumption profiles of a representative sample of the consumer population. This is the focus of this chapter and includes the details of the preparatory work of defining the consumer sample, the selection and sourcing of equipment, the actual site work together with important site observations and lastly the analysis of the data.

3.1 Preparatory Work

3.1.1 Consumer Sampling

Johannesburg Water reads its water meters monthly and the actual read meters for any month are based on an agency agreement with the City of Johannesburg which is responsible for consolidated billing of all services. As a result, this number always varies from month to month and the actual consumers in any month are not always the same. This is illustrated in Figure 3-1, which was only available from November 2009.

The period leading up to June 2010 coincides with a billing system changeover that affected several activities such as meter reading, resulting in the lowest recorded number of read meters. An evaluation of the highest number of consumers with active meters in various consumption categories over a six-month period was done to obtain an indication of the total number of consumers per category. The categories were as defined within the billing system but were reclassified into four for simplicity

comprising the business, residential (single dwelling), public benefit organisations and multi-residential categories. There was however limited information on household properties that could be used for finer disaggregation of consumers into homogeneous segments as suggested in in the literature. The method and table described in Krejcie & Morgan (1970) was utilised to determine the required sample size. The 95% confidence level, with an error margin of ±10% per consumer category and an error margin of ±5% city-wide was adopted due to budget constraints. Table 3.1 summarises the number of consumers per category and the targeted sample quantities for meter logging for each category.

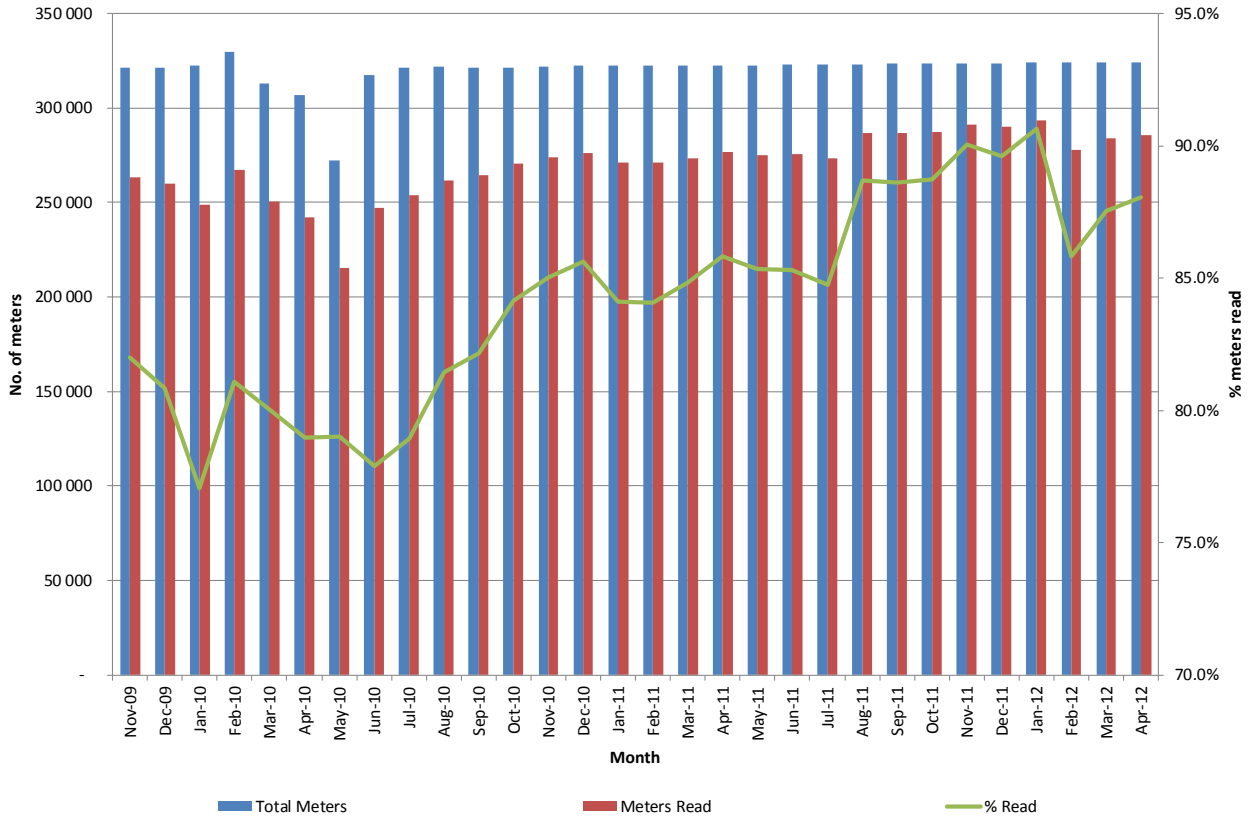


Figure 3-1: Meter Reading Patterns

Table 3.1: Water Consumption Sample Selection

Reclassified Consumption Categories	*No. of Consumers*	*Target Sample Size*
Business (Bus)	40,816	96
Single Residential Dwelling (Res)	325,020	97
Public Benefit organisations (Utility.)	3,014	94
Multiple Residential Dwelling (Multi)	9,405	96
Total	***378 255***	***383***

To account for unforeseen field challenges there was an oversampling by about 20% and the sampling was performed in R, a statistical program (R Core Team, 2017) that had been selected for statistical analysis within the study. The two categories that were of concern were the business category and the public benefit organisation as these would likely not be adequately homogeneous. From the author's experience with some of the consumers it was known that business properties were largely office buildings while public benefit properties included government offices, religious properties, and schools, among others. A record was kept on site in order to ascertain that the categories of the properties visited coincided with the filed records. The location of sampled properties is shown in Figure 3-2, which are dotted throughout the city.

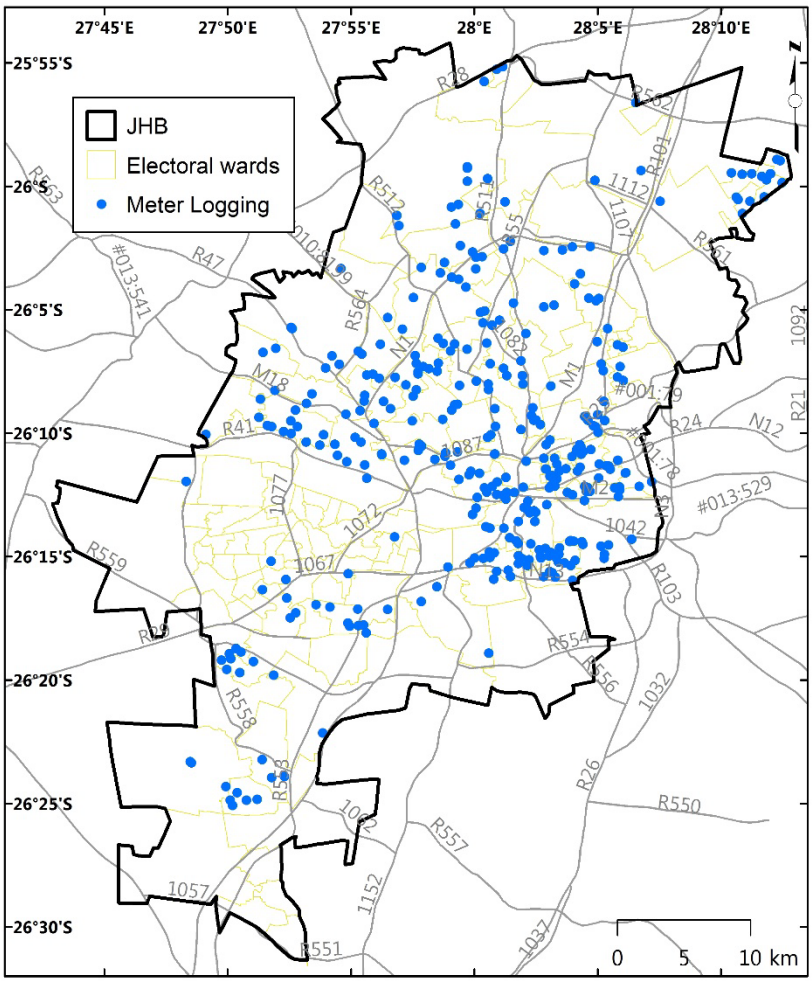


Figure 3-2: Location of Sample Points for Consumption Characterisation

3.1.2 Logging Equipment Selection

Consumption profiling requires the capturing of consumption data at a very high resolution with high levels of accuracy. On the other hand, security of the logging equipment and meters was also of considerable concern as theft and vandalism are prevalent within the study area. A secondary consideration was the compatibility of the equipment with the meters on site. Inspection of some of the sampled meters was done to evaluate the status of the meters to be able to make an informed procurement decision.

Bulk meters with nomimal sizes ranging from 32mm to 150mm and with flanged connections generally come with a pulse output as standard. As such the only requirement to obtain a flow profile is to connect such a pulse output using a meter-specific reed switch or pulse reader to a data logger. A summary of all the meter models on the meter database was compiled and the most appropiate and most accurate of the pulse readers were ordered from the relevant meter model supplier. There were four main manufacturers with each having about three different meter models. Sensus Cosmos Data Loggers (CDL) and its accompanying software, CDLWin, were selected and purchased as the prefered data logger. The memory of the logger was adequate for the storage of at least seven days of consumption data for the majority of the properties. There were however some bulk meters that could not be logged from the sampled meters as they did not have pulse outputs while a few others were too old such that they were either unreadable or suscepted to be highly inaccurate.

The small bore meters (15 – 25mm threaded connections) were much more complicated. The bulk of those initially visited either did not have a pulse output or were Class B meters which were not adequate for this exercise due to higher starting flows. In some instances some connections were oversized, and therefore would also give misleading results. As a result, it was necessary to purchase highly accurate meters to ensure the integrity of flow data that are collected. Taking into account the number of small meters that has to be sampled, it because obvious that it would take a couple of years to complete to sample these meters if the same approach adopted for bulk meters of using the same loggers on more accurate meters was pursued. Furthermore, the majority of the existing meters comprised of above ground connections in meter boxes that had limited internal

space to house an additional logger inside. This therefore meant that the logger could not be inserted within the meter box and having the loggers external to the meter box would have compromised the security of the instrumentation. Therefore, highly accurate meters with internal logging and adequate memory for data storage for at least seven days and that would not interfere with the measuring ability of existing meters were sought from prominent local suppliers. The best alternative was a solid state Sensus iPerl smart meter which has an MID approval as an R800 (Q_3/Q_1 ratio) meter, shown in Figure 3-3 with its almost flat error curve. It has no spacing requirements before and after the meter and does not influence the flow measurement of any in-situ mechanical meters. The choice of a meter had a number of implications that included the cost of having a contractor install and remove the meters after the requisite logging period.

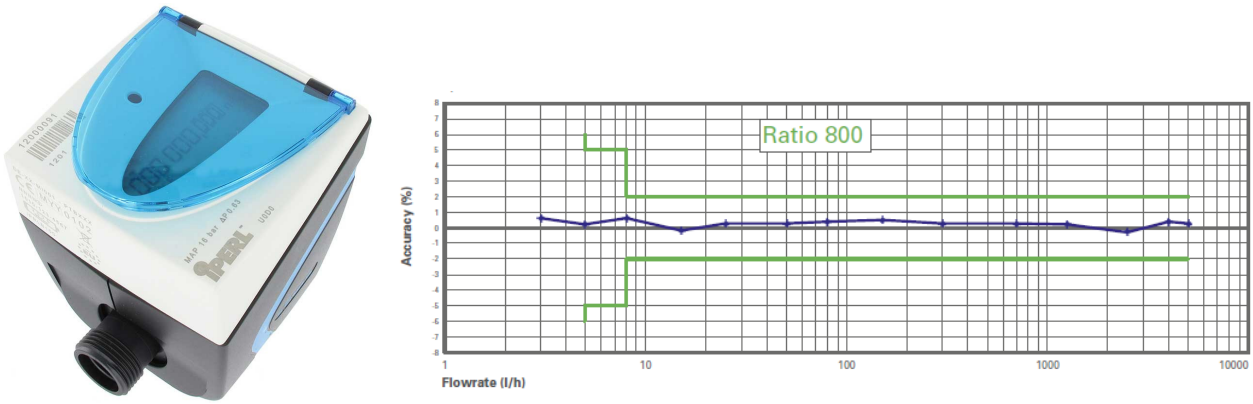


Figure 3-3: iPerl Meter and its error curve

Source: Sensus Meters (South Africa)

The iPerl meter has a minimum flowrate (q_{min} or Q1) of 3.13ℓ/h, compared to 11.25ℓ/h for a Class D meter and a starting flowrate of 1ℓ/h, and with the capability to emit data every 15sec to a receiver from either of its two radios. However, despite its high accuracy, its data logger is not as equally impressive with a minimum logging frequency of 1min and a data storage capacity accommodates meter reading of about 4.2days. While there is a functionality that combines the flow variation and the reading, its use reduces by almost a half the time for the logger capacity to be exceeded, implying that at least four visits will be required to each property to achieve the compilation of consumption data of 7 days.

After considering all the pros and cons, and further evaluation of typical data obtained from the meter (as discussed in greater detail in §3.3), the iPerl was chosen as the most optimum choice under the circumstances and a total of 50 meters were obtained for this exercise. As such, for each of the small meter properties, two visits post installation were done in order to collate a minimum of seven continuous days of data, thereby increasing the travelling time and effort required.

3.1.3 Logistics Planning

Due to the sample size and the randomness of the location of the property, some concerted planning had to be done in the office prior to any site visit. This included dividing the city into smaller areas to be done in stages and matching the available equipment with the meter models on site for each stage for the bigger meters. Thereafter, there was also a need to use route planning and optimisation software to minimise traveling time, and a Garmin Nuvi 2495 was used for this purpose.

The smaller meters were however more complicated in that the coordination of the contractor also had to be factored in and, in some instances, it was not possible to work within predetermined areas hence a greater need for route optimisation. As an illustration, to cover about 30 meters in a single day, which was the most the contractor could do, would normally require 250 - 300km of travelling without optimisation. Route optimisation was able to reduce this distance to about 150km in certain instances, also considering peak traffic in the various areas.

3.2 Site Activities

Meter logging was implemented during the period July 2014 – November 2015 and had to be split between small and large meters. The total number of meters that were eventually logged is summarised in Table 3.2. For the purposes of the study a minimum of 7days was the adopted logging period and deemed as sufficiently representative of the consumption of the consumers. The variation in consumption flow distribution (not volume) between seasons was also assumed to be minimal. The occurrence of on-site tanks that have been reported elsewhere in literature (Arregui et al., 2006;

Mutikanga et al., 2010; Mutikanga et al., 2011) is very limited in Johannesburg and is therefore not evaluated in this study.

Table 3.2: Successful Meter Logging Summary

Meter Type	Logging Period	Consumer Category				Total
		Business	Utility	Residential	Multi-Res	
Large Water Meters	July 2014 – Nov 2014	16	-	-	30	46
Small Water Meters	May 2015 – Nov 2015	68	76	141	77	362
	Totals	***84***	***76***	***141***	***107***	***408***

3.2.1 Bulk Meter Logging

A total of 102 large Class B meters were logged by installing an appropriate reed switch or pulse reader connected to a data logger on to an existing meter, as shown in Figure 3-4.



Figure 3-4: Meter Logging Setup for Large Meters

The logging interval was pulse-based ranging from 1 pulse per litre to 1 pulse per 10 litres, depending on the specific meter model. This variation in the litres per pulse for the logging was a concern as the basis for comparison was not the same. Furthermore, the logged bulk meters are generally Class B meters that are expected to have poor recording ability of low flows, hence obtaining a pulse every 10 litres was not necessarily the level of detail required.

When setting up the loggers, the current meter reading was used as the starting point of the logging and on removal the integrity of the logging data was also verified against manual readings from the meter. Because of these checks, it was realised that a set of subcontracted loggers were problematic and gave erroneous results. Furthermore, several meters did not register any flow over the entire

7day period because they were either faulty or were in fact metering a fire connection but were not flagged or marked as such on the billing system. In addition, there were meters that could not be logged because of site conditions such as the meter being not readable, meter chamber being flooded or some other obstruction within the chamber, with examples shown in Figure 3-5. This resulted in only successfully logging 46 meters out of 120 meters that were visited, a relatively low success rate.



(a) Incorrectly installed meter with obstructions



(b) Rock obstructions



(c) Class A meter for a multi-residential unit, with no pulse output



(d) Damaged meter with no glass and misaligned register

Figure 3-5: Examples of meters that could not be logged

One major challenge encountered was the inaccuracy of meter records relating to the meter make, model and size and the proper categorisation of each connection. This made it very difficult to properly plan the field work and led to repeat visits. Such omissions substantially increased the effort and the cost of implementing consumption characterisation, and they revealed the state of poor

management of the meters. Additionally, some meters did not have any appropriate output device for meter logging and were therefore also omitted – this is despite the metering standard of Johannesburg Water that expressly required all meters to have at least one output for meter logging. A few Class A meters, which are not allowed for custody transfer, were discovered during the evaluation exercise and were not included in this exercise as they did not have pulse outputs (see Figure 3-5(c)) and would have been inaccurate at lower flows. About 50 Class A meters were unearthed in the process and were accordingly reported and removed at the manufacturer's expense. These incidences points to poor management of meter and in particular the processes related to the selection and installing of meters. Security for the data loggers was also problematic, with a few of the data loggers being stolen or tampered with during the exercise.

3.2.2 Small Meter Logging

The iPerl meter was installed in series with an existing meter as shown in Figure 3-6 and left for about 7 days after which it was removed and moved to the next property on the sample list. By outsourcing the installation, removal and reinstatement of meters, the task became easier as some of the risks and inconveniences were transferred to the contractor.



Figure 3-6: Meter Logging for Small Meters

The lessons of logging the bulk meters also came in handy as several checks were introduced that included ascertaining the property ID irrespective of the meter record to ensure that the right property was logged. The use of an additional logging meter independent of the existing meter also meant

that logging would commence regardless of the state of the connection or the meter, including damaged meters of which a few were found and reported.

In order not to confuse the meter readers during the logging period, despite having communicated extensively, it was necessary to cover the iPerl meter so that is was not in clear sight for its reading being mistakenly taken. The meter was therefore buried in most instances or placed in a covered underground meter box. The radio function of the meters was quite strong for both downloading the meter reading range and the logger data – the latter required a strong connection due to the data packet size.

The data was downloaded at most every 4.2 days, with the second visit done when the meters were being removed to be transported to another site to ensure a minimum of 7 days was recorded. On average, the logging interval was 8.4 days. Due to the challenge with bulk meters, the sample size of the smaller meters was increased to make up for the large quantities of these meters. It was however not possible to control in detail which of the properties to visit within the provided sample as there were several faulty meters found on site, refusal of access by some consumers, and the fact that at any given point the 50 available logging meters needed to be used. This meant the contractor also had to adapt as he went along leading to the higher number of domestic properties, as shown in Table 3.2, as these were the most prevalent and the easiest to locate and install a meter.

Similar to the bulk meters, there were several incorrect installations of small meters, although to a lesser extent as most new meters came in pre-assembled meter boxes from the suppliers. A few of the meters were damaged and as such their readers were not legible, indicating inadequate meter management.

3.3 Smart Metering Consumption Data

As part of the process of selecting an appropriate meter for logging purposes (described in §3.1.2) consumption data from properties of where the iPerl meters were in use was provided for evaluation. The main concern was whether by selecting to use the iPerl meter at a logging frequency of 1min compared to finer resolutions would differ substantially when compared with the approach of other

studies of using pulse-based meters at a resolution of a pulse per 0.1l and logging intervals of 10 – 15 sec. The provided data was from a fixed radio setup that read the meters every 15 seconds, which is a similar resolution to other studies in terms of time. The evaluation was therefore to check if there is a substantial difference in consumption patterns between a logging resolution of 1min compared to a 15sec interval. A random sample of the provided residential properties was taken, and the data analysed using the 15sec time step and thereafter a 1min time step. While there we subtle differences, they were not statistically different and considering that the objective is to calculate the weighted meter error, such subtle differences did not matter.

The local meter manufacturer and the relevant property managers of where the iPerl meters were utilised thereafter provided permission for the use of the complete dataset for this research. This comprised of a total of 278 consumption records of sub-metering data with a length ranging from 70 days to 235 days from Johannesburg and Cape Town. The categories of consumers in the dataset are shown in Table 3.3 of which 256 records were used.

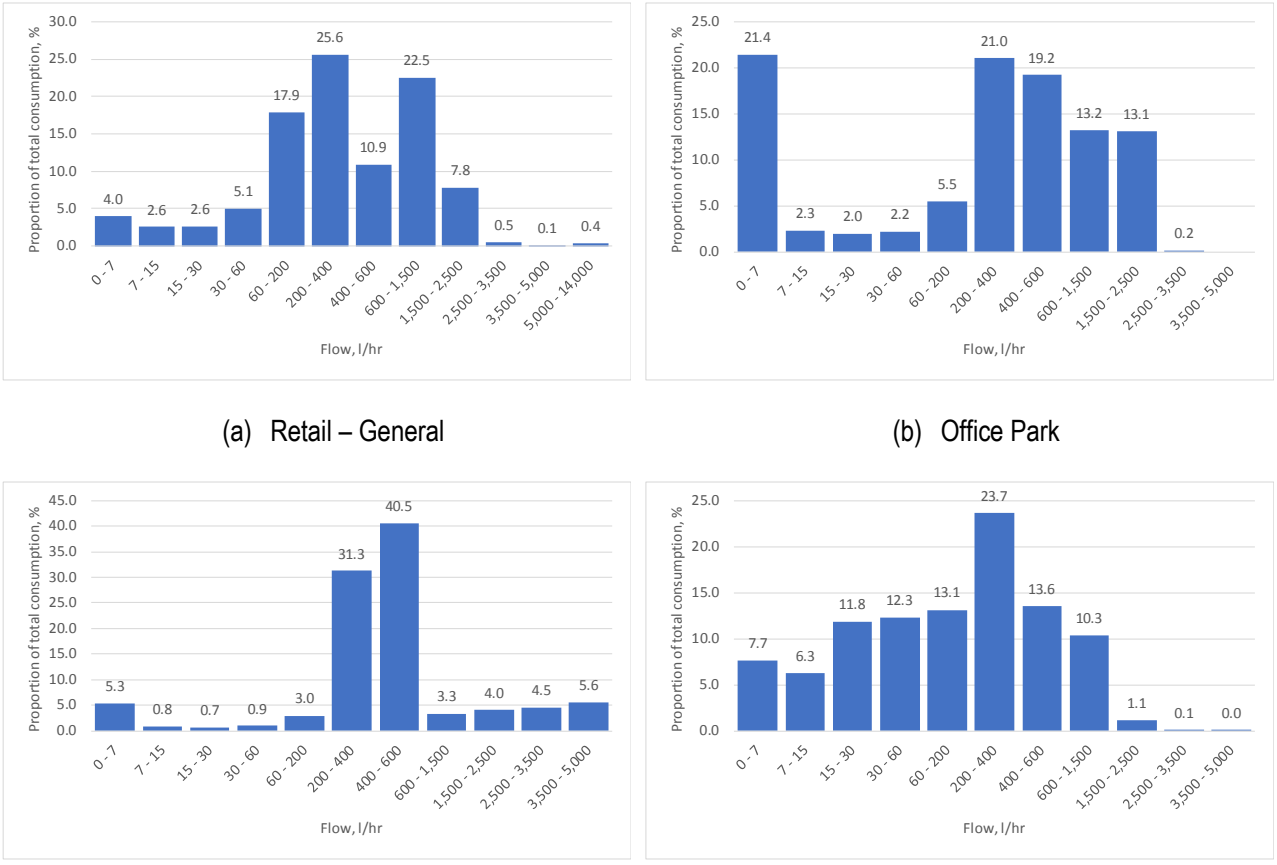
Table 3.3: Smart Metering Data

Utilised Dataset		Discarded Dataset	
Category	**No. of**	**Category**	**No. of**
Office Park	14	Residential - Students Toilets	2
Residential	21	Retail - Aircon	5
Residential - Students	77	Retail - Cellular	7
Retail - General	144	Retail - Financial	4
		Retail - Toilets	4
Totals	***256***		***56***

Consumption profiles were determined by finding the proportion of flowrates between specified values for each of the 256 records and aggregating the data to come up with four consumption profiles shown in Figure 3-7.

The retail consumers are uncategorised shops in shopping malls that use their water mainly for kitchen and toilet facilities. There is a limited amount of consumption at low flows for this type of consumer, with the bulk of the flow being in the medium flowrate range. The office park consumers

have a peculiar pattern with more than 21% of the flow occurring at very low flowrates of up to 7ℓ/h, which is most likely due of leaks. There is evidence of occurrences of night flow in offices that are not expected to have any water use at that time, thus confirming this. The other low flows up to 60ℓ/h is in the similar ranges as retail shops, which would be expected as both are in a work/office environment with similar usage types.



(a) Retail – General

(b) Office Park

(c) Students Accommodation

(d) Residential

Figure 3-7: Consumption Pattern: Smart Metering Data

However, as expected there is a wide variation from meter to meter for the office category due to the relatively small sample size. The student consumption profile has a peculiarity of being dominated by mostly medium to high flow rates when compared to the other categories. There is however still some flow at the low end which could also be indicative of leakage.

The residential consumers in this dataset consist of individual units in an up-market townhouse complex in Johannesburg. Despite the small number of households there is very little variability in the data and the profile has distinct characteristics with consumption spread throughout the meter range.

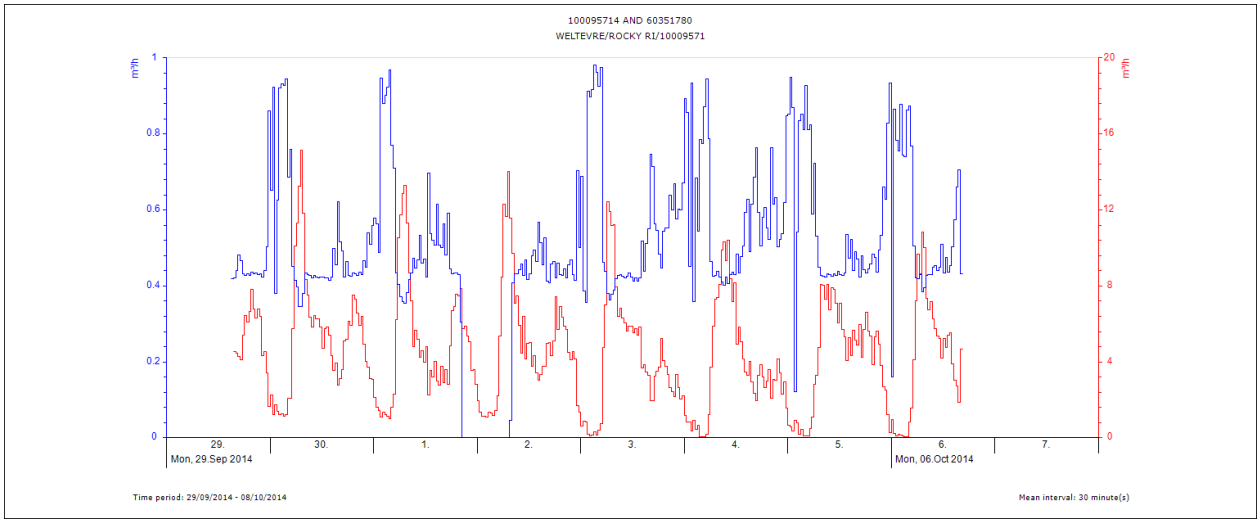
This segment has 25.8% proportion of flow in the region up to 30ℓ/h, which when compared against Class B meters commonly in used within Johannesburg Water in 15mm to 25mm sizes (q_p of 1,500ℓ/h to 3,500ℓ/h) clearly demonstrates that a Class B 15mm is not the most appropriate for this category of consumers as its minimum flow is 30ℓ/h. There are also some clear differences between this profile with those of Figure 2-6 for other parts of the world, which also reinforces the need to evaluate consumption patterns locally.

The residential and office park results are only indicative due to the small number of samples while the student accommodation and retail consumers with adequate sample size permit the use of the data in its current form. Such data are useful for sub-metering entities who typically deal with such consumers in South Africa, and for utilities in setting up appropriate standards that consider the variation of flow as well as in selecting bulk meters for such consumers.

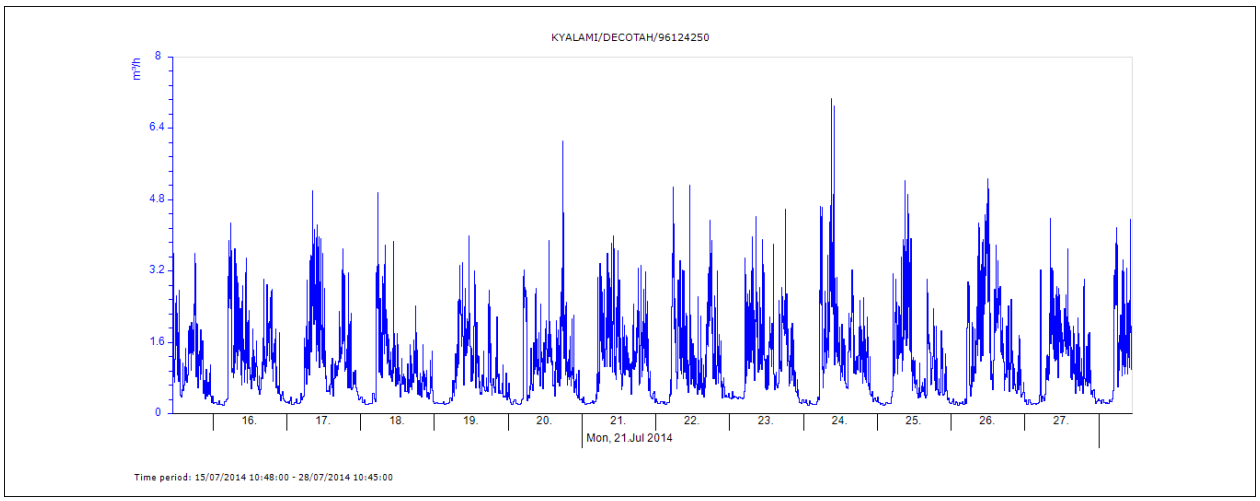
3.4 Consumption Patterns

3.4.1 Bulk Consumers

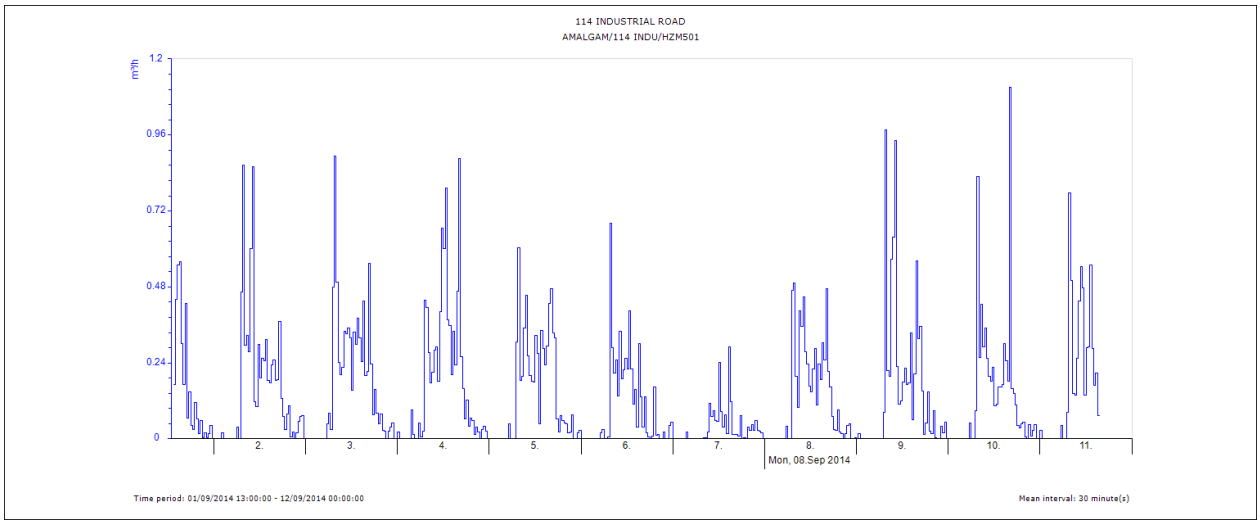
Examples of the meter logging results are shown in Figure 3-8. Without much analysis it was immediately apparent that there was a very high prevalence of oversized meters and on-site leakage from observing the output such as (b) and (c) from meters with a permanent flowrate greater than 40m^3/hr. In fact, 56% of the logged meters were of at least nominal diameter of 100mm with $q_p \geq 180$m^3/hr. Only 20% of the logged meters had a q_p of 40m^3/hr (and q_s of 80m^3/hr), which preliminary observations showed could be the most appropriate meter for most of the properties. Fire flow requirements were cited by the utility staff as one of the reasons why the meters were that big but this however does not justify sacrificing the proper measure of continuous flow for a fire event that has less chance of occurrence. Furthermore, there are better available options for catering for such a requirement such as have separate connections for fire and consumption or sizing meters to fail during a fire event. Another observation was that the meters were generally sized according to the connection size as opposed to evaluating the required flowrate.



(a) Combination Meter



(b) Oversized Meter and On-site Leakage



(c) Oversized Meter and no leakage

Figure 3-8: Examples of Bulk Meter Logging Results

That as it may be, the utility failed to use other options available such as low flow meter for a 40mm connection such as a threaded 40mm meter with a q_p of 10m^3/hr as opposed to a flanged 40mm with a q_p of 40m^3/hr.

Figure 3-9 summarises the consumption distribution for both the business and multi-residential consumers, together with the uncertainty of the proportions of flow. As expected, due to the small sampled numbers, particularly for the business consumers, there is low confidence in the consumption distribution for this category of consumers. There is also less than 1% of flow occurring below 0.06m^3/hr and above 40m^3/hr for both category of consumers.

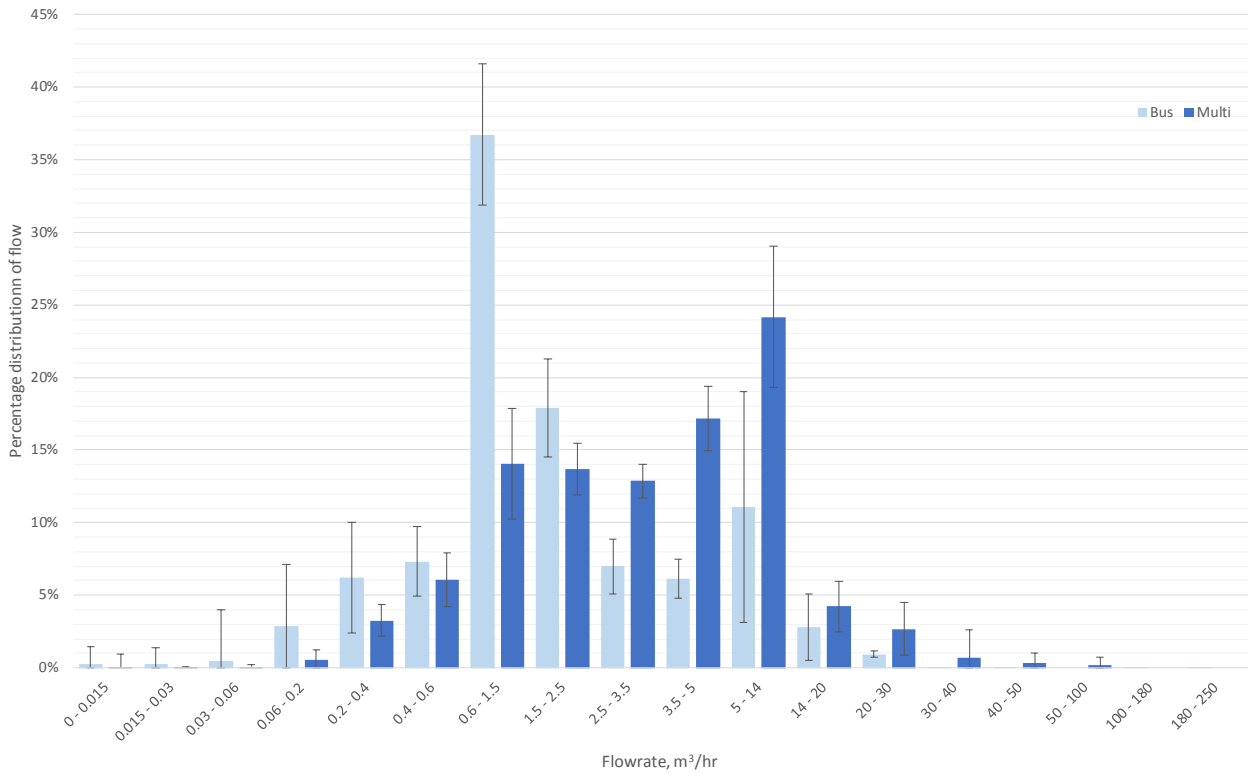


Figure 3-9: Consumption Distribution for Large Connections.

It is however very likely that the low flows are understated due to meter oversizing and the progressed age of some of the in-situ meters. This therefore implies that the consumption profile is only indicative and should be considered as such. The limited flow rates above 40m^3/hr also further reinforce the observation that the bulk of the meters were oversized with considerable scope for the use of 32mm and 40mm meters with range of q_p values from 5m^3/hr to 20m^3/hr, which are not common

within Johannesburg. The flow of the business consumers is also more skewed towards smaller flows compared to the multi-residential consumers, which is expected as these businesses are generally office accommodation with not water intensive activities nor do they have significant high flow rate water uses.

Further analysis of the multi-residential consumers shows that there is a possibility of using the results to better size meters for comparable properties. Figure 3-10 shows the variation of the recorded flowrates with the number of units in residential complexes which follows a linear trend. Despite the small sample size, there is meaningful relationships which suggest the viability of the approach. Obtaining further meter logging results for multi-residential properties, which are prevalent within Johannesburg and other areas in South Africa, will facilitate further development of a methodology to size meters with possible incorporation of property sizes and values and the use of the probabilities of exceedance with some other appropriate benchmarks.

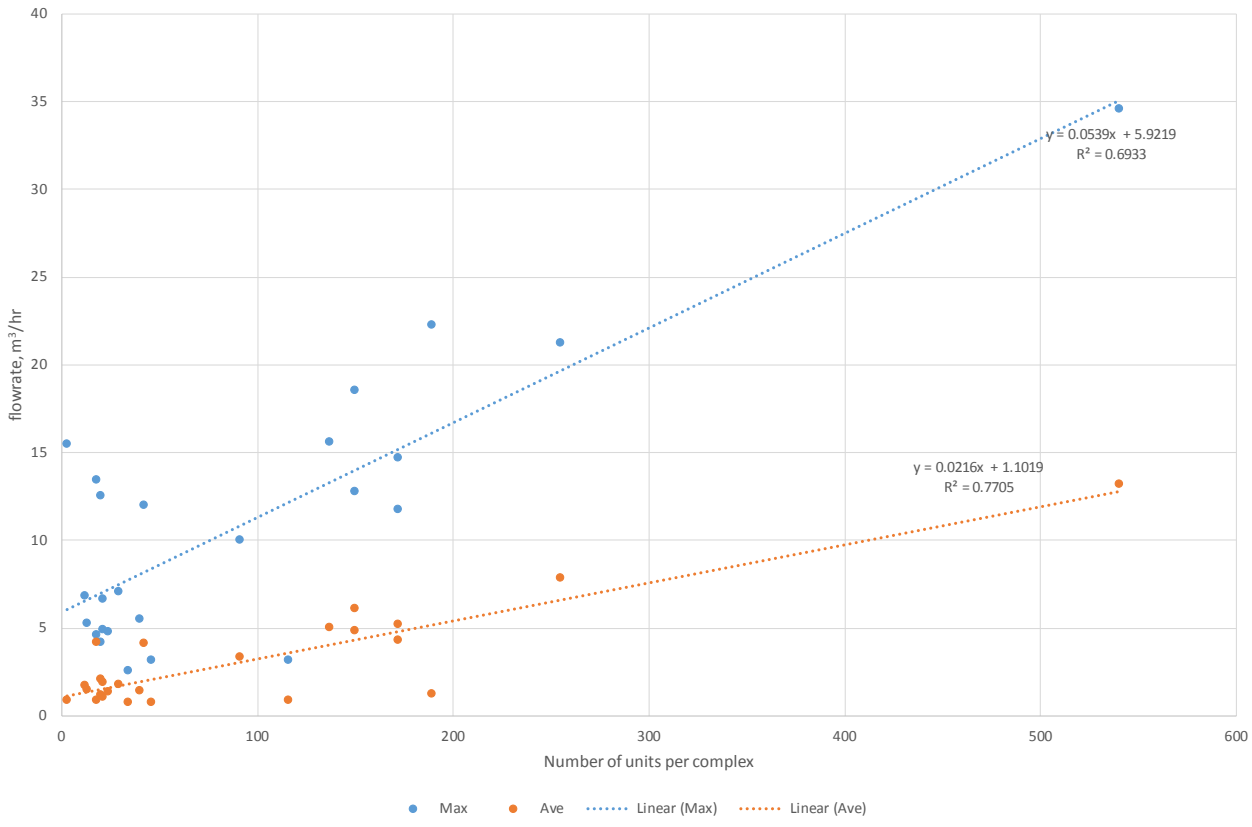


Figure 3-10: Variation of Flowrate with Units in Residential Complexes

Another practice unearthed during the exercise and other engagement with utility staff was that bulk meters were being sizing using the fire flow requirements as the permanent flowrate, q_p, of the meter not the maximum flowrate, q_s, as would be expected. The fire flow can never be the realistically the permanent flow but rather the maximum flow, if and when required.

3.4.2 Small Connections

Several assumptions were made with respect to the analysis of the recorded data for small connections. All recorded flows were presumed to have occurred within the minute of sampling as it was possible to have several events happen within a minute. Secondly, because of the prior assumption and the fact that the starting flowrate of the meters used was 1ℓ/h, another check was made of the calculated value to ensure that if it is less than 1ℓ/h it is discarded as it would not have been feasible to read. The impact of these assumptions was evaluated and found not to eliminate a significant amount of data nor produce undesirable results that impacted the results negatively.

Figure 3-11 summarises the consumption distribution for consumers with small meters. Residential consumers show a considerably higher proportion of consumption at low flows compared to other categories with 14.5% of the flow occurring below 15ℓ/h, the q_{min} for a Class C 15mm meter. An additional 15% is below 30ℓ/h meaning that about 31% of flow is below the q_{min} of a Class B 15mm meter. There is very little flow that occurs above 1,500ℓ/h, which together with the low flows, clearly discourage the use of 20mm meters of any class which are unfortunately common within the Johannesburg system. Although such meters will register some of the low flow, the fact that it predominantly falls outside the regulated part of the meters' error envelope is undesirable. The residential consumption patterns are also like the residential category obtained from the smart metering data plotted in Figure 3-7, which is expected.

The utility segment comprises government buildings, religious organisations, schools and other public organisations. In some case the meter logged was part of a few meters for the property and therefore not representative of the entire property. This is therefore clearly not a homogeneous grouping of properties as would typically be required for such exercises. However, for the sake of

completeness and to create a consolidated profile that can describe an average Johannesburg consumer this category was included. The reader must therefore only consider the utility category results as purely indicative and not conclusive due to the lack of homogeneity.

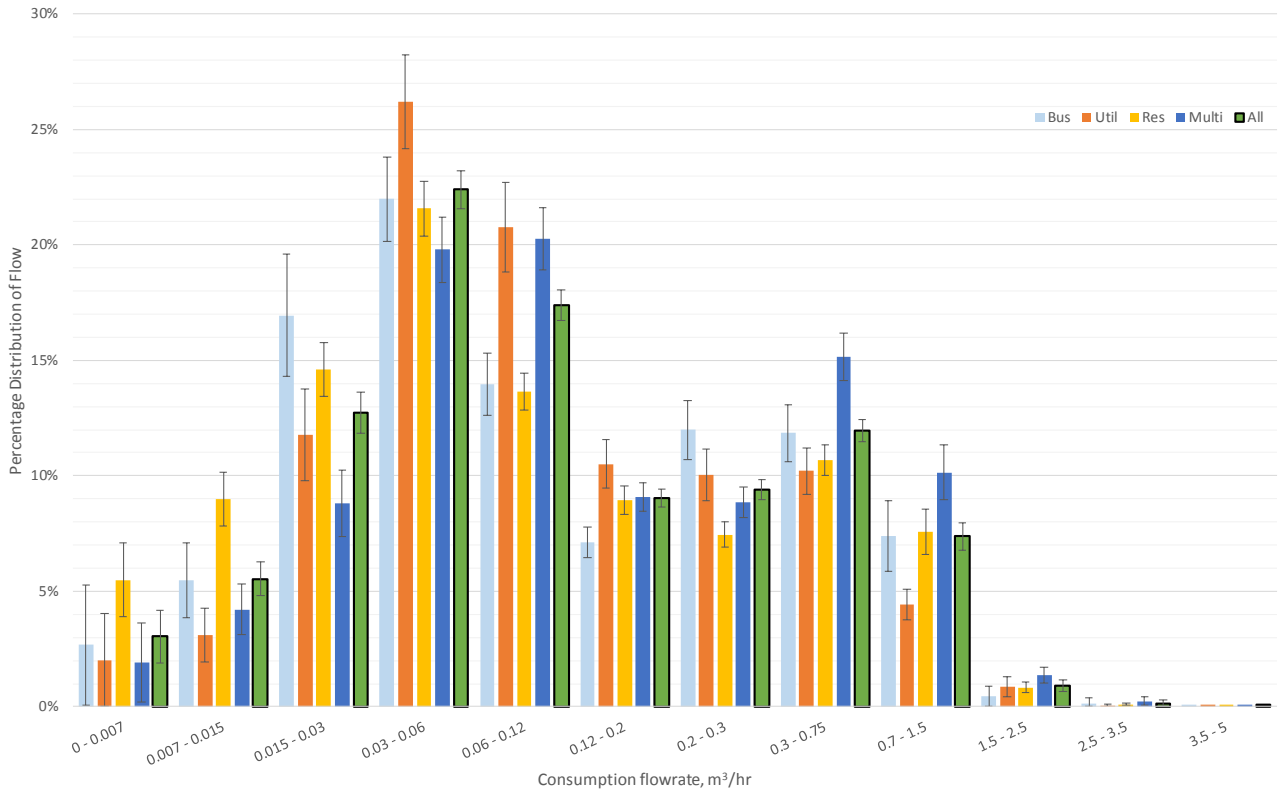


Figure 3-11: Consumption Distribution for Small Connections

The multi-residential consumers profile is very similar to those from the bulk meter logging exercise in terms of description with the only difference being the use of a smaller meter (usually 25mm, q_p of 3.5m^3/hr, and q_s of 7m^3/hr) serving up 24 units. Up to 43% of the flow in the multi-res consumers occurs below 60ℓ/h compared to the 1% for the bulk multi-res consumers, and this further highlights the high probability of under recovering from the oversized bulk meters. Similar comments can also be made concerning the business category. There are obviously exceptions to this rule where meters are properly sized with the use of combination, as shown in Figure 3-8(a).

Figure 3-12 shows the further disaggregation of the residential consumption by average annual daily demand (AADD), as recorded on the billing database. The results show an expected trend that the lower the AADD the higher the proportion of flow at low flowrates, and vice versa. The results also

indicate that the low flow measurement requirement is of greater significance for consumers with lower AADD values than for those with higher AADD, and conversely for the measurement requirement at higher flow rates. For example, consumers with an AADD of 0.5kℓ/day have over 31% of their flow at less than 15ℓ/h while at least 62% of the flow is below 30ℓ/h. This emphasises the requirement for due care when dealing with consumers with low AADD to ensure as much of the flow is recorded.

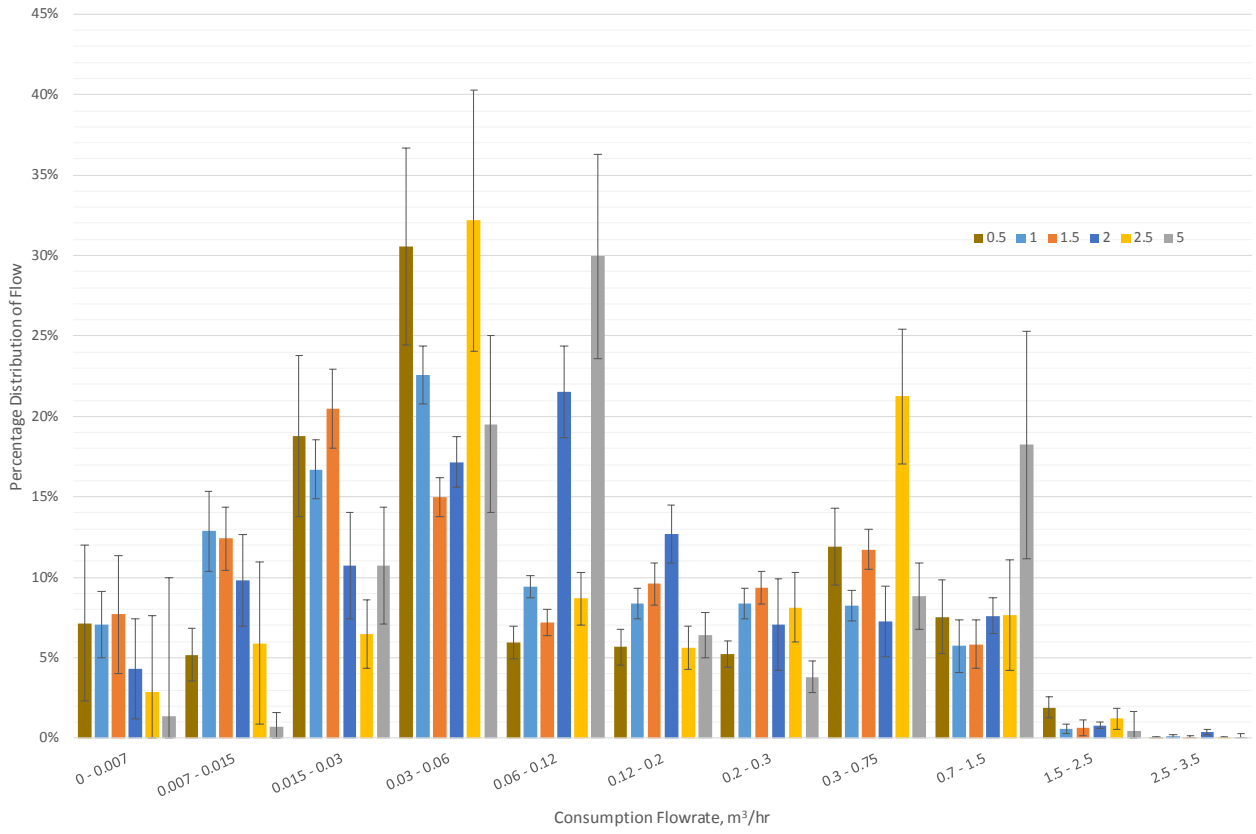


Figure 3-12: Residential Consumers Consumption Distribution by AADD

In overall, the results indicate that for Johannesburg, water meters with a q_p of 1.5m^3/hr (and q_{min} of 15 ℓ/hr, i.e. Class C) are the optimum meters for most consumers with small connections, with the likely exception of some business and multi-residential consumers. These results have been adopted and now been used by Johannesburg Water in revising of their metering standard. This may also contribute to the high proportion of low flow rates than are experienced in other countries. That as it may be, the contribution of the high prevalence of onsite leakage, as discussed in the following section, to low flow rates is a reality in the study area.

3.5 Onsite Leakage

From the meter logging results, a further analysis was done to characterise the consumption that occurs during the hours of 00:00 to 04:00 following the Minimum Night Flow Analysis (MNF) assessment approach as detailed in García, Cabrera and Enrique (2006).

3.5.1 Bulk Consumers

The occurrence of on-site leakage for the bulk consumers was limited to 57% of the multi-residential properties and 19% of business properties investigated. For the properties with leaks, the average minimum night flow for business consumers was found to be 0.02m^3/hr (median 0.005m^3/hr) while that for multi-residential properties was much higher at 0.47m^3/hr (median 0.11m^3/hr). These cumulative losses translated to 0.39m^3/day and 11.25m^3/day per business and multi-residential property respectively, which is significant for the latter.

Considering the likelihood for unrecorded low flows for bulk consumers due to the oversizing of meters, there is high likelihood of a higher level of leakage. This calls for enhanced meter management of bulk meters for reduced apparent losses and increased water use efficiency, particularly for multi-residential properties.

3.5.2 Small Connections

Onsite leakage for this category was assumed to be from events that occurred over a time duration longer than 10% of the legitimate night use during the hour of 00:00 to 04:00 (that is 24mins in total). The overall occurrence of on-site leakage, excluding properties with deemed legitimate night use, constitutes 63% of all properties comprising of 82% leakage of multi-residential, 73% of utility, 65% of business and 48% of residential properties. The overall leakage occurrence is similar to the values recorded by Lugoma et al. (2012) of 64% in Johannesburg and Mayer et al. (1999) of 67% throughout the United States of America. Table 3.4 summarises the average statistics of water usage during the night.

Table 3.4: Night flow average statistics for small connections

Category	Total Night use, m^3/night	Weighted[γ] night flow, m^3/hr	Night flow to total flow, %	Weighted[ψ] night consumption to total use, %	Weighted[ψ] night losses to total use, %
Business	0.2674	0.1601	9%	19%	17%
Utility	0.2590	0.1387	8%	12%	11%
Residential	0.1010	0.1165	6%	13%	11%
Multi-Res	0.3888	0.1819	7%	10%	9%

[γ] weighted by the number of flow events for each consumer (where a flow event corresponds to a meter reading change)

[ψ] weighted by the total night flow for each consumer

Multi-residential consumers recorded the highest night flowrate of 0.18m^3/hr which contributes 7% to the total demand of this category. The business category contributes the highest night use of 17% and occurs at a flow rate of 0.16m^3/hr, which is quite high. This is despite the fact that none of the sampled business properties had activities that warrant night use, a clear indication of leakage.

Multi-residential properties have the highest monthly leakage of 40.8m^3, followed by business (28.1 m^3), then utility (27.2 m^3) and lastly residential (10.6m^3), assuming a 50% reduction in losses during the day due to corresponding pressure reduction. The 50% reduction is based on the average range of pressure difference in JW from an average of 6bar to a maximum of 12 bar, with a leakage exponent of 1.0 for various pipe material and failure types – an example of such calculations can be found in van Zyl & Clayton (2007). The total leakages recorded by Lugoma et al. (2012) range from 12m^3/month to 29 m^3/month which is very similar to those of found by this study. While the residential category has the lowest leakage, it is by far the most populous and therefore has the greatest impact on water losses and should accordingly receive most attention in terms of reducing these losses.

The distribution of the night time consumption is shown in Figure 3-13 which is similar to Figure 3-11, with the only exception that the lower flows contribution is more significant during the night, as would be expected. The distribution is however quite different to that of Couvelis & van Zyl (2015) who found that 50 – 90% of the leakage occurred below 0.02m^3/hr which is too high. This difference is likely attributed to the difference in methodology, with the current methodology based on actual measurements over several days unlike theirs. What is however clear in all cases is that there is an appreciable volume of leakage at low flows that require attention and the need for proper choice of meters to minimise the under-registration of these flows. It is important to note that a considerable

proportion of domestic consumption (about 25%) occurs below 15ℓ/h, further emphasising the prudent choice of appropriately sized meters.

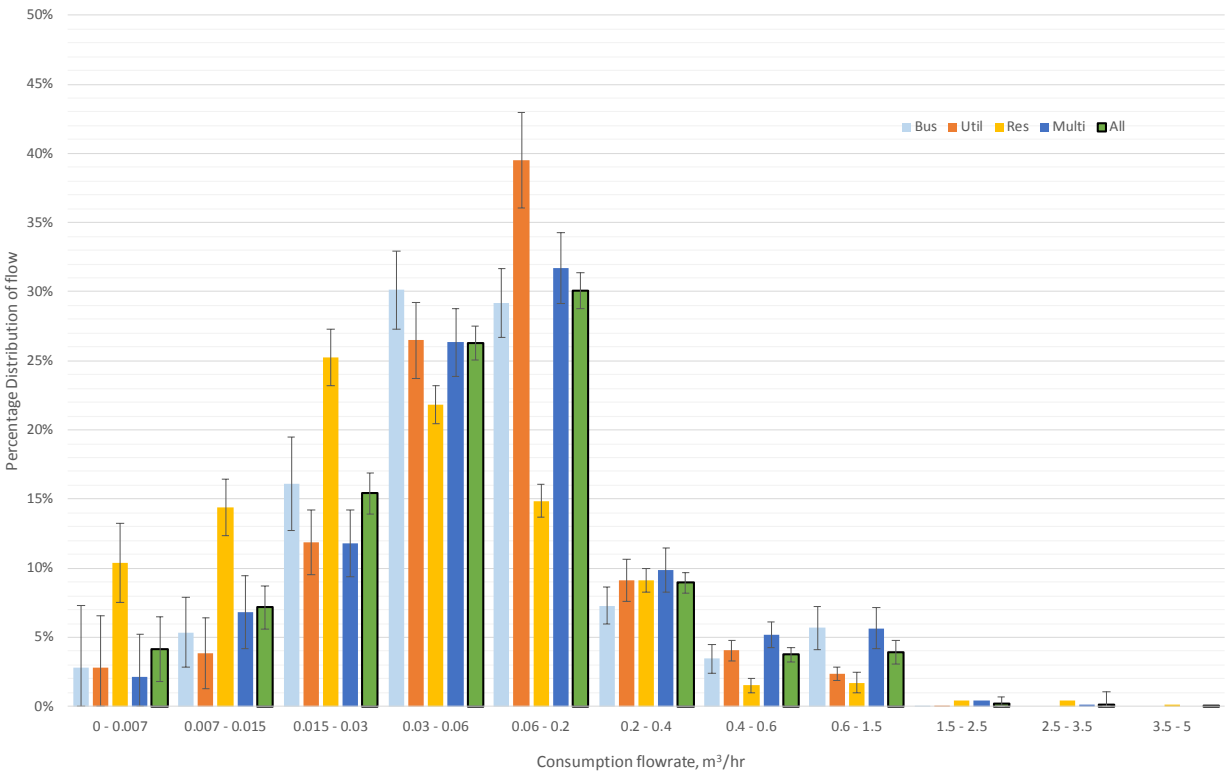


Figure 3-13: Distribution of night time consumption

The distribution of the ratio of night consumption to total consumption is depicted in Figure 3-14. Business and public benefit properties have the highest proportions of night flow exceeding 10% of the total flow. It is interesting to note that the multi-residential properties had no properties with zero-night flow, while the business category had the highest percentage of properties with no night flows, as would be expected from this type of properties. The elevated levels of night flow, particularly for multi-residential properties, warrant further attention that is aimed at substantially reducing these flows. Super-imposing the results of multi-residential properties with small connections against the results of the bulk connections of the same category, there is a clear case that there is indeed a substantial portion of losses that are not being measured in the latter and would contribute to apparent losses overall. There is therefore a case to be made for better sizing of bulk meters in multi-residential properties.

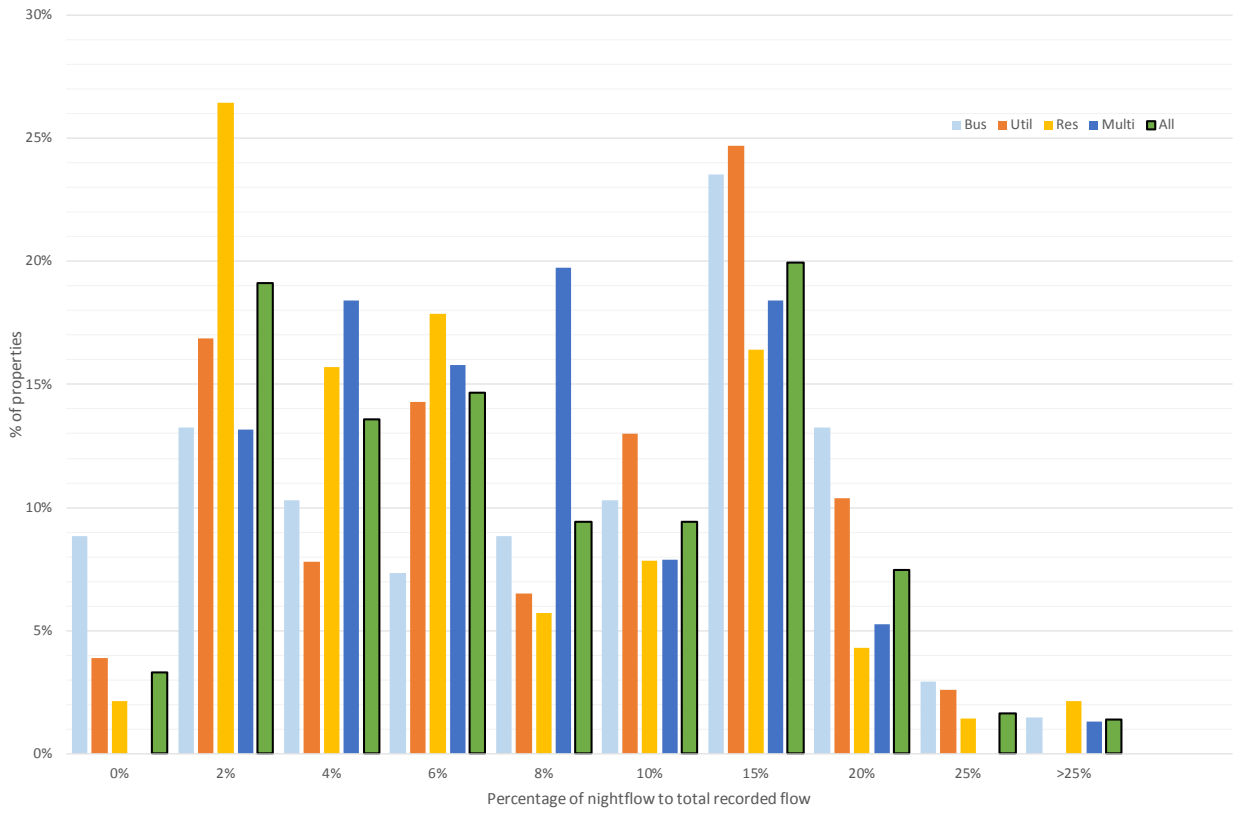


Figure 3-14: Distribution of the ratio of night flow to total flow

3.6 Summary

This Chapter discussed the consumption characterisation of 408 properties within the City of Johannesburg comprising residential, multi-residential, business and public benefit properties. These properties were logged over an average of 8.4days. In addition, analysis of an additional 256 properties that were measured separately using an even finer resolution of data spanning a much longer period was done to determine consumption profiles for four categories comprising residential, student accommodation, retail and office accommodation properties. This is the first known South African study to produce comprehensive consumption profiles based on representative consumers supplied by one of the largest utilities in the country. These consumption patterns are likely applicable to similar municipal areas both within and outside South Africa. In comparison with other studies, it is an American study that used more than 664 properties in the generation of consumption profiles that is better than the current one, while the only known study within developing countries (Uganda) only evaluated 78 properties. The consumption profiles of the various consumer segments which are deemed

representative, i.e. those from the small meter logging excluding the utility category and the students' accommodation and general retail from the smart metering data, are reproduced in Table 3.5.

Table 3.5: Consumption Profiles

Flowrate, ℓ/h	Residential	Multi-residential	Business	Students Accommodation	General Retail
0 - 7	5.5	1.9	2.7	5.3	4.0
7 - 15	9.0	4.2	5.5	0.8	2.6
15 - 30	14.6	8.8	16.9	0.7	2.6
30 - 60	21.6	19.8	22.0	0.9	5.1
60 - 200	13.6	20.3	14.0	3.0	17.9
200 - 400	9.0	9.1	7.1	31.3	25.6
400 - 600	7.4	8.9	12.0	40.5	10.9
600 - 1,500	10.7	15.2	11.9	3.3	22.5
1,500 - 2,500	7.6	10.1	7.4	4.0	7.8
2,500 - 3,500	0.8	1.4	0.4	4.5	0.5
3,500 - 5,000	0.1	0.2	0.1	5.6	0.1
5,000 - 14,000					0.4

Residential consumers have up to 37% (including uncertainties) of flow consumption occurring at less than 0.03m^3/hr, indicating a substantial proportion of low flow rates. While the other consumer categories do not have correspondingly significant low flows, all consumer groups have less than 0.25% of flow above 2.5m^3/hr indicating that meters with a q_p of 1.5m^3/hr are generally applicable for these types of consumers. The typical consumption profile of the residential consumers was also represented based on the average annual daily demand (AADD) of the individual consumer with clear indications that the use of AADD in consumption profiling can also enhance the process of meter selection and sizing, particularly for residential and multi-residential properties. While the sample for multi-residential consumers with bulk connections did not have adequate number of properties, it clearly showed that there are high occurrences of meter oversizing. The results also showed that it is very plausible to acquire additional data to support the development of relationships between the consumption and the number of units in a residential complex, which subsequently can be used to better size meters for multi-residential consumers.

On-site leakage has also been evaluated. There is a 63% occurrence of leakage within Johannesburg with the estimated monthly leakage ranging from 11 – 41m^3/month, which is relatively high and warrants further attention, particularly for the multi-residential and business properties.

Chapter 4 Water Meter Testing

This chapter is based, in part, on the publications *"Meter Accuracy Degradation and Failure Probability based on Meter Tests and Meter Change Data"* and *"Assessment of Apparent Water Losses - A Comparative Approach"* referenced as Ncube & Taigbenu (2015) and Ncube & Taigbenu (2019), respectively.

The previous chapter derived consumption profiles for representative segments of consumers in Johannesburg. This was the first step of two in determining the weighted meter accuracy of the study area, as illustrated in Figure 2-4. The second requirement involves the laboratory testing of homogeneous grouping of meters in a flow laboratory to reproduce the meter error curve as accurately as possible. This chapter summarises the work done in selecting such representative groupings of meters and the subsequent meter testing to evaluate meter accuracy at various flows to derive the related meter error degradation. In addition, an evaluation of historical meter test data and the possibility of utilising the data in determining key trends such as meter degradation and the probability of failure of meters is also done. The chapter concludes with recommendations for the modification of the weighted meter accuracy equation to make it more representative of reality.

4.1 Historical Meter Test Data

Johannesburg Water operates a water flow laboratory that is accredited by the South African National Accreditation System (SANAS) for flow metrology. All tests conducted in the laboratory comply with the requirements of South African National Standard SANS 1529 suite of standards on water metering. Most of the meter tests performed have historically been for consumers who request a test on suspicion of their meter being faulty. A few additional tests were also done by the water utility as

part of meter replacement projects to evaluate meter accuracy. The Utility provided all their historical data to the research, the summary of which is shown in Table 4.1.

Table 4.1: Summary of Meter Testing Data.

Category	No. of meters	% Passing	No. of Multi Jet	No. of Volumetric
Consumer meter data (2001 – 2004)	654			
Consumer meter data (2004 - 2014)	2511	53.5%	996	1345
Planned Meter Replacement - Area 1 (2005)	210	65.7%	210	
Planned Meter Replacement - Area 2 (2009)	307	63.5%	307	
Planned Meter Replacement - City wide (2010)	251	52.2%	152	89
Total	***3933***			

In accordance with SANS 1529-1, meters that pass the test are those that have a relative error of less than ±8% for q_{min} and ±3.5% for the remainder of the flow rates, considering that the meters are already in use. The data was provided as individual test reports with each report having three tests for each of the following five flowrates;

0.9 - 1.0q_s 0.9 -1.0q_p 0.45 - 0.5q_p 1.0 - 1.1q_t 1.0 - 1.1q_{min}

The first task was to consolidate the individual test reports using the average test results for the five test flowrates into a single spreadsheet for analysis. Only the average test results from tests that were within the uncertainty of measurement of less than 5% were used to provide a level of confidence of at least 95%. However, test reports for tests conducted prior to 2004 were inconsistent in format and therefore not amenable to automated analysis and were excluded. Table 4.2 shows the stratification by meter type for the final dataset that excluded a few incomplete records.

Table 4.2: Stratification of Meters

Meter Type	Meter Size							
	15mm	20mm	25mm	32mm	45mm	80mm	100mm	150mm
Multi Jet	1192	213	149	1	86	23		1
Single-Jet					49	10	1	1
Turbine						7	7	19
Volumetric	550	739	141					
Woltmann					1	44	9	23
Total	***1742***	***952***	***290***	***1***	***136***	***84***	***17***	***44***

A further review of the information on large meters (>32mm) showed that there were many inconsistences with the records and therefore the data could not be relied upon. Together with the uncertainty introduced by the relatively small numbers of meters involved, all bulk meters were excluded from further analysis leaving only the multi-jet and volumetric meters. 94% of the remaining meters were all from one manufacturer.

While all the meters have a recorded reading at the time of the test, a sizeable proportion did not indicate the age of the meter and this information was inferred from the manufacturer serial number on the meter. The respective meter manufacturers were approached, and they provided the serial number logic that included the year of manufacture for each meters. It was assumed that meter installation occurred within the same year of manufacture and therefore the meter age was the difference between the date of meter manufacture and the test date. The distribution of the meter age and meter reading are summarised in Figure 4-1, showing an expected skewed distribution to the left for both age and reading but with an unusually high number for meters aged 21 – 22 years. No explanation could be provided for this anomaly but may be due to some rigorous metering programs around that particular time.

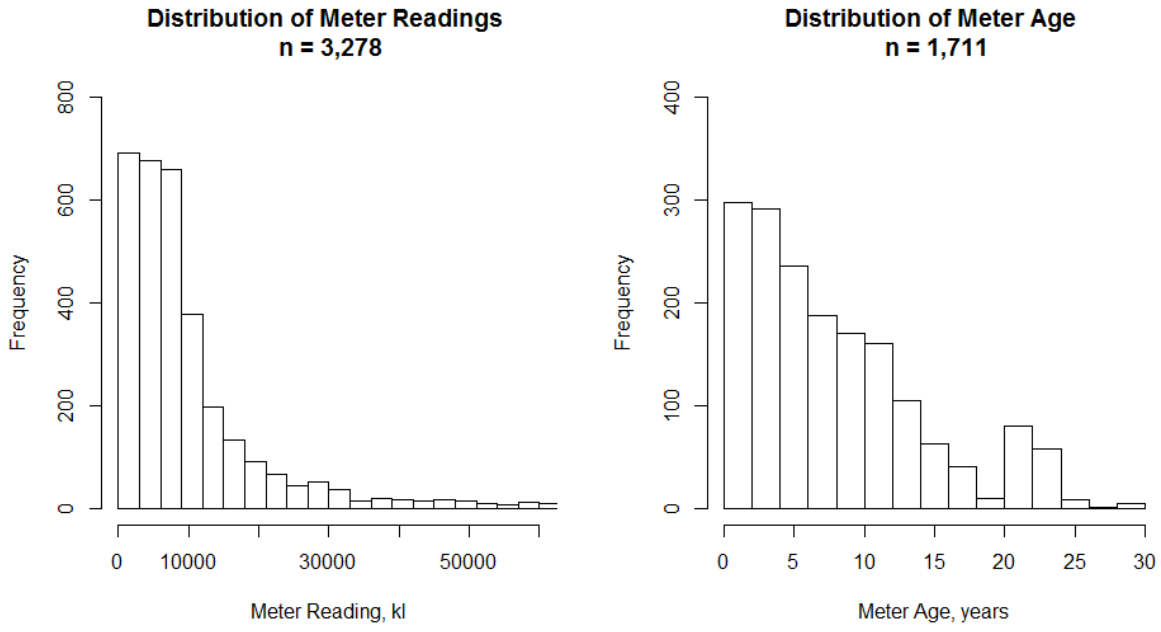


Figure 4-1: Distribution of Meter Test Data.

Meters were accordingly grouped per age group and by meter reading quantiles, with the mean reading of the quantile used to represent all meters in that quantile. Statistical analysis was thereafter performed using the R statistical software package (R Core Team, 2017) to determine the relationships between meter age, meter reading and the meter error for the volumetric meters and the velocity meters (multi-jet and single-jet).

4.1.1 Likelihood of Meter Failure

The likelihood of a meter failing a laboratory test in relation to both the age of the meter and the volume that has registered on that meter was determined. The probability of failure was calculated separately for the volumetric and velocity meters for both the meter age and meter reading. For each category of meter age and reading, the probability of failure was determined as

$$P_f = \frac{No.\ of\ meters\ in\ category\ failing\ meter\ test}{Total\ number\ of\ meters\ in\ category}$$

Equation 4.1: Probability of Failure

The probability of failure was thereafter evaluated against meter age and meter reading for the different meter types. Figure 4-2 shows the likelihood of meter failure against meter age for volumetric meters of up to 18 years of age;

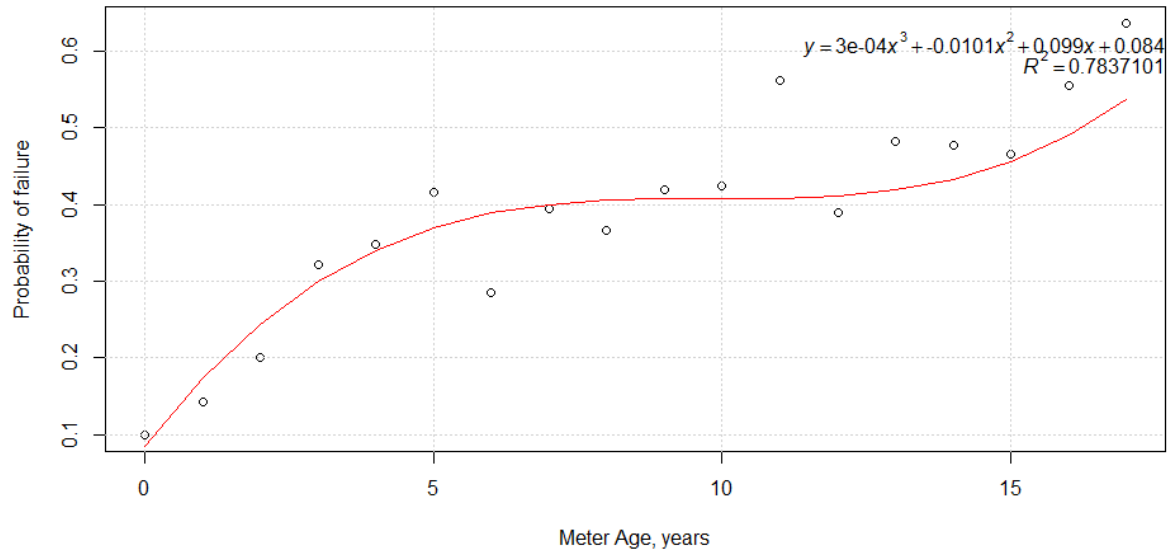


Figure 4-2: Probability of Failure with Age - Volumetric Meters

The probability of failure increases rapidly in the first five years of a meter and stabilises for about seven years before it begins to increase rapidly again. The rate of increase of the probability of failure is about 8% per year for the first 5 years of a meter's lifespan. The curve does seem to make sense intuitively as the meter settles in to its environment. This is in line with the observation in Arregui et al. (2018) that meter accuracy diminishes rapidly in the first few years and slows down with time.

The probability of failure profile for volumetric meters against the indicated meter reading is shown in Figure 4-3.

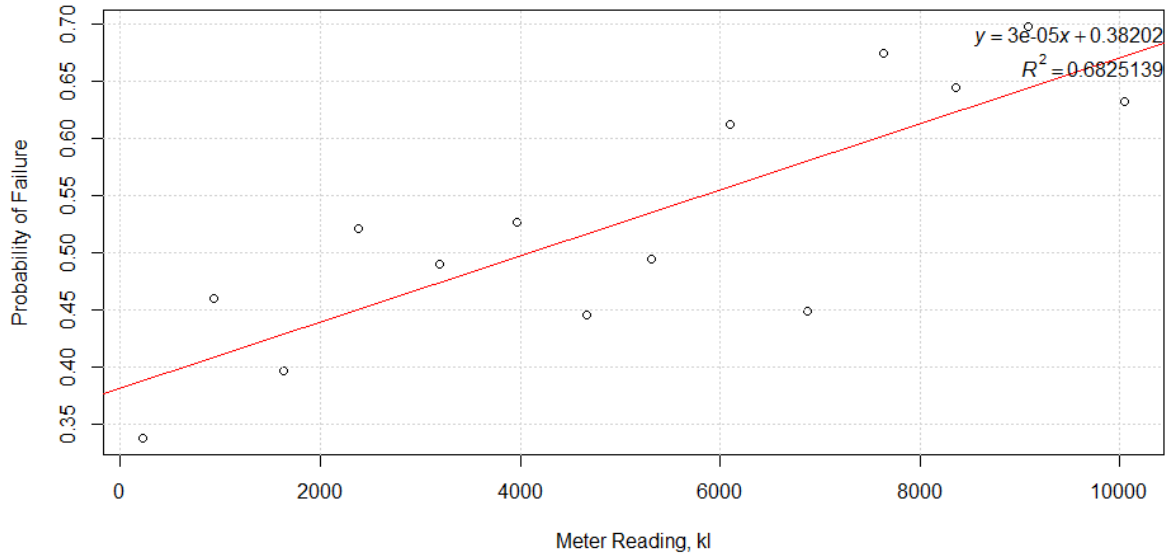


Figure 4-3: Probability of failure with Meter Reading - Volumetric Meters

The relationship is a linear with a 3% increment in the probability of failure for every 1,000kℓ on the meter counter for the first 10,000kℓ of a meter. This is, again, an expected result as the meter loses its accuracy with an increasing amount of water flowing through it leading to an increased probability of failure.

For the velocity meters, the probability of failure against meter age or meter reading did not produce any meaningful relationships. A further disaggregation of the meters against meter model and meter manufacturer did not yield any results. This was unexpected and may indicate the limitations of relying on this approach exclusively. However, it must be noted that velocity meters are more susceptible greater variation in meter accuracy with time depending on the water quality, installation position and a number of other environmental factors. Such factors may be the reason the lack of relations.

It is expected that for the volumetric meters, the actual probability of meter failure should be a combination of both the age of the meter and the meter reading at any point in time. Such combined probability of failure, according to probability theorem, is given by the joint probability mass function that the meter will fail. This can accordingly be found by the following relationship;

$$P_f(Age, Reading) = P(X = Age, Y = Reading)$$

Equation 4.2: Joint Probability of Failure

Despite a lack of results from the velocity meter, the volumetric meter results do indicate that routine meter testing data can provide additional value to utilities and may also inform and influence metering policies and the general approach to meter testing in response to consumer queries. In the case of Johannesburg, the current approach is that they will test any meter at the request of a customer, even in the meter is new, but charge a very small fee which is much less than the costs involved in removing and replacing the meter and testing it at the meter laboratory. Although this fee is credited to the consumer should the meter fail, the entire process is expensive for JW and could be significantly improved. The determination of the probability of failure, as done in this study, can be used to determine the ages and reading at which the meter can be simply replaced at no cost to the consumer. Meters with a reading of over 4,000kℓ and an age of over 10 years could be such candidates. Conversely, only meter that are not expected to fail, such as those less than 5 years only with an average meter reading could only be tested at the full economic test of the exercise.

The use of probabilities can also be used as a proxy for determining meter reliability in instances where a comprehensive evaluation of the weighted meter accuracy has not been carried out. Generally, the relationships give insight into the likely probabilities of failure of meters that could be related to particular models that are widely used throughout South Africa.

4.1.2 Meter Degradation

An evaluation of the degradation in the accuracy within the flowrates of the meter was evaluated using the average meter accuracy at the five test flow rates against both the meter reading and the meter age. Only meters with a relative error of less than ±30% were included, as higher errors were

found to be outliers. Furthermore, meters that were older than 19 years and/or with readings higher than 35,000 were also omitted, as they had very few records and introduced higher unreliability. Figure 4-4 shows error degradation rates with meter reading at the minimum flowrate for volumetric meters.

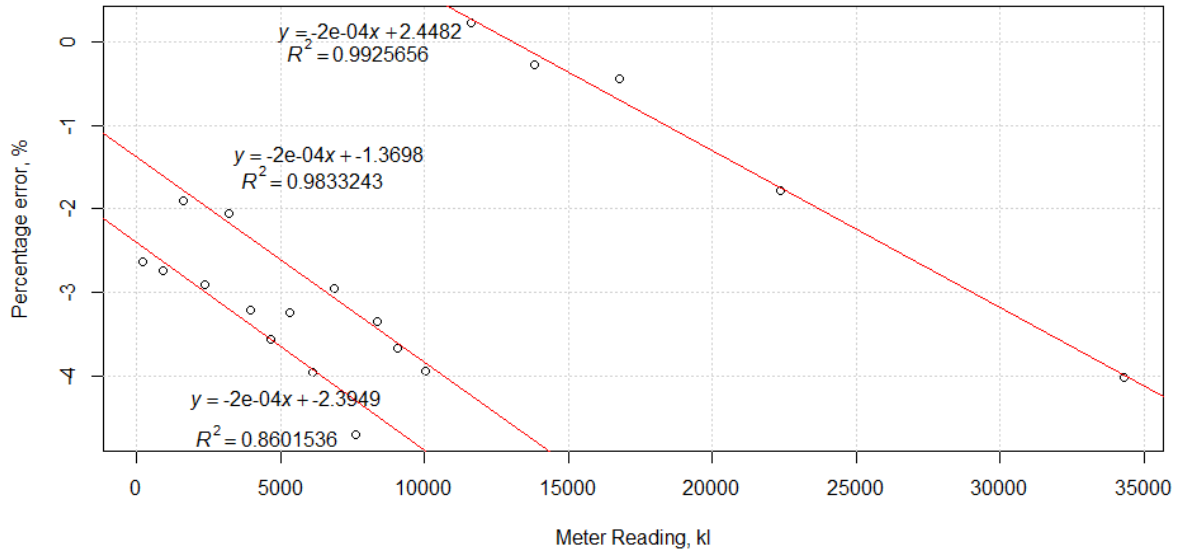


Figure 4-4: Accuracy degradation at q_{min} with meter reading

The data shows three distinct regions can be observed, and it is interesting to note that the degradation rate in either region is very similar, ranging from -0.19% to -0.25% per 1,000kℓ registered by the meter. The lower two of the three regions on the graph may be attributed to meter clock-overs for meter with four digits (up to 9999) while the upper graph might be of different meter sizes and/or batches. There was unfortunately no further detail on the meters to verify either of these assumed explanations. The degradation rate shown is on the low side as the meter only loses about 2% of its initial accuracy at a reading of 10,000, with the starting error being a key factor. In the most pessimistic case, even if a meter begins at the required ±5% for q_{min}, it will be still within the ±8% error at q_{min} for used meters.

Figure 4-5 shows error degradation rates with meter age at the same minimum flowrate for volumetric meters. Error degradation is slightly higher at 0.32% increase per year, but not so substantially different from the meter reading degradation. After ten years, for a meter with a starting q_{min} of -5% will be above the required -8% error at the flow rate and therefore non-compliant.

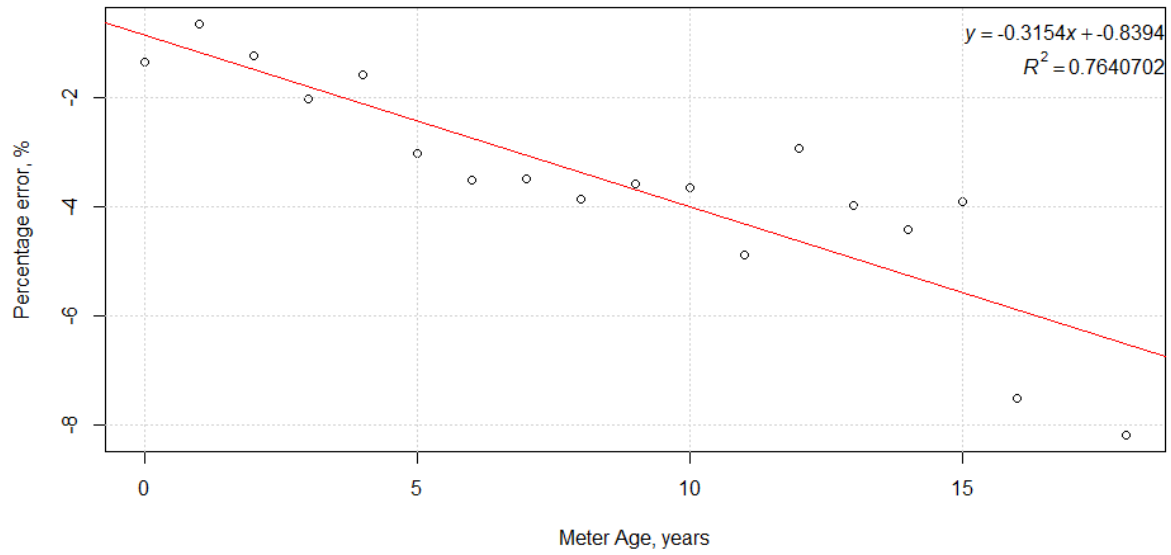


Figure 4-5: Accuracy degradation at q_{min} with meter age

When comparing the correlation strength between meter accuracy and meter age or meter reading the meter readings showed a better correlation, particularly true for volumetric meters. This aligns with expectation that the meter reading should in fact better represent meter degradation than age as it is a better indicator of the level of wear and tear of a meter.

Table 4.3 summarises all the variations of the meter error degradation rates with a coefficient of determination (R^2) value greater than 0.5 by meter type, including those that were illustrated in Figure 4-4 and Figure 4-5.

Table 4.3: Error Degradation Rates.

Category	Degradation Rate	Volumetric Meters				Multi-Jet Meters	All meters			
		q_{min}	q_t	q_p	q_s	q_t	q_{min}	q_t	$0.5q_p$	q_p
Meter Age	Rate/yr.	-0.32					-0.13		-0.06	-0.05
	R^2	0.76					0.66		0.57	0.69
Meter Reading	Rate/1,000kℓ	-0.19 to -0.25	0.07	-0.1	-0.1	0.08		0.06		
	R^2	≥ 0.86	0.57	0.67	0.65	0.74		0.83		

From Table 4.3, multi-jet meters only exhibited a relationship between the transitional flow and meter reading. For all the meters, there is an expected increase to the positive meter error at the transitional flow rate indicating a positive degradation rate. This is because the meter error at the transitional flowrate is usually positive as depicted in the typical meter error curve of Figure 2-5 with a general

tendency of over-registration at the transition flowrate. There is a minimal difference in the quantum of the degradation rate between velocity and volumetric meters at the transitional flowrate as well. It can also be observed that the degradation rate at the permanent flow rate is about half that at the minimum flowrate, which is again expected as the meter accuracy is better at the permanent flow rate (refer to Figure 2-5 for a typical meter error curve). This emphasises the fact that the ability of a meter to measure low flows should be a thoughtful consideration as it degrades the most at that flow rate.

It can also be observed that, as was the case with the likelihood of failure analysis, the volumetric meters can be better represented by analysing historical data while velocity meters hardly show any relationship.

Despite the usefulness of the historical meter testing data, its use in the estimation of the weighted meter accuracy by superimposing the consumption profile was not implemented. This is because the flow rates used in the test while complying with the requirements of legislation, do not produce a representative error curve and could lead to misleading results as was discussed in §2.2.2. in addition, by their very nature, these tests are not representative and therefore cannot be simply used to derive a representative value of apparent losses and meter degradation. However, while that might be the case, these results demonstrate that in the absence of better information, the routine but reactive meter test data of utilities do offer valuable data that can be used to estimate the metering error and the overall management of meters. There is therefore a case to be made for leveraging the use of such data for the improved management of apparent water losses.

4.2 Controlled Meter Testing

4.2.1 Preparatory Work

The database described in §3.1.1 was used to define a representative sample of meters that were tested in the laboratory. In this case, only active meters were selected and a review of fields that pertain to meter make, model and size was done prior to arriving at the sample size. Together with

prior field experience from the Consumption Characterisation, it became clear that these fields (meter make, model and size) were inconsistent and could not be relied. This made it difficult to prepare homogeneous groupings of meter samples for testing. Fortunately, the fact that historically Johannesburg Water mainly used two meter manufacturers, one for smaller meters and the other for bulk meters, was viewed as a mitigant to this challenge. Meter make information was therefore not considered much in the sampling process.

There were about 317,000 meters that were considered active in the JW database with adequate information on current meter reading and indicative meter size. A sample of 325 small diameter meters (15 – 25mm, presumably) and 115 bulk meters (presumed 40mm) were selected, stratified per every 2,000m^3 and 5,000m^3 reading on the meter dial respectively, to achieve a global minimum of 95% confidence level. An example of the stratified sampling is presented in Table 4.4 for small meters based on active meters as of January 2015. Oversampling was done in anticipation of field complications and to ensure that each reading category had at least 5 meters that could be tested.

Table 4.4: Stratified Sampling for Small Meters

Reading Range	Total No. of Meters	Small Meters %	Minimum Sample Size.
2,000	91,803	29.5%	88
4,000	61,345	19.7%	59
6,000	42,259	13.5%	41
8,000	32,865	10.5%	32
10,000	26,130	8.4%	25
12,000	13,570	4.3%	13
14,000	9,773	3.1%	9
16,000	7,554	2.4%	7
18,000	5,723	1.8%	5
20,000	4,348	1.3%	4
22,000	3,493	1.1%	3
24,000	2,553	0.8%	2
26,000	2,049	0.6%	2
28,000	1,615	0.5%	1
30,000	11,972	2.6%	8
Grand Total	***317,052***	***100.10%***	***299***

These meters were then removed from the field by a contractor and subsequently tested at the SANAS accredited flow laboratory of Johannesburg Water. The removed meters were replaced with new meters, and this constitutes part of the cost associated with the weighted accuracy methodology of determining apparent losses. Replacing the meters is the most practical way of dealing with removing meters for testing as all other options are cumbersome and can compound billing challenges. Figure 4-6 shows the location of properties from which meters were removed for testing.

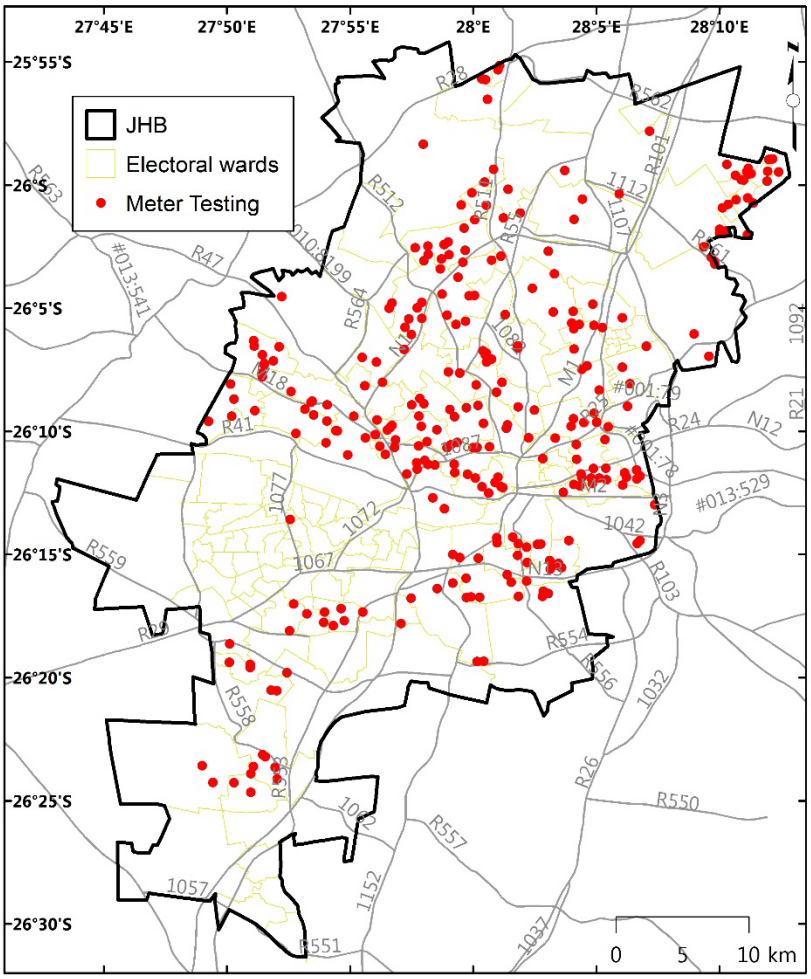


Figure 4-6: Location of Sample Point for Meter Testing

As shown in §2.2.2, a non-trivial aspect of meter testing for the purposes of determining the weighted meter error is the determination of the appropriate test flowrates to reproduce the meter error curve as accurately as possible. The recommendations of Arregui et al. (2013) were slightly modified to suit meters common in South Africa. The final test flowrates are indicated in Table 4.5, together with the derived relationship for any meter size due to the uncertainty of the actual meter sizes in the field. For each of the flowrates of Table 4.5, the test flow rate was repeated three times to evaluate

for consistency using statistical variables. Only tests that were consistent were used to calculate the average accuracy at the specified flowrate as shown in Figure 4-7 for 15mm Class C volumetric meter with a q_p of 1.5m^3/h. A sample of complete meter test reports is included in Appendix A.

Table 4.5: Test flow rates

Size	*q_p, ℓ/h*	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*	*Q7*	*Q8*	*Q9*	*Q10*
15mm	1,500	7	15	22.5	30	60	120	750	1,500	2,250	3,000
20mm	2,500	12	25	37.5	50	100	200	1,250	2,500	3,750	5,000
25mm	3,500	17	35	52.5	70	140	280	1,750	3,500	5,250	7,000
Any other	q_p	$0.5q_{minC}$ - 1	q_{minC}	q_{tC}	q_{minB}	$0.5q_{tB}$	q_{tB}	$0.5q_p$	q_p	$1.5q_p$	$2q_p$

Where q_{minC} is the minimum flow rate for a Class C meter, q_{tB} is the transitional flow rate of a Class B meter, and so forth.

Test Bench Used			Vol B		Reference Quantities					Test Item			Test Results		
Test Bench Postion				Test	Test	Water	Test	Ref	Ref	Start	End	Indicated	Error		Variance
	(TXXX)	T005		No	Press	Temp	Duration	Volume	Flow Rate	Reading	Reading	Volume	ABS	Rel	To Ave
					kPa	°C	s	I_{REF}	l/h	m^3	m^3	I_{IND}	l	%	%
0.9	to	1.0	Q_s	Test 1	482	22.9	186.89	151.70	2922.15	376.36482	376.51543	150.61	-1.09	-0.72	-0.04
2700	to	3000	l/h	Test 2	482	22.9	188.17	151.45	2897.49	376.51543	376.66592	150.49	-0.96	-0.63	0.05
± 2%	New & Refurb			Test 3	482	22.9	188.03	151.75	2905.39	376.66592	376.81663	150.71	-1.04	-0.69	-0.01
± 3.5%	In Use			Ave					2908.34			150.60	-1.03	-0.68	
												Test Result	Pass		
0.9	to	1.0	Q_p	Test 1	648	23.0	242.82	151.60	2247.59	376.81663	376.96744	150.81	-0.79	-0.52	-0.16
2025	to	2250	l/h	Test 2	648	23.0	242.71	151.60	2248.61	376.96744	377.11844	151.00	-0.60	-0.40	-0.03
± 2%	New & Refurb			Test 3	648	21.0	242.69	151.60	2248.79	377.11844	377.26978	151.34	-0.26	-0.17	0.19
± 3.5%	In Use			Ave					2248.33			151.05	-0.55	-0.36	
												Test Result	Pass		
0.45	to	0.5	Q_p	Test 1	800	22.4	377.92	151.450	1442.69	377.26978	377.42137	151.59	0.14	0.09	0.00
1350	to	1500	l/h	Test 2	800	22.4	376.83	151.550	1447.81	377.42137	377.57300	151.63	0.08	0.05	-0.04
± 2%	New & Refurb			Test 3	800	22.4	374.21	151.650	1458.91	377.57300	377.72484	151.84	0.19	0.13	0.04
± 3.5%	In Use			Ave					1449.80			151.69	0.14	0.09	
												Test Result	Pass		

Figure 4-7: Meter Testing Results Extract

The average accuracy for each of the flowrates was thereafter used to estimate the meter accuracy curves for each meter, assuming a straight line between the respective flowrates.

4.2.2 Results

During the site work, many meters could not be removed for a variety of reasons ranging from illegible meters, recently changed meters without updated details on the meter database, and meters found on site had a much different meter reading range as per the sampling strata. In addition, due to budgetary constraints and time limitation it was no longer possible to proceed with the testing of bulk meters, particularly because of the discrepancies between meter information on site and those in the meter database.

Slightly over 210 meters were removed from site and taken to the flow laboratory for testing. However, only 123 meters consisting of six models and three sizes from mostly two manufacturers were successfully tested. This represented a higher number of meter models than was initially envisaged, and it is expected to further reduce the confidence of the results. This, unfortunately, could not be avoided because it was not possible to pre-inspect all the meters in an attempt to achieve the anticipated sample size. The unsuccessful tests included plastic meters that developed leaks when tested at high flowrates and those meters that were damaged on arrival at the testing laboratory. This relatively low success rate of 59% for testing meters was very concerning and highlighted that using meter testing is not without additional challenges over and above the cost and effort associated with the exercise. These difficulties are further compounded by the poor state of the database with incorrect meter information.

Figure 4-8 graphically shows the test results for the Model 2 (15mm, Class C) which had the highest number of meters tested and is also known to be a common plastic meter in many centres around South Africa and beyond.

There is a wide variation of accuracy at the low flow regions, with many meters failing the ±8% error requirement at q_{min} although they were still in service. This demonstrates that the ongoing verification requirement of the Trade Metrology Act holds merit but had not been complied with in the study area. The wide accuracy variation at low flows is indicative of significant decline of accuracy at these low flowrates, which underscores the importance of testing the meter at more point that just the legislated flow rates to adequately capture such variation. Such test flow rates should, as of necessity, include flow rates below the minimum flow rate to cater for the rapid drop in accuracy which would otherwise not be accounted for. To illustrate the impact of such omission, Figure 4-9 illustrates the degradation rate of the Q1 error ($0.5q_{minC}$ - 1) for all the tested 15mm volumetric meters.

The error degradation rate at this flow rate is very high at -10% per 1,000kℓ. This is an order of magnitude higher than what was found for the q_{min} using the historical meter testing data. Other than the fact that this flow rate is lower than q_{min} and therefore it is expected for the error to be higher, the results show that using reactive meter test data that does not measure lower flows has its shortfalls.

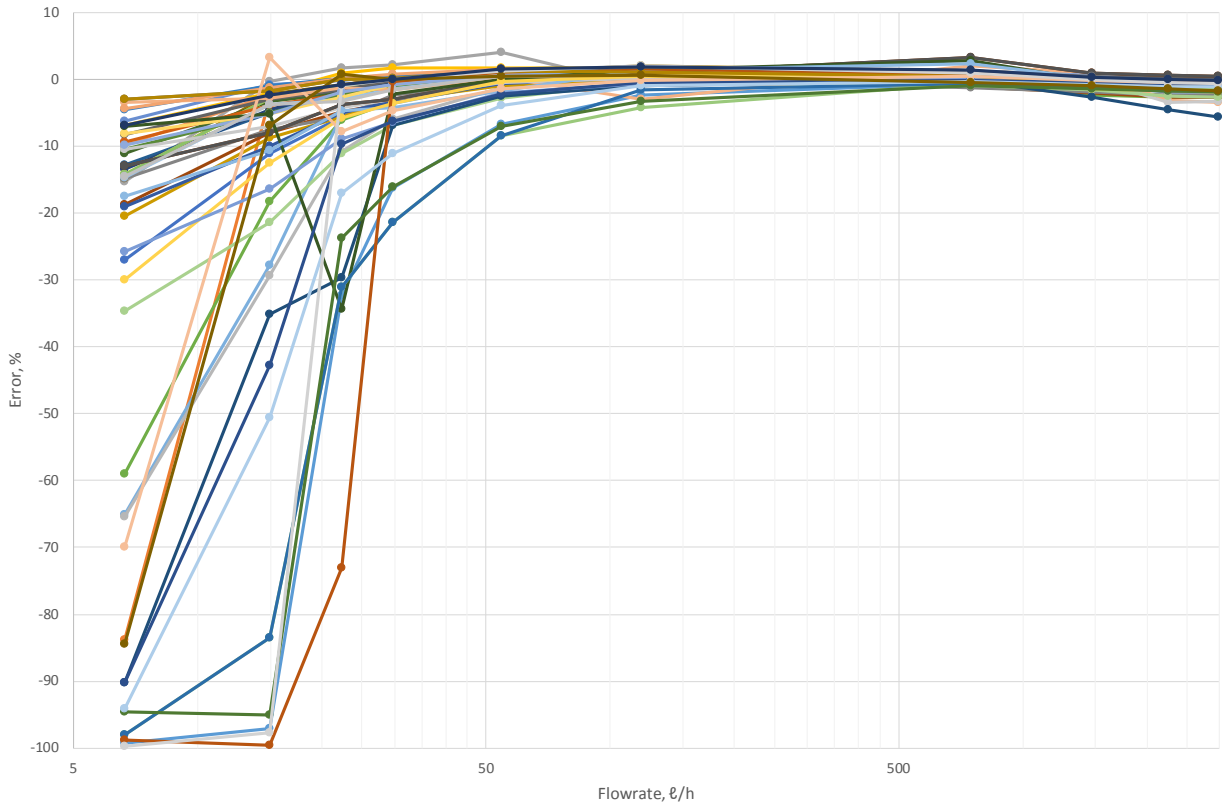


Figure 4-8: Model 2 – 15mm Test Results

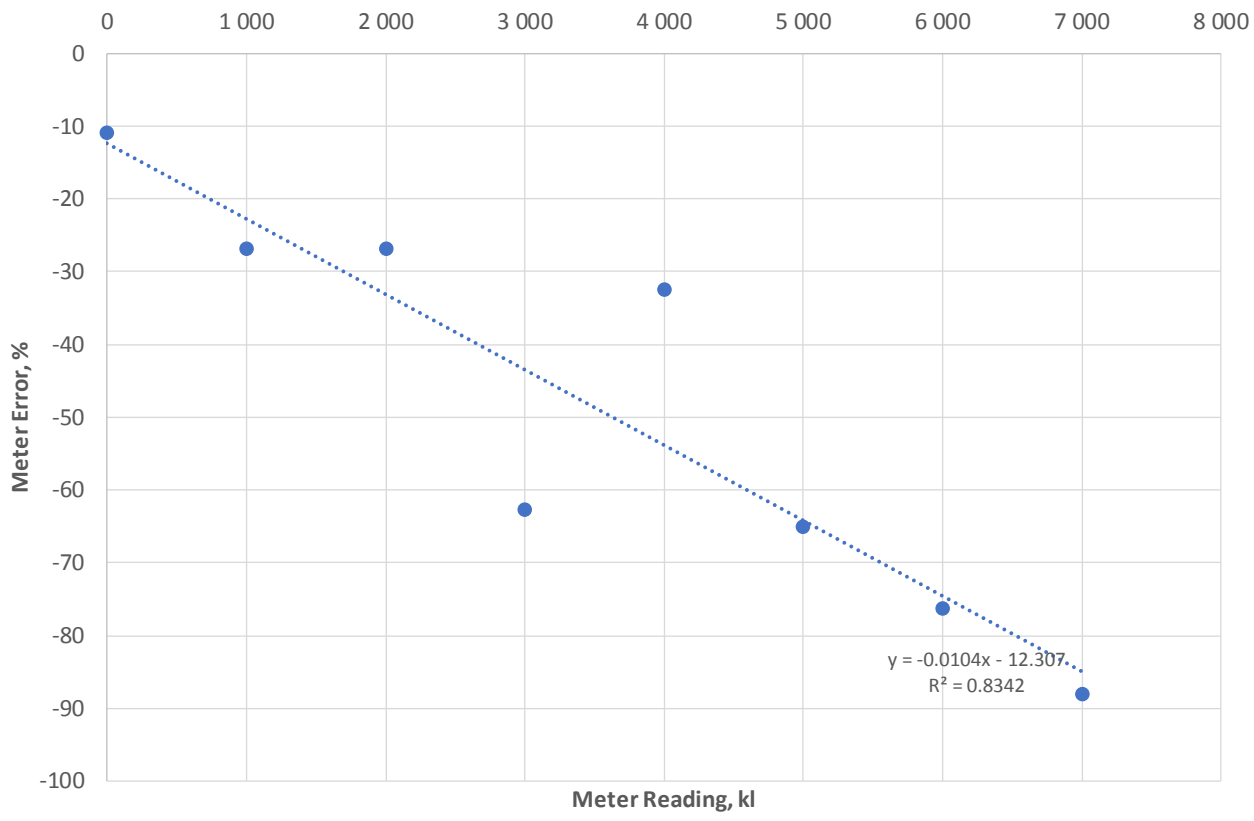


Figure 4-9: Q1 Error Degradation – Volumetric Meters

While the use of historical data does have its value, the user must be aware that such data may mask certain trends, such as these higher degradation rates. While there were few points per category of meter reading and the age in the generation of Figure 4-9, thereby creating some uncertainty, the high R^2 value does indicate that the relationship is sufficiently strong.

Figure 4-8 also shows that the variation in the meter error is however much more constrained with increase in the flow from the transitional flow to the permanent flow rate. This is expected as this is the region where meters are generally the most accurate. Therefore, the sizing of water meters should be done such that the highest proportion of the consumption lies around the permanent flow rate of the meter.

The average accuracy for all the meters, and other details, are summarised in Table 4.6 for each of the test flow rates while Figure 4-10 graphically shows the average error per meter model and meter size. All meters in Figure 4.9, except where indicated otherwise, are all 15mm in size.

Table 4.6: Meters Test Results

Size and model	No.	Ave Age	Ave Read	Average Errors per flowrate, %									
				Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
15mm	**85**	**10.9**	**3129**	**-35.80**	**-20.89**	**-11.22**	**-4.80**	**-1.29**	**-1.00**	**0.00**	**-0.81**	**-1.36**	**-1.73**
Model 1	10	14.8	1193	-42.27	-24.23	-14.17	-8.53	-3.92	-2.00	-2.35	-4.16	-4.20	-6.56
Model 2	51	9.6	2774	-34.54	-20.08	-9.45	-5.22	-1.78	0.43	0.41	-0.45	-1.05	-1.23
Model 3	3	22.4	11039	-58.91	-16.11	-22.65	-6.87	-0.36	0.08	-0.10	-0.69	-0.55	-0.68
Model 4	3	26.9	7824	-42.43	-39.45	-24.18	-0.71	0.90	1.64	-0.57	-1.96	-2.17	-2.61
Model 5	18	7.5	3110	-30.82	-19.06	-10.51	-1.87	1.07	-5.13	0.25	0.20	-0.67	-0.49
20mm	**35**	**12.6**	**6776**	**-40.10**	**-20.35**	**-13.32**	**-7.49**	**-5.87**	**-5.03**	**-6.04**	**-3.73**	**-3.92**	**-3.99**
Model 2	17	8.7	2867	-32.83	-25.83	-20.51	-13.14	-11.65	-11.18	-6.40	-7.01	-7.25	-7.30
Model 3	14	19.9	11565	-49.28	-11.22	-1.02	1.43	0.10	0.63	-7.56	-0.79	-1.00	-1.15
Model 5	2	13.7	6977	-15.05	-0.95	2.16	3.07	1.44	0.99	0.41	0.30	0.37	0.39
Model 6	2	3.3	6271	-62.67	-57.09	-53.73	-32.51	-5.88	1.67	1.11	-0.48	-0.32	-0.04
25mm	**3**	**22.9**	**1959**	**-11.96**	**-1.00**	**-0.14**	**1.30**	**1.34**	**1.57**	**0.29**	**-0.11**	**-0.34**	**-0.55**
Model 2	2	-	454	-1.65	0.51	0.61	1.33	1.32	1.64	0.44	-0.04	-0.35	-0.65
Model 3	1	22.9	4968	-32.60	-4.01	-1.62	1.25	1.39	1.41	0.00	-0.26	-0.32	-0.36
Global	***123***	***11.5***	***4138***	***-36.44***	***-20.25***	***-11.54***	***-5.42***	***-2.53***	***-2.08***	***-1.72***	***-1.63***	***-2.06***	***-2.34***

The 15mm meters generally perform better than the bigger meters at lower flows, which is expected as their measuring ability for lower flows is better. For the 20mm and 25mm meters, it is also clear

from Figure 4-10 that while the test flowrates adopted are an improvement to three test flowrates reported in the literature, there is still a need to include additional low flow rates in keeping with those of the 15mm meters particularly because these meters are used for consumers who could be served adequately with 15mm meters. The impact of meter sizing can also be clearly seen in this case where higher inaccuracies are observed with bigger meters compared with the 15mm sized meters, particularly at lower flowrates.

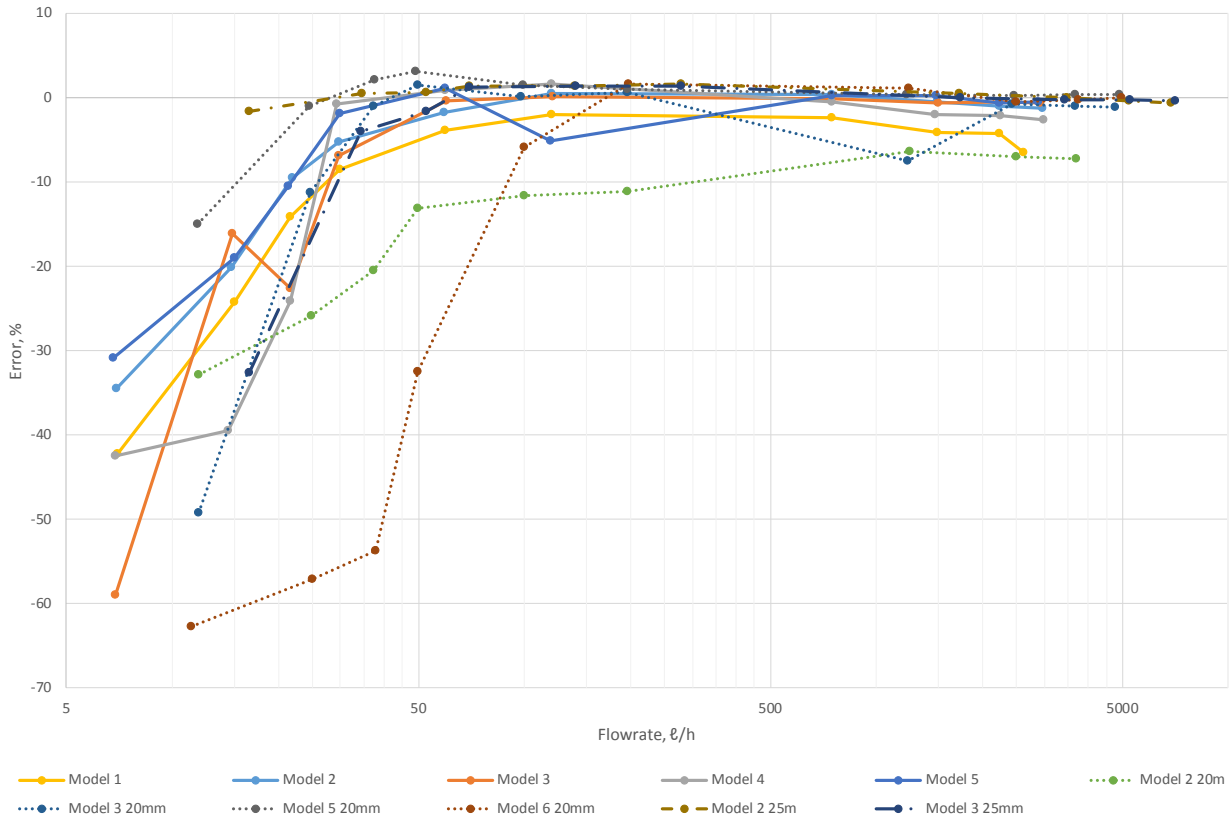


Figure 4-10: Average Error per Meter Model and Size

The rapid variations of the error curve between successive test flow rates, particularly at low flows, in Figure 4-10 is also instructive on the use of the group average accuracy test results in calculating the weighted meter accuracy using Equation 2.4. Some authors have taken the group average to be purely the average of the accuracy for meters tested at a particular flow rate without considering the variation of the average accuracy between successive test flow rates to capture the huge differential. The average of the group accuracy for the flow range under consideration will therefore be a better

representation of the accuracy of the meter. This is in fact in keeping with the principle of finding the area under the graph to estimate the likely volume.

A modified equation to better estimate the weighted meter accuracy without leaving any room for ambiguity is proposed as follows;

$$\bar{\varepsilon}_{w} = \sum_{q=1}^{n} \mathrm{PTC}_{q} \bullet 0.5(\mathrm{GAAL}_{q} + \mathrm{GAAL}_{q-1})$$

Equation 4.3: Modified Weighted Meter Accuracy

Where GAAL_0 (when q = 1) represents the meter accuracy at the starting flowrate of the meter. In cases where the accuracy at the starting flowrate is not determined, as was the case in this study, GAAL_0 may be assumed to be -100% at just above the zero flowrate. The GAAL_0 term is important in that in captures the point at which the meter is most inaccurate and assumes linear progression to the next measured error term (as observed in Figure 4-10) thereby reducing the over-estimation of the weighted meter accuracy. In practical terms, Equation 4.3 only modifies the weighted meter error in the low flow regions and remains about the same in the higher flow regions.

4.3 Summary

The evaluation of historical meter testing has shown that such records, despite being largely from random tests done at the request of consumers, do offer valuable insight into the degradation of the metering error and the overall probability of a meter failing to meet the required accuracy levels. The starting flowrate is the worst affected with up to 0.25% increase in error per every 1,000kℓ passing through the meter while the permanent and overload flowrates degrade at almost half the rate of the starting flowrate. On the other hand, the error at the transition flow rate is positive and degrades at lower levels of about 0.06% per 1,000kℓ. In the absence of elaborate and carefully designed and implemented projects to determine metering errors at a flow laboratory, utilities can make use of such data in the preliminary and indicative determination of the metering error. This in turn can be enhanced by the availability of data on the initial metering error from verification exercises. Furthermore, the data also shows that the probability of failure can be used to make decision on meter

testing requirements and the acceptable level that can be used for billing purposes. However, it is important to note that such results are incomplete on their own and should therefore be used with caution for indicative assessments.

Detailed meter testing was successfully done for 123 small bore meters out of a planned 394 meters. A relative low rate of 59% for successfully testing meters was achieved demonstrating the challenges that might be encountered in removal of meters for testing. The average error for each of the six meter models in three sizes are shown in Table 4.6 for each of the 10 test flowrates that were evaluated for each meter, together with the average meter age and reading. Due to the rigorous sampling done, the sample is representative of the Johannesburg Water meter fleet, albeit at a lower error margin to that initially envisaged. While not all the objectives of the controlled meter testing were achieved, the process was instructive, as were the results. Prior studies have tested more meters than was currently done, such as the 250 water meters implemented by Mutikanga et al. (2011a) in Uganda, but at only three flowrates as opposed to the ten flowrates adopted in this study. In essence, this study performed well over 1,230 individual tests compared to the 750 tests in Mutikanga et al. (2011a), which is significant. The value of having more test flowrates has been demonstrated in this study by observing the variations of accuracy at low flowrates.

While test flow rates have been suggested in this study for any meter size, the gaps observed for smaller flowrates for 20mm and 25mm meters show that it is necessary to determine the correct flow rates to use for the intended consumer group. In this study the error curve of the 20mm and 25mm meters would have been better represented by adopting the 15mm flowrates, particularly at low flowrates.

It has also been observed in this study that the current weighted meter error equation does not adequately represent the actual metering error within a flowrate range as it only utilises the upper accuracy value of a flow range. As such, an alternative equation that better represents the weighted accuracy of meters, particularly at low flowrates, is also introduced and will be used in determining apparent losses. The modification allows for the use of the midpoint accuracy of a flow range as opposed to using the highest flow range accuracy to represent the entire flow range. The modification

is even more valuable in instances where the meter error curve is not adequately represented due to the selection of a few meter test points.

It has also been shown that despite good intentions and design, the testing of meters from the field, other than the cost involved, is fraught with many difficulties some of which are related to how well a meter database is structured and managed. The small number of meters that could be successfully tested compared to those removed from the field also shows that the exercise is not guaranteed to be successful, and utilities need to factor this into their plans. Most importantly, this has also highlighted the need to explore viable alternatives of evaluating metering errors, without the current risk of not obtaining adequate returns from a meter testing process.

Chapter 5 Apparent Water Loss Assessment

This chapter is based, in part, on the publications *"Meter Accuracy Degradation and Failure Probability based on Meter Tests and Meter Change Data"* and *"Assessment of Apparent Water Losses - A Comparative Approach"* referenced as Ncube & Taigbenu (2015) and Ncube & Taigbenu (2019), respectively.

Chapter 3 produced consumption profiles for several consumer categories representative of the City of Johannesburg water consumers in formal non-prepayment areas. On the other hand, Chapter 4 evaluated the metering error of representative samples of meters from the same area. This chapter combines the consumption profiles of Chapter 3 with the meter accuracy of Chapter 4 to determine the weighted meter accuracy and an estimate of apparent losses for the City of Johannesburg. Based on the results, a further discussion and assessment on the reference annual apparent losses is done and proposes appropriate values, as part of the traditional method of assessing AL.

The chapter also evaluates apparent losses using two main alternative methods for the same study and evaluates and compares the results to those obtained through the traditional weighted meter accuracy methodology. Gaps in the methodologies are also discussed with recommendations for further development.

5.1 Traditional Assessment Method

The traditional assessment of apparent losses follows the methodology described in §2.2.1 and schematically represented in Figure 2-4 to estimate the weighted meter error. The output of consumption patterns from Table 3.5 and the meter errors of Table 4.6 are combined to give the weighted meter accuracy as detailed in the following sections.

5.1.1 Weighted Meter Accuracy

The meter errors of Table 4.6 are translated into meter accuracy (1 – ε) in Table 5.1 for all the respective meter models and sizes.

Table 5.1: Average Meter Accuracy

Size and model	Average Accuracy per flowrate, %									
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
15mm	**64.20**	**79.11**	**88.78**	**95.20**	**98.71**	**99.00**	**100.00**	**99.19**	**98.64**	**98.27**
Model 1	57.73	75.77	85.83	91.47	96.08	98.00	97.65	95.84	95.80	93.44
Model 2	65.46	79.92	90.55	94.78	98.22	100.43	100.41	99.55	98.95	98.77
Model 3	41.09	83.89	77.35	93.13	99.64	100.08	99.90	99.31	99.45	99.32
Model 4	57.57	60.55	75.82	99.29	100.90	101.64	99.43	98.04	97.83	97.39
Model 5	69.18	80.94	89.49	98.13	101.07	94.87	100.25	100.20	99.33	99.51
20mm	**59.90**	**79.65**	**86.68**	**92.51**	**94.13**	**94.97**	**93.96**	**96.27**	**96.08**	**96.01**
Model 2	67.17	74.17	79.49	86.86	88.35	88.82	93.60	92.99	92.75	92.70
Model 3	50.72	88.78	98.98	101.43	100.10	100.63	92.44	99.21	99.00	98.85
Model 5	84.95	99.05	102.16	103.07	101.44	100.99	100.41	100.30	100.37	100.39
Model 6	37.33	42.91	46.27	67.49	94.12	101.67	101.11	99.52	99.68	99.96
25mm	**88.04**	**99.00**	**99.86**	**101.30**	**101.34**	**101.57**	**100.29**	**99.89**	**99.66**	**99.45**
Model 2	98.35	100.51	100.61	101.33	101.32	101.64	100.44	99.96	99.65	99.35
Model 3	67.40	95.99	98.38	101.25	101.39	101.41	100.00	99.74	99.68	99.64
Global	**63.56**	**79.75**	**88.46**	**94.58**	**97.47**	**97.92**	**98.28**	**98.37**	**97.94**	**97.66**

The consumption profiles of Table 3.5 are also recalculated from the raw data obtained from the field to correspond with the meter testing flow rates of Table 4.5 for subsequent calculation of the weighted meter accuracy. The consumption profiles corresponding to the flowrates of the 15mm and 20mm are shown in Table 5.2 and Table 5.3 respectively, where the flowrates correspond to the equivalent flowrate of Q1 to Q10 in Table 5.1. It must be noted that this recalculation does not change the consumption pattern but only reapportions it to suit the flowrates adopted in the meter testing process.

The starting flowrate of each of the individual meters tested nor the accuracy at that starting flowrate were not evaluated, as recommended in Arregui et al. (2017), due to time and resource constraints.

It was therefore assumed that the meters had an error of -100% (i.e. no registered flow) at a flowrate just above zero and that the error was linear to the first evaluated test flowrate.

Table 5.2: Consumption Profile for 15mm

Category	Proportion of flow at flowrate										
	7	**15**	**22**	**30**	**60**	**120**	**750**	**1500**	**2250**	**3000**	**5000**
Residential	5%	9%	6%	9%	22%	14%	27%	8%	1%	0%	0%
Business	3%	5%	7%	10%	22%	14%	31%	7%	0%	0%	0%
Public Benefit Organisation	2%	3%	3%	8%	26%	21%	31%	4%	1%	0%	0%
Multi-Residential	2%	4%	3%	6%	20%	20%	33%	10%	1%	0%	0%
All	**3%**	**6%**	**5%**	**8%**	**22%**	**17%**	**30%**	**7%**	**1%**	**0%**	**0%**

Table 5.3: Consumption Profile for 20mm

Category	Proportion of flow at flowrate										
	12	***25***	***38***	***50***	***100***	***200***	***1250***	***2250***	***2650***	***3700***	***5000***
Residential	12%	9%	10%	1%	22%	21%	25%	2%	0%	0%	0%
Business	6%	9%	11%	0%	23%	19%	30%	1%	0%	0%	0%
Public Benefit Organisation	4%	5%	9%	0%	26%	30%	24%	2%	0%	0%	0%
Multi-Residential	4%	5%	7%	0%	20%	28%	33%	3%	0%	0%	0%
All	**7%**	**7%**	**9%**	**0%**	**23%**	**25%**	**28%**	**2%**	**0%**	**0%**	**0%**

The weighted meter accuracy is obtained by using the modified Equation 4.3 and involves the addition of the products of the average accuracy at a particular flowrate and the proportion of flow consumed at that same flow range. As an illustration, the weighted accuracy for the residential category when supplied by a Model 6 20mm meter is as follows;

0.5(0+37.33) * 0.12 + 0.5(37.33+42.91) * 0.09 + 0.5(42.91+46.27) * 0.1 + 0.5(46.27+67.49) * 0.01
+ 0.5(67.49+94.12) * 0.22 + 0.5(94.12+101.67) * 0.21 + 0.5(101.67+101.11) * 0.25
+ 0.5(101.11+99.52) * 0.02 + 0 + 0 + 0
= 2.24 + 3.61 +4.26 +0.57 + 17.78 + 20.56 + 25.35 + 2.01 = **76.38%**

On the other hand, if the unmodified weighted error equation, Equation 2.4, is used the result is as follows;

37.33 * 0.12 + 42.91 * 0.09 + 46.27 * 0.1 + 67.49 * 0.01 + 94.12 * 0.22 + 101.67 * 0.21

+ 101.11 * 0.25 + 99.52 * 0.02 + 0 + 0 + 0

= 4.48 + 3.86 + 4.63 + 0.67 + 20.71 + 21.35 + 25.28 + 2.0 = **82.98%**

Overall, there is a difference in the weighted meter accuracy with the unmodified equation giving a better value of meter accuracy, which might be overstated. The differences in the estimates of the accuracy are mostly at the low flowrates where there is a rapid change in meter accuracy. From a practical point of view, the uncorrected equation does not make sense as the flow does not occur at a point but over a range of values, with the mid-point accuracy being a better representation than either end of the range. The smaller the difference between successive flowrates the better and more accurate are the results but with the trade-off of increased cost and time. The modified equation is therefore seen to be a better representation of the weighted meter losses but both equations can be improved by factoring in the meter starting flowrate and the accuracy at that flowrate.

For the estimation of the apparent losses within Johannesburg, the calculation of the weighted meter accuracy for the six meter models assumes for simplicity that all the consumers exclusively use each meter model in turn. These results are shown in Table 5.4 and graphically in Figure 5-1 and Figure 5-2 but exclude 25mm meters as these were too few for meaningful consideration.

Table 5.4: Weighted Meter Accuracy

Meter size & model	No. of	Ave Age	Ave Read	Res	Business	Utility	Multi	All
15	**85**	**10.9**	**3129**	**90.95**	**93.57**	**95.16**	**95.17**	**93.70**
Model 1	10	14.8	1193	88.23	90.96	92.68	92.73	91.14
Model 2	51	9.6	2774	91.46	94.08	95.59	95.66	94.19
Model 3	3	22.4	11039	88.86	92.02	94.18	94.33	92.34
Model 4	3	26.9	7824	89.54	92.54	95.40	95.22	93.19
Model 5	18	7.5	3110	91.61	94.00	95.46	95.28	94.08
20	**35**	**12.6**	**6776**	**83.33**	**86.76**	**89.43**	**89.23**	**87.19**
Model 2	17	8.7	2867	80.51	83.73	85.52	85.91	83.90
Model 3	14	19.9	11565	86.35	90.03	93.60	92.45	90.61
Model 5	2	13.7	6977	93.38	96.65	98.34	97.74	96.50
Model 6	2	3.3	6271	76.00	79.78	84.51	86.37	81.75
Global	***123***	***11.51***	***4138***	***88.85***	***91.71***	***93.61***	***93.54***	***91.92***

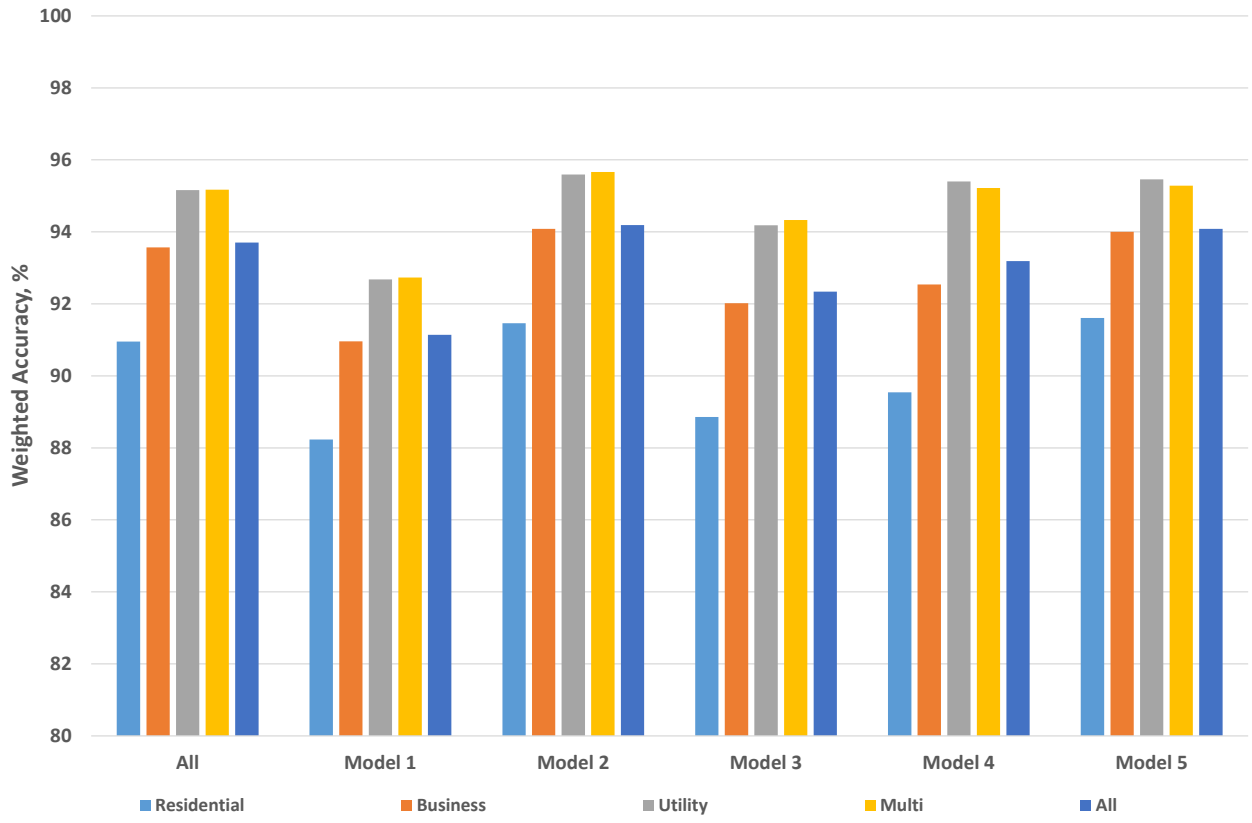


Figure 5-1: Weighted Meter Accuracy – 15mm Meters

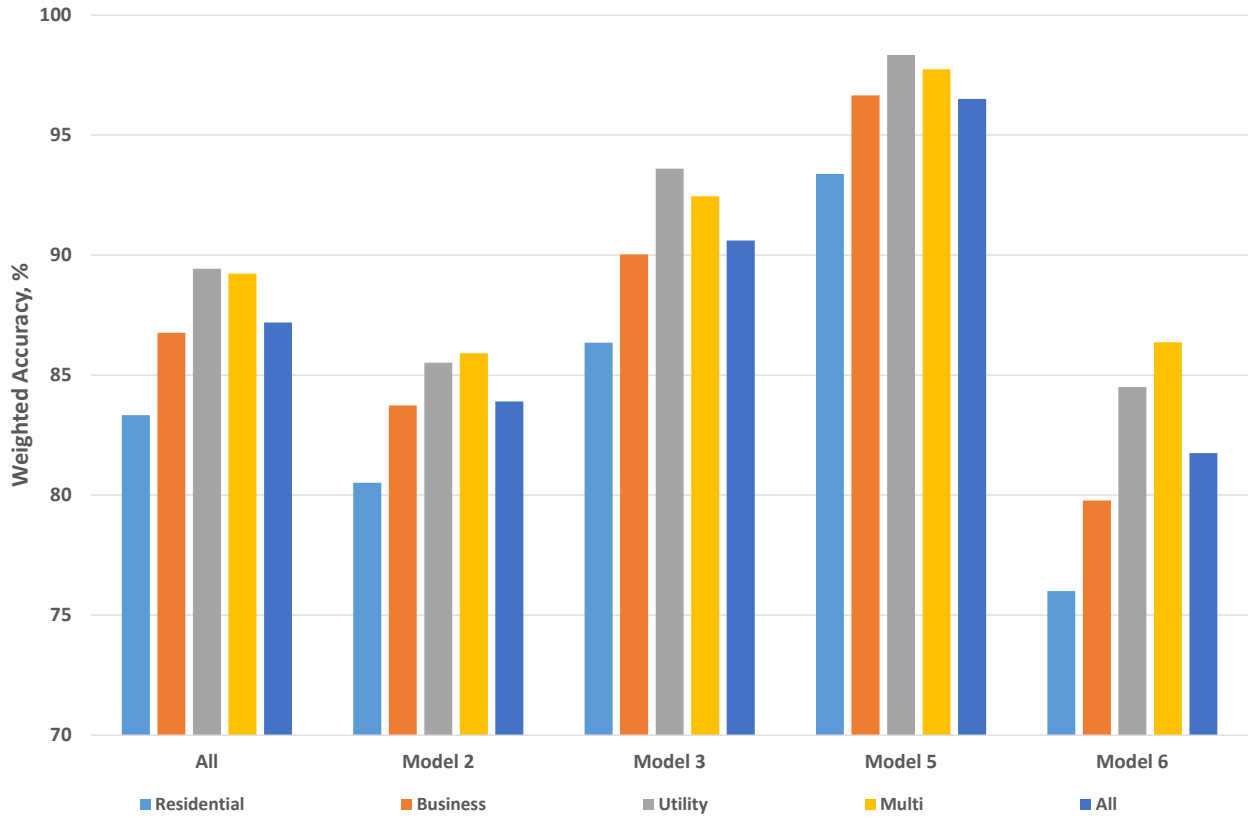


Figure 5-2: Weighted Meter Accuracy – 20mm Meters

As would be expected 15mm meters generally perform better than 20mm meters with the differences in accuracy between the two being indicative of the error due to meter sizing. Residential consumers also have the least meter accuracy for any meter model due to their consumption pattern that has more proportion of flow in the low flow regions.

In order to estimate the metering error, the accuracy values in Table 5.4 are superimposed with the consumption profiles in Table 5.2 and Table 5.3. Due to the lack of adequate information of the meter models and sizes, the estimation of apparent losses for the Johannesburg Water fleet of meters is carried out by assuming an "average" meter model for each of the meter sizes. In addition, the recorded consumption per category, in contrast to the numbers of meter models that are adopted in some studies, is used to estimate the global weighted apparent losses. This is because consumption provides a better basis for comparison, taking into account the fact that apparent losses are a proportion of the volume passing through the meter. In fact, there are instances where the number of meters has little bearing on the volume consumed or unrecorded, such as in the case of fire flow connections that may be many but hardly ever register any flow. The results are summarised in Table 5.5.

Table 5.5: Estimated Meter Error

Consumption Category	Ave Monthly Consumption, kℓ	Meter Error, % (15mm only)	Meter Error, % (20mm only)
Residential	12,188,587	11.15	16.67
Business	3,830,024	8.29	13.24
Public Benefit Organisation	1,179,206	6.39	10.57
Multiple Residential Dwelling	4,275,743	6.46	10.77
Total/Weighted Averages	***21 619 060***	***9.44***	***14.55***

Apparent water losses due to metering errors in Johannesburg are estimated to be between 9.4% and 14.6% of the measured water volume, depending on the adopted meter sizes. The use of 25mm meters, as has been observed in the field, could give even higher losses of at least 23% based on the assumption of the 55% increase from upsizing a meter as is the case between the 15mm and 20mm meters. In the conservative case of an equal split of 15mm and 20mm meters, the meter under-registration can be estimated at 12%, which is higher than the default value of 10% for cities

with direct supply suggested in Mutikanga et al. (2011a) or the 8% estimated for Utilities with good quality water but poor meter age and accuracy as in Seago et al. (2004).

As expected, domestic consumers, being the most numerous and with higher proportion of low flows through the meter, contribute the most to apparent water losses and therefore proper sizing and management of their meters has the greatest impact on the reduction of these losses. The results also demonstrate the need for keeping and maintaining a comprehensive database of meter, consumer, and consumption information as in the absence of such information, the use of global average results (as in Table 5.4) can mask salient results that can inform decision-making and deployment of resources by the Water Utility. Furthermore, the challenges experienced in failing to test all the meters that were scheduled underscore the fact that not only is the traditional method costly, but it is also intricate and may not always yield the required results.

5.1.2 Reference Annual Apparent Losses (RAAL)

The guideline of Vermersch et al. (2016) provides a default value for the RAAL as 5% of the billed authorised consumption. No basis is provided for the choice of 5% in the guideline document. The provision of guideline or reference values is useful and has its place in practice, but these need to be informed by reality and must be applicable to local variables and conditions.

The permissible tolerance or error allowable in South Africa, provided in Table 2.6, is 8% for q_{min} and 3.5% for all other flowrates for used meters and 5% and 2%, respectively, for new meters. This essentially means that this is the standard that utilities and meter manufacturers are required to comply with for both new and used meters. The task of utilities is therefore to manage their meters to remain within the 5% - 8% tolerance for q_{min} and 2% - 3.5% for the other flowrates. Essentially, the meter accuracies at these boundaries, together with the utility specific consumption profile, should ideally inform the quantum of an annual reference value of apparent losses that a utility should aim to achieve.

To illustrate this further, if a Class C 15mm meter is adopted as standard for residential consumers in Johannesburg, an analysis of the RAAL would consider the accuracy of a new 15mm meters

against the consumption profile of the domestic category. Since no tests on new meters were carried out in this study, the error curve is reproduced of a random meter of 15mm size with a reading of less than 500kℓ that was tested as part of this research and presented in Figure 5-3.

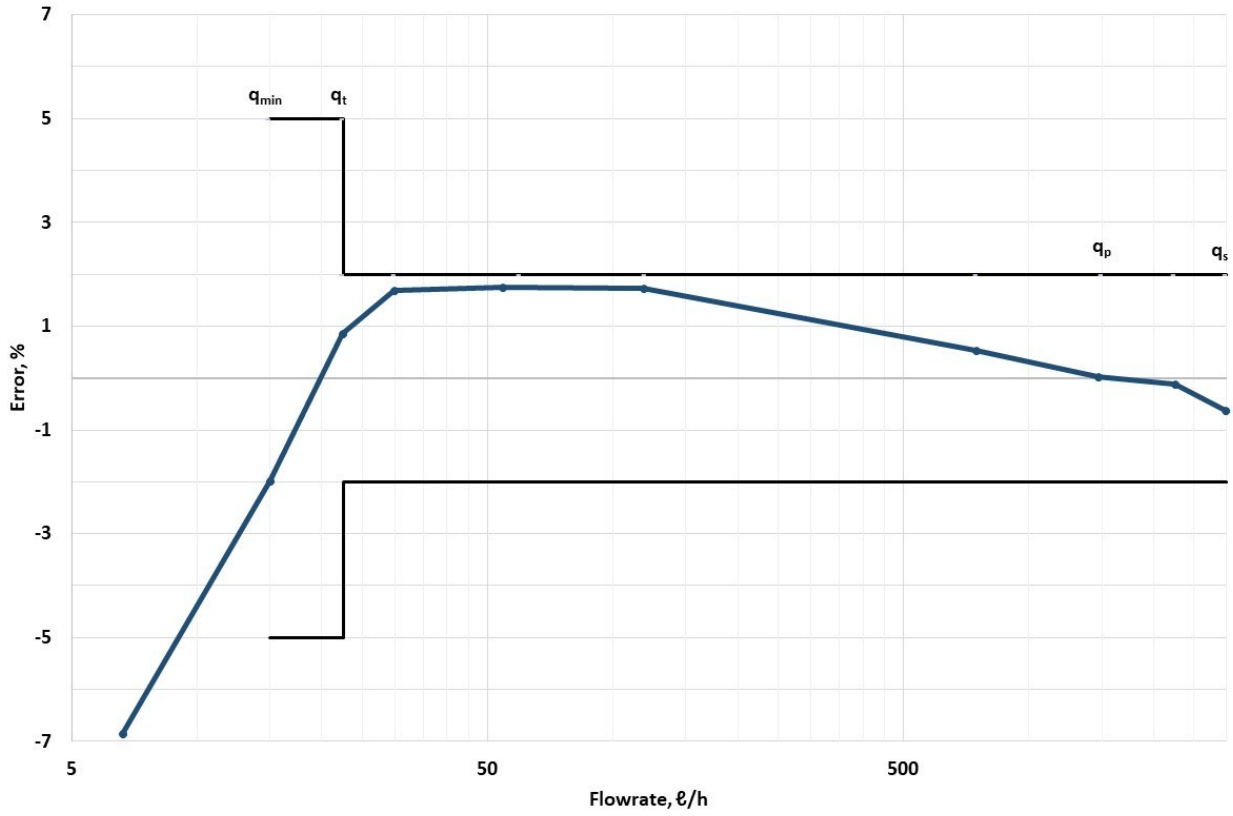


Figure 5-3: Error Curve for a random 15mm meter

Superimposing the consumption profile of the residential consumers, shown in Table 5.2, against this relatively new meter enables the calculation of the weighted meter accuracy. The estimated meter under-registration error is **4.2%**. In a year's time, assuming a degradation rate of 0.8%, the estimated meter error would be already at 5%. This therefore raises questions as to whether a value of RAAL of 5% is a reasonable reference value if the meters at a utility will only be a year or two old, as in this example. Another example is research done in Spain described in Arregui et al. (2018) where new Class B 15mm meters were tested and the weighted meter error was found to be −3.5% and −3.9% for two meter models. It is important to note that their consumption profile is different and permits the use of Class B meters, whereas for the consumption profile of Johannesburg with higher proportion of consumption at low flows, the use of Class B meters will give a higher weighted meter

error of, say, about 6 – 7%. In either case, the reference RAAL value would be breached in two to three years, hence an unreasonable target for utilities to adopt in their meter replacement.

A more scientific approach is therefore required for a more realistic value of RAAL that considers the weighted meter accuracy of new meters in an area that is based on both the metrology standard and the applicable consumption profile. It is therefore recommended that the RAAL value be based on the best available estimate of the weighted meter error of a new Class C 15mm meter or suitably equivalent approved meter in that jurisdiction, plus and an additional allowance for some meter degradation of at least three to four years. In the case of Johannesburg, an appropriate RAAL value would be 7.4% of the metered consumption.

5.1.3 Unauthorised Consumption

The estimation of the contribution of unauthorised consumption to apparent losses involves the sampling of consumers and determining the percentage of irregular connections from that sample, or the use of historical information of reported cases of unauthorised consumption (Mutikanga, 2012). In this study, historical information was utilised together with data from the sampled properties for both meter testing and consumption characterisation.

The historical record of reported illegal connections within Johannesburg Water is shown in Figure 5-4 from 2002 to 2014. The peaks in the data coincide with meter audits that were done by the municipality and they tended to be within specific areas of a suburb or township. In overall, the occurrence level of confirmed unauthorised connections was estimated at 2% of the number of connections surveyed, which were predominately within the residential sector. The contribution of unauthorised connections estimated as a percentage of the total monthly volume in Table 5.5 translates to 1.13% of the authorised consumption. Unfortunately, the data for the actual confirmed illegal connections from the conventionally metered areas (non-prepayment) was very limited to allow for further analysis that permits the use of data mining applications on such data.

From the slightly more than 650 properties that were visited during the meter testing and meter logging period, only 3 illegal connections were found which constitutes a very low value of 0.46% of

the sampled properties. It is likely that this small occurrence is linked to the perception that illegal connections tend to be concentrated in specific areas and therefore a purely random sample may fail to capture the actual prevalence of illegal connections.

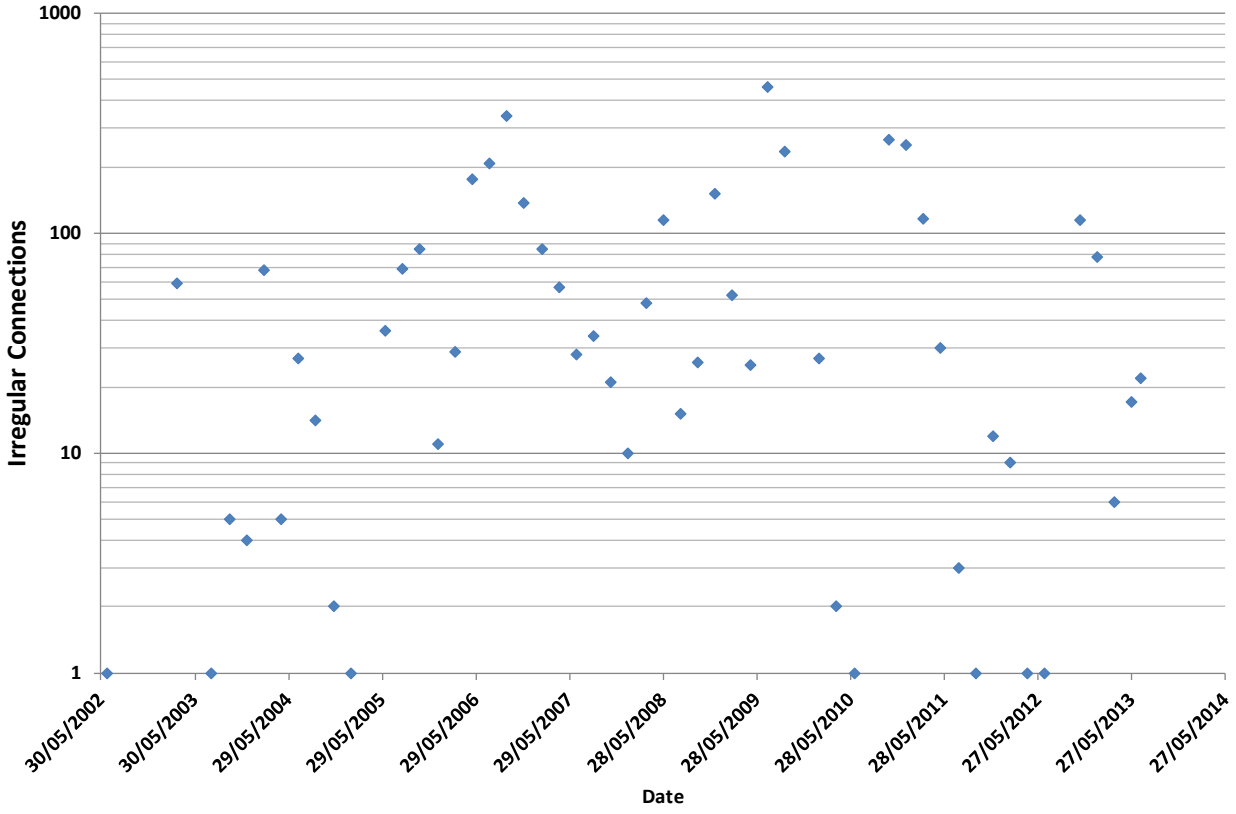


Figure 5-4: Historical Record of Unauthorised Connections

In the absence of better data, the occurrence rate of 2% was adopted as the estimate for unauthorised consumption within the study area to include all other consumer categories. It is important to note that this study excluded prepayment and flat-rate areas such as the townships of Soweto, Cosmos City, Ivory Park, and Orange Groove. Some of these areas have reported high incidences of unauthorised connections and therefore the estimated 2% may not apply there.

5.1.4 Data Acquisition Errors and Unmetered Consumption Errors

Using the meter sample of over 650 properties, an evaluation on site was done on the recorded meter reading and the actual meter reading. There was a general agreement between the two in

most of the cases with the exception where meters have been changed and the meter database was not yet updated.

However, it is public knowledge and also reported in literature such as Mazibuko (2013) that there is a billing crisis within the City of Johannesburg that is systemic. Due to these reported challenges, data acquisition errors and the estimates for unauthorised consumption errors were not evaluated further.

5.2 Meter Change Method

The meter change method is described in detail in §2.2.4 and follows the methodology proposed in Couvelis & van Zyl (2015). In this study, the method is implemented on a sub-set of the data that was clearly labelled as meter change data and secondly on an unlabelled dataset with all the consumption record of Johannesburg Water. All statistical analysis was done using a combination of the R Statistical Software (R Core Team, 2017) and MS Excel 2016.

5.2.1 Labelled Dataset

Johannesburg Water provided a list of over 58,000 properties that had changed meters in the period July 2010 – June 2013 from their maintenance management system. After preliminary data cleansing and elimination of invalid data such as duplicates or work orders without adequate property information, the composition of the meter changes is shown in Table 5.6.

Most of the properties (91%) have the expected one meter change over the three-year period while a surprising 101 properties have had more than ten meter changes over the same period. While the latter might include properties with multiple meters, the number of meters remains unusually high considering the period of analysis. In fact, within three years, it is not expected to have more than one meter change unless the consumer in questions is a high volume consumer.

However, these findings were not entirely surprising considering the observations and findings from field surveys for the meter logging and meter testing activities. In many respects, the meter management process within Johannesburg Water was poor thereby undermining the credibility of the meter change database.

Table 5.6: Meter changes per property

No. of meter over period	*No. of properties*	*Percentage*
2	41562	91.45%
3	2921	6.43%
4	582	1.28%
5	144	0.32%
6	69	0.15%
7	29	0.06%
8	25	0.06%
9	8	0.02%
10	8	0.02%
>10	101	0.22%
Total	***45449***	***100%***

An evaluation of a few of the properties with at least ten meter changes revealed that in some instances fire meters were be changed a few times in a year because it registered the same reading which was flagged as a "stuck meter". Such observation necessitated further data cleansing to ensure that only records that met the following criteria remained;

- a 1:1 match of meters was identified, that is an old meter with a matching replacement meter of a different serial number and an initial reading date in proximity to the last reading date of the old meter;
- At least six months of non-zero data were recorded before and after the meter change event.

This reduced the dataset to 39,634 records and, for those meters without any installation dates, the serial number logic from two known manufacturers was used to determine the age for a total of 6,560 meters within the dataset. The distribution of the data is summarised in Figure 5-5.

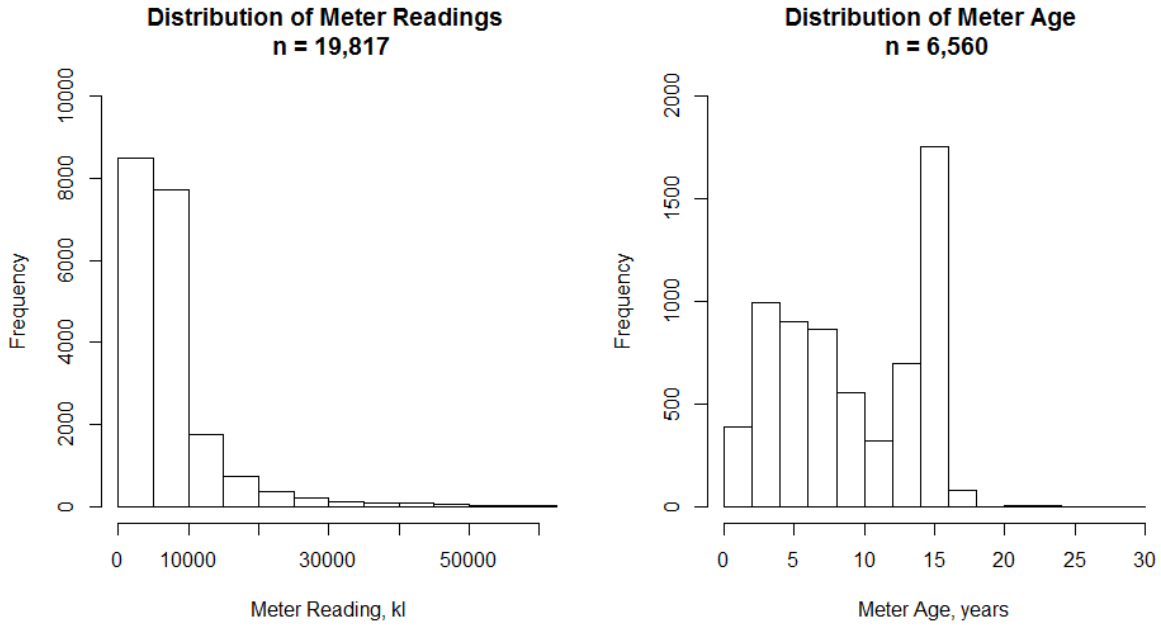


Figure 5-5: Distribution of meter change data.

The readings show an expected skew to the left while the distribution of meter age was atypical with a high number of meters in the 14 – 16 years range. The metering error was calculated using Equation 2.6 and the relationship of the age and the reading of the old meters with the meter error were analysed. For all records with an absolute meter error of at most 100%, the degradation of the meter error with the meter reading is depicted graphically in Figure 5-6.

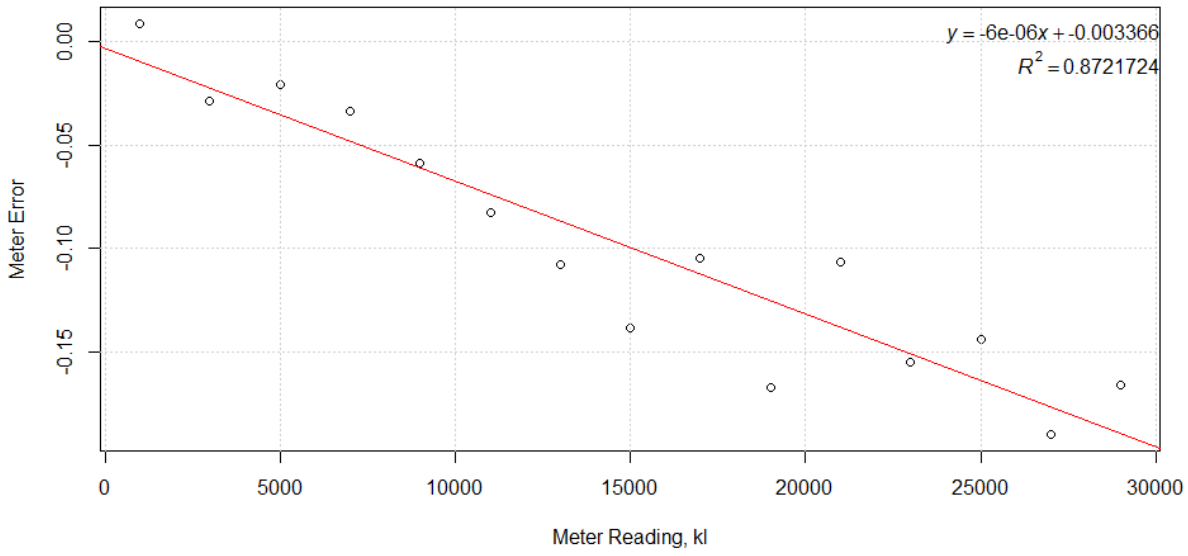


Figure 5-6: Error degradation with meter reading - Meter Change

The degradation rate with meter reading is 0.64% per 1,000kℓ of flow through the meter which is lower than the 0.9% of consumption per 1,000kℓ in Couvelis & van Zyl (2015) for the eThekwini Municipality meters. The meter composition in eThekwini and Johannesburg is not expected to differ that much to warrant such differences. With the average reading from the Johannesburg Water meter database as at end of June 2015 of 7,329kℓ, the estimated apparent losses due to metering error is 4.7%, which is similar with the estimated 5% error in Couvelis & van Zyl (2015). However, Couvelis & van Zyl (2015) includes 2.2% as the estimated meter under-registration due to on-site leakage while the estimate in this study does not. The under-registration of 2.2% is likely understated as the weighted error of a new meter has been shown to be 4.2% in §5.1.2 and increases with meter usage and time. In any case, even allowing for similar levels of meter under-registration both these figures are much lower than the likely estimate from the empirical method.

A plot of the degradation rate against meter age showed randomly scattered points with no meaningful relationship, which may highlight the correct use of meter readings as opposed to meter age for the assessment of meter degradation. It is however possible that such relationships are masked by the meter change practises of JW.

5.2.2 Complete Consumption Dataset (Unlabelled)

From the meter reading records, whose compilation and processing is described in detail in Chapter 6, an extract of properties that had exactly one meter change, at least a one-year record and no meter clock-overs or abnormal readings were extracted from the meter readings database comprising over 600,000 meter reading records. These limitations were imposed to reduce the amount of data cleansing required. The meter change was presumed to have occurred if, for an individual property, the last reading the record of a specific meter number (the old meter) is within proximity of the first reading of a different meter number (the new meter).

The data was processed and patched using the Zeileis & Grothendieck (2005) package within R with further processing done within MS Excel to give a total of 111,448 records (55,744 properties). This

is almost three times the amount of data that was used in §5.2.1. The distribution of the data is shown in Figure 5-7.

The distribution of meter reading follows the expected skewedness towards meters with low readings due to the dominant residential consumers. The meter age data is much more random with the various peaks corresponding to the meter installation/replacement activities of the company such as the meters that are five years old

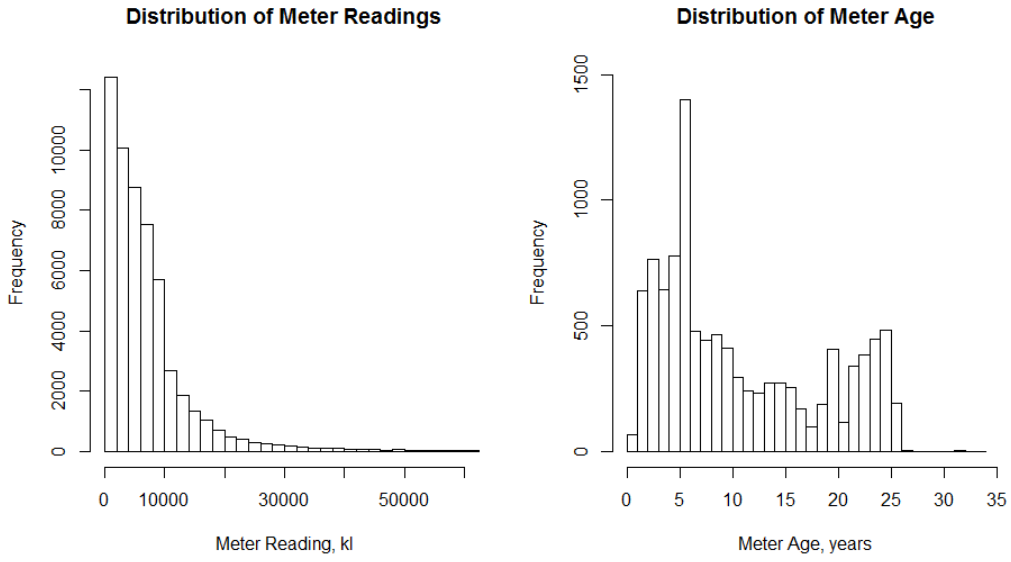


Figure 5-7: Distribution of meter reading data – Meter Change Data

Despite the increase in data, several variations in the analysis of the complete dataset did not uncover any credible relationships between meter error and both meter age and meter reading. Coupled with the absence of accurate information on meter properties such meter model, it was not possible to disaggregate the analysis any further.

5.2.3 Key Observations

For the smaller labelled sample, the estimated global value of 4.7% for the apparent loss was found to be less than 50% of the 11.15% value from the average weighted meter accuracy method for 15mm meters. Comparison with the apparent losses value of 16.67% for residential consumers, who are served by 20mm meters, shows an even greater difference with weighted accuracy method being

137% higher. The meter change method therefore understates the error quite significantly. Some measure of underestimation was expected as the method presupposes new meters are accurate without any under-registration errors, but the magnitude is surprisingly large. Allowing for an initial meter error to the calculation, such as 4% in Arregui *et al.* (2017), will bring the results closer but there is still over 3% difference that is not accounted for.

The difference in result between the label and unlabelled dataset was also quite surprising but can be attributed to several factors. For instance, due to the erratic nature of how meter serial number information is captured and maintained in the Johannesburg database, it is probable that several of the records were in fact not meter changes but the indiscriminate changing of a serial number by adding some other character. While a lot of data processing was done such variations were still possible. In both the labelled and unlabelled dataset about 50% of the records had lower consumption after a meter change. It is therefore plausible that there a sizeable number of meters in the larger sample that masked possible relationship through the cancelling out of increasing and decreasing consumption.

To understand the meter change dynamics in more detail the data of the smaller sample was analysed further, with records extracted to understand how and why specific meters were changed. Figure 5-8 shows a histogram of the frequency of meter change in relation to the average monthly consumption after change compared to before change. The data follows a clear normal distribution with mode being the meters where there was no financial gain or loss. In total, 310,790kℓ/month were gained while 406,564kℓ/month were lost with a net volume -96,000kℓ/month after meter replacement, which is in overall a substantial loss amounting to over R800,000 per month at the base tariffs for 2015/16 for each of the consumer categories.

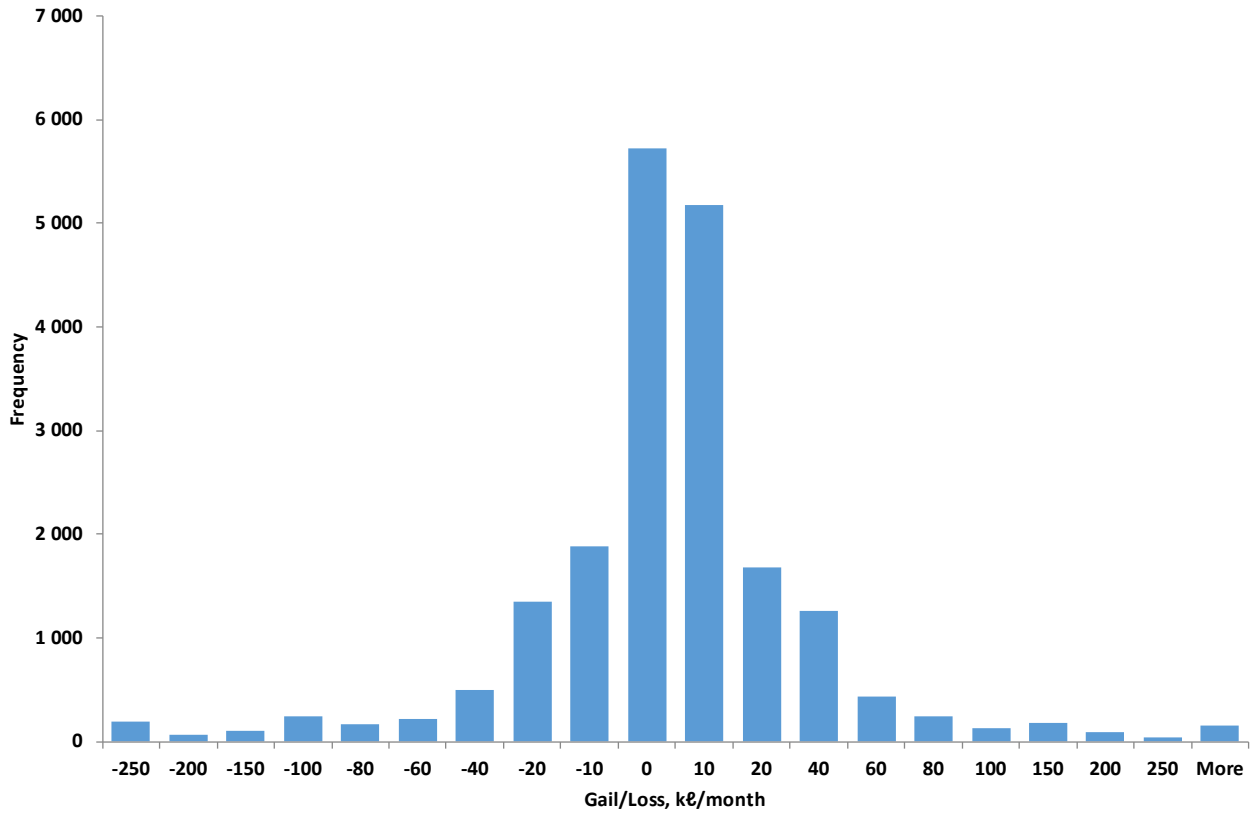


Figure 5-8: Histogram of the Impact of Meter Change

The objective of meter replacement should be at the least to maintain the consumption volume. Further root-cause analysis revealed some unpleasant realities that included the following;

- Meter connection policy was based on pipe sizes of 25mm and 20mm which to most also meant that meters were to be the same size.
- Many meters were upsized during replacement for no apparent reason other than the fact that the replacement meter size was in stock. This was observed in one particular street where bulk meters of 40mm size were all replaced with 80mm sized meters.
- Class A meters were also used in replacing old meters by contractors without adequate supervision. This is despite the fact that this class is illegal for billing purposes.
- Class C meters were also replaced with Class B meters in many instances. The lack of factual meter information on the database also meant that issuing of meters for replacement did not account for the class change.

Figure 5-9 details the time it takes to fix meter related jobs within the City of Johannesburg. More than 60% of the meters get to be replaced or attended to in more than 60 days from the day of the request. Only about 30% of the work is done within 30 days. Such long turnaround times are detrimental to the accurate measurement of water consumption and impact on revenue. These delays were also experienced when a few meters that we logged were non-functional and needed replacement – some of them were still not replaced by the time of completion of the logging exercise.

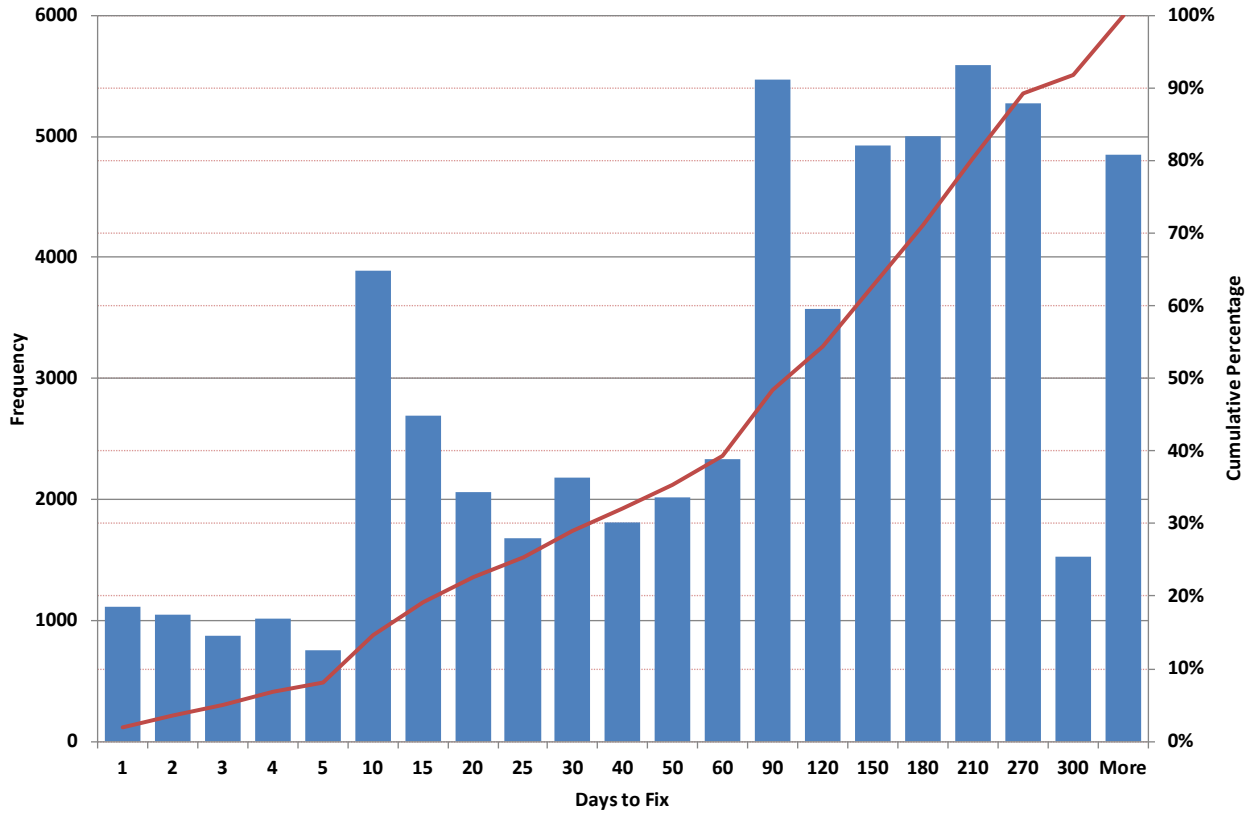


Figure 5-9: Days to Fix Meter Related Jobs

Depending on the nature of the request, there is abundance of evidence of unbilled authorised consumption as arising from delays in replacement of malfunctioned meters. There is no knowledge of the quantification of such water within the municipality. This delay in meter maintenance would therefore also affect the ability to estimate apparent losses using the meter change method.

Quality control for meter replacement was therefore not adequate to protect the revenue of the municipality, which made the meter change methodology unreliable at best. However, the analysis of the meter change records is a very good indicator of the efficacy of meter replacement programs

and may also give insight into the level of improvement or regression in benchmarked apparent loss figures as demonstrated in this case.

5.3 Comparative Billing Assessment Method

The Comparative Billing Assessment method is described in detail in §2.2.4 and follows the methodology proposed in Mbabazi et al. (2015) as interpreted without any variations.

5.3.1 Data Preparation

Similar to §5.2.2, the database of over 600,000 meter reading records from July 2003 to June 2015 was used for this exercise. Properties with only one meter and whose meter readings had not clocked over (i.e. not started again from the beginning) and did not have abnormally high readings were selected for analysis to minimise the impact of outliers and data cleansing requirements. The final number of records used was slightly above 106,000. The data was processed and patched using the Zeileis & Grothendieck (2005) "zoo package" within R statistical software (R) with further processing done within MS Excel. The patching was necessary as the required processing in R could handle any gaps in the data while the procession in MS Excel was to align with the required input into the R packages used. Two twelve-month periods (in 2008 and 2012) with the greatest amount of overlapping meter reading data (before filling) were selected and key information such as the last reading, annual consumption and the age of the meter at each interval were calculated and were presumed to represent the total meter fleet of Johannesburg Water. This information was consolidated into one dataset whose distribution is shown in Figure 5-10.

The data follows similar distribution to that of the meter change method dataset, despite the different numbers of meters analysed. Similarly, no reliable information on meter types and models was available and these variables were therefore not considered further in the analysis.

The records were further classified into similar meter reading ranges of 0 – 2000m^3, 2000 – 4000m^3, and so forth, and the average annual consumption per range was computed for each age category.

Only records with up to a reading of 10,000m^3 and up to an age of 15 years were considered in the analysis to exclude outliers and fewer observation points with higher uncertainties.

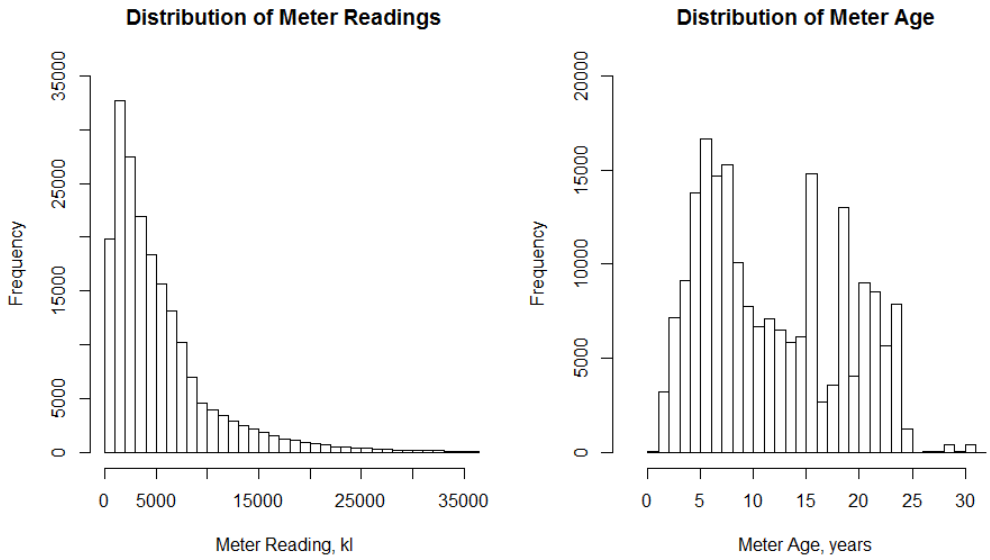


Figure 5-10: Meter Reading Data Distribution – Comparative Billing

5.3.2 Analysis and Results

The variation of the average annual consumption with meter age for the various categories of meter reading, reduced to linear relationships for simplicity, is shown in Figure 5-11. The actual relationship between the two follows an inverse power relationship starting with a rapid decline in annual volumes within the first five years and thereafter declining slowly with time.

To describe the linear relationships shown in Figure 5-11, Equation 2.7 and Equation 2.8 are used separately for each portion of the graphs and the parameters determined for each meter reading ranges. The results from the linear regression are summarised in Table 5.7, with all ranges having a coefficient of determination (R^2) values greater than 96%.

Higher degradation rates of 9 – 14% are evident in the first five year and thereafter reducing to 3 – 5%. The degradation rates in the first few years is very excessive and can amount to an unusual total reduction of 39 – 54% of the annual volume of water that is flowing through the meter in the first five years. While this behaviour may likely by explained by other compounding factors which are not

known at this stage, such degradation rates are in fact excessive and unrealistic as observed from the meter testing results of Table 5.5.

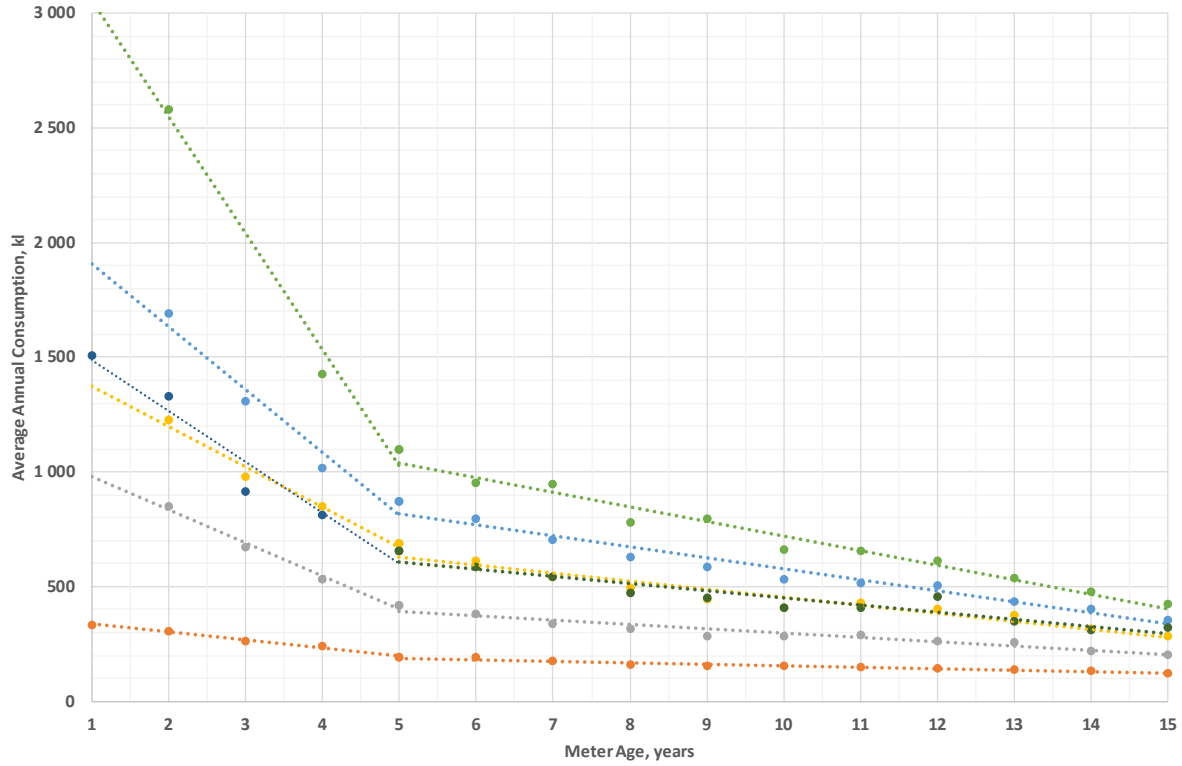


Figure 5-11: Annual Consumption Variation with Meter Age

Table 5.7: Annual Volume Degradation with Age

Meter Reading Range	0 - 5 years			From 5 years		
	β_1	β_1	*d*	β_1	β_1	*d*
0 – 2,000	-35.19	372.55	-9.4%	-6.651	222.83	-3.0%
2,000 – 4,000	-143.74	1121.9	-12.8%	-18.671	483.05	-3.9%
4,000 – 6,000	-174.03	1544.9	-11.3%	-35.318	807.38	-4.4%
6,000 – 8,000	-274.70	2183.5	-12.6%	-47.822	1054.30	-4.5%
8,000 – 10,000	-506.19	3559.2	-14.2%	-63.258	1353.90	-4.7%
Average Curve	***-222.58***	***1711.8***	***-13.0%***	***-31.063***	***761.88***	***-4.1%***

The degradation rates of 1.45% to 6.67% beyond year 5 are within the ranges found by Mbabazi et al. (2015) for different metering technologies used in Uganda. However, applying the degradation of 4.1% from the average curve to the average meter age of 11.5 years as in Table 5.4 and assuming the same rate from year 1 with an initial accuracy of 100%, gives an accuracy of only 61.8%. Such an accuracy value is very low and does not compare with the worst-case scenario of 85.5% of the empirical method. Such results are therefore not acceptable.

Disregarding the different meter reading categories and evaluating the combined uncategorised dataset for the degradation rate of the annual consumption volumes gives the results shown in Figure 5-12. While the coefficient of determination (R^2) is lower at 75% than the previous 96%, the degradation rate is a reasonable 1.25%. The 1.25% annual degradation rate is lower than those found in Mbabazi et al. (2015) which were 6.67%, 4.68%, and 1.45% per year for Models 1 to 3 meters but higher than some guideline values such as in Arregui et al. (2010) of 0.1% to 0.9% per year. Applying the degradation rate of 1.25% per year to the Johannesburg Water meter fleet of an average age of 11.5 years gives a meter accuracy of 86%. This is comparable to the worst-case scenario of 85.5% when using only 20mm meters to determine the weighted meter accuracy. Considering that the method also does not take into consideration the initial meter error, being in line with the worst-case scenario is not adequate. It is however much closer and within range than the meter change method which was much lower and not in the range of values found in other studies.

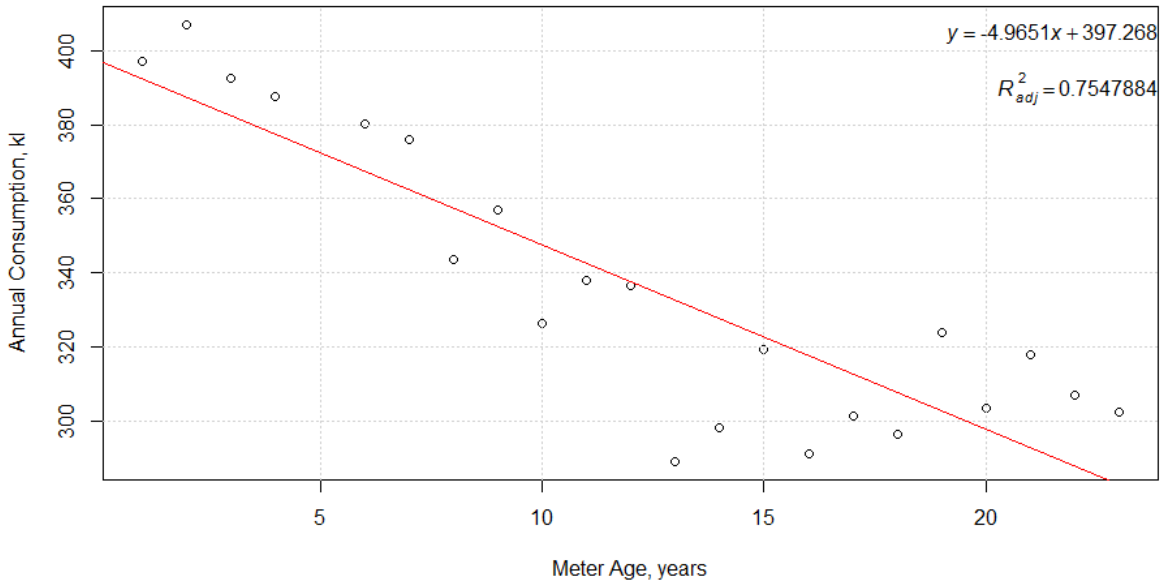


Figure 5-12: Annual Consumption Variation with Meter Age

The comparative billing analysis, as currently implemented provides results which are in proximity the estimates of the empirical method. There is, however, uncertainty in its application as different slight variations in the analysis can produce different results thereby reducing the confidence in the methodology. Its proximity to the weighted accuracy methodology also makes it a better candidate

for further development compared to the meter change method. The method was therefore selected for further development, which is described detail in Chapter 6.

5.4 Summary

This chapter provided a comparative evaluation of apparent losses due to metering errors using three different methodologies, all based on the metering data from Johannesburg Water. This is the first known attempt where the traditional weighted accuracy method is evaluated and compared against two alternative methods for the same study area.

Apparent water losses due to metering errors are estimated by the weighted accuracy method to be, on average, about 12% of the billed consumption, which is higher than the currently used default value of 10%. The actual estimated range of the apparent losses is between 9.4% to 14.6% depending on the comparative number of 15mm and 20mm meters in use. Using the comparative billing analysis method, apparent losses were found to be 14%, similar with the worst-case scenario of the weighted accuracy method of 14.6%. The meter change analysis understated the best-case scenario of apparent losses of 9.4% by about 50% but was in line with previous studies using similar methodology. While the comparative billing method was within the range of the weighted accuracy method, it does not take into consideration the initial error of the meter and therefore the actual value is higher. The different variations of the analysis that provided different results also suggest the need for further refinement of the methodology. As a result of being in proximity of the best available method, this alternative method has been chosen for further development.

The meter change method was also very susceptible to variation depending on how much data was analysed, particularly if the data was not explicitly labelled as meter change data. Further analysis showed that the Johannesburg Water meter change database and processes introduced variables into the analysis that made the output unreliable. The method may therefore be only suitable for utilities with very good data, which is typically not the case in the developing world. However, the evaluation of meter change data provides a lot of useful insights that should be used by utilities to influence how meter changes are implemented and to evaluate the efficiency of such programs.

The suitability of the reference annual apparent losses (RAAL) default value of 5% was also evaluated using the examples of data produced in this study and that from literature. The results show the current default value may not be adequate for many utilities. A proposal is also made that enables estimates of RAAL to be based on local meter regulations and consumption profiles.

Chapter 6 Comparative Billing Assessment 2.0

This chapter is based, in part, on the publications *"Assessment of Apparent Water Losses due to Meter Inaccuracy using an Alternative Validated Methodology"* referenced as Ncube & Taigbenu (2018a) and Ncube & Taigbenu (2019).

The preceding chapters have estimated the various components of apparent water losses, particularly due to metering errors, using currently available methods and applying them to the Johannesburg Water dataset. The weighted meter accuracy methodology has been extensively used over the years and proved reliable. It is however also very involving and expensive for most utilities in the developing world where there is hardly any mention of its use in such areas in the literature. The consensus is evident, even from this research, that despite its accuracy it remains difficult to implement and will therefore not be readily and widely adopted. Chapter 5 evaluated two methods that have attempted to fill the gap of providing alternatives to the weighted accuracy method. One of the methods, the meter change methodology, was shown to understate the losses and is also highly dependent on the integrity of the meter change database and meter change processes within a utility. As the current study uses a database that has been found wanting in this regard, the method was not taken forward for further development.

The Comparative Analysis Method, evaluated in §5.3, performed better and produced results within the range of the weighted meter accuracy methodology. It however requires further development to make it repeatable, reliable and capable of producing results that are closer to those of the weighted meter accuracy methodology. In particular, the following weaknesses need to be resolved;

- The lack of clarity in the steps that culminate at the selection of the final dataset, and other various sub-steps of data cleansing and analysis. This introduces a lack of reproducibility in methodology application leading to subjectivity and verification challenges.
- The failure to maximise the utilisation of the complete consumption dataset, with only 0.4% of the available data over eight months of consumption records used in Mbabazi et al. (2015). In this study (§5.3), only twelve months of the data were used. The method therefore does not consider the full water consumption time series, including any variations that may occur from year to year.
- While the analysis in Mbabazi et al. (2015) catered for different meter models and types, the requirement is that they be labelled accordingly, which is a challenge in many utilities. Coupled with the minimal use of available data, the use of small sample sizes of meters per model introduces uncertainties.
- There is omission on how the increasing consumption over time is handled in the analysis, as was observed in analysis of the meter change method, with only reference to declining consumption.

What most utilities in the developing countries have is a consumption database with monthly, bi-monthly, and so forth, readings that form the basis of billing. Such database may also not have adequate information relating to meter-type characteristics and the consumer categories which is traditionally a requirement for apparent water loss assessment in deriving errors for the different consumption categories and the different metering devices in use. In such contexts it is relevant to find out how unclassified, sub-annual time series databases can be interrogated and analysed methodologically to address the apparent losses assessment challenge. Part of the solution to this problem will require some form of time-series clustering and analysis.

This chapter proposes an improved alternative method for the assessment of apparent loses that addresses the weaknesses of prior methods. The alternative method follows a methodological and repeatable process that includes data preparation to compile a requisite consumption database, data cleansing through several routines, clustering analysis and meter accuracy degradation analysis.

6.1 Data Preparation

Johannesburg Water is responsible for the water metering function, including meter reading while the City of Johannesburg was responsible for performing the billing function on behalf of all the City's service entities. As a result of both organisations maintain two different but related databases, with Johannesburg Water maintaining the meter reading database while the City maintained the billing database. Because of the reported billing challenges within the city, a decision was made to use the meter reading database to minimise any further errors that might be introduced when meter reading information is transferred into billing records.

However, on interrogating the meter reading database it became clear that the structure of the database was not setup to capture the meter reading history per consumer but rather the monthly reading data for all the meter reading routes. It was therefore a necessary prerequisite to create the required time series for all consumers from existing data as part of this study.

6.1.1 Database Creation

A total of 144 monthly meter reading flat files, in text format, that were submitted for billing purposes from July 2003 to June 2015 were made available by Johannesburg Water. The files were in two different formats owing to a change of systems in June 2010, with one format corresponding to the period prior to the change and the other to the period after the change. Each monthly reading file was first converted to an MS Excel file to extract only the required fields and thereafter was processed within SQL Server Integration Services (SSIS) in Microsoft Visual Basic for aggregation into a new database within Microsoft SQL Server 2016.

For each of the monthly flat files, a series of activities were repeated that included opening a file, looking at specific fields and importing all data with non-zero values for the meter reading in that month. No personal data was collected from the files but only the meter ID, the property ID, consumer category and the meter reading. The property ID consisted of an alphanumeric 21-digit ID of the Cadastral Land Parcel from the Survey General and referred to as the SGID. In the case of Johannesburg, the SGID codes prior to June 2010 had more than 21 digits and were changed to the 21

digits in 2010. Due to this change, for the data prior to June 2010, an additional step in the processing was required to map the old SGIDs to the new ones using information obtained from Johannesburg Water. Data Quality Services (DQS) tools that are found within SSIS were used to map the different properties in different periods that had similar meter numbers for consolidation. The dataflow process for the pre-2010 data is shown in Figure 6-1 while Figure 6-2 summarises the process for import meter reading data collected after June 2010.

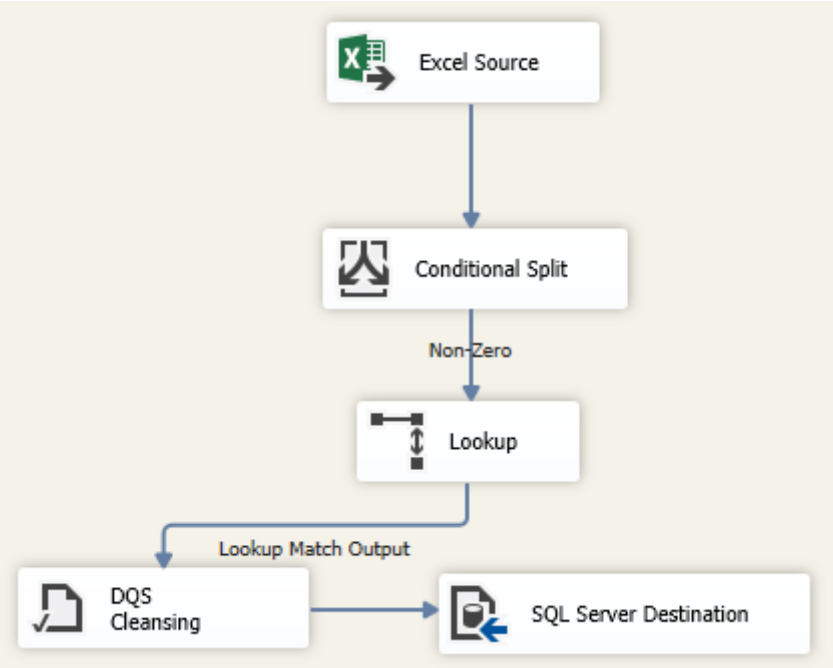
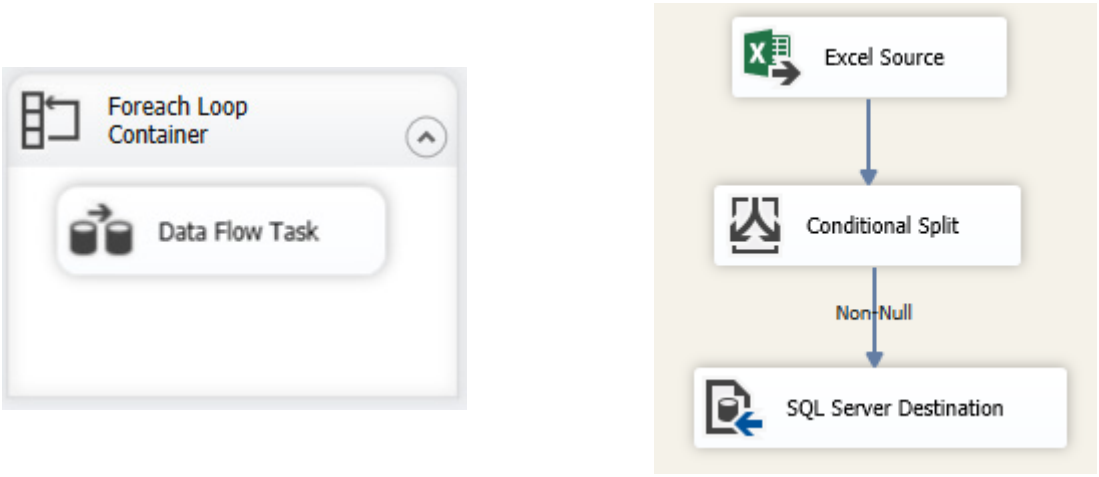


Figure 6-1: Data Import within SSIS – Pre 2010 Data

The data from both processes illustrated in Figure 6-1 and Figure 6-2 were all consolidated in one staging database (a temporary database) where the output data was stored before further processing. A sample of the staging database is shown in Figure 6.3 and the full database had over 35.4 million records in total. After fully populating the staging database with the required meter reading data, it was necessary to transform the data to a time series for each individual and unique combination of property ID and meter number. This was achieved using a stored procedure that was developed and coded in SQL Server to pivot the data for each unique property and meter. An extract of the output database is shown in Figure 6-4 and had a total of 738,804 unique time series of meter reading data. The SQL stored procedure used to achieve this is reproduced in §B.1.1 in the appendices.



(a) Control Flow (b) Data Flow Tasks

Figure 6-2: Data Import within SSIS – Post 2010 Data

SGID	MeterNo	Reading	ReadingDate	ReadID	Period
T0IQ130A00000000699000140001400000003	CKWC056	3628	2003-07-04	9990	200307
T0IQ063A00000000186000010000100000003	80514	35043	2003-07-04	9991	200307
T0IQ064A00000000711002790027900000003	261861	76438	2003-07-04	9992	200307
T0IQ063A00000000150000010000100000003	232335	9404	2003-07-04	9993	200307
T0IQ058A00000001227000720007200000004	79087	6615	2003-07-04	9994	200307
T0IR00290000000004730000000	238966	31044	2003-07-04	9995	200307
T0IQ345A00000000046100000RE699000000004	4677	1410	2003-07-04	9996	200307
T0IQ344A00000000005000000000D01000000	KZN810	22332	2003-07-04	9997	200307
T0IQ269A00000000591000020000200000004	94026448	72	2003-07-04	9998	200307
T0IR00580000000000550000000	3200	88544	2003-07-04	9999	200307
T0IR00640000000008510000000	68688	30604	2003-07-04	10000	200307

Figure 6-3: Staging Database Fields

SGID	MeterNo	201106	201107	201108	201109	201110	201111	201112	201201	201202
T0IQ0216002000036660000000	#00000012	8102	8132	NULL	8203	8226	8257	8285	8307	8331
T0IR0253000000003280000000	#00000027	94745	95181	95690	96102	96464	NULL	NULL	97644	154
T0IQ0220000000003360000300	#00000037	76071	76240	76465	76718	76920	77139	77357	77573	77697
T0IR0493000000002080000100	#00000067	11643	11669	NULL	11736	NULL	11774	11810	11836	11832
T0IR0431000000000870000000	#00000075	38054	38080	38151	38234	38435	38519	38615	NULL	38633
T0IR0378000000006230000000	#00000080	28895	28946	28978	29032	29076	29124	29156	29184	NULL
T0IR0045000000006340000000	#00000106	10238	10759	11324	11907	12411	12939	13476	14221	14767
T0IQ0217000000015460000000	#00000111#...	NULL	6091	6137	NULL	71	120	NULL	NULL	NULL
T0IQ0217000000014730000000	#00000111#...	5909	5926	5938	5955	5973	5985	5997	NULL	6026
T0IQ0217000000014100000000	#00000111#...	5429	5454	5476	5512	5540	5578	5608	5631	5651

Figure 6-4: Consumption Time Series

The transposed time series illustrated in Figure 6-4 revealed that despite the initial matching of prior SGIDs with current ones, there were still several duplicates that required data cleansing. Most of the duplicates emanated from subtle changes in meter serial number or the change in the SGID number due to stand consolidation, subdivision or some other reasons.

6.1.2 Data Cleansing

Due to the enormity of the database, it was not possible to eyeball all records individually to be able to clean it adequately. The SQL Data Quality Client was again utilised to identify and combine duplicate meter numbers and property IDs. This meant iteratively finding meters with similar meter numbers from the same property IDs and thereafter finding properties with the same meter number until there was minimal change. In the process, a few lookup tables had to be created and used to repeat the processes described in §6.1.1 to derive the final dataset. An extract of the look-up tables used is illustrated in Figure 6-5.

Inevitably, there was a lot of manual verification that was required after each sub-process, which was a laborious and time-consuming exercise. In the end, the easiest verification was of properties with the same property IDs but with slightly different meter numbers such as a change in just one digit. Such meter numbers were verified by looking at the actual consumption and checking if the new number represents a continuation of the old one or not. At some point, a call had to be made to proceed with the data with the assumption that adequate cleaning had been done. The final database after this initial cleansing had slightly over 600,000 meter reading records, a reduction of over 120,000 duplicate records.

Several additional criteria had to be introduced to improve on data accuracy owing to the known additional duplicates and other irregularities. After several iterations, only records which met the following criteria were retained, in order of application;

- A sufficiently long record of at least 60 individual readings, that is, a minimum of five years of observations. Five years was deemed a sufficiently long record for a meter to exhibit degradation trends;

- Less than 20% patching of missing data required within the dataset, i.e. less than 1 year of data was missing. The five years of observations prior to filling will be supplemented by filling any gaps in the data;
- Meters that did not clock-over (i.e. did not start again from the beginning) and whose minimum/maximum meter reading was at the beginning/end of the record. This was to ensure that the analysis is based on an expected increase of the meter reading and to minimise the cleansing requirement introduced by very high consumption from meter clock-overs.

SGID	MeterNo	MeterNoAlias
T0IQ00300000000005080000000	##ERA1601	ERA1601
T0IQ00330000000004000000000	140004	14004
T0IQ00330000000007330000000	4000391#WD1	400391#WD1
T0IQ00330000000010010000000	130944494	13094494
T0IQ00420000000008230000000	11004269	110044269
T0IQ00450000000003860000000	30283513	302835513
T0IQ00450000000006190000000	30200838	302200838
T0IQ00450000000011770000000	30283536	302835366
T0IQ00490000000001680000000	12999	129999
T0IQ00590000000001340000000	#00268453#WD1	#00268454#WD1
T0IQ00640000000007110009800	1000012411	100012411

(a) Meter Lookup based on SGID

MeterNo	SGID	SGIDAlt
#00000293#WD1	T0IR03160000000098900000RE	T0IR031600000000989000000
#00036839	T0IR011400000001558000030O	T0IR011400000001558000000
#00057654	T0IR050900000000207000010O	T0IR050900000000207000020
#0010095	T0IQ027600000000160000030O	T0IQ02760000000016000000RE
#00100954	T0IR008900000000483000000O	T0IR008900000000430000000
#00101600	T0IR030400000000020000000O	T0IR030400000000022000000
#00112175	T0IR011200000000201000140O	T0IR01120000000020100001RE
#00163351	T0IR050900000000538000030O	T0IR05090000000053800000RE
#00178907	T0IR050900000000176000010O	T0IR05090000000017600000RE
#00180109	T0IR00640000000097400000RE	T0IR006400000000974000010O
#00180173	T0IQ00680000000002920000000	T0IQ00680000000002960000000

(b) SGID Lookup based on Meter

Figure 6-5: Lookup Table Examples

Records with constant readings, such as zero consumption or entirely negative consumption, and records with abnormally high readings were also eliminated in the process. These included several

records of fire meters with the expected zero consumption, and some meters with unexplained negative readings throughout. This reduced the dataset to 185,000 records which were deemed adequate for the purposes of further analysis and method development.

What was clear in the processes leading up this stage was that the data management practises applied to meter reading data in Johannesburg were poor. This is similar to the experience described in previous chapters on the lack of adequate and correct meter information. Such poor data quality practices are a major deterrent in the analysis of meter reading databases, despite the opportunities that such data represent.

6.1.3 Final Dataset

Statistical analysis was done on the data using R Statistical Software (R Core Team, 2017) to categorise each record according to its average consumption quantiles and other variables such as the initial meter reading on record. The data was patched using the "*zoo package*" (Zeileis & Grothendieck, 2005), also within R. The patching was performed using linear interpolation for gaps only within the record and not the interpolation of data before or after the start or end of the original record. While cubic spline interpolation would have been better for such cyclical data, it introduced negative consumption figures which were not ideal. Data patching was necessary for the clustering analysis as the method does not allow for any gaps in the data.

An important variation to the original comparative billing analysis method was the adoption of the running annual totals as opposed to monthly data. This was to ensure that the time series property of consumption is retained at a scale that is not susceptible to monthly and seasonal variations. The running annual consumption for all records was generated using the "*zoo package*" and was used for subsequent analysis.

A visual first-order assessment was performed on the dataset by plotting the running average annual consumption for all records as shown in Figure 6-6. The plot clearly indicates unusual variations from month 60 which was about the time when the billing system was changed. Superimposing this data

against the variation in the number of meter read, as shown in Figure 3-1, confirms that there was a meaningful change in the meter reading process in general around this period.

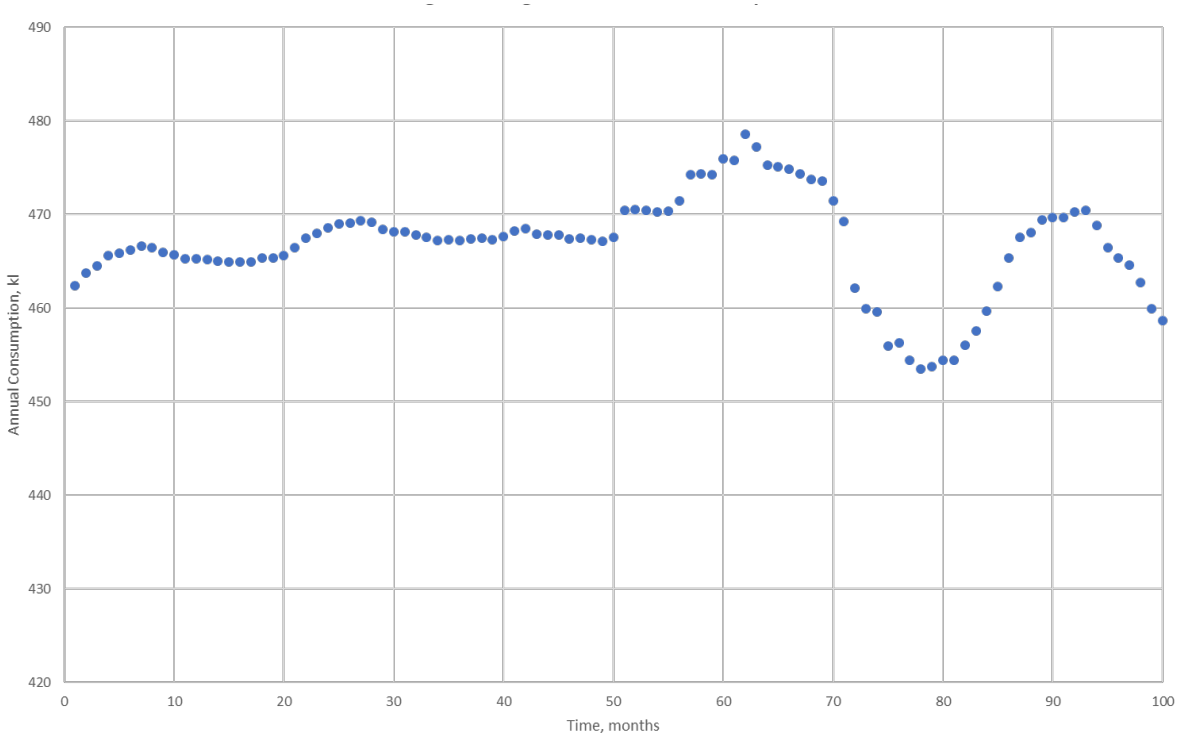


Figure 6-6: Average Running Annual Consumption

The anomalies of Figure 6-6 were not conducive for method development and therefore all post-2009 data were excluded from further analysis to eliminate these variations. The following additional criteria were introduced to the data cleansing algorithm discussed in §6.1.2;

- Individual meters had to have at least 60 months prior to the end of 2009;
- Individual meters had to have less than 10% of the annual running consumption readings falling outside of the statistical interquartile range of the record to ensure that outliers with any unusual readings are removed.

The above data cleansing criteria was reapplied, and a sample of 87,589 (about 15%) records remained. This sample size was considered adequate for method development as it is not only a considerable proportion (about 25%) of the population of active consumers in Johannesburg of conventionally metered consumers (about 375,000 consumers) but is also comparable to the number of

consumers in many small to medium sized utilities. The distribution of the first reading on record for the surviving dataset is shown in Figure 6-7.

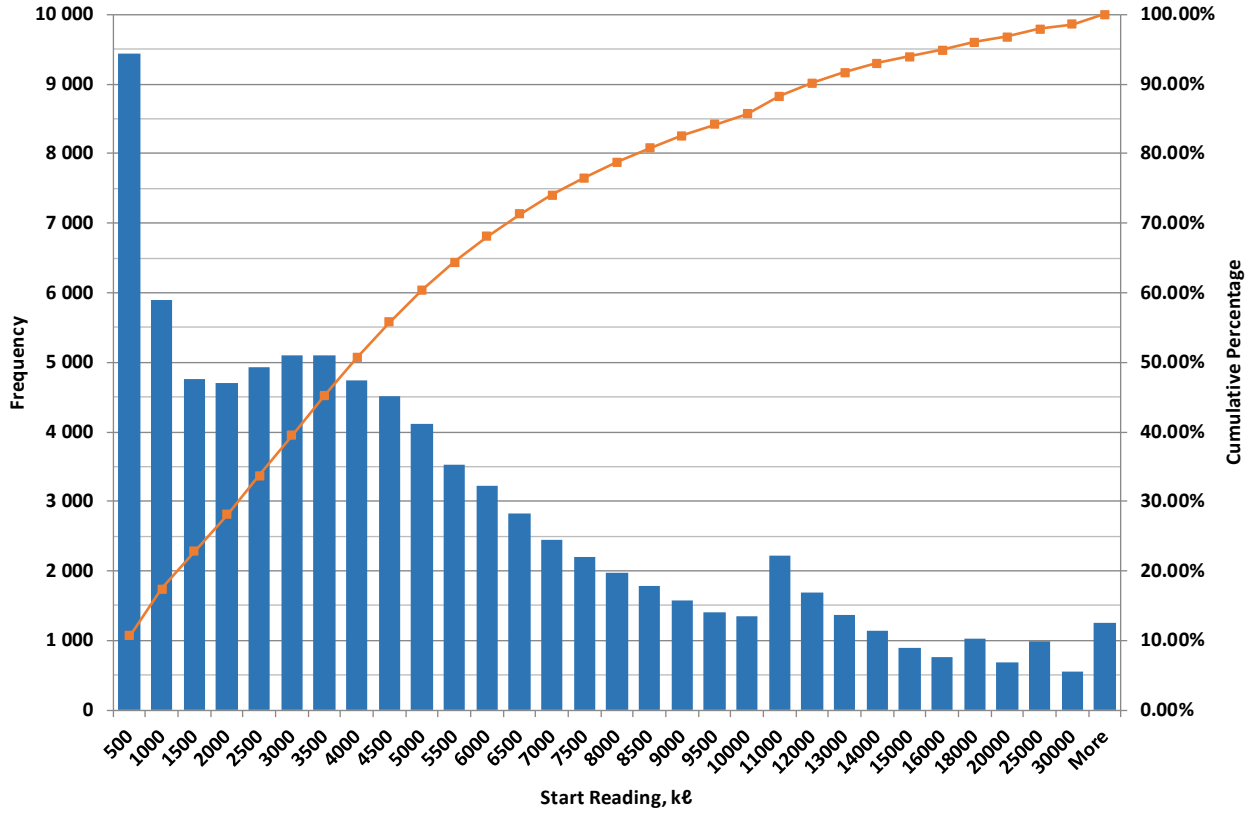


Figure 6-7: Start Reading Distribution

There is an expected higher proportion of meters with low meter readings at the start of the individual records. There was also a sizeable proportion of old meters with high initial meter readings which was expected as the starting point of the database was different from the actual installation date of the meters. This necessitated bringing all the records to a common arbitrary start date for the sub-sequent clustering analysis and factoring the initial meter reading of the individual records for the degradation analysis This highlighted another variation from the previous methodology to factor in the difference in the actual volume that has passed through the meter in the degradation analysis at a later stage. It is important to note that the clustering process is only concerned with the pattern of the time series rather than the starting dates.

The mean monthly consumption of the consumers is illustrated in Figure 6-8 with a skewed distribution towards low consumption values, as would be expected. The average consumption is 30kℓ/month. The difference in the consumption was used at a later stage in the degradation analysis.

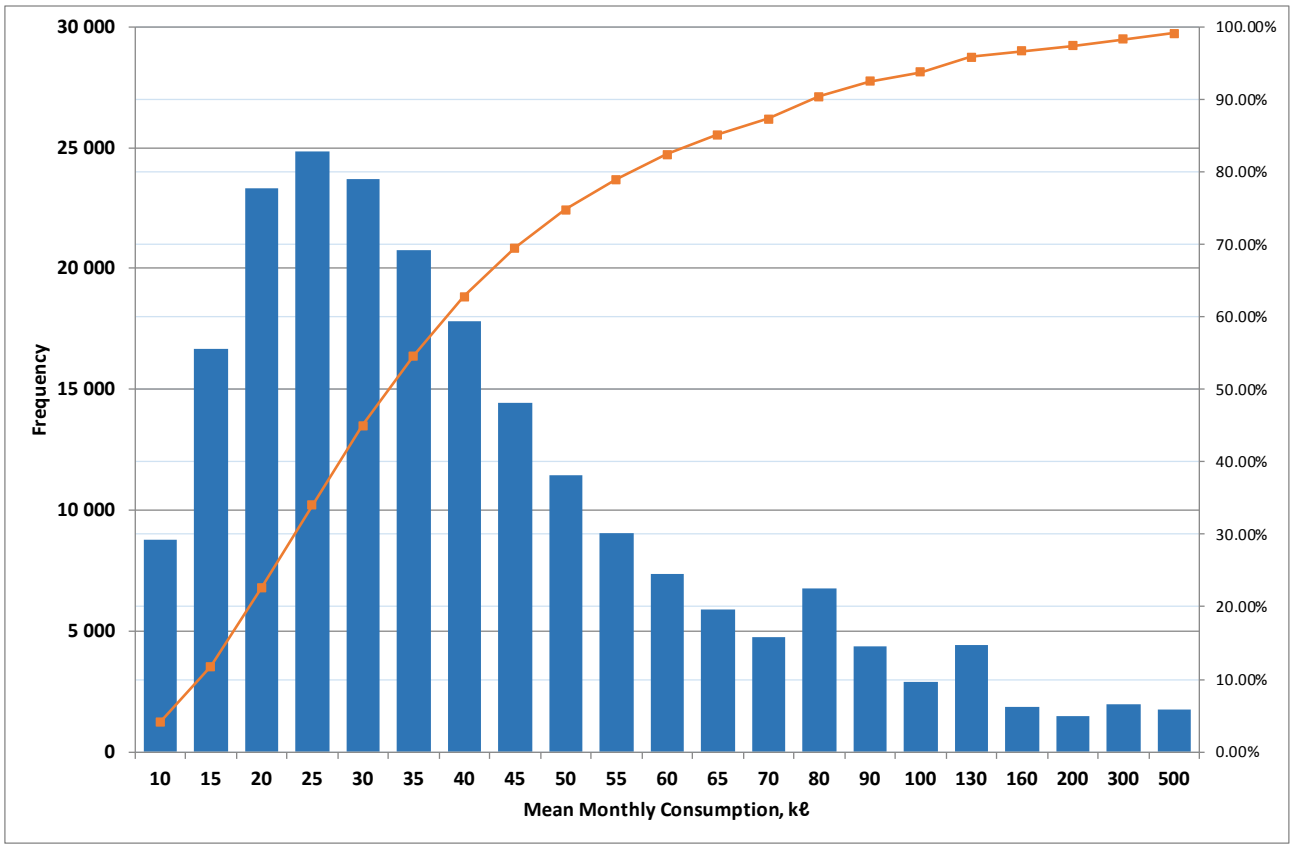


Figure 6-8: Average Monthly Consumption Distribution

The surviving dataset, together with the composition of the consumer categories from the billing data are summarised in Table 6.1. While there is an over-representation of the residential customers and an underrepresentation of the business category, the overall distribution results were deemed adequately indicative for the purposes of this study.

Table 6.1: Surviving Dataset

Consumption Category	**No. of**	**%**	**% in JW**
Residential	81,211	92.7	85.9
Multiple Residential Dwelling	2,278	2.6	2.5
Business	3,563	4.1	10.8
Utility (Public Benefit Organisation)	537	0.6	0.8
Total	***87,589***	***100***	***100***

6.2 Clustering Analysis

Aghabozorgi *et al.* (2015) provides a review of time series clustering and lists the following as important properties that necessitate the need to cluster time series datasets;

- Time series databases contain valuable information that can be obtained through pattern discovery. Clustering is a common solution performed to uncover these patterns.
- Time series databases are very large and cannot be handled well by human inspectors. As a result, time series data are represented as a set of groups of similar time series by aggregation of data in non-overlapping clusters.
- Time series clustering is the most-used approach as an exploratory technique, and as a subroutine in complex data mining algorithms, such as rule discovery, classification, and anomaly detection
- Visualization of time-series data can help users quickly understand the structure of data, clusters, anomalies, and other regularities in the datasets

These properties are all applicable to the task at hand, hence the choice of clustering as a tool of choice to uncover any patterns in the water consumption data, with a focus on the degradation of the meter accuracy with time and volume recorded.

Clustering is generally classified into five groups, with the two most common groups being (Aghabozorgi, Shirkhorshidi & Wah, 2015);

1. *Hierarchal clustering* which makes a hierarchy of clusters using either agglomerative or divisive algorithms. The advantages of this type of clustering is that it does not require the number of clusters as an initial input and it can handle series of unequal lengths. Its main weakness, however, is that it is incapable of dealing with large time series and its inability to adjust the clusters after splitting a cluster in divisive method, or after merging in agglomerative method.
2. *Partitioning clustering* makes *k* groups from *n* unlabelled objects with a requirement of the number of clusters as input, which is not an easy task. It includes algorithms which provide hard or crispy clusters, such as k-means and k-medoids, and the soft or fuzzy cluster of the Fuzzy c-

means algorithm. Partitioning algorithms require a defining prototype for each of clusters and could be a mean of all its objects or a member of the cluster. They are much faster than the hierarchal clustering and therefore find greater application with bigger datasets.

The choice of clustering algorithm is therefore an important consideration particularly when considering that time series data present a challenge due to their large size and high dimensionality. For this study specifically, there are about 88,000 records with each of at least 60 readings, which is considerable. To aid in the evaluation and choice of appropriate clustering algorithms to use, the "*dtwclust package*" of the R Statistical Software (R Core Team, 2017) was chosen as it provides a common platform on which classical and new clustering algorithms can be evaluated and compared against each other (Sard, 2015). Its convenience and the range of supported implementations were its main attractions. However, only partitional clustering algorithms were evaluated in this study due to the time and memory complexity of $O(N^2)$ of hierarchal clustering, where N is the total number of records in the dataset.

6.2.1 Preliminary Clustering Analysis

For partitional clustering algorithms, also due to the high dimensionality of the dataset, the first step of the analysis was to compile a small random sample of records to use for the evaluation of the number of clusters and different algorithms. 5,000 records were chosen as an ideal sample number after several iterations. The algorithms evaluated include (Sard, 2015);

1. *Classical Partitional Clustering* using variations of the distance measure between the Dynamic Time Warping (DTW) with the classic Euclidean distance, which is a straight line between two points. The partition around medoids (PAM), also referred to as the mean of the time series, was used as the centroid function in both cases.
2. *TADPole Clustering,* a recent improvement on the Density Peak Algorithm based on the DTW distance measure by exploiting the upper and lower bounds of DTW to prune unnecessary distance computations (Begum et al., 2015).

3. *k-Shape Clustering* is another recent algorithm based on a custom centroid function (shape extraction) and a custom distance measure (shape-based distance, SBD). The function is stochastic in nature and requires *z*-normalization of the input data. The centroid series is randomly chosen from the data unless the time-series have the same length, in which case no centroid is required prior.
4. *Fuzzy C-means Clustering* uses the Euclidean distance within each of the *k* clusters to find the degree of belongingness for each record with the sum of such degrees being constrained to equal to across all clusters.

Each of the clustering algorithms was run for the sample dataset with *k=2:10* (inclusive), where *k* is the number of clusters. For uniformity, the data was pre-processed with z-normalisation which was a requirement for the *k*-Shape clustering algorithm.

A non-trivial aspect of partitional clustering is the determination of the number of clusters, *k*, and evaluating the clustering algorithm performance. To this end, the recommended cluster validity indices (CVI) identified in Arbelaitz et al. (2013) were used to evaluate the output of the algorithms, with the number of clusters determined through a majority vote of the different CVIs used (Sard, 2015). Arbelaitz et al. (2013) provides a comparison of 30 cluster validity indices on an extensive set of configurations with a statistical significance analysis of the results showing that the Silhouette, Davies–Bouldin, Calinski–Harabasz, generalized Dunn, COP and SDbw indexes behave better than the other indices. Accordingly, the indices used in this study are the Silhouette index (Sil), the generalised Dunn index (D), COP index, Davies-Bouldin index (DB), Calinski-Harabasz index (CH) together with the Modified Davies-Bouldin index (DBstar) (Kim & Ramakrishna, 2005) and the Score Function (SF) (Saitta, Raphael & Smith, 2007). Additionally, the agreement between the different algorithms in terms of the ideal number of clusters was also compared. The only exception to the evaluation of the CVIs was the Fuzzy Clustering algorithm as the appropriate CVIs were not yet implemented within the "*dtwclust package*". A sample of the output for the evaluation of the *k*-shape algorithm using the cluster validity indexes is presented in Table 6.2.

Table 6.2: Cluster Validity Indices Output Example

CVI	No. of Clusters, *k*								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sil	0.2205	***0.2351***	0.2287	0.2209	0.2031	0.1978	0.1784	0.1776	0.1700
SF	***0.5148***	0.4539	0.3971	0.3552	0.3199	0.2867	0.2647	0.2331	0.2160
CH	**6621.646**	3488.182	2287.017	1750.072	1434.331	1344.347	2002.952	1043.618	948.882
DB	***0.9993***	1.0899	1.0591	1.1129	1.1502	1.3475	1.4282	1.4674	1.5288
DBstar	***0.9993***	1.0899	1.1042	1.2198	1.4046	1.6518	1.8077	1.8391	2.0400
D	0.0042	0.0044	0.0079	0.0084	0.0109	***0.0135***	***0.0135***	0.0098	0.0114
COP	0.3646	0.2907	0.2613	0.2399	0.2289	0.2173	0.2100	0.2048	***0.1975***

The objective is to maximize CVIs 1 to 3 and 6 while minimising the rest. Through the majority vote, two clusters were the “purest” number of clusters from all the evaluated algorithms. As an example, in Table 6.2 the required maxima or minima are bolded and in italics, with a clear indication that *k* = *2* has the most common number of clusters, hence a winner of the majority vote. The two clusters in all cases consisted of increasing consumption and decreasing consumption clusters.

For the selection of the clustering algorithms to use further, a consideration of the time and resource requirements and the ease of implementation of each algorithm was considered. Methods with DTW were considerably slower as was expected due to the need to compute the difference between individual time series. The TADPole algorithm is deterministic in some way, depending on the value of a cut-off distance that must also be specified by the user. This is not very easy to use and is a deterrent to users. Through this elimination process, the *k*-Shape and Fuzzy clustering algorithms were eventually selected for implementation due to their relative speed and ease of implementation.

6.2.2 Final Clustering Analysis

The results of the preliminary clustering analysis of two clusters, out of up to ten clusters evaluated, were carried forward to the final clustering analysis. These two clusters consist of one with records of decreasing annual consumption over time while the other cluster had increasing consumption. It must be noted that even with up to ten clusters, the two main characteristics remained a combination of clusters with increasing and decreasing consumption, albeit at different amplitudes. This shows

that there is a possibility of further disaggregation of the analysis to find other patterns if the data is adequately labelled, which unfortunately was not the case for this study.

The k-Shape algorithm was thereafter implemented for the entire dataset with the two selected clusters. The randomly selected centroids for the analysis are shown in Figure 6-9 with the length and the shape of the centroids corresponding to the selected time-series. Despite some unusual variations at the beginning and the end of the time series, there is clearly a generally upward trend for the first cluster and a downward trend for the second one. Figure 6-9 also clearly illustrates the weakness of the algorithm in that the selected random series may not always be the best presentation of the shape and length of the overall cluster.

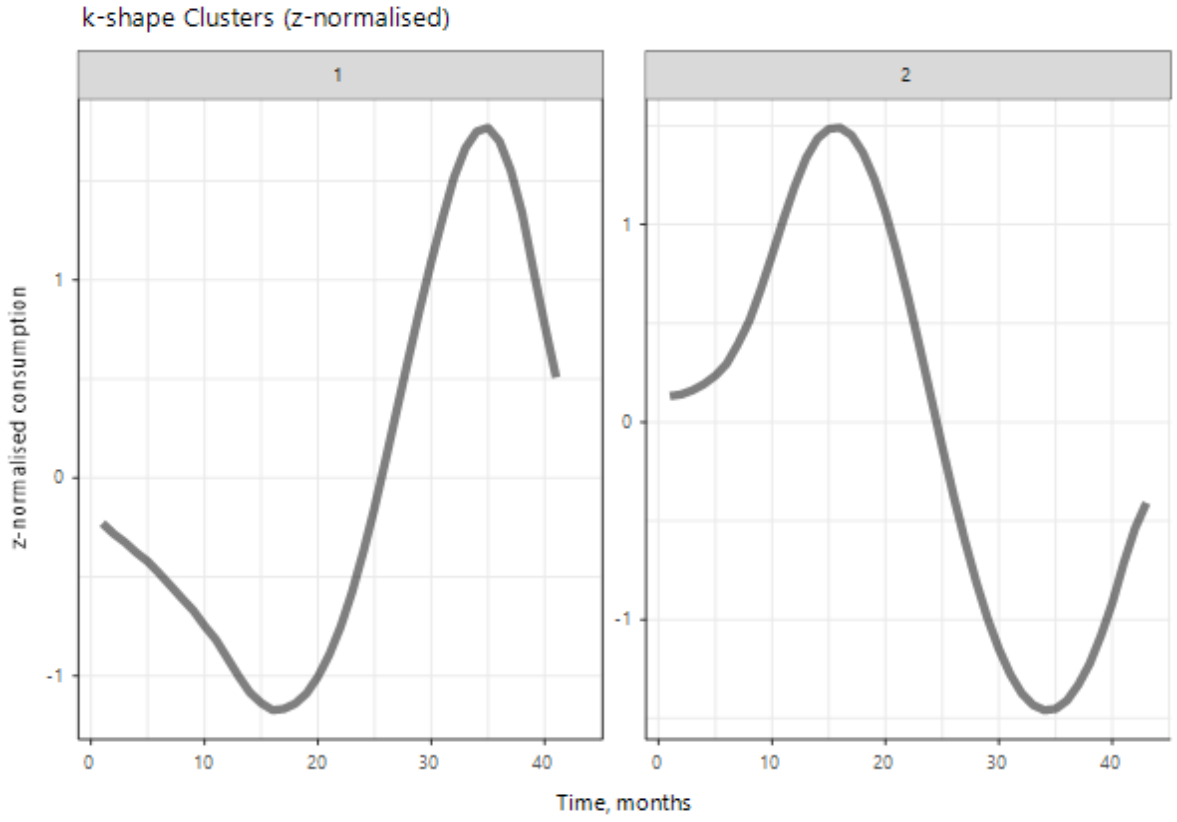


Figure 6-9: k-Shape Clusters

Unlike the k-Shape algorithm, the Fuzzy c-means algorithm computes and uses the mean of the cluster as a representation of the individual cluster member. This therefore means the resultant centroid will in most cases be a better representation and is particularly suited for the purposes of this

research. The fuzzy clustering algorithm was therefore implemented, but with an additional requirement of re-interpolation of the existing data to obtain time series of equal length as required by the algorithm. The output of the algorithm is a membership matrix which defines the degree of membership of each record to each of the two clusters. These non-crisp partitions of the algorithm can also be made crisp by taking the maxima of each cluster as shown in Figure 6-10 for the selected crisp centroids of each cluster.

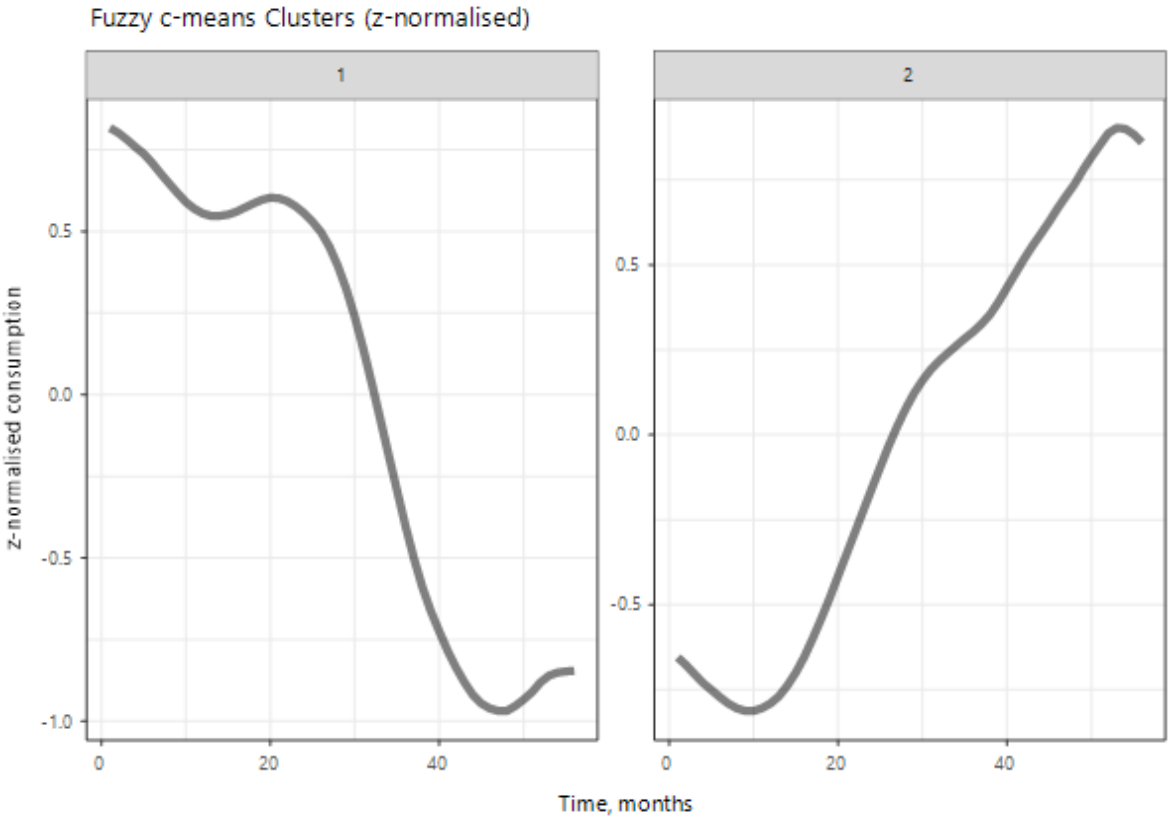


Figure 6-10: Fuzzy Clusters

Similar to the k-Shape algorithm, there are clear trends in both clusters, one of increasing consumption and the other decreasing. There is however a better presentation of the clusters by the centroids in this case due to the use of an average synthetic time-series. A further examination of the data shows that the end of the clusters in both implementations correspond to fewer records of consumption data hence the variability in the shape and values of the normalised consumption. There is also a clear indication that the beginning of either cluster does not always follow the general pattern of the cluster, which is instructive in instances where users use data at only specific periods of the time series as was done with the current implementations of alternative methods.

Due to the stochastic nature of the algorithms there were expected differences in the composition of clusters, but with comparable results, depending on the reference centroid for k-Shape. The fuzzy clustering implementation was therefore the preferred algorithm due to the use of average centroid which better represents the data and the fact that in typical water demand problems, it is also possible for a consumer to belong to a few clusters at various points in time.

6.3 Accuracy Degradation Analysis

6.3.1 Exploration of the Dataset

Using the final clustering results, the final dataset (without z-normalisation) was accordingly classified into the two clusters for accuracy degradation analysis. For each record, the information pertaining to the average consumption quantile, the consumer category and a categorical classification of whether a meter is new or old were incorporated into the dataset. The classification of a new meter was on the assumption that the meters had readings of less 1,000kℓ at the beginning of the record. Based on the average domestic consumption of 30kℓ/month, this assumption considers meters of up to 3 years of usage as new for an average consumer, although age is not the primary focus. Old meters were taken to be those with 10,000kℓ. For other consumer segments new meters were assumed to have an initial reading of up to twice the average annual consumption of that category while old meters were those with an initial reading of at least five times the average annual consumption. The proxies for whether a meter was old were varied later in the analysis to evaluate their sensitivity on the result but there was no meaningful impact other than increased variability due to fewer records as the cut-off year increased.

The variation of the average running annual consumption for each category of the final dataset is shown in Figure 6-11. As expected, non-residential consumption categories have a much higher annual consumption, up to an order of magnitude than the residential sector. The residential consumers however, as the clear majority of the consumer base, significantly lower the overall average consumption to a value closer to the domestic average. The residential consumption category has

therefore the greatest impact on the consumption in Johannesburg and should be managed accordingly and not neglected in favour of individual bulk consumers. In addition, the importance of normalising (z-normalisation in this case) the data removes the differences in amplitude between the various consumption categories to reveal the shape similarity. It can also be noted from Figure 6-11 that the average consumption for any consumer category does not follow any trend – this also demonstrates that the averaging effect of looking at consumers without unpacking the underlying consumption patterns can lead to unfavourable results.

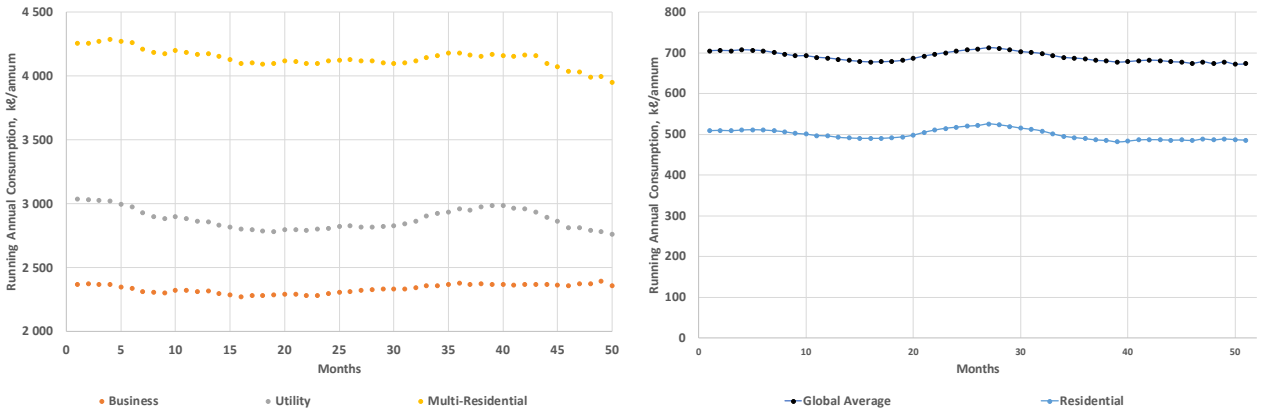


Figure 6-11: Running Average Annual Consumption per Category

The running annual consumption for each of the clusters is shown in Figure 6-12 confirming the output of the cluster analysis of declining and increasing consumption clusters. For the residential category, Figure 6-13 shows the increasing consumption cluster for “new meters” with the different consumption quantiles, all with R^2 values above 0.9 for a linear trend line except the 100% quantile (>95% quantile) with an R^2 of 0.81. The graph also shows the downside of calculating the degradation rate at only a point (i.e. for only a year), particularly if the results are not validated, as the annual consumption is not static. This is further amplified when observing the declining consumption cluster for “new meters” of the same residential category in Figure 6-14.

For the first year or so, particularly for high consumption quantiles, the consumption does not actually decrease but increases slightly before decreasing. This further emphasises the pitfall of taking different points from different quantiles as it does not give an accurate representation of what is transpiring with the measurement of consumption and any associated trends. The incorporation of the

different consumption quantiles into the analysis also allows the combination of both age and meter reading into the assessment of the meter error. This is achieved through the evaluating of the meter error degradation for consumers with specific consumption patterns (provided by the volumetric reading) over the life of the meter (age). This has also not been clear in prior studies where the two factors were analysed separately. There is however limited variation from the different consumption quantiles, except for the higher consumption quantiles whose annual consumption tended to have greater variability.

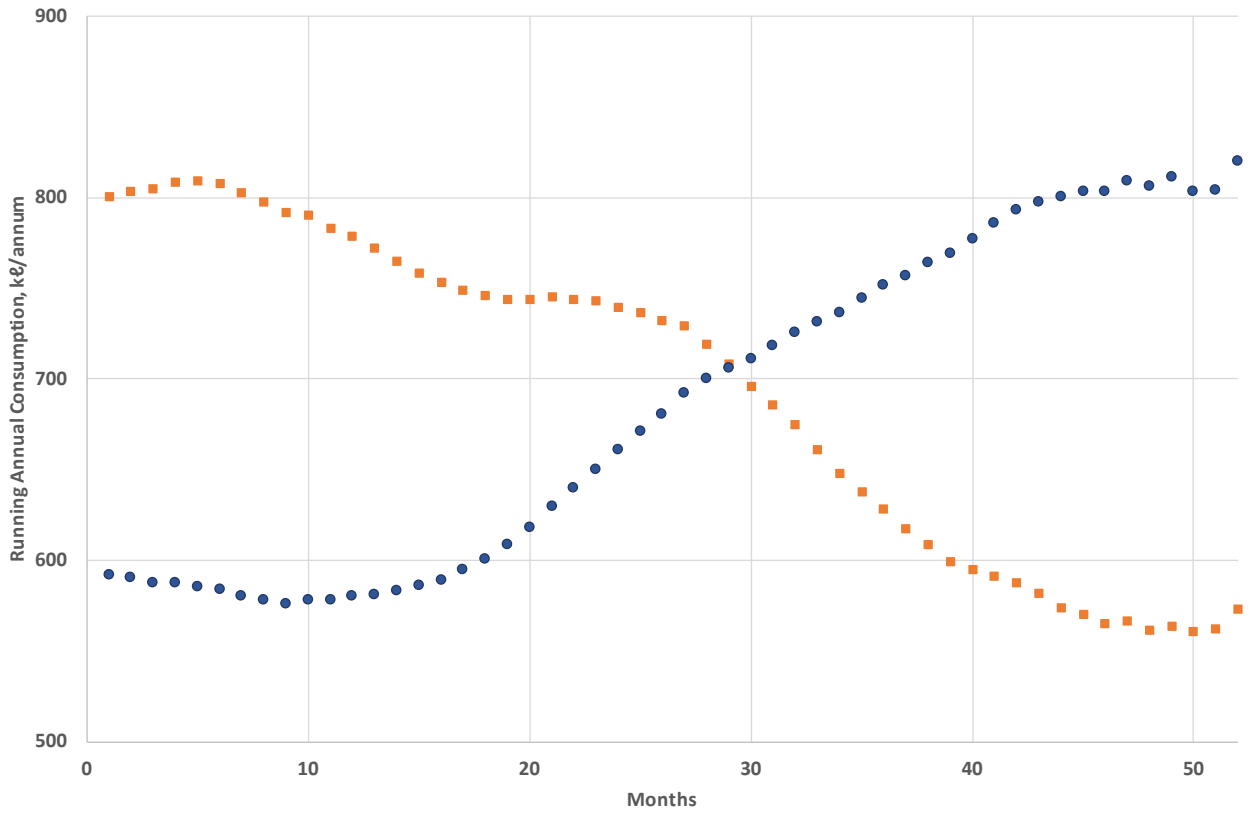


Figure 6-12: Cluster Average Consumption

6.3.2 Error Degradation

For each combination of the derived clusters, the average consumption quantile, the consumer category and a categorical age of a meter (as either new or old), the variation of the annual running consumption was evaluated using Equation 2.7 as suggested by Mbabazi et al. (2015).

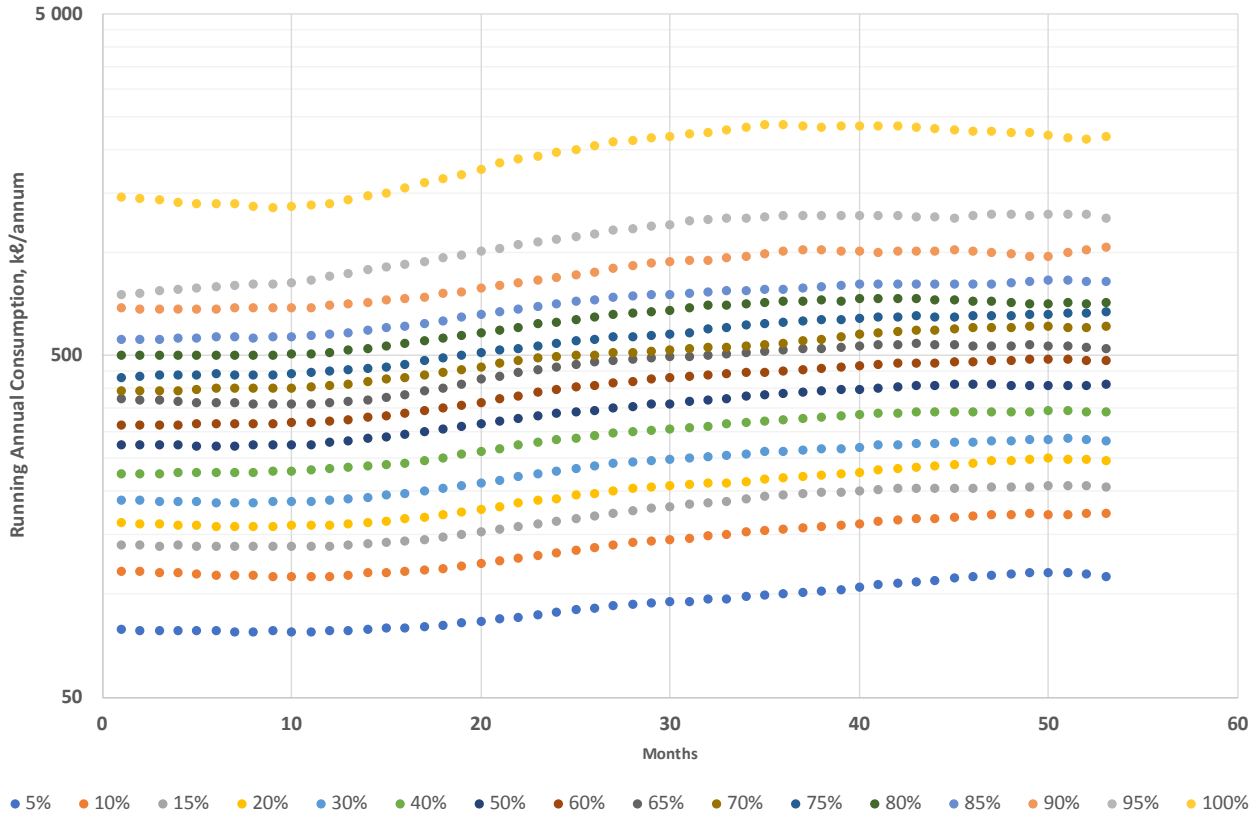


Figure 6-13: Increasing Consumption Cluster

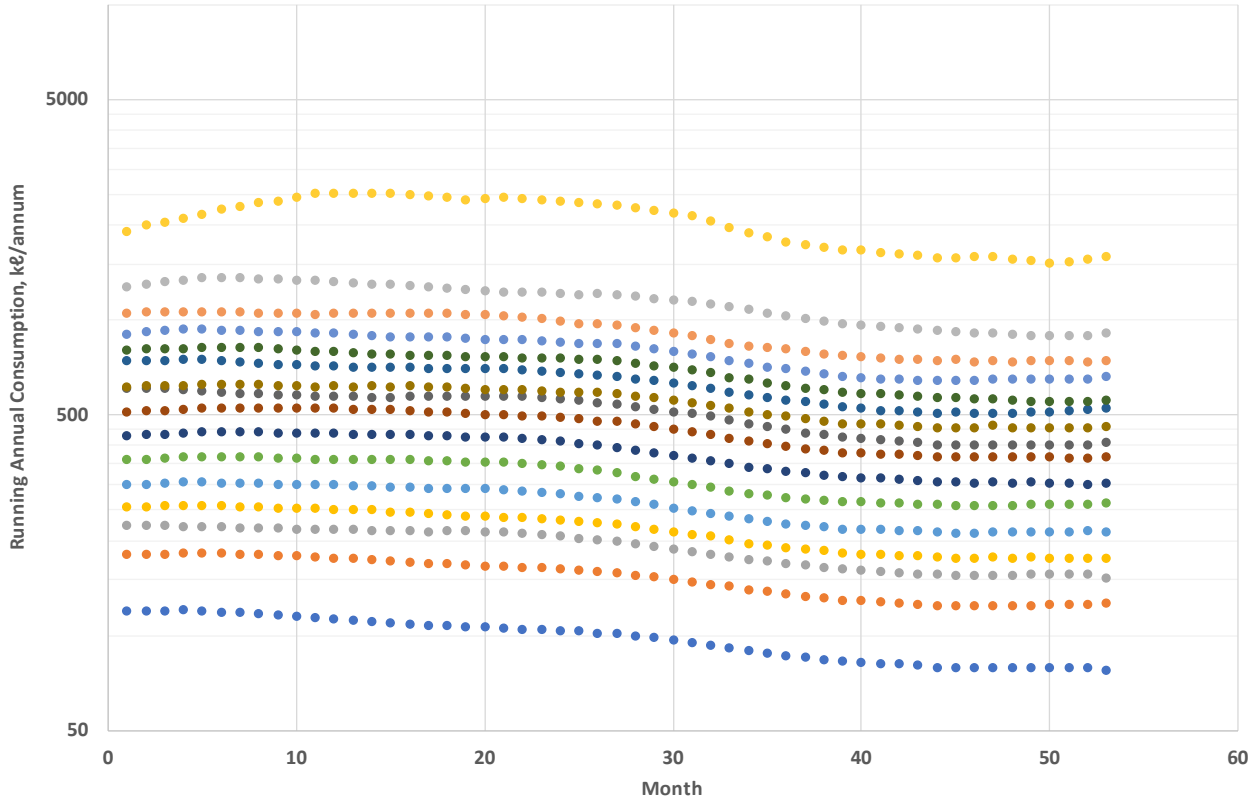


Figure 6-14: Decreasing Consumption Cluster

Other than the difference in approach thus far, another main variation between in this research and that of Mbabazi et al. (2015) is the use of the annual running total every month as opposed to the annual consumption for meters of various calendar years. The method adopted in this research maximises the use of each individual time series without taking only a small component of data for analysis. In addition, by using the complete time-series in the analysis the new methodology avoided the pitfall of evaluating the degradation at a point.

The degradation rate was thereafter calculated for each consumption quantile in each cluster using Equation 2.8.The degradation rates for the new meters and that for old meters were thereafter compared to evaluate the progression of the degradation, or the lack thereof. This was implemented within MS Excel using the LINEST function to obtain the linear regression coefficients. Table 6.3 shows the tabulation of each of constants from the regression curves for the quantiles of Figure 6-13 and Figure 6-14. The constants β_0 and β_1 are from Equation 2.7 and d is the ratio of the two from Equation 2.8. The average values of d at the bottom of Table 6.3 are of all quantiles with an $R^2 \geq 0.7$, demonstrating high correlation. This obviously excluded a few and insignificant quantiles of consumption categories that had few data points because of the low reliability of their results.

Table 6.3: Calculation of Degradation Rate

Quantile	New Meters				Old Meters			
	R^2	β_0	β_1	d	R^2	β_0	β_1	d
5%	0.9752	-0.9437	123.797	-0.8%	0.6882	-1.1603	151.531	-0.8%
10%	0.9569	-1.3604	190.158	-0.7%	0.6658	-1.8555	212.255	-0.9%
15%	0.9366	-1.6545	235.660	-0.7%	0.8233	-1.7895	244.714	-0.7%
20%	0.9514	-1.9959	271.637	-0.7%	0.9501	-2.9190	309.417	-0.9%
30%	0.9259	-2.3213	323.557	-0.7%	0.9519	-3.2587	361.849	-0.9%
40%	0.9078	-2.7283	390.685	-0.7%	0.9265	-3.2134	417.554	-0.8%
50%	0.9150	-3.3919	470.816	-0.7%	0.9492	-4.1587	506.749	-0.8%
60%	0.9152	-3.9614	561.031	-0.7%	0.9319	-4.6307	589.819	-0.8%
65%	0.9224	-4.7381	638.028	-0.7%	0.9294	-4.6957	648.779	-0.7%
70%	0.9047	-4.2033	664.239	-0.6%	0.9202	-5.1652	702.270	-0.7%
75%	0.9319	-5.5912	783.391	-0.7%	0.9066	-5.5026	774.824	-0.7%
80%	0.9400	-6.1992	865.453	-0.7%	0.9083	-6.2162	867.871	-0.7%

Quantile	New Meters				Old Meters			
	R^2	β_0	β_1	d	R^2	β_0	β_1	d
85%	0.9253	-6.8518	980.371	-0.7%	0.8905	-6.2931	953.040	-0.7%
90%	0.9094	-8.0847	1138.047	-0.7%	0.9014	-7.0562	1109.712	-0.6%
95%	0.9386	-10.3938	1426.271	-0.7%	0.9069	-9.9021	1453.570	-0.7%
100%	0.6314	-18.6079	2564.458	-0.7%	0.8982	-16.394	2487.975	-0.7%
			Average	***-0.7%***				***-0.7%***

The consolidated results of the degradation rates for all consumption categories are tabulated in Table 6.4, with an inclusion of the average meter which is essentially all meters of that category irrespective of the meter reading. The decline columns comprise the difference between the degradation rates of new meter and old meters within each consumer category.

Table 6.4: Degradation Rates

Consumption Category	Declining Consumption Cluster				Increasing Consumption Cluster			
	New	***Ave***	***Old***	***Decline***	***New***	***Ave***	***Old***	***Decline***
Residential	-0.7%	-0.7%	-0.7%	0.0%	1.4%	1.1%	0.8%	0.5%
Multiple Residential Dwelling	-0.7%	-0.7%	-1.1%	0.3%	1.4%	1.3%	0.6%	0.8%
Business	-0.9%	-0.9%	-1.0%	0.1%	1.9%	1.7%	1.4%	0.5%
Public Benefit Organisation	-0.9%	-0.9%	-0.9%	0.1%	1.8%	1.8%	1.2%	0.7%
Combined	***-0.7%***	***-0.7%***	***-0.8%***	***0.1%***	***1.5%***	***1.1%***	***0.8%***	***0.7%***

The properties with declining consumption show a clear trend of decrease in consumption ranging from 0.7% to 1.1% per annum, with the residential sector (small meters) having a consistent rate for both old and new meters. The categories with higher consumption have a higher degradation rate, which is expected due to the higher wear and tear of mechanical meters from the high-volume throughput. Similarly, the multi-residential sector also shows the largest differential between new and old meters and this again is very logical as these are meters that would typically be the most used at varying flowrates due to the many permutations of users using water at various times and in varying volumes.

What had previously been missing in such analysis conducted in previous studies is how to handle increasing consumption, where it is either not mentioned or considered as an anomaly by some

authors. From Table 6.4, it can be observed that for the consumers with increasing consumption there is a marked difference between the degradation rate of new and old meters of 0.5% to 0.8%. A closer scrutiny shows that the decline in the degradation rate for the increasing consumption cluster is similar, albeit lower, than the overall decline experienced by meters in the declining consumption cluster. Such a result could be indicative of what should be evaluated for consumers within the increasing consumption cluster, which is the decline in the degradation rate from new to old meters. The lower decline rates can also be explained by the expected combined effect of increasing losses and decreasing accuracy that mask the actual loss in accuracy. It is for this reason that the use of the degradation rates from the increasing clusters should be used cautiously. Considering the high extent of on-site leakage within the study area, as found by Lugoma et al. (2012) and corroborated by this study, leaks tend to get bigger with time and would make the increasing consumption cluster very plausible. The use of the average rate in this case would be misleading. The decline in the degradation rate is also the highest for multi-residential consumers as was the case in the decreasing consumption cluster for likely similar reasons.

In overall, the degradation rate of meter accuracy with time for both clusters is estimated at 0.7% per year. This degradation rate is much lower than that of 1.45% to 6.67% in Mbabazi et al. (2015) and 2.1% per year in Arregui et al. (2006b), which both used comparative billing analysis. Both studies however concede that such rates are much higher than those found using the weighted meter accuracy method with ranges from 0.1% to 0.7% in Arregui et al. (2006b) and 0.1% to 0.6% in Noss *et al.* (1987). The results of the developed method are therefore a significant improvement from similar studies producing degradation rates comparable with those in other studies using the weighted meter accuracy methodology.

6.4 Method Validation

For validation purposes, the weighted meter accuracy results produced in §5.1.1, with an average meter age 11.5 years, were used as the basis of comparing the results of the developed methodology. It is however important to note the fundamental differences between the two methods in terms of what they account for prior to the comparison;

- The estimates from the weighted meter accuracy method, which are field-based, inherently measure meter under-registration through laboratory testing of meter accuracy. The new data-based methodology, on the other hand, cannot account for such meter under registration. This is because the billing records that underpin this approach do not include such flows as by definition they are not recorded on the meter register but can only be found through meter testing.
- The weighted accuracy methodology focused on only small meters of up to 20mm in size with the results excluding all bulk meters, which coincidentally were shown to be oversized. Additionally, the non-residential consumers included in the weighted accuracy method are essentially high-volume consumers with small meters. The new methodology includes all meters, irrespective of size, in the analysis.

For each of the respective clusters of the new method, the error degradation rate was applied over 11.5years, the average age of the meters in Johannesburg water, assuming 100% accuracy of the meters from the onset. The accuracy of each consumer category was determined using the average degradation rate for the declining consumption cluster and the decline in the degradation rate for the increasing consumption cluster. The validation results are reproduced in Table 6.5, together with the best-case scenario of the weighted accuracy method (field estimates).

Table 6.5: Estimated Apparent Water Losses

Consumption Category	*Field Estimates**	*Declining Consumption*	*Increasing Consumption*
Residential	11.2	8.1%	6.2%
Multiple Residential Dwelling	6.5	8.2%	9.8%
Business	8.3	10.6%	6.0%
Public Benefit Organisation	6.4	10.2%	7.5%
Average	***9.4***	***8.2%***	***7.9%***

* Best case scenario

On average, the differences between the declining and increasing consumption clusters are very small and they are comparable with field estimates. This is a meaningful improvement from the results in §5.3 of the unrefined methodology that gave a much higher estimate of 14%, closer to the worst-case scenario which is considered an overestimate. This is an indication of the improved accuracy of the new methodology.

For the residential category, which presents the best basis for comparison due to the common use of small meters, both the declining and increasing consumption clusters understate the losses as determined through the field-based approach by 3.1% and 5% respectively. This difference, particularly for the declining consumption cluster, can be attributed to meter under-registration that is only accounted for in the laboratory-based approach. Arregui, Balaguer & Soriano (2017) have shown that the expected initial weighted error of ISO velocity meters, under the most favourable working conditions ranges from -0.71% to -3.87%, while oscillating pistol meters range from 0% to -1%. Arregui et al. (2018) found that the weighted meter error for new meters was around −4%. The difference of -3.1% for the decreasing consumption cluster is within this range and can therefore be attributed to the initial meter error, validating the results for the residential category. The larger differences for the increasing consumption cluster is potentially the complex combined effects of meter under-registration and increasing level of leakages with time and should therefore not be used other than as a sanity check for similarities.

As expected and due to the inclusion of bulk meters for the error estimates from the developed method, all other categories have higher estimates of AL than from the field-based method, particularly for the declining consumption cluster. This agrees with field observations of meter sizing challenges within the Johannesburg Water metering fleet. However, due to the lack of adequate and accurate information, these results should only be taken as indicative.

While the developed methodology has been successfully validated for the estimation of apparent losses due to metering errors for residential consumers, it can be further enhanced for related activities. For example, in addition to the clustering algorithms, it is possible to train classification algorithms to recognise the different meter and consumer categories within the same data. However,

classification is a supervised learning algorithm that requires labelled data, which was not available for this study. An additional potential application is anomaly detection for the automatic identification of illegal connections. This also requires labelled data and the cooperation of the utility in the confirmation of the identified anomalies. In the case of the study area, the reported illegal connections of §5.1.3 showed very few confirmed cases of illegal connections making it impossible to implement anomaly detection.

6.5 Generalised Equation for AL Estimation

In keeping with other studies which provide a generalised form for the estimation of apparent losses, such as Mutikanga *et al.* (2011) and Seago *et al.* (2004), a generalised equation is proposed. For utilities with similar meters and consumption characteristics as Johannesburg, the results of the new methodology are typical degradation rates of 0.7% per annum. Factoring in meter under-registration, in cases where no data is available for the weighted accuracy of new meter, the estimates of Arregui *et al.* (2018) and the initial error estimates in §5.1.2 can be adopted with allowances of 4% - 5% for the initial meter error. It has also been shown in this study, in Table 5.5, that upsizing a 15mm meter to 20mm can also increase the metering error by up to 50% and therefore utilities with a legacy of incorrect meter sizing should account for the additional under registration margin of error. In this study the under registration due to under sizing was estimated in the region of 5% of the metered consumption. The following relationship for apparent losses due to metering errors is proposed;

$$AL_{ME} = 0.007a + 0.04b + 0.05c = \frac{0.7a + 4b + 5c}{100}$$

Equation 6.1: Apparent Loss Estimate

Where ***a*** is the average meter age, ***b*** is the estimated probability of meter under-registration and ***c*** is the probability of meter oversizing. The recommended default value of 4*b* has been used for the second term in Equation 6.1 where no data is available but where there is available data it should be replaced with its actual value. The third term is a subjective estimate of how pervasive meter oversizing is in the utility with the proposed maximum of 5% being found as the worst-case scenario

for Johannesburg. No allowance has been made for unauthorised connections, data acquisition errors, and additional errors such as the influence of storage tanks as these have not been evaluated in this study.

6.6 Summary

This chapter introduced the improved comparative billing methodology which has been developed using a combination of data preparation techniques, clustering analysis and classical regression analysis. The methodology uses commonly available monthly billing data to find the meter accuracy degradation rate and the estimated AL due to metering inaccuracies. Central to this novel methodology, is the adequate preparation of the data, including the use of the annual running total and thereafter the disaggregation of underlying consumption patterns into clusters that are analysed separately to minimise the averaging effect. Clustering is implemented in two parts; the initial part finds the appropriate number of clusters using cluster validity indices on a random sample of the entire dataset. The final clustering component implements the appropriate number of clusters on the whole dataset to determine the members of each cluster. Finally, classical regression analysis is performed on the dataset with cluster information to obtain degradation rate and ultimately the estimate for AL.

The method has been applied and validated for Johannesburg Water producing comparable results to the traditional field-based method at a fraction of the cost and effort. Apparent losses due to metering error for the residential category were estimated to be 8.1% of billed consumption compared to 11.2% found using the field-based method, using an error degradation rate of 0.7% per annum. The difference in the values is attributed to meter under-registration which is not accounted for with the new developed method and can be higher than 4% for new meters. About 15% of the available data was used in the development of the method, which was a significant improvement from previous attempts of related work. Average results are even closer with only a difference of 1.2% which is also better than the previous estimates of the unrefined method in the previous chapter. This is the first known comprehensive alternative method for assessing apparent water losses that has been developed and validated. With better data quality, it is possible to further enhance the method for

detailed classification purposes to capture consumer segment identification, different metering technologies, and other categories of interest, which was not possible in this study.

A general equation has also been developed to provide a simple estimation of apparent losses due to metering errors in utilities with similar conditions and environmental factors as Johannesburg Water. The equation includes the estimation of the under-registration errors and those due to oversizing that is endemic in Johannesburg Water.

Data-driven discovery processes, as developed and applied in this research, therefore do offer viable alternatives that can be used in conjunction with the traditional apparent loss assessment processes by utilities for better management of water losses. The developed method overcomes the difficultly and costs involved in field assessments and laboratory work as it is easy and cheap to implement without the unintended negative consequences that accompany the traditional approaches.

Chapter 7 Decision Enhancement for the Management of Apparent Losses

Parts of this chapter are based on presentations by the author to the 7th Regional African Water Leakage Summit 2017 in Johannesburg and Cape Town, South Africa in August 2017 entitled "*Meter Management Imperatives for the Management of Apparent Losses*" and the publication *"Decision Enhancement for the Management of Apparent Losses"* referenced as (Ncube & Taigbenu, 2018b).

This research has thus far delved in detail into the methods of assessment of apparent water losses, including their application within the City of Johannesburg. An improved method, which is faster and reliable with comparable results to the trusted but costly and laborious traditional weighted meter accuracy method has also been developed as an alternative to fill the gap of viable, cheaper and reliable AL assessment methods. What has been missing thus far is a discussion on the user-friendly tools that empower utility users within which to apply the various activities and methods of assessing and managing AL. This is because such tools are very limited outside of commercial applications, which are outside the reach of resource-constrained municipalities. The lack of appreciation and need to estimate apparent losses has made such tools of low priority. This chapter introduces a suite of tools developed as part of this research to assist municipalities and utilities in understanding the requirements for assessing and managing apparent losses and applying them within a common and user-friendly environment. Key to this is the application of an all-encompassing meter management framework that is also introduced and discussed in this chapter.

7.1 Meter Management Framework

Various authors offer different but related perspectives of what makes for good meter management within a water utility. As discussed in §2.3, Arregui, Cabrera Jr. & Cobacho (2006) provides a comprehensive overview of the requirements for integrated management with the key components summarised in Figure 2-7, with the text clearly indicating additional underlying components that are critical such as meter management databases. This is however not immediately apparent to users who might be keen on starting the apparent loss management journey. Van Zyl (2011) provides an alternative view that clearly puts meter management databases at the core as shown in Figure 7-1, while leaving out components such as consumer auditing for the determination of consumption profiles. Both frameworks clearly capture the cyclic and integrated nature of the management of meters. This has the unintended side-effect of communicating that everything needs to be done in the manner suggested without breaking the chain, which cannot be further from the truth.

The context of many utilities and municipalities is that they do implement several activities in both Figure 2-7 and Figure 7-1, albeit for different reasons other than to quantify apparent water losses and may not be as recurrent as they ought to. For example, the analysis of §4.1, that draws on the historical meter testing done in Johannesburg Water, is often carried out on a reactive basis. The output of this analysis showed that the data are important for various purposes that include the management of apparent losses. Over and above this, some municipalities also perform other activities that include various forms of meter audits, consumer audits and meter logging to solve specific service problems. It then becomes important how such routine activities can be leveraged to benefit the assessment and measurement of apparent water losses and promote a proactive approach towards meter management and the broader apparent water loss management imperatives

The recognition that many utilities do in fact carry out a variety of metering related activities without the express objective of quantifying and/or managing apparent water losses prompted an investigation into what form of framework can be put in place to leverage such routine activities of utilities. Central to such a framework is adequate recognition of the various individual activities while also

emphasizing the importance of integration to achieve better meter management and ultimate the management of apparent water losses.

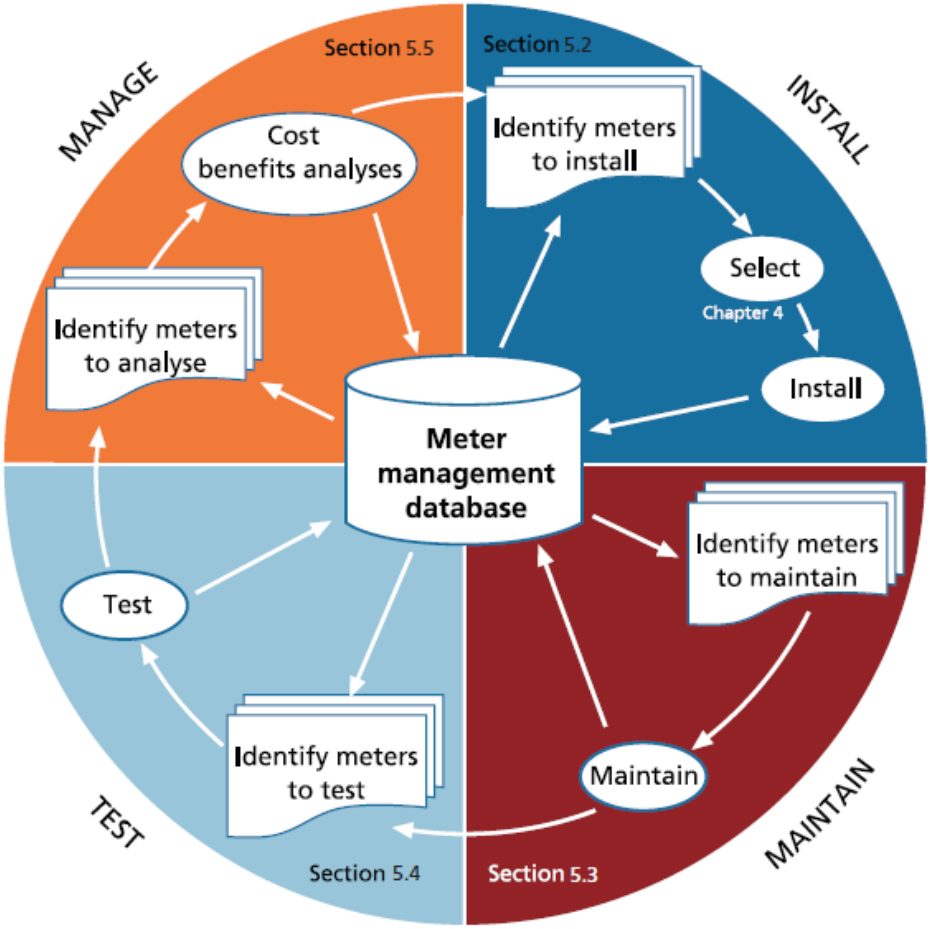


Figure 7-1: The Meter Management Process

Source: van Zyl (2011)

A review of the common activities, including those mentioned in the frameworks of Arregui, Cabrera Jr. & Cobacho (2006) and van Zyl (2011), showed that most of the activities comprised four aspects that relate to systems, audits, analysis and actual metering work in the field. This has been summarised into the Systems, Audits, Analysis and Metering (SAAM) framework illustrated in Figure 7-2. The four aspects of Figure 7-2 correspond to the management (systems, analysis and metering), fieldwork (audits and metering) and laboratory work (audits and metering) components of Figure 2-7. The realignment is also meant to match the practical structuring and segregation of duties of some water utilities. This will enable them to implement the identified activities individually but within the context and emphasis of integrated systems, which have the meter management databases of Fi-

gure 7-1 as an essential component. Within SAAM framework, the respective activities can be implemented separately but they have greater impact within the ambit of interdependence, vertical and horizontal integration with other activities.

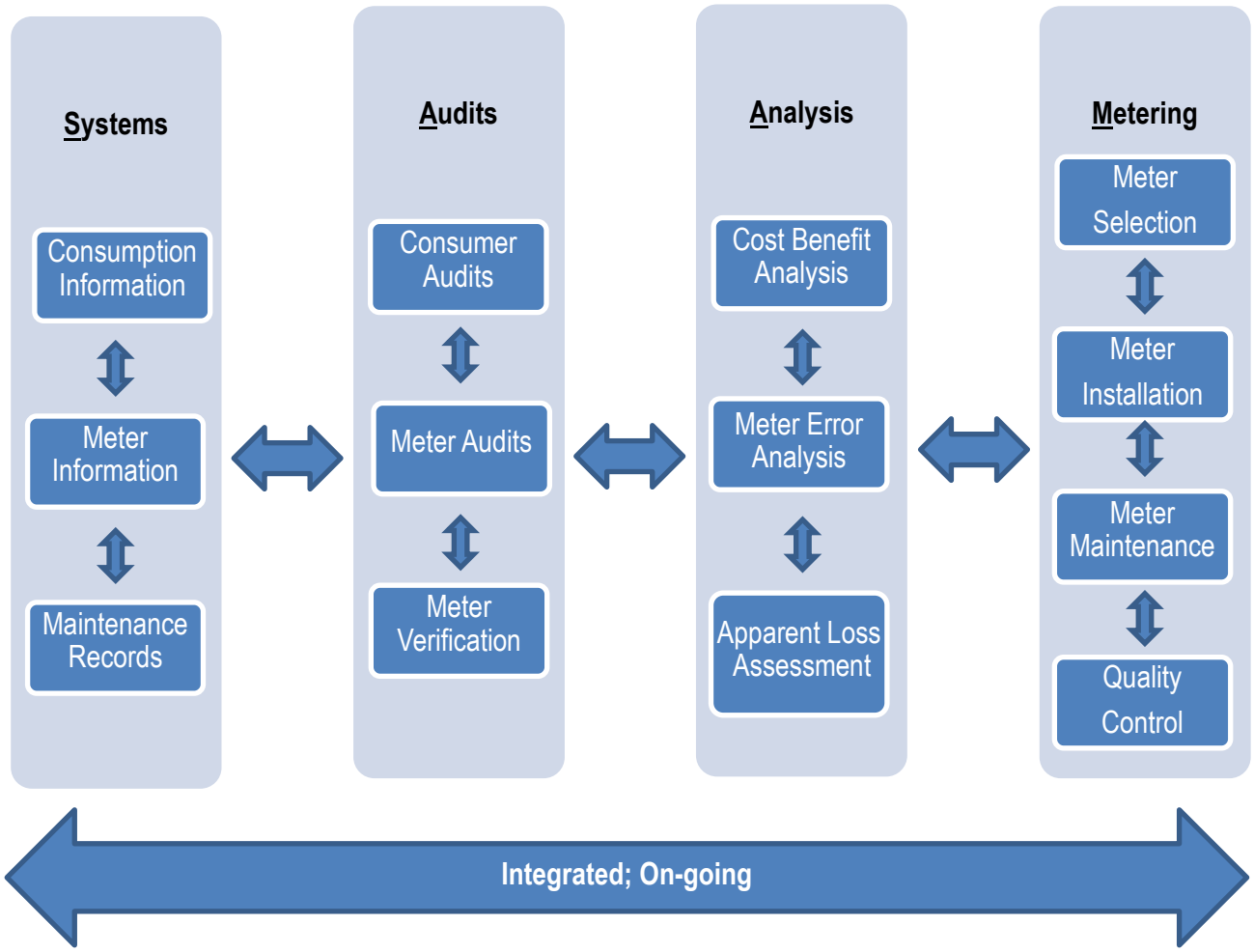


Figure 7-2: Proposed Meter Management Framework

To assist utilities in better managing apparent losses and related activities, a management tool, aptly named "*Apparent Loss Management Tool (ALMAT)*" and is based on the proposed meter management framework of Figure 7-2. The landing page of ALMAT is shown in Figure 7-3 with the grey buttons and black being the activities for which toolsets have been developed and will be explained in greater detail in the following sections.

The management tool has been developed within the widely used and user-friendly Microsoft Excel software. All algorithms are developed using Visual Basic for Applications (VBA) within Excel. The Apparent Loss Assessment, in addition to Visual Basis is also integrated with the R Statistical Software (R Core Team, 2017) through the Basic Excel R Toolkit (BERT), a tool for connecting Excel

with the statistics language R and supports the running of R functions from Excel spreadsheet cells (Structured Data LLC, 2018). This allows the running of algorithms for the developed comparative billing analysis method within Excel. Clicking on any of the activities with defined tools takes the user to a separate sheet for input of data and implementation of the required analysis.

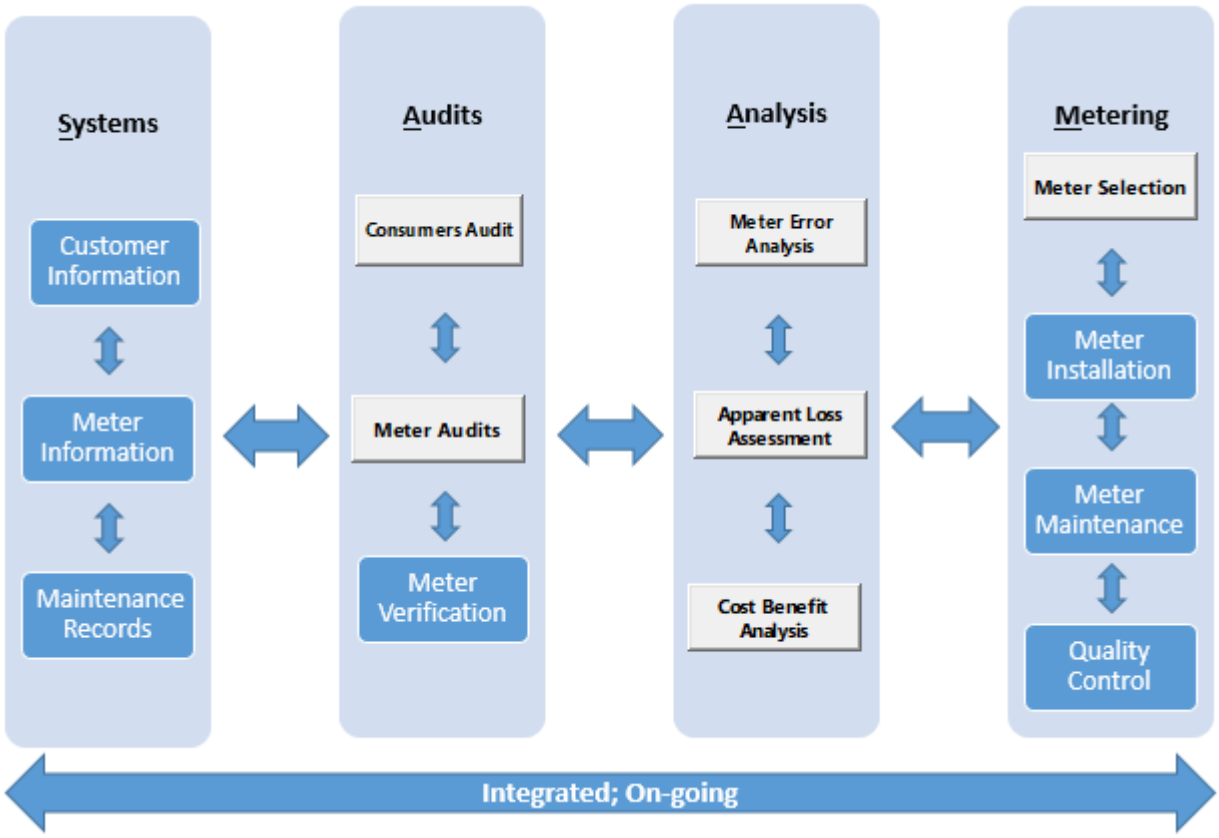


Figure 7-3: Introduction to the Management Tool

7.2 Systems for Metering and Consumption Information

Considerable data can potentially be generated on water consumers and meters over and above the monthly meter readings and such data need to be collected and consolidated by water utilities (Johnson, 2007b). Software systems facilitate such collection and typically comprise the backbone of any management system that provides the bigger picture perspective and related analytics to better inform the decision-making process. What is required is that utilities carefully consider and curate appropriate data systems and processes that will enable them to achieve long-term goals and requirements. Unfortunately, in South Africa metering, billing and costs recovery are largely a finance function that preclude the technical issues addressed in this research (McKenzie, Siqalaba &

Wegelin, 2012). This functional allocation usual means that the interest tends to be very narrow and limited to monthly billing and receipt of revenue to the total neglect of the technical metering aspects that underpin everything else. This is despite the legislative requirements such as the Compulsory National Standards and Measures to Conserve Water (Regulation 509 of 2001) under the Water Services Act (Act 108 of 1997) that requires universal metering, the reporting of information and calculations relating to water losses and compliance with the erstwhile Trade Metrology Act (Act No. 77 of 1973) requirements. The proper performance of such activities needs more that the simple storage of a monthly meter reading, but fully functional and well-designed systems.

While the respective systems have been listed separately there is nothing wrong with meeting all the requirements within one database system. The only challenge is that each of the systems tend to be the responsibility of different functionaries within municipalities.

No tools have been developed for this segment of metering systems as their development falls outside the scope of this research. The output of these systems is however invaluable to the task of assessing and managing apparent losses.

7.2.1 Consumption Information

The main objective of consumer and consumption databases should be to facilitate the proper management of both the consumer and the related consumption. This therefore requires that such information be up to date and accurate. Utilities must be able to have the complete consumption history for a consumer and consumption categories that differentiate between account ownership and the actual water usage type. As an example, there were several properties visited during the meter logging and meter testing exercises in this study whose water use was registered as business simply because the property belonged to a business, but the consumption was residential. Management strategies for either differ significantly.

Demographics information on consumers with updated property information relating to property type, property size and market value will assist in the appropriate segmentation of consumers and the

development of appropriate tariffs and water conservation policies. The ease of consumption segmentation is also dependent on the ability to have the complete consumption history for each member of the segment, which is currently difficult in many jurisdictions as municipalities and utilities tend to keep monthly meter readings per reading cycle.

This research has demonstrated the immense potential of using data mining tools to analyse the consumption to derive useful information such as water losses estimates, which could be the most cost-effective way of extracting useful information. However, such analysis is only as good as the quality of the data.

7.2.2 Meter Information

The water meter is at the heart of all water management activities and is therefore the most vital component. To emphasise its importance, it is commonly referred to as the cash register of a utility. Unfortunately, it tends to be the most neglected asset within the water distribution system. Frequent practices are also that the main attributes that are kept in the database, as was experienced during this research, are those related to the meter reading, that is the meter size and some form of reference to its serial number. The correctness of some these attributes is also dependant on whether the initial meter connection fee relates to the meter size installed and whether there are routine processes related to the checking and maintaining of the correctness of the meter reading and the meter serial numbers.

Critical meter information that should be kept and maintained include;

- *Manufacturer and product specification details* such as the permanent flowrate, the serial number, meter manufacturer, meter model, meter type (multi-jet, single jet etc.), meter class, meter length, number of dials, meter factor (if any), direction of flow and the meter approval number. Meter diameter is the least important and should never be the focus as is commonly practiced, to the detriment of utilities. The serial number, whether it refers to that of the manufacturer or the utility's code, should not be changed under any circumstance, as was observed within Johan-

nesburg Water, and should be valid for the lifespan of the meter. Such information can be obtained directly from the manufacturer and imported into the system to avoid finger trouble. The bulk of the information is static per meter model and such records can be used to populate the database as well.

- Meter verification details, particularly the initial verification information which, despite being a legal requirement in South Africa, is hardly made a meter delivery prerequisite by utilities.

- Meter installation/connection details such as date, reading, geolocation, connection type (underground, above ground, vertical on wall etc.), service type (zonal meter, fire connection, temporary connection etc.), and type of supply (direct, pump, tank etc.).

- Meter maintenance and related requirements such re-calibration. For example, meters with strainers will need specific routines or event-based cleaning of the strainers.

Support for maintaining the integrity of actual meter reading and throughput is very important as it is the final arbiter in cases of any disputes. It is also recommended that no meter clock-overs be allowed but rather that an appropriate meter for the consumer be selected that will be adequate for its entire useful life. In any case, frequent meter clock overs are a good indication that the meter might not be appropriate for the consumer. Should the clock-over be permissible for a utility appropriate management flags to indicate the number of clock-overs and the lifetime reading of the meter must also be instituted. Without this precaution any analysis on the meter data will not produce factual results. Like the consumption information, a meter must also have a complete consumption history with the dates of different consumer, in case of changes over its lifespan.

Failure to have correct meter information will not only frustrate water management efforts but will make meter changes very expensive due to incorrect device information, particularly if the work is subcontracted out. This was also experienced during this research where incorrect meter information frustrated the implementation of meter logging and meter testing based on samples from the meter information systems. Therefore, an up-to-date and comprehensive meter database will be a great asset for a utility and would enable it to do any of its meter related activities with precision while also

opening vast opportunities in terms of what can be done with the data, including the use of data mining and other analytic tools.

7.2.3 Meter Maintenance

Meter maintenance records are an integral part of the meter information systems and has been put separately for emphasis only. They must contain all the maintenance related activities on individual meters and properties, including meter changes. Such records should be the source of historical information on the performance, durability and accuracy of meters that have been installed within a utility. The complete meter maintenance history, when superimposed with financial data relating to the costs related to each meter model, meter manufacturers and so forth will provide a greater understanding of the financial implications of specific meter choices. In overall, this provides valuable input into the evaluation of failure rate of meters, the efficacy of meter changes, meter selection and related activities per meter model per consumer segments.

The use of the consumption data of consumers, such the average annual daily demand (AADD), in prioritising meter maintenance and meter changes makes logical sense and should be utilised by utilities to minimise unbillable consumption when meters fail. This has been demonstrated in this study in §3.4.2.

7.3 Consumers and Meter Audits

The activities of auditing consumers and meters for the purposes of assessing apparent losses has been extensively discussed in this research in Chapter 3 and Chapter 4 respectively. These activities are therefore at the core of the management of apparent losses and are all about building up a knowledge base that provides the required insight in understanding customers and meters better within a specific utility.

As mentioned previously, all the activities are related, and the implementation of the meter and consumer audits provides an opportunity to verify and update meter and customer information, together

with generating additional information for the estimation of the apparent losses. Meter Audit is concerned with the auditing of a specific meter for one consumer while the Consumers Audit looks at a grouping of consumers and their meters.

7.3.1 Meter Audits

Utilities occasionally log meters of consumers for several reasons and the output includes the variation of flowrate with time as shown in Figure 7-4. The objective of the meter audit tool is to assist utilities in better understanding the consumption pattern passing through the meter, to check for the presence of on-site leakage and the adequacy of the size of the meter.

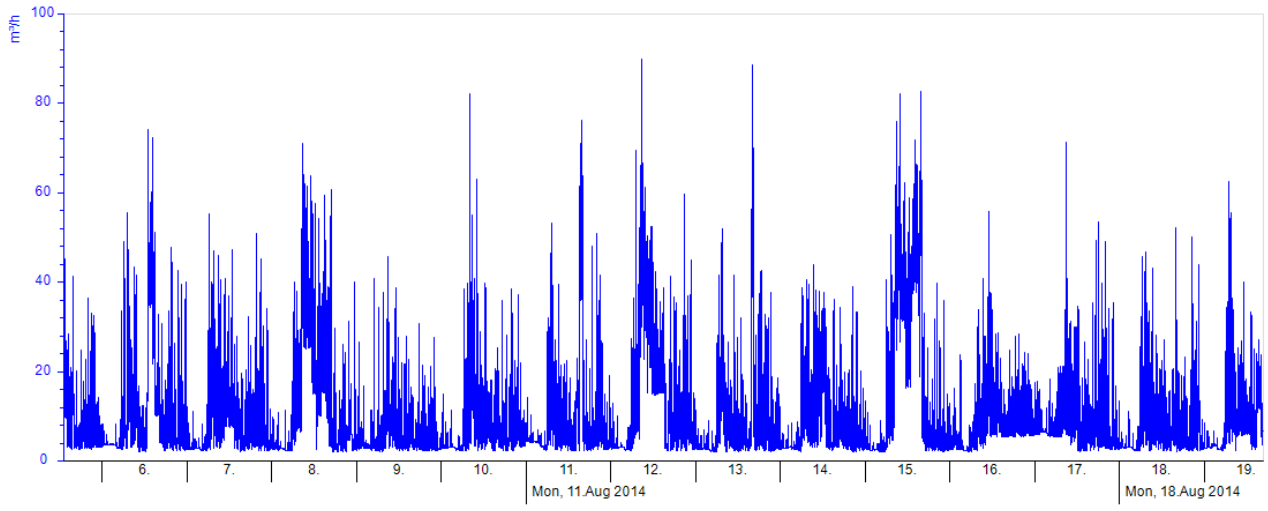


Figure 7-4: Meter Logging Result

The meter in Figure 7-4 is a Class B 40mm Woltmann meter with a q_p = 50m^3/hr and q_s = 100m^3/hr that services a residential estate. It is observed that the highest flow recorded over the 14 days is about 90m^3/hr which is within the maximum flow rate allowable for the meter; there is on-site background leakage of approximately 3m^3/hr throughout the entire period; and that there seems to be a high proportion of low flow which might be a concern. The meter audit tool allows the user to objectively evaluate all this at a click of a button. The landing page for the tool with instructions for the user is shown in Figure 7-5, with the main requirement being the formatting of the file following the structure of the output common to popular meter loggers. It is important to note that fire flow requirements are excluded in the analysis as these are normally legislated. However, users are encouraged

to think of the implication of providing a meter for such higher flows – this toolset has the necessary output to evaluate such impacts.

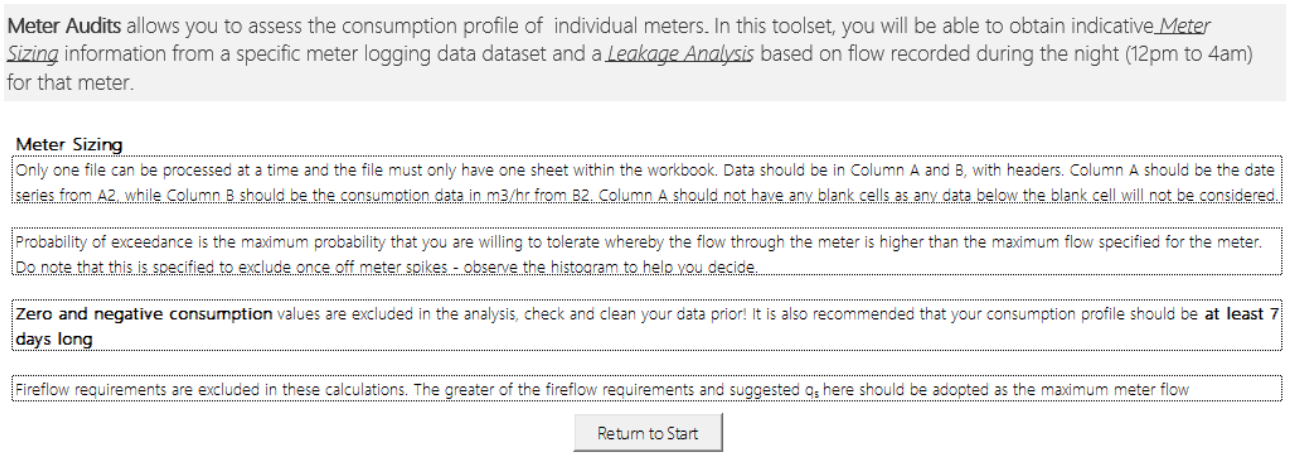


Figure 7-5: Introduction to Meter Audit

The user may choose a specific file to be analysed, with a recommended minimum length of at least seven days. The results of the analysis are shown in Figure 7-6. The statistics of the flow profile are displayed on the tool together with a dynamic histogram where the user can change the number of bins to a maximum of 20, which should be more than adequate for most purposes. All cells highlighted in yellow accept user input, such as the title of the histogram, among others.

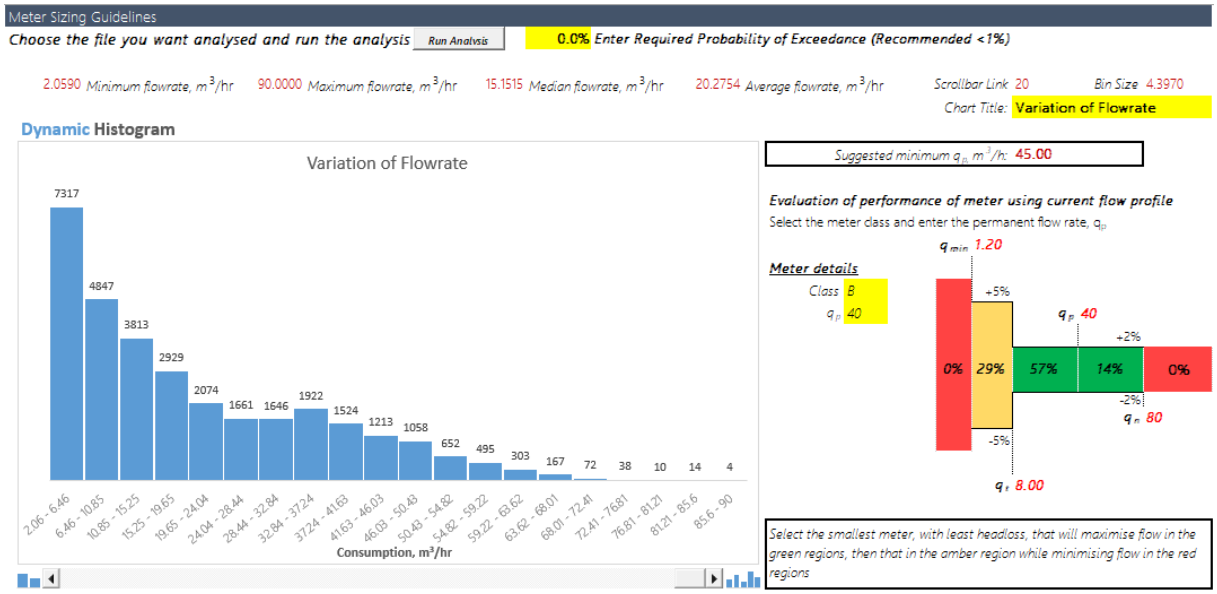


Figure 7-6: Consumption Profile and Meter Sizing Guideline

In addition to the histogram and statistics of the flow profile, meter sizing guidelines are also provided based on the SANS 1529 metering standards. The user can choose to insert the percentage exceedance of the maximum flow that he is willing to tolerate, with a recommended maximum limit of

1% which is to say for 1% of the time the maximum flow of the meter may be exceeded. This is meant to cater for very unusual flows and unexplained spikes when logging the meter. Based on this limit, the minimum limit of a permanent flowrate for a meter is suggested. In the case analysed, a recommended minimum for q_p is 45m^3/hr which confirms the current size of the meter at q_p of 50m^3/hr, with a sufficient safety margin. The user then specifies the characteristics of the meter from which the logging results are from in terms of the meter class (B, C or D options) and the permanent flow rate. The proportion of the flow between the legislated flow rates of q_{min}, q_t, q_p and q_s are also displayed. The meter details may be varied to see what the most appropriate meter that maximises the proportion of flow between q_t and q_s while minimising the flow outside the meter operating range.

As alluded to previously, fire flow requirements are not included, and the user must take the greater of the fire flow requirement and twice the recommended permanent flowrate. It is important that q_p is not used but rather q_s for fire flow requirements as the only requirement in terms of South African standard (SANS 10400) is that the meter does not restrict the required flow for a maximum of four hours. A compliant meter will be able to do this quite easily. In cases where there is a substantial difference between fire flow requirements and normal usage it is suggested that separate connections (fire and consumption) be investigated to accurately measure the daily consumption of the consumer.

Night-flow analysis is also performed for the same consumption profile with results including the night flow statistics, a dynamic histogram and the proportion of flow within the specified meter's flow ranges as shown in Figure 7-7. The night flow analysis is indicative proxy of on-site leakage on the property, as in the most ideal of circumstances night time flow must be close to zero. In this case, the flow is a significant minimum of 2.3m^3/hr and an average of 4.5m^3/hr. The proportion of flow within the meter flow range graphically shows whether such leakages happen at flows that would be of concern to the utility, such as low flows which will likely not be measured accurately due to the poor capabilities of meters to record low lows. In the example above, all the flows occur within the green region, which is better from a measurement perspective as most of this flow will be adequately

measured. The analysis could therefore be the basis for alerting the consumer of likely onsite leakage that needs to be addressed.

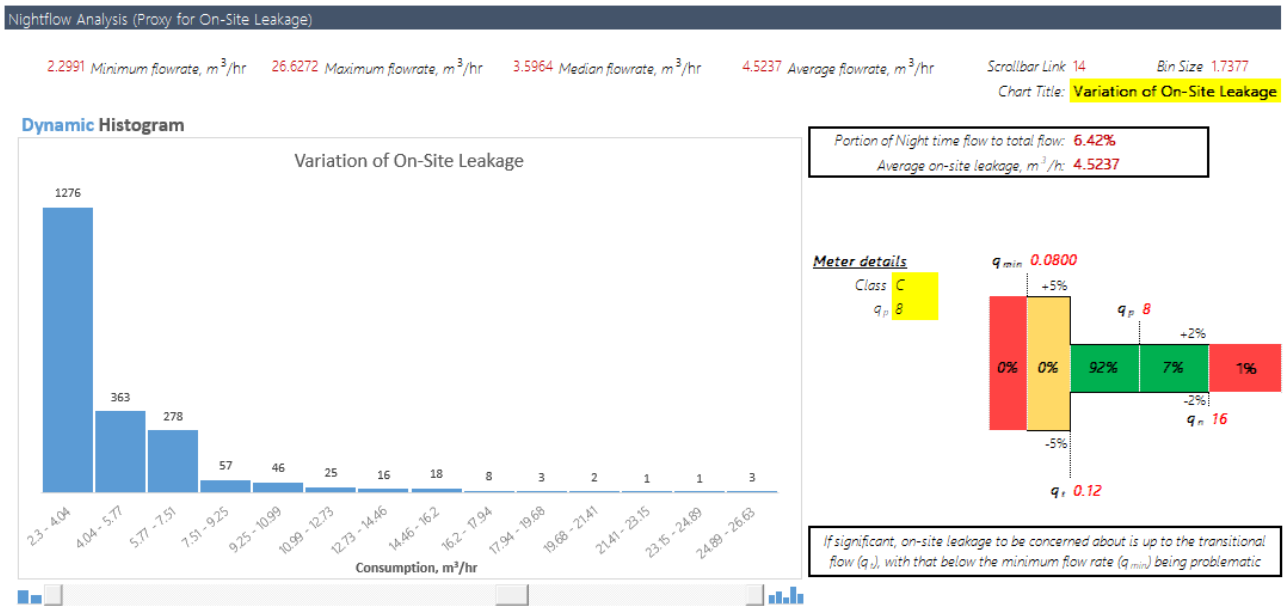


Figure 7-7: Night flow Analysis and Meter Sizing Implications

The tool therefore summarises the consumption profile of any consumer given meter logging data to provide the consumer's consumption pattern, the occurrence of on-site leakage and an indication of the correct permanent flow rate for a meter measuring that consumer. Lastly, the proportion of flow within the legislated flow ranges, based on a user specified permanent flow rate, graphically illustrates whether the user should be concerned with flow distribution for the specified meter both for general consumption and night consumption. While the meter sizing component is useful, the user is referred to the cost benefit analysis and meter selection tools for a more complete resolution of the meter selection problem.

7.3.2 Consumers Audit

Other than verifying consumer information, the audits allow the aggregation of the consumption data of customers over time, such as those from the meter audit, to develop consumption profiles. The latter is the objective of the consumers audit tool and it allows the user to consolidate meter logging data from a grouping of similar consumers to determine their consumption profiles. The specific requirements of how consumers can be selected, as was done in this study, is covered in §3.1.1.

Although such an elaborate approach is commendable, it is not a requirement to use the tool. Figure 7-5 shows the landing page to consumers audit tool.

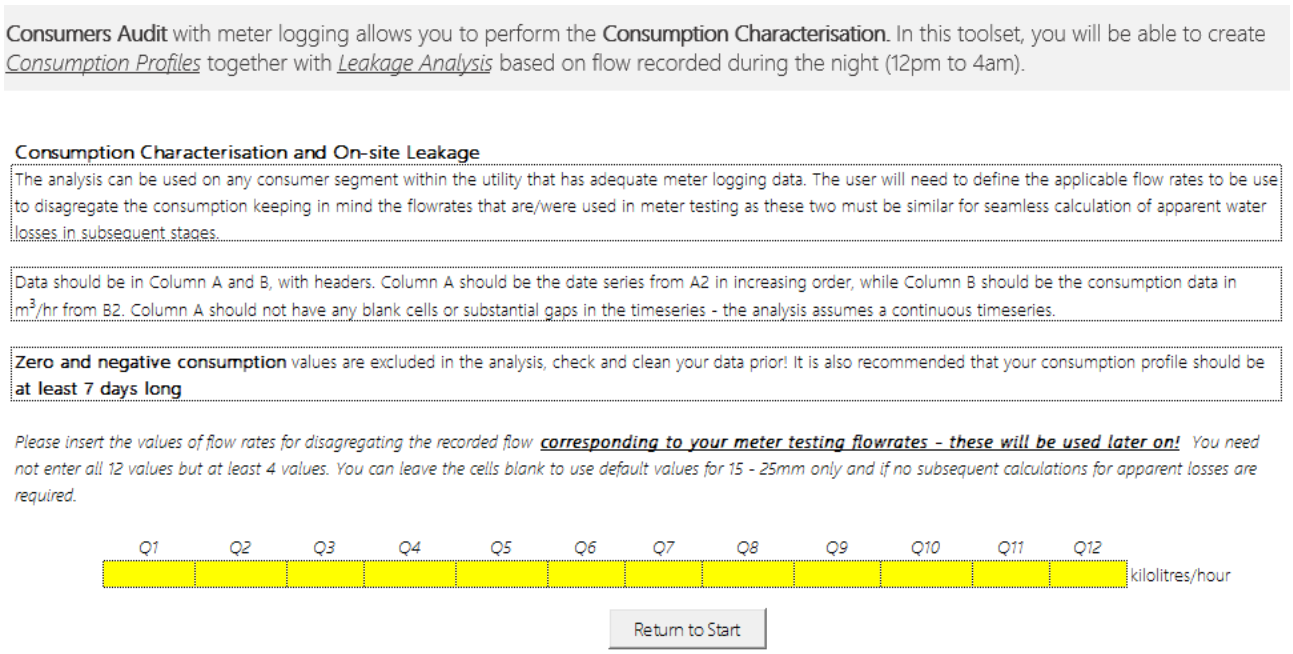


Figure 7-8: Introduction to Consumers Audit

The main requirement is on the formatting of the native files, with only two columns of data required per file, which is standard output of common loggers used in utilities. There is the option to specify at least four flowrates required for data aggregation. If these are not entered, the flow rates for the meter testing of 15mm in Table 4.5 are utilised. Unlike the Meter Audit tool, because of the number of files, the analysis time and the subsequent use of the information for apparent loss assessment, it is more practical to fix the flow rates ranges as opposed to the dynamic histogram output in the previous tool.

The user may select the location of the folder for all the meter logging files, each with a recommended minimum length of seven days, to perform the consumption characterisation and night-flow analysis, as shown in Figure 7-9. In the background, each individual file is parsed and the proportion of flow within the specified flowrates is determined, excluding all zero and negative flow values. Provision is made to allow a user to estimate the acceptable night-time use. Hamilton & McKenzie (2014) provides a guideline value for the assessment of night flow of a district metered zone as 6% of the population being active at night using about 10 litres each. This is rather complicated for this case

and a percentage of time between the hours of 00:00 to 04:00, a total of 240 minutes, has been adopted instead. It is suggested that 10% of that time (24mins) be considered sufficient for allowable night time use.

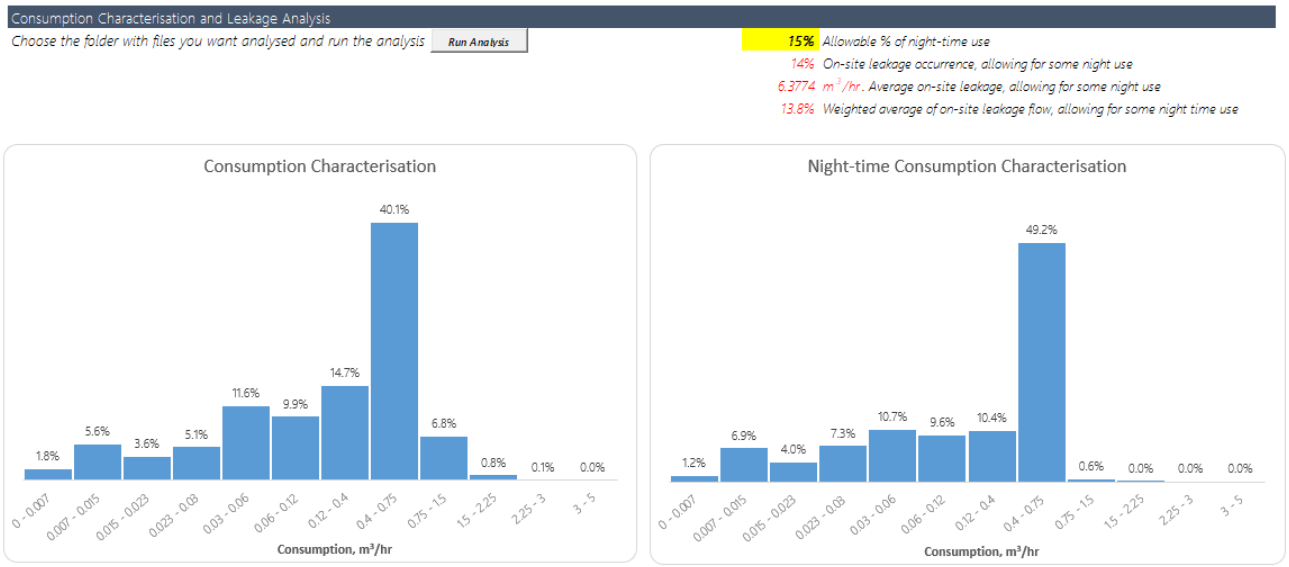


Figure 7-9: Consumption Characterisation

The obtained consumption profile can be utilised later in the estimation of apparent losses and it also provides meaningful data on what could be the most appropriate meter sizes for the consumer category in question. Secondly, specific interventions may also be designed for the possible occurrence of on-site leakage as demonstrated by the night-time usage for the same consumers and others like them. For instance, Figure 7-9 shows that there is a high occurrence of leakage even at high flows, with a global average flow of 6m^3/hr, which is substantial. The specific consumers may be notified accordingly and given notices to attend to the leaks in accordance with utility's by-laws. All the graphs within the tool can be selected, modified and copied to be used in any other document as required.

7.3.3 Meter Verification

Meter verification refers to the testing of meters to verify if they still fall within the prescribed accuracy limits. Because meters are used for custody transfer, the verification of their accuracy is a requirement in many, if not all, jurisdictions. In South Africa, in accordance with the Legal Metrology Act, all measuring instruments are subject to initial verification and subsequent verification in accordance with the relevant legal metrology technical regulation. While these regulations are currently not in

place yet, it is safe that the on-going verification requirement of the repealed Trade Metrology Act will be maintained or further enhanced.

The initial verification requirement spelled out in regulations compels meter manufacturers to test the meter at a SANAS accredited laboratory before they sell it to a user or utility. The required three test flow rates according to SANS 1529-1:2006 are $1 – 1.1q_{min}$, $1 – 1.1q_t$ and at least $0,9q_p$ with the allowable tolerances being 5% for flow below q_t and 2% for the rest. Such test certificates should therefore come with the meter when purchased, which usually does not happen and when it does some manufacturers only put the overall test result for the meter not the detailed information. Utilities should therefore require the full test results and include such information as part of the meter information to be included in §7.2.2 at no cost to them. Utilities should also consider the additional testing of samples of new meters per meter model to establish the starting flow rates or other low flow rates and other flow rates to be able to estimate the initial error of the meters. An example of such work is detailed in Arregui et al. (2017) and provides useful information for decision making in the selection of meters. Such additional initial verification may be better placed within a quality control program discussed in §7.5.4 thereby achieving two objectives from the same effort.

In addition to the initial verification of meters, utilities should institute routine programs for the testing of meters over and above the reactive tests done in response to consumer complaints. These should be based on sampling theory as was done in §3.1.1 and be part of a routine program to manage costs. Such programs will provide verification information for groupings of meter models according to different ages and meter reading required for the determination of the performance of homogeneous grouping of meters with time and consumption. Like initial verification, the legislative test flow rates remain the same with the only variation being the allowable tolerance of 8% for flows below q_t and 3.5% for the rest of the flow rates. For apparent losses assessment purposes, it is important to evaluate the accuracy at additional flow rates, particularly low flow rates. Such additional flow rate consideration needs to be considered when dealing with reactive consumers' test requests to be able to derive an additional benefit at a minimal additional cost. This is because the bulk of the cost of the meter test is and should be borne by the consumer in such requests. §4.1 demonstrated the

usefulness of consumer testing data, and therefore a few enhancements of additional flow rates to the routine testing data will be invaluable to the task of assessing and managing apparent water losses. It is important to remember that the higher the number of records available for meter error degradation, the higher the reliability of the results obtained from the data. As such, the consolidating of test results from various possible sources, such as neighbouring utilities, will leverage each utility's dataset to for a greater benefit to the wider sector.

Some accredited flow laboratories in South Africa, such as Johannesburg Water, have adopted five standard test flow rates for all meter tests conducted within the lab based on the flow rates used for initial efficiency tests in SANS 1529-1:2006. These prescribed flow rates are 1 - 1.1q_{min}, 1.0 - 1.1q_t, 0.45 - 0.5q_p, 0.9 - 1.0q_p and 0.9 - 1.0q_s. Such practices are commendable and generate useful information and can be supplemented by the addition of lower flowrates than q_{min} to complete the picture. Table 4.5 can be used as the basis for the determination of such additional flow rates with the added caution to utilise the consumption category in question to determine the flow rates to use. For example, in cases where domestic consumers are connected via 20mm and 25mm meters, only the flow rates recommended for the 15mm meters in Table 4.5 should be used.

Considering that all the meter verification being done is for the primary objective of complying with the requirements of regulations, the use of an accredited laboratory will therefore be important as such results are trusted and can be better defended in a court of law as reliable. There are also provisions for newly established laboratories in South Africa to be part of a proficiency testing scheme that utilities can be part of to enhance the credibility of their results.

The evaluation of meter verification results is included in §7.4.1 in the analysis of meter errors of a grouping of similar meter models.

7.4 Analysis and Assessments

The Audits of the SAAM meter management framework generates data that are a required input into the Analysis component of the framework, which is the subject of this section. The quality of the data invariably impacts on the analysis and as such users should take time to prepare their datasets to

ensure that the output is valuable and usable. It is also important to format the individual meter test reports to achieve the compilation of the consolidated summary to be used in subsequent analysis. Due to the variance in the formatting of test reports, no tool for consolidation has been included in the ALMAT.

7.4.1 Meter Error Analysis

Figure 7-10 shows the landing page for the meter error analysis based on meter verification data collected through the activities of meter verification (§7.3.3). The main requirement is the format of the consolidated summary of the meter testing results that allow for up to ten test flowrates. Both the meter age and meter reading for each individual record should also be provided whenever available to be able to evaluate the degradation of the meter error. The user can then select the appropriate input file to commence with the analysis.

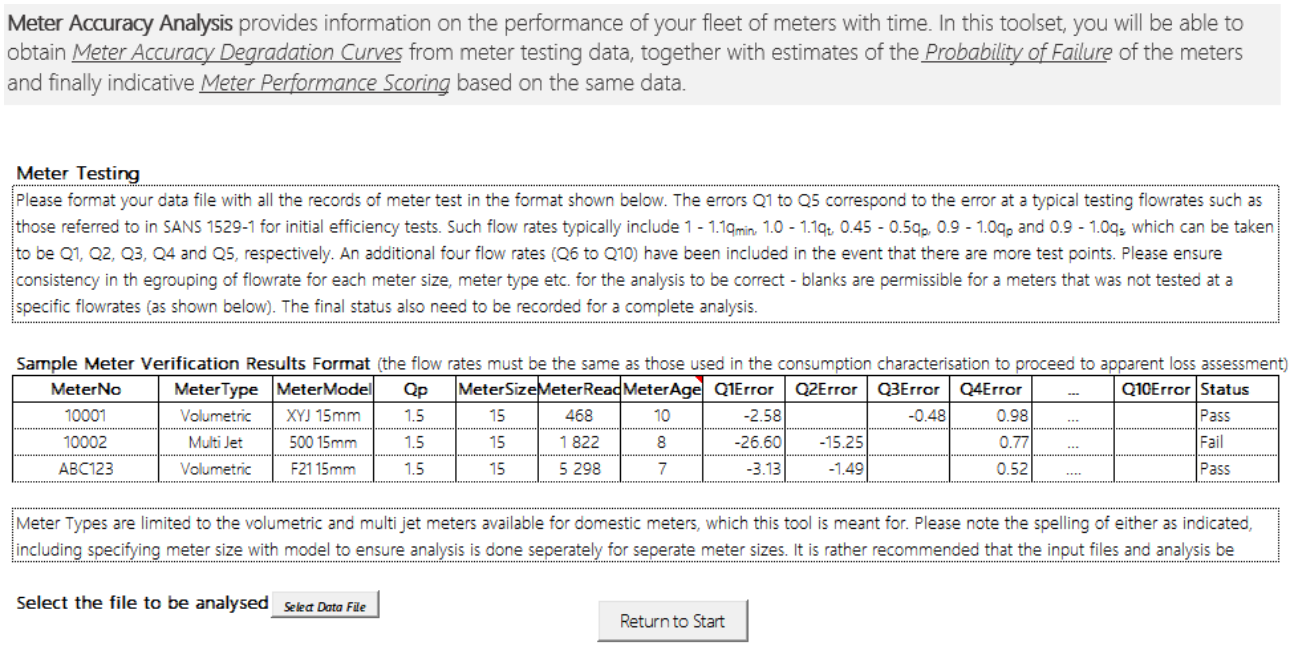
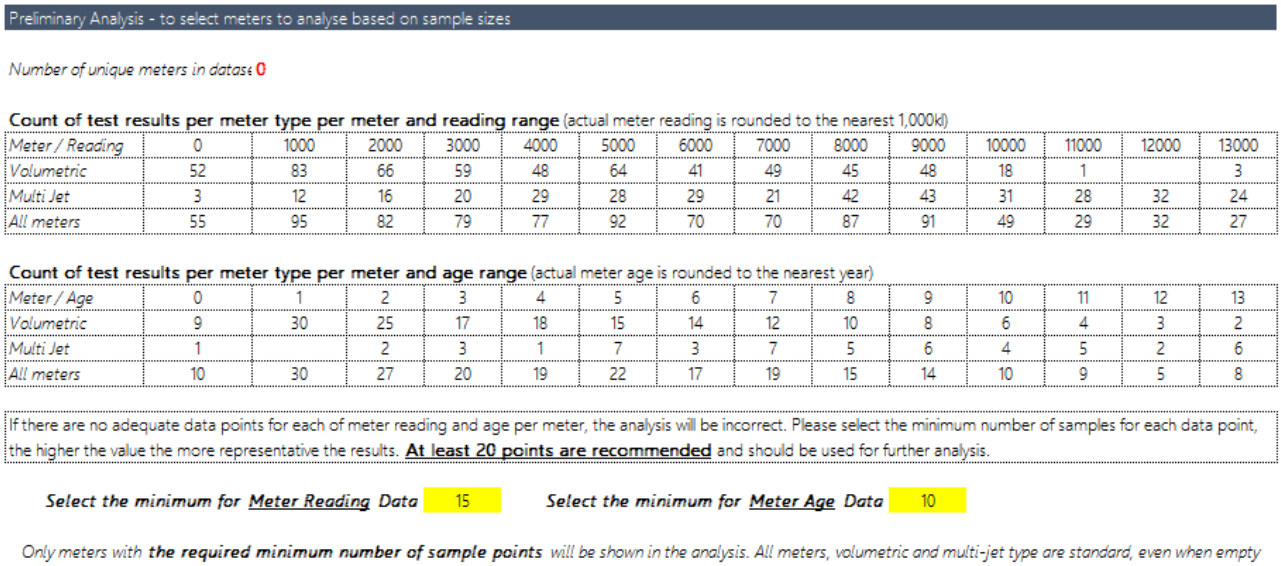


Figure 7-10: Meter Error Analysis Introduction

Care must be taken to ensure that for each meter model and size, the test flow rate for Q1 to Q10 are consistent within each grouping, particularly if the analysis is to be done for multiple meters at the same time. The analysis is done according to both meter type and meter model and therefore these fields must be correctly filled. In the provided example, the meter size is included as part of the meter model to allow for the combined analysis of a wider grouping of meters.

The first part of the analysis starts by summarising the data to allow the user to observe the available meter type and meter model with respect to the categories of age and meter reading. Figure 7-11 shows the preliminary analysis screen with the default meter type information. When the user selects a file, several rows, corresponding to the number of unique meter models in the database, are inserted per reading range and/or age of the meter. The meter age is rounded to the nearest year while the meter reading is rounded to the nearest one thousand.



Preliminary Analysis - to select meters to analyse based on sample sizes

Number of unique meters in datase 0

Count of test results per meter type per meter and reading range (actual meter reading is rounded to the nearest 1,000kl)

Meter / Reading	0	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000	11000	12000	13000
Volumetric	52	83	66	59	48	64	41	49	45	48	18	1		3
Multi Jet	3	12	16	20	29	28	29	21	42	43	31	28	32	24
All meters	55	95	82	79	77	92	70	70	87	91	49	29	32	27

Count of test results per meter type per meter and age range (actual meter age is rounded to the nearest year)

Meter / Age	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Volumetric	9	30	25	17	18	15	14	12	10	8	6	4	3	2
Multi Jet	1		2	3	1	7	3	7	5	6	4	5	2	6
All meters	10	30	27	20	19	22	17	19	15	14	10	9	5	8

If there are no adequate data points for each of meter reading and age per meter, the analysis will be incorrect. Please select the minimum number of samples for each data point, the higher the value the more representative the results. **At least 20 points are recommended** and should be used for further analysis.

Select the minimum for Meter Reading Data 15 *Select the minimum for Meter Age Data* 10

*Only meters with **the required minimum number of sample points** will be shown in the analysis. All meters, volumetric and multi-jet type are standard, even when empty*

Figure 7-11: Meter Error Preliminary Analysis

Depending on the data source, it is possible that there will be insignificant counts of meter age or meter reading for certain categories. For instance, in Figure 7-11 there are only 2 multi jet meters with an age of 2 and one volumetric meter with a reading of 11,000. The use of averages from such small values in the analysis will lead to inconclusive results. To curb against this, the user is required to set the minimum number of records per range for that specific range to be considered in the analysis. While at least twenty records are recommended, the user can enter a minimum value from ten upwards as it is recognised that some utilities may initially not have adequate data.

The first result of the meter error analysis is the probability of failure of all meter models and types with the required minimum number of records per range. Figure 7-12 shows a typical output filtered for volumetric meters. As with any Excel chart included in this management tool, the user can click on any of the graphs, filter accordingly, modify the graph type, insert a trend line and many other activities that can be done on a typical Excel chart. As an example, the probability of failure with

meter reading is reproduced in Figure 7-13 with a trend line and few modifications from the same graph in Figure 7-12.



Figure 7-12: Probability of Failure Analysis

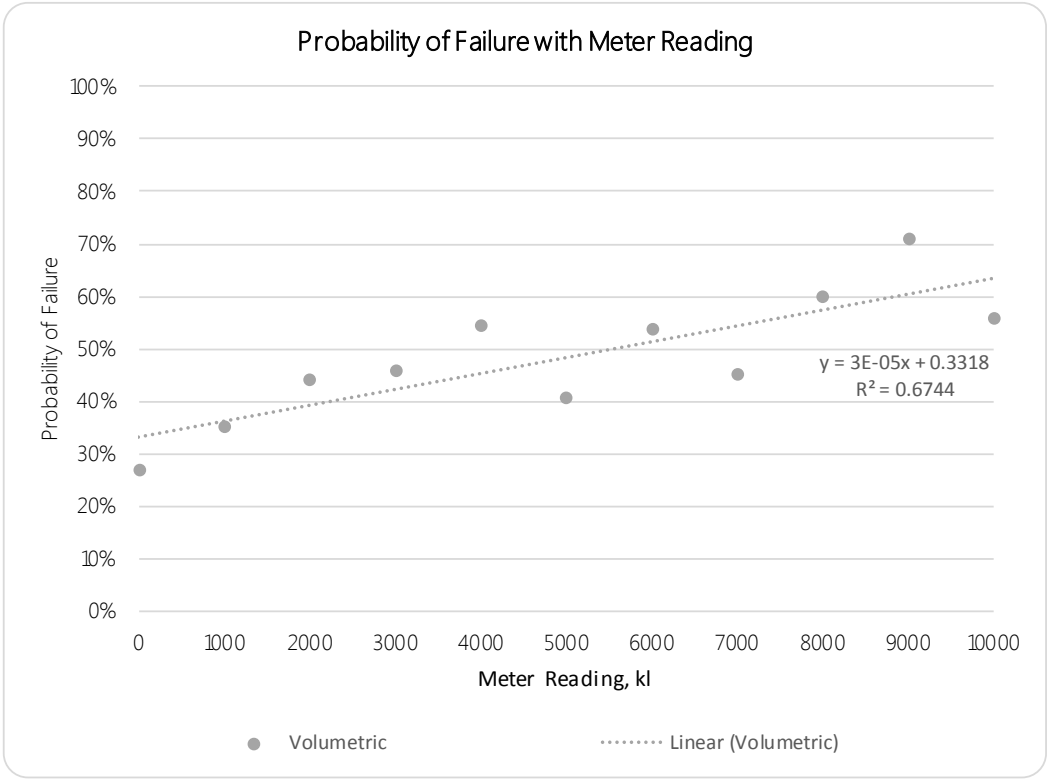


Figure 7-13: Probability of Failure with Meter Reading

The next part of the analysis is the error degradation with respect to meter types, namely volumetric, multi jet and Woltmann meters. While this analysis should not, in the strictest sense, combine differ-

ent meter models/manufacturers using similar physical measuring philosophy (e.g. volumetric or velocity), it is a needful requirement for as long as African utilities do not have adequate datasets to analyse. This analysis produces results that assist utilities in assessing the performances of various meter types. Figure 7-14 illustrates the output of the analysis filters for Q1 and Q3 errors. As can be observed, the range of the meter reading is adequate to realise meaningful plots, while the meter age chart has fewer values which give meaningful plots despite the gaps.



Figure 7-14: Degradation of Error Analysis – Meter Type

In comparison, Figure 7-15 shows that when you go to specific meter model the data becomes less and less, with no data for the meter age graph on the right. The user is cautioned, in view of the results of the preliminary analysis shown in Figure 7-11, to interpret the error degradation analysis based on few observations per range to be indicative.

In cases of adequate data, it becomes a simple exercise to combine the meter degradation rates with the consumption profiles to obtain the weighted average degradation rate which is important for a few other tasks such as the cost benefit analysis and meter selection. This is discussed in greater detail in §7.4.3.1.

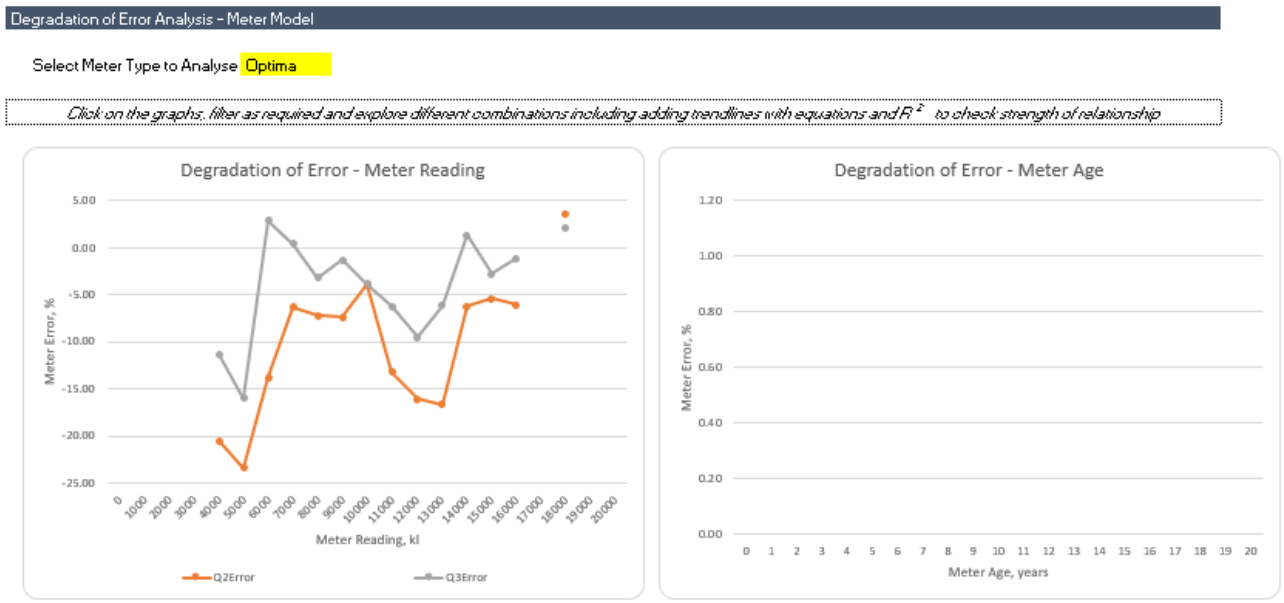


Figure 7-15: Degradation of Error Analysis – Meter Model

7.4.2 Cost Benefit Analysis and Optimum Replacement Period

The cost-benefit analysis (CBA) is an economic assessment based on the discussions covered under §2.3.1. Key to this analysis is an assumption that the optimum meter replacement period is unknown and needs to be found. This will be different for utilities that are required to change meters at specified periods. The calculation of the optimum meter replacement is based on the concept of the net present value of infinity meter replacements, with the year of the lowest cost being the most optimum year. This is calculated using Equation 2.12.

The key inputs to this equation include the real discount rate, a function of the nominal discount rate and inflation as provided in Equation 2.10; the costs of the meter and associated costs such as installation costs, administration cost and other related costs; the price of the water which is assumed constant over time; the annual consumption of the consumer; the initial weighted error of the meter and the weighted error degradation rate. The input screen for the CBA tool is shown in Figure 7-16, with all yellow cells being input cells whose values need to be supplied by the user.

One important inclusion within the tool that had previously not received attention in other similar tools is the use of an indicative tariff based on a rising block tariff. This omission might seem trivial but

leads to errors where some estimates are only based on the lower rung of the rising block tariff or the bulk purchase rate of potable water. The indicative tariff is calculated by assuming a monthly average consumption from the annual consumption and applying the rising block tariff to it to calculate the total cost. This cost is thereafter divided by the volume to obtain the indicative tariff.

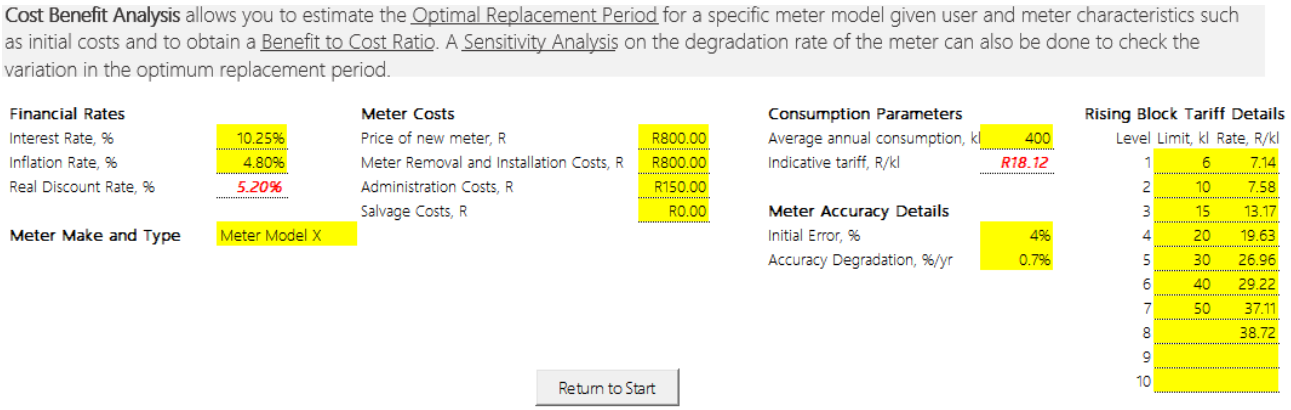


Figure 7-16: Cost-Benefit Analysis Input Screen

The initial weighted error of specific meter models will likely not be known in many utilities and will need to be evaluated. In the interim, values from literature such as the 4%, which was also confirmed for a random meter in this study in §5.1.2 could be used. The weighted meter error degradation curve per meter model is another unknown variable that also requires careful consideration and determination. This will require the analysis of meter error with time combined with the consumption profiles of consumers or the use of alternative methods, such as the one developed in this study. Values such as 0.7% found in the alternative method of this study may be used, together with other similar values from literature, in cases where the utility specific values are not known. Therefore, it is important to implement the recommendation of meter verification in §7.3.3 so that such values are determined and used for local meters.

The tool to obtain indicative replacement periods both in terms of age and the meter reading at the optimum replacement period together with the benefit to cost ratio and percentages of lost revenue are shown in Figure 7-17. While a value is provided for the optimum replacement period, a closer analysis shows that there are several years close to the minima that have similar net present value of replacement costs. The user is therefore cautioned that while the provided year is the mathematical optimum, surrounding years provide comparable results and therefore should be considered

accordingly within the broader plans of the utility. The analysis is performed over a 20-year period which should be sufficient for most intents and purposes. However, for analysis where there is very low annual consumption the optimum period might be longer.

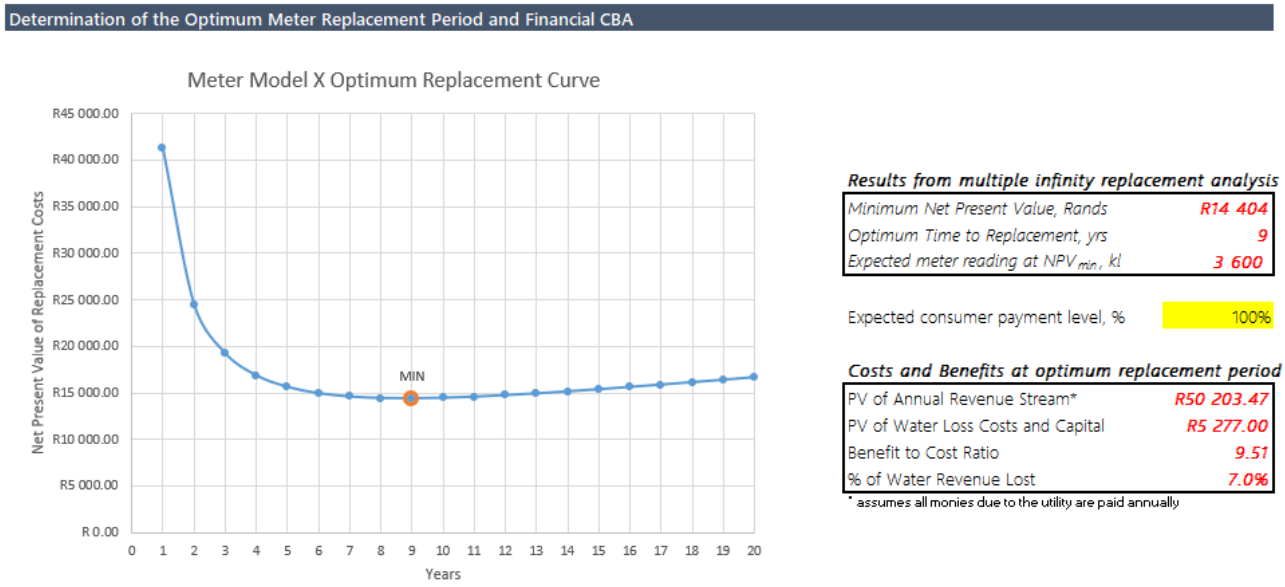


Figure 7-17: Optimum Replacement Period and CBA

Also included in Figure 7-17 is the provision for a realistic benefit to cost ratio that takes into account payment defaulters to the utility. This was not included in previous similar tools such as in Mutikanga *et al.* (2009). Another addition previously not included also is the estimate of the percentage of revenue lost over the proposed lifespan of the meter. Considering that meters tend to be a regulatory requirement and that water is a precious resource, certain utilities may want to minimise the percentage of revenue lost to say a maximum of 10% of total revenue, leading to either better meters from a measurement point of view but with lower cost-benefit ratio. Due to the nature of the spreadsheet environment, all variations in the input data readily update the results so that it is therefore trivial to determine the variables which give optimum result to the utility.

Because of the uncertainty regarding the meter degradation rate, which is one of the most important variables in the equation as it is applicable throughout the life of the meter, it is important to consider its influence on the optimum replacement period. To cater for this requirement a sensitivity analysis has also been included as part of the tool as illustrated in Figure 7-18 where the uncertainty is provided as a percentage of the degradation rate. The minimum and maximum of the degradation rate

are used to plot the corresponding curves of optimum replacement period. In the provided example, while a degradation rate of 0.7% has been specified, it is estimated that it could vary by 40% in either direction meaning it could range from 0.42% to 0.98%. This leads to a variation in the optimum replacement to be in the range of 7 – 12 years, from the initial estimated 9 years.

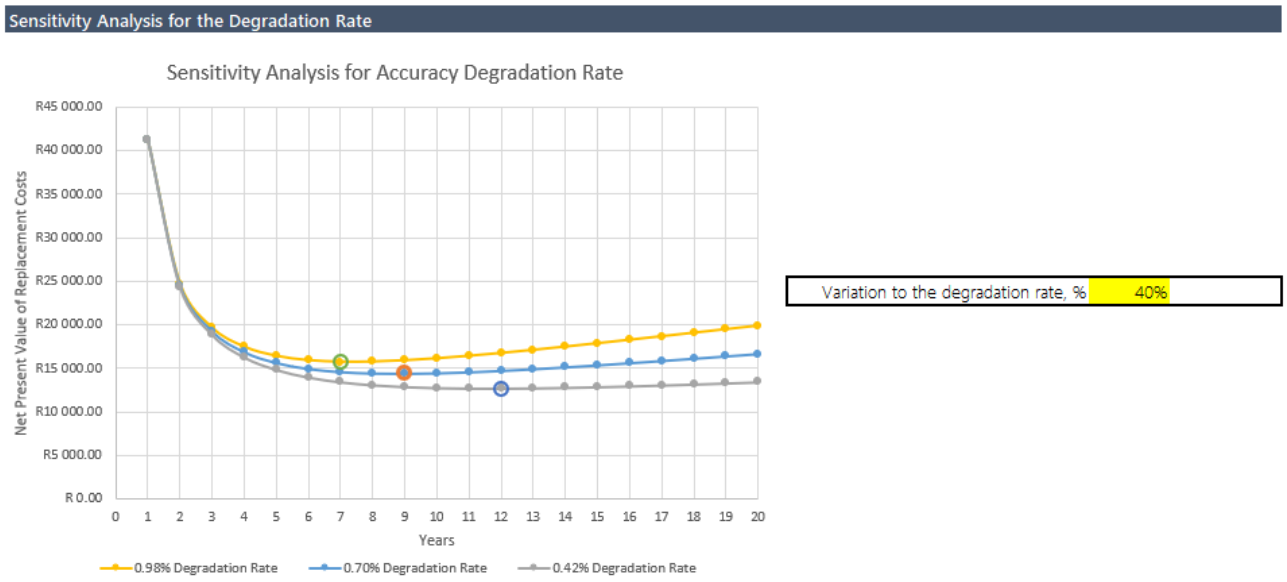


Figure 7-18: Degradation Rate Sensitivity Analysis

While sensitivity analysis does serve a purpose, the best way of minimising uncertainty is to generate adequate data to properly characterise the consumers and meters within a utility area. This can also be done by using an appropriately validated methodology as was demonstrated in this research.

The sensitivity analysis and the preceding CBA analysis are only applicable to one meter type at a time. If a comparative assessment for multiple meters is required, this can be achieved as described in §7.5.1 for a longer period of 30 years. However, the user would need to return to this part of the tool to perform a sensitivity analysis to confirm or evaluate the results for a chosen meter.

7.4.3 Apparent Loss Assessment

Apparent water loss assessment is a key component of the management tool as it allows for the standardisation of the calculation of the apparent water loss based on the weighted meter accuracy methodology and the new comparative analysis method. For the weighted error method, the data

used is a combination of the output of meter accuracy analysis (§7.4.1) and consumption characterisation from consumers audit (§7.3.2), with the latter being optional as the proportions can be inserted manually. It is therefore important that users only attempt this component when they have done the component with similar flow rates used in aggregating the consumption patterns and in meter testing. It is a very simple exercise to rerun the consumers audit component of the tool with the meter test flow rates.

The landing page of the assessment is shown in Figure 7-19 which introduces the user to the various aspects of the tool and the requirement in each case. For simplicity, this overview has been split into two components covering both methodologies.

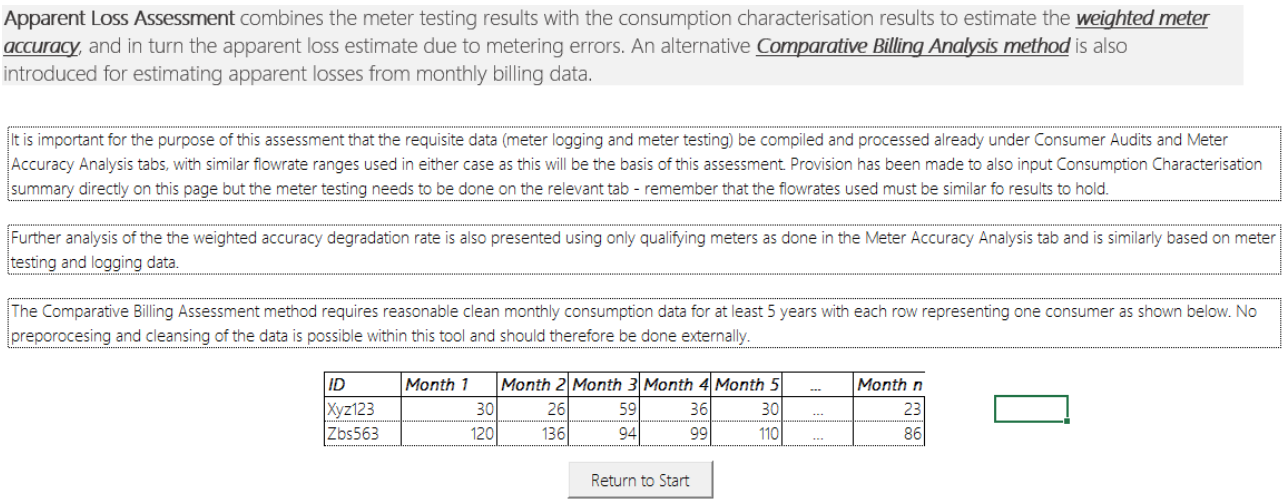


Figure 7-19: Apparent Loss Assessment Introduction

7.4.3.1 Traditional Approach

The first component of the traditional approach applies the modified weighted accuracy equation, Equation 4.3, to determine the weighted accuracy for each of the unique meters in the database by superimposing the specified or default consumption profile. The results are shown in Figure 7-20.

Other than having done prior analysis on the meter error, the user can choose to specify the proportion of flow that falls within each meter testing range manually in the yellow cells shown in Figure 7-20. This is in cases where the user has not implemented consumption characterisation but wishes to adopt values from a similar environment. However, when these values are blank, the used data

used are drawn from the consumers audit analysis and therefore it is important to verify the correctness of the data.

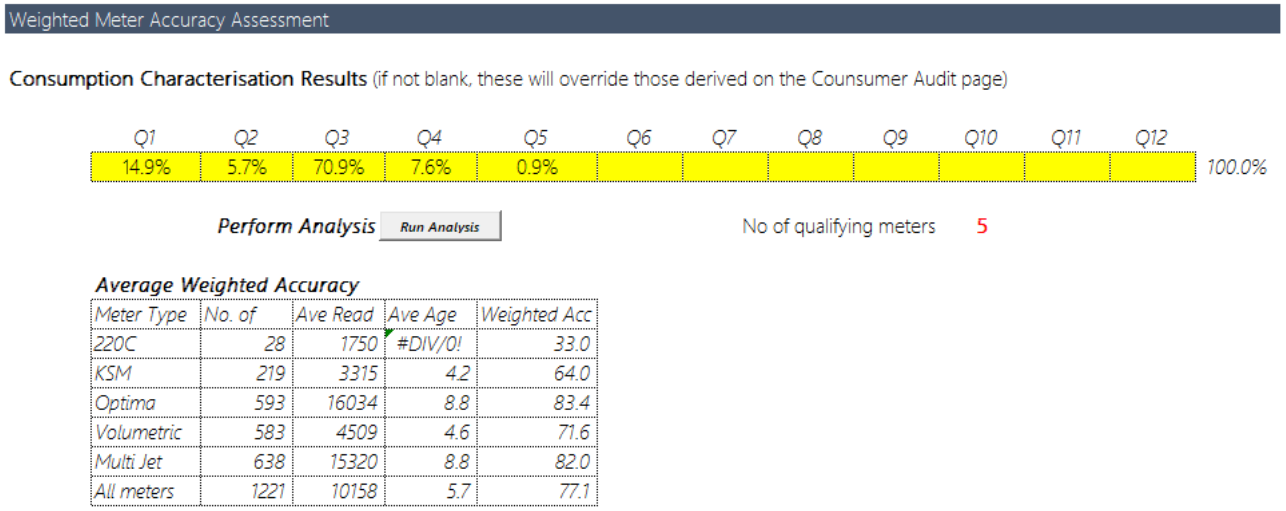


Figure 7-20: Weighted Accuracy Estimates

The second component is the determination of the degradation of the weighted accuracy of the meter by meter type. Like the meter error analysis, this category is only included for preliminary analysis while the utility gathers data on their meter types. The output is shown in Figure 7-21 and is also based on the minima set for both meter age and reading in §7.4.1.

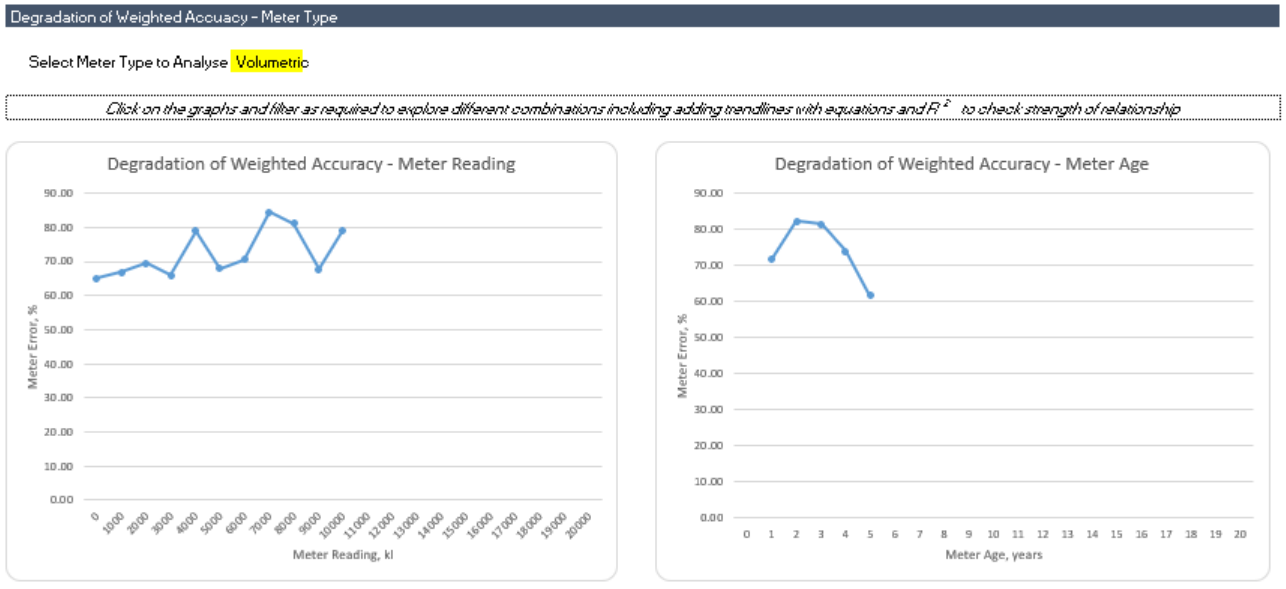


Figure 7-21: Degradation of Accuracy - Meter Type

Evaluation of the results of Figure 7-21 should be done while observing the number of meters per category (as in Figure 7-11) to decide on the appropriateness of the data and any outliers. As with any analysis, the results must make sense. In the above example the degradation of the error with

meter reading seems to be improving for volumetric meters which suggests that the meters get more accurate with increased volume. This might be explained by a concentration of consumption in a region where the meter error increases towards positive errors, which should be confirmed, otherwise the data do not make logical sense. On the other hand, the degradation with meter age could be more meaningful. The more appropriate analysis is the degradation of the weighted error for each unique meter model, as illustrated in Figure 7-22, and is also included within the management tool.

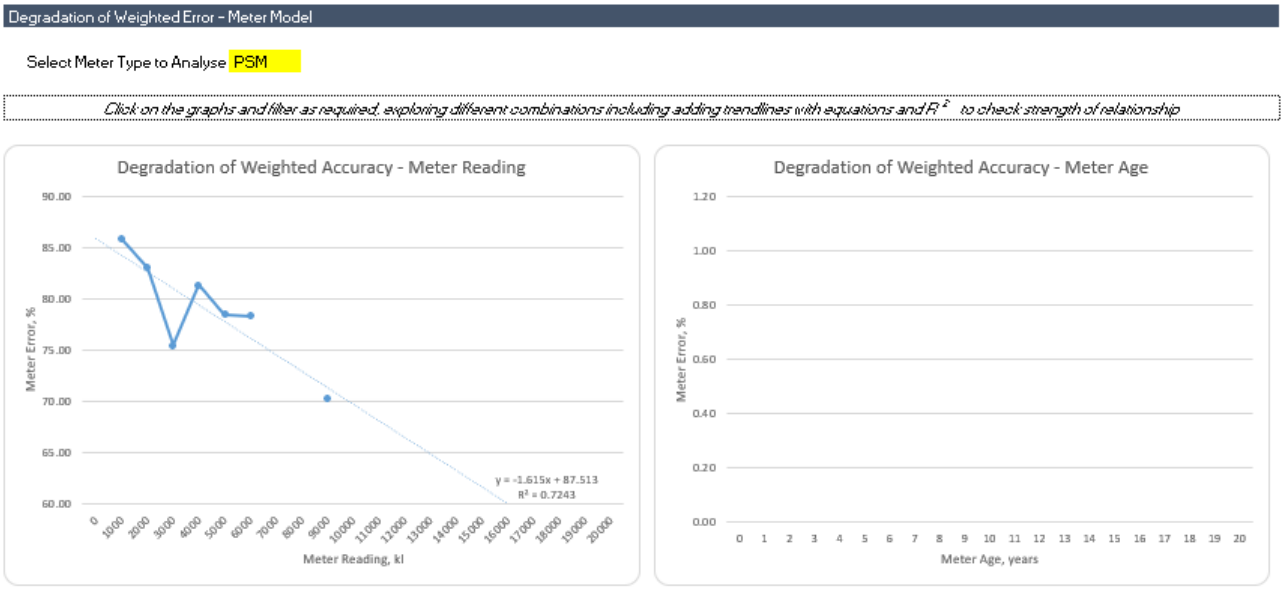


Figure 7-22: Degradation of Accuracy - Meter Model

The degradation of accuracy with meter age does not have adequate data and hence the plot is empty while with meter reading there is a strong relationship with an $R^2 = 0.72$ after the removal of two outliers. The degradation rate is -1.62% per 1,000kℓ consumed, which is rather high but possible.

7.4.3.2 Comparative Analysis Approach

All the components of the management tool have been done entirely within the Excel environment with data being kept in the background of the tool. The application approach for the alternative method is slightly different – Excel still remains the interface but a bulk of the work is implemented in external Excel files which are processed in, or are an output of, the R Statistical Software (R Core Team, 2017). The user does not need to understand the R language as only the familiar Excel tools are used. The main requirement is that the user installs BERT (Structured Data LLC, 2018) that is freely available from https://bert-toolkit.com/ and to copy the code written as part of this study from

§B.1.2 and copy them to the bottom of the “functions.R” file (found in the BERT installation folder) which can be opened by any text editor such as Notepad. The code is required within the management tool to implement the various analyses following the process described and implemented in Chapter 6.

The initial step within this alternative method is data preparation, starting with the required file format in Figure 7-19 with the month being a date field in Excel, which can then be selected and processed from the tool as illustrated in Figure 7-23. It is important that the data preparation is limited to what is provided in the input, but it is diverse and dependent on a variety of factors that precludes the development of a tool that accounts for all possible situations.

The processing done within the tool refers to the filling of any gaps in the data using linear interpolation between the start and end date of each individual record. No interpolation is done for periods before and after the start and end date of the record, respectively. The filled data are transformed into running annual totals and saved as a *.csv* file in the same folder as the initial data file. The global average annual running total is also plotted within the tool as shown in Figure 7-23 to allow the user to see if the data makes sense. Because of the transformation of the data into annual running totals, it is not expected that there be any unexplainable drastic variations in the data. The variation in the first ten to twelve months in Figure 7-23 is due to the presence of very few records at the start with high consumption and is acceptable. The user must also review the *.csv* file and cross check that the running annual totals, among other things, have been calculated properly and if everything is in order the file must be saved as an Excel file prior to the next stage.

After the preparation of the annual running consumption the next step is to determine the number of clusters through a preliminary clustering analysis. The user selects the relevant annual consumption file and the analysis is done using classical partitional clustering with the Euclidian distance and the k-Shape algorithm for up to ten clusters which are then evaluated using the cluster validity indices. The analysis is only done on a random sample of 5,000 records from the data to minimise the time

requirement. The output files are saved in a new sub folder in the same folder as the annual consumption file for the user to review. The number of recommended clusters from each method are the main output from this step as shown in Figure 7-24.

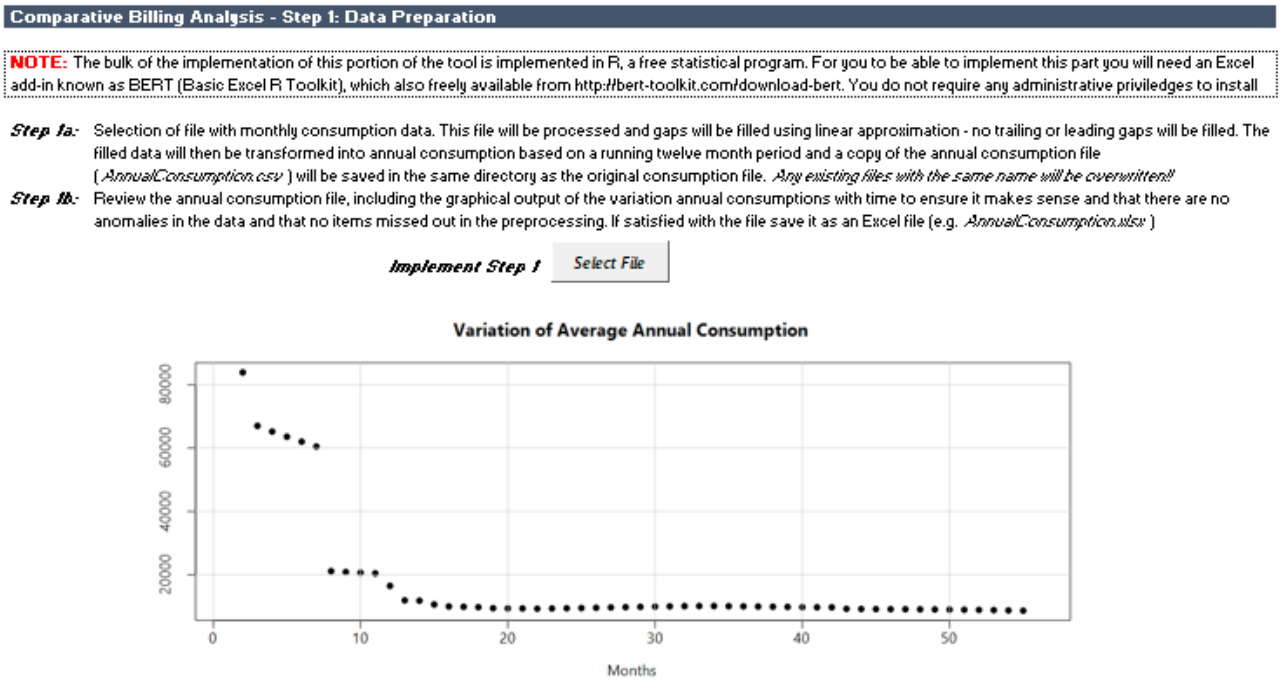


Figure 7-23: Data Preparation

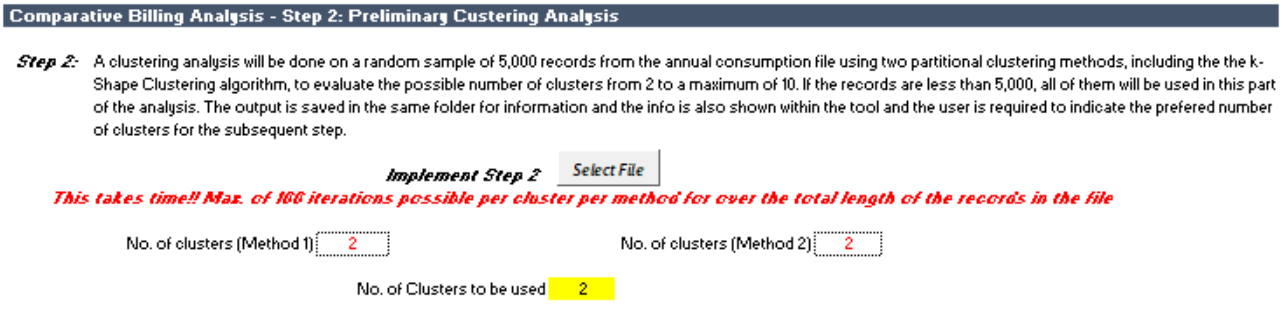


Figure 7-24: Preliminary Clustering

After the preliminary clustering, the next step is the final clustering analysis which is implemented with k-Shape and Fuzzy c-means algorithms using the user-defined number of clusters. The user is required to confirm the number of clusters required for subsequent analysis as opposed to purely accepting the results of the clustering analysis. This is inserted for flexibility but unless there is better information, the number of clusters should not be arbitrarily changed.

The user selects the annual consumption file again to implement the analysis with the typical output from the clustering analysis shown in Figure 7-25 together with a *.csv* file output and a folder with the clustering output in the same folder as the input file. The output file is like the annual consumption, only that all the records have the same starting point in the first month, which is required for the subsequent error degradation analysis.

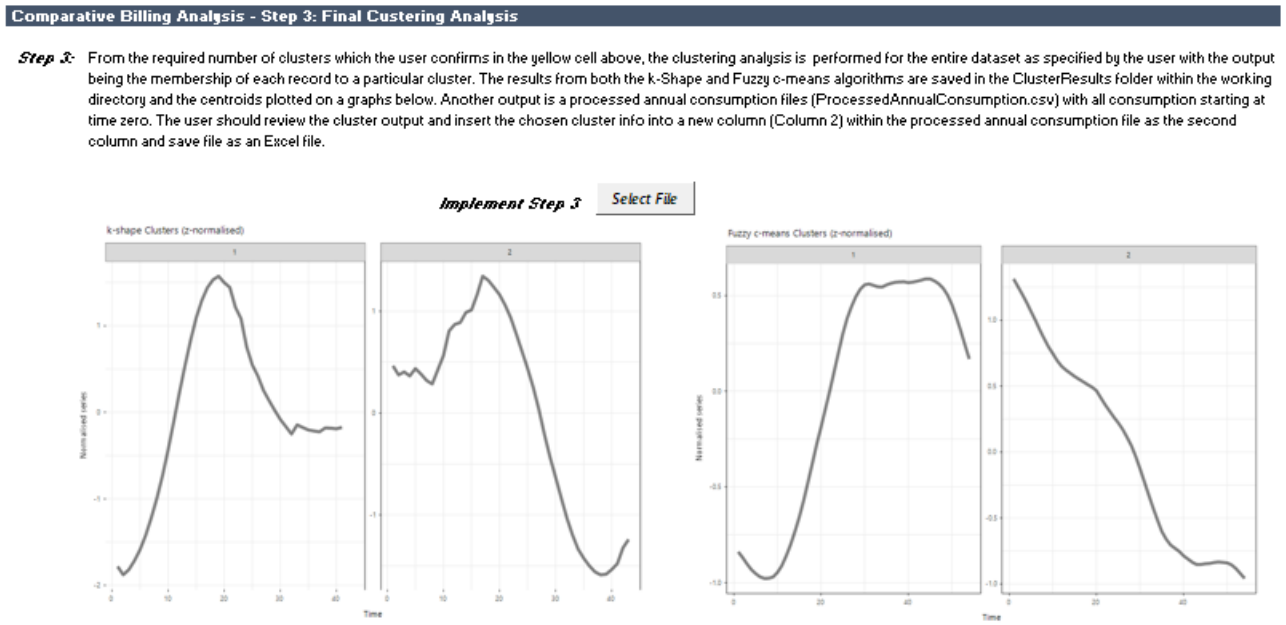


Figure 7-25: Final Clustering Analysis

The final clustering is a bit faster as the CVIs are not run in this part, but the overall speed will also depend on the number of records being analysed. The user must note that the k-Shape algorithm chooses one of the records as its centroid in the calculation and this has disadvantages, the most important being that the shape and length of the centroid determines both the length and shape of the cluster. The fuzzy algorithm however uses the mean of the records as the centroids, which are well defined. It is for this reason, and others, that the use of the results of the fuzzy algorithm are recommended.

When satisfied with the output of the results, the user must then insert the cluster membership details into the processed annual consumption file together with the consumption category (or meter model, or any other variable the user wants to evaluate) and the first reading of each record as shown in Figure 7-26. The final file with the correctly formatted data must then be saved as an Excel file for the error degradation analysis, which is done within that Excel file as well.

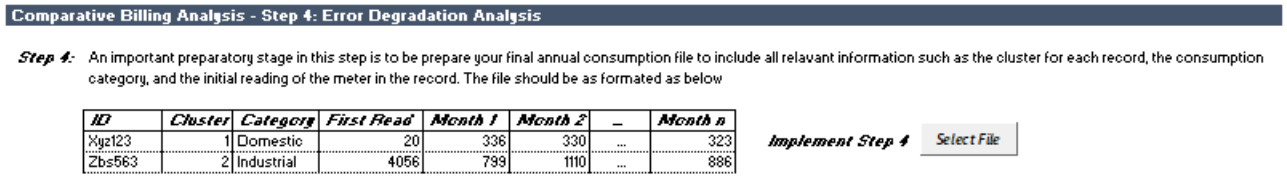


Figure 7-26: Error Degradation Analysis

When satisfied with the file, the user may select it from within the tool as shown in Figure 7-26 to run the analysis. With the analysis complete, the user may open the processed annual consumption file which now has a results tab where the user can input the cluster, the category and age category (new meter and old meter), as shown in Figure 7-27.

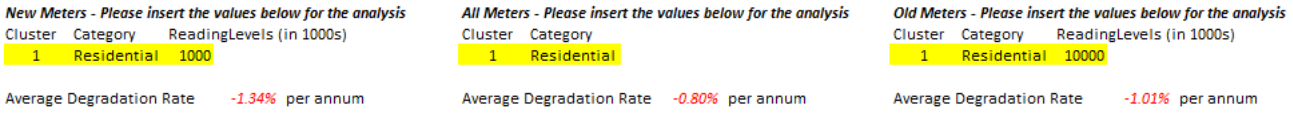


Figure 7-27: Error Degradation Results

The analysis calculates the average consumption for each record, the corresponding quantile to the nearest ten and the category of the readings to the nearest thousand. The rest of the analysis follows that of §6.3.2 which provides the rational of certain choices. The data tab of the processed Excel file contains the raw data and all calculations that may be reviewed by the user.

The degradation rates from the analysis are the ultimate results that may be used in other parts of the analysis and for estimating the level of apparent water losses based on the age of the meters within the utility.

7.5 Meter Management

With the implementation of appropriate consumer and meter information systems that enable and facilitate the auditing of consumers and meters, together with the analysis of data generated in those audits, a utility is better placed to manage its meter fleet more effectively.

Meter management has to do with the field work required to have the meters' measure flow accurately, together with the supporting office and laboratory work that provide the insight into the choice of meters, their validation and installation quality.

7.5.1 Meter Selection

Meter selection is an extension of the determination of the optimum period of replacement for a meter as in §7.4.2, with the only difference being to choose an optimum meter model from comparison of several ones. Figure 7-28 shows the input screen for the analysis, which has fixed parameters and variables specific to each meter model being evaluated.

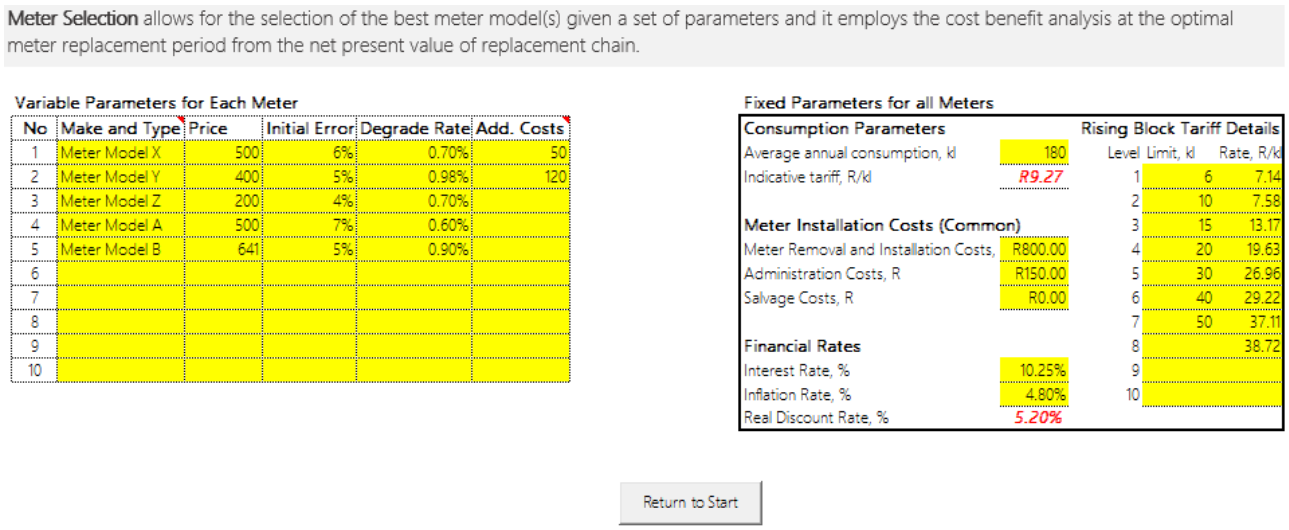


Figure 7-28: Meter Selection Input

When all the input data have been entered for up to ten meters, the analysis, which is now done over a 30-year period, can be implemented as shown in Figure 7-29. The output of the results includes all the meter models, the optimum period to replacement and the associated financials at the determined optimum replacement period. The recommended model is based on maximising the benefit to cost ratio which in this case also coincides with the least percentage of revenue lost.

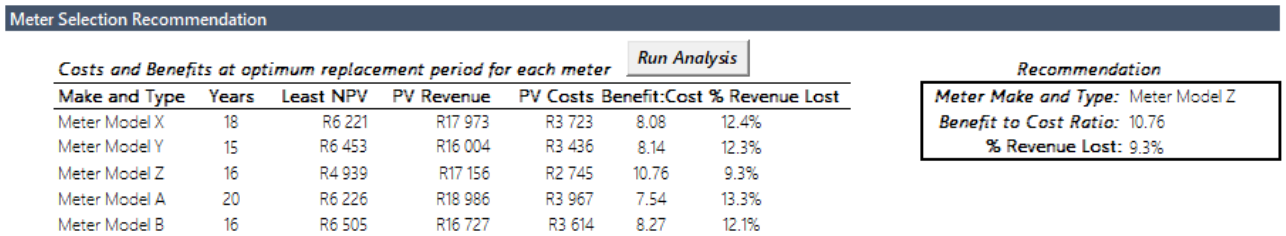


Figure 7-29: Meter Selection Recommendations

Figure 7-29 provides a good example of why a utility may choose to base their replacement not on the age but on the percentage of revenue lost, particularly for consumers with very low consumption

as in this example. For instance, in a water scarce region it does not make sense to wait for 20 years to replace Model A when the total revenue lost at that point would have been 13.3%.

The use of the degradation rate and the initial error still have uncertainties as these values are difficult to evaluate. The user is therefore encouraged to utilise the sensitivity analysis of §7.4.2 to see how this uncertainty affects the viability of the selected optimum meter. A few iterations of the sensitivity analysis may be required should the selected meter overlap with another meter within the meter selection toolset.

7.5.2 Meter Installation

Despite the best efforts in undertaking the work of the prior sections, it is common for utilities to undo any gains by failing to install meters properly or not to timeously undertake the required maintenance work at the required levels of quality workmanship. As precision instruments, meters require adherence to strict installation criteria, typically defined by meter manufacturers and standards such as the length of straight pipeline before and after the meter and the correct alignment. Van Zyl (2011) provided detailed information on the installation requirements for each meter type and gave examples of improper installations, and lists the main installation requirements as follows;

- Correct orientation of both the pipe work and the meter, as specified on the meter needs to be followed (i.e. horizontal or vertical). The accuracy of the meter is adversely affected when it is not in the right orientation. This is particularly true for velocity meters such as multi jet meters while volumetric meters are not susceptible to this. Typically, due to meter reading requirements, meters are installed horizontally to be read facing downwards

- Minimum straight lengths of pipe (with no fittings) directly upstream and downstream of the meter must be provided, as specified by the manufacturer. Where straight pipes sections are not possible, flow straighteners or meter types with no pipe length requirements should be considered. This requirement is to ensure that the flow profile is linear with no turbulence or pulses when it reaches the measuring devices within the meter. A secondary consideration is that when meters are installed in series care must be taken to ensure that the meters do not influence each other.

As an example, the installation of a multi jet meter next to a piston meter will be problematic for the multi jet due to the pulsing effect of the piston meter.

- Installation in the correct direction of flow as typically indicated by an arrow on the meter body. Other than the recording of negative flows, the accuracy of many meters is not the same in either direction of flow.

- Installation of a separate meter strainer upstream of the meter, where required and not coming pre-installed within the meter body. Such strainers should be accessible to facilitate cleaning and replacement.

- Installation of isolating valves, ideally both upstream and downstream of the meter, to be able to carry out any maintenance on the meter and within the consumer's property.

- Meters connected to electricity power supply may require lightning and electrical surge protection particularly in lighting prone areas.

- Plastic bodied meters are required to be protected from sunlight and should therefore be installed in appropriate housings. In areas where brass meter theft is a challenge the use of plastic bodied meters in encouraged but the utility must have some stringent quality to ensure that such meters are adequately protected.

- Appropriate corrosion protection should be considered for all steel pipes used within the installation, particularly if above ground, with suitable UV protection.

Special consideration should also be placed on the associated fittings connecting the meter to the rest of the network as these may compromise the connection and the meter. Utilities must prepare and adopt standards and policies that specify in detail such installation requirements to standardise on the work and quality of their installations. The inclusion of a suitable non-return valve is one such consideration which is important for water safety and to ensure that consumers, particularly those with on-site storage are not charged twice as most meters as not as accurate in reverse flow.

Another aspect of meter installation that requires control is meter size selection. There must be very clear guidelines and control on how meter sizes are selected for an particular consumer, with the use of either consumption profiles, for which a tool is provided in §7.3.1, or the use of modified unit fixtures method as discussed in §2.3.2. Failing to adopt processes and controls for the proper selection and sizing of meters at installation inevitably leads to higher apparent losses.

Meter information is usually generated during the meter installation process and, paying attention to detail at this stage, presents the least-cost method of gathering accurate meter and consumer information for the utility. For utilities with systems that manage the acquisition of material, with meters being one such material, a seamless integration of systems allows the selection of a specific meter with all its details during the work/job management and installation. In such a case, only the installation details need to be added which reduces the work load and the chances of errors.

Quality control of all installation is another important requirement and will be discussed further in §7.5.4.

7.5.3 Meter Maintenance

Water meter maintenance requires simple but important actions, such as cleaning of strainers, cleaning and repair of meter boxes, fixing leaks and replacing damaged registers and register covers, with larger meters requiring visual inspection of the meter body and chamber for visible damage (van Zyl, 2011). In most of the cases common sense and the proper management and maintenance of water pipe networks is also beneficial to water meters. Examples include proper scouring of the pipelines after a repair to ensure that than little to no suspended solids get trapped in the strainers of meters; the proper (slow) opening of valves and use of air valves after service interruptions minimises the flow of high velocity air through the meters which can damage them; and the maintenance of adequate flow rates within the water networks minimises the settling of suspended matter within both the networks and the meters leading to clogged strainers.

What is important is that in the implementation of maintenance, meter seals should not be disturbed and if broken should be reinstated by a qualified verification officer after confirming the adequacy and accuracy of the meter.

Meter replacement is a key component of meter maintenance whereby all meters that have gone past their useful life, and most appropriately are losing more money for the utility than the cost of replacing them, should be promptly replaced. This is where some utilities fall short and do not respond as promptly as they should. The use of the annual average daily demand (AADD) in prioritising all meter replacement work is an appropriate strategy that will protect the revenue of a utility. Based on the use of tools such as §7.4.2, a utility should also have plans for the planned replacement of its meter fleet on an ongoing basis.

7.5.4 Quality Control and Assurance

Quality control and assurance provides the much-needed confirmation and assurance whether what is claimed by the meter manufacturers, meter installation teams, meter readers and various other staff involved in the meter management process, is as it should be. It seeks to answer questions such as;

- Are meters supplied meeting the defined specifications and requirements?
- Are meters installed to specification and is the information provided correct?
- Do the systems data reflect reality?
- Do the right consumers have the right meters?
- Can the verification processes and the preservation of custody transfer data withstand legal scrutiny?
- Are the right meters being changed at the right time?

A utility that can answer in the affirmative to all these questions would have done a commendable job of managing its water meters and would be having rigorous and functional quality control systems

in place. As can be observed from the questions, quality control is just not about the meters being delivered from a supplier, but it looks at the entire process.

7.5.4.1 Meter Acceptance Sampling

Arregui, Cabrera Jr., *et al.* (2006) provides a detailed quality control procedure for the acceptance sampling and testing of meters from manufacturers, evaluating the compliancy of each batch of meters. The procedure is based on the ISO 3951 standard that deals with sampling procedures for inspection by variables. The process involves the selection of the acceptable level of quality and the level of inspection required. The third parameter is specification limits for the meters, which are already predefined for water meters. The output of the procedure determines whether a batch of meters is acceptable based on the predetermined variables.

While the reader is referred to the ISO standard for greater detail on acceptance sampling, what is critical is that utilities must adopt some form of processes based on sampling techniques that will assist them in evaluating the adequacy of all the meters being supplied to the utility. This is the only way of being certain that all incoming meters meet the prescribed requirements.

Beyond acceptance sampling, a bit of common sense will also assist utilities in avoiding costly mistakes. The simple checking of the details of the meters being supplied, without any testing, will eliminate the installation of wrong meters such as the Class A meter in Figure 3-5(c).

7.5.4.2 Meter Installation

Section 7.5.2 demonstrated the exacting requirements for meter installations, with the slightest of variation affecting the accuracy of the meter. Poor plumbing workmanship is a reality in many areas and is the major cause of on-site leakage (Lobanga et al., 2008). There is therefore a need to institute quality control/assurance processes for the evaluating of the correctness of meter installations as they happen throughout the utility area.

Figure 7-30 illustrates a process developed for Johannesburg Water to introduce system generated work orders (W/O) from completed and in-progress metering jobs to be evaluated by the metering

foreman daily. To be able to implement this appropriately, a new works order form was also designed, an extract of which is shown in Figure 7-31. The process compels the foreman to inspect a minimum of 10% of the work done by plumbing teams and not only verify the captured data but the installation details in terms of alignment, fittings, corrosion protection and other reinstatement work required.

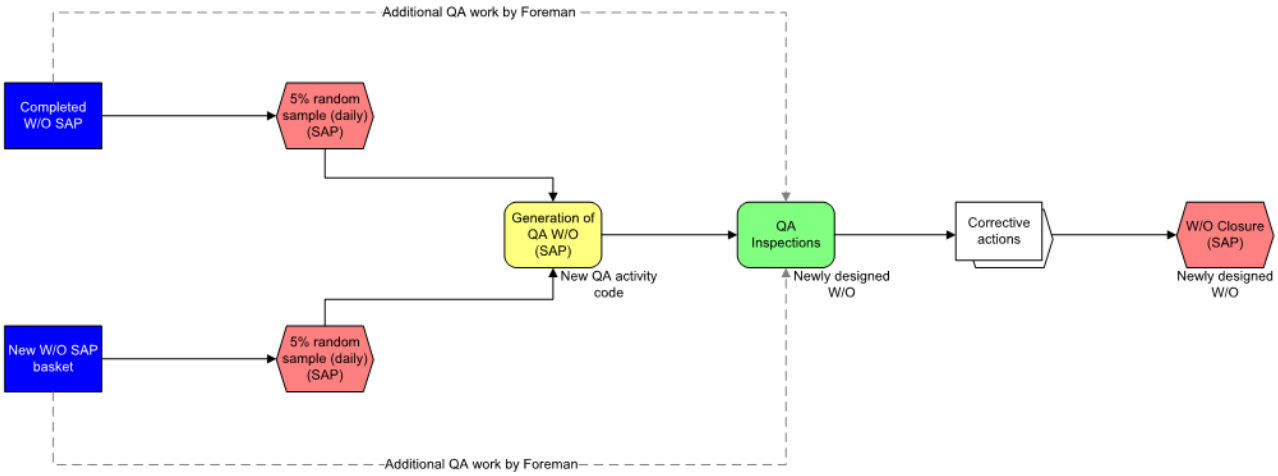


Figure 7-30: Meter Quality Assurance Process

Visibility and supervision of meter installation minimises the occurrence of incorrect installation and other related information such as the data capturing of the completed work. The fact that proper quality assurance is based on random samples also assists the utility, if properly designed, in ensuring that all possible types of work gets to be checked and verified by usually a more experienced and senior person than the installer.

7.5.4.3 Meter Replacement

The quality assurance for the installation part of meter replacements is the same as described in §7.5.4.2. The additional work required is to be able to verify if the underlying assumptions that led to meter replacement were in fact valid and that the envisaged gains have been achieved. Examples of such evaluation is described in §5.2 in evaluating meter change data. The results of such analysis will provide valuable insight to the utility in how it can improve its processes in as far as meter replacement is concerned.

QA job type:	In-progress		Partially completed	[Specify outstanding work]	Complete	

SAFETY PRECAUTIONS TAKEN												
Cones		Barricading		Road Signs		Shoring		Correct PPE		Other	[Specify]	

CAPTURED/OLD METER DETAILS				OBSERVED/NEW METER DETAILS (MAIN METER)			
Make	[Prepopulated]	Type	[Prepopulated]	Make	e.g. Kent	Type	e.g. Single Jet
Number	[Prepopulated]	Class	[Prepopulated]	Number		Class	B C D
Size	[Prepopulated]mm	Factor	[Prepopulated]	Size	mm	Factor	
Reading	[Prepopulated]	Digits	[Prepopulated]	Reading		Digits	
Position	[Prepopulated]			Position			

Combination Meter (Smaller Meter Details)		Type	e.g. Single Jet	Number	
Size	mm	Class		Factor	

Digits		Reading	

WORK CARRIED OUT						
Correct fittings used?	Yes		No		Type of fittings	[Specify]
Correct installation, alignment and depth?	Yes		No		Comments	[Specify]
Corrosion prevention?	Yes		No		Type of protection	[Specify]

Service reinstated?	Yes		No		Comments	[Specify]

Reinstatement	Paving		Garden		Road		Rubble removal		Other	[Specify]

Figure 7-31: Job Card Extract - Meter QA

7.6 Summary

This chapter introduced a simplified meter management framework comprising four main categories of Systems, Audits, Analysis and Metering, abbreviated as SAAM, to help simplify the main requirements for improved assessment and management of apparent water losses, with meters being its central component. Systems comprise meter and consumer information systems that house all pertinent information related to the consumers and meters within a utility and thus forming the backbone of all other activities. Audits facilitate the collection of consumption profiles and meter accuracy data that enable the utility to know its consumers and meters better and thereby be better placed to select appropriate meters for each consumer segment and grouping. The Analysis component is about the evaluation and analysis of all data information gathered in all the components to derive critical parameters such as the optimum replacement period of meter replacement, estimates of apparent losses and the degradation of meter accuracy. Finally, the metering component relates to the select-

ing the correct meters, installing and maintaining them appropriately with the necessary quality assurance activities. This simpler and consolidated framework is seen as an initial step in facilitating better understanding of managing apparent losses.

Based on the SAAM meter management framework, an Apparent Loss Management Tool (ALMAT) was developed and introduced in this chapter. The tool was developed within the widely known and used Microsoft Excel software using Visual Basic for Applications (VBA) and some integration with the R Statistical Software through the Basic Excel R Toolkit (BERT). The tool facilitates meter audit analysis with meter sizing guidelines and on-site leakage assessment; deriving of consumption profiles and on-site leakage assessment for groupings of consumers; meter error analysis from meter testing data; determination of the optimum meter replacement period and associated cost benefit analysis; assessment of the apparent losses and meter accuracy degradation analysis using the weighted accuracy method and the comparative billing analysis method developed in this study; and the selection of an optimum meter from a number of alternatives. This is a first known tool that provides for the consolidated assessment of parameters essential for the management of apparent losses.

Together with the algorithms provided with ALMAT, the chapter provides guidance on how the tool can be used to better assess and manage apparent water losses with additional recommendations and examples for each of the activities within the SAAM framework. This includes the minimum requirements for meter and information management databases, recommendation on suitable meter testing programs and test flow rates, minimum requirements for meter installation and maintenance and quality assurance requirements for improved meter management. The consolidated tool and recommendations provide essential decision enhancement to the management of apparent water losses, particularly in the developing world context.

Chapter 8 Conclusion and Recommendations

This chapter summarises the main findings and conclusions from this research, together with a review of how well the stated objectives have been achieved. Key contributions of this research to the body of knowledge on the assessment and management of apparent water losses are also summarised. Lastly, several gaps and further research needs are also summarised with recommendations for future research.

8.1 The Context

In Chapter 1 it was shown that there will be a projected 40% gap between demand and supply in 2030 (UNEP, 2011) if water use efficiency is not improved. This requirement is captured in Goal 6.4 of the Sustainable Development Goals which implores all United Nations member states to substantially increasing water-use efficiency in all sectors by 2030 (United Nations, 2016). However, non-revenue water (NRW) in various jurisdictions remains high such as 25% in London, 32% in Norway (WWAP, 2016) and 41% in South Africa (Department of Water and Sanitation, 2017). In addressing NRW, most of the work, both in practice and research, has focused on the real loss component of water losses (Johnson & Vermersch, 2011).

In addressing apparent water loss, particularly metering errors which are the leading cause, it has been shown that the traditional method of assessment is costly and very laborious and viable alternatives are required. In addition, while a few tools exist for managing meters and apparent losses, the majority focus on only specific aspects without integrating the assessment of the various component of apparent water losses. There is therefore a need for tools that facilitate the integrated assessment and management of apparent water losses within the context of the developing world.

Chapter 2 provided a detailed review of the literature and application of apparent loss, together with the relevant context on how it fits within the greater water loss management paradigm. The bulk of current research is largely based on the collecting of field observations and laboratory measurements while other sectors have increasingly been adopting data-driven discovery processes, such as data mining techniques, which presents an opportunity and possibility that data mining of available meter and billing information could reveal insightful trends that could be used for apparent loss management

8.2 Achieved Results

8.2.1 Consumption Characterisation

Chapter 3 details the consumption characterisation of 408 properties within the City of Johannesburg comprising residential, multi-residential, business and public benefit properties and an additional 256 properties from various areas within South Africa. Residential consumers have up to 37% (considering uncertainties) of flow consumption occurring at less than 0.03m^3/hr, indicating a substantial proportion of low flow rates. While the other consumer categories do not have correspondingly significant low flows, all consumer groups have less than 0.25% of flow above 2.5m^3/hr indicating that meters with a q_p of 1.5m^3/hr are generally applicable for these types of consumers. The typical consumption profile of the residential consumers was also represented based on the average annual daily demand (AADD) of the individual consumer with clear indications that the use of AADD in consumption profiling can also enhance the process, particularly for residential properties, and may have additional benefits from a meter sizing guideline perspective.

On-site leakage was also evaluated and there was a 63% occurrence of leakage within Johannesburg with the estimated monthly leakage ranging from 11 – 41m^3/month, which is relatively high and warrants further attention particularly for the multi-residential and business properties.

Consumption characterisation contributed towards Objective 1 of quantifying apparent losses due to metering errors using the weighted meter accuracy methodology as a key input into the method.

8.2.2 Meter Accuracy Evaluation

Chapter 4 details the evaluation of historical meter testing that shows that the starting flowrate has the highest degradation rate of up to 0.25% per every 1,000kℓ passing through the meter and 0.32% per year while the permanent and overload flowrates degrade at almost half the rate of the starting flowrate. The error at the transition flow rate is positive and degrades at lower levels of about 0.06% per 1,000kℓ. In the absence of elaborate and carefully designed and implemented projects to determine metering errors at a flow lab, utilities can make use of such data in the preliminary determination of the metering error. The probability of failure concept is also introduced from the historical testing data and may be used in making decision on meter testing requirements and its acceptable level for meters used for billing purposes.

Detailed meter testing was successfully done for 123 small bore meters out of 210 meters taken to the laboratory from an initially planned sample of 394 meters. The testing was done at 10 test flowrates to properly represent the error curve. The value of having more test flowrates has been demonstrated in this study by observing the variations of accuracy at low flowrates and the need to add smaller flowrates for 20mm and 25mm meters to better represent the error curve at these low flowrates. It was also demonstrated that despite good intentions and design, the testing of meters from the field, other than the cost involved, is fraught with many difficulties highlighting the need to explore viable alternatives of evaluating metering errors.

An alternative equation that better represents the weighted accuracy of meters, particularly at low flowrates, is also introduced and used in determining the apparent losses. The modification allows for the use of the midpoint accuracy of a flow range as opposed to using the highest flow range accuracy to represent the entire flow range. The requirement for the modification is necessary in instances where the meter error curve is not adequately represented due to the selection of a few meter test points.

The meter error evaluation also contributed towards Objective 1 of quantifying apparent losses due to metering errors using the weighted meter accuracy methodology as a key input into the method.

8.2.3 Apparent Loss Assessment and Method Comparison

Chapter 5 evaluated the apparent losses due to metering errors using three different methodologies. Using the weighted meter accuracy method, the errors were estimated to be, on average, about 12% of the metered consumption, which is higher than the currently used default value of 10% which incidentally incorrectly refers to the proportion of water losses. The actual estimated range of the apparent losses is between 9.4% to 14.6% depending on the ratio of 15mm and 20mm meters used. Using the comparative billing analysis method, apparent losses were found to be 14%, similar with the worst-case scenario of the weighted accuracy method of 14.6%. The meter change analysis understated the best-case scenario of apparent losses of 9.4% by about 50% but was in line with previous studies using similar methodology.

While the comparative billing method was within range of the weighted accuracy method, it did not take into consideration the initial error of the meter and therefore the actual value is higher. The different variations of the analysis that provide different results also showed that it required refinement and improvement. The meter change method was susceptible to variation depending on how much data was analysed, particularly if it is not labelled data. Further analysis showed that the Johannesburg Water meter change database and processes introduced variables into the analysis that made the output unreliable. The method may therefore be only suitable for utilities with very good data, which is typically not the case in the developing world. However, the evaluation of meter change data provides a lot of useful insights as demonstrated herein and should be used by utilities to influence how meter changes are implemented and to evaluate the efficiency of such programmes.

The suitability of the reference annual apparent losses (RAAL) default value of 5% was also evaluated using the examples of data produced in this study and that from literature, also with due regard for the legal requirements for meters in South Africa. The results show the current default value may not be adequate for many utilities. A proposal is also made to consider the local meter regulations and the consumption profiles and base the RAAL on the best available estimate of a new meter's accuracy requirements in a locale with some margin that allows for degradation after a few years.

The comparative assessment of apparent losses addresses Objective 1 of quantifying apparent losses due to metering errors using the weighted meter accuracy methodology and Objective 2 related to the evaluation of viable alternative methodologies in estimating apparent losses due to metering errors.

8.2.4 Development of an Alternative Apparent Loss Assessment Method

Chapter 6 introduced the improved comparative billing methodology which was developed using a combination of data preparation techniques, clustering analysis and classical regression analysis. Apparent losses due to metering error for the residential category were estimated to be 8.1% of billed consumption compared to 11.2% found using the field-based method, using an error degradation rate of 0.7% per annum. The difference in the values are attributed to meter under-registration which is not accounted for with the new developed method and can be higher than 4% for new meters. About 15% of the available data was used in development of the method, which was a significant improvement from previous attempts of related work. Average results are even closer with only a difference of 1.2% which is also better that the previous estimates of the unimproved method.

A general equation has also been developed to provide a simple estimation of apparent losses due to metering errors in utilities with similar conditions and environmental factors as Johannesburg Water.

Objective 3 and 4 of the research of developing and validating a data mining based methodology for the assessment of apparent water losses were met in this component.

8.2.5 Apparent Loss Management Tool

Chapter 7 introduced a simplified meter management framework comprising of four main categories of Systems, Audits, Analysis and Metering, abbreviated as SAAM, to help simplify the main requirements for improved assessment and management of apparent water losses, with meters being a central part of that. Based on the SAAM meter management framework, an Apparent Loss Management Tool (ALMAT) was also developed. The tool facilitates meter audit analysis with meter sizing

guidelines and on-site leakage assessment; deriving of consumption profiles and on-site leakage assessment for groupings of consumers; meter error analysis from meter testing data; determination of the optimum meter replacement period and associated cost benefit analysis; assessment of the apparent losses and meter accuracy degradation analysis using the weighted accuracy method and the comparative billing analysis method developed in this study; and the selection of an optimum meter from a number of alternatives.

The apparent loss management tool addressed Objectives 5 and 6 of developing a meter management framework and an integrated decision enhancement tool for the management of apparent water losses.

8.3 Contribution to Knowledge

The following are some of the main contributions of this study;

1. The determination of consumption profiles for a representative sample of consumers, mostly from the metropolitan area of Johannesburg, South Africa, from a total of 664 properties. This is the largest known dataset of consumption profiles within Sub-Saharan Africa and the developing world and is only bettered by America studies. The only known study in developing areas evaluated only 78 properties. This research has therefore substantially increased the depth and breadth of information with respect to the consumption profiles of consumers within a developing nation perspective.

2. Detailed assessment of on-site leakage based on the minimum night flow analysis of 408 properties with continuous measurements over an average of 8.4 days. This included the breakdown of the leakage in terms of its flow distribution which has marked impact on apparent water losses. The properties surveyed were a representative sample of the City of Johannesburg, with meters of various ages and readings. Prior studies within the same area were based on at least two subsequent meter reading over a 5-minute period and repeated two or three times on different days during the day for 182 properties with meters less than 5 years old. The assessment in this

study is therefore more in-depth, more representative and broader, with pertinent information on on-site leakage as it influences apparent water loss.

3. The useful use of historical meter testing data from other purposes other than determination of apparent water losses has been demonstrated, with the introduction of the probability of failure of meters as an important variable to determine and apply for better meter management. This is the first known study where such a concept has been introduced and implemented.

4. The determination of meter accuracy from the laboratory testing of 123 meters at 10 flow rates for meters commonly used within South Africa, which has not been recorded in literature before. Within a developing nation context, this is the only known study where meter testing has been implemented over a wide range of flow rates compared to three flow rates of another African study using 250 meters. It has also been shown why the testing of meters at three flow rates is inadequate for the assessment of apparent water losses, particularly considering the very high error degradation rate at low flows.

5. An improved and unambiguous equation that better represents the weighted accuracy of meters, particularly at low flowrates, had been developed and applied in the determination of the apparent losses. The requirement is particularly necessary in instances where the meter error curve is not adequately represented due to the selection of a few meter test points.
6. A means of testing the applicability of the reference annual apparent losses (RAAL) using the study data and literature, including regulatory prescriptions and the consumption profile of a specific area has been introduced and applied for the study area with results showing that the default value of 5% is not always applicable.

7. A comparative analysis of three different methodologies of calculating apparent losses has been done for the same study area, with an attempt to validate the two alternative methods against the weighted meter accuracy method. This is the first known study that has attempted to do that, with lessons that are applicable to the wider sector on how the alternative methods can and cannot be used.

8. The development and application of a novel comparative billing methodology that incorporates data mining algorithms of time series clustering together with regression analysis as an alternative method for the assessment of meter accuracy degradation and apparent water losses. The method has been successfully validated against the improved traditional weighted meter accuracy methodology for the study area. This is the first know attempt of an alternative method for assessing apparent losses that uses data mining tools and was also validated with field data.

9. The development of a meter management framework and an accompanying apparent loss management tool that offers the enhanced and integrated management and assessment of various components of apparent water losses due to metering errors. The management tool includes all the above contributions within one user-friendly environment and is therefore also novel.

8.4 Recommendations and Future Research Needs

This study has attempted to answer some important research questions related to the assessment and management of apparent water losses, together with the development and enhancement of methodologies for addressing the same. However, there were number of gaps that the research was unable to fill and other areas that require additional research. These include;

- The assessment of consumption profiles was based on a sampling rate of one minute and may have averaged the actual consumption pattern at low flows. Further evaluation at much lower sampling rates is recommended.

- The degradation of meter accuracy with both time and meter reading could not be adequately defined due to the small numbers of meters that were eventually tested per meter model and this will be important to improve the knowledge base for meters locally used in South Africa and the region.

- The starting error of the various meter models is another key factor that will benefit for quantification to better aid meter selection and determination of the optimum period of meter replacement. This was not evaluated in this study and is recommended.

- The weighted error of new meters was also not evaluated in this study and will further enhance a variety of activities related to the management and assessment of apparent water losses.

- Only apparent losses from consumers with small meters were evaluated in this study, although some consumption patterns were derived for bulk consumers. It would therefore be important to extend the study to cover consumers with larger meters particularly of multi-residential category which are common and share many similar characteristics.

- The dataset used throughout the study had many imperfections that led to a lot of data cleansing and the failure to further drill down for each of the analyses that were performed. It would be important to perform the similar analysis with different datasets to further validate the methods developed within this study and look at how different metering technologies, models and consumer segment can be proactively identified using data mining technologies.

- Other components of apparent water losses, such as unauthorised consumption, have not been covered in this study. It is however an ideal candidate for fraud analysis and anomaly detection, together with data acquisition errors, that needs to be explored and added on to the developed methodology thereby further enhancing the use of data mining applications for water loss assessment.

- The meters in this study all receive water from the same water source and provider and therefore the relationships and values derived are very specific. Extending similar work to other locales and investigating how the developed methodology can be used to cater for other variables such as water quality would be useful.

- The part of the study on the weighted meter accuracy method focused on consumers with small connections with limited work on bulk metered consumers. While the developed comparative billing method and the management too can be applied for bulk consumers, no verification was done for this category. Furthering the study to focus on bulk consumers is therefore recommended.

- The developed framework and management tool can also be further enhanced and disseminated to utilities and municipality in South Africa and other regions of the developing world.

- This study contributed to the body of knowledge in apparent water loss with is relatively very small compared to real loss assessment. Additional research to build up this knowledge base would serve the water industry well.

References

Abrahart, R.J., See, L.M. & Solomatine, D.P. Eds. 2008. *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications*. Springer-Verlag Berlin Heidelberg.

Adachi, S., Takahashi, S., Kurisu, H. & Tadokoro, H. 2014. Estimating Area Leakage in Water Networks Based on Hydraulic Model and Asset Information. *Procedia Engineering*. 89:278–285. DOI: 10.1016/j.proeng.2014.11.188.

Aghabozorgi, S., Shirkhorshidi, A.S. & Wah, T.Y. 2015. Time-series clustering - A decade review. *Information Systems*. 53(May):16–38. DOI: 10.1016/j.is.2015.04.007.

Ajbar, A.H. & Ali, E. 2012. Water Demand Prediction for Touristic Mecca City in Saudi Arabia using Neural Networks. *World Academy of Science, Engineering and Technology*. 65:36–40.

Al-Washali, T.M. 2011. Non- - Revenue Water Management in Sana'a Water Distribution System. University of Cologne for Applied Sciences.

AL-Washali, T., Sharma, S. & Kennedy, M. 2016. Methods of Assessment of Water Losses in Water Supply Systems: a Review. *Water Resources Management*. 30(14):4985–5001. DOI: 10.1007/s11269-016-1503-7.

Algotsson, E., Murombo, T., Davis, M. & Poole, M. 2009. *Water Supply and Sanitation in South Africa: Environmental Rights and Municipal Accountability*. Available: https://cer.org.za/wp-content/uploads/2011/11/LHR-DBSA_Water_Report.pdf [2018, January 06].

Allender, H. 1996. Determining the economical optimum life of residential water meters. *Water Engineering & Management*. 143(9):20–24.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., P??rez, J.M. & Perona, I. 2013. An extensive comparative study of cluster validity indices. *Pattern Recognition*. 46(1):243–256. DOI: 10.1016/j.patcog.2012.07.021.

Arregui, F., Cabrera, E., Cobacho, R. & Palau, V. 2003. Management strategies for optimum meter selection and replacement. *Water Science & Technology: Water Supply*. 3(1/2):143–152.

Arregui, F., Cabrera, E., Cobacho, R. & García-Serra, J. 2006a. Reducing Apparent Losses Caused By Meters Inaccuracies. *Water Practice and Technology*. 1(4):1–8. DOI: 10.2166/wpt.2006.093.

Arregui, F., Gavara, F.J., Soriano, J. & Cobacho, R. 2014. Analysis of domestic water meters field performance. In *Proceedings of the Water Loss 2014 Conference*. Vienna, Austria. 1–8.

Arregui, F., Gavara, F., Soriano, J. & Pastor-Jabaloyes, L. 2018. Performance Analysis of Ageing Single-Jet Water Meters for Measuring Residential Water Consumption. *Water*. 10(5):612. DOI: 10.3390/w10050612.

Arregui, F.J., Cabrera Jr., E. & Cobacho, R. 2006. *Integrated Water Meter Management*. IWA Publishing.

Arregui, F.J., Cabrera, E., Cobacho, R. & García-Serra, J. 2006b. Reducing Apparent Losses Caused By Meters Inaccuracies. *Water Practice and Technology*. 1(4):1–8. DOI: 10.2166/WPT.2006093.

Arregui, F.J., Martinez, B., Soriano, J. & Parra, J.C. 2009. Tools for Improving Decision Making in Water Meter Management. *Proceedings of the 5th IWA Water Loss Reduction Specialist Conference*. (April):225–232.

Arregui, F.J., Cobacho, R., Cabrera Jr., E. & Espert, V. 2010. Graphical method to calculate the optimum replacement period for water meters. *Journal of Water Resources Planning and Management*. (February):143–146.

Arregui, F.J., Soriano, J. & Gavara, F.J. 2012. An integrated approach to large customer water meters management. In *IWA Water Loss 2012*. International Water Association (IWA). Available: http://www.iwa-waterloss.org/2012/Final_Papers_1/31.pdf [2013, August 27].

Arregui, F.J., Soriano, J., Cabrera, E. & Cobacho, R. 2012. Nine steps towards a better water meter management. *Water Science and Technology*. 65(7):1273–80. DOI: 10.2166/wst.2012.009.

Arregui, F.J., Soriano, J., García-Serra, J. & Cobacho, R. 2013. Proposal of a systematic methodology to estimate apparent losses due to water meter inaccuracies. *Water Science & Technology: Water Supply*. 13(5):1324. DOI: 10.2166/ws.2013.138.

Arregui, F.J., Balaguer, M. & Soriano, J. 2017. Quantifying measuring errors of new residential water meters considering different customer consumption patterns. *Urban Water Journal*. 9006(October):1–13. DOI: 10.1080/1573062X.2014.993999.

AWWA. 1966. Determination of Economic Period for Water Meter Replacement. *Journal - American Water Works Association*. 58(6):642–650. Available: http://www.jstor.org/stable/41264779.

AWWA. 2004. *Sizing Water Service Lines and Meters: Manual of Water Supply Practices - M22*. 2nd ed. American Water Works Association.

AWWA. 2009. *Water Audits and Loss Control Programs - M36*. Third ed. American Water Works Association.

AWWA Water Loss Committee. 2003. Committee Report: Applying worldwide Best Management Practices in water loss control. *Journal AWWA*. 95(8):65–79. Available: www.cuwcc.org/WorkArea/downloadasset.aspx?id=8818.

Babovic, V. 2005. Data mining in hydrology. *Hydrological Processes*. 19(7):1511–1515. DOI: 10.1002/hyp.5862.

Babovic, V., Drécourt, J.-P., Keijzer, M. & Friss Hansen, P. 2002. A data mining approach to modelling of water supply assets. *Urban Water*. 4(4):401–414. DOI: 10.1016/S1462-0758(02)00034-1.

Bakırlı, G., Birant, D. & Mutlu, E. 2012. Data Mining Solutions for Local Municipalities. *… on e-Government*. 10(2):97–106.

Barfuss, S.L., Johnson, M.C. & Neilsen, M.A. 2011. *Accuracy of In-Service Water Meters at Low and High Flow Rates*. Denver, CO, USA: Water Research Foundation.

Beal, C. & Stewart, R.A. 2011. *South East Queensland Residential End Use Study: Final Report*. Available: http://www98.griffith.edu.au/dspace/handle/10072/46802.

Begum, N., Ulanova, L., Wang, J. & Keogh, E. 2015. Accelerating Dynamic Time Warping Clustering with a Novel Admissible Pruning Strategy. In *Proceedings of the 21th ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining - KDD '15. 49–58. DOI: 10.1145/2783258.2783286.

Bezdek, J.C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer, Boston, MA. DOI: https://doi.org/10.1007/978-1-4757-0450-1.

Bidgoli, A.M. 2009. Water losses reduction programme in Iran. In *Proceedings of international workshop on drinking water loss reduction: Developing capacity for applying solutions*. R. Ardakanian & J.L. Martin-Bordes, Eds. UN Campus, Bonn, Germany: UN-Water Decade Programme on Capacity Development (UNW-DPC). 80–87.

Bierman, H. & Smidt, S. 2007. *The capital budgeting decision : economic analysis of investment projects*. Routledge. Available: https://www.routledge.com/The-Capital-Budgeting-Decision-Ninth-Edition-Economic-Analysis-of-Investment/Jr-Smidt/p/book/9780415400046 [2018, January 12].

Bolton, R.J. & Hand, D.J. 2002. Statistical Fraud Detection: A Review. *Statistical Science*. 17(3):235–255. Available: http://0-projecteuclid.org.innopac.wits.ac.za/euclid.ss/1042727940 [2014, January 14].

Bowen, P.T., Harp, J.F., Hendricks, J.E. & Shoeleh, M. 1991. *Evaluating residential water meter performance*. Denver, CO, USA.

Bowen, P.T., Harp, J.F., Baxter, J.W. & Shull, R.D. 1993. *Residential Water Use Patterns*. Denver, CO, USA: American Water Works Association Research Foundation.

Boyle, C.E., Eskaf, S., Tiger, M.W. & Hughes, J.A. 2011. Mining water billing data to inform policy and communication strategies. *Journal AWWA*. 103(November):45–58.

Browne, A., Medd, W. & Anderson, B. 2013. Developing Novel Approaches to Tracking Domestic Water Demand Under Uncertainty—A Reflection on the "Up Scaling" of Social Science Approaches in the United Kingdom. *Water Resources Management*. (August):1–38. Available: http://link.springer.com/article/10.1007/s11269-012-0117-y [2013, August 27].

Cabral, J.E., Pinto, J.O.P., Martins, E.M. & Pinto, A.M. a. C. 2008. Fraud detection in high voltage electricity consumers using data mining. *2008 IEEE/PES Transmission and Distribution Conference and Exposition*. (April):1–5. DOI: 10.1109/TDC.2008.4517232.

Cardell-Oliver, R., Wang, J. & Gigney, H. 2016. Smart Meter Analytics to Pinpoint Opportunities for Reducing Household Water Use. *Journal of Water Resources Planning and Management*. 142(6):1–9. DOI: 10.1061/(ASCE)WR.1943-5452.0000634.

Chauhan, A. & Rajvanshi, S. 2013. Non-Technical Losses in Power System: A Review. In *Proceedings of the 2013 International Conference on Power, Energy and Control (ICPEC)*. 558–561. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6527720 [2014, January 14].

Chen, C.-H., Chiang, R.-D., Wu, T.-F. & Chu, H.-C. 2013. A combined mining-based framework for predicting telecommunications customer payment behaviors. *Expert Systems with Applications*. 40(16):6561–6569. DOI: 10.1016/j.eswa.2013.06.001.

Cheng, B. & Titterington, D.M. 1994. Neural Networks: A Review from a Statistical Perspective. *Statistical Science*. 9(1):2–30.

Chou, J.-S., Hsu, Y.-C. & Lin, L.-T. 2014. Smart meter monitoring and data mining techniques for predicting refrigeration system performance. *Expert Systems with Applications*. 41(5):2144–2156. DOI: 10.1016/j.eswa.2013.09.013.

Couvelis, F. 2012. Apparent water loss due to consumer meter inaccuracies in selected areas of

South Africa. University of Cape Town.

Couvelis, F. a & van Zyl, J.E. 2015. Apparent losses due to domestic water meter under-registration in South Africa. *Water SA*. 41(5):698–704. DOI: http://dx.doi.org/10.4314/wsa.v41i5.13.

Criminisi, A., Fontanazza, C.M., Freni, G. & La Loggia, G. 2009. Evaluation of the apparent losses caused by water meter under-registration in intermittent water supply. *Water Science and Technology*. 60(9):2373–82. DOI: 10.2166/wst.2009.423.

Darsana, P. & Varija, K. 2018. Leakage Detection Studies for Water Supply Systems—A Review. Springer, Singapore. 141–150. DOI: 10.1007/978-981-10-5711-3_10.

Davis, S. 2005. Residential water meter replacement economics. In *IWA Leakage Conference, September 2005*. International Water Association (IWA). 1–10. Available: http://rash.apanela.com/tf/leakage/Residential Water Meter Replacement Economics.pdf [2013, August 27].

DeOreo, W.B., Mayer, P., Dziegielewski, B. & Kiefe, J. 2016. *Residential End Uses of Water, Version 2: Executive Summary*. Available: http://www.waterrf.org/Pages/Projects.aspx?PID=4309.

Department of Water Affairs. 2013. *National Water Resource Strategy: Water for an Equitable and Sustainable Future - Second Edition*. Pretoria, South Africa. Available: http://www.dwaf.gov.za/nwrs/LinkClick.aspx?fileticket=u_qFQycClbI%3D&tabid=91&mid=496.

Department of Water and Sanitation. 2017. *Benchmark of Water Losses, Water Use Eficiency and Non Revenue Water in South African Municipalities (2004/05 - 2015/16)*. Pretoria, South Africa.

Depuru, S.S.S.R., Wang, L. & Devabhaktuni, V. 2011. Support vector machine based data classification for detection of electricity theft. In *Proceedings of the 2011 IEEE/PES Power Systems Conference and Exposition (PSCE)*. 1–8. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5772466 [2014, January 14].

Depuru, S.S.S.R., Wang, L., Devabhaktuni, V. & Green, R.C. 2013. High performance computing for detection of electricity theft. *International Journal of Electrical Power & Energy Systems*. 47:21–30. DOI: 10.1016/j.ijepes.2012.10.031.

Dhanya, C.T. & Nagesh Kumar, D. 2009. Data mining for evolution of association rules for droughts and floods in India using climate inputs. *Journal of Geophysical Research*. 114(D2):D02102. DOI: 10.1029/2008JD010485.

Dimaano, I. & Jamora, R. 2010. Embarking on the World ' s Largest Non-Revenue Water Management Project. In *Proceedings of the 6th IWA Water Loss Reduction Specialist Conference*. Sao Paulo, Brazil.

Dobbs, R., Pohl, H., Lin, D.-Y., Mischke, J., Garemo, N., Hexter, J., Matzinger, S., Palter, R., et al. 2013. *Infrastructure productivity: how to save $1 trillion a year*. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Infrastructure+productivity:+How+to+save+$1+trillion+a+year#0.

Esling, P. & Agon, C. 2012. Time-series data mining. *ACM Computing Surveys*. 45(1):1–34. DOI: 10.1145/2379776.2379788.

Fanner, P., Thornton, J., Liemberger, R. & Sturm, R. 2007. *Evaluating Water Loss and Planning Loss Reduction Strategies*. (AWWARF Report Series). AWWA Research Foundation.

Fawcett, T. & Provost, F. 1997. Adaptive fraud detection. *Data mining and knowledge discovery*. 316:291–316. Available: http://link.springer.com/article/10.1023/A:1009700419189 [2014,

January 14].

Ferréol, E. 2005. How to Measure and Reduce the Water Meter Park Inefficiency? *Proceedings of the IWA Specialised Conference "Leakage 2005"*. 1–4.

García-Escudero, L.A., Gordaliza, A., Matrán, C. & Mayo-Iscar, A. 2010. A review of robust clustering methods. *Advances in Data Analysis and Classification*. 4(2–3):89–109. DOI: 10.1007/s11634-010-0064-5.

García, V.J., Cabrera, E. & Enrique, C.J. 2006. The minimum night flow method revisited. In *Water Distribution Systems Analysis Symposium 2006*. Reston, VA: American Society of Civil Engineers. 1–18. DOI: 10.1061/40941(247)35.

Garzon-Contreras, F., Uribe-Preciado, A., Yepes-Enriquez, L. & Agredo-Perdomo, A. 2009. Unauthorized consumption: The key component of Apparent Losses in Colombia's major cities. In *Proceedings of the 5th IWA Water Loss Reduction Specialist Conference*. Cape Town, South Africa. 29–35.

Gauteng Provincial Treasury. 2012. *Provincial Economic Review and Outlook (PERO)*. Johannesburg, South Africa. Available: www.finance.gpg.gov.za.

Gavara, F. & Arregui, F. 2016. Analysis of metering inaccuracies of residential meters in a water supply system. In *Proceedings of the Water Loss 2016 Conference*.

Gorunescu, F. 2011. *Data Mining: Concepts, models and techniques*. J. Kacprzyk & L. C.Jain, Eds. Springer-Verlag Berlin Heidelberg.

Graeser, H.. . 1958. Detecting Lost Water at Dallas. *Journal - American Water Works Association*. 50(7):925–932. Available: http://www.jstor.org/stable/41255868.

Green Cape. 2017. *Water – 2017 Market Intelligence Report* –. Cape Town, South Africa. Available: https://www.greencape.co.za/assets/Uploads/GreenCape-Water-MIR-2017-electronic-FINAL-v1.pdf [2017, August 07].

Guibentif, H., Rufenacht, H.P., Rapillard, P. & Ruetschi, M. 2007. Acceptable level of water Conference, losses in Geneva. In *Proceedings of the 4th IWA Water Loss Reduction Specialist Conference*. Bucharest, Romania. 138–147.

Hamilton, S. & McKenzie, R. 2014. *Water Management and Water Loss*. London, United Kingdom: IWA Publishing.

Hudson, W.D. 1964. Reduction of Unaccounted-for Water. *Journal - American Water Works Association*. 56(2):143–148. Available: http://www.jstor.org/stable/41264131.

Humaid, E. 2012. A Data Mining Based Fraud Detection Model for Water Consumption Billing System in MOG. Islamic University of Gaza. Available: http://library.iugaza.edu.ps/thesis/106989.pdf [2014, January 10].

Humaid, E.H. & Barhoum, T. 2013. Water Consumption Financial Fraud Detection: A Model Based on Rule Induction. *2013 Palestinian International Conference on Information and Communication Technology*. (April):115–120. DOI: 10.1109/PICICT.2013.28.

Hunter, R.B. 1940. *BMS65 Methods of estimating loads in plumbing systems*. Washington DC.

Iglesias, F. & Kastner, W. 2013. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns. *Energies*. 6(2):579–597. DOI: 10.3390/en6020579.

Jain, A., Murty, M. & Flynn, P. 1999. Data clustering: a review. *ACM computing surveys (CSUR)*. 31(3):264–323.

Johannesburg Water. 2016. *Integrated Annual Report 2015/16*. Johannesburg, South Africa.

Johnson, E. 2001. Optimal water meter selection system. *Water SA*. 27(4):481–488. Available: http://www.ajol.info/index.php/wsa/article/view/4961 [2013, August 27].

Johnson, E. 2007a. Estimating uncertainty of flow measurement for the in-situ calibration of large water meters. *Australian Journal of Water Resources*. 11(1):41–52. Available: http://search.informit.com.au/documentSummary;dn=663985337999122;res=IELENG [2013, August 27].

Johnson, E.H. 2007b. *Management of Non-revenue and Revenue Water Data*. Sydney, Australia: Engineers Media.

Johnson, E.H. 2012. Case Studies Demonstrate the Implications of Metering Errors. In *IWA Water Loss 2012*. International Water Association (IWA). 1–10. Available: http://www.iwa-waterloss.org/2012/Final_Papers_1/7.pdf.

Johnson, E. & Vermersch, M. 2011. IWA Guidance for apparent loss reduction: filling the knowledge gap. *Water21*. (August):48–50.

Kalteh, A.M., Hjorth, P. & Berndtsson, R. 2008. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling & Software*. 23(7):835–845. DOI: 10.1016/j.envsoft.2007.10.001.

Kanakoudis, V. & Tsitsifli, S. 2010. Results of an urban water distribution network performance evaluation attempt in Greece. *Urban Water Journal*. 7(5):267–285. DOI: 10.1080/1573062X.2010.509436.

Kanakoudis, V., Tsitsifli, S., Samaras, P. & Zouboulis, A. 2013. Assessing the performance of urban water networks across the EU Mediterranean area: The paradox of high NRW levels and absence of respective reduction measures. *Water Science and Technology: Water Supply*. 13(4):939–950. DOI: 10.2166/ws.2013.044.

Karadirek, I.E., Kara, S., Yilmaz, G., Muhammetoglu, A. & Muhammetoglu, H. 2012. Implementation of Hydraulic Modelling for Water-Loss Reduction Through Pressure Management. *Water Resources Management*. 26(9):2555–2568. DOI: 10.1007/s11269-012-0032-2.

Kim, M. & Ramakrishna, R.S. 2005. New indices for cluster validity assessment. *Pattern Recognition Letters*. 26(15):2353–2363. DOI: 10.1016/j.patrec.2005.04.007.

Kingdom, B., Liemberger, R. & Marin, P. 2006. *The Challenge of Reducing Non-Revenue (NRW) Water in Developing Countries. How the Private Sector can Help: A Look at Performance-Based Service*. (8). Washington DC. Available: http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2007/04/11/000020953_20070411153243/Rendered/PDF/394050Reducing1e0water0WSS81PUBLIC1.pdf [2013, August 27].

Klingel, P. & Knobloch, A. 2015. A Review of Water Balance Application in Water Supply. *Journal - American Water Works Association*. 107(July):E339–E350. DOI: 10.5942/jawwa.2015.107.0084.

Kohonen, T. & Oja, E. 1998. Visual feature analysis by the self-organising maps. *Neural Computing & Applications*. 7:273–286. Available: http://link.springer.com/article/10.1007/BF01414888 [2014, January 20].

Kossieris, P., Kozanis, S., Hashmi, A., Katsiri, E., Vamvakeridou-Lyroudia, L.S., Farmani, R., Makropoulos, C. & Savic, D. 2014. A web-based platform for water efficient households. In *Procedia Engineering*. V. 89. 1128–1135. DOI: 10.1016/j.proeng.2014.11.234.

Kotsiantis, S.B. 2007. Supervised machine learning: A review of classification techniques. *Informatica*. 31:249–268.

Krejcie, R. V & Morgan, D.W. 1970. Determining Sample Size for Research Activities Robert. *Educational and Psychological Measurement*. 38(1):607–610. DOI: 10.1177/001316447003000308.

Kuranz, A.P. 1942. Meter Maintenance Practice. *Journal - American Water Works Association*. 34(1):117–120. Available: http://www.jstor.org/stable/41232764.

Kusiak, A., Zeng, Y. & Zhang, Z. 2013. Modeling and analysis of pumps in a wastewater treatment plant: A data-mining approach. *Engineering Applications of Artificial Intelligence*. 26(7):1643–1651. DOI: 10.1016/j.engappai.2013.04.001.

Lambert, A. 1994. Accounting for Losses: The Bursts and Background Concept. *Water and Environment Journal*. 8(2):205–214. DOI: 10.1111/j.1747-6593.1994.tb00913.x.

Lambert, A. 2000. What Do We Know About Pressure: Leakage Relationships in Distribution Systems? In *IWA Conference on Systems Approach to Leakage Control and Water Distribution SystemManagement*.

Lambert, A.O. 2009. Ten Years Experience in Using the UARL Formula to Calculate Infrastructure Leakage Index. In *Proceedings of the IWA Water Loss 2009 South Africa*. Cape Town, South Africa. 189–196.

Lambert, A. & Hirner, W. 2000. *The IWA Blue Pages: Losses from Water Supply Systems: Standard Terminology and Recommended Performance Measures*.

León, C., Biscarri, F., Monedero, I., Guerrero, J.I., Biscarri, J. & Millán, R. 2011. Variability and Trend-Based Generalized Rule Induction Model to NTL Detection in Power Companies. *IEEE Transactions on Power Systems*. 26(4):1798–1807.

Liao, S.-H., Chu, P.-H. & Hsiao, P.-Y. 2012. Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*. 39(12):11303–11311. DOI: 10.1016/j.eswa.2012.02.063.

Liemberger, R., Brothers, K., Mckenzie, R., Rizzo, A. & Waldron, T. 2007. Water Loss Performance Indicators. In *Water Loss 2007*. 1–12. Available: http://www.miya-water.com/user_files/Data_and_Research/miyas_experts_articles/2_NRW/15_Water Loss Performance Indicators.pdf [2013, August 27].

Lobanga, K.P., Lugoma, F.T., van Zyl, J.E. & Ilemobade, A.A. 2008. THE EXTENT OF NON-COMPLIANT PLUMBING COMPONENTS AND ON-SITE LEAKAGE IN SOUTH AFRICA. 782–789.

Loureiro, D., Alegre, H., Coelho, S.T., Martins, A. & Mamade, a. 2014. A new approach to improve water loss control using smart metering data. *Water Science & Technology: Water Supply*. 14(4):618. DOI: 10.2166/ws.2014.016.

Lugoma, M.F.T., Van Zyl, J. & Ilemobade, A.A. 2012. The extent of on-site leakage in selected suburbs of Johannesburg. *Water SA*. 38(1):127–132. DOI: 10.4314/wsa.v38i1.15.

Maier, H.R., Jain, A., Dandy, G.C. & Sudheer, K.P. 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*. 25(8):891–909.

Makara, L. 2009. Successful Outcomes from Water Loss Reduction Eda Ranu, Papua New Guinea. *Proceedings of the 5th IWA Water Loss Reduction Specialist Conference*. (April):73–84.

Manyaka, R.K. 2014. Collection of municipal own revenue in South Africa : challenges and prospects. *Journal of Public Administration*. 49(1):127–139. Available: https://journals.co.za/content/jpad/49/1/EJC153633 [2018, January 05].

Mayer, P., Deoreo, W., Opitz, E.M., Kiefer, J.C., Davis, W.Y., Dziegielewski, B. & Nelson, J.O. 1999. *Residential end uses of water*. AWWA Research Foundation and American Water Works Association.

Mazibuko, G.P. 2013. The impact of municipal billing system on revenue collection in selected South African cities. University of Pretoria. Available: http://repository.up.ac.za/bitstream/handle/2263/41277/Mazibuko_Impact_2014.pdf?sequence=3.

Mazumdar, A., Jaman, H. & Das, S. 2014. Modification of Hunter's Curve in the Perspective of Water Conservation. *Journal of Pipeline Systems Engineering and Practice*. 5(1):04013007. DOI: 10.1061/(ASCE)PS.1949-1204.0000150.

Mbabazi, D., Banadda, N., Kiggundu, N., Mutikanga, H. & Babu, M. 2015. Determination of domestic water meter accuracy degradation rates in Uganda. *Journal of Water Supply: Research and Technology - AQUA*. 64(4):486–492. DOI: 10.2166/aqua.2015.083.

McIntosh, A.C. 2003. *Asian Water Supplies: Reaching the Urban Poor*. Asian Development Bank and International Water Association. Available: http://www.adb.org/sites/default/files/asian-water-supplies.pdf.

McKenzie, R. & Seago, C. 2005. Assessment of real losses in potable water distribution systems: some recent developments. *… Science & Technology: Water …*. 5(1):33–40.

McKenzie, R., Siqalaba, Z. & Wegelin, W. 2012. *The State of Non-Revenue Water in South Africa (2012)*. Gezina, Pretoria, South Africa: Water Research Commission.

de Medeiros, M.F., Gondo, J.Y., do Rego, A.C. & Padilha, P.S. 2010. Apparent Loss in Sabesp: Water Meter Management System - Under measurement and Inevitable Loss. In *Proceedings of the 6th IWA Water Loss Reduction Specialist Conference*. 1–8.

Morais, J., Pires, Y., Cardoso, C. & Klautau, A. 2009. An Overview of Data Mining Techniques Applied to Power Systems. In *Data Mining and Knowledge Discovery in Real Life Applications*. J. Ponce & A. Karahoca, Eds. Intech. 419–438. Available: http://www.intechopen.com/books/data_mining_and_kno.

Mousavi, S.A., Shahbazi, I., Janjani, H., Veysinejad, R., Sobhani, A.A. & Bakhti, M. 2017. Study of non-revenue water status and enforcement measures to reduce water loss: case study in villages of Kermanshah Province of Iran. *Chinese Journal of Population Resources and Environment*. 00(00):1–6. DOI: 10.1080/10042857.2017.1406744.

Mutikanga, H. 2012. *Water Loss Management: Tools and Methods for Developing Countries*. CRC Press/Balkema. Available: http://repository.tudelft.nl/assets/uuid:361e3285-606a-4d0d-9e9a-34dc0ee0fef4/UNESCO-IHE_PhD_MUTIKANGA_THESIS.pdf [2013, August 27].

Mutikanga, H.E., Vairavamoorthy, K., Kizito, F. & Sharma, S.K. 2009. Decision Support Tool for Optimal Water Meter Replacement. In *Proceedings of the Second International Conference on Advances in Engineering and Technology*. 649–655.

Mutikanga, H.E., Sharma, S.K. & Vairavamoorthy, K. 2010. Customer Demand Profiling for Apparent Water Loss Reduction. In *Proceedings of the 6th IWA Water Loss Reduction Specialist Conference*. Sao Paulo, Brazil. CD-ROM.

Mutikanga, H.E., Sharma, S.K. & Vairavamoorthy, K. 2011a. Assessment of apparent losses in

urban water systems. *Water and Environment Journal*. 25(3):327–335. DOI: 10.1111/j.1747-6593.2010.00225.x.

Mutikanga, H.E., Sharma, S. & Vairavamoorthy, K. 2011b. Investigating water meter performance in developing countries: A case study of Kampala, Uganda. *Water SA*. 37(4):567–574. Available: http://www.scielo.org.za/scielo.php?pid=S1816-79502011000400018&script=sci_arttext&tlng=es [2013, August 27].

Mutikanga, H.E., Sharma, S.K. & Vairavamoorthy, K. 2013. Methods and Tools for Managing Losses in Water Distribution Systems. *Journal of Water Resources Planning and Management*. 139(April):166–174. DOI: 10.1061/(ASCE)WR.1943-5452.0000245.

Nagi, J., Yap, K.S., Tiong, S.K., Ahmed, S.K. & Mohammad, a. M. 2008. Detection of abnormalities and electricity theft using genetic Support Vector Machines. In *TENCON 2008 - 2008 IEEE Region 10 Conference*. Ieee. 1–6. DOI: 10.1109/TENCON.2008.4766403.

Nagi, J., Yap, K.S., Tiong, S.K., Ahmed, S.K. & Mohamad, M. 2010. Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines. *IEEE Transactions on Power Delivery*. 25(2):1162–1171. DOI: 10.1109/TPWRD.2009.2030890.

Ncube, M. & Taigbenu, A. 2015. Meter Accuracy Degradation and Failure Probability based on Meter Tests and Meter Change Data. In *Proceedings of the 4th YWP-ZA Biennial Conference and 1st African YWP Conference*. Pretoria, South Africa.

Ncube, M. & Taigbenu, A. 2019. Assessment of apparent losses due to meter inaccuracy – a comparative approach. *Water SA*. 45(2):174–182. DOI: 10.4314/wsa.v45i2.03.

Ncube, M. & Taigbenu, A.E. 2016. Consumption Characterisation and On-Site Leakage in Johannesburg, South Africa. In *Proceeding of the IWA Water Loss Conference 2016*.

Ncube, M. & Taigbenu, A.E. 2018a. Assessment of apparent losses due to meter inaccuracy using an alternative, validated methodology. *Water Supply*. 19(4):1212–1220. DOI: 10.2166/ws.2018.178.

Ncube, M. & Taigbenu, A.E. 2018b. Decision Enhancement for the Management of Apparent Losses. In *Proceeding of the Biennial WISA 2018 Conference*. Cape Town, South Africa.

Ngai, E.W.T., Xiu, L. & Chau, D.C.K. 2009. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*. 36(2):2592–2602. DOI: 10.1016/j.eswa.2008.02.021.

Nguyen, K. a., Stewart, R. a. & Zhang, H. 2013a. An intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Environmental Modelling & Software*. 47:108–127. DOI: 10.1016/j.envsoft.2013.05.002.

Nguyen, K.A., Zhang, H. & Stewart, R.A. 2013. Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-environment Research*. (March). DOI: 10.1016/j.jher.2013.02.004.

Nguyen, K.A., Stewart, R. a. & Zhang, H. 2013b. An autonomous and intelligent expert system for residential water end-use classification. *Expert Systems with Applications*. (August). DOI: 10.1016/j.eswa.2013.07.049.

Nicolini, M., Giacomello, C. & Deb, K. 2010. Calibration and optimal leakage management for a real water distribution network. *… Planning and Management*. (February):134–142. DOI: 10.1061/(ASCE)WR.1943-5452.0000087.

Noss, R.R., Newman, G.J. & Male, J.W. 1987. Optimal Testing Frequency for Domestic Water Meters. *Journal of Water Resources Planning and Management*. 113(1):1–14. DOI:

10.1061/(ASCE)0733-9496(1987)113:1(1).

OFWAT. 2010. *Service and delivery – performance of the water companies in England and Wales 2009-10*. Available: http://www.ofwat.gov.uk/regulating/reporting/rpt_los_2009-10.pdf.

Ouyang, R., Ren, L., Cheng, W. & Zhou, C. 2010. Similarity search and pattern discovery in hydrological time series data mining. *Hydrological Processes*. 24(9):1198–1210. DOI: 10.1002/hyp.7583.

Palau, C. V., Arregui, F.J. & Carlos, M. 2012. Burst Detection in Water Networks Using Principal Component Analysis. *Journal of Water Resources Planning and Management*. 138(1):47–54. DOI: 10.1061/(ASCE)WR.1943-5452.0000147.

Paparrizos, J. & Gravano, L. 2015. k-Shape: Efficient and Accurate Clustering of Time Series. *ACM Sigmod*. 1855–1870. DOI: 10.1145/2723372.2737793.

Pearson, D. & Lambert, A. 2013. Accounting for Water Leakage and Measuring. *Sustainable Cities*. (June):1–5.

Puleo, V., Fontanazza, C.M., Notaro, V., Marchis, M. De, Loggia, G. La & Freni, G. 2014. Definition of water meter substitution plans based on a composite indicator. *Procedia Engineering*. 70:1369–1377. DOI: 10.1016/j.proeng.2014.02.151.

Puust, R., Kapelan, Z., Savic, D.A. & Koppel, T. 2010. A review of methods for leakage management in pipe networks. *Urban Water Journal*. 7(1):25–45. DOI: 10.1080/15730621003610878.

R Core Team. 2017. Available: https://www.r-project.org.

Räsänen, T., Ruuskanen, J. & Kolehmainen, M. 2008. Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. *Applied Energy*. 85(9):830–840. DOI: 10.1016/j.apenergy.2007.10.012.

Rizzo, A. & Cilia, J. 2005. Quantifying meter under-registration caused by the ball valves of roof tanks (for indirect plumbing systems). In *Proceedings of the Leakage 2005 Conference, ….* International Water Association (IWA). 1–12. Available: http://rash.apanela.com/tf/leakage/Quantifying Meter Under-Registration Caused by the Ball Valves of Roof Tanks.pdf [2013, August 27].

Rizzo, A., Pearson, D., Stephenson, M. & Harper, N. 2004. Apparent water loss control: A practical approach. *Water21*. (June 2004):44–45. Available: http://www.iwaponline.com/w21/00603/w21006030044.htm [2013, August 27].

Rizzo, A., Vermersch, M., John, S., Micallef, G. & Pace, R. 2007. Apparent water loss control: the way forward. *Water21*. 45–47. Available: http://www.iwaponline.com/w21/00904/w21009040045.htm [2013, August 28].

Rossman, L.A. 2000.

Saitta, S., Raphael, B. & Smith, I.F.C. 2007. A Bounded Index for Cluster Validity. In *Machine Learning and Data Mining in Pattern Recognition: 5th International Conference, MLDM 2007, Leipzig, Germany, July 18-20, 2007. Proceedings*. P. Perner, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg. 174–187. DOI: 10.1007/978-3-540-73499-4_14.

SANS. 2006. *SANS 1529-1 : 2006 SOUTH AFRICAN NATIONAL STANDARD Water meters for cold potable water Part 1 : Metrological characteristics of mechanical water meters of nominal bore not exceeding 100 mm*.

Sard, A. 2015. Available: https://cran.r-project.org/web/packages/dtwclust/index.html.

Schouten, M. & Halim, R.D. 2010. Resolving strategy paradoxes of water loss reduction: A synthesis in Jakarta. *Resources, Conservation and Recycling*. 54(12):1322–1330. DOI: 10.1016/j.resconrec.2010.05.006.

Seago, C., Bhagwan, J. & McKenzie, R. 2004. Benchmarking leakage from water reticulation systems in South Africa. *Water SA*. 30(5):573–580.

Sharma, S.K. & Chinokoro, H. 2010. Estimation of ELL and ELWL for Lusaka Water Distribution System. In *Proceedings of the 6th IWA Water Loss Reduction Specialist Conference (CD-ROM)*. Sao Paulo, Brazil.

Sharma, S.K. & Nhemafuki, A. 2009. Water Loss Management in Bhaktapur and Dhulikhel Cities in Nepal. *Proceedings of the 5th IWA Water Loss Reduction Specialist Conference*. (April):546–552.

Shields, D. 2011. Revenue recovery through meter replacement. Utah State University. Available: http://digitalcommons.usu.edu/etd/1111/ [2013, August 27].

Shields, D.J., Barfuss, S.L., Johnson, M.C. & Hannah, C.C. 2013. Meter-testing methods matter. 664–670.

Shin, E., Park, H., Park, C. & Hyun, I. 2005. A Case Study of Leakage Management in the City of Busan , Korea. *Leakage Conference Proceeding*. 1–10.

Siew, C. & Tanyimboh, T.T. 2012. Pressure-Dependent EPANET Extension. *Water Resources Management*. 26(6):1477–1498. DOI: 10.1007/s11269-011-9968-x.

De Silva, D., Yu, X., Alahakoon, D. & Holmes, G. 2011. A data mining framework for electricity consumption analysis from meter data. *IEEE Transactions on Industrial Informatics*. 7(3):399–407. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5893960 [2013, August 27].

Solomatine, D.P. 2003. Applications of data-driven modelling and machine learning in control of water resources. In *Computational Intelligence in Control*. M. Mohammadian, R.A. Sarker, & X. Yao, Eds. Hershey, PA: Idea Group Publishing. 197–217.

Structured Data LLC. 2018. Available: http://bert-toolkit.com/.

Szilveszter, S., Beltran, R. & Fuentes, A. 2015. Performance analysis of the domestic water meter park in water supply network of Ibarra, Ecuador. *Urban Water Journal*. 14(1):85–96. DOI: 10.1080/1573062X.2015.1057181.

Tabesh, M., Yekta, A.H.A. & Burrows, R. 2009. An Integrated Model to Evaluate Losses in Water Distribution Systems. *Water Resources Management*. 23(3):477–492. DOI: 10.1007/s11269-008-9284-2.

van Thienen, P. & Vertommen, I. 2016. Automated feature recognition in CFPD analyses of DMA or supply area flow data. *Journal of Hydroinformatics*. 18(3):514–520. DOI: 10.2166/hydro.2015.056.

Thornton, J., Sturm, R. & Kunkel, G. 2008a. Controlling Apparent Losses from Systematic Data Handling Errors in Customer Billing Systems. In *Water Loss Control*. Second ed. McGraw-Hill Professional. 225–246.

Thornton, J., Sturm, R. & Kunkel, G. 2008b. Controlling Apparent Losses — Unauthorized Consumption. In *Water Loss Control*. Second ed. McGraw-Hill Professional. 247–258.

Trifunovic, N., Sharma, S. & Pathirana, A. 2009. Modelling Leakage in Distribution System Conference, using EPANET. In *Proceedings of the 5th IWA Water Loss Reduction*

Specialist Conference. Cape Town, South Africa. 482–489.

Tung, A.K.H., Zhang, R., Koudas, N. & Ooi, B.C. 2006. Similarity search: a matching based approach. In *Proceedings of the 32nd International Conference on Very Large Databases*. Seoul, Korea: VLDB Endowment. 631--642. Available: http://dl.acm.org/citation.cfm?id=1164182 [2014, January 14].

UNEP. 2011. *Towards a Green Economy: Pathways to Sustainable Development and Poverty Eradication*. V. www.unep.o. DOI: 10.1063/1.3159605.

United Nations. 2016. *The Sustainable Development Goals Report*. DOI: 10.18356/3405d09f-en.

Vassiljev, A., Koppel, T. & Puust, R. 2010. Use of error analysis for calibration of water distribution systems. (2007).

Vermersch, M., Carteado, F., Rizzo, A., Johnson, E., Arregui, F. & Lambert, A. 2016. *Guidance notes on apparent losses and water loss reduction planning*.

Vesanto, J. & Ahola, J. 1999. Hunting for correlations in data using the self-organizing map. In *Proceedings of the International ICSC Congress on Computational Intelligence Methods and Applications (CIMA)*. Academic Press. 279–285.

World Economic Forum. 2017. *The Global Risks Report 2017: 12th Edition*. Geneva, Switzerland. Available: http://www3.weforum.org/docs/GRR17_Report_web.pdf.

Wu, Y. & Liu, S. 2017. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water Journal*. 14(9):972–983. DOI: 10.1080/1573062X.2017.1279191.

Wu, C.G., Liu, B.C., Yuan, Y.X., Pei, W.Z., Zhao, P., Wang, M.X. & Yang, X.D. 2010. Water distribution network leakage analysis modeling based on pattern recognition theory. *Pattern Recognition*. 599–603.

Wu, Z.Y., Farley, M., Turtle, D., Kapelan, Z., Boxall, J., Mounce, S., Dahasahasra, S., Mulay, M., et al. 2011. *Water Loss Reduction*. Exton, Pennsylvania USA.: Bentley Institute Press.

WWAP. 2016. *The United Nations World Water Development Report 2016 - Water and Jobs*. DOI: ISBN 978-92-3-100146-8.

Xin, K., Tao, T., Lu, Y., Xiong, X. & Li, F. 2014. Apparent Losses Analysis in District Metered Areas of Water Distribution Systems. *Water Resources Management*. 28(3):1–14. DOI: 10.1007/s11269-013-0508-8.

Yaniv, S., Freni, G. & Fantozzi, M. 2005. Analysis method for apparent losses estimation and identification of technical solutions. V. 3.

Yee, M. 1999. Economic analysis for replacing residential meters. *Journal-American Water Works Association*. 91(7):72–77. Available: http://cat.inist.fr/?aModele=afficheN&cpsidt=1878870 [2013, August 27].

Yin, Y., Kaku, I., Tang, J. & Zhu, J. 2011. *Data Mining: Concepts, Methods and Applications in Management and Engineering Design*. R. Roy, Ed. Springer-Verlag London.

Zeileis, A. & Grothendieck, G. 2005. zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*. 14(6):1–27. DOI: 10.1017/CBO9781107415324.004.

van Zyl, J. 2011. *Introduction to Intergrated Water Meter Management*. 1st ed. Gezina, Pretoria, South Africa: Water Research Commission.

van Zyl, J.E. & Clayton, C.R.I. 2007. The effect of pressure on leakage in water distribution systems.

Proceedings of the Institution of Civil Engineers - Water Management. 160(2):109–114. DOI: 10.1680/wama.2007.160.2.109.

Appendix A Meter Test Results Sample Data

Following is a random selection of the full test reports from the accredited JW laboratory.

Water Meter Test Report — Test Report No: 20151118 T005 — Report Date: 17/Dec/2015

Applicable Regulation: SANS 1529-1 — File Name: 20151118T005--Std-C-HRA 6376 — (dd-Mmm-yyyy)

Procedure No. 2030

Johannesburg Water

Meter Details		Meter Details		Test Details		Installation Details	
Meter Number	C-HRA 6376	Manufacturer	Kent	Test Date (yyyymmdd)	20151118	Stand No	
Meter Reading m³	376	Meter Make / Model	KSM	Flow Orientation	H	Township	Johannesbug Flow Laboratory
Meter Size mm	15	Meter Type	Volumetric	Test Area Temp (°C)	24.5	Street No	No 2 Ffennell road
Q_t (m³/h)	1.5					Street Name	Cnr Rosettenville road
Q_4 (m³/h)	5	Meter Status	In Use	Client Details		Region	Village Main
Meter Resolutio (l)	0.02	(New & Refurb)		Name			Tel : 011 493 2270
Meter Class	C	(In Use)		Address		V Order No	

Test Bench Used	Vol B	Reference Quantities					Test Item			Test Results		
Test Bench Postion (TXXX)	Test No	Test Press	Water Temp	Test Duration	Ref Volume	Ref Flow Rate	Start Reading	End Reading	Indicated Volume	Error ABS	Error Rel	Variance To Ave
		kPa	°C	s	l_{REF}	l/h	m³	m³	l_{IND}	l	%	%
0.9 to 1.0 Q_t	**Test 1**	482	22.9	186.89	151.70	2922.15	376.36482	376.51543	150.61	-1.09	-0.72	-0.04
2700 to 3000 l/h	**Test 2**	482	22.9	188.17	151.45	2897.49	376.51543	376.66592	150.49	-0.96	-0.63	0.05
± 2% New & Refurb	**Test 3**	482	22.9	188.03	151.75	2905.39	376.66592	376.81663	150.71	-1.04	-0.69	-0.01
± 3.5% In Use	**Ave**					2908.34			150.60	-1.03	-0.68	
									Test Result	Pass		
0.9 to 1.0 Q_t	**Test 1**	648	23.0	242.82	151.60	2247.59	376.81663	376.96744	150.81	-0.79	-0.52	-0.16
2025 to 2250 l/h	**Test 2**	648	23.0	242.71	151.60	2248.61	376.96744	377.11844	151.00	-0.60	-0.40	-0.03
± 2% New & Refurb	**Test 3**	648	21.0	242.69	151.60	2248.79	377.11844	377.26978	151.34	-0.26	-0.17	0.19
± 3.5% In Use	**Ave**					2248.33			151.05	-0.55	-0.36	
									Test Result	Pass		
0.45 to 0.5 Q_t	**Test 1**	800	22.4	377.92	151.450	1442.69	377.26978	377.42137	151.59	0.14	0.09	0.00
1350 to 1500 l/h	**Test 2**	800	22.4	376.83	151.550	1447.81	377.42137	377.57300	151.63	0.08	0.05	-0.04
± 2% New & Refurb	**Test 3**	800	22.4	374.21	151.650	1458.91	377.57300	377.72484	151.84	0.19	0.13	0.04
± 3.5% In Use	**Ave**					1449.80			151.69	0.14	0.09	
									Test Result	Pass		
1.0 to 1.1 Q_t	**Test 1**	248	22.4	52.68	10.930	746.92	377.72558	377.73660	11.02	0.09	0.82	0.09
675 to 750 l/h	**Test 2**	248	22.4	52.52	10.905	747.49	377.73660	377.74758	10.98	0.08	0.69	-0.05
± 2% New & Refurb	**Test 3**	248	22.4	52.80	10.895	742.84	377.74758	377.75855	10.97	0.07	0.69	-0.04
± 3.5% In Use	**Ave**					745.75			10.99	0.08	0.73	
									Test Result	Pass		
1.0 to 1.1 $Q_{...}$	**Test 1**	273	17.6	330.22	10.885	118.67	378.16045	378.17150	11.05	0.16	1.52	0.00
108 to 120 l/h	**Test 2**	273	18.3	331.51	10.875	118.10	378.17150	378.18255	11.05	0.18	1.61	0.09
± 5% New & Refurb	**Test 3**	273	19.8	329.76	10.855	118.50	378.18255	378.19356	11.01	0.15	1.43	-0.09
± 8% In Use	**Ave**					118.42			11.04	0.16	1.52	
									Test Result	Pass		
	Test 1	275	20.8	656.32	10.85	59.51	378.19356	378.20450	10.94	0.09	0.83	0.00
54 to 60 l/h	**Test 2**	275	21.2	656.29	10.85	59.52	378.20450	378.21545	10.95	0.10	0.92	0.09
± 2% New & Refurb	**Test 3**	275	21.6	656.19	10.85	59.53	378.21545	378.22638	10.93	0.08	0.74	-0.09
± 3.5% In Use	**Ave**					59.52			10.94	0.09	0.83	
									Test Result	Pass		
	Test 1	277	21.9	1316.49	10.85	29.66	378.22638	378.23718	10.80	-0.04	-0.41	-0.09
27 to 30 l/h	**Test 2**	277	22.3	1315.83	10.83	29.63	378.23718	378.24798	10.80	-0.03	-0.28	0.05
± 2% New & Refurb	**Test 3**	277	22.8	1315.29	10.85	29.70	378.24798	378.25880	10.82	-0.03	-0.28	0.05
± 3.5% In Use	**Ave**					29.66			10.81	-0.03	-0.32	
									Test Result	Pass		
	Test 1	277	23.4	1782.38	10.850	21.91	378.25880	378.26942	10.62	-0.23	-2.12	-0.51
20 to 22 l/h	**Test 2**	277	24.2	1782.59	10.835	21.88	378.26942	378.28010	10.68	-0.16	-1.43	0.18
± 2% New & Refurb	**Test 3**	277	25.2	1782.38	10.840	21.89	378.28010	378.29080	10.70	-0.14	-1.29	0.32
± 3.5% In Use	**Ave**					21.90			10.67	-0.18	-1.61	
									Test Result	Pass		
	Test 1	277	25.9	2678.12	10.825	14.55	378.29080	378.30112	10.32	-0.50	-4.67	-0.19
13.5 to 15 l/h	**Test 2**	277	20.7	2680.22	10.855	14.58	378.30112	378.31150	10.38	-0.48	-4.38	0.10
± 2% New & Refurb	**Test 3**	277	21.9	2679.33	10.855	14.58	378.31150	378.32188	10.38	-0.48	-4.38	0.10
± 3.5% In Use	**Ave**					14.57			10.36	-0.48	-4.47	
									Test Result	Pass		
	Test 1	277	24.2	5612.61	10.860	6.97	378.32188	378.33103	9.15	-1.71	-15.75	-2.28
6.2 to 7 l/h	**Test 2**	277	22.4	5620.13	10.860	6.96	378.33103	378.34052	9.49	-1.37	-12.62	0.86
± 5% New & Refurb	**Test 3**	277	25.3	5616.39	10.870	6.97	378.34052	378.35008	9.56	-1.31	-12.05	1.42
± 8% In Use	**Ave**					6.96			9.40	-1.46	-13.47	
									Test Result	Pass		

Error Calculation

$((I_{IND} - I_{REF})/I_{REF})100$ — **Final Result** Pass

Key		
	Under-Register	U-R
	Over-Register	O-R
	Defective Register	D-R
	Meter Stopped	M-S

Testing Officer: F.H. Termets - VO number 569 — **Signed** _____ — **Date** 17/Dec/2015

Water Meter Test Report

Test Report No: 20160113 T001 | Report Date: 6/Apr/2015 (dd-Mmm-yyyy)

Applicable Regulation: SANS 1529-1 | File Name: 20160113T001--Std-150069443

Procedure No. 2030

Johannesburg Water

Meter Details		Meter Details		Test Details		Installation Details	
Meter Number	150069443	Manufacturer	Sensus	Test Date (yyyymmdd)	20160113	Stand No	
Meter Reading m³	50	Meter Make / Model	XNP	Flow Orientation	H	Township	Johannesbug Flow Laboratory
Meter Size mm	15	Meter Type	Multie Jet	Test Area Temp (°C)	23	Street No	No 2 Ffennell road
Q_t (m³/h)	1.5					Street Name	Cnr Rosettenville road
Q_n (m³/h)	3	Meter Status	In Use	Client Details		Region	Village Main
Meter Resolutio (l)	0.02	(New & Refurb)		Name			Tel : 011 493 2270
Meter Class	B	(In Use)		Address		W Order No	

Test Bench Used	Vol B	Reference Quantities					Test Item			Test Results		
Test Bench Postion		Test No	Test Press	Water Temp	Test Duration	Ref Volume	Ref Flow Rate	Start Reading	End Reading	Indicated Volume	Error ABS	Error Rel	Variance To Ave
(TXXX)	T001		kPa	°C	s	l_{REF}	l/h	m³	m³	l_{IND}	l	%	%
		Test 1	60	25.1	185.62	151.55	2939.23	51.20618	51.35675	150.57	-0.98	-0.65	-0.01
2700 to 3000 l/h		Test 2	60	25.1	185.41	152.00	2951.30	51.35675	51.50782	151.07	-0.93	-0.61	0.02
± 2%	New & Refurb	Test 3	60	25.0	185.55	151.65	2942.28	51.50782	51.65850	150.68	-0.97	-0.64	-0.01
± 3.5%	In Use	Ave					2944.27			150.77	-0.96	-0.63	
										Test Result	Pass		
		Test 1	137	25.0	242.55	151.50	2248.61	51.65850	51.80912	150.62	-0.88	-0.58	0.02
2025 to 2250 l/h		Test 2	137	25.0	242.77	151.60	2248.05	51.80912	51.95982	150.70	-0.90	-0.59	0.01
± 2%	New & Refurb	Test 3	137	25.0	242.69	151.65	2249.54	51.95982	52.11052	150.70	-0.95	-0.63	-0.03
± 3.5%	In Use	Ave					2248.73			150.67	-0.91	-0.60	
										Test Result	Pass		
		Test 1	290	24.6	365.29	151.70	1495.03	52.11052	52.26210	151.58	-0.12	-0.08	0.01
1350 to 1500 l/h		Test 2	290	24.7	364.68	151.70	1497.53	52.26210	52.41366	151.56	-0.14	-0.09	0.00
± 2%	New & Refurb	Test 3	290	24.7	364.39	151.70	1498.72	52.41366	52.56522	151.56	-0.14	-0.09	0.00
± 3.5%	In Use	Ave					1497.10			151.57	-0.13	-0.09	
										Test Result	Pass		
		Test 1	187	24.6	52.46	10.890	747.31	52.56522	52.57615	10.93	0.04	0.37	-0.08
675 to 750 l/h		Test 2	187	24.6	52.42	10.895	748.23	52.57615	52.58710	10.95	0.06	0.50	0.06
± 2%	New & Refurb	Test 3	187	24.6	52.30	10.900	750.29	52.58710	52.59805	10.95	0.05	0.46	0.02
± 3.5%	In Use	Ave					748.61			10.94	0.05	0.44	
										Test Result	Pass		
		Test 1	378	24.5	324.95	10.920	120.98	52.81334	52.82426	10.92	0.00	0.00	-0.03
108 to 120 l/h		Test 2	378	25.1	324.36	10.866	120.60	52.82426	52.83512	10.86	-0.01	-0.06	-0.08
± 5%	New & Refurb	Test 3	378	25.2	323.32	10.865	120.98	52.83512	52.84600	10.88	0.01	0.14	0.11
± 8%	In Use	Ave					120.85			10.89	0.00	0.03	
										Test Result	Pass		
		Test 1	276	24.9	650.10	10.74	59.47	52.85450	52.86524	10.74	0.00	0.00	-0.25
54 to 60 l/h		Test 2	276	25.8	649.73	10.74	59.51	52.86524	52.87602	10.78	0.04	0.37	0.12
± 2%	New & Refurb	Test 3	276	26.1	649.68	10.74	59.51	52.87602	52.88680	10.78	0.04	0.37	0.12
± 3.5%	In Use	Ave					59.50			10.77	0.03	0.25	
										Test Result	Pass		
		Test 1	278	26.8	1288.18	10.74	30.00	52.88680	52.89736	10.56	-0.18	-1.63	0.40
27 to 30 l/h		Test 2	278	27.4	1290.13	10.75	30.00	52.89736	52.90782	10.46	-0.29	-2.70	-0.67
± 2%	New & Refurb	Test 3	278	27.8	1290.31	10.74	29.96	52.90782	52.91837	10.55	-0.19	-1.77	0.26
± 3.5%	In Use	Ave					29.99			10.52	-0.22	-2.03	
										Test Result	Pass		
		Test 1	279	28.8	1828.22	10.740	21.15	52.91837	52.92844	10.07	-0.67	-6.24	-2.05
20 to 22.5 l/h		Test 2	279	29.7	1819.68	10.740	21.25	52.92844	52.93886	10.42	-0.32	-2.98	1.21
± 2%	New & Refurb	Test 3	279	30.7	1818.53	10.740	21.26	52.93886	52.94924	10.38	-0.36	-3.35	0.84
± 3.5%	In Use	Ave					21.22			10.29	-0.45	-4.19	
										Test Result	Pass		
		Test 1	279	31.6	2581.24	10.740	14.98	52.94924	52.95940	10.16	-0.58	-5.40	1.02
13.5 to 15 l/h		Test 2	279	26.2	2582.31	10.740	14.97	52.95940	52.96960	10.20	-0.54	-5.03	1.40
± 2%	New & Refurb	Test 3	279	26.8	2582.19	10.740	14.97	52.98137	52.99116	9.79	-0.95	-8.85	-2.42
± 3.5%	In Use	Ave					14.97			10.05	-0.69	-6.42	
										Test Result	Pass		
		Test 1	279	28.3	5700.16	10.735	6.78	52.99116	53.00059	9.43	-1.30	-12.16	0.16
6.2 to 7 l/h		Test 2	279	29.7	5706.33	10.735	6.77	53.00059	53.01000	9.41	-1.33	-12.34	-0.03
± 5%	New & Refurb	Test 3	279	31.6	5703.45	10.735	6.78	53.01000	53.01940	9.40	-1.34	-12.44	-0.12
± 8%	In Use	Ave					6.78			9.41	-1.32	-12.31	
										Test Result		Fail	
								Error Calculation					
								$((I_{IND} - I_{REF})/I_{REF})100$		Final Result	Pass		
										Key	Under-Register		U-R
											Over-Register		O-R
Testing Officer: F.H. Termets - VO number 569		Signed						Date	6/Apr/2015		Defective Register		D-R
											Meter Stopped		M-S

Water Meter Test Report

Applicable Regulation: SANS 1529-1			**Test Report No**	20160517 T004	**Report Date**	6/Jun/2016	
Procedure No. 2030			**File Name**	20160517T004-Suideroord-Std0-C-LUJ013		(dd-Mmm-yyyy)	
Meter Details			**Meter Details**		**Test Details**		**Installation Details**
Meter Number		C-LUJ013	Manufacturer	Elster	Test Date (yyyymmdd)	20160517	Stand No 0
Meter Reading	m³	5329	Meter Make / Model	Optima	Flow Orientation	H	Township Suideroord
Meter Size	mm	20	Meter Type	Multi Jet	Test Area Temp (°C)	14.6	Street No 5
Q_t (m³/h)		2.5					Street Name Conroy
Q_s (m³/h)		5	Meter Status	In Use	**Client Details**		Region
Meter Resolutio	(l)	0.02	(New & Refurb)		Name		
Meter Class		B & C	(In Use)		Address		**W Order No**

Johannesburg Water

Johannesbug Flow Laboratory
No 2 Ffennell road
Cnr Rosettenville road
Village Main
Tel : 011 493 2270

Test Bench Used	Vol B			**Reference Quantities**					**Test Item**			**Test Results**		
Test Bench Postion			**Test No**	**Test Press**	**Water Temp**	**Test Duration**	**Ref Volume**	**Ref Flow Rate**	**Start Reading**	**End Reading**	**Indicated Volume**	**Error ABS**	**Error Rel**	**Variance To Ave**
(TXXX)	T004			kPa	°C	s	l_{REF}	l/h	m³	m³	l_{IND}	l	%	%
			Test 1	274	17.3	413.17	10.84	94.41	5331.71075	5331.72160	10.85	0.01	0.14	0.31
90 to 100	l/h		**Test 2**	274	17.3	412.86	10.84	94.48	5331.72160	5331.73238	10.78	-0.05	-0.51	-0.34
± 2%	New & Refurb		**Test 3**	274	17.5	412.45	10.84	94.57	5331.73238	5331.74320	10.82	-0.02	-0.14	0.03
± 3.5%	In Use		**Ave**					94.49			10.82	-0.02	-0.17	
											Test Result	Pass		
			Test 1	274	17.5	779.95	10.830	49.99	5331.74320	5331.75425	11.05	0.22	2.03	1.57
45 to 50	l/h		**Test 2**	274	17.5	780.21	10.830	49.97	5331.75425	5331.76505	10.80	-0.03	-0.28	-0.74
± 2%	New & Refurb		**Test 3**	274	17.5	780.09	10.830	49.98	5331.76505	5331.77584	10.79	-0.04	-0.37	-0.83
± 3.5%	In Use		**Ave**					49.98			10.88	0.05	0.46	
											Test Result	Pass		
			Test 1	274	18.2	1031.09	10.830	37.81	5331.77584	5331.78661	10.77	-0.06	-0.55	-0.52
28 to 38	l/h		**Test 2**	274	18.2	1030.27	10.830	37.84	5331.78661	5331.79742	10.81	-0.02	-0.18	-0.15
± 2%	New & Refurb		**Test 3**	274	18.2	1030.17	10.830	37.85	5331.79742	5331.80832	10.90	0.07	0.65	0.68
± 3.5%	In Use		**Ave**	274				37.83			10.83	0.00	-0.03	
											Test Result	Pass		
			Test 1	274	18.3	1585.33	10.830	24.59	5331.80832	5331.81894	10.62	-0.21	-1.94	0.98
18 to 25	l/h		**Test 2**	274	18.3	1586.37	10.830	24.58	5331.81894	5331.82950	10.56	-0.27	-2.49	0.43
± 2%	New & Refurb		**Test 3**	274	18.3	1586.41	10.830	24.58	5331.82950	5331.83986	10.36	-0.47	-4.34	-1.42
± 3.5%	In Use		**Ave**					24.58			10.51	-0.32	-2.92	
											Test Result	Pass		
			Test 1	275	18.5	3 263.82	10.850	11.97	5331.83986	5331.84696	7.10	-3.75	-34.56	-1.85
11 to 12	l/h		**Test 2**	275	18.5	3 262.79	10.840	11.96	5331.84696	5331.85440	7.44	-3.40	-31.37	1.34
± 5%	New & Refurb		**Test 3**	275	18.8	3 262.59	10.840	11.96	5331.85440	5331.86175	7.35	-3.49	-32.20	0.51
± 8%	In Use		**Ave**					11.96			7.30	-3.55	-32.71	
											Test Result		Fail	
			Test 1	168	16.6	118.38	152.25	4630.01	5329.87525	5330.02608	150.83	-1.42	-0.93	0.00
4500 to 5000	l/h		**Test 2**	168	16.7	118.53	152.25	4624.15	5330.02608	5330.17690	150.82	-1.43	-0.94	-0.01
± 2%	New & Refurb		**Test 3**	168	16.7	119.02	152.25	4605.11	5330.17690	5330.32776	150.86	-1.39	-0.91	0.02
± 3.5%	In Use		**Ave**					4619.75			150.84	-1.41	-0.93	
											Test Result	Pass		
			Test 1	370	17.0	149.22	152.10	3669.48	5330.32776	5330.47726	149.50	-2.60	-1.71	-1.02
3375 to 3750	l/h		**Test 2**	370	17.0	149.62	152.00	3657.27	5330.47726	5330.62862	151.36	-0.64	-0.42	0.27
± 2%	New & Refurb		**Test 3**	370	17.3	148.33	152.00	3689.07	5330.62862	5330.78070	152.08	0.08	0.05	0.75
± 3.5%	In Use		**Ave**					3671.94			150.98	-1.05	-0.69	
											Test Result	Pass		
			Test 1	673	17.6	223.14	151.55	2445.01	5330.78070	5330.93036	149.66	-1.89	-1.25	0.03
1125 to 2500	l/h		**Test 2**	673	17.6	222.64	151.55	2450.50	5330.93036	5331.07990	149.54	-2.01	-1.33	-0.05
± 2%	New & Refurb		**Test 3**	673	17.6	222.97	151.35	2443.65	5331.07990	5331.22936	149.46	-1.89	-1.25	0.03
± 3.5%	In Use		**Ave**					2446.39			149.55	-1.93	-1.27	
											Test Result	Pass		
			Test 1	217	17.7	446.98	151.55	1220.59	5331.22936	5331.37901	149.65	-1.90	-1.25	-0.02
1125 to 1250	l/h		**Test 2**	217	17.7	446.89	151.55	1220.84	5331.37901	5331.52876	149.75	-1.80	-1.19	0.05
± 2%	New & Refurb		**Test 3**	217	17.7	446.82	151.55	1221.03	5331.52876	5331.67840	149.64	-1.91	-1.26	-0.03
± 3.5%	In Use		**Ave**					1220.82			149.68	-1.87	-1.23	
											Test Result	Pass		
			Test 1	274	16.7	202.46	10.885	193.55	5331.67840	5331.68916	10.76	-0.12	-1.15	-0.34
180 to 200	l/h		**Test 2**	274	17.0	203.63	10.855	191.91	5331.68916	5331.69997	10.81	-0.05	-0.41	0.40
± 5%	New & Refurb		**Test 3**	274	17.0	202.69	10.875	193.15	5331.69997	5331.71075	10.78	-0.09	-0.87	-0.06
± 8%	In Use		**Ave**					192.87			10.78	-0.09	-0.81	
											Test Result	Pass		
									Error Calculation					
									$((I_{IND} - I_{REF})/I_{REF})100$		**Final Result**	Pass		
												Key	Under-Register	U-R
													Over-Register	O-R
Testing Officer: F.H. Termets - VO number 569			**Signed**						**Date**	6/Jun/2016			Defective Register	D-R

Appendix B Comparative Billing Analysis Code

B.1.1 Stored Procedure for Pivoting Data in SQL

The stored procedure below allows the pivoting of data from the staging meter reading database to the final database with the entire consumption record for each combination of property id and meter number in one row. The syntax is obviously specific to the names used in the database.

```
USE [JWMetering]
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
ALTER PROCEDURE [dbo].[sproc_pivot_meter_readings]
AS
BEGIN
DECLARE @columns NVARCHAR(MAX), @sql NVARCHAR(MAX);
SET @columns = N'';
SELECT @columns = N',' + STUFF((SELECT DISTINCT ',' + QUOTENAME([Period])
                                FROM [MeterReadings] AS rt
                                FOR XML PATH('')),1,1,'');
SET @sql = N'
SELECT SGID, MeterNo,' + STUFF(@columns, 1, 1, '') + ' INTO MeterReadTrans
FROM ( SELECT rt.SGID, rt.MeterNo, rt.[Period], Max(rt.Reading) as Reading
          FROM [MeterReadings] AS rt
               GROUP BY rt.SGID, rt.MeterNo, rt.[Period]
       ) as j
PIVOT
(
  MAX(j.Reading)
  FOR [Period] IN ('
  + STUFF(@columns, 1, 1, '')
  + ')
) AS pvt;';
--PRINT @sql;
EXEC sp_executesql @sql;
END
```

B.1.2 BERT Functions

The following functions have been written for inclusion within the BERT functions folder to be able perform the requisite R analysis within Excel. The functions are called within Excel when the appropriate button is clicked.

```
processconsumption <- function(strFilePath, strFolder, initDate, LastDate){


  list.of.packages <- c("ggplot2", "zoo", "openxlsx", "plyr" )
  new.packages <- list.of.packages[!(list.of.packages %in% installed.pack-
ages()[,"Package"])]
  if(length(new.packages)) install.packages(new.packages)

  library(ggplot2)
  library(zoo)
  library(openxlsx)
  library(plyr)

  setwd(strFolder)

  ##Reading in the data and naming correcting dates
  Data <- read.xlsx(strFilePath, sheet = 1, startRow = 1, colNames = TRUE,
rowNames = TRUE)
  colnames(Data) <- as.Date(as.integer(colnames(Data)), origin="1899-12-30")
#rename columns to dates

  TData <- as.data.frame(t(Data))  #transpose data frame

  #Tranforming to zoo data
  td <- seq(as.Date(initDate), as.Date(LastDate), "months") #creating a data se-
quence for zoo
  ZooData <- zoo(TData, td)

  # filling gaps in monthly data
  #ZooData <- na.spline(ZooData) + 0*na.approx(ZooData, na.rm = FALSE) correct
if no negative results, which is not the case here
  ZooData <- na.approx(ZooData, na.rm = FALSE)

  #write filled file
  write.csv(as.data.frame(t(ZooData)), "./FilledMonthlyConsumption.csv", na="")

  #obtaining runnining annual total
  AnnualConsumption <- as.data.frame(t(rollapply(ZooData, 12, sum))) # creating
annual consumption

  #obtain mean of rows
  AveAnnualConsumption <- sapply(AnnualConsumption, mean, na.rm = TRUE)
  BERT.graphics.device("AveAnnualConsumption")
  plot(AveAnnualConsumption, main = "Variation of Average Annual Consumption",
                ylab="Annual Consumption, kl", xlab = "Months", col="black",
pch=19)
  grid()

  # write the annual consumption file
  write.csv(AnnualConsumption, "./AnnualConsumption.csv", na="")

}

preclusteringanalysis <- function(strFilePath, strFolder){

  list.of.packages <- c("ggplot2", "zoo", "openxlsx", "plyr", "dtwclust" )
  new.packages <- list.of.packages[!(list.of.packages %in% installed.pack-
ages()[,"Package"])]
  if(length(new.packages)) install.packages(new.packages)

  library(ggplot2)
  library(zoo)
```

```
  library(openxlsx)
  library(plyr)
  library(dtwclust)

  setwd(strFolder)

  # Annual Data
  AnnualConsumption <- read.xlsx(strFilePath, sheet = 1, startRow = 1, colNames
= TRUE, rowNames = TRUE)
  colnames(AnnualConsumption) <- as.Date(as.integer(colnames(AnnualConsump-
tion)),origin="1899-12-30") #rename columns to dates

  # Transposing the dataframes for analysis
  YData <- as.data.frame(t(AnnualConsumption))  #transpose yearly data frame

  # converting dataframe to lists of values in each property/meter combination
  AnnualData <- lapply(YData, function (x) x[!is.na(x)]) # annual data

  # Reinterpolate data to same length for functions that require same length
  annual.series <- reinterpolate(AnnualData, new.length = max(lengths(Annual-
Data)))

  # Creating samples for the testing of number of clusters
  set.seed(250)
  if (nrow(AnnualConsumption)<5001) {
    samplea.series <- annual.series
    sampleAData <- AnnualData
  } else {
    samplea.series <- annual.series[sample(5000)]
    sampleAData <- AnnualData[sample(5000)]
  }

  pc.l2 <- tsclust(samplea.series, k = 2L:10L, distance = "L2", centroid =
"pam", seed = 3247, preproc = zscore, trace = TRUE)

  pc.kshapeClust <- tsclust(sampleAData, preproc = zscore, type = "partitional",
distance = "sbd",
                            centroid = "shape", k =2L:10L, seed = 93, trace = TRUE)

  names(pc.l2) <- paste0("k",2L:10L)
  names(pc.kshapeClust) <- paste0("k",2L:10L)

  CVIl2 <- sapply(pc.l2, cvi, type="internal")
  CVIKshape <- sapply(pc.kshapeClust, cvi, type="internal")

  write.csv(CVIl2, "./CVIOutput/PartitionalTable.csv", na="") #output to csv
  write.csv(CVIKshape, "./CVIOutput/KShapeTable.csv", na="") #output to csv

}

clusteringanalysis <- function(strFilePath, strFolder, NumClusters) {

  list.of.packages <- c("ggplot2", "zoo", "openxlsx", "plyr", "dtwclust" )
  new.packages <- list.of.packages[!(list.of.packages %in% installed.pack-
ages()[,"Package"])]
  if(length(new.packages)) install.packages(new.packages)

  library(ggplot2)
  library(zoo)
  library(openxlsx)
  library(plyr)
  library(dtwclust)
```

```
  #set work directory
  setwd(strFolder)

  #consumption data
  AnnualConsumption <- read.xlsx(strFilePath,sheet = 1, startRow = 1, colNames =
TRUE, rowNames = TRUE)
  colnames(AnnualConsumption) <- as.Date(as.integer(colnames(AnnualConsump-
tion)),origin="1899-12-30") #rename columns to dates

  # Transposing the dataframes for analysis
  YData <- as.data.frame(t(AnnualConsumption))  #transpose yearly data frame

  # converting dataframe to lists of values in each property/meter combination
  AnnualData <- lapply(YData, function (x) x[!is.na(x)]) # annual data
  TSData <- rbind.fill(lapply(AnnualData, function(x) as.data.frame(t(x))))

  #write the output of processed annual data
  write.csv(TSData, "./ProcessedAnnualConsumption.csv", na="") #output to csv

  # Reinterpolate data to same length for functions that require same length
  annual.series <- reinterpolate(AnnualData, new.length = max(lengths(Annual-
Data)))

  Kshape <- tsclust(AnnualData, preproc = zscore, type = "partitional", distance
= "sbd", centroid = "shape", k =NumClusters, seed = 777, trace = TRUE)

  KShapeTable <- as.matrix(Kshape@cluster)
  write.csv(KShapeTable, "./ClusterOutput/KShapeClusters.csv", na="") #output to
csv

  #### Fuzzy c-means Clustering ####
  FuzzyC <- tsclust(annual.series, type = "f", k= NumClusters, preproc = zscore,
distance = "L2", seed = 36, trace = TRUE, control = fuzzy_control(iter.max =
1000L))

  # Write output to CSV
  FuzzyTable <- as.matrix(FuzzyC@fcluster)
  write.csv(FuzzyTable, "./ClusterOutput/FuzzyClusters.csv", na="") #output to
csv

  # Plot the series
  BERT.graphics.device("KShape")
  plot(Kshape, type="centroid", labs.arg = list(title = "k-shape Clusters (z-
normalised)", x = "Time", y = "Normalised series"))
  BERT.graphics.device("Fuzzy")
  plot(FuzzyC, type = "centroids", labs.arg = list(title = "Fuzzy c-means Clus-
ters (z-normalised)", x = "Time", y = "Normalised series"))

}
```