

Household water end-use identification in the presence of rudimentary data

by

Bettina Elizabeth Meyer  
(née Botha)



Dissertation presented for the degree of Doctor of Philosophy in the Faculty of Engineering  
at Stellenbosch University

Supervisor: Prof H.E. Jacobs

March 2021

## Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third part rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

This dissertation includes 4 original papers published in peer-reviewed journals or conference proceedings and 3 unpublished publication. The development and writing of the papers (published or unpublished) were the principal responsibility of myself and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors.

---

B E Meyer

March 2021

---

Date

## Abstract

Detailed and accurate information regarding residential water use is essential for targeted water demand management (WDM) strategies and water security, and yet most utilities have limited information regarding household water demand at end-use level. Flow trace analysis software has been successfully deployed to disaggregate household water end-uses from high resolution smart meter data in various earlier studies, however, water utilities from a range of socio-economic settings, especially in developing countries, typically measure household water consumption data at resolutions too low for commercially available disaggregation software. The aim of this research was to identify and develop methods to evaluate and quantify household water demand at an end-use level, in the absence of high resolution data.

Numerous end-use studies were conducted using direct methods (i.e. water meters) and indirect methods (e.g. temperature loggers) to record residential water demand at the point of entry and at the point of use. Valuable information was extracted from the recorded time series data by applying the automated temperature analysis algorithm, with end-use event durations and event frequencies being derived from the results. Numerous benefits and limitations regarding temperature loggers as indirect method were addressed as part of this research.

Additionally, measurements were taken at a single entry point on a residential property. An automated end-use extraction tool (PEET) and classification model (WEAM) were developed to identify and categorise residential end-use events from a rudimentary data set. Despite the coarse resolution of the measured data making it impossible to separately classify background leakage and relatively low flow water use events (consequently categorising both instances as minor events), PEET was able to extract notable end-use events from the study site. The WEAM model was able to correctly classify the notable end-use events into indoor use and outdoor use categories.

The methods and models proposed as part of this research could enable utilities to broadly classify household end-use events as being indoor or outdoor, without relying on pre-trained models. By applying the developed models on rudimentary data sets, water managers could improve water security through better informed demand management programmes.

## Acknowledgements

To my husband, Marnu Meyer, and parents, Joan and Willem Botha, thank you for your endless support, patience, love and motivation. I appreciate you.

I would also like to sincerely thank my sister and brother, Jutami and Ockert Augustyn, my direct family, Ina Smith and Mercia Botha, my extended family, the Meyers and Tasje van Jaarsveld, friends and loved ones for their continued support, reviewing my work and being soundboards.

I would like to thank Prof. H. E. Jacobs for providing valuable advice, direction and positive reassurance that led to the successful completion of this work.

I am grateful for the opportunity to have worked with Prof Ilemobade, for his valuable insights and advice.

I am grateful to Johannesburg Water for their financial support contributing in part to my postgraduate studies.

## Table of Content

Declaration.....	i
Abstract.....	ii
Acknowledgements.....	iii
Table of Content .....	iv
List of Tables .....	vi
List of Figures .....	vii
List of Symbols .....	viii
Abbreviations and acronyms .....	ix
Chapter 1. Introduction .....	10
Chapter 2. Meyer B.E. and Jacobs H.E. 2019. Garden irrigation as household end-use in the presence of supplementary groundwater supply. <i>Water SA</i> , 45(3): 447-455 .....	16
Chapter 3. Meyer B.E., Jacobs H.E , Biggs B. and Ilemobade A. 2017. Analysis of shower water use and temperature at a South African university campus. 15th International Computing & Control for the Water Industry Conference, CCWI 2017, The University of Sheffield Journal contributions, F64.....	37
Chapter 4. Meyer B.E., Jacobs H.E , Terblanche U. 2018. Clothes washing as household end-use: comparison of different appliance models in view of expected water savings. 1st International WDSA/CCWI Joint Conference, Kingston, Ontario, Canada. Conference Proceedings. Vol. 1. 2018, F020.....	47
Chapter 5. Assessing different measurement methods, data resolutions and end-use event characteristics .....	57
Chapter 6. Meyer B.E., Jacobs H.E , and Ilemobade A. 2020. Extracting household water use event characteristics from rudimentary data. <i>Journal of Water Supply: Research and Technology - Aqua</i> , 69(4): 387–397.....	75
Chapter 7. Meyer B.E., Nguyen K., Beal C.D., Jacobs H.E. and Buchberger S. Submitted. Household water end-use classification model for implementation on coarser data: toward improving the benefits of lower resolution end-use data sets. <i>Journal Water Resources Planning and Management</i> , paper reference number WRENG-4842.....	91

Chapter 8. Meyer B.E., Jacobs H.E , and Ilemobade A. Submitted. Classifying household water use into indoor and outdoor use from a rudimentary data set – A case study in Johannesburg, South Africa. Journal of Water, Sanitation & Hygiene for Development .....	113
9 Discussion.....	131
10 Conclusion.....	140
References .....	144
Appendix A: Declarations of candidate and co-authors .....	147

## List of Tables

Table 1.1. Summary of data sets used during this research study .....	14
Table 2.1. Measurement methods for water end-use data collection.....	21
Table 2.2. Statistical distribution descriptions.....	26
Table 2.3. DFI model input parameters for study site in autumn .....	30
Table 3.1. Residence B flow rates and pressures .....	42
Table 3.2. Statistical parameters for the shower flow rates .....	42
Table 3.3. Statistical parameters for the shower durations .....	43
Table 4.1. Top loader washing machine sales and water consumption.....	50
Table 4.2. Front loader washing machine sales and water consumption .....	50
Table 4.3. Summary of washing cycle frequency per household .....	51
Table 5.1. End-use statistical distribution parameters .....	62
Table 6.1. Collected data set format in MS Excel .....	78
Table 6.2. MS Excel format of PEET output .....	81
Table 6.3. List of variables.....	82
Table 6.4. Comparison between evaluated time gaps settings.....	85
Table 7.1. End-use events in each WEAM data subsets .....	97
Table 7.2. Training data subsets .....	99
Table 7.3. Training data subsets performance on SVM fit .....	102
Table 7.4. Confusion matrix on classification results (train set).....	104
Table 7.5. Model performance on the training data set .....	104
Table 7.6(a). WEAM confusion matrix.....	106
Table 7.6(b). WEAM performance metrics on the test set.....	106
Table 7.6(c). WEAM performance on total event volume .....	106
Table 8.1. Residential indoor and outdoor water consumption .....	115
Table 8.2. Theoretical demand estimates for residences properties in Johannesburg .....	118
Table 8.3. Dates with reported water use from meter measurements .....	121
Table 8.4. Classification of end-use events.....	121
Table 8.5. End-use event classifications and household information .....	122
Table 8.6. Correlation between PPH and indoor use as proportion of total demand .....	123
Table 8.7. Correlation between property size and outdoor use .....	123
Table 8.8. Comparison between billing data and smart meter data.....	125
Table 9.1(a). Indirect measurement methods implemented in this study.....	134
Table 9.1(b). Direct measurement methods implemented in this study .....	134

## List of Figures

Figure 2.1. Measured pipe wall and derived baseline temperatures.....	24
Figure 2.2(a). Temperature difference between pipe wall and baseline temperatures.....	24
Figure 2.2(b). Filtered temperature differences for pumping duration .....	24
Figure 2.3. Cumulative distribution function for garden irrigation duration .....	28
Figure 2.4. Cumulative distribution function for garden irrigation event frequency.....	28
Figure 2.5. Cumulative distribution function for groundwater flow intensity .....	29
Figure 2.6. Cumulative distribution function of the average daily garden irrigation event volume .....	31
Figure 3.1. Identifying shower events.....	41
Figure 3.2. Cumulative distribution function (CDF): shower flow rate .....	42
Figure 3.3. Shower event frequency.....	43
Figure 3.4. Cumulative distribution function (CDF): shower durations.....	43
Figure 3.5. Cumulative distribution function (CDF): water consumption per shower event ..	44
Figure 4.1. CDF washing cycles per household per annum .....	52
Figure 4.2. CDF of average washing machine water consumption based on model results...	53
Figure 4.3. CDF of average washing machine water consumption based on field experiments .....	54
Figure 4.4. CDF comparison of model results and field experiments.....	55
Figure 5.1. Flow pattern of a typical shower event at different recording resolutions .....	71
Figure 5.2. Flow pattern of various minor tap events at different recording resolutions .....	72
Figure 6.1. Two single events with a 45 s time gap .....	79
Figure 6.2. Schematic of a meter spike/lagged reading .....	80
Figure 6.3. Schematic of end-use extraction tool procedure .....	81
Figure 6.4. Schematic of low flow events .....	82
Figure 6.5. Schematic example of lumping multiple events.....	83
Figure 6.6. End-use volume for the three different time gap settings .....	85
Figure 6.7. End-use duration for the three different time gap settings .....	86
Figure 6.8. End-use intensity for the three different time gap settings.....	86
Figure 7.1. Ellipsoidal decision surface .....	100
Figure 7.2. Example of reducing square error using the sigmoid function .....	101
Figure 7.3. Error in the EDS model (Equation 7.1) as a function of $D_r$ .....	103
Figure 7.4. Comparing the ROC curves of the classification models .....	105
Figure 8.1. Monthly outdoor consumption at Homes H03, H08, H11.....	125
Figure 9.1. Proposed data resolution categories.....	132

## List of Symbols

$a$	minimum
$b$	maximum
$D$	event duration
$d_o$	event start time
$d_e$	event end time
$F$	event frequency
$I$	event intensity (flow rate)
$m$	most likely value
$R^2$	coefficient of determination
$\Delta r$	difference between two consecutive water meter pulses
subscript $p$	best-fit probability distribution
subscript $WM$	washing machine event
subscript $r$	ellipsoid principal semi-axes
$t$	date-time stamp
$\Delta t$	absolute difference in temperature
$v_o$	event start water meter reading
$v_e$	event end water meter reading
$V$	total event volume
$\alpha$	significance level
$\beta$	regression coefficients
$\mu$	mean
$\rho$	density of the liquid
$\sigma$	standard deviation
$\lambda$	average

## Abbreviations and acronyms

AUC	Area Under the receiver operating Curve
CCWI	Computer and Control for the Water Industry
CDF	Cumulative Distribution Function
EDS	Ellipsoidal Decision Surface
JW	Johannesburg Water
GOF	Goodness-Of-Fit
GAP	Groundwater Abstraction Point
PEET	Python End-use Extraction Tool
PPH	People Per Household
PRP	Poisson Rectangular Pulse
REUM	Residential End-Use Model
RF	Random Forests
ROC	Receiver Operating Curve
SIMDEUM	SIMulation of water Demand, an End-Use Model
SVM	Support Vector Machine
TGS	Time Gap Setting
UBL	Upper Bound Limit
WEAM	Water End-use Apportionment Model
WDM	Water Demand Management
WDSA	Water Distribution System Analysis

# Chapter 1.

## Introduction

### 1.1 BACKGROUND AND MOTIVATION

In view of the growing population and high urbanisation rates, especially in developing countries, the need to accurately estimate water demand is now more crucial than ever. Understanding water demand at individual end-use scale is useful for urban water planners to develop water management schemes and to provide an effective approach to preserve water resources (Jorgensen et al. 2013). It is valuable for a water utility to have detailed and accurate information regarding water use, allowing for targeted water demand management (WDM) strategies as well as economic incentives (Nguyen et al. 2013). Household water demand is typically divided into indoor and outdoor use, and then further categorised into single end-use components. Household end-uses could be classified as being indoor or outdoor, based on the physical location of the water use event in-and-around the home. The most notable indoor end-uses include the shower, bath, washing machine, toilet, dishwasher and indoor taps (Nguyen et al. 2018). Garden irrigation, swimming pool, car washing and outdoor tap are the most notable outdoor end-uses at a household level (Beal and Stewart 2013). Note that leaks, while not strictly an “end-use”, are both indoor and outdoor events (Britton et al. 2013).

End-use water demand models, such as SIMulation of water Demand, an End-Use Model (SIMDEUM) presented by Blokker et al. (2010) and Residential End-Use Model (REUM) presented by Jacobs and Haarhoff (2004), rely on various parameters to populate the models. These parameters include end-use characteristics, such as event duration, event intensity (flow rate), event volume, frequency of use, and time of day. The quality of the input data used for end-use modelling is of the highest importance to assure accurate results of the analysis (Van Zyl et al. 2003). Researchers in various developed regions have conducted accurate (high spatial and temporal resolution) household end-use studies. Identifying individual end-uses was pioneered by Buchberger and Wu (1995), and recent subsequent studies include Beal et al. (2011), DeOreo et al. (2011), Beal and Stewart (2013), Arregui (2015), Nguyen et al. (2013, 2018) and Pastor-Jaboloyes et al. (2018). Some of these studies involve the use of expensive data logging technology (smart meters) paired with flow trace analysis software (end-use classification methods). A summary of completed water end-use studies, presented by Beal and Stewart (2011), demonstrates that a sub-10 second metering resolution for data capturing, with pulse measurements of less than 0.1 L/pulse, is needed for end-use disaggregation and classification. Water meters with such high recording resolutions are uncommon in practice.

Water utilities generally measure household water consumption data at coarser resolutions, due to the resource intensive and costly nature of higher resolution smart meters (Ilemobade et al. 2018). Existing water meters are commonly set to record water consumption at 1 L/pulse (Nguyen et al. 2013). End-use studies based on rudimentary end-use data include, Cole and Stewart (2013) and Pretorius et al. (2019). Although these studies provide insight into anomalous events, such as peak hour demand and leakage, they have not reported on specific household end-uses and end-use event characteristics. Cominola et al. (2018) investigated the trade-off between data resolution and end-use identification and concluded that a sub-minute measuring frequency is needed for end-use classification. The degree to which household water consumption at end-use level can be identified with a higher volume per pulse setting on a meter (e.g. 1 L/pulse) was the focus of this study. Knowledge regarding indoor or outdoor water consumption can be valuable, as it allows for the evaluation of water saving potential for different WDM strategies.

A number of studies have proposed indirect flow sensing approaches for characterising and quantifying household water end-use events. These indirect approaches measure either at the point of use, or at a single entry point at a property and include ultrasonic- (Makwiza and Jacobs 2017, Fogarty et al. 2006), thermal- (Massuel et al. 2009; Nel et al. 2015a, Nel et al. 2015b), vibration- (Evans et al. 2004, Kim et al. 2008, Sterne 2019), pressure- (Froehlich et al. 2009, 2011, Larson et al. 2012), motion-sensing (Srinivasan et al. 2011), or a combination of these (Pirow et al. 2018, Cloete 2017). Indirect approaches are attractive for investigating household water end-use events because of the relatively lower cost when compared to conventional or smart water meters and the methods are generally unobtrusive. The data obtained from indirect flow metering are more detailed and reliable compared to consumer surveys. Thus, in the absence of high resolution smart meters, indirect flow sensing approaches could be explored to better understand water consumption at end-uses level.

## **1.2 CONTEXT**

Numerous former end-use studies have contributed significantly to understanding household water demand at end-use level. Household end-uses were first recorded at point of use by Edwards and Martin (1995) in the UK, using water meters at each end-use point. More recent studies developed indirect flow sensing approaches for application at point of use, some of which were employed in this research. No new measurement or sensing methods were developed in this study.

Extracting and identifying end-uses from high resolution water meter data at entry to the property (e.g., measured at the consumer meter) was pioneered by De Oreo et al. (1996) and Mayer et al. (1999). End-use models were developed in parallel. Household end-uses were first described as rectangular pulses and modelled theoretically by Buchberger and Wu (1995). These early end-use studies employed high resolution smart meters, which are not commonly available, especially not in developing countries such as South Africa.

Water consumption data are often only available in practice at a reduced temporal or spatial resolution, which was referred to as rudimentary data in this study. This research focussed on extracting knowledge from rudimentary end-use data.

### **1.3 RESEARCH PROBLEM**

Water consumption data would normally be collected with high resolution smart meters in order to disaggregate end-uses from the recorded data in end-use studies. Specialised software, based on the high resolution inputs, would be used to disaggregate and classify the data into single end-use events. Developing countries have identified the need for smart meters, however, studies involving such relatively expensive technology has not yet been deployed (Gupta et al. 2016). Water utilities, from a range of socio-economic settings, generally measure household water consumption data at resolutions too low for commercially available disaggregation software.

This research study set out to answer the following question: To what extent can measured rudimentary data, in terms of temporal and spatial resolution, be utilised to classify water use at a household level in order to extract useful knowledge? This research question addressed the issue of limited household water end-use information available to utilities, because of the coarser end-use data captured from existing water distribution infrastructure.

### **1.4 AIM AND OBJECTIVES**

The aim of this research was to identify and develop methods to evaluate and quantify household water demand at an end-use level, in the absence of high resolution data. This was achieved by dividing the main goal into key objectives, namely:

- Conduct a thorough literature review related to all aspects of end-use disaggregation and classification;
- Research and test the suitability of indirect flow sensing to identify and quantify residential end-use events;
- Evaluate the physical characteristics of single end-use events in South Africa;
- Record residential end-use data and determine the trade-off between rudimentary data and end-use classification;
- Develop an apportionment model to classify household water demand into indoor use and outdoor use. The apportionment model must be applicable on coarser data sets;
- Employ the developed apportionment model in a South African case study and conduct empirical analyses of actual water use using measured rudimentary data.

## **1.5 SCOPE AND LIMITATIONS**

This study focussed on the most notable household end-uses. Indoor end-uses included in this study are the shower, washing machine, dishwasher, toilet and indoor tap. The only outdoor end-use assessed as part of this research was garden irrigation. Outdoor use was assumed to be predominantly driven by garden irrigation, as suggested in published literature (Roberts 2005). This assumption was considered acceptable, as all the homes in the study samples had gardens.

In addition to smart metering at the consumer supply connection, multiple indirect flow sensing approaches were considered for data collection. The indirect flow sensing approaches considered were relatively small in size, unobtrusive, rugged, accurate, and inexpensive. Ultimately, temperature loggers were implemented during this study, due to its costs and availability at the time of the study.

Per definition, it was impossible to distinguish between minor (low flow) events in this research. Background-leakage flows in the plumbing system, minor leaks at the point-of-use (e.g. a dripping tap) and relatively low flows from valid water use events (e.g. filling a 0.2 L glass with water), would appear similar. Consequently, all these events were categorised as minor events as part of this study. The classification model developed as part of this research made no provision for independent variables describing the region per se, such as climatological- or socio-economical inputs. The model was limited to domestic (residential) water use and focussed on three end-use event identifying characteristics, namely event duration, event volume, and event intensity.

## **1.6 DATA SAMPLING**

Data used for this research were obtained from various sources and are discussed in detail in the respective chapters. The specifics and characteristics of each data set are summarised in Table 1.1. A combination of high and low resolution data measured at different locations were purposefully introduced to meet the research objectives.

## **1.7 BRIEF CHAPTER OVERVIEW**

This manuscript includes a combination of published articles, unpublished articles and written chapters, in line with the requirements for a doctoral dissertation according to the published rules and policies' guidelines (Section 2.1.2 updated 2019) of Stellenbosch University. The page numbers, table numbers and figure numbers of the published works were re-formatted in a consistent manner, with no changes made to the content, as required by Stellenbosch University. The tables and figures were formatted according to the requirements of the respective journals they were published in.

Table 1.1. Summary of data sets used during this research study

Characteristics of data set	Targeted end-use				
	Garden irrigation	Shower	Washing machine	All end-uses (Australia)	All end-uses (South Africa)
Recording Period	April 2016 to May 2016 (22 days)	February 2017 to March 2017 (10 days)	August 2017 to September 2017 (20 days)	2010-2012	September 2016 to January 2018 (217 days)
Region	Cape Town, South Africa	Stellenbosch, South Africa	Stellenbosch, South Africa	Gold Coast, Australia	Johannesburg, South Africa
Measurement method	Thermochron iButton temperature loggers	Thermochron iButton temperature loggers	Mechanical water meter	Smart meters	Sensus iPERL water meters
Installation location	Point of use	Point of use	Point of use	Central location for entire house	Central location for entire house
Data resolution	Temperature measurements at 2 min intervals	Temperature measurements at 1 min intervals	Flow measurements per event; 1 L/pulse	Flow measurements every 5 s; 0.014 L/pulse	Flow measurements every 15 s; 1 L/pulse
Sample Size	10 homes	2 university residences	1 university residence	252 homes	63 homes
Total number of end-uses	68	759	54	200,266*	1,107,547*

\*Note: Estimated from household consumption time-series data

The manuscript consists of 10 chapters, of which two are published articles in ISI-listed journals, namely *Water SA* and the *Journal of Water Supply: Research and Technology – Aqua*. Two articles were published as international conference proceedings (CCWI and WDSA/CCWI) and two articles have been submitted for review and possible publication in the *Journal Water Resources Planning and Management* and the *Journal of Water, Sanitation and Hygiene for Development*. The contribution made by each co-author for the respective articles were documented and is presented in Appendix A.

The dissertation starts by employing indirect flow sensing approaches to measure and analyse the physical characteristics of household water end-uses. Chapter 2 focussed on the most notable outdoor use, garden irrigation, in an unrestricted scenario. The novelty of the contribution made by Meyer and Jacobs (2019) is that the residential consumption of water from boreholes or well points, for garden irrigation use, has not previously been reported on. Chapter 3 and Chapter 4 focussed on two of the most notable indoor end-uses, namely the shower and washing machine (clothes washer), respectively. The indirect flow sensing approach was employed on showers, and the findings were presented by Botha et al. (2017).

This was the first time temperature loggers were deployed on showers (at the point of use) at residential properties to measure indoor hot water consumption. A direct measuring approach was utilised to measure and analyse the physical characteristics of washing machines and was presented by Botha et al. (2018). Both papers presented in Chapter 3 and Chapter 4 contribute to the understanding of water demand at individual end-use scale.

The characteristics of the remainder of the notable household end-uses are discussed in Chapter 5. Chapter 5 presents an in-depth view of the physical characteristics of all household end-uses, and is based on research conducted as part of this dissertation (Chapter 2 through Chapter 4), as well as research from previously published literature. The chapter also reports on the different water demand measurement methods and analysis software and discusses the trade-off between data resolution and practicality of use. Based on the findings from Chapter 5, a case study was conducted in the City of Johannesburg, South Africa. The physical characteristics of household end-uses were extracted from a rudimentary data set, to ultimately conduct empirical analyses of actual water use. The novel automated end-use extraction tool developed by Meyer et al. (2020), termed PEET, is presented as Chapter 6. The extraction tool presented in the published article has the potential to improve the usefulness and value of rudimentary data, collected from household water meters.

In order to disaggregate extracted end-use into indoor use and outdoor use, a classification model was required. The newly developed apportionment model was described by Meyer et al. (submitted) and is presented as Chapter 7 in this manuscript. The automated extraction tool (PEET) and the apportionment model (WEAM) are the first methods developed that can be employed on coarser data sets in order to classify household data. The next step was to extend WEAM to categorise household water use from a Johannesburg case study.

The case study is presented in Chapter 8, and was submitted for review and possible publication in an ISI journal. The manuscript ends with a comprehensive discussion (Chapter 9) and conclusion (Chapter 10), which comments on the novelty of this dissertation and suggests topics for future research. The dissertation is structured in such a way that the references for each published chapter is included at the end of each chapter. References for unpublished chapters are collectively listed at the end of the manuscript.

## Chapter 2.

### Garden irrigation as household end-use in the presence of supplementary groundwater supply

Bettina Elizabeth Meyer<sup>1</sup>; Heinz Erasmus Jacobs<sup>1\*</sup>

1. Department of Civil Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa, [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za), [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za)

\*Corresponding author: [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za)

Reproduced from Water SA, 45(3), 447-455.

Available online: <http://dx.doi.org/10.17159/wsa/2019.v45.i3.6741>

#### ABSTRACT

Garden irrigation is a significant and variable household water end-use, while groundwater abstraction may be a notable supplementary water source available in some serviced residential areas. Residential groundwater is abstracted by means of garden boreholes or well points and - in the study area - abstracted groundwater is typically used for garden irrigation. The volume irrigated per event is a function of event duration, frequency of application and flow rate, which in turn are dependent on numerous factors that vary by source - including water availability, pressure and price. The temperature variation of groundwater abstraction pipes at residential properties was recorded and analysed as part of this study in order to estimate values for three model inputs, namely, pumping event duration, irrigation frequency, and flow rate. This research incorporates a basic end-use model for garden irrigation, with inputs derived from the case study in Cape Town, South Africa. The model was subsequently used to stochastically evaluate garden irrigation. Over an 11-d period, 68 garden irrigation events were identified in the sample group of 10 residential properties. The average garden irrigation event duration was 2 h 16 min and the average daily garden irrigation event volume was 1.39 m<sup>3</sup>.

**Keywords:** garden irrigation, end-use, groundwater, residential water demand

## INTRODUCTION

Residential water consumption is typically categorised into indoor end-uses and outdoor end-uses. Previous studies suggest outdoor use to be seasonal, driven by weather-related variables, whilst indoor use has been found to be relatively constant (Fisher-Jeffes et al., 2015). Outdoor use is also considered more unpredictable than indoor use (Hemati et al., 2016). Howe and Linaweaver (1967), in an early study of residential water demand, reported on the inelastic nature of indoor water use versus the elastic nature of outdoor use, meaning that outdoor use was found to be more sensitive to a change in inputs than indoor use. Jacobs and Haarhoff (2007) used elasticity and a sensitivity parameter to identify pan evaporation, an irrigation factor, lawn surface area (lawn size) and the vegetation crop factor (lawn grass genotype) as the most notable parameters when modelling outdoor water use.

Various parameters describing outdoor use have received attention as part of earlier work, including garden irrigation (Beal et al., 2011), lawn size (Runfola et al., 2013), swimming pools (Fisher-Jeffes et al., 2015), and water use from the outside tap (Makwiza and Jacobs, 2017). Household water leakage was also addressed in earlier work (Britton et al., 2008; Lugoma et al., 2012). The most notable outdoor end-use in an unrestricted scenario is garden irrigation. Garden irrigation is often reported as a notable part of the total per-capita consumption (Willis et al., 2011). It is unsurprising that outdoor use is the primary target during water restrictions, with earlier studies reporting on reduced water use during water restrictions, mainly due to reduced outdoor use (Jacobs et al., 2007).

Despite the attention to various facets of outdoor use in earlier work, end-use studies have paid limited attention to water supply from supplementary household water sources (Nel et al., 2017). This research focuses on modelling garden irrigation as an end-use in an unrestricted scenario, where groundwater was abstracted from privately owned groundwater abstraction points (GAPs) as supplementary water source. Residential GAPs include garden boreholes and relatively shallow wellpoints.

Consumers may turn to alternative non-potable water sources such as rainwater, groundwater or greywater during stringent water restrictions. The quality of these resources typically limits application to nonpotable uses, such as garden irrigation (MacDonald and Calow, 2009). According to Nel et al. (2017), groundwater use is the most notable supplementary source in terms of the expected supply volume. Many privately owned GAPs are in use across South Africa, with at least one notable case study in the Cape Town region (Wright and Jacobs, 2016). Monitoring of household groundwater abstraction in South Africa is poor and published information regarding yield, flow rate, and/or the pumping event duration of household GAPs is limited.

## **Garden irrigation as outdoor end-use**

The contribution of garden irrigation to the total household water use varies by season (Parker and Wilby, 2013) and also varies from country to country and even from house to house. Garden irrigation tends to be higher during dry, hot seasons, and increases with reduced rainfall (Jacobs and Haarhoff, 2004; Parker and Wilby, 2013) and increased maximum daily temperatures (Rathnayaka et al., 2015), for example. The garden event duration and number of occurrences are contingent on the method of irrigation. Roberts (2005) identified three main irrigation methods, namely, hand-held hose, manual sprinkler and automated sprinkler. The latter contributed most to garden irrigation volumes from the end-use study conducted by Roberts (2005) in Australia. The same three irrigation methods were found in the study area during this research.

Literature includes various reports of garden irrigation expressed as a percentage of the total household water demand, in order to explain the significant contribution of garden irrigation to total household water use. The perceived percentage of residential water demand used for garden irrigation in South Africa, based on an annual average, was reported to vary between 0% and 70% (Veck and Bill, 2000). More recent end-use studies conducted in South Africa reported the percentage of average annual household water demand ascribed to garden irrigation as 40% to 60% (Du Plessis and Jacobs, 2015) and 58% (Du Plessis et al., 2018) in different South African study samples.

End-use studies conducted in other parts of the world also report a wide range of values expressing garden irrigation as a percentage of the total household water use. In Australia, the percentage of household water demand used for garden irrigation ranges from 5% (Beal et al., 2011) to 54% (Loh and Coghlan, 2003). Arbon et al. (2014) reported a strong seasonal impact in Adelaide, Australia, with a 2013 winter mean of 153 L/person per day increasing to 498 L/person per day in the summer of 2013/14. The average annual use was 245 L/person per day and 289 L/person per day in 2013 and 2014 respectively, that could indicate a garden irrigation contribution of 50% to 70% of the total annual household demand; a significant shift in the diurnal pattern was noted, with an afternoon peak more prominent during summer. A lower outdoor use contribution of 15% was reported at high-income detached houses by Ghavidelfar et al. (2018) in Auckland, New Zealand. Wasowski (2001) conducted an end-use study in the United States of America and stated that between 40% and 60% of annual average residential water demand is attributed to garden irrigation.

## **Rationale**

Suburban households in the case study area of Cape Town, South Africa, are accustomed to a reliable supply of potable water from the pressurised water distribution system. However, the rising block-based water tariff was relatively high and, also, outdoor water use from the distribution system was banned during water restrictions in the study area - for the period

June 2017 to December 2018. Consumers subjected to emergency water restrictions turned to alternative sources of water to maintain gardens during this 18-month period. Little is known about garden water use by consumers with access to groundwater from garden GAPS; the restrictions provided the opportune time to investigate the matter. The main challenge in this study was to obtain data regarding actual groundwater use by private homeowners, who were often reluctant to share any information regarding uncontrolled and unmetered household water sources.

### **Research problem**

An end-use model was needed to assess garden irrigation in relation to supplementary groundwater supply, while populating the model with data that could realistically represent the key unknowns.

### **METHODS**

Parameters describing the quantity and quality of household groundwater abstraction form important inputs to end-use models of household water use. Groundwater use for garden irrigation was modelled in this study, with inputs based on measured values. Data were collected from a relatively small case study site in Cape Town, South Africa. Direct measurement of groundwater abstraction was not considered feasible and an alternative method to assess the volume of groundwater abstracted for garden irrigation was employed. Groundwater pumping event start times and durations were derived from continuously recorded pipe wall temperatures at each of the 10 residential properties. Ad hoc volumetric measurements were subsequently conducted at each home to gain insight into flow rates at each study home. Stochastic end-use modelling was employed to estimate the expected garden irrigation event volume of the 10 properties in the research sample. Based on information obtained during the site survey, garden irrigation volume was considered to be equal to the groundwater abstracted from GAPS for all homes in the case study.

### **Overview of residential end-use models**

The focus of this study was on modelling water demand at a small spatial scale of single residential homes - and garden irrigation as a specific end-use of water. Numerous residential end-use models have been developed in the past; however, a model to evaluate garden irrigation in relation to groundwater abstraction as supplementary source has not yet been developed. Some examples of earlier end-use models include the Poisson Rectangular Pulse (PRP) model developed by Buchberger et al. (1996; 2003), the SIMulation of Demand End-Use Model (SIMDEUM) by Blokker et al. (2010) and the Residential End-Use Model (REUM) by Jacobs and Haarhoff (2004). REUM and SIMDEUM incorporate garden irrigation as end-use.

## **Experimental field tests and data analysis**

### ***Study site selection and sample group***

A map of verified residential properties with GAPs in the Cape Town Metropolitan area was developed by Wright and Jacobs (2016). The sample group of 10 homes for this study was based on sub-regions where clustering of GAPs (as reported by Wright and Jacobs, 2016) was observed, followed by personal invitation to participate in the study. Relatively small sample sizes are not unusual for end-use studies. Former end-use studies had sample sizes of 28 homes (Butler, 1991), 16 homes (DeOreo et al., 1996), 37 homes (DeOreo et al., 2001), 21 homes (Buchberger et al., 2003), 12 homes (Heinrich, 2007), 10 homes (Jacobs, 2007) and 10 homes (Mead and Aravinthan, 2009).

The manageable sample size in this study also enabled the authors to inspect individual pump installations for leaks and to conduct follow-up inspections. All the houses in the sample were single residential properties, with property plot sizes ranging from 600 m<sup>2</sup> to 1 400 m<sup>2</sup>. Prominent, well-irrigated gardens and lawns were present at all homes. Two residential properties from the study site each had a swimming pool; however, the homeowners assured that the abstracted groundwater was explicitly used for garden irrigation at the time of this study. The assumption that groundwater supply equalled garden irrigation was thus considered valid for the study sample. The addresses and suburb names of the study homes were omitted for anonymity, in line with ethical requirements.

### ***Data collection methods***

Residential water demand patterns should preferably be obtained by measuring actual water use (Scheepers and Jacobs, 2014); however, empirical investigations involving data collection are often faced with several logistical, time and financial constraints. Various data collection methods were considered for this study in order to collect sensitive information that was needed to assess household groundwater abstraction. A list of empirical measurement methods is presented in Table 2.1, including the key advantages and disadvantages in each case, as well as a reference to earlier application. Each method was categorised in terms of feasibility as it relates to the case study. Two categories were included, namely: (1) considered for this case study and also implemented in the study; (2) considered for this study, but not used.

Table 2.1. Measurement methods for water end-use data collection

Measurement method/device	Gathered information	Advantages	Disadvantages	Literature	Applicability to this study*
Consumer Surveys	Any information (within ethical constraints)	Flexibility, relatively simple to implement	Lower accuracy, ethical restrictions, post-processing of data required	Roberts, 2005; Colvin and Saayman, 2007	1
Temperature recorders (iButtons)	Time stamp, temperature	Non-intrusive, relatively low cost, no plumbing changes needed	Post-processing of data required	Chapmin et al., 2014; Massuel et al., 2009	1
Mechanical water meter (no logger)	Consumption, meter reading data	Accuracy	Manual readings, plumbing changes needed, relatively expensive	Turrall et al., 2005	2
Smart water meter with data logger	Flow rate, pressure, time stamp	Accuracy, automated readings	High cost, plumbing changes needed	Ngunyen et al., 2013	2
Watt-hour meter (electrical)	Time stamp, pump power, on-off state	Non-intrusive, no plumbing required	High cost, electrical changes needed	Massuel et al., 2009	2

\*Note: (1) feasible and implemented; (2) considered, but not-used.

### ***Equipment and temperature recording***

The project plan involved recording pipe wall temperature at case study homes in an unobtrusive way, with no plumbing requirements, a short installation time and relatively low cost. The DS1922 Thermochron Hi Resolution iButton was selected for this study, based on the relatively small size, ruggedness, accuracy, cost, and availability. The iButtons were used in this study to measure the variation in temperature of the groundwater pump delivery pipe - that is the delivery pipe of the GAP pump supplying water directly for garden irrigation. The temperature variations were subsequently used to assess water use events by determining the event duration of groundwater pumping, start and stop times.

The iButtons were preconfigured to set the start time and sample rate. ColdChain ThermoDynamics software was used for preconfiguration and to extract and save the recorded data. All the iButtons were programmed to have a sampling rate of 2 min, which was considered sufficient when compared to the relatively long events. The period of 2 min was the shortest interval available when programming the equipment. The iButtons were synchronised to start at the same time on the same date. The internal iButton memory allowed for a total recording duration of 11 d and 9 h (sample count of 8 192 records per iButton). After the iButtons were activated and before the specified start time, the iButtons were installed on the outside wall of the outlet pipe, using adhesive electrical tape. Each GAP was equipped with two synchronous iButtons to record temperature in parallel. The sample included 10 homes and data were recorded during April and May 2016. Subsequently the total data set included 110 test days, representing 11 actual calendar days for each of the 10 homes.

The iButtons were placed in three different environments (A, B, and C). Each environment type was linked to an installation that affected the temperature changes of the iButtons differently. In Environment A the pump and outlet pipes were located in an enclosure that was not exposed to any sunlight. A typical Environment A would be described as a well-insulated concrete pump house with an access door. Due to the insolation, the ambient temperature fluctuation within the enclosure was moderate. Environment B would have the pump and outlet pipes protected from direct sunlight and precipitation by means of a four-walled, wooden or steel enclosure. Access to the equipment was provided via a removable roof. The shape and size of the enclosure is similar to that of a typical medium-sized doghouse. Environment B was found to be relatively similar to Environment A in terms of temperature fluctuation within the enclosure. In Environment C, the pump and outlet pipes were exposed to direct sunlight and therefore experienced more notable ambient temperature changes compared to the other two environments. The sample group had two GAPs located in an Environment A, six GAPs in an Environment B and two GAPs in an Environment C.

### ***Flow intensity measurements***

The intensity (flow rate) was determined at each GAP, using on-site volumetric measurements. The measurement entailed filling a container with water at the endpoint of the irrigation pipe. The container was filled for 45 s and subsequently weighed. The container was weighed pre- and post-fill and the flow rate was calculated. The manageable sample size allowed for a sufficient number of volumetric measurements. Each measurement was repeated 10 times at each GAP, resulting in 100 flow rate measurements. The measurements were used to create a distribution graph, representing the flow intensities for the study site. This method was easily executed, cost effective and caused little disturbance to the residents.

### ***Consumer surveys***

Surveys have been used in the past as an indication of indoor (Blokker et al., 2010) and outdoor water use (Roberts, 2005; Veck and Bill, 2000). A site survey was conducted as part of this study to obtain relevant information regarding water use activities, including identification of the irrigation method (e.g. hand-held hose, manual sprinkler, automatic sprinkler), system connectivity, pump placement environment and water leakage. Although the method of irrigation was documented in the survey, it was not incorporated into the end-use model due to the limited sample size. The site surveys were also used to confirm that the residents used the irrigation systems at the maximum flow rate in each case. The pump flow rate was assumed equal to the garden irrigation flow rate in each case, with no leakage reported at any site.

### ***Identifying pumping events and durations***

Adopting terminology from Jacobs and Haarhoff (2004), the number of events over a given time period was described by using the term 'event frequency', expressed as the number of events per day. The term 'event duration' was used to describe the time lapse from an event start to event end. The recorded pipe wall temperature was analysed in order to identify pumping events and to extract the event frequency and event duration. The procedure was termed temperature variation analysis. Since temperature on each pipe was separately measured and analysed, there were no overlapping events. Each pumping event represented a single garden irrigation occurrence and was characterised by the pump start operation (water flowing through the pipe with corresponding temperature change) and the pump being turned off again. A Visual Basic macro, for implementation in MS Excel, was written to implement the temperature variation calculations. The baseline temperature, needed to identify significant interruptions in the expected graph pattern, was first established. Each interruption (difference between pipe wall temperature and baseline temperature) corresponded to a pumping event.

The daily ambient temperature fluctuated over the study period. The fluctuations varied per installation, because each iButton was placed in a different environment. Consequently, the baseline temperature at each GAP varied. The developed baseline temperature time-series graph at each GAP represented the typical daily temperature cycle per installation. The coefficient of determination ( $R^2$ ) was used as a measure of similarity in shape between the baseline temperature at each GAP, and the temperature measured on the pipe wall. Thus each GAP had a specific baseline temperature corresponding to the particular environment and the ambient temperature of the specific day. After the baseline temperature was developed, pumping events and durations were identified. Figure 2.1 shows an example of the pipe wall temperature measured by an iButton, and the corresponding baseline temperature curve, for one property over a 2-d period. The selected time series shows two events. A pumping event is noticed at about 06:00 on both days.

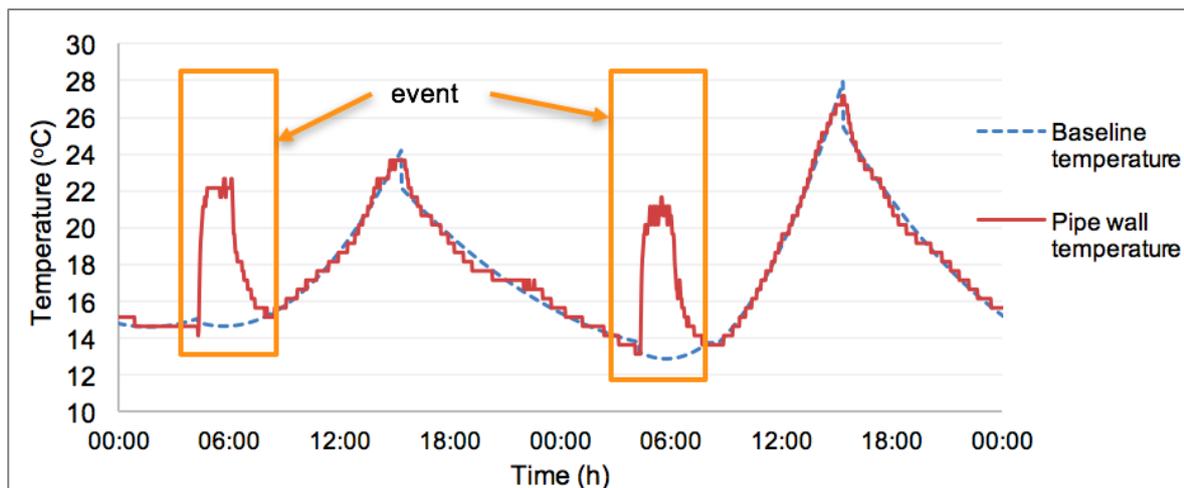


Figure 2.1. Measured pipe wall and derived baseline temperatures

During time steps where the measured pipe wall temperature and baseline temperature deviated notably, water was likely flowing through the pipe (evidence of an event). Firstly, the temperature noise in each environment had to be separated from notable temperature deviations. Figure 2.2(a) shows the difference between the derived baseline temperature and the pipe wall temperature. Temperature noise is clearly visible around the zero y-axis value. In order to automatically detect pumping events, a conditional filter, incorporating a threshold temperature, was applied. The threshold value was determined with consideration for the different environments in which the iButtons were placed, being informed by earlier studies. Massuel et al. (2009) used a threshold of 2.6°C to detect pumping events as part of a study in India. In this study, the threshold was set equal to 2.0°C for Environment A and Environment B and to 3.0°C for Environment C. Implementing the threshold allowed for pumping events to be identified with an algorithm, which is significantly less time consuming than manual interpretation of the recorded data. With reference to the temperature noise visible in Figure 2.2(a), all values not exceeding the threshold temperature were set equal to zero and the result is plotted in Figure 2.2(b). Figure 2.2(b) shows the two individual events in the selected time series, excluding temperature noise below the selected threshold values.

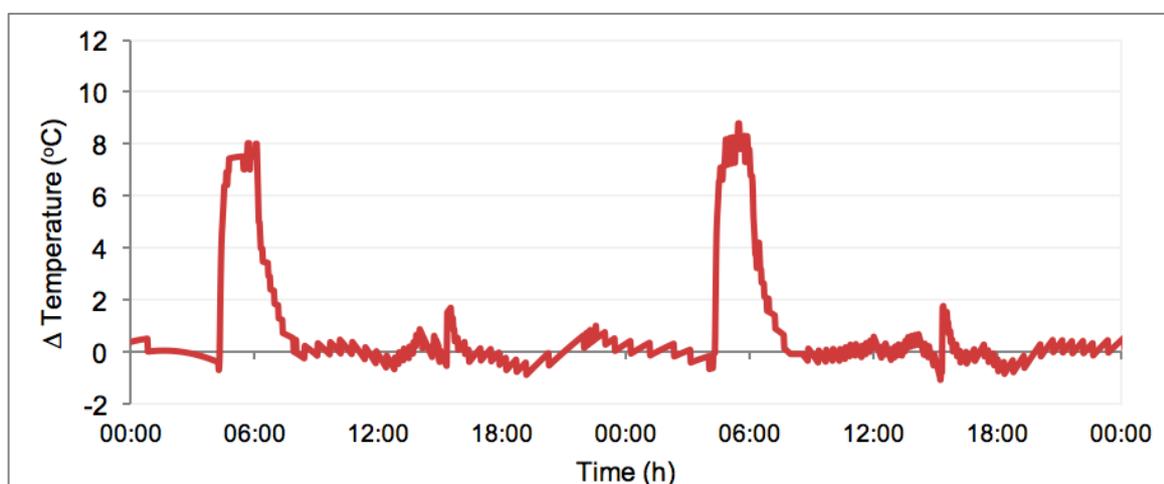


Figure 2.2(a). Temperature difference between pipe wall and baseline temperatures

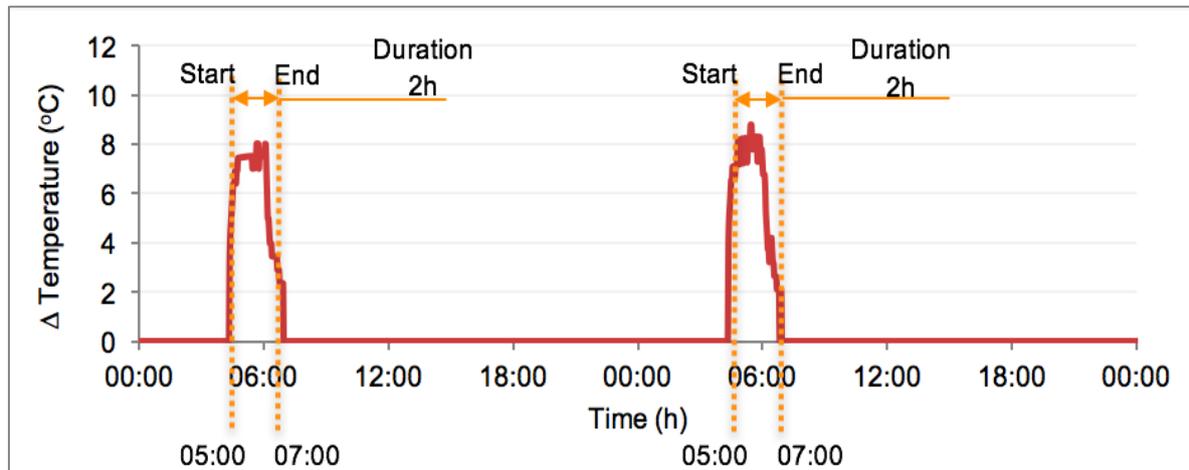


Figure 2.2(b). Filtered temperature differences for pumping duration

All recorded data were analysed in this manner by employing the algorithm. In total, 68 individual events were identified on 59 test days, considering the full data set of 110 test days. Multiple irrigation events per day were detected at only one home. The highest event frequency was 3 events per day, reported only once; 2 events per day were reported 7 times in the full data set at the same home. The limited number of events reported on in this study also allowed for subsequent visual inspection of the temperature difference at each event.

### Basic model structure

A rudimentary model was developed to stochastically determine the average daily volume of groundwater pumped for garden irrigation. The model included three independent parameters (duration, frequency, intensity) and was termed the DFI model. The DFI model adopted notation from the SIMDEUM model developed by Blokker et al. (2010), and is based on the assumption that all the input parameters are independent and statistically distributed random variables.

The DFI model structure, for a single residential property, is described by Equation 2.1:

$$V_p = D_p * F_p * I_p \quad (2.1)$$

where,

$V$  = average daily garden irrigation event volume ( $\text{m}^3/\text{d}$ )

$D$  = event duration (h/event)

$F$  = event frequency (events/d)

$I$  = flow intensity (flow rate) at GAP ( $\text{m}^3/\text{h}$ ).

The subscript  $p$  represents the best-fit probability distribution type of the respective variables in the DFI model. The procedure of setting up a stochastic model with the known distributions for each variable involved an evaluation of each model input parameter in terms of suitable statistical distributions.

### Stochastic description of parameters

The best-fit statistical distributions for  $D$ ,  $F$ , and  $I$ , were selected by implementing goodness-of-fit (GOF) tests to the measured data, using @Risk software. The GOF tests included the Anderson-Darling, Kolmogorov-Smirnov and Chi-Squared statistic. The best-fit distribution was chosen based on a combined scoring system of the GOF tests, similar to the selection tool developed by Masereka et al. (2015). The fitted statistical distributions used in the DFI model are presented in Table 2.2, along with the corresponding mathematical descriptions and parameters.

Table 2.2. Statistical distribution descriptions

Distribution	Probability distribution function, $f(x)$ Cumulative distribution function, $F(x)$	Parameters
<b>Log-logistic</b> <b>LL (<math>\beta, \alpha</math>)</b> <b>Continuous</b>	$f(x) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{(1 + (x/\alpha)^\beta)^2}$	$\alpha > 0$ (scale) $\beta > 0$ (shape) $x \geq 0$
	$F(x) = \frac{1}{1 + (x/\alpha)^{-\beta}}$	
<b>Lognormal</b> <b>LN(<math>\mu, \sigma</math>)</b> <b>Continuous</b>	$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$	$\mu > 0$ $\sigma > 0$ $\mu = \text{mean}$ $\sigma = \text{standard deviation}$
	$F(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[ \frac{\ln x - \mu}{\sigma \sqrt{2}} \right]$	
<b>PERT</b> <b>PERT(<math>a, m, b</math>)</b> <b>Continuous</b>	$f(x) = \frac{1}{B(\alpha_1, \alpha_2)} \frac{(x-a)^{\alpha_1-1} (b-x)^{\alpha_2-1}}{(b-a)^{\alpha_1+\alpha_2-1}}$ $B(\alpha_1, \alpha_2) = \text{Beta function}$	$b > a$ boundary $b \geq m \geq a$ $a = \text{minimum}$ $b = \text{maximum}$ $m = \text{most likely value}$
	$\alpha_1 = \frac{4m + b - 5a}{b - a}$	
	$\alpha_2 = \frac{5b + a - 4m}{b - a}$	
	$F(x) = \frac{B_z(\alpha_1, \alpha_2)}{B(\alpha_1, \alpha_2)} = I_z(\alpha_1, \alpha_2)$ $z = \frac{x-a}{b-a}$ $B_z(\alpha_1, \alpha_2) = \text{incomplete Beta function}$	
<b>Binomial</b> <b>B(<math>n, p</math>)</b> <b>Discrete</b>	$f(x) = \frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{(1 + (x/\alpha)^\beta)^2}$	$n > 0$ $0 < p < 1$ $n = \text{count}$ $p = \text{success probability}$
	$F(x) = \frac{1}{1 + (x/\alpha)^{-\beta}}$	

## RESULTS

### Experimental field test results

Results discussed in this section were obtained from temperature variation analysis and volumetric measurements. A total of 68 irrigation events were identified over the 11-d study period by means of the temperature variation analysis. The average garden irrigation event duration was 2 h 16 min, with a relatively large standard deviation of 1 h 17 min. The longest irrigation event measured was 6 h and 59 min, and the shortest was 22 min. Some events were found to be relatively long in comparison with garden irrigation events reported elsewhere. The consumers of this study sample confirmed that, in some cases, the GAP would be operated until the aquifer was (temporarily) depleted and the event had to be terminated to allow for recharge, resulting in relatively long events. The probability of an irrigation event occurring on a specific day during the 110 test days in the sample was 54%, meaning that consumers irrigated roughly every second day, on average. The flow intensities at the GAPs ranged between 1.14 m<sup>3</sup>/h and 1.25 m<sup>3</sup>/h, with a most likely value of 1.16 m<sup>3</sup>/h. These values were considered to be typical for groundwater abstraction at the household scale in South Africa. Local borehole contractors often use a thumb rule of 1 L/s (3.6 m<sup>3</sup>/h) as a relatively good flow rate from a garden borehole pump. Tennick (2000) reported that garden borehole flow rates in Hermanus, South Africa, ranged between 1.0 m<sup>3</sup>/h and 2.0 m<sup>3</sup>/h. Naidoo and Burger (2017) also reported on groundwater abstraction in South Africa. The average pump flow rate was found to range between 0.36 m<sup>3</sup>/h and 2.7 m<sup>3</sup>/h (Naidoo and Burger, 2017). Flow intensity values from this study were thus within the range reported earlier.

### Stochastic results

#### *Event duration*

Garden irrigation event duration  $D$  was determined by means of the temperature variation analysis procedure described earlier. Many factors may contribute to the duration of irrigation, including the method of irrigation, property size, rainfall, aquifer yield, ambient temperature and time of day. These factors were not considered in the DFI end-use model; event duration was modelled as an independent variable. If multiple events occurred on the same day, each event duration was analysed separately. A cumulative distribution function (CDF) of the measured duration is presented in Figure 2.3, along with the CDF of the stochastic distribution with the best fit. The log-logistic distribution provided the best fit, slightly outperforming the lognormal distribution that ranked second in terms of fit. However, the parameters of the lognormal distribution (mean and standard deviation) are more readily available than the shape and scale parameters of the log-logistic distribution. Therefore, the lognormal distribution was selected to model the irrigation duration variable, thus simplifying practical application in the future.

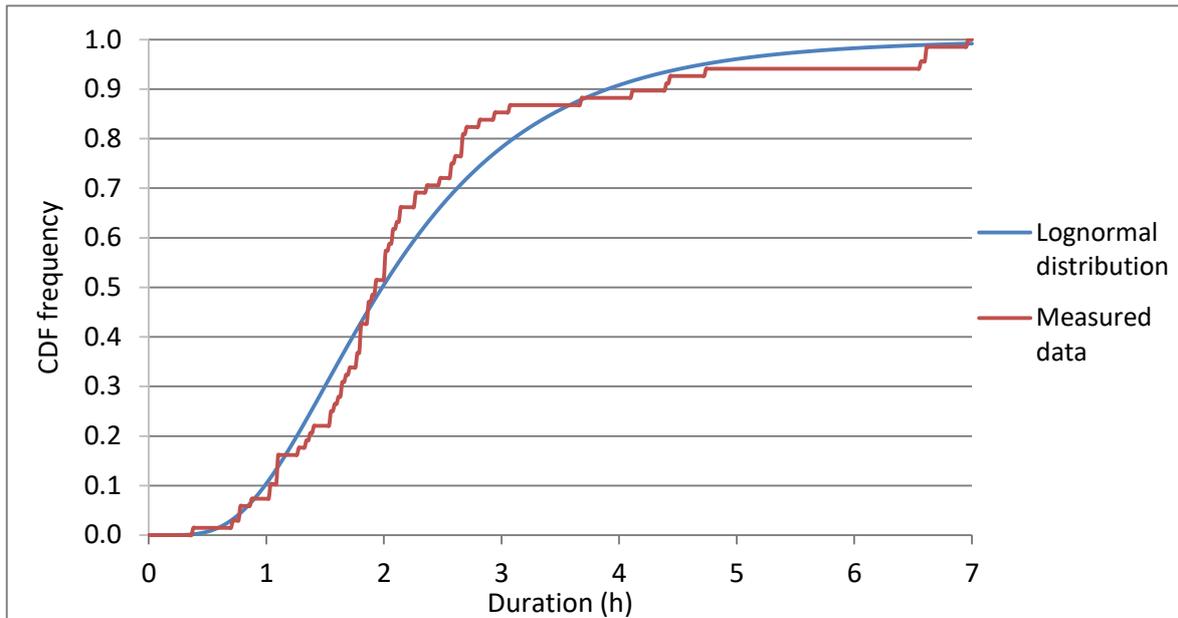


Figure 2.3. Cumulative distribution function for garden irrigation duration

### **Event frequency**

Event frequency  $F$  is often described using a discrete statistical distribution (Blokker et al., 2010) and is typically expressed as a Poisson distribution, in which case only one parameter is needed ( $\lambda$  = average) to populate the distribution. The event frequency was modelled as the probability of a garden irrigation event occurring on a specific day, with a maximum of 1 pumping event per day. Consequently, the binomial distribution was used to describe event frequency over the 11-d study period. Figure 2.4 shows the CDF of the measured irrigation frequencies with the fitted binomial distribution curve.

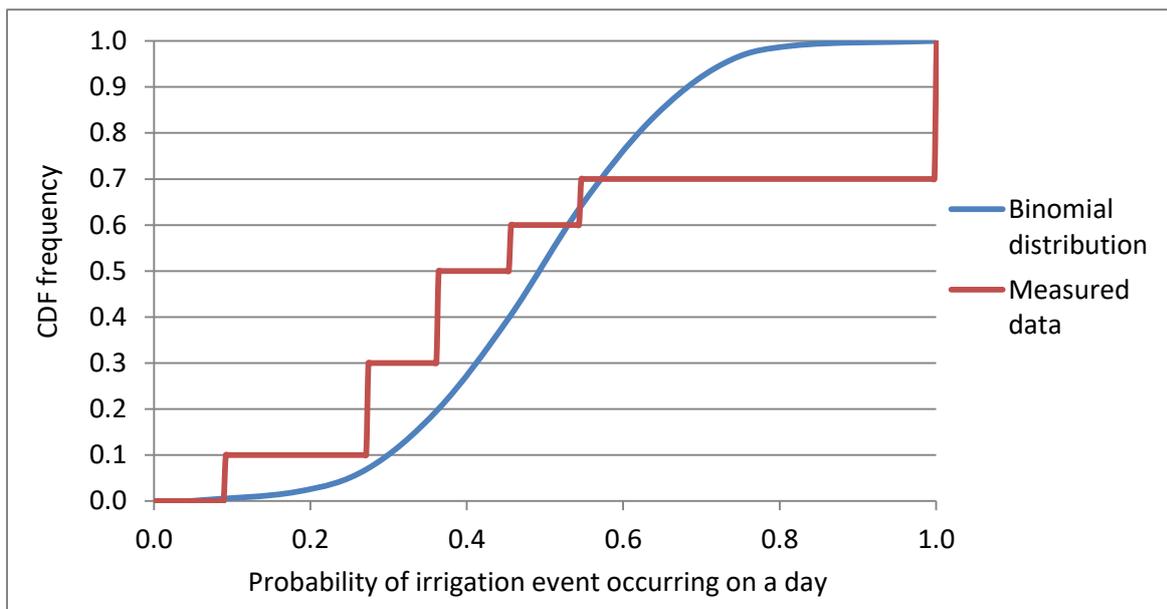


Figure 2.4. Cumulative distribution function for garden irrigation event frequency

No distribution fitted the measured data well, partly because of the small sample size and relatively sustained consumer habits and/or the use of automated programmed irrigation timers. Many different events from a particular home would thus report the same frequency and would be lumped in the CDF. The significant difference between event irrigation frequencies could also be ascribed in part to consumer behaviour and also to the different types of irrigation systems used at the study homes. The irrigation method was, however, not included as independent variable in the DFI model. Additionally, no rain days occurred during the study period. The binomial distribution provided the best fit to the data and was considered adequate to illustrate application of the model, with appreciation that future research in this regard is needed.

### ***Flow intensity***

The site survey confirmed that all GAPs were operated at full capacity while irrigating. Thus the flow intensity  $I$  at each GAP was measured at the maximum flow rate. A CDF, containing 100 data points (10 measurements at each of the 10 residential properties) was plotted in Figure 2.5. The Beta-Program Evaluation and Review Technique (PERT) distribution, identified as the best fit to the actual data, was superimposed on the actual data. The PERT distribution incorporates three parameters: the minimum, maximum, and the most likely value.

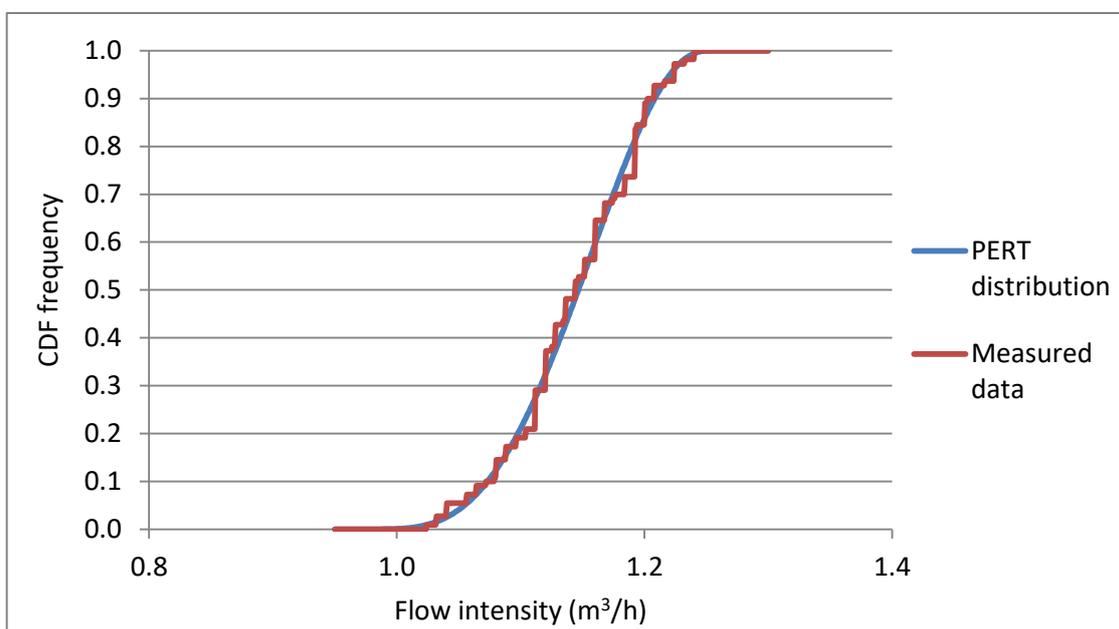


Figure 2.5. Cumulative distribution function for groundwater flow intensity

### Application of DFI model to study site

Statistical distributions were fitted to measurements obtained from the iButtons and volumetric measurements. Equation 2.1 was modified to include the identified best-fit distributions for each model variable. Equation 2.2 represents the stochastic DFI model:

$$V(\text{LN} \sim \mu, \sigma) = D(\text{LN} \sim \mu, \sigma) \times F(\text{B} \sim n, p) \times I(\text{PERT} \sim a, m, b) \quad (2.2)$$

Table 2.3 summarises the variables of the DFI model, as well as the parameter values of the specific study site. The parameter values in Table 2.3 represent garden irrigation in autumn for the specific Cape Town study site.

Table 2.3. DFI model input parameters for study site in autumn

Variable	Average ( $\mu$ )	Distribution	Parameter	Parameter value
Duration (h)	2.273	Lognormal	$\mu$	2.372
			$\sigma$	1.289
Frequency (events/d)	0.536	Binomial	$n$	11.000
			$p$	0.536
Intensity ( $\text{m}^3/\text{h}$ )	1.143	PERT	$a$	0.978
			$m$	1.157
			$b$	1.252

The DFI model was implemented on the study site by populating Equation 2.2 with the values presented in Table 2.3. A total of 1 000 000 iterations were simulated using the Monte Carlo method to stochastically determine the average daily volume (in  $\text{m}^3/\text{d}$ ) of groundwater pumped for garden irrigation. The CDF of the average daily garden irrigation event volume supplied from GAPS at the study site is shown in Figure 2.6. A comparison of the DFI model's stochastic results (based on GOF tests) and the study site measurements is presented in Figure 2.6.

The results presented in Figure 2.6 relate to the study site over the study period (April/May) and should not be generalised. The average daily groundwater abstraction for garden irrigation could simply be calculated by multiplying the average values of  $D$ ,  $F$ , and  $I$ . The stochastic results also show that a daily average of  $1.39 \text{ m}^3/\text{d}$  is used for garden irrigation, as would be expected. Due to the relatively large variation in garden irrigation volume, from one home to the next, one region to the other and by season, the average value alone does not provide sufficient insight. The stochastic results provide more detail. An additional sensitivity analysis was conducted in order to explain the relative contribution of different parameter values. The sensitivity analysis showed that garden irrigation volume was the most sensitive to event duration. The significant contribution of event duration in the model is explained by the notable parameter variability coupled to a relatively wide range in event duration amongst residents.

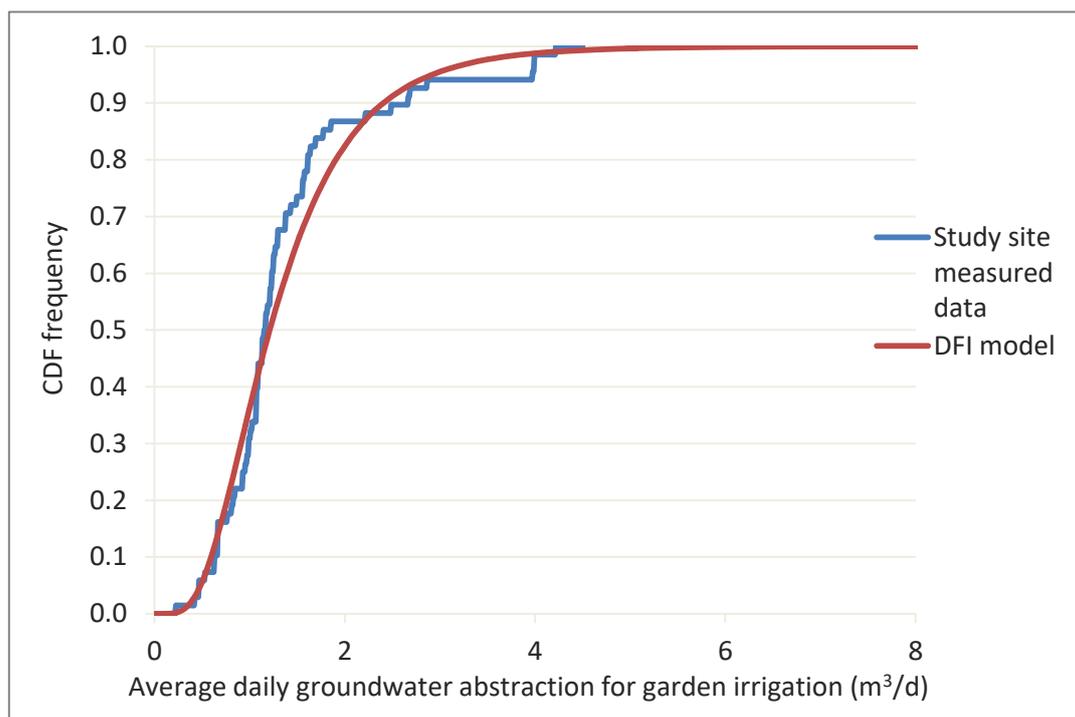


Figure 2.6. Cumulative distribution function of the average daily garden irrigation event volume

## DISCUSSION

Utilising iButtons as indirect method for measuring water usage at privately owned GAPs proved useful. The method was simple, cost effective and caused relatively little disturbance to the homeowners. The average pumping event duration at the study site was 2 h and 16 min, with the shortest event being 22 min. The recording interval of 2 min ensured that irrigation events could successfully be identified, because event duration significantly exceeded the recording interval. Expanding the application of iButtons to include different household end-use components, such as the bath, shower, washing machine and dishwasher could be explored. However, iButtons would be unable to detect events with a relative short duration (less than 2 min), such as basin taps and toilet flushing, or events where the temperature variation is expected to be small. The temperature variation method has been applied to hot water end-uses, such as the shower (Botha et al., 2017), where temperature variation is expected to be relatively large.

The average irrigation duration and frequency measured in this study are higher than values reported by Roberts (2005). This research project focused on groundwater as supplementary water source, meaning that an unrestricted irrigation scenario was considered. Consequently, it could be expected that residents with GAPs (this study) would irrigate more regularly and for longer durations compared to a sample group of residents using water from the potable water distribution system for garden irrigation.

The DFI model can serve as a useful, rudimentary means to investigate garden irrigation by researchers and utility managers. Based on the relatively small sample of 59 measurement points, a probability distribution function (PDF) cannot be defined with sufficient representativeness for longer time periods, or other regions. The combination of literature values and the 59 data points was used in this study to compile PDFs as a means to illustrate the method and obtain results from the study sample. The results are not representative of a larger region, or consumers beyond the study site. However, the DFI model is scalable over different study sites, as the parameters of the distribution curves could be populated with values corresponding to another region, or time, as applicable. The DFI model could be expanded in the future to incorporate seasonal variability, different irrigation methods and also other types of supplementary household water supply, such as rainwater and greywater.

## CONCLUSION

Unique garden irrigation events from groundwater abstraction points were identified by means of temperature variation analysis in the Cape Town case study site. A relatively high garden irrigation event occurrence was observed at all 10 homes and the recorded duration of the 68 detected events was relatively long. The DFI model was based on data measured in the Cape Town study site and was subsequently used to illustrate stochastic modelling of garden irrigation. The temperature variation analysis could be employed elsewhere to populate the DFI model with values for event duration, frequency and intensity (flow rate) in order to assess garden irrigation from groundwater abstraction points in other regions.

## ACKNOWLEDGEMENTS

The authors would like to thank the National Research Foundation (NRF) of South Africa and GLS Consulting for financial support towards a post-graduate student bursary.

## REFERENCES

ARBON N, THYER M, HATTON MACDONALD D, BEVERLEY K and LAMBERT M (2014) Understanding and predicting household water use for Adelaide. Technical Report Series No. 14/15. Goyder Institute for Water Research, Adelaide, South Australia.

BEAL C, STEWART RA, HUANG T and REY E (2011) South East Queensland residential end use study. *J. Aust. Water Assoc.* **38** (1) 80-84.

BLOKKER EJM, VREEBURG JHG and VAN DIJK JC (2010) Simulating residential water demand with a stochastic end-use model. *J. Water Resour. Plann. Manage.* **136** 19-26.

[https://doi.org/10.1061/\(asce\)wr.1943-5452.0000002](https://doi.org/10.1061/(asce)wr.1943-5452.0000002)

BOTHA BE, JACOBS HE, BIGGS B and ILEMOBADE AA (2017) Analysis of shower water use and temperature at a South African university campus. *CCWI 2017 - Computing and Control for the Water Industry*, 5-7 September 2017, Sheffield, United Kingdom.

<https://doi.org/10.15131/shef.data.5364562.v1>

BRITTON T, COLE G, STEWART R and WISKAR D (2008) Remote diagnosis of leakage in residential households. *J. Aust. Water Assoc.* **35** 89-93.

BUCHBERGER SG and WELLS GJ (1996) Intensity, duration, and frequency of residential water demands. *J. Water Resour. Plann. Manage.* **112** (1) 11-19.

[https://doi.org/10.1061/\(asce\)0733-9496\(1996\)122:1\(11\)](https://doi.org/10.1061/(asce)0733-9496(1996)122:1(11))

BUCHBERGER SG, CARTER JT, LEE YH, and SCHADE TG (2003) Random demands, travel times, and water quality in dead ends. Rep. No. 90963F. AWWARF, Denver.

BUTLER JJ (1991) A stochastic analysis of pumping tests in laterally nonuniform media. *Water Resour. Res.* **27** (9) 2401-2414.

<https://doi.org/10.1029/91wr01371>

CHAPMIN T, TODD AS and ZEIGLER MP (2014) Robust, low-cost data loggers for stream temperature, flow intermittency, and relative conductivity monitoring. *Water Resour. Res.* **50** (8) 6542-6548. <https://doi.org/10.1002/2013wr015158>

COLVIN C and SAAYMAN I (2007) Challenges to groundwater governance: a case study of groundwater governance in Cape Town, South Africa. *Water Polic.* **9** (2) 127-149.

<https://doi.org/10.2166/wp.2007.129>

DEOREO WB, HEANEY JP and MAYER PW (1996) Flow trace analysis to assess water use. *J. Am. Waste Water Assoc.* **88** (1) 79-90.

DEOREO WB, DIETERMANN A, SKEEL T, MAYER PW, LEWIS DM and SMITH J (2001) Retrofit realties. *J. Am. Water Works Assoc.* **93** (3) 58-72.

DU PLESSIS JL and JACOBS HE (2015) Procedure to derive parameters of stochastic modelling of outdoor water use in residential estates. *Proced. Eng.* **119** 803-812.

<https://doi.org/10.1016/j.proeng.2015.08.942>

DU PLESSIS JL, FAASEN B, JACOBS HE, KNOX AJ and LOUBSER C (2018) Investigating wastewater flow from a gated community to disaggregate indoor and outdoor water use. *J. Water Sanit. Hyg. Dev.* **8** (2) 238-245. <https://doi.org/10.2166/washdev.2018.125>

FISHER-JEFFES L, GERTSE G and ARMITAGE NP (2015) Mitigating the impact of swimming pools on domestic water demand. *Water SA* **41** (2) 238-246.

<https://doi.org/10.4314/wsa.v41i2.9>

GHAVIDELFAR S, SHAMSELDIN AY and MELVILLE BW (2018) Evaluating spatial and seasonal determinants of residential water demand across different housing types through data integration. *Water Int.* **43** (7) 926-942. <https://doi.org/10.1080/02508060.2018.1490878>

HEMATI A, RIPPY M, GRANT S, DAVIS K and FELDMAN D (2016) Deconstructing demand: the anthropogenic and climatic drivers of urban water consumption. *Environ. Sci. Technol.* **50** (23) 12557-12566. <https://doi.org/10.1021/acs.est.6b02938>

HEINRICH M (2007) Water end use and efficiency project (WEEP) - Final report. BRANZ Study report 159. BRANZ, Judgeford, New Zealand.

HOWE CW and LINAWEAVER FP (1967) The impact of price on residential water demand and its relation to system design and price structure. *Water Resour. Res.* **3**(1) 13-32.

<https://doi.org/10.1029/wr003i001p00013>

JACOBS HE (2007) The first reported correlation between end-use estimates of residential water demand and measured use in South Africa. *Water SA* **33** (4) 549-558.

JACOBS HE and HAARHOFF J (2007) Prioritisation of parameters influencing residential water use and wastewater flow. *J. Water Supply: Res. Technol.* **56** (8) 495-514.

<https://doi.org/10.2166/aqua.2007.068>

JACOBS HE AND HAARHOFF J (2004) Structure and data requirements of an end-use model or residential water demand and return flow. *Water SA* **30** (3) 293-304.

<https://doi.org/10.4314/wsa.v30i3.5077>

JACOBS HE, GEUSTYN LC, FAIR K, DANIELS J and DU PLESSIS JA (2007) Analysis of water savings: a case study during the 2004/2005 water restrictions in Cape Town. *J. S. Afr. Inst. Civ. Eng.* **49** (3) 16-26.

LOH M and COGHLAN P (2003) Domestic water use study in Perth, Western Australia (1998-2001). March 2003. Water Corporation, Perth, Australia. ISBN 1 74043 1235

LUGOMA MFT, VAN ZYL JE and ILEMOBADE AA (2012) The extent of on-site leakage in selected suburbs of Johannesburg. *Water SA* **38** (1) 127-132.

<https://doi.org/10.4314/wsa.v38i1.15>

MACDONALD AM and CALOW RC (2009) Developing groundwater for secure water supplies in Africa. *Desalination* **248** 546-556. <https://doi.org/10.1016/j.desal.2008.05.100>

MAKWIZA C and JACOBS HE (2017) Sound recording to characterize outdoor tap water use events. *J. Water Suppl.: Res. Technol.* **66** (6) 392-402.  
<https://doi.org/10.2166/aqua.2017.120>

MASEREKA EM, OTIENO FAO, OCHIENG GM and SNYMAN J (2015) Best fit selection of probability distribution models for frequency analysis of extreme mean annual rainfall events. *Int. J. Eng. Res. Dev.* **11** (4) 34-53.

MASSUEL S, PERRIN J, WAJID M, MASCRE C and DEWANDEL B (2009) A simple low-cost method to monitor duration of groundwater pumping. *Groundwater* **41** (1) 141-145.  
<https://doi.org/10.1111/j.1745-6584.2008.00511.x>

MEAD N and ARAVINTHAN V (2009) Investigation of household water consumption using smart metering system. *Desalination Water Treat. J.* **11** (1-3) 115-123.  
<https://doi.org/10.5004/dwt.2009.850>

NAIDOO V and BURGER M (2017) Groundwater abstraction assessment for development on the farm Vlakfontein 523 JR Portion 25. Report compiled for JCJ Developments. Geo Pollution Technologies Report No. ICEAS-17-2246. JCJ Developments, Gauteng, South Africa.

NEL N, JACOBS HE, LOUBSER C and DU PLESSIS JA (2017) Supplementary household water sources to augment potable municipal supply. *Water SA* **43** (4) 553-562.  
<https://doi.org/10.4314/wsa.v43i4.03>

PARKER JM and WILBY RL (2013) Quantifying household water demand: A review of theory and practice in the UK. *Water Resour. Manage.* **27** (4) 981-1011.  
<https://doi.org/10.1007/s11269-012-0190-2>

RATHNAYAK K, MALANO H, MAHEEPALA S, GEORGE B, NAWARATHNA B, ARORA M, ROBERTS P (2015) Seasonal demand dynamics of residential water end-uses. *Water J.* **7**, 202-216. <https://doi.org/10.3390/w7010202>

ROBERTS P (2005) Yarra Valley Water 2004 residential end use measurement study. Final report, June 2005. Yarra Valley Water. Melbourne, Australia.

RUNFOLA DM, POLSKY C, NICOLSON C, GINER NM, PONTIUS RG, KRAHE J and DECATUR A (2013) A growing concern? Examining the influence of lawn size on residential water use in suburban Boston, MA, USA. *Landscape Urban Plann. J.* **119** 113-123.

<https://doi.org/10.1016/j.landurbplan.2013.07.006>

SCHEEPERS HM and JACOBS HE (2014) Simulating residential indoor water demand by means of a probability based end-use model. *J. Water Supply: Res. Technol.* **63** (6) 476-488.

<https://doi.org/10.2166/aqua.2014.100>

TENNICK FP (2000) Preliminary report on hydrocensus of groundwater resources in the Eastcliff and Northcliff suburbs, Hermanus, Western Cape, South Africa. Compiled by FPT Consulting Services for the Overstrand Municipality, May 2000.

TURRAL HN, ETCHELLS T, MALANO HMM, WIJEDASA HA, TAYLOR P, MACMAHON TAM and AUSTIN N (2005) Water trading at the margin: The evolution of water markets in the Murray-Darling Basin. *Water Resour. Res.* **41** W07011. <https://doi.org/10.1029/2004wr003463>

VECK GA and BILL MR (2000) Estimation of the residential price elasticity of demand for water by means of a contingent evaluation approach. WRC Report No. 790/1/00. Water Research Commission, Pretoria.

WASOWSKI A (2001) Changing the American landscape. *J. Am. Wastewater Assoc.* **3** 40-44.

WILLIS RM, STEWART RA, PANUWATWANICH K, WILLIAMS PR and HOLLINGSWORTH AL (2011) Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. *J. Environ. Manage.* **92** (8) 1996-2009.

<https://doi.org/10.1016/j.jenvman.2011.03.023>

WRIGHT T and JACOBS HE (2016) Potable water use of residential consumers in the Cape Town metropolitan area with access to groundwater as supplementary household water source. *Water SA* **42** (1) 144-150.

<https://doi.org/10.4314/wsa.v42i1.14>

## Chapter 3.

### Analysis of shower water use and temperature at a South African university campus

Bettina Elizabeth Meyer<sup>1</sup>; Heinz Erasmus Jacobs<sup>1\*</sup>, Ben Biggs<sup>2</sup>, Adeshola Ilemobade<sup>3</sup>

1. Department of Civil Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa, [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za), [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za)

2. JG Afrika, 14 Central Square, Pinelands, 7430, Cape Town, South Africa, [BiggsB@jgafrika.com](mailto:BiggsB@jgafrika.com)

3. School of Civil and Environmental Engineering, University of the Witwatersrand Johannesburg, Private Bag 3, WITS, 2050, South Africa, [Adesola.Ilemobade@wits.ac.za](mailto:Adesola.Ilemobade@wits.ac.za)

\*Corresponding author: [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za)

Reproduced from the Computer and Control for the Water Industry (CCWI), The University of Sheffield Journal contributions, F64 (2017).

Available online: <https://doi.org/10.15131/shef.data.5364562.v1>

#### ABSTRACT

Earlier research has underlined that household end-uses form building blocks of the residential water demand pattern. Numerous end-use studies have been presented in the past, but none have reported on shower end-uses at Universities in South Africa. This research focuses on shower water use, as part of the first detailed end-use field study conducted in South Africa. An extensive desktop study was conducted regarding South African end-uses, focusing on shower use. Shower flow rate was physically measured under different conditions, while actual shower duration for the same showers was derived from water temperatures recorded over two periods of 5 days each, at 1 min frequency. The changes in temperature were used to estimate actual shower duration and event start times. The total shower event volume was stochastically determined by using Monte Carlo analysis. The average shower duration of the 759 shower events identified as part of this study was 9 min and 33 sec, with a flow rate of 8.7 L/min. The average shower event volume was 83 L/event.

**Keywords:** Water end-use, shower, temperature

## INTRODUCTION

### Shower as end-use of water

End-uses of water are considered to be building blocks of the residential water demand pattern (Buchberger and Wells 1996). The most notable indoor end-uses have been reported to be the toilet, shower, bath, and clothes washing machine (Scheepers and Jacobs 2014). End-uses of water can be extracted from time series data recorded at high spatial and temporal resolution by a consumer meter, employing software tools such as Flow Trace Analysis (Mayer et al. 1999) or intelligent pattern recognition models (Nguyen et al. 2013). Two models available for theoretical end-use analysis include SIMDEUM (Blokker et al. 2010) and REUM (Jacobs and Haarhoff 2007). According to earlier reports (Blokker et al. 2010, Jacobs and Haarhoff 2007) the three parameters that influence water use for showers include event duration ( $D$ ), frequency of use ( $F$ ), the flow rate or intensity ( $I$ ) - and of course the household size. Assuming a known household size, the focus shifts to  $D$ ,  $F$  and  $I$ . Investigation into the type of geysers used to heat the water was not deemed necessary to meet the objectives of this study. The frequency of use was also not evaluated during this phase of the project due to various constraints.

### Research focus and method

The research team set out to assess shower event volume in communal bathrooms at two University residences. The focus was on the shower duration, intensity, and the derived shower event volume. Event volume was modelled stochastically, because it was considered impractical to measure the actual event volumes due to various constraints. The two key unknown model parameter values were the shower duration,  $D$ , and intensity,  $I$ . Shower duration notably impacts shower water use and it was considered necessary to assess actual shower durations. Shower duration was assessed for 15 different shower heads at two Stellenbosch University residences. Temperature loggers were used to record shower water temperature over two periods of five days each, at 1 min intervals. The variation in temperature of shower heads was used to derive shower duration and timing of each event. The temperature recorders provided estimates for duration instead of actual duration, but had numerous advantages when compared to other methods of measurement.

The intensity (flow rate in L/s) is a function of shower head design and the water pressure, which in turn is often influenced by the type of water heater (geyser). An additional uncertainty is introduced by the fact that consumers could throttle the flow rate by adjusting the shower taps, or outlet valve(s) at any time during the shower event. Three types of shower heads and sizes were tested in this study with fully open valves. Some shower heads control the flow rate (typically modulated to 9 L/s), for example with a pressure compensating flow restrictor, so it was not considered necessary to measure actual water pressure to obtain suitable results.

## Objectives

The study was structured around the following objectives: (i) conduct an extensive desktop study regarding South African end-uses, focusing on shower use; (ii) record shower water temperature at various showers in two residences over two periods of 5 days each, at 1 min frequency, in order to estimate the actual shower duration and event start times; (iii) record selected shower flow rates and (iv) set up a stochastic model for shower event volume, using Monte Carlo analysis.

## REPORTED SHOWER WATER USE

Published values for the per capita shower event volume vary notably, from about 30 L/p/d to over 100 L/p/d. The variation is also dependent on whether an individual prefers to bath or shower. The bath and shower share the same purpose - for personal hygiene - and are often reported on integrally in terms of end-use volume. The term “shower/bath” is used in literature and was adopted in this text as well. Earlier research (Makki et al. 2013) noted that the average shower/bath use volume from a number of other studies was 46 L/c/d, and reported 45 L/c/d for the shower event volume in South East Queensland, Australia. The mean shower event volume in a North American study (DeOreo et al. 1996) was 59 L/event. The shower end-use category has been identified as the largest portion of indoor consumption in another Australian study (Beal et al. 2011), with an average of 42.7 L/p/d, representing 33% of the total indoor consumption. Various other international end-use studies have also identified the bath/shower as the most notable indoor end-use. A summary of formerly published values noted that the shower/bath contributed between 26% and 46% to the total indoor water use (Scheepers and Jacobs 2014). The shower has been noted to contribute about 15% to the total peak hourly indoor flow (Funk and DeOreo 2011). One study reported on shower water use in South Africa, suggesting an event volume of 92 L/p/d (Shutte and Pretorius 1997). Some Municipal by-laws in South Africa stipulate 10 L/min as the maximum shower head flow rate, with the most common shower head based on sales volume of a leading supplier providing 9 L/min according to specification - the latter has a pressure compensating flow restrictor. A shower event volume of 81 L/p/d was assumed (Jacobs et al. 2017) in a recent end-use analysis for homes in Johannesburg, based on 9 L/min intensity and 9 min shower duration.

The highest reported shower event volume in a study group of 28 homes in the UK (Butler 1991) was 156 L/p/d, and another UK-study reported 150 L/p/d (Lauchlan and Dixon 2003). A reasonable upper limit for shower event volume is ~400 L/event, considering a relatively high shower nozzle flow rate (20 L/min) combined with a 20 minute shower duration (this study noted numerous shower event durations between 20-30 min, so even with low flow nozzles the total event volume may exceed 300 L/event).

## DESCRIPTION OF FIELD TESTS

### Study site

Two residences were investigated at Stellenbosch University - one male and one female residence. The women's residence was labelled Residence A and the men's residence was labelled Residence B. Residence A is a 4-story building and Residence B is a 9-story building. The two study sites were chosen based on information provided by the facilities manager. The two residences represent typical women's and men's residences for the Stellenbosch University campus. Residence A has roughly 16 residents per shower and Residence B has roughly 6 residents per shower. The total duration of the shower investigation study was 3 weeks, from 27 February 2017 to 20 March 2017.

### Volumetric measurement

Flow rate measurements were taken using a measuring cylinder. For university residences, it is not uncommon for students to remove the shower head to increase water flow, but showers without showerheads were repaired and not included in the measurement sample. Selected showers were inspected and flow rates measured volumetrically. In this study, 3 measurements were taken on 5 floors (floor 2, 3, 5, 7 and 9) at Residence B, using 3 different shower heads (existing large flat, oxygenics, and shower power) in each case. The flow rates at the Residence A (floor 2 and 3) were assumed to correspond to measurements at Residence B, as confirmed with two volumetric tests at Residence A. The static pressure was not taken into account when simulating the flow rates, since the flow rate was measured and stochastic analysis accounted for uncertainties in variables.

### Temperature measurement

Shower head temperature was recorded by means of temperature loggers. The temperature variation was used to derive shower durations. The method of using temperature logging to identify event durations has been used successfully in the past (Massuel et al. 2009, Botha 2017, Meyer and Jacobs 2019) for groundwater flow from boreholes. Events are identified based on the difference between the recorded end-use outlet temperature (on the shower head in this case) and some baseline temperature, generally recorded simultaneously as control. However, having multiple showers in the same room meant that a separate baseline temperature was not needed and the events were determined by analysing the difference between the measured temperature and the room temperature. When there was an increase in temperature (positive temperature gradient over a certain time step) an event was flagged, and the duration was determined using MS Excel and the time-series graph from Coldchain Thermodynamics. The end of each event was identified by the first subsequent negative temperature gradient. A relatively long event would thus show a temperature increase, followed by a sustained temperature during the shower event, and a temperature decline after the end of the shower event.

Relatively shorter events would not have a sustained component (due to the logging frequency of 1 min). A sequence of events from the time series data at one of the shower heads is shown in Figure 3.1.

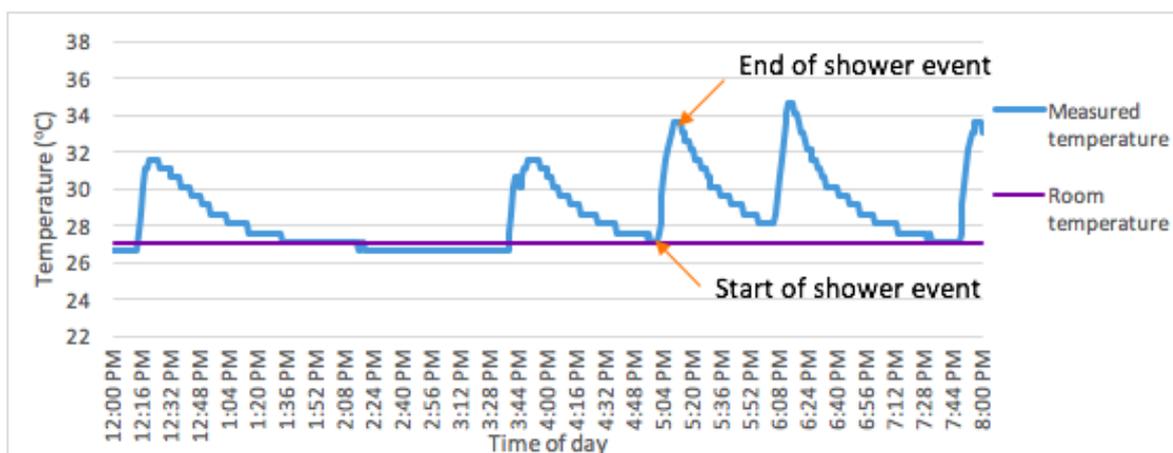


Figure 3.1. Identifying shower events

## STOCHASTIC ANALYSIS

Monte Carlo analysis was used in this study to transform uncertainties in input variables to stochastic model outputs. Numerous software applications employ Monte Carlo method for risk analysis. @Risk, a commercial spreadsheet-based software programme by Palisade Corporation is available in South Africa and was used in this study. Input data are specified in Microsoft Excel and computed by adding different probabilistic distribution functions to the specified input cells. The software imitates the uncertain input variables using Monte Carlo simulations. The best distribution fit for the input variables and generated outputs are determined in @Risk by means of a goodness of fit (GOF) test, such as the Anderson-Darling test, chi-squared test and the Kolmogorov-Smirnov test. A basic shower end-use model was set up in Excel, using the @Risk add-in, to run a Monte Carlo analysis. The model stochastically describes the shower event volume as an appropriately selected value for duration multiplied by intensity.

## RESULTS

### Shower flow rate

Table 3.1 summarises the flow measurements taken at the respective shower heads and the static pressure at each floor. The values were used as input variables for the shower intensity. The large flat shower heads are the existing shower heads at the residences. Although only the existing shower heads are analysed in this study, the other two shower heads were tested to evaluate potential future installations. The measured flow rates for all shower heads were less than locally reported typical values for showers, of 9-10 L/min (Jacobs et al. 2017).

The cumulative distribution function (CDF) of all the measured flow rates is plotted in Figure 3.2 and the best fit distribution for the simulated flow rates (uniform distribution) was superimposed on the graph. The statistical parameters of the flow rate variable are summarised in Table 3.2.

Table 3.1. Residence B flow rates and pressures

Floor	Static Pressure (kPa)	Flow (L/min)		
		Large flat (existing)	Oxygenics	Shower power
2	250	8.8	8.3	8.9
3	230	9.0	7.8	8.2
5	180	8.1	6.6	6.9
7	100	8.8	5.6	4.9
9	0	8.3	4.3	3.7
<b>Average</b>	152	8.6	6.5	6.5

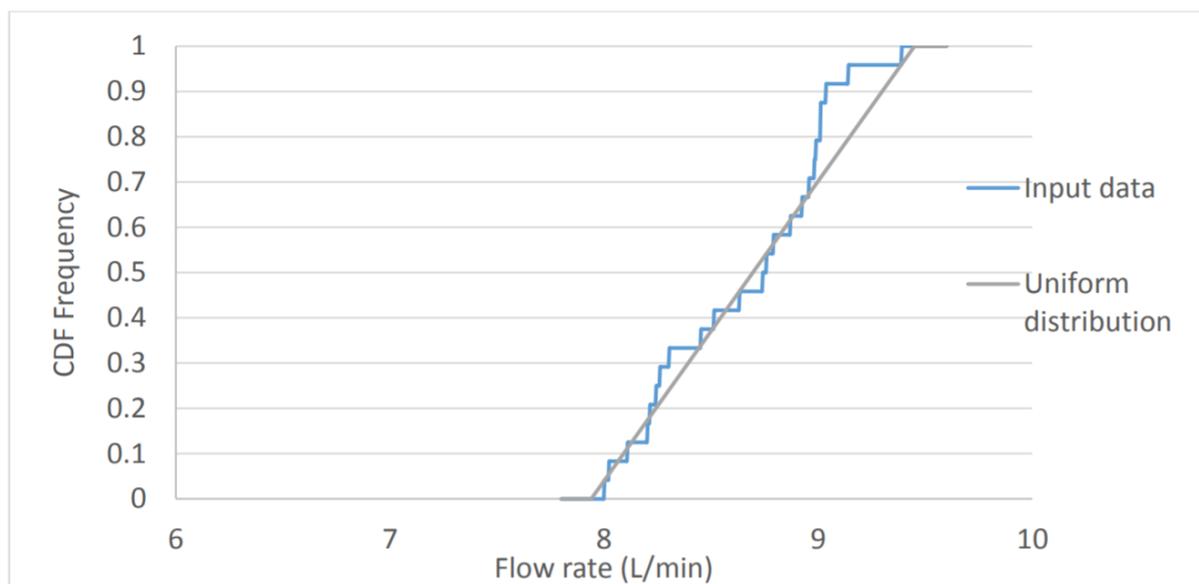


Figure 3.2. Cumulative distribution function (CDF): shower flow rate

Table 3.2. Statistical parameters for the shower flow rates

Variable	Distribution	Minimum (a)	Maximum (b)
Flow rate (Q)	Uniform	7.940	9.450

### Shower duration

The temperature analysis, discussed earlier, successfully identified 759 shower events. The average duration was 9 min 33 s with a standard deviation of 4 min 15 s and maximum of 30 min. The peak shower frequency occurred during the morning (6AM – 9AM) and late afternoon (7PM – 10PM). The shower duration histogram is presented in Figure 3.3. The gamma distribution was selected as the best fit based on the overall ranking determined by the GOP tests. The CDF of all events is plotted in Figure 3.4 and corresponding statistical parameters are summarised in Table 3.3.

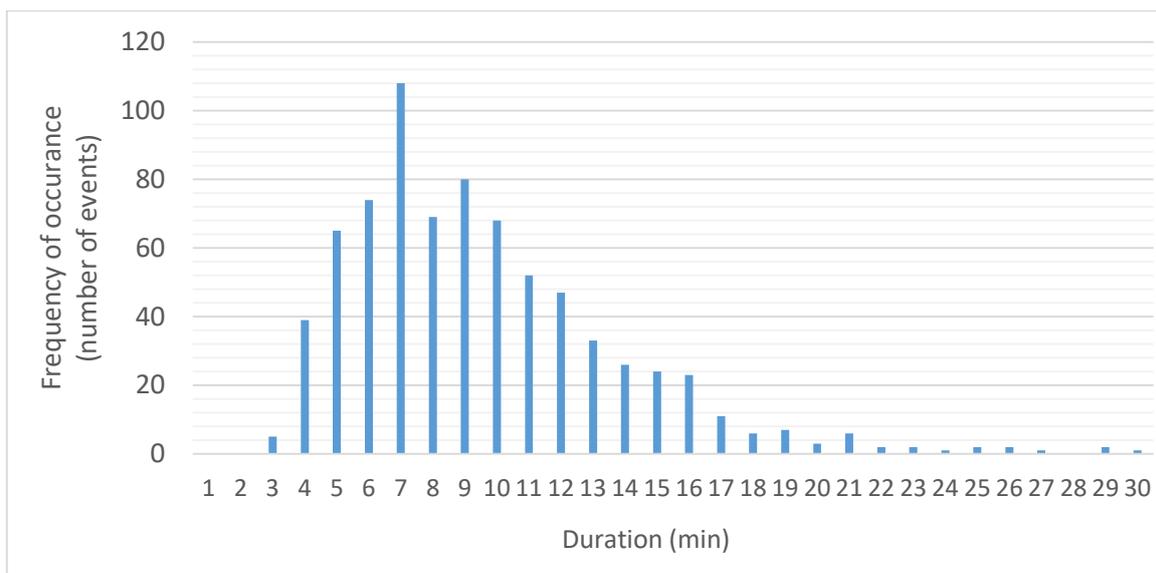


Figure 3.3. Shower event frequency

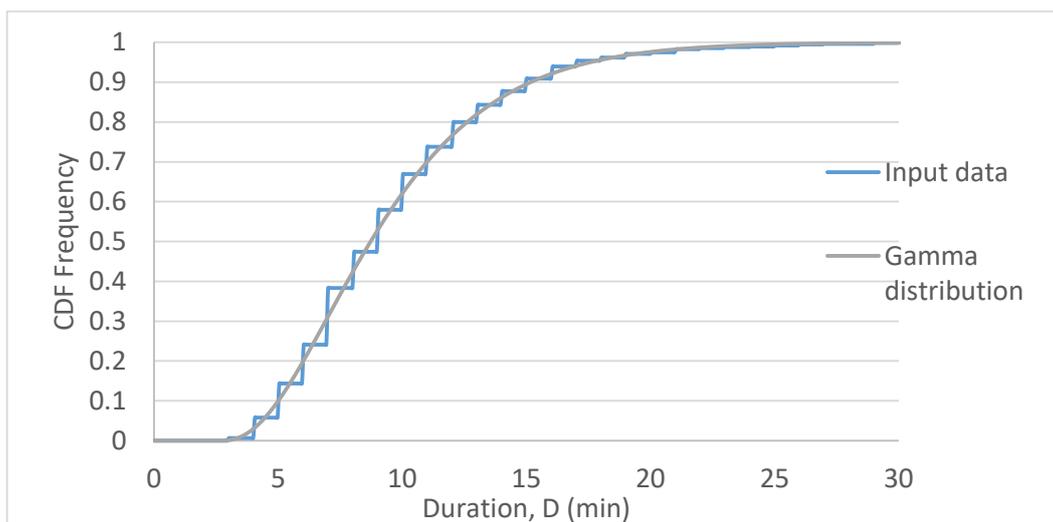


Figure 3.4. Cumulative distribution function (CDF): shower durations

Table 3.3. Statistical parameters for the shower durations

Variable	Distribution	Significance level ( $\alpha$ )	Regression coefficients ( $\beta$ )	Shift factor
Duration (D)	Gamma	2.735	2.536	2.616

## Shower event volume

The simulated input values were used to populate the shower end-use model. The CDF of the model results is plotted in Figure 3.5. Average shower event volume was found to be 83 L/event, with 90% of the events using more than 42 L/event and 10% of the events using more than 135 L/event. The results are similar to reported values from previous studies.

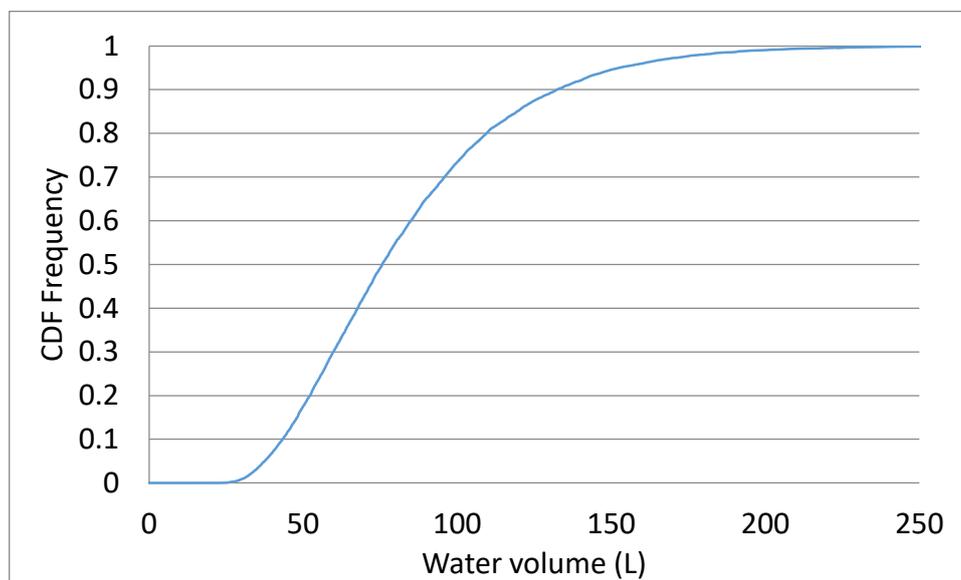


Figure 3.5. Cumulative distribution function (CDF): water consumption per shower event

## CONCLUSION

The shower duration was derived from recorded temperature at shower heads in two University residences. Temperature loggers proved to be a simple, cost effective method to determine shower durations. The majority of shower events took place in the morning between 6 AM and 9 AM; and at night between 7 PM and 10 PM. The average shower duration was measured to be 9 min 33 s with a standard deviation of 4 min 15 s, in line with typical values reported in literature. The average intensity, or flow rate, was 8.7 L/min. The average shower event volume for the specific study site was determined to be 83 L/event.

## REFERENCES

C. Beal, R.A. Stewart, T.T. Huang and E. Rey, "SEQ residential end use study". *Journal of the Australian Water Association*, 38 (1), pp. 80-84, 2011

E.J.M. Blokker, J.H.G. Vreeburg, and J. C. van Dijk, "Simulating Residential Water Demand with a Stochastic End-Use Model", *J. Water Resour. Plann. Manage.*, 136:pp. 19- 26, 2010

B.E. Botha (2017) "Estimating household groundwater abstraction by means of a probability model", Master's thesis, Stellenbosch University, 2017 [Online], Available: <http://scholar.sun.ac.za/handle/10019.1/100900>

S.G. Buchberger and G.J. Wells, "Intensity, duration, and frequency of residential water demands", *J. Water Resour. Plann. Manage.*, Jan/Feb, pp. 11-19, 1996

D. Butler, "A small-scale study of wastewater discharges from domestic appliances", *Water and Environmental Management Journal*, 5, pp.178-185, 1991

W.B. DeOreo, J.P. Heany and P.W. Mayer, "Flow trace analysis to assess water use", *Journal American Wastewater Association*, January 1996: pp.79-90, 1996

A. Funk and W.B. DeOreo, "Embedded Energy in Water Studies Study 3: End-use Water Demand Profiles", CALMAC Study Volume 1 of 1 Main Report, Prepared by Aquacraft, Inc. for the California Public Utilities Commission Energy Division, April 29, 2011

HE Jacobs, AA Illembade, BE Botha 2017. End-use estimates of water demand within certain residential categories in the city of Johannesburg. *Computing and Control for the Water Industry (CCWI)*, Sheffield, 5th-7th September 2017. Journal contribution. <https://doi.org/10.15131/shef.data.5364556.v1>

H.E. Jacobs and J. Haarhoff, "Application of a residential end-use model for estimating cold- and hot water demand, wastewater flow and salinity.", *Water SA J.*, 30(3), pp. 305- 316, 2007

C.S. Lauchlan and A. Dixon, "Modelling of water discharges for houses with conventional and low water use appliances", *Advances in Water Supply and Management - Maksimovic, Butler and Mermon (Editors)*, Swets and Zeitlinger, Lisse, pp.517-524, 2003

A.A. Makki, R.A. Stewart, K. Panuwatwanich and C. Beal, "Revealing the determinants of shower water end use consumption: enabling better targeted urban water conservation strategies", *Journal of Cleaner Production*, Dec 1;60: pp.129-46, 2013

S. Massuel, J. Perrin, M. Wajid, C. Mascré, and B. Dewandel, "A simple low-cost method to monitor duration of groundwater pumping", *Groundwater*, January/February 41(1), pp. 141- 145, 2009

P.W. Mayer, W.B. DeOreo, E.M. Opitz, J.C. Kiefer, W.Y. Davis, B. Dziegielewsji, and J.O. Nelson, "Residential end uses of water", AWWA Research Foundation and American Water Works Association, 1999

Meyer BE, Jacobs HE. 2020 Garden irrigation as household end-use in the presence of supplementary groundwater supply. *Water SA*, 45(3), 447-455.  
doi: 10.17159/wsa/2019.v45.i3.6741

K.A. Nguyen, R.A. Stewart, H. Zhang, "An intelligent pattern recognition model to automate the categorisation of residential water end-use events" *Environmental Modelling & Software* 47 pp. 108-127, 2013

H.M. Scheepers and H.E. Jacobs, "Simulating residential indoor water demand by means of a probability based end-use model", *J.of Water Supply: Research and Technology-Aqua* 63 (6), pp. 476-488, 2014

C.F. Schutte and W.A. Pretorius, "Water demand and population growth", *Water SA J.* 23(2), pp. 127-133

## Chapter 4.

### Clothes washing as household end-use: comparison of different appliance models in view of expected water savings

Bettina Elizabeth Meyer<sup>1</sup>; Heinz Erasmus Jacobs<sup>1\*</sup>, Ulrich Terblanche<sup>2</sup>

<sup>1</sup>Department of Civil Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa, [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za), [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za),

<sup>2</sup>Center of Renewable and Sustainable Energy Studies, Stellenbosch University, Banghoek Rd, 7600, South Africa, [ute@3e.eu](mailto:ute@3e.eu)

\*Corresponding author: [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za)

Reproduced from the Water Distribution System Analysis / Computer and Control for the Water Industry (WDSA/CCWI) Joint Conference Proceedings. Vol. 1. 2018

Available online: <https://ojs.library.queensu.ca/index.php/wdsa-ccw/article/view/12009>

#### ABSTRACT

The washing machine has been identified in earlier research as one of the most notable indoor household end-uses of water. Water is also commonly heated for clothes washing, with added implications regarding energy use. The paper reports on the first study from Africa based on measurements at the point of use and forms part of the first detailed end-use field study conducted in South Africa. This research focused on washing machine water use (laundry). The approach consisted of (i) conducting an extensive desktop study regarding washing machine water use; (ii) measuring the actual water use per wash cycle for two different appliance types over a period of 4 weeks, in a controlled environment at a University residence in Stellenbosch. Washing machine event volume was stochastically determined for various appliance types by means of Monte Carlo analysis and a rudimentary end-use model. The simulated water use for the two appliance types that were included in the field experiment were compared to the actual water use for the appliances. The study included a total of 54 washing machine events with an average event volume of 147 L/cycle for the top loader and 62 L/cycle for the front loader appliance. The stochastically derived expected water saving due to appliance change (as per this study) is 85 L/cycle. A change of washing machine type may hold enormous water saving potential of up to 58% of the water used for clothes washing.

**Keywords:** Water consumption, household water savings, washing machine

## BACKGROUND

The washing machine has been identified in earlier research as one of the most notable indoor household end-uses of water (Scheepers and Jacobs 2014). A summary table constructed by Scheepers and Jacobs (2014) shows that on average clothes washing contributes 24% to the total indoor household water demand when a washing machine is present. Research focusing on South African end-use studies estimated the contribution of washing machines to the total indoor household water demand to be 12% (Jacobs et al. 2017). Water consumption of a washing machine at a residential property is dependent on various factors, namely: appliance brand, model and type; number of wash cycles per household, the chosen washing temperature and program; as well as the load size (weight) being washed. An end-use study conducted by (Jacobs et al. 2017) identified the need to quantify washing machine cycle frequency and volume per cycle.

A washing machine's energy data (electricity and water consumption) is generally made available by the manufacturer. Previous studies have shown that there are large differences between actual measured consumptions per wash cycle and what is rated or listed on the appliance. It is thus important to measure actual in-use water consumption data. However, accurate water consumption measurements at the point of use are hard to obtain. This paper reports on the first study from Africa based on measurements at the point of use and forms part of the first detailed end-use field study conducted in South Africa. The first objective of this research was to conduct field experiments for two appliance types (top loader and front loader) to determine and compare the actual water use of the appliances. Measuring the water consumption at the point of use can elucidate demand estimates.

In addition to the water consumption per cycle, the number of cycles per household largely affects the water demand for clothes washing. Energy efficiency for washing machines is a main focus for regional and local regulations (DEWHA 2008). Globally, consumer habits vary considerably, and little research has been done regarding actual consumer behaviour. Past publications have commented on the number of wash cycles per household per year in various countries around the world, however, no data of Africa, or South Africa, was available at the time of this study. Therefore, the wash cycles for residential homes in South Africa were stochastically determined by using the global data as input variables in a rudimentary end-use Monte Carlo model simulation. Monte Carlo analysis is a statistical method for numerical integration and consists of three key steps; (i) defining and generating input variables, (ii) constructing a probability model, and (iii) solving the model by running numerous simulations to develop an output distribution (Polkoradi and Molnar 2011). The National Statistical office of a country generally uses survey data for residential washing machine water consumption publications. A comparison was made by Bocken et al. (2017) between customer interviews and actual washing cycles and showed that 70% of customers wash more than they think. For the purpose of this paper @Risk, a spreadsheet based Monte Carlo analysis software program, was used as software tool for risk analysis.

The load size for washing is dependent on the appliance capacity and washing program selected. Whites and colours' recommended load capacity are typically double the recommended load capacity of easy care and delicate programs. The vast majority of European households tend to use more than 75% of the washing machine's capacity (Almeida and Fonseca 2006). Apart from Europe, very little data is available regarding the load sizes. No clothes washing load size data was available at the time of this study and subsequently load sizes were considered out of scope for this research. Also, the performance of washing machines, besides the water consumption, was not included in the research.

The objectives of this research were thus to (i) determine the washing machine event volume for various appliance models, (ii) conduct field experiments for two appliance types (top and front loader) to determine and compare the actual water use for the appliances, and (iii) to compare the field experiments to results from the rudimentary end-use model.

## METHODOLOGY

### Rudimentary end-use statistical model

Monte Carlo analysis is commonly used in statistics and economics to stochastically determine consumer demand values. The stochastic method uses a large number of uncertainty variables to solve mathematical problems that cannot otherwise be solved Nathan and Weinmann (2003). Monte Carlo analysis was used in this research to simulate a rudimentary end-use model to ultimately estimate the washing machine water consumption per household and per annum. The rudimentary end-use model is presented in Equation 4.1:

$$Q_{WM} = V_{WM} \times F \quad (4.1)$$

where,

$Q_{WM}$  = average washing machine water consumption (L/household/annum)

$V_{WM}$  = water consumption per washing cycle (L/cycle)

$F$  = frequency of wash cycles (cycle/household/annum)

The water consumption per cycle input variable was determined by conducting an extensive desktop study regarding washing machine water use on a global level with a specific focus on washing machines sold in South Africa. The specific appliance brands and models chosen for the statistical model were based on the sales numbers of one of South Africa's leading appliance retailers (who requested to stay anonymous). This paper did not compare the washing machine models and/or brands, but simply used the water consumption data of each machine. Table 4.1 and Table 4.2 list the sales (frequency of purchase) of the top ten washing machines (top loaders and front loaders respectively) sold in 2017 as well as the average manufactures reported water consumption.

Table 4.1. Top loader washing machine sales and water consumption

Top Loader			
Brand	Model	Water consumption per cycle (L)	Sales (Units)
SAMSUNG	WA90H4200SW	160	807
SAMSUNG	WA13F5S2UWW	146	729
SAMSUNG	WA13J5710SG	154	473
SAMSUNG	WA15J5730SS	154	290
LG	T1450TEFT	168	179
WHIRLPOOL	3SWTW4800YQ	68	177
SPEED QUEEN	LWS21NW	150	161
WHIRLPOOL	WTL1300SL	70	124
LG	T1449TEFT1	168	111
DEFY	DTL146	147	89

Table 4.2. Front loader washing machine sales and water consumption

Front loader			
Brand	Model	Water consumption per cycle (L)	Sales (Units)
BOSCH	WAB20268ZA	53	1441
BOSCH	WAK2428SZA	56	835
BOSCH	WAT2848XZA	65	721
SIEMENS	WM10K200ME	54	665
AEG	L34173W	54	470
DEFY	DAW373	50	306
BOSCH	WAB16061ZA	50	290
WHIRLPOOL	FSCR90426	49	248
SMEG	WM128SSA	58	235
PANASONIC	NA-148MB1LZA	50	222

The water use of each specific appliance was obtained through product specification sheets, as well as personal correspondence with the relevant brand manufacturers. Different appliance programs (cottons or delicates for example) for different water temperatures were considered during the Monte Carlo analysis. The following factors may cause the manufacturer's consumption values to differ from on-site installation values: water pressure, quality; water inlet temperature; ambient temperature; type and amount of laundry; type and amount of detergent used; fluctuations in the power supply and additional functions selected.

The sales frequency of purchase and the water consumption per washing machine presented in Table 4.1 and Table 4.2 were used to generate stochastic water consumption estimates. The rudimentary model thus accounted for sales volume of different machine types. The average simulated water consumption for the top loader was 148 L/cycle and 55 L/cycle for the front loader. These results were compared to that of the field experiment.

The number of cycles per household is proportional to the household size (Pakula and Stamminger 2010). An increase in household size will result in an increase in washing cycles per household, but result in a decrease in washing cycles per person living in the household. Research conducted by Berkholz et al. (2006) shows a nearly linear regression between household size and washing cycles. The washing cycles vary from 110 cycles per annum for single households to 364 yearly cycles for households with 6 persons. According to the Living Conditions Survey 2014/2015 conducted by Statistics South Africa (2015), the average South African household size is 3.3. Because the relationship between household size and washing machine ownership is not known in South Africa, the distribution curve of the number of washing cycles per household per year was stochastically determined using a variety of international published data as input variables. Table 4.3 was adapted from Pakula and Stamminger (2010) and documents the yearly number of wash cycles per household for different regions around the world. The cumulative distribution function (CDF) curve of the estimated washing cycles per household per year is graphed in Figure 4.1.

Table 4.3. Summary of washing cycle frequency per household

Region	Washing cycles (cycles/household/annum)
West Europe	165
East Europe	173
Turkey	211
North America	289
Australia	260
China	100
South Korea	208
Japan	520

Figure 4.1 shows the average washing cycles per household per year to be 197. Washing machines sold in South Africa must be labelled with an energy efficiency stamp in order for consumers to know the energy and water efficiency of each appliance. Manufacturers base their yearly water consumption calculations on an estimated value of 220 standard washing cycles per year. The distribution curve for the cycles per year are thus in line with industry estimates.

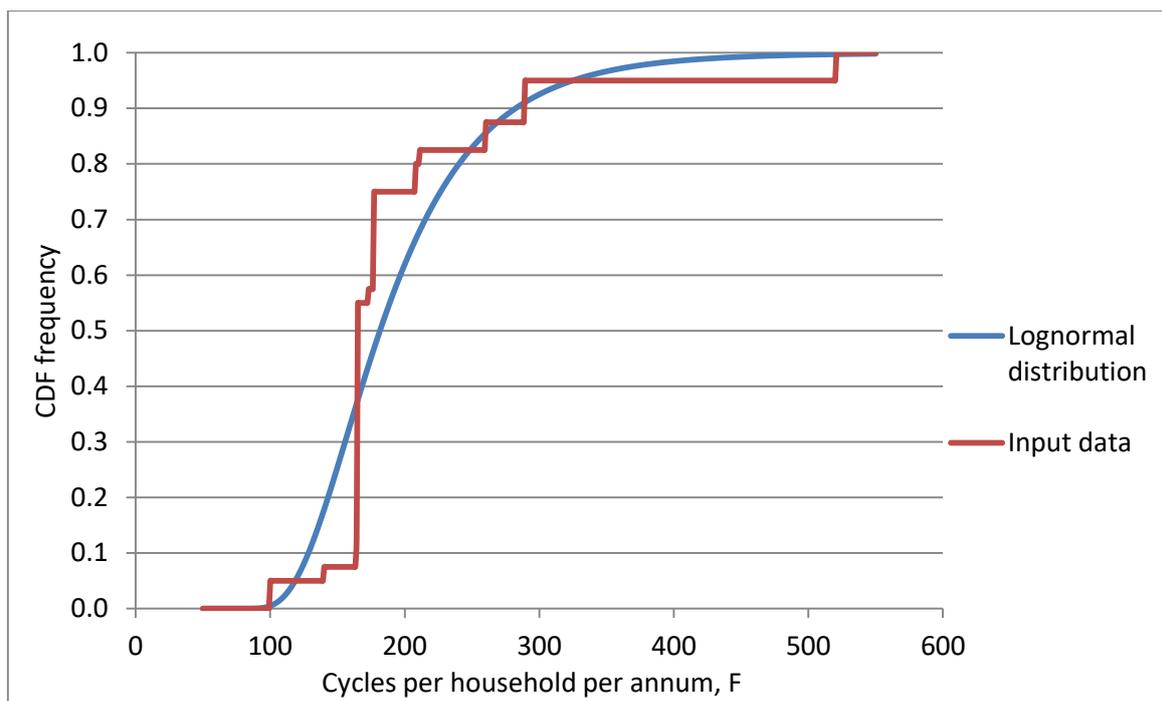


Figure 4.1. CDF washing cycles per household per annum

## FIELD EXPERIMENTS

The actual water use per wash cycle was measured at the point of use as part of this study. Two different appliance types (top loader and front loader) were monitored over a period of 4 weeks, in a controlled environment at a University residence in Stellenbosch, South Africa. The 4-week period was considered a sufficient amount of time to ascertain the typical operating benchmark for each machine. A water meter was installed at both appliances and meter readings were taken before and after each wash cycle. Estimates of event frequency, needed to estimate the washing event volume per person, were based on access control to the washing room. For the purpose of these field experiments, the water consumption for each washing cycle (washing, rinsing and drying) was not disaggregated. Additionally, due to the scope of the field work, the following factors contributing to the water consumption per cycle were not considered: water temperature; cycle duration; amount and type of detergent; size of washing loads.

## RESULTS

### Statistical model

The purpose of the rudimentary end-use model developed for Monte Carlo analysis was to generate a distribution curve representing the average water consumption by washing machines. Equation 4.1 was used to model the average water demand per household for top loaders and front loaders separately. For each type of washing machine, 1 000 000 iterations were performed and the resultant CDF curves of the front loader and top loader are plotted in Figure 4.2.

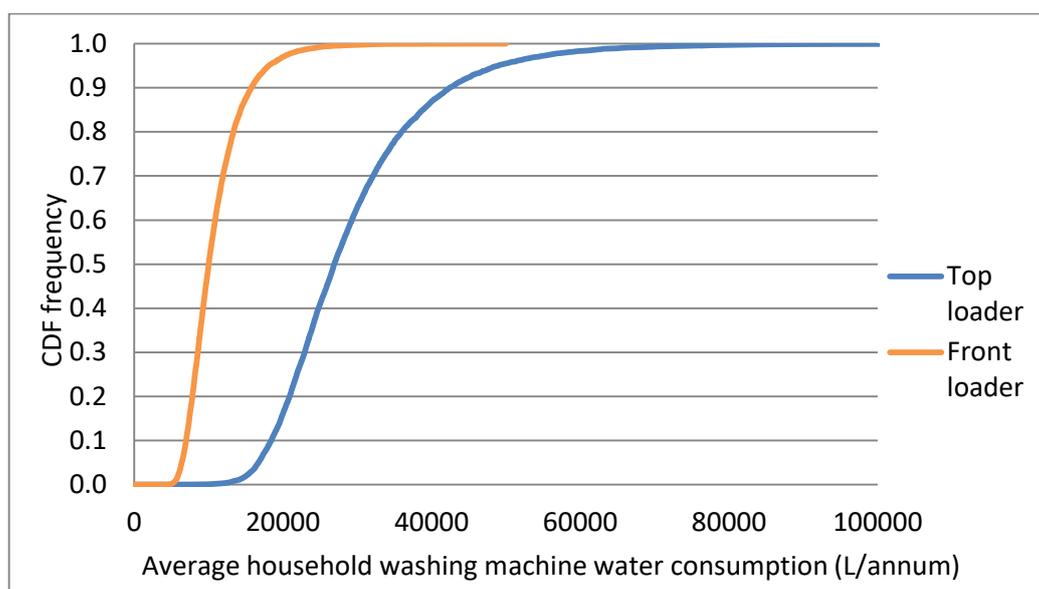


Figure 4.2. CDF of average washing machine water consumption based on model results

Figure 4.2 clearly shows the potential of significant water savings if a top loader were to be replaced by a front loader. The stochastic results estimate the average water consumption of a top loader to be 29 kL per household per annum. The average water consumption for a front loader was calculated to be 11 kL per household per annum. The average household indoor water demand at residential properties in Johannesburg, South Africa, was calculated to be 159 kL per annum (Jacobs et al. 2017). Thus, if a total indoor water demand of 159 kL per annum is assumed, the statistical estimates of the washing machines water consumption would contribute 18% (top loader) and 7% (front loader) to the total indoor water demand. These values are within the estimated range of 3% to 21% for residential houses as estimated by previous published literature (Jacobs et al. 2017).

The statistical model also illustrates that using a front loader instead of a top loader can potentially save a household 63% of washing machine water consumption on average per year (assuming constant load size). The main reason for the higher water usage of the top loaders is the bigger drum/bin sizes. The specific top loaders analysed had drum sizes ranging between 9 kg to 14 kg, whereas the front loaders' bin sizes ranged between 6 kg and 9 kg.

If the model were to include washing load weight as an independent parameter then the advantage (in terms of water saving) of a front loader would be less pronounced - almost double the number of washing events would be needed in a front loader, compared to a top loader, to clean the same weight of clothing.

### Field experiments

The Stellenbosch field experiments included a total of 54 washing machine events with an average measured event volume of 147 L/cycle for the top loader and 62 L/cycle for the front loader. The average duration of the washing cycles for the respective washing machines was 45 min for the top loader and 36 min for the front loader. All the recorded event volumes, based on water meter readings at the point of use, were sorted and recalculated to plot on a unit-scale. The subsequent CDF curves that fit the actual data are shown in Figure 4.3.

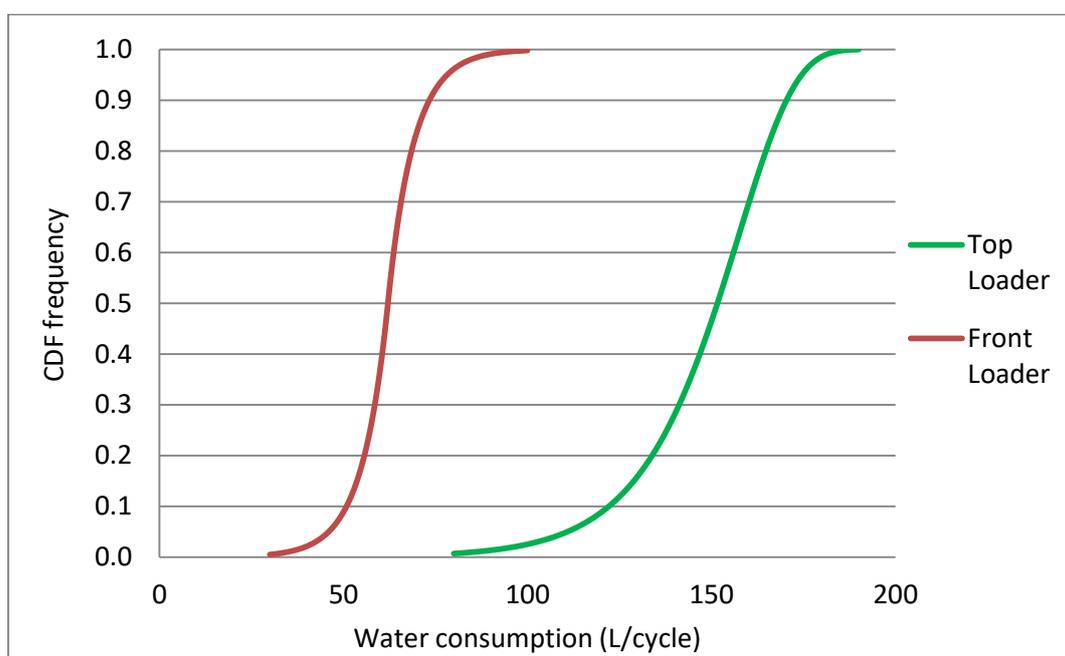


Figure 4.3. CDF of average washing machine water consumption based on field experiments

The stochastically derived expected water saving due to appliance change (as per this study) is 85 L/cycle. A potential water saving of 58% is thus possible if top loaders were to be replaced by front loaders.

### Comparison between rudimentary model and field experiments

The field experiments were compared to the stochastic results from the rudimentary end-use model for both the top loader and front loader. Figure 4.4 shows a superimposed graph of the resulting water consumption values.

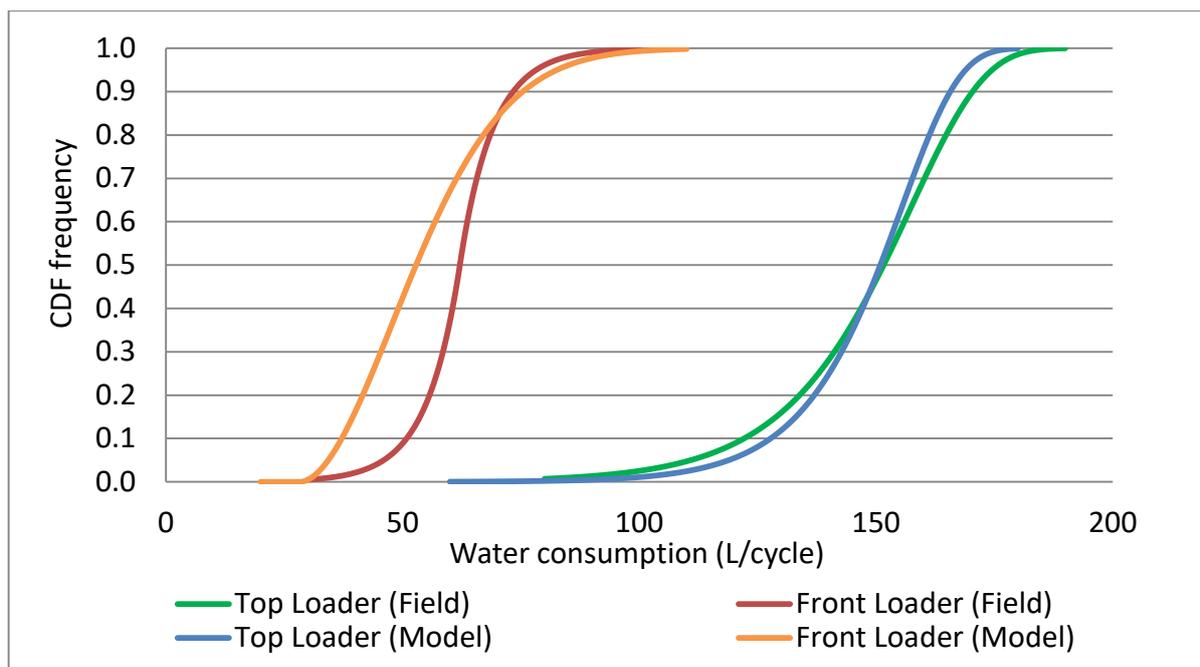


Figure 4.4. CDF comparison of model results and field experiments

The model overestimates the average water consumption per washing cycle for the top loader by <1% and underestimates the water consumption for the front loader by 14% when compared to the actual field experiment readings. The statistical model can thus be considered a good representation of top loader washing machine water consumption for the respective study site, while the model was able to provide a reasonable estimate for front loaders.

## CONCLUSION

A change of washing machine type may hold notable water saving potential assuming that wash frequency is independent of load size. The quantification of washing machine event frequency and water consumption per cycle for end-use studies are important. Measuring the water usage at the point of use can elucidate demand estimates. The field experiments conducted during this study measured the average water consumption of a top loader and front loader washing machine to be 147 L/cycle and 62 L/cycle respectively. The expected water saving due to appliance change as per this study is 85 L/cycle.

The rudimentary end-use model presented in the study provided acceptable results. The model overestimated the average water consumption per washing cycle by <1% for the top loader and underestimates the average water consumption by 14% for the front loader when compared to the field tests. The model estimated the average household water consumption for a top loader to be 29 kL/annum and for a front loader to be 11 kL/annum. Future research should include washing frequency field measurements to ultimately improve modelled results.

## REFERENCES

- P. Berkholz, R. Stamminger, A. Kruschwitz, A. Brükner, "Definition und Ermittlung verhaltensabhängiger Energiesparpotentiale beim Betrieb Elektrischer Haushaltswaschmaschinen. Bonn., 2006
- N.M.P. Bocken, C.A. Born and H.J. Lemstra, "Business experiments as an approach to drive sustainable consumption: The case of HOMIE", Product Lifetimes and the Environment (PLATE) Conference, Delft University of Technology, Netherlands, 8<sup>th</sup>– 10<sup>th</sup> November, 2017
- A. de Almeida and P. Fonseca, "Residential monitoring to decrease energy use and carbon emissions in Europe", Portugal: ISR-University of Coimbra, Dep. Electrical engineering, Paper ID 6273, 2006
- Department of the Environment, Water, Heritage and the Arts, "Equipment energy efficiency program", Volume 1, Annual Report 2007-2008
- H.E. Jacobs, A.A. Ilemobade and B.E. Botha, "End-use estimates of water demand within certain residential categories in the City of Johannesburg", Computing and Control for the Water Industry (CCWI) Conference, Sheffield, UK, 5<sup>th</sup>-7<sup>th</sup> September, 2017
- R. Nathan and E. Weinmann, "Monte Carlo simulation techniques. Australia Rainfall and Runoff", Engineers Australia, Water engineering, AR&R D2, 2013
- C. Pakula and R. Stamminger, "Electricity and water consumption for laundry washing by washing machine world wide", J.Energy Efficiency, 2010, DOI:10.1007/s12053-009-9072-8
- L. Polkoradi and B. Molnar, "Monte Carlo simulation of the pipeline system to investigate water temperature's effects", UPB Sci. Bull., Series D, 73(4), 223-232, 2011
- Statistics South Africa, "Living Conditions of Households in South Africa", Statistical release, P0310, Survey 2014/2015
- H.M. Scheepers and H.E. Jacobs, "Simulating residential indoor water demand by means of a probability based end-use model", J. of Water Supply: Research and Technology-Aqua, 63 (6), pp. 476-488, 2014

## Chapter 5.

### Assessing different measurement methods, data resolutions and end-use event characteristics

Bettina Elizabeth Meyer

1. Department of Civil Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa, [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za)

#### 5.1 INTRODUCTION

##### 5.1.1 Background

Water demand management (WDM) of the municipal supply network improves water security and forms part of the consideration for sustainable cities. Population growth (Vörosmary et al. 2005), urbanisation (McDonald et al. 2014) and higher standard of living (Mead and Aravinthan 2009) have, amongst other factors, caused water demand in cities to increase over the years. Residential water demand is influenced by various factors, including residential income (Willis et al. 2011), household size (Rathnayaka et al. 2015), property size (Fox et al. 2009), socio-demographics (Willis et al. 2013), water pressure (Meyer et al. 2018), garden size (Gato 2006), irrigation methods (Rathnayaka et al. 2015) and climate (Beal et al. 2011). Water demand also varies geographically. A change in residential water demand factors and water use trends have been observed over the years. Therefore, it is essential to update predictive water demand models for a specific geographical region, in the present time. Utilities make use of demand models to estimate water consumption for planning purposes. Demand models are also used as a tool to ensure the sustainability of water resources. Creaco et al. (2017) distinguishes between 2 types of demand models that operate at different spatial scales. The first type refers to models that predict water demand for the entire house as a whole, making no distinction between specific end-uses contributing to the water demand. The second type of demand model predicts water demand of specific end-uses and combines the end-use demands to construct a diurnal pattern for a house. Demand models at end-use scale allow for the evaluation of different WDM measures to predict the water savings potential as well as the impact of the WDM measures on waste water flow. End-use prediction models thus give insight into indoor water use, outdoor water use, and waste water flow.

Models that predict water demand based on end-use components include the Residential End-Use Model (REUM) presented by Jacobs and Haarhof (2004), and the SIMulation of water Demand, an End-Use Model (SIMDEUM) presented by Blokker et al. (2010). REUM makes demand predictions based on 111 different input parameters. The input parameters can be populated with data obtained from surveys, historic billing, monthly rainfall data, subjective evaluation and knowledge regarding the end-uses. The indoor end-uses included in REUM are the shower, washing machine, bath, toilet (large and small flush), kitchen tap, bathroom tap, miscellaneous indoor. The outdoor end-uses included in REUM are garden irrigation, swimming pool, leaks, and miscellaneous outdoor usage.

Although SIMDEUM is a predictive model, the model could also be described as a descriptive model, as it indicates the arrival time of each end-use and constructs the diurnal pattern for a home. SIMDEUM incorporates 8 end-uses, namely the shower, washing machine, bath, toilet, dishwasher, kitchen tap, bathroom tap and outdoor tap. The duration, intensity and frequency of use for each of these end-use are described by probability distributions. The distributions were developed using a variety of data. Information on flow intensities of appliances were obtained from technical information provided by manufacturers. Information regarding the duration and frequency of end-uses was obtained from 3,200 surveys conducted in the Netherlands. SIMDEUM assumes that each end-use event pattern is in the shape of a rectangle. Thus, the volume (L) of each end-use can be calculated by multiplying the duration (s) and intensity (L/s), since  $V = D \times I$ . In addition to the end-use characteristics, the surveys also provided insight into other model parameters. Other parameters incorporated into SIMDEUM are the number of people per household (PPH), age, gender, appliance ownership and hours spent awake at the home. Blokker et al. (2010) suggests that if the parameters of the statistical probability distributions can be populated, SIMDEUM can be applied to water networks at different locations.

The quality of the input data used for end-use modelling is thus of the highest importance to assure accurate results of the analysis (Van Zyl et al. 2003). One major limit of using demand models populated with surveys, verbal estimates, manufacturer specifications, or diaries, is the discrepancies between the perceived water use and actual water use. Inaccuracies in the input parameters could possibly lead to mismanagement of water operations due to prediction results being inaccurate. Mead and Aravinthan (2009) found that measured shower intensities at the study site differed from the manufacturer's estimates. Reasons for this could include varying water pressures or the shower tap not being fully closed. The water consumption of washing machines, as per manufacturer specifications, ranged from 68 L/cycle to 168 L/cycle for top loaders (Botha et al. 2018). Thus, if knowledge regarding the type of washing machine present at the home is not known (which is often the case), the substantial difference in water use (100 L) could contribute to inaccurate water predictions.

Roberts (2005) compared water usage estimates obtained from surveys, to the measured consumption data. Roberts (2005) reported that homeowners underestimate their garden irrigation duration by 33% to 40% (depending on the irrigation method), underestimate their shower duration by 13%, and underestimate the weekly frequency of washing machine use by 10%. Measuring the water usage at the point of use could elucidate demand estimates. Measured data at end-use level is thus important for accurate demand predictions and measuring actual water use is considered the most robust method for demand model population. Roberts (2005) reported that end-uses with the largest proportion of indoor consumption, for the end-use study conducted in Australia, were the shower (21.7%) and washing machine (18.7%). Multiple end-use studies reported similar results and concluded that the shower was the largest component of indoor water demand, with washing machine being the second largest. Shower use contributed to 43.5% (Mead and Aravinthan 2009), 30.9% (Beal et al. 2011) and 31% (Willis et al. 2011) of the total indoor demand at residential properties at the respective study sites. For the same end-use studies, washing machines contributed to 22.7%, 22.4% and 20% of the total indoor consumption. Toilets and taps (kitchen and bathroom combined) were the third and fourth biggest contributors to indoor water demand. Beal and Stewart (2011) reported that the three end-uses that vary remarkably between regions are showers, washing machines, and garden irrigation.

Outdoor use varies with season, with garden irrigation being a substantial proportion of outdoor use when a garden is present (Beal and Stewart 2013). Garden irrigation is more prevalent during hot and dry seasons (Mead and Aravinthan 2009). As part of an end-use study conducted in Australia, Roberts (2005) reported that garden irrigation's proportion of the total outdoor water demand is 87.1%. Due to the demand variability and large portions of total household water demand, garden irrigation, showers and washing machines were targeting during the first three end-use studies conducted as part of this research.

### **5.1.2 Aim and objectives**

An investigation into household water consumption should start with the end-uses which take up the largest component of the total demand. Therefore, the first part of this dissertation focussed on garden irrigation, shower and washing machine as household end-uses. Physical characteristics of these end-uses were measured and reported on. This chapter reports on these end-uses and addresses the remainder of the notable end-uses, namely the toilet, bath, tap and dishwasher. Different measurement methods for household end-uses were also evaluated.

The objectives were to:

- Give a renewed understanding of water end-uses and their physical characteristics;
- Conduct a comprehensive literature review reporting on different methods to retrieve household end-use data;
- Investigate the impact of different measurement resolutions on the usefulness of the data (what information can be extracted from the data).

## 5.2 END-USE EVENT CONSUMPTION: STATISTICAL DISTRIBUTIONS AND CONTRIBUTIONS OF DIFFERENT PARAMETERS

Buchberger and Wells (1996) proposed two types of end-uses, namely deterministic and random end-uses. Deterministic end-uses referred to end-uses such as the washing machine, toilet, dishwasher and bath, which have set volumes. The duration and intensity of an event would not affect the water consumption of deterministic end-uses. Multiple events of the same end-use would have similar flow patterns. Random end-uses, on the other hand, referred to end-uses that have high varying flow patterns and were very dependent on the event duration and event intensity. Random end-uses are thus largely influenced by residential habits. The volume consumed by random events could differ significantly from one event to another. Examples of random end-uses are the shower, garden irrigation and taps. In order to understand individual end-use consumption, the parameters that mostly influence the water consumption of each end-use, first needs to be understood. Makki et al. (2015) states that physical properties of end-uses, such as duration, intensity and frequency, are significant determinants of end-use consumption.

An end-use study conducted by Meyer and Jacobs (2019) found that event duration had the most notable contribution towards water consumed for garden irrigation. Although event frequency and event intensity contributed to the total consumption per garden irrigation event, a sensitivity analysis showed that garden irrigation event volumes were most sensitive to the event duration. The significant contribution of event duration to the consumption volume of garden irrigation could be explained by the notable parameter variability, with a relatively wide range of event duration values amongst different households. The wide range of durations could be ascribed in part to residents' behaviour and also to different types of garden irrigation methods. Roberts (2005) reported that the event duration of sprinklers on average, was 45% longer in duration than the hand-held hose irrigation method. Meyer et al. (2019) recommended garden irrigation duration to be assessed with a lognormal probability distribution. Blokker et al. (2010) also specified that the duration of garden irrigation events followed a lognormal probability distribution. Garcia et al. (2014) reported the exponential distribution to best fit garden irrigation end-use event durations.

Showers are also considered a random end-use event and event volumes differ significantly between residents. Subsequently, the flow patterns of different shower events would not be identical. Mead and Aravinthan (2009) found that flow intensities at showers are dependent on the type of shower head, water pressure and to what extent a shower tap is opened. Additionally, Mead and Aravinthan (2009) reported that shower duration is the parameter that most affects water consumption. Although the event duration of most random events could be assessed with a lognormal probability distribution, it is not necessarily true for all cases. Blokker et al. (2010) suggested measurements of end-use events are needed to better estimate durations of random events. The end-use study conducted by Botha et al. (2017) suggested the gamma distribution to best describe shower duration.

Washing machines and dishwashers use a fixed volume of water per event and were thus considered deterministic end-uses (Buchberger and Wells 1996). The flow intensities of washing machines and dishwashers do not affect the volume of water consumed per event. Botha et al. (2018) list various factors influencing washing machine consumption, namely the type of washing machine (top or front loader), machine model (brand), size of the washing machine (kg of clothes being washed) and washing cycles selected (e.g. delicates vs speed wash). The different factors influencing dishwashing event volumes include the machine model (brand) and the dishwasher setting (e.g. intensive, 70°C vs quick, 45°C). A linear regression exists between household size and washing cycles per household (Berkholz et al. 2006). The same correlation exists between household size and average number of dishwashing cycles (Roberts 2005). Subsequently, washing cycles per capita in a household decreases with an increase in household size (Botha et al. 2018). A comparison between measured results and surveys showed that residents overestimate their dishwashing frequencies by 1 load per week (Roberts 2005), which roughly equates to 23.9 L/week.

The duration of filling a toilet cistern is dependent on the bowl volume and water pressure, not the user. Toilets, although considered to be deterministic events, have various factors influencing the total consumption of a single event. Jacobs and Haarhoff (2004) distinguishes between conventional toilets (single flush) and dual flush toilets. Both types of toilets can further be categorised into large flush volumes (i.e. 9 L cisterns for dual flush toilets or 13 L cisterns for conventional toilets), and smaller flush volumes (i.e. 6 L cisterns). Foekema and Engelsma (2001) presented the relationship between flushing frequency and the person's age, with older people having a larger flush frequency than younger people. The frequency of toilet use can be described by a Poisson distribution, having only one parameter that needs to be populated, namely average frequency (Blokker et al. 2010).

Some end-use studies distinguish between kitchen taps, bathroom taps, and outdoor taps. The duration of a tap being opened is dependent on the user, which is why taps are considered random events. Similar to garden irrigation, Blokker et al. (2010) described tap duration as a lognormal distribution. Tap use was considered a high frequency and low volume end-use (Roberts 2005). Due to the low volume per tap event, WDM strategies generally do not focus on limiting tap usage. A summary of published literature reporting on the probability distributions of end-use characteristic parameters, are summarised in Table 5.1. The parameters that have the most significant influence on end-use water consumption are also depicted in Table 5.1.

Table 5.1. End-use statistical distribution parameters

End-use	Parameter	Statistical distributions	Literature	Most significant parameters
Shower	Duration	Exponential	Garcia et al. 2004	Duration, intensity, frequency
		Normal	Wong and Mui 2007	
		Geometric		
		X2	Blokker et al. 2010	
		Gamma	Cahill et al. 2013, Botha et al. 2017	
	Lognormal	Roberts 2005, Hand 2005, Blokker et al. 2010, DeOreo et al. 2011		
	Intensity	Weibull	Garcia et al. 2004	
		Normal	Wong and Mui 2007	
		Geometric		
		Log-logistic	Scheepers and Jacobs 2014	
	Volume	Uniform	Rosenberg 2007, Blokker et al. 2010, Hussien et al. 2016, Botha et al. 2017	
		Binomial	Blokker et al. 2010	
	Frequency	Log-logistic	Scheepers and Jacobs 2014	
Lognormal		Roberts 2005, Athuraliya et al. 2008		
Bath	Duration	Exponential	Garcia et al. 2004	Volume, frequency
		Normal	Hendron and Burch 2008	
		(fixed)	Blokker et al. 2010	
	Intensity	Weibull	Garcia et al. 2004, Scheepers and Jacobs 2014	
		Normal	Hendron and Burch 2008	
		(fixed)	Blokker et al. 2010	
	Volume	Uniform	Hand et al. 2005, Roberts 2005, Blokker et al. 2010, Grafton et al. 2011, Hussien et al. 2016	
		Rayleigh	Scheepers and Jacobs 2014	
	Frequency	Lognormal	Roberts 2005	
		Poisson	Blokker et al. 2010	
Washing machine	Duration	Exponential	Garcia et al. 2004	Volume, frequency
		(fixed)	Blokker et al. 2010	
		Beta general	Scheepers and Jacobs 2014	
	Intensity	Weibull	Garcia et al. 2004, Scheepers and Jacobs 2014	
		(fixed)	Blokker et al. 2010	
	Volume	Poisson	Blokker et al. 2010	
		Weibull	Scheepers and Jacobs 2014	
	Frequency	Normal	Buchberger and Wu 1995, DeOreo et al. 2001, Roberts 2005, Blokker et al. 2010, Pakulu and Stamminger 2010	
		Lognormal	Botha et al. 2018	

Table 5.1. End-use statistical distribution parameters (continued)

End-use	Parameter	Statistical distributions	Literature	Most significant parameters
Dishwasher	Duration	Exponential	Garcia et al. 2004	Volume, frequency
		(fixed)	Blokker et al. 2010	
		Log-logistic	Scheepers and Jacobs 2014	
	Intensity	Weibull	Garcia et al. 2004	
		(fixed)	Blokker et al. 2010	
		Erlang	Scheepers and Jacobs 2014	
	Volume	Poisson	Blokker et al. 2010	
Log-logistic		Scheepers and Jacobs 2014		
Frequency	Uniform	Roberts 2005, Blokker et al. 2010		
Toilet	Duration	Exponential	Garcia et al. 2004	Volume, frequency
		Normal	Wong and Mui 2007	
		Geometric	Wong and Mui 2007	
		(fixed)	Blokker et al. 2010	
	Intensity	Weibull	Garcia et al. 2004, Scheepers and Jacobs 2014	
		Normal	Wong and Mui 2007	
		Geometric	Wong and Mui 2007	
		(fixed)	Blokker et al. 2010	
	Volume	(fixed)	Blokker et al. 2010	
		Weibull	Scheepers and Jacobs 2014	
Frequency	Poisson	Mayer et al. 1999, Roberts 2005, Rosenburg 2007, Blokker et al. 2010, DeOreo 2011, Hussien et al. 2016		
Tap (bathroom)	Duration	Exponential	Garcia et al. 2004	Duration, intensity, frequency
		Normal	Wong and Mui 2007	
		Geometric	Wong and Mui 2007	
		Lognormal	Blokker et al. 2010	
	Intensity	Weibull	Garcia et al. 2004	
		Normal	Wong and Mui 2007	
		Geometric	Wong and Mui 2007	
		Uniform	Blokker et al. 2010	
	Volume	Gamma	Scheepers and Jacobs 2014	
		Lognormal	Scheepers and Jacobs 2014	
Frequency	Poisson	Blokker et al. 2010		

Table 5.1. End-use statistical distribution parameters (continued)

End-use	Parameter	Statistical distributions	Literature	Most significant parameters
Tap (kitchen)	Duration	Exponential	Garcia et al. 2004	Duration, intensity, frequency
		Lognormal	Blokker et al. 2010	
	Intensity	Weibull	Garcia et al. 2004	
		Uniform	Blokker et al. 2010	
		Gamma	Scheepers and Jacobs 2014	
	Volume	Lognormal	Scheepers and Jacobs 2014	
Frequency	Negative binomial	Blokker et al. 2010		
Garden irrigation	Duration	Exponential	Garcia et al. 2004	Duration, intensity, frequency
		Lognormal	Blokker et al. 2010, Meyer and Jacobs 2019	
	Intensity	Uniform	Blokker et al. 2010	
		Weibull	Garcia et al. 2004	
		Lognormal	Roberts 2005	
		PERT	Meyer and Jacobs 2019	
	Frequency	Poisson	Blokker et al. 2010	
		Triangular	Roberts 2005, Hussien et al. 2016	
Binomial		Meyer and Jacobs 2019		
Tap (outside)	Duration	Exponential	Garcia et al. 2004	Duration, intensity, frequency
		Lognormal	Blokker et al. 2010	
	Intensity	Weibull	Garcia et al. 2004	
		Uniform	Blokker et al. 2010	
	Frequency	Poisson	Blokker et al. 2010	
Swimming pool	Volume	Lognormal	Jacobs and Haarhoff 2004, Fisher-Jeffes et al. 2015	Volume, frequency
	Frequency	Triangular	Jacobs and Haarhoff 2004, Fisher-Jeffes et al. 2015	
Car wash	Volume	Uniform	Rosenberg 2007, Janik and Kupiec 2007, Smith and Shilley 2009	Volume, frequency
	Frequency	Discrete	Hussien et al. 2016	
Leaks	Duration	Exponential	Garcia et al. 2004	Duration, intensity
	Intensity	Weibull	Garcia et al. 2004	
		Gaussian	Cody et al. 2020	
	Volume	Uniform	Mayer et al. 1999, Liu et al. 2016	

### 5.3 END-USE MEASUREMENT METHODS

Water conservation methods, such as restricting flow intensities at taps, smart irrigation systems and monitoring systems for automatic leak detection, are implemented by utilities in an attempt to reduce the gap between an increase in water demand and a decrease in water resources. However, prior to undertaking any of these initiatives, knowledge regarding where and how water is used at a home first needs to be understood (Hauber-Davidson and Idris 2006). End-use event characteristics can be investigated in multiple ways. Measurements can be taken at a single point at a residential property, or at the point of use (at the end-use). Measuring approaches can further be divided into direct flow sensing methods and indirect flow sensing approaches, or a combination of these.

#### 5.3.1 Single point direct end-use measurements

Utilities typically measure household consumption at the point of entry, recording the entire property consumption with a single water meter. Utility meters are read manually and readings are commonly taken at daily, monthly or even quarterly frequencies. Typical resolutions of these types of meters are set at 0.5 L/pulse, 1.0 L/pulse, or 1.0 kL/pulse (Roberts 2005, Nguyen 2013). Mechanical meters provide insight into monthly household water consumption, but no further information is available due to the limited and delayed water consumption information. Some studies have recorded residential water use at hourly intervals (Cardell-Oliver et al. 2016), providing information on peak-hour demands, peak-day demands and anomalous events such as leaks, which is important to ensure resilient water network infrastructure. However, due to the meter pulse volume and low frequency of reading, no insights were given into household consumption at end-use level. Mechanical water meters were not utilised nor able to identify where and how water is used at a home.

Smart meters have received attention in end-use studies to better understand how and where water is used at a residential property. Smart meters are mechanical water meters linked with loggers, allowing for automated data measurement readings and real time monitoring. In addition, smart meters record at higher resolution frequencies ( $< 10$  s), with data accessible via the internet (Giurco et al. 2008). End-use studies employing high resolution smart meters (0.014 L/pulse) at recording frequencies of 1 s (Kowalski and Marshallsay 2005, Buchberger and Wells 1996), 5 s (Nguyen et al. 2013, Beal and Stewart 2013, Roberts 2005), and 10 s (Mead and Aravinthan 2009, Willis et al. 2009, Mead 2008, Heinrich et al. 2007, Mayer et al. 1999), are able to disaggregate flow patterns into different end-uses. The high resolution data allow for sophisticated analysis, with less assumptions needed for accurate end-use modelling (Beal and Stewart 2013).

End-uses can be identified from high resolution measured data by utilising disaggregation software. Trace Wizard™ (Aquacraft 2010), a flow trace analysis tool, is one of the most popular disaggregation software tools on the market (Nguyen et al. 2013). Trace Wizard™ requires a substantial amount of input parameters, including the minimum, maximum and most frequent values of duration, volume and intensity for each of the end-use categories. Cross-checking actual end-use events classified with entries from user diaries showed that Trace Wizard™ had an approximate accuracy of 72 % (Nguyen et al. 2013). Higher accuracies could be achieved when experienced analysts manually check each end-use event being apportioned, which would be extremely resource intensive and not practical for large end-use studies. The concept of flow pattern recognition was further refined, and more accurate, automated water end-use disaggregation tools have been developed for commercial applications. Subsequent flow trace analysis software include Identiflow (Kowalski and Marshall 2003), Autoflow (Nguyen et al. 2013), BuntBrain (Arregui 2015), REU2016 (Vitter and Webber 2018), SmartH2O (Cominola et al. 2018) and Autoflow<sup>U</sup> (Nguyen et al. 2018). These commercially available automated flow trace analysis tools are able to disaggregate end-uses achieving accuracies of 72-93% (Nguyen et al. 2018). These end-use studies have made significant contributions towards understanding household water demand and end-use level. However, smart meters with such high recording resolutions are uncommon. Employing high resolution smart meters over a large spatial scale is not (yet) viable, especially in developing countries. Consumption data are often only available at a reduced temporal or spatial resolution due to large data storage capacity requirements, and the resource intensive and costly nature of recording high resolution data (Ilemobade et al. 2018, Nguyen et al. 2013).

### **5.3.2 Single point indirect flow sensing approaches**

Different types of indirect measurement methods have been investigated in previous studies. The application of indirect flow sensing approaches have received attention in the past due to the devices typically being small, non-intrusive and more cost effective than direct measurement devices.

Evans et al. (2004) employed accelerometer to the surface of outflow pipes in order to investigate the possibility of identifying flow intensities at particular end-uses. Events were identified based on vibrations in the pipe induced by water flowing through the pipe. Distinguishing between vibrations caused by water flowing through the pipe and the background noise proved challenging. Other studies, which also investigated vibration sensors as indirect flow sensing approach, reported similar challenges with regards to vibration or noise interference from nearby pipes or devices (Pirow et al. 2018, Kim et al. 2008, Fogarty et al. 2006). Additionally, the vibration patterns were dependant on the pipe materials and the connection between the sensor and the pipe. Devices thus have to be calibrated for accurate estimations (Kim et al. 2008). Consequently, knowledge regarding pipe specifications are needed prior to installation (Evans et al. 2004).

Pipe vibrations have also been recorded using sound recording devices. The sound recorded is essentially the vibrations of water flowing through a pipe (Young et al. 2012). Microphone-based sensing was conducted by Fogarty et al. (2006), placing 4 microphones at strategic positions in a home in order to classify recorded pipe noise into individual hot and cold water end-use events. One microphone was placed on the cold water inlet pipe, another on the outlet pipe of a geyser, and the remaining two microphones were placed on the kitchen and bathroom drain pipes, respectively. The feasibility study conducted by Fogarty et al. (2006) demonstrated that recorded sound patterns can successfully be used to classify end-use events. However, one major limitation to the recorded sound method proposed by Fogarty et al. (2006) is that the algorithm can only be implemented on pre-segmented data. Meaning, the start and end times of an event had to be flagged before the algorithm could classify an event. The start and end times of events were flagged by measuring end-use events at the point of use. The algorithm is thus not able to extract end-uses from the recorded set, and cannot be implemented without the aid of point of use measurements.

Pressure sensors have been employed at residential properties, monitoring continuous water pressures at a single point in a home. HydroSense, proposed by Froehlich et al. (2009), is able to quantify and estimate household end-use frequencies and volumes. HydroSense, however, cannot distinguish between multiple events occurring at the same time (overlapping of events). Additionally, HydroSense was only implemented in a controlled environment, and the application thereof in a “real world” scenario has not been verified. Subsequently, Froehlich et al. (2011) developed a pressure based inference algorithm to identify end-use events in a real world scenario, and were able to successfully identify between 76% and 98% of end-use events at the study sites. Although the accuracy of the classification rate is high, the algorithm requires extracted end-use events prior to classification. Similar to the study conducted by Fogarty et al. (2006), the start and end times of end-uses first have to be identified by means of point of use measurements, before the algorithm can accurately classify the events into specific end-uses.

Ultrasonic water meters can be used to identify end-use events. The device uses ultrasonic transducers to measure the velocity of the water flowing through a pipe. External software is then used to characterise the end-use events (Paulsen et al. 2001). Although the method is accurate and non-intrusive (no plumbing changes are required), the cost of ultrasonic meters is roughly 7 times more expensive than the other indirect flow sensors (Sterne 2019), and is thus not an attractive alternative.

Specifically focussing on hot water end-uses, Nel et al. (2015a) and Nel et al. (2015b) placed temperature sensors at the outflow pipes of geysers. The main purpose of these studies were to measure hot water consumption patterns to ultimately estimate energy usage of electric water heaters. The indirect flow sensing approach was paired with an inline meter installed at the inflow pipe of the geyser in order to quantify the volume of water consumed per event.

Similar to the approach followed by Massuel et al. (2009), Nel et al. (2015a, 2015b) identified a hot water end-use event by recognising a change in temperature in the outflow pipe. An increase in the pipe wall temperature would constitute the start of an event (water flowing through the pipe), and a decrease in the pipe wall temperature would indicate the end of an event. The temperature threshold used to identify events were derived empirically. The algorithm developed to detect hot water usage was able to identify 91% of the events recorded by the water meter, however, no consideration was made for overlapping of events (two hot water events occurring at the same time). Events were classified into three different categories, small, medium and large events. Although both Nel et al. (2015a) and Nel et al. (2015b) provided value information regarding hot water usage patterns at a home, recorded events were not categorised into specific end-uses. Thus, specific end-use characteristics were not reported on in these studies.

### **5.3.3 Point of use direct flow sensing approaches**

Indirect measurement methods could be employed at the point of use to obtain ground truth end-use characteristics. To the best of the author's knowledge, no published end-use studies have been conducted in the past measuring end-use water consumption at the point of use with a direct flow measuring device, such as a water meter. The first end-use study conducted, measuring washing machine water consumption at the point of use with a mechanical water meter, was presented in Chapter 4 (Botha et al. 2018). Although employing water meters at the point of use is the most accurate means of quantifying water consumption at end-use level, the application thereof over a large scale is not practical. Water meters would have to be installed between pipe segments, requiring plumbing expertise. Additionally, such installations would be intrusive, and retrofitting pipes over a large scale would be tedious and expensive.

### **5.3.3 Point of use indirect flow sensing approaches**

Indirect flow sensing devices employed at the point of use include sound recording devices (Makwiza and Jacobs 2007), vibration sensors (Stern 2019, Froehlich et al. 2011), and temperature loggers (Massuel et al. 2009). Makwiza and Jacobs (2017) were able to identify recognisable sound waves from sound recording devices installed at outdoor taps in Malawi. The sound recording devices were able to determine the start and end times of outdoor tap events. The data were subsequently analysed to derive event durations and event frequencies. Although no audio (conversations) were recorded during the study, people were still very sceptical and reluctant to install sound recording devices inside their homes, for privacy reasons. Makwiza and Jacobs (2017) achieved precision and recall rates of 80%.

Sterne (2019) developed a tool that recorded vibrations at an inlet pipe as water was flowing through the pipe. The devices were placed at the inlet pipe of a dishwasher, washing machine, shower and outdoor tap, and on the showerheads at two residential properties. Similar to challenges reported earlier by Evans et al. (2004), Sterne (2019) found that a lot of “noise” vibrations were recorded in addition to the vibrations induced by actual events. In addition to the vibration sensors, a water meter was installed at a single point of entry at each house, recording at a volumetric resolution of 0.5 L/pulse. The water meter was used in order to distinguish between actual water use events and “noise” recorded by the vibration sensors. The vibration sensors can thus not be implemented on its own without the water meter. Sterne also noted that if two pipes are adjacent to each other (which are often the case in household plumbing), the vibrations from the one pipe will be picked up by the nearby vibration sensor, resulting in false positives. When paired with the smart meter, the vibration sensors were able to successfully identify between 76.5% and 88.0% of the end-use events.

Froehlich et al. (2011) combined 7 different indirect flow sensing approaches, employing the most suitable indirect flow sensing approach to each of the different end-uses, in order to identify household end-use events. The flow sensing approaches included accelerometers, reed switches, magnets and ball switches. Using this combination of flow sensing approaches, Froehlich et al. (2011) were able to distinguish between hot water events, cold water events, and a combination of hot and cold water use events. The main purpose of this point of use measurements was to calibrate, verify and validate the pressure based identification algorithm. The algorithm was used to analyse the data recorded by the pressure sensor installed at a single point at the property. This combined flow sensing approach is not recommended for point of use end-use studies. The setup of all the devices is tedious, resource intensive and impractical for larger application (the installation takes 2 people 2 full working days per residential property).

Massuel et al. (2009) employed temperature loggers at the outflow pipes of groundwater abstraction points, in southern India, evaluating pumping durations of water withdrawals for agricultural irrigation. Temperature loggers were placed on the outflow pipes of wells, and a pumping event was identified by temperature variation in the pipe. The robust, small and relative inexpensive temperature loggers are quick and easy to implement, as the loggers was simply taped to the outflow pipes. The method was tested against the electricity usage of the pumps at the boreholes and wells. Massuel et al. (2009) could successfully identify the duration of pumping events, with a 1.2% difference in duration from the electrical pump durations. The first implementation of temperature loggers at the point of use at residential properties were conducted by Botha et al. (2017) and Meyer et al. (2019).

#### **5.4.1 Temperature logger as indirect point of use measurement method at residential properties**

Previous chapters explored the use of temperature loggers as an indirect method to identify the duration, frequency and time of day of water end-use events. Temperature loggers (iButtons) were used to estimate event start and finish times by measuring the pipe wall temperature of outflow pipes. Temperature loggers were selected based on availability and budget. The temperature loggers were also easy to install, and calibration prior to installation was not needed, since the effect the pipe material and the logger-to-pipe connection have on the results are negligible. Meyer and Jacobs (2019) and Botha et al. (2017) placed temperature loggers at the point of use and recorded the change in temperature at the outflow pipe of ground water abstraction points (GAPs) and on shower heads, respectively.

Similar to the method employed by Massuel et al. (2009), the start and finish times of each event was successfully identified by analysing the temperature change experienced when water was flowing through the outflow pipe and shower head. The event duration and frequency of use of the shower and garden irrigation events were subsequently successfully identified. The implementation of temperature loggers are, however, limited to hot water end-uses, or end-uses with long duration events where a temperature variation is evident (e.g. garden irrigation) (Meyer and Jacobs 2019). The temperature loggers are not expected to be successful when employed on end-uses with relatively short duration events (less than 2 min) or on events where the temperature variation is small. It would thus not make sense to expand the use of temperature loggers to identify household end-use components such as the toilet or tap.

#### **5.5. MEASUREMENT RESOLUTION AND THE IMPACT ON EVENT SIGNALS**

Cominola et al. (2018) investigated the trade-off between information gained from high resolution data and the cost of the smart meters needed to record at such high resolutions. The study reported that end-uses cannot be extracted from a time series if the recording resolution is longer than 1 min. Only studies that use smart meters with sub-minute recording frequencies are able to extract and classify end-uses. At the time this research paper written, to the author's best knowledge, no study has investigated the trade-off between the meter pulse volume and the extent of information gained from the metered data. Future research should investigate to what extent measured data, which are too coarse for commercially available end-use disaggregation tools, could be used to obtain water end-use demand information at a household level. This section was included to briefly explain the impact of coarser resolution measurements on end-use event flow patterns. The shower, a typical large event (long duration and total volume) and the kitchen tap (various minor events in succession while washing up) were assessed.

A typical shower event, measured at a resolution of 0.014 L/pulse at 10 s recording frequencies, is presented in Figure 5.1(a). The shower event has a duration of 410 s and total event volume of 51.2 L. The intensity (flow rate) – after opening both taps – was relatively constant at about 0.127 L/s for the duration of the event. During the event, 1.274 L of water would pass the water meter each 10 s, on average. As the recording resolutions are changed to 0.1 L/pulse, 0.5 L/pulse and 1.0 L/pulse, a change in flow pattern is observed, as depicted in Figure 5.1(b), Figure 5.1(c), and Figure 5.1(d), respectively.

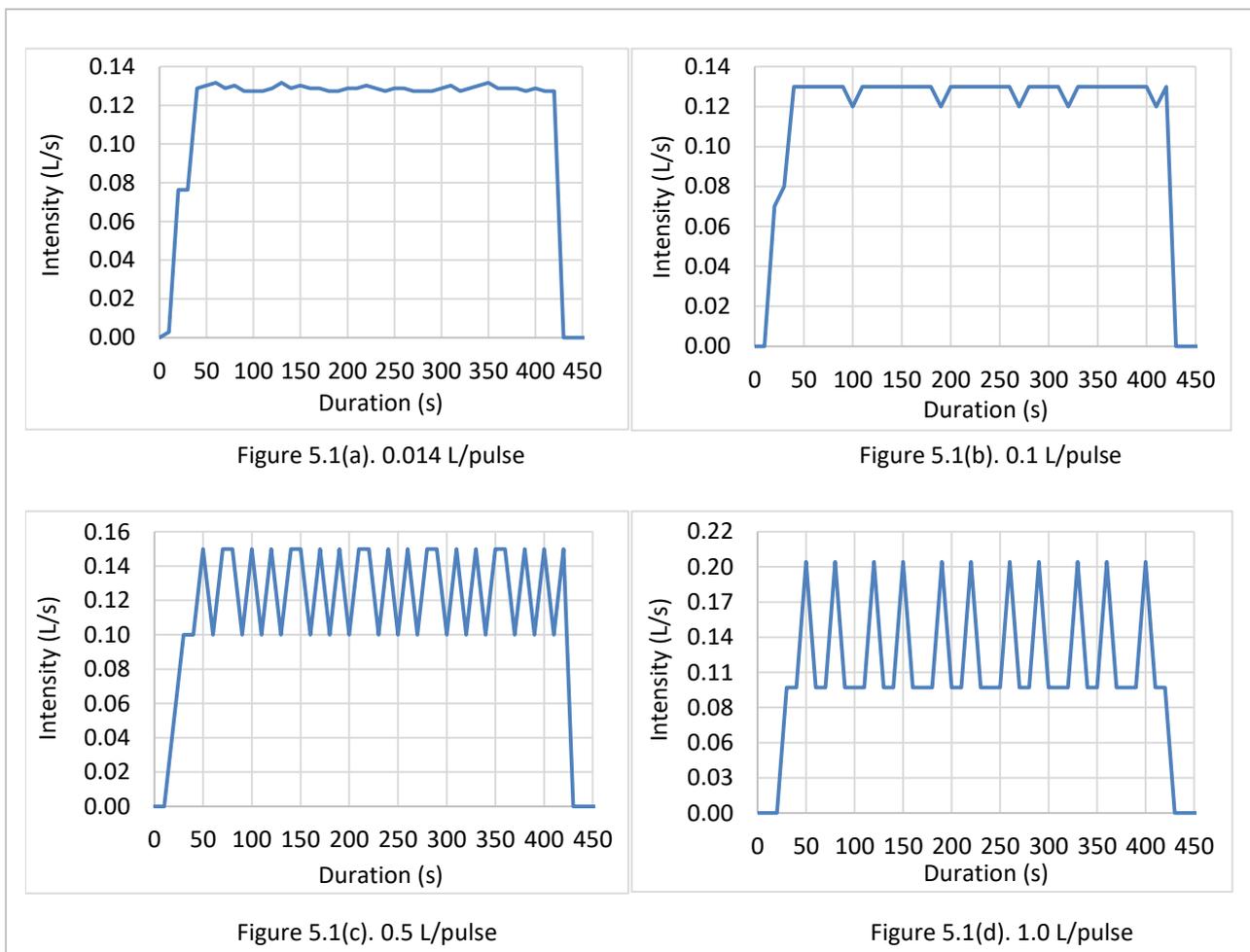


Figure 5.1. Flow pattern of a typical shower event at different recording resolutions

Due to the rudimentary nature of the 0.5 L/pulse recorded data, either 0.5 L or 1.0 L would pass through the meter every 10 s during the shower event. Similarly, the volume that would pass the meter every 10 s for the 1.0 L/pulse resolution, fluctuates between 1.0 L and 2.0 L, for example. This best explains why the flow patterns appear to be “spiky” for coarser data sets.

The physical characteristics of the shower event over the different recording resolutions differed slightly. Because the logging frequency stayed constant at 10 s, the duration of the shower event stayed the same over the different recording resolutions, at 410 s. The actual shower volume ranged between 51.0 L and 51.7 L, and the average shower intensity ranged between 0.124 L/s and 0.128 L/s. Similar changes in flow pattern could be expected from a typical filling-of-the-bath event.

A typical series of kitchen tap events during dish washing (total cumulative duration 320 s and total cumulative volume 9.35 L) is shown in Figure 5.2. The different flow patterns for each of the recording resolutions, namely 0.014 L/pulse, 0.1 L/pulse, 0.5 L/pulse and 1.0 L/pulse, are shown in Figure 5.2(a), Figure 5.2(b), Figure 5.2(c), and Figure 5.2(d), respectively.

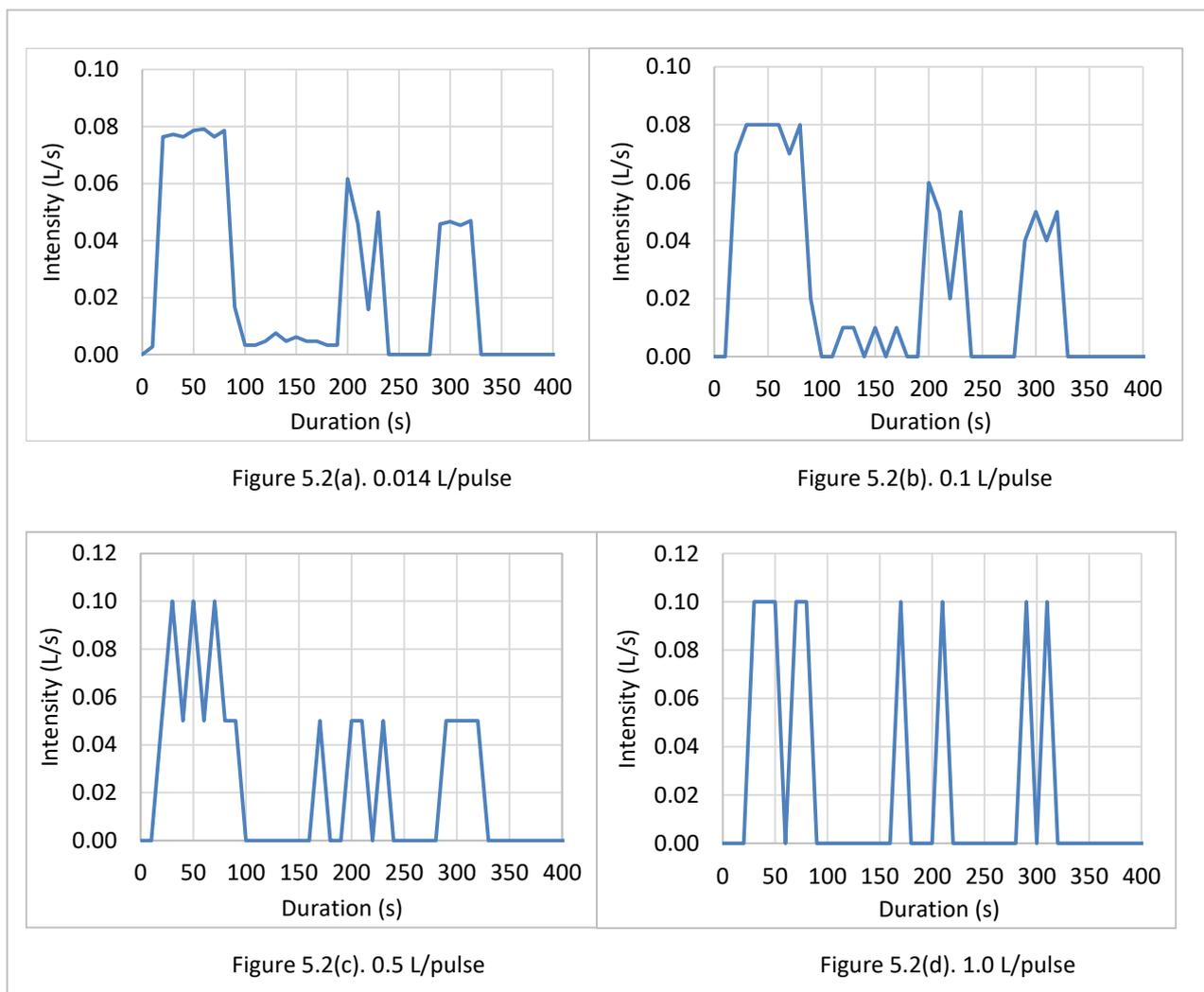


Figure 5.2. Flow pattern of various minor tap events at different recording resolutions

Two different minor tap events, with short durations and low flow intensities, are shown in Figure 5.2(a). The first event indicates a tap being opened fully, then partly closed for a period of time (extremely low intensity), and opened again for a short stint before being closed

completely. The second event shows the tap being opened for 30 s, with a relative constant intensity of 0.047 L/s, and closed again. A significantly different flow pattern is observed in Figure 5.2(c) and Figure 5.2(d). The tap events comprise small segments, making it difficult to identify periods when the tap would be open/closed, resulting in typical spikes. The rudimentary data resolution of 1.0 L/pulse is only able to record a measurement if 1.0 L of water has passed the meter. As a result, the flow pattern in Figure 5.2(d) looks like 6 minor end-use events occurring at the residential property, instead of the actual 2 events. This difference in flow patterns can further be explained by the following example.

If pulse measurements are only read every 1.0 L, an actual tap event of 0.9 L would not be recorded. If another end-use event takes place later during the day, say a tap event of 0.25 L (filling a glass), the second event would be recorded as a 1.0 L event, with an intensity of 0.1 L/s (a 1 L event passed over 10 s). Figure 5.1 suggests that longer and larger events remain recognisable, even when the volumetric resolution is reduced to 1.0 L/pulse. Figure 5.2 explains why the identification of small minor events, such as taps, would be difficult to identify with rudimentary data. The assessment of the different pulse volume resolutions suggests that relatively large events (long duration and total volume) would be detectable with coarser data, while smaller events could go missing.

## 5.6 CONCLUSION

Knowledge regarding household water consumption at end-use level is important for effective WDM strategies and water security. Household water consumption can be measured with direct or indirect flow sensing approaches, at a single point on a property, or at the point of use. Point of entry direct flow sensing approaches, such as smart meters, have been used in the past to record household water consumption. Coupled with flow trace analysis software, the time series data can be disaggregated into individual end-use events. Developing countries have identified the need for smart meters, however, it has not yet been implemented. A few reasons for this could include the costs, data storage requirements, resource needed for analysis and limited product availability.

Indirect flow sensing methods hold numerous advantages, especially if the event duration and frequency of use is required as key parameters. The three end-uses targeted in earlier chapters of this dissertation include garden irrigation, shower, and clothes washing (washing machine). Temperature loggers were used to determine the event duration of end-uses with a long duration (irrigation, showers) or with hot water (showers). Temperature loggers were chosen based on cost and availability. Data obtained from indirect flow sensing approaches, such as temperature loggers, provide valuable input parameters to populate theoretical demand models such as REUM or SIMDEUM. Temperature loggers, however, were not applicable on all end-uses.

The resolution of data recorded by conventional mechanical meters is not intended for end-use classification. A knowledge gap exists when it comes to household end-use level water consumption, in the presence of rudimentary data. Future research should determine what level of detail could be extracted from rudimentary data recorded by mechanical water meters. Distinguishing between indoor use and outdoor use could improve the development, implementation and monitoring of WDM strategies. As a first step, future research could investigate whether or not end-use events could be extracted from rudimentary data sets. The subsequent goal should be to determine the possibility of classifying the extracted events into specific end-uses such as the shower, washing machine, toilet, bath, dishwasher, tap, garden irrigation, and leaks. If end-use disaggregation is not possible, household water consumption behaviour at end-use level, in terms of indoor use and outdoor use needs to be investigated.

## Chapter 6.

### Extracting household water use event characteristics from rudimentary data

Bettina Elizabeth Meyer<sup>1</sup>, Heinz Erasmus Jacobs<sup>1\*</sup> and Adeshola Ilemobade<sup>2</sup>

1. Department of Civil Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa, [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za), [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za)

2. School of Civil and Environmental Engineering, University of the Witwatersrand Johannesburg, Private Bag 3, WITS, 2050, South Africa, [Adesola.Ilemobade@wits.ac.za](mailto:Adesola.Ilemobade@wits.ac.za)

\*Corresponding author: [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za)

Reproduced from the Journal of Water Supply: Research and Technology - Aqua (2020), 69(4): 387–397.

Available online: <https://doi.org/10.2166/aqua.2020.153>

#### ABSTRACT

Household water end-uses have been extracted from high resolution smart water meter data in various earlier studies. However, research on end-use disaggregation from rudimentary data is limited. Rudimentary data are defined as data recorded in intervals longer than one minute, or data recorded with resolutions larger than 0.1 L/pulse. Developing countries typically deal with rudimentary data, due to the high cost and high resource investment associated with high resolution data. The aim of this study was to extract useful event characteristics from rudimentary data, without identifying the actual end-uses per se. A case study was conducted in the City of Johannesburg, South Africa, where 63 homes were equipped with iPERL smart water meters. The meters recorded flow measurements every 15 s at a 1 L/pulse resolution, rendering the recorded data rudimentary. A total of 1 107 547 event pulses were extracted over the 217-day study period. Although the method presented is limited in the sense that water use events cannot be identified, the method allows for disaggregation of event pulses in the presence of rudimentary data. Using this tool, it is possible to lift valuable information from rudimentary data that would subsequently benefit service providers in setting water demand strategies.

**Keywords:** data resolution, end-use, household water demand, smart meters, water use

## CHARACTERISTICS OF WATER END-USE EVENTS

End-uses of water, such as the shower, toilet, tap and washing machine, are considered the building blocks of the residential water demand pattern (Buchberger and Wu 1995). The relationships between an end-use event's characteristics, namely duration, intensity, and volume, create a unique end-use "fingerprint". Each end-use event "fingerprint" is typically represented by a rectangular pulse (Buchberger and Wu 1995, Alcocer-Yamanaka et al. 2012). Extracting and identifying end-uses from high resolution data was pioneered by De Oreo et al. (1996), and subsequent investigations include Mayer et al. (1999), Loh and Coghlan (2003), Beal et al. (2011), DeOreo et al. (2011), Beal and Stewart (2013), Arregui (2015), Nguyen et al. (2013, 2018) and Pastor-Jaboloyes et al. (2018). However, extracting end-use events from rudimentary data sets, and utilizing the relationships between event characteristics to categorise extracted end-use events, has yet to be explored.

## DATA RESOLUTION

Developing effective demand management strategies requires a clear understanding of household water consumption (Jorgensen et al. 2013). Water consumption at a home is typically measured using water meters. Meter readings could be time-based or event-based. In the case of time-based recordings, flow volume through the meter would be averaged over time and recorded at fixed intervals of say 1 s (Kowalski and Marshallsay 2003, Buchberger and Wells 1996), 5 s (Beal and Stewart 2013, Roberts 2005), 10 s (Stewart et al. 2009, Mayer et al. 1999), 15 min (Pretorius et al. 2019), or 1 h (Cardell-Oliver et al. 2016). Disaggregation of end-uses from a time series requires water end-use data to be collected at a sub-minute resolution (Cominola et al. 2018).

Alternatively, event-based recording involves meter readings taken per water meter pulse – a water meter producing one pulse per litre would not be able to record end-use events smaller than 1 L, for example. Domestic consumer water meters typically used in South Africa, where the case study was undertaken, provide one pulse per litre; the smallest pulse volume commercially available in South Africa at the time of this study was 0.5 L/pulse. At the time of this study, two of the most accurate pulse volumes reported were 0.014 L/pulse (Beal et al. 2011) and 1000 pulses per L, or 1 mL per pulse (Otaki et al. 2011). To date, the lowest data resolution used for end-use disaggregation and classification was found in a study conducted by Pastor-Jaboloyes et al. (2018), employing volumetric water meters generating a pulse every 0.1 L. Data obtained from meter readings with pulse volumes higher than 0.1 L/pulse (which was the case for this study) were thus deemed rudimentary.

## **MOTIVATION AND AIM**

The required resolution for end-use analysis is not typically available or accessible to service providers in developing countries, due to various constraints (financial, human resources, limited technical expertise, etc.). In order for developing countries to utilise rudimentary data as a vital tool for water demand strategies, a method is needed to classify end-uses into indoor use and outdoor use based on the relationships that exist between basic event characteristics (i.e. duration, intensity, volume). Before end-use events can be classified, the end-use needs to be extracted from a rudimentary data set. This paper addresses the latter problem. The aim of this study was therefore to extract event characteristics from a rudimentary time series data set, without the need to identify the end-use in question. Also, this study set out to develop a procedure that will identify major end-uses from rudimentary data.

## **STUDY SITE AND CONSUMER SURVEY**

The study site was located in Lonehill, a suburb north of Johannesburg, South Africa. The study sample comprised 63 suburban homes, of which 9 were stand-alone single family homes and 54 were single-family, semi-detached town houses, located inside a gated community. Gated communities are common in South Africa and earlier studies provide more detail about this relatively high-income dwelling type (Du Plessis and Jacobs 2018).

Following an ethical approval process, the project team embarked on a comprehensive consumer survey and water audit process by visiting selected homes, interviewing selected individuals and distributing survey questionnaires to all homes in the sample. Thirty-two completed survey responses were received, with the team visiting 6 homes as part of the research process. The average household size for the survey respondents was 1.9 people per household (PPH), with the maximum household size being 4 PPH. Roughly half of the sample reported single-person dwellings, and 29% of the survey respondents reported a household size of 2 PPH.

## **DATA COLLECTION AND SORTING**

Each home in the sample was equipped with a smart water meter, recording the total consumption of each property. In order to identify household end-uses from the recorded flow rate profile, a relatively small volume per pulse and a relatively short time interval would be required. As part of this study, the Sensus iPERL (International) smart water meters were used. The iPERL has integrated bi-directional communications capability and high measurement accuracy. Data were collected between 5 September 2016 and 29 January 2018 from all 63 homes. The iPerL smart meters measured flow volume to a resolution of 1 L/pulse. The smart meters inbuilt data loggers were programmed to transmit pulse counts at 15 s intervals. Thus, the minimum temporal resolution for recording meter pulses was 15 s.

Although the data were metered with a sub-minute resolution, the data set was considered rudimentary due to the relatively large meter pulse volume (1 L/pulse). All water use events smaller than 1 L would thus be reported as part of a larger event, or as part of a set of smaller events, which exceed 1 L when combined. Similarly, events with durations <15 s would be reported at regular intervals of 15 s (not less). Each measurement was reported in terms of the metered volume ( $\geq 1$  L) and the time stamp ( $\geq 15$  s), to the nearest 15 s. The recorded values were set to be reported each minute and every 15 s afterwards (00:00:00, 00:00:15, 00:00:30, 00:00:45 and so on).

Ilemobade et al. (2018) reported on the complexities of dealing with high-resolution data in the context of a developing country. While the intention with this study was to record only end-use data from the 63 smart meters, in reality, data from various nearby devices (e.g. other household smart meters, security system remotes, and some toys) that were transmitting at the same frequency as the designated smart meters, although unwanted events, were also recorded. The data generation rate ( $\sim 500$  kb/h) led to about 10 000 to 16 000 records being reported per hour. The raw data were filtered and organised into a format appropriate for analysis, using algorithms developed for the particular purpose. After undertaking several iterations of sorting the data, an algorithm was developed to filter and sort the data, as presented by Ilemobade et al. (2018).

After downloading and processing the relevant water use data, the recorded data were sorted chronologically. The final set comprised 63 separate MS Excel files with each file containing the filtered water meter recordings of a single property. Each MS Excel file contained 3 fields, namely the unique identifier (meter number), date-time stamp and the recorded meter reading (L), from which the pulse volume (L) and intensity (L/s) over the said time interval was deduced. Table 6.1 summarises the format of the MS Excel files, which were later used as input files for the extraction process.

Table 6.1. Collected data set format in MS Excel

<b>MS Excel field</b>	<b>Field 1 [Column A]</b>	<b>Field 2 [Column B]</b>	<b>Field 3 [Column C]</b>
<b>Data description</b>	Unique identifier (meter number)	Date-time stamp	Recorded meter readings (L)
<b>Data format</b>	1010-001-xxxx	YYYY/MM/DD hh:mm:ss	325 xxx
<b>Variable assigned</b>	-	t	r

## PROCEDURE FOR EVENT EXTRACTION

A single event was identified by investigating the sequence of measured pulse readings. Event start times ( $d_0$ ) and event end times ( $d_e$ ) were derived by evaluating the time difference between recording intervals. If a gap occurred between readings, in other words, if consecutive pulse readings were recorded at intervals larger than the temporal resolution of the meter (15 s), the start/end time of an event was identified. The time passed between measured events was termed a time gap. Figure 6.1 shows an example of 2 single events, with a time gap of 45 s. The second event is thus preceded by a 45 s-gap.

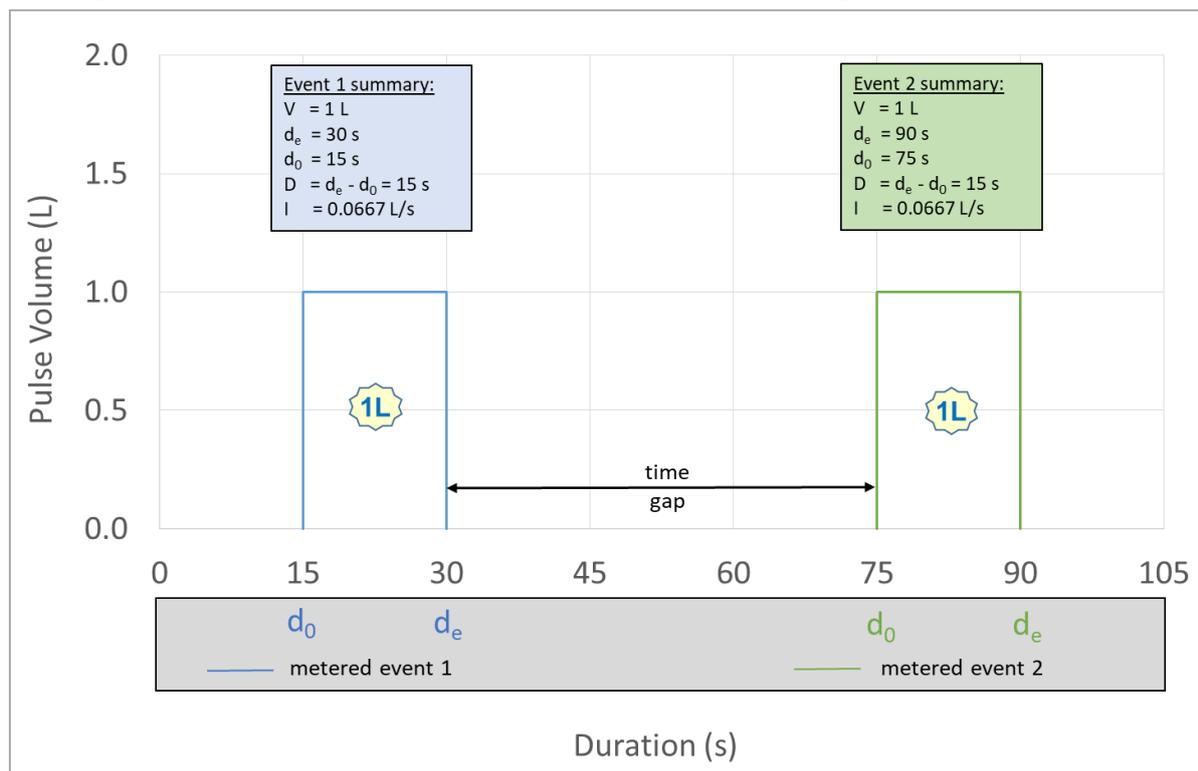


Figure 6.1. Two single events with a 45 s time gap

The event duration ( $D$ ) was calculated by subtracting  $d_0$  from  $d_e$ . The water meter reading difference between two consecutive water meter pulses ( $\Delta r$ ) represented the volume consumed between the two pulses. The difference between the event start water meter reading ( $v_0$ ) and the event end water meter reading ( $v_e$ ), derived from ( $\Delta r$ ), represents the total event volume ( $V$ ). The average intensity ( $I$ ) of an identified event was calculated using the total event volume ( $V$ ) and the event duration ( $D$ ).

## TIME GAP SETTINGS

In some cases, event pulses were lumped despite a delay of  $\pm 30$  s (preceded by a  $\pm 15$  s-gap), or even  $\pm 45$  s (preceded by a  $\pm 30$  s-gap). Inspection of the data set confirmed that some lumped readings formed part of a single end-use event. Inconsistencies with recorded meter readings (e.g. lagged meter reading) and data gaps in water meter readings have been reported on in earlier studies (e.g. Cominola et al. 2018).

Some lagged readings may be superimposed onto the subsequent reading, which is called a lumped reading in this text. Such lumped records would typically be reported once during a relatively long water use event, with a relatively constant flow rate. Figure 6.2 shows a schematic of an interrupted single end-use event (with lumped reading), with the resulting meter spike occurring after a 15 s time gap.

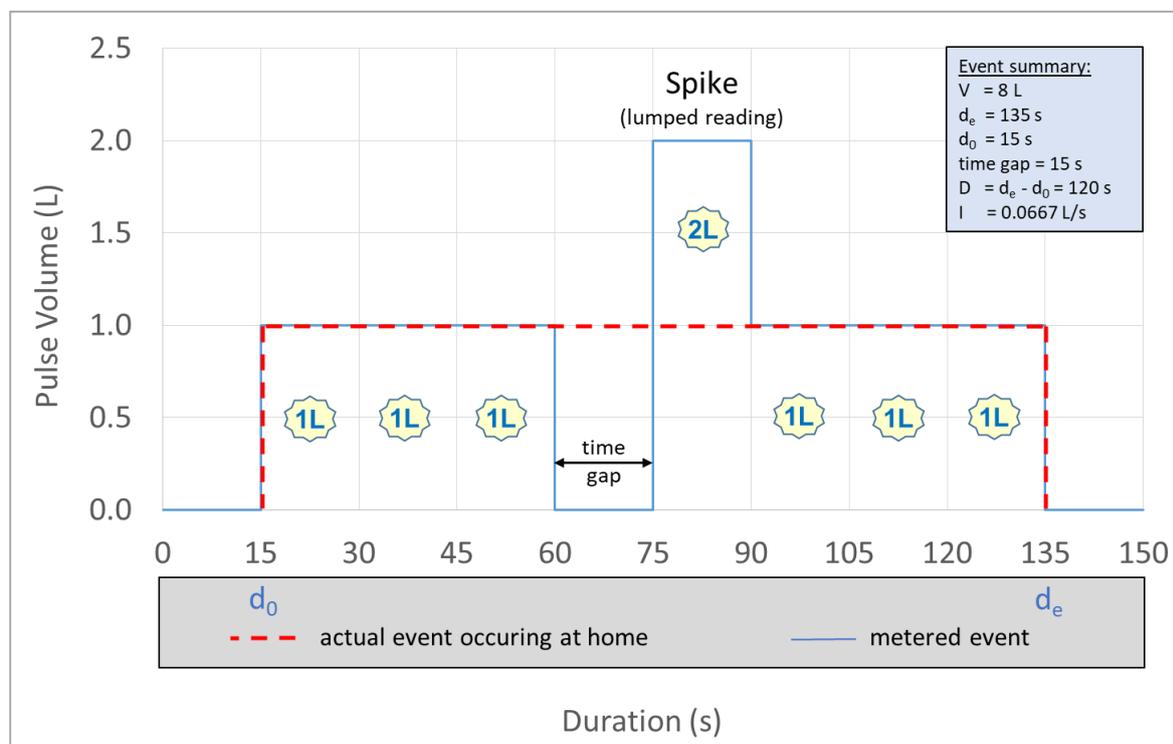


Figure 6.2. Schematic of a meter spike/lagged reading

In order to address this problem, a procedure was developed to incorporate lumped values as part of a single event, instead of incorrectly splitting the readings into two or more separate events. Subsequently, a time gap setting (TGS) was incorporated into the extraction tool to determine a suitable time gap between consecutive events. No earlier research was available on which to base an initial time gap estimate. Consequently, time gaps were chosen based on intervals of 15 s. Only 3 TGS were considered, since a preliminary assessment showed that a TGS > 45 s resulted in excessive lumping of end-use events. Thus, the time gaps assessed between separate events were: 15 s-gap, 30 s-gap and 45 s-gap.

### END-USE EXTRACTION TOOL

A Python End-use Extraction Tool (PEET) was developed as part of this study so that end-use characteristics could be extracted from the recorded water meter data series. PEET's input and output are MS Excel files. The format of PEET's input are summarised in Table 6.1. The TGS also had to be defined, in order to determine which consecutive pulse readings must be lumped together to represent a single event. Figure 6.3 depicts the decision pattern of PEET that extracts end-use events at a single residential property.

The resulting PEET output as an MS Excel file, contains 5 fields (see Table 6.2). Figure 6.3 is the schematic decision pattern for 1 property, thus, the process was repeated for each of the 63 properties in the data set. For each TGS, PEET generated 63 MS Excel files, one file per property. Smart meter serial numbers were used as unique identifiers, in order to link the extracted end-use events to the different homes and the corresponding consumer survey results, which were available for selected homes only. Table 6.3 summarises all variables defined during this study.

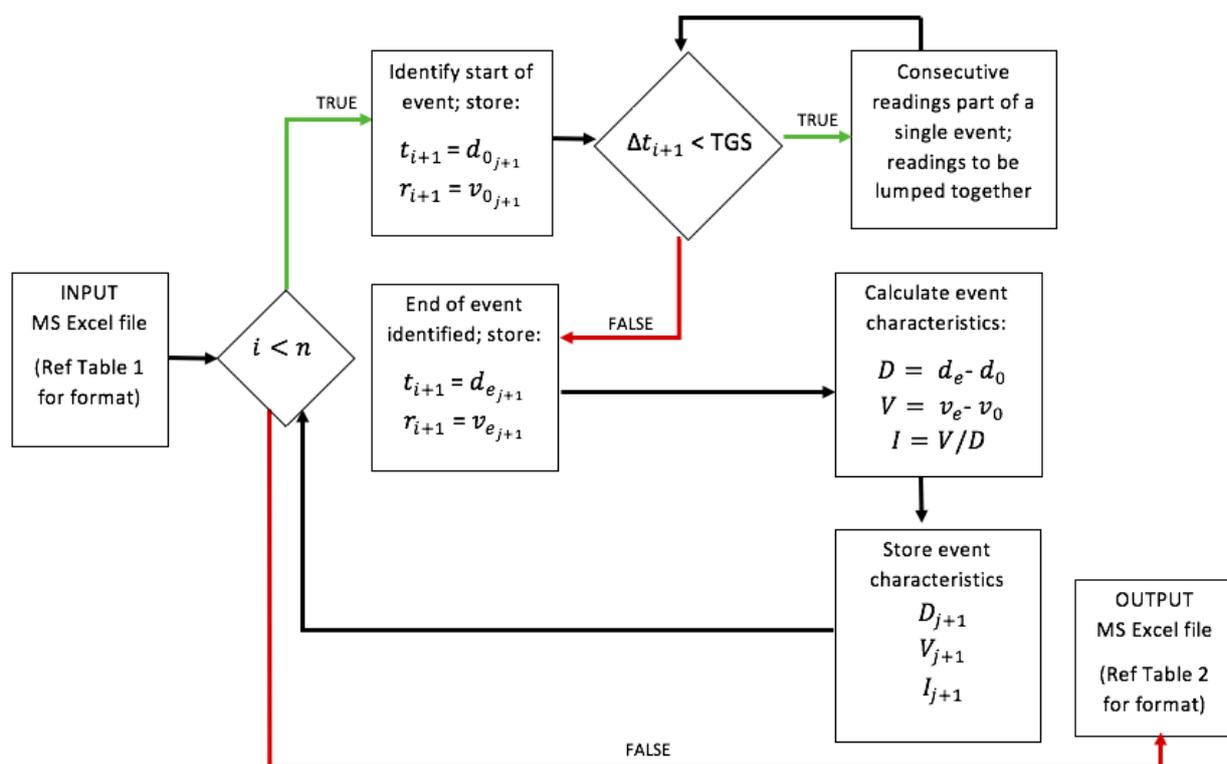


Figure 6.3. Schematic of end-use extraction tool procedure

Table 6.2. MS Excel format of PEET output

MS Excel field	Field 1 [Column A]	Field 2 [Column B]	Field 3 [Column C]	Field 4 [Column D]	Field 5 [Column E]
Data description	Date-time stamp	Event volume	Event duration	Event intensity	Unique identifier
Format/units	YYYY/MM/DD hh:mm:ss	(L)	(s)	(L/s)	1010-001-xxxx
Variable assigned	t	V	D	I	-

Table 6.3. List of variables

Variable	Description
d	Timestamp at start/end of event
D	Extracted event duration (s)
i	Metered pulse count in time series: $i = 0, 1, 2, \dots, n$
I	Extracted event intensity (L/s)
j	Event count identified at a home: $j = 0, 1, 2, \dots, m$
m	Total number of end-uses identified at a home
n	Final pulse reading in MS Excel file
$\Delta r$	Volume difference between two consecutive meter readings
$\Delta t$	Time difference between two consecutive meter readings
t	Meter pulse reading timestamp
Subscript 0	Start of event
Subscript e	End of event
V	Extracted event volume (L)

### CHARACTERISATION AND IDENTIFICATION OF MINOR EVENTS

Two types of events were categorised during this study, namely minor events and major events. Figure 6.4 represents a schematic of 4 low flow events ( $I < 0.035$  L/s) occurring at a home (say a tap being opened and closed).

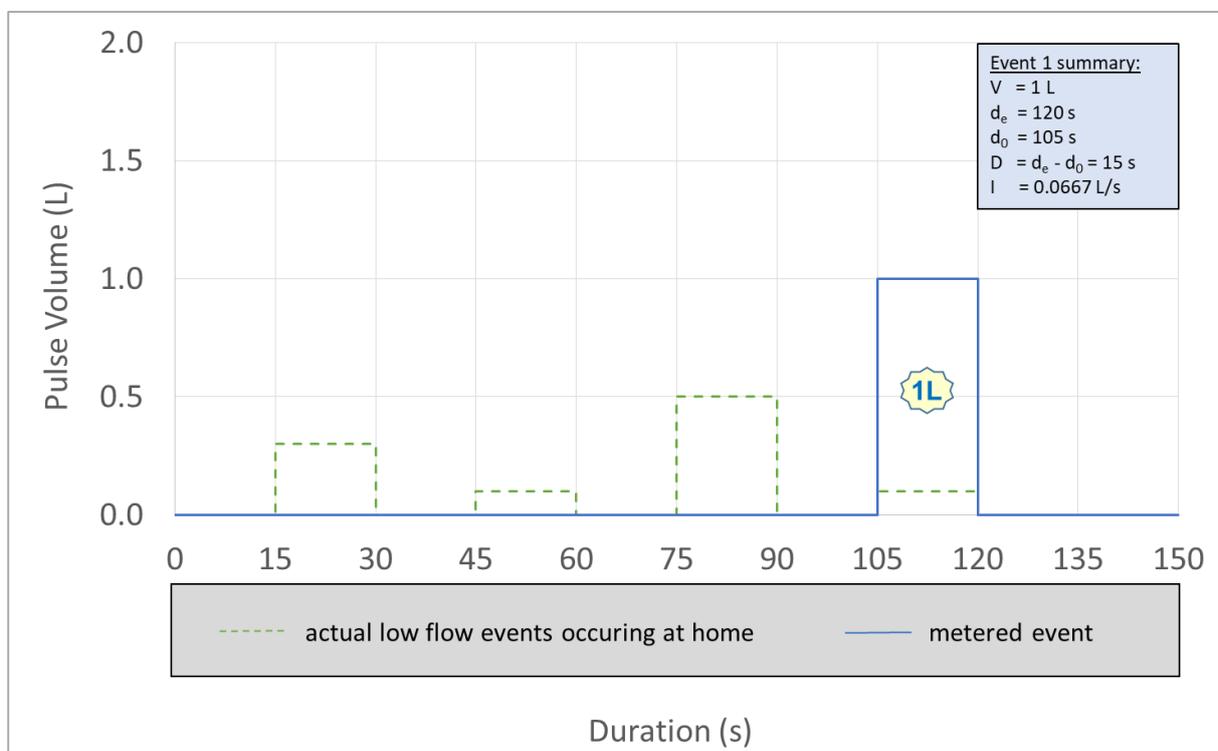


Figure 6.4. Schematic of low flow events

Due to the limiting 1 L pulse volume, a single event of 1 L in volume was recorded at 105 s on the time series, with a total duration of 15 s. Multiple low flow events would be reported by the water meter as a single event, at a later time. The recorded event is thus not a true representation of the actual events occurring at the home. Consequently, all events with a 1 L pulse volume and a 15 s duration (preceded and followed by a delay larger than the time gap), were categorised as minor events and grouped together.

Screening for realistic low flow events involved assumptions regarding the minimum flow rate of a valid end-use event. Since the flow rate resolution of the meters were 0.067 L/s (1 L pulse over a 15 s recording period), all events with intensities  $\leq 0.067$  L/s were investigated. Consider an end-use with a constant flow rate of 0.04 L/s being active for a certain period of time (for example 75 s) – until (say) the consumer closes the running tap. During the active period, the event would produce one pulse (of 1 L) intermitted at 15 s intervals. Figure 6.5 represents a schematic of this example.

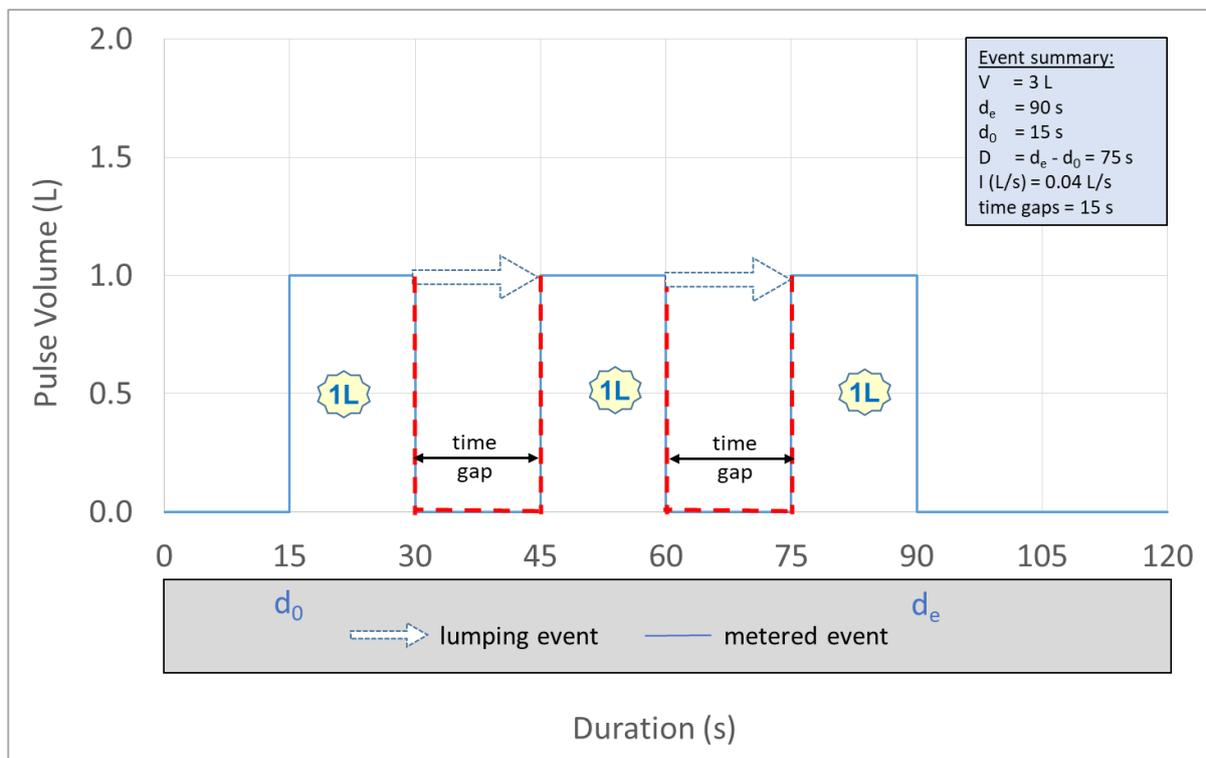


Figure 6.5. Schematic example of lumping multiple events

Due to the TGS incorporated in the extraction process, the three recorded pulses would be lumped together as one single event, with an event volume of 3 L over a duration of 75 s. However, a genuine water use event of 1 L, used in 15 s intervals, would also report one pulse (1 L) at 15 s intervals. The time series of a genuine 1 L event reported over 15 s and that of numerous small events that were reported by the measurement system as 1 L over 15 s, would appear identical. Due to the rudimentary nature of the time series data, the two instances could not be distinguished.

Consequently, all extracted events with intensities  $< 0.067$  L/s were categorised as minor events, and grouped together. Earlier work by Otaki et al. (2011) used water meters that were able to measure intensities of 0.0167 L/s (1 L/min), but the same authors also note that no in-house activity in the Thailand study area required such a low flow rate.

## READING AND ROUNDING ERRORS

Some extracted events were filtered out of the data set, including all zero values and negative values, which were considered to be reading errors, or rounding errors. Relatively high values could be explained as being either a valid event – possibly spread over a relatively long duration, or a meter reading error. Consequently, a meter verification exercise was conducted to evaluate typical maximum flow rates. The highest flow rate recorded at a single end-use in this study was  $\sim 0.4$  L/s, but a total flow rate at the consumer meter of  $\sim 0.5$  L/s was recorded at a home of one of the authors with various taps open simultaneously. Flow rates of  $> 0.5$  L/s were reportedly uncommon in Australia, with manual sprinkler systems reporting the highest flow rate of  $\sim 0.4$  L/s in one study (Roberts, 2005). An upper limit of 1.0 L/s was considered appropriate for the study sample and all readings where the intensity exceeded 1.0 L/s for  $\geq 15$  s were filtered out. In other words, events with a total volume difference  $> 15$  L in a single recording interval of 15 s, were considered to be errors. A summary of all filtered values is presented in Table 6.4.

## RESULTS AND DISCUSSION

### Time gap setting

PEET was employed to extract single end-use events from rudimentary data, considering a time gap of 15 s, 30 s, and 45 s, before and after a recorded pulse. A comparison of the events extracted, for all three TGS, are tabulated in Table 6.4. Due to the rudimentary nature of the data, minor events were grouped together. All events not categorised as minor events, were considered major events.

As was expected, the 15 s-gap setting extracted the most end-use events from the raw data set, with the 45 s-gap reporting the lowest numbers. For all three TGS, the total volume of major events comprised  $>74\%$  of the total volume of all extracted events. Only major events were considered for further analysis. This method was considered acceptable due to the large percentage of total volume representing major events, as well as the uncertainty surrounding minor events.

Table 6.4. Comparison between evaluated time gaps settings

Description		All extracted events	Reading / rounding errors	Minor events	Major events
TGS = 15 s	# Extracted events	1 288 373	5 377	971 032	311 064
	# Extracted events (%)	100.00	0.42	75.44	24.14
	Total Volume (L)	4 429 578	78	971 950	3 457 550
	Total Volume (%)	100.00	0.00	21.92	78.06
TGS = 30 s	# Extracted events	1 107 547	5 238	890 249	212 060
	# Extracted events (%)	100.00	0.47	80.38	19.15
	Total Volume (L)	4 429 578	86	1 072 735	3 356 757
	Total Volume (%)	100.00	0.00	24.22	75.78
TGS = 45 s	# Extracted events	1 022 290	5 177	827 501	189 612
	# Extracted events (%)	100.00	0.51	80.95	18.55
	Total Volume (L)	4 596 464	87	1 152 296	3 444 081
	Total Volume (%)	100.00	0.00	25.07	74.93

### Characteristics of events

The cumulative distribution functions (CDFs) of event characteristics for the three TGS were compiled and are presented in Figure 6.6 (event volume), Figure 6.7 (event duration) and Figure 6.8 (event intensity).

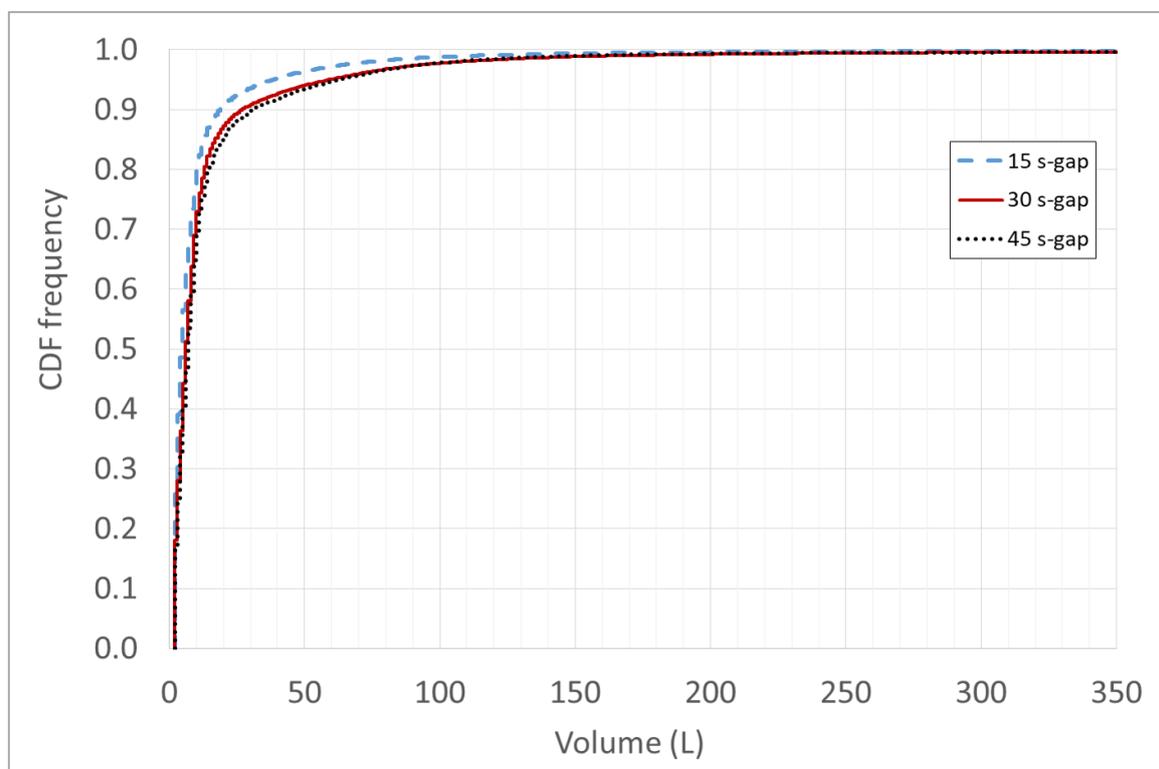


Figure 6.6. End-use volume for the three different time gap settings

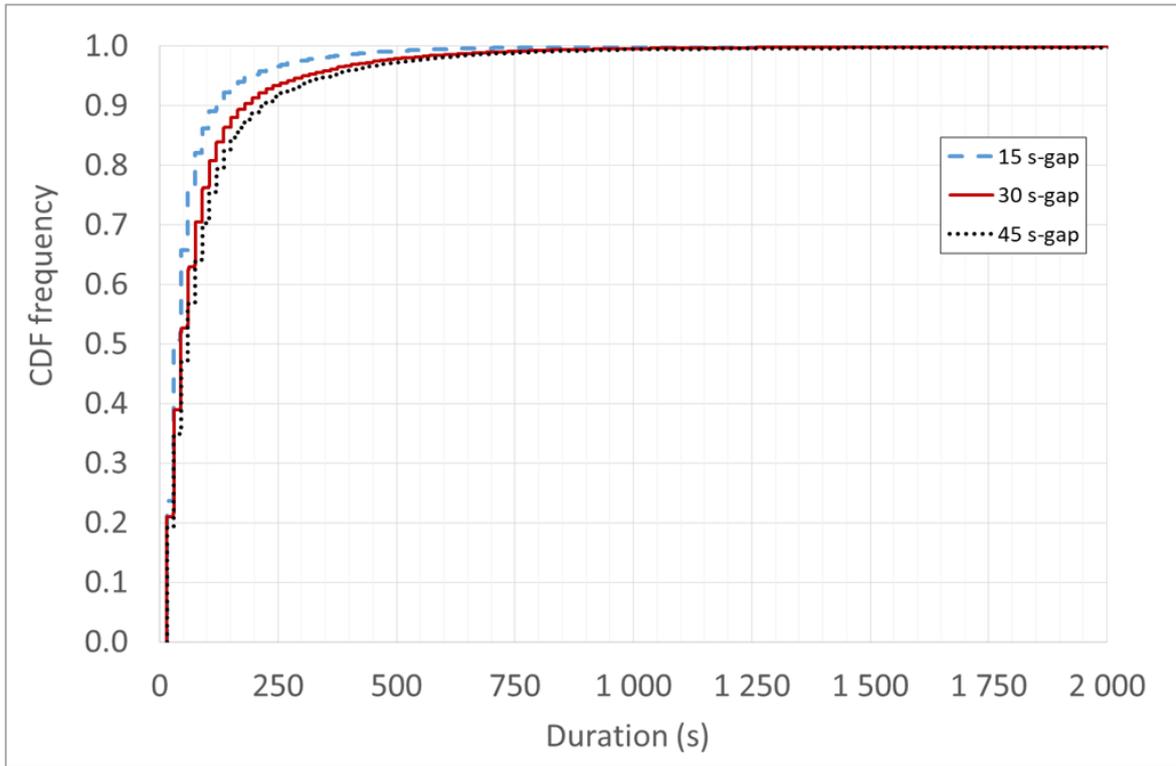


Figure 6.7. End-use duration for the three different time gap settings

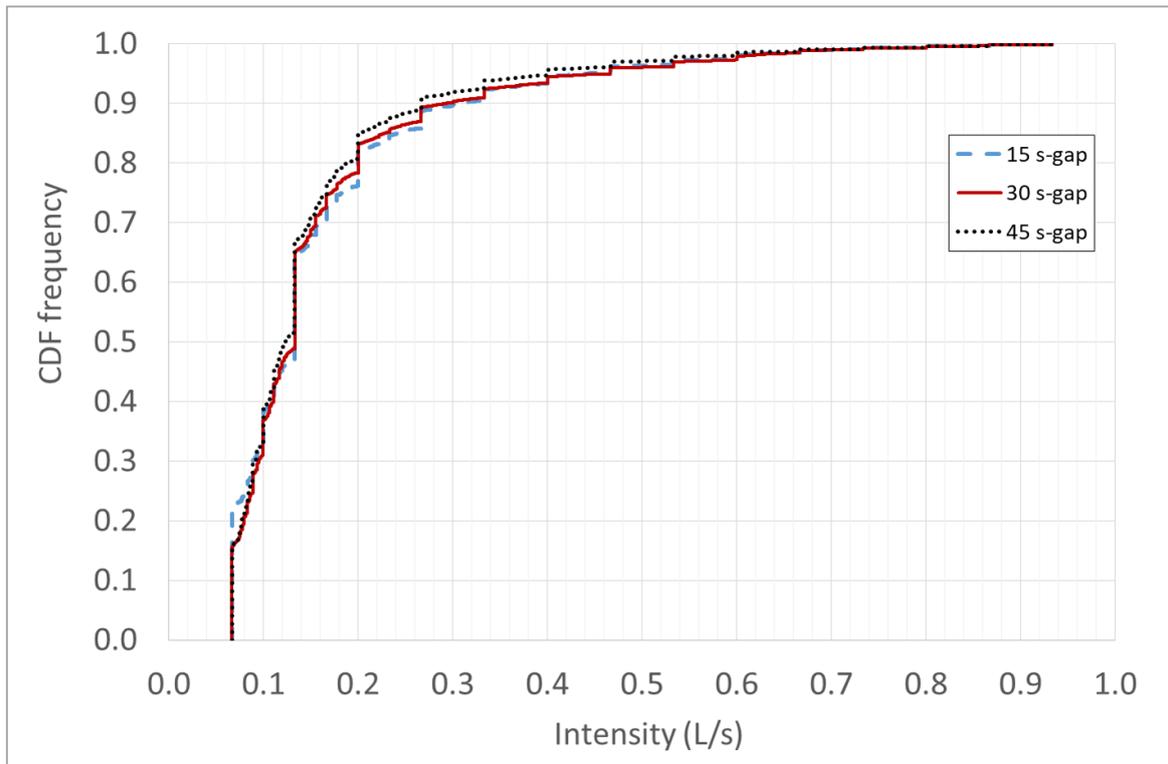


Figure 6.8. End-use intensity for the three different time gap settings

Above a given threshold on Figures 6.6 and 6.7, the shortest TGS resulted in the lowest event volumes and shortest event durations, as could be expected. In contrast, the shortest TGS produced the highest intensities. The CDF also showed that the lowest 70% of event volume- and duration-values were almost identical for all TGS. The lowest 70% event volume values were less sensitive to the TGS, compared to the upper 30%. Half of all extracted events had durations of less than 60 s, and event volumes of less than 7 L. This was true for the three different TGS in PEET. The median and most frequent intensity was 0.14 L/s.

Earlier studies (Buchberger and Wells 1996) considered event volume <210 L to be a reasonable limit for classifying indoor events. Approximately 99% of the extracted end-use events for the three TGS had volumes <210 L. Based on the assumed limits, 99% of the events at the 63 homes would thus be considered indoor events – which was unlikely when compared to the survey responses regarding frequency of outdoor irrigation. Simply apportioning end-use events based on arbitrary values is thus not sufficient, and future research should develop a robust method to classify end-uses.

With reference to Figure 6.6, the largest volume for a single event was 2.6 kL, 3.6 kL and 4.7 kL for the 15 s-gap, 30 s-gap, and 45 s-gap settings respectively. The 45 s TGS reported the longest event of 57 660 s (almost 16 h in duration), while the longest event for the 30 s TGS was 39 136 s. The relatively long durations for the 45 s-gap setting were considered excessive, suggesting that the 45 s-gap setting may be invalid – in the sense that separate events were combined. The 30 s-gap setting showed the most reasonable values for household end-uses when dealing with rudimentary data, and was consequently selected as the optimal TGS for this study.

### **Final data set**

Using the 30 s-gap setting, a total of 1 107 547 events were extracted from 63 homes over the 217 days, prior to cleaning and filtering the data set. After filtering, the final data set comprised 212 060 single end-use events. About 24% of the total volume of all events was attributed to minor events, representing 80% of the number of events extracted.

The average number of events per home per day was 16 for 1 PPH, 18 for 2 PPH and 28 for 4 PPH. The number of notable end-uses equates to 9 events per person per day, on average over the study period and for all homes. This value was considered realistic, considering that all minor events were filtered out and many homes had a low occupancy of one or two persons.

## CONCLUSION

Household water end-use event characteristics were extracted from rudimentary data – in this study the resolution was 1 L per water meter pulse at a recording interval of 15 s. Various assumptions were employed in the process and three time-gap settings were investigated in attempt to eliminate data lumping problems in the raw data. PEET, a Python End-use Extraction Tool, was developed as part of this study in order to automate the process. PEET was able to extract three water use characteristics, namely event duration, event volume and event flow intensity, from a rudimentary data set. One of the limitations encountered when dealing with rudimentary data is the fact that minor events had to be grouped together and could not be further analysed. Nonetheless, major end-use events were extracted, and valuable information was deduced from the results. Unfortunately, it was impossible to separately classify background leakage flows in the plumbing system, minor leaks at the point-of-use (e.g. a dripping tap) and relatively low flows from valid water use events (e.g. filling a 200 mL glass with water), so all had to be categorised as minor events. The extracted characteristics of major events could in future be used to classify end-uses as being either indoor events or outdoor events. Such a classification would benefit service providers in setting water demand strategies when faced with rudimentary data.

## ACKNOWLEDGEMENTS

The authors would like to thank Johannesburg Water for funding this project, contributing in part to the PhD bursary of Mrs Meyer.

## REFERENCES

- Alcocer-Yamanaka, V., Tzatchkov, V.I., Arreguin-Cortes, F. 2012 Modelling of drinking water distribution networks using stochastic demand.” *Water Resour. Manage. EWRA*. 26(7): 1779-1792. doi: 10.1007/s11269-012-9979-2
- Arregui F. 2015 New software tool for water end-uses studies. Presented at the 8th IWA International Conference on Water Efficiency and Performance Assessment of Water Services, Cincinnati, OH, USA, 20–24 April 2015.
- Beal C., Stewart R.A., Huang T. and Rey E. 2011 South East Queensland residential end use study. *J. Australian Water Association*. **38** (1): 80-84.
- Beal C. and Stewart R.A. 2013 Identifying residential water end uses underpinning peak day and peak hour demand. *J. Water Resour. Plann. Manage., ASCE*, **136**: 19-26. doi:10.1061/(ASCE)WR.1943-5452.0000357

Buchberger, S.G. and Wells, G.J. 1996 Intensity, duration, and frequency of residential water demands. *J. Water Resour. Plann. Manage.*, **122**(1): 11-19.  
doi: 10.1061/(ASCE)0733-9496(1996)122:1(11)

Buchberger S.G. and Wu L. 1995 A model for instantaneous residential water demands. *J. Hydr. Engrg., ASCE*, **121**(3): 232-246. doi: 10.1061/(ASCE)0733-9429(1995)121:3(232)

Cardell-Oliver R., Wang J. and Gigney H. 2016 Smart meter analytics to pinpoint opportunities for reducing household water use. *J. Water Resour. Plann. Manage., ASCE*, **142**(6).  
doi: 10.1061/(ASCE)WR.1943-5452.0000634, 04016007

Cominola A., Giuliani M., Castelletti A., Rosenberg D.E., Abdallah A.M. 2018 Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management. *Environmental, Modelling and Software* **102**(2018): 199-212.  
doi: 10.1016/j.envsoft.2017.11.022

DeOreo W.B., Heaney J.P. and Mayer P.W. 1996 Flow trace analysis to assess water use. *Am. Water Works Assoc.*1996, **88**, 79–90. doi: 10.1002/j.1551-8833.1996.tb06487.x

DeOreo W.B, Mayer P.W, Martien L., Hayden M., Funk A., Kramer-Duffield M. and Davis R. 2011. California single family water demand efficiency study. A report compiled for the California Department of Water Resources, 20 July 2011.

Du Plessis J.L. and Jacobs H.E. 2018 Analysis of water use by gated communities in South Africa, *Water SA* **44**(1), 130-135. doi: 10.4314/wsa.v44i1.15

Jorgensen, B.S., Martin, J.F., Pearce, M.W., Willis, E.M. 2013 Predicting household water consumption with individual-level variables. *Env. Behaviour*, **46**(7), 872-897.  
<https://doi.org/10.1177/0013916513482462>

Ilemobade A., Jacobs H.E. and Botha B. E. 2018 Sorting high resolution end-use data – A pleasurable headache. Proceedings. 1st International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada – July 23-25.

Kowalski M. and Marshallsay D. 2003 A System for Improved Assessment of Domestic Water Use Components. II International Conference Efficient Use and Management of Urban Water Supply; International Water Association: Tenerife, Spain, 2003.

Kowalski M. and Marshallsay D. 2005 Using measured microcomponent data to model the impact of water conservation strategies on the diurnal consumption profile. *Water Science & Technology Water Supply* 5(3):145-150. doi: 10.2166/ws.2005.0094

Loh M. and Coghlan P. 2003 Domestic water use study. Water Corporation, Perth.

Mayer P.W., DeOreo W.B., Opitz E.M., Kiefer J.C., Davis W.Y., Dziegielewsji B. and Nelson J.O. 1999 Residential end uses of water. AWWA Research Foundation and American Water Works Association, Denver, CO, USA.

[http://www.waterrf.org/PublicReportLibrary/RFR90781\\_1999\\_241A.pdf](http://www.waterrf.org/PublicReportLibrary/RFR90781_1999_241A.pdf)

Nguyen K. A., Stewart R.A., Zhang H. and Sahin O. 2018 An adaptive model for the autonomous monitoring and management of water end use. *Smart water Journal* 3(5).

doi: 10.1186/s40713-018-0012-7.

Nguyen, K.A., Zhang, H., Stewart, A.R. 2013 Development of an intelligent model to categorise residential water end use events. *J. Hydro-environ. Res.* 7(2013), 182-201.

doi: 10.1016/j.jher.2013.02.004

Otaki, Y., Otaki, M., Sugihara, H., Mathurasa, L., Pengchai, P. & Aramaki, T. 2011 Comparison of residential indoor water consumption patterns in Chiang mai and Khon Kaen, Thailand. *Journal AWWA* 103: 104–110. doi: 10.1002/j.1551-8833.2011.tb11457.x

Pastor-Jaboloyes L., Arregui F.J. and Cobacho R. 2018 Water end use disaggregation based on soft computing techniques. *Water* 2018, 10, 46. doi: 10.3390/w10010046

Pretorius, D., Crouch, M.L., Jacobs, H.E. 2019 Diurnal water use patterns for low-cost houses with indigent water allocation: a South African case study. *J. Water Sanitation Hyg. for Develop.* 9(3): 513-521. doi: 10.2166/washdev.2019.165

Roberts P. 2005 Yarra Valley Water: 2004, Residential end use measurement study. Melbourne, Yarra Valley Water, 1-77, June 2005.

Sensus 2018 iPERL™ (International) Water Meters. Available from: <https://sensus.com/products/iperl-international/> [Accessed 07 June 2018].

Stewart R.A., Willis R. and Capati B. 2009 Closing the loop on water planning: an integrated smart metering and web-based knowledge management system approach. 10<sup>th</sup> IWA International

## Chapter 7.

### Household water end-use classification model for implementation on coarser data: toward improving the benefits of lower resolution end-use data sets

Bettina E. Meyer<sup>1\*</sup>, Khoi Nguyen<sup>2</sup>, Cara D. Beal<sup>2</sup>, Heinz E. Jacobs<sup>1</sup> and Steven Buchberger<sup>3</sup>

1. Department of Civil Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa, [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za), [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za)
2. Cities Research Institute and School of Engineering and Built Environment, Griffith University, Australia, [c.beal@griffith.edu.au](mailto:c.beal@griffith.edu.au), [k.nguyen@griffith.edu.au](mailto:k.nguyen@griffith.edu.au)
3. Civil Engineering Program, University of Cincinnati, Cincinnati, Ohio, USA 45221-0071, [buchbesg@ucmail.uc.edu](mailto:buchbesg@ucmail.uc.edu)

\*Corresponding author: [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za)

Submitted for review and possible publication to the Journal Water Resources Planning and Management. Paper reference number WRENG-4842.

#### ABSTRACT

Previously, flow trace analysis on high resolution smart meter data sets have been described to identify individual end-use events. This research presents a method to classify relatively low resolution household water use data that is more commonly available to water utilities, into indoor and outdoor consumption. The relationships between the most notable characteristics of end-use events, namely event duration, volume, and intensity (flow rate), were investigated in order to categorize the water use as being indoor or outdoor. Three classification models were developed, calibrated and compared using over 200,000 household end-use events, recorded independently in Australia and South Africa. The three methods were also compared to an arbitrary classification scheme currently being implemented. The classification model recommended in this paper correctly classifies between 60.7% and 96.2% of end-use events, thus reinforcing the value of low resolution data as a source of useful information for water demand management. It is hoped that by applying this method on coarser data sets, water utilities from a range of socio-economic settings can have greater opportunities to improve water security through better informed demand management programs.

**Keywords:** demand management, end-use classification, household water use, hydro-informatics, modelling, smart meters

## INTRODUCTION

### End-use disaggregation and classification methods

Previous studies, measuring indoor and outdoor end-uses as distinct components of total water consumption, demonstrated the benefit of having consumption data at household scale (Makwiza and Jacobs 2017, Beal et al. 2011, Makki et al. 2011). Resource intensive mathematical models, developed using machine learning algorithms, could be used to disaggregate water use events. Recent studies regarding notable disaggregation methods include Pastor-Jabaloyes et al. (2018), *Autoflow* (Nguyen et al. 2018), *SmartH2O* (Cominola et al. 2018), *REU2016* (Vitter and Webber 2018), and *BuntBrain-ForEndUses* (Arregui 2015). These disaggregation models require high resolution data for model application. A summary of completed water end-use studies, presented by Beal and Stewart (2011), demonstrate that a sub-10 seconds (s) metering resolution for data capturing, with pulse measurements of less than 0.026 gallons/pulse, is considered high resolution data. Water meters with such high recording resolutions are uncommon. Employing high resolution smart meters over a large spatial scale is not (yet) viable, due to the resource intensive and costly nature of such projects (Ilemobade et al. 2018, Nguyen et al. 2013).

Utilities around the world are investing in advance smart metering systems with automatic meter reading (AMR), producing data that are not meant for end-use disaggregation, e.g. application of flow trace analysis tools such as *Autoflow* (Nguyen et al. 2013). The AMR meters typically measure water consumption at 15 seconds (s) intervals with 0.264 gallons/pulse (Meyer et al. 2020), 15 minutes (min) intervals with 0.264 gallons/pulse (Pretorius et al. 2019), or hourly intervals with 1.321 gallons/pulse (Cole and Stewart 2013). Although lower resolution data provide insights into anomalous events (especially leakage), this coarser data resolution prevents the identification of individual household end-uses (Cominola et al. 2018).

Alternatively, other studies used sensors at the point of use in order to identify event location, for example shower events (Botha et al. 2017) and garden irrigation (Meyer and Jacobs 2019). End-use sensing techniques are impractical for application on a large scale in the same way as high-resolution flow trace analyses. The question arises whether it would be possible to classify water use events as being either indoor or outdoor, given end-use measurements with a resolution too low for application on current disaggregation methods (e.g. measurement frequency intervals longer than 10 s and pulse measurements coarser than 0.026 gallons/pulse).

The most basic method to distinguish indoor events from outdoor events in a time series would be to establish limits for selected demand characteristics. During an extensive study of residential water use in Milford, a small town in North America (Buchberger et al. 2003), water end-use events were divided into indoor and outdoor consumption on the basis of fixed but arbitrary limits. In this regard, maximum values for duration and volume were used as upper bounds for indoor uses. Based on an examination of seasonal water use measurements, household water demands with durations exceeding 30 min or volumes exceeding 80 gallons (gal) were categorised as outdoor use. These arbitrary limits were assumed based on water demands over the winter months of 1997 (during which it was assumed water use was indoor only). The method can thus be considered regionally limited, and may not transfer to other regions. Although Buchberger et al. (2003) used this upper bound limit (UBL) method on high resolution data (1 s interval measurements), the method can also be applied to coarser data sets, once an end-use extraction tool, such as PEET (Meyer et al. 2020), is implemented on the time series. The accuracy of the values selected for the UBL outside of Milford has never been verified. Additionally, the classification method presented by Buchberger et al. (2003) only considers two input parameters, namely water pulse duration and volume. Testing this method on a supervised data set will give insight into the accuracy thereof on regions outside of Milford, and possibly show that a more complex method is needed to attain accurate classifications with coarse data sets.

### **Household water end-uses**

Understanding water consumption at a household level is vital for developing effective demand management strategies (Jorgensen et al. 2013). Household end-uses could be classified as being indoor or outdoor, based on the physical location of the water use event in and around the home. Some examples of typical indoor end-uses are the toilet, shower, bath, clothes washing machine, dishwasher and indoor tap (Nguyen et al. 2018, Scheepers and Jacobs 2014, Blokker et al. 2010). Typical outdoor uses include garden irrigation (Makwiza and Jacobs 2017, Survis and Root 2012), outdoor washing (Beal et al. 2018), water use for swimming pools and ponds (Fisher-Jeffes et al. 2015, DeOreo and Mayer 2012) and pet care (Beal et al. 2018). Irrigation can be a substantial proportion of outdoor use when a garden, lawn or large outdoor area, is present, particularly in dry conditions (Beal and Stewart 2013).

Consumer behaviour in terms of indoor water use and outdoor water use differs notably, especially in regions where garden irrigation is prevalent. Indoor water consumption is primarily influenced by the number of people in the household and fixture efficiency, whereas outdoor use involves additional parameters, such as effective rainfall, evaporation, humidity, temperature, plant species, soil moisture content and irrigation system design, - maintenance and - operation (Glenn et al. 2015). Several factors which influence indoor water demand include water pressure (Meyer et al. 2018, Inman and Jeffry 2006), socio-demographics (Willis et al. 2013), environmental and scarcity of supply conditions (Beal et al. 2018) and the presence of water efficient appliances (Athuraliya et al. 2008).

Recently, serious water restrictions in the City of Cape Town specifically targeted outdoor water use under the “Day Zero” water restrictions (Nel and Jacobs 2019), recognizing the fact that the most behaviourally-driven outdoor water use activities hold more promise for saving water than indoor use. The classification of household end-uses, as being either indoor or outdoor, can empower regulators and aid planners to overcome various challenges regarding water saving interventions and programs. With new water wise products being implemented in homes, water use in the residential setting has decreased over the years creating new challenges for water utilities (DeOreo et al. 2016). Understanding residential end-use water consumption behaviour could help establish new water end-use benchmarks. Notable outdoor events such as garden irrigation and pool water use are typically targeted separately when water restrictions are applied (Nel and Jacobs 2019, Survis and Root 2012, Jacobs et al. 2007). Note that leaks, while not strictly an “end-use”, are both indoor and outdoor events and can usually be detected with intermediate sampling resolution data (Britton et al. 2013).

### Characteristics of end-uses

A household water end-use can be represented by a rectangular pulse on the recorded time series (Alcocer-Yamanaka et al. 2012, Buchberger and Wu 1995). The rectangular pulse includes three event characteristics, namely event duration, event volume, and event intensity (flow rate). When a rectangular pulse event is assumed, the event intensity is found as the ratio of event volume to event duration, so that  $I=V/D$ . When a rectangular pulse is not assumed, intensity fluctuates during an event. Intelligent end-use disaggregation tools, such as the software *Autoflow*, determine an event intensity based on various parameters obtained from the time series (Nguyen et al. 2018). During the disaggregation process, event intensities are determined based on maximum intensities for shorter events (such as a toilet flushing), or most frequent intensity for longer events (a shower for instance).

At a single residential property, the intensity of different end-use events will vary significantly (DeOreo 2011, Buchberger and Wells 1996), and the relationship between duration, volume and intensity gives insight into the type of end-use. Beal and Stewart (2011) categorized end-uses into three clusters of intensity ranges. End-uses associated with intensities less than 0.44 gallons per minute (gpm) were considered to be mainly leaks and low flow indoor events. Intensities ranging between 0.44 gpm and 4.41 gpm were associated with both indoor and outdoor use. Consequently, intensity cannot be used independently to distinguish indoor use from outdoor use. Irrigation, some high flow indoor uses, and service break leaks were associated with flow rates ranging between 4.41 gpm and 7.93 gpm (Beal and Stewart 2011).

End-uses with relatively constant volumes, such as the toilet, washing machine and dishwasher, were termed deterministic end-uses by Buchberger and Wells (1996). Bath use can also be considered deterministic (Blokker et al. 2010). In the case of a bath, the volume depends on the physical dimensions of the bath tub, the bath water level and the size of the submerged body, suggesting that volume is relatively constant from one use to the next for a

particular consumer. The volume values for deterministic end-uses depend on the type of fixture and the setting used. Water pressure could affect event duration and intensity, but event volume varies little with changed pressure from one deterministic use to the next.

Washing machines and dishwashers use water in cycles and each load may have between 2 and 7 water use cycles (Botha et al. 2018, Makki et al. 2015, Nguyen et al. 2013). Each water use cycle would show up as a single end-use pulse. The complete event cycle could be identified by intelligent flow trace analyses software, so that the total volume per wash cycle could be expressed as one water end-use event (Nguyen et al. 2013).

All non-deterministic end-uses, such as the shower, taps and garden irrigation, are termed random or discretionary end-uses (Buchberger and Wells 1996). The rectangular pulses of discretionary end-use events vary notably and are largely determined by consumer behaviour. The duration of random water use events is highly variable and cannot be solely relied on to classify end-use events as being indoor or outdoor. Creaco et al. (2015) showed that incorporating the correlation between the different pulse characteristics (such as the duration and intensity) could improve the simulation of water demands at the household level. Consequently, the relationships between different event characteristics need to be explored in order to classify the events as being indoor or outdoor.

## **Aim**

The research aim was to develop and validate a mathematical model for binary classification of lower resolution end-use data. Specifically, the key objective was to develop a model that could categorize water use events as being either indoor or outdoor, based on three input parameters, namely event duration, volume, and intensity. The model was developed with the purpose of being a useful tool for a much broader number of utilities e.g. ones that only had access to coarser end-use data sets. This would enable water utilities from a range of socio-economic settings to broadly classify household end-use events without relying on pre-trained models.

## **Scope and limitations**

The scope involved household water end-use data from two different sources, recorded independently at specific locations in two countries, in two different continents – Australia (Gold Coast) and South Africa (Cape Town). The data from the different sources were collected at different resolutions. Outdoor water use was prevalent in both sample sets. Post-processing of the collected data, in the form of *Autoflow* (Australian data set) and temperature variation analysis (South African data set), was required in order to extract end-use characteristics from a time series of meter readings.

The classification model made no provision for independent variables describing the region per se, such as climatological- or socio-economical inputs. The model is limited to domestic (residential) water use and focusses on three end-use event identifying characteristics, namely event duration, volume, and intensity. Lastly, some small outdoor water event (e.g., opening a tap to fill a dog bowl) might be classified as an indoor event and vice versa for large indoor use (e.g., a prolonged shower).

## **CLASSIFICATION MODEL DEVELOPMENT**

### **Approach**

The aim was to develop a classification model (or decision boundary) to solve a binary (2-class) classification problem. The objective of the model was to correctly classify indoor events and outdoor events from coarser end-use data, while minimizing the classification error. The model categorizes end-use events as being either indoor or outdoor, with the three predictors being the three event characteristics, namely event duration ( $D$ ), event volume ( $V$ ) and event intensity ( $I$ ). Three models were developed, calibrated and compared. Model performance was evaluated based on the Receiver Operating Curve (ROC) and the Area Under the receiver operating Curve (AUC). The ROC graphically illustrates the diagnostic ability and performance of the binary classification model, by plotting the true positive rate (recall) against the false positive rate ( $1 - \text{specificity}$ ). The AUC calculates the area under the ROC. The AUC value ranges from 0 to 1, with 1 being a perfect model. The best fit model was termed the Water End-use Apportionment Model (WEAM).

Two models were developed using supervised machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest (RF). The third model assumed the best decision surface to be an ellipse, and optimized the decision surface by minimizing the squared error. The data set was split into training, testing and validation subsets. In order to account for the class imbalanced data set (99.63% indoor vs 0.37% outdoor), pre-processing of the data was required prior to model development. Class reweight, undersampling and oversampling were evaluated for pre-processing. Although WEAM was developed using high resolution data, the application of the model had to be suitable for use on coarser data sets. WEAM was thus considered suitable to be employed on data sets with resolutions too low for disaggregation tools such as flow trace analysis and *Autoflow*.

### **Data sample**

Two data sets were combined for model development and calibration. The first data set was a combination of end-uses recorded in South Africa as part of earlier studies, spanning over three years (2016-2018), where the event characteristics of showers, washing machines, and garden irrigation were measured using temperature loggers (Meyer and Jacobs 2019, Botha et al. 2017), vibration sensors (Sterne 2019) and direct metering methods (Botha et al. 2018).

Meyer and Jacobs (2019) and Botha et al. (2017) used temperature variation analysis to identify end-use events from a time series. The method is based on the difference between the recorded pipe wall temperature and the ambient temperature, to ultimately identify end-use events and quantify each event's duration. The underlying assumption is that water temperature in the pipe varies notably from the baseline (e.g., ambient) temperature, so the method is ideal for hot water end-uses, such as a shower. The South African data set included 1,631 measured indoor events and 70 measured outdoor events collected at 12 single residential properties and 2 residential flats, over a total recording period of 77 days. The data sets included selected end-uses (as reported on separately in each study cited here), so the total number of events listed above is not representative of all the end-use events at all homes over all days. The  $D$ ,  $V$  and  $I$  of each event were known.

The second data set was collected from 252 homes located in South East Queensland, Australia, in 2010-2012 and used in the Southeast Queensland Residential End-use Study (Beal and Stewart 2011). A mixed method approach, employing smart water meter, data logger, stock survey and water audit, was utilised to obtain high resolution data of 272 pulses/gal representing a pulse every 0.0036 gal at five second intervals. The unprocessed flow trace series was subsequently segregated into single classified end-use events using the flow trace software Trace Wizard™ (Aquacraft 2010). The Australian data set contained detailed information with  $D$ ,  $V$  and  $I$  of 199,586 indoor and 679 outdoor events and played a vital role in conceptualising and verifying the model developed in this study.

The South African and Australian data sets were combined for model development and comprised of 201,966 end-use events, of which 201,217 were indoor (99.63%) and 749 were outdoor (0.37%). The combined data, termed the WEAM data set, were randomly split into three subsets, in order to train, test and validate the model. The percentage of the WEAM data sample apportioned to the training-, testing-, and validating data subset, was 64%, 16%, and 20% respectively. The ratio of indoor events to outdoor events were preserved during the split process. The training data set was used to develop and calibrate (tune) the model. The final WEAM model was employed on the test set to evaluate the model performance. Table 7.1 summarises the total number of indoor and outdoor events in each subset.

Table 7.1. End-use events in each WEAM data subsets

WEAM subsets	Number of indoor events	Number of outdoor events
Training (64%)	128,779	480
Testing (16%)	32,194	119
Validating (20%)	40,244	150
Total (100%)	201,217	749

## Pre-processing of data

Outdoor use normally represents a smaller percentage of a household's total water demand compared to indoor use. Willis et al. (2011) found outdoor use to range between 10%-18% of the total water demand. With outdoor use typically consuming larger volumes of water per event compared to indoor use, the number of outdoor events over a study period should thus theoretically be significantly less than the number of indoor events over the same period. Imbalanced data sets are thus expected in end-use studies, with indoor events being in the majority class and outdoor events in the minority class. Imbalanced data sets can cause classification problems to many machine learning algorithms. An accuracy driven algorithm can simply ignore the minority class and still achieve a high accuracy. For example, using the WEAM data set, an algorithm can achieve 99.63% accuracy if the algorithm classifies all events as indoor, since 99.63% of the data set consists of indoor events. In order to avoid model bias, the WEAM data set was balanced prior to model development.

Farquad and Bose (2012) summarizes different approaches to balance data sets. Data sets differ, and there is no balancing technique that works best on all data sets. Consequently, three balancing techniques were evaluated as part of this study, namely upsampling, downsampling, and class reweight. Upsampling randomly samples the outdoor events (with replacement) to be the same size as the indoor events (120,729 data points). Downsampling reduces the number of indoor events to match the sample size of the outdoor events (449 data points). Class reweight assigns a specific weight to the outdoor samples (the minority set), and the indoor samples (the majority set), corresponding with the sample size of each class. The smaller class size, outdoor, is given a much larger weight (121.94) than the larger class size, indoor (0.50), in order to balance the data set.

Similar to the method proposed by Farquad and Bose (2012), this study implemented a 2-step method to handle the imbalanced WEAM data set. First, the training data set was balanced using all three approaches mentioned previously. Thus, three different training subsets were generated in addition to the imbalanced training set. Secondly, a SVM model was constructed in R (R Core Team 2020) to ultimately determine which balancing method worked best for the WEAM data set. SVM identifies an optimal decision boundary to classify the end-use events as being indoor or outdoor. SVM is a very effective binary classification problem solver, and is one of the most efficient techniques proposed in literature (Wu et al. 2008). The SVM model with the best prediction accuracy on the corresponding training subset indicated which data balancing approach worked best on the WEAM data set. The maximum number of data points for acceptable computational time for SVM is 20,000. Consequently, to improve the computational time, the imbalanced training and upsampling training data subsets were reduced to include only 10% and 5% of the data points for SVM fit. The smaller sample sizes will reduce processing times, but still achieve satisfying results. Table 7.2 summarizes the training subsets.

Table 7.2. Training data subsets

Balancing method	Class size		Class size for improved computational time	
	Indoor	Outdoor	Indoor	Outdoor
Imbalanced	128,779	480	12,873	53
Upsampling	128,779	128,779	6,439	6,439
Downsampling	480	480	480	480
Weighted	128,779	480	12,873	53

### Model development: Machine learning algorithms

Developing a classification model to identify indoor events and outdoor events, based on the characteristics of an event ( $D$ ,  $V$  and  $I$ ), was the aim of this study. Machine learning algorithms are often used to solve binary classification problems on supervised data sets. Two popular machine learning algorithms, SVM and RF, were trained and evaluated using the balanced training data set. A SVM algorithm develops a decision boundary that aims at maximizing the margin, which is the minimum distance between the decision boundary and the data points. Due to the training data set not being linearly separable, a SVM with a non-linear kernel, radial basis function (RBF), was used to enhance SVM flexibility and robustness to fit the training data set. Different model parameters and hyperparameters were evaluated to tune the optimal SVM model. Two of the main parameters for model tuning include the soft margin error ( $C$ ), also known as the cost value, and the kernel parameter sigma. The  $C$  parameter controls the trade-off between the correct classification of the training data and maximizing the margin, by assigning a large penalty for errors. Repeated cross-validation was employed to tune the hyperparameters of both the SVM and RF algorithms, in order to select the best classification model with the highest performance. SVM does not perform well with large data samples due the computational complexity of the algorithm. The training time for SVM becomes impractical for data sets larger than 20,000 points, whereas RF runs efficiently on large data sets.

RF is intrinsically a large number of decision trees built out of randomly selected data samples and parameters. Each tree makes a prediction, and the class with the most votes is selected by the model. Due to the large number of decision trees forming a RF model, over fitting is highly unlikely. RF is also difficult to interpret since the classification decision information is hidden inside the model structure. Nonetheless, RF is an accurate algorithm and provides a good indicator of input parameter importance.

### Model development: Alternative algorithm

A third model was developed to provide an easily interpretable decision surface for comparison against the more abstract machine learning algorithms. Upon visual inspection of the data samples, it was hypothesized that an ellipsoid, in 3D space, with center (0, 0, 0), would best encapsulate most of the indoor events (Figure 7.1).

The ellipsoidal decision surface (EDS) can be written as:

$$\text{EDS} = \frac{x}{D_r^2} + \frac{y}{V_r^2} + \frac{z}{I_r^2} \left\{ \begin{array}{l} \leq 1 \text{ indoor use} \\ > 1 \text{ outdoor use} \end{array} \right\} \quad (7.1)$$

Where  $D_r$ ,  $V_r$  and  $I_r$  are the ellipsoid principal semi-axes, in the x, y and z direction, with units min, gal and gpm. The variables  $x$ ,  $y$ , and  $z$ , represent the event characteristics of the data point being classified, and correspond with  $D$  (min),  $V$  (gal) and  $I$  (gpm) of end-use events. All events plotting inside the ellipsoidal-surface were classified as indoor ( $\text{EDS} \leq 1$ ), while events beyond the surface would be considered outdoor ( $\text{EDS} > 1$ ).

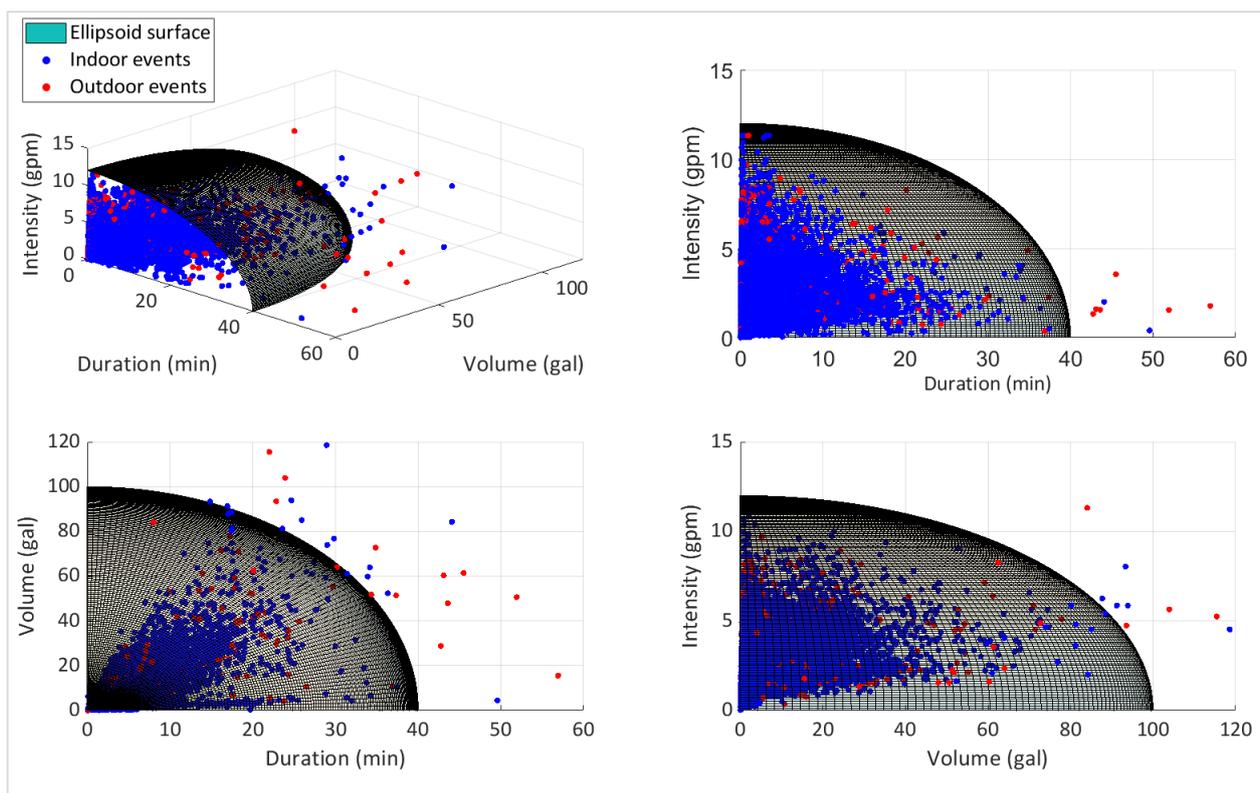


Figure 7.1. Ellipsoidal decision surface

A least square error algorithm was developed, to minimize the volume of the ellipsoid while simultaneously maximizing the classification performance of the model. The algorithm thus optimized EDS by finding the optimal values representing  $D_r$ ,  $V_r$  and  $I_r$ . As a first step, a sigmoid function was used to smooth the EDS function and eliminate the discontinuities. A grid search was done to determine the best starting point for the  $D_r$ ,  $V_r$  and  $I_r$  values.

Since the values of  $x$ ,  $y$ , and  $z$  represent the intercepts of an ellipse, the bounds of the grid search were chosen to be twice the greatest value of the maximum duration, volume and intensity found in the training data set (this was done to ensure that the algorithm allowed the ellipse to extend beyond the data set if necessary). The point (the triplet value of  $D_r$ ,  $V_r$  and  $I_r$ ) with the lowest least square error was chosen as the starting point. A couple iterations were performed using the smoothed function to refine the point, each time taking steps in the direction that provided the greatest reduction in square error. The values of  $D_r$ ,  $V_r$  and  $I_r$  were thus either increased (positive direction) or decreased (negative direction). Taking Figure 7.2a as an example, increasing  $I_r$  from 4 gpm to 4.1 gpm would provide the greatest reduction in square error. The effect is shown in Figure 7.2b, requiring the next iteration to take a step in the positive direction for  $D_r$ , with a step size of 0.1. When the least square error could not be reduced any further by taking steps in either direction (Figure 7.2c), the point (the triplet value of  $D_r$ ,  $V_r$  and  $I_r$ ) was recorded.

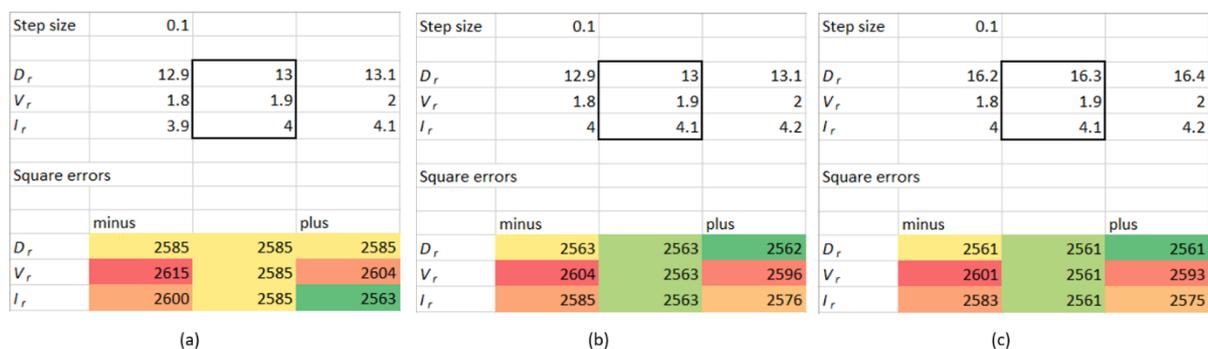


Figure 7.2. Example of reducing square error using the sigmoid function

The point (16.3, 1.9, 4.1) in Figure 7.2(c) was then transferred to the discontinuous model and further refined by taking steps in the direction of greatest reduction in square error, reducing the step size until an optimal accuracy was achieved. The smallest step size selected was 0.001. This was considered sufficient, as smaller values offered no improvement in accuracy, and are not friendly to future users of the model.

The EDS model is not expected to outperform the machine learning algorithms, however, due to the simplistic nature of the decision surface, the EDS model could be utilized as a quick estimate for end-use classification. The EDS model will thus also be compared to the UBL method proposed by Buchberger et al. (2003), as a more accurate alternative for quick classification of indoor and outdoor water use.

## Performance metrics

A confusion matrix was constructed for all the classification models, to give a picture of model performance. Two accuracy indices were adopted to review model performance, namely recall (also known as the sensitivity) and specificity. Recall is a measure of how well the model correctly predicts the positive class, in this case, indoor end-uses, and specificity measures how well the model predicts outdoor events, which is the negative class. The objective of the classification models was to maximize both recall and specificity. However, a trade-off exists between recall and specificity, and one cannot be maximized without negatively affecting the other. Therefore, the AUC of each classification model was calculated, which combines specificity and recall into a single number. The AUC is the area under the ROC. The ROC is a graphical plot of the true positive rate and the false positive rate at various threshold settings. The true positive rate is equivalent to the recall, and the false positive rate is equivalent to one minus the specificity. The AUC is a useful measure for imbalanced data sets, since it optimizes the classification of both classes.

## Model validation

The final WEAM model was employed on the unseen 20% of the data set (not used to train or test the model), in order to validate the model.

## RESULTS AND DISCUSSION

### Final training data set

Table 7.3 summarizes the performance of the SVM fit on the imbalanced and balanced training subsets. The desired value for each metric is 1.

Table 7.3. Training data subsets performance on SVM fit

Balancing method	Recall	Specificity	AUC
Imbalanced	1.000	0.000	0.500
Upsampling	0.856	0.789	0.822
Downsampling	0.860	0.754	0.807
Weighted	1.000	0.037	0.519

The weighted class balancing method did not perform well, and had similar model performance metrics compared to the imbalanced data set. Results from the SVM fit shows that both upsampling and downsampling balancing methods significantly improve the performance of the SVM model. Upsampling showed the best model performance from the SVM model, with an AUC value of 0.822 and specificity value of 0.789, and was subsequently used as the training data set for the rest of this study. The final WEAM data set thus consisted of 128,779 indoor events and 128,779 outdoor events.

### Best model fit and final model selection

As part of this study, three classification models were developed and trained on the balanced training data subset. The best performing SVM model had a C (soft margin error) value equal to 1, sigma value of 15.525, and a 17.76% training error. The best performing RF model consisted of 500 decision trees and had a training error of 0.36%. The input parameter importance ranking, calculated by the RF model, is duration, intensity, and volume. Evaluating the error as a function of  $D_r$  (Figure 7.3) showed that no significant improvements are made in the EDS model outcome when  $D_r$  is increased above the critical point.

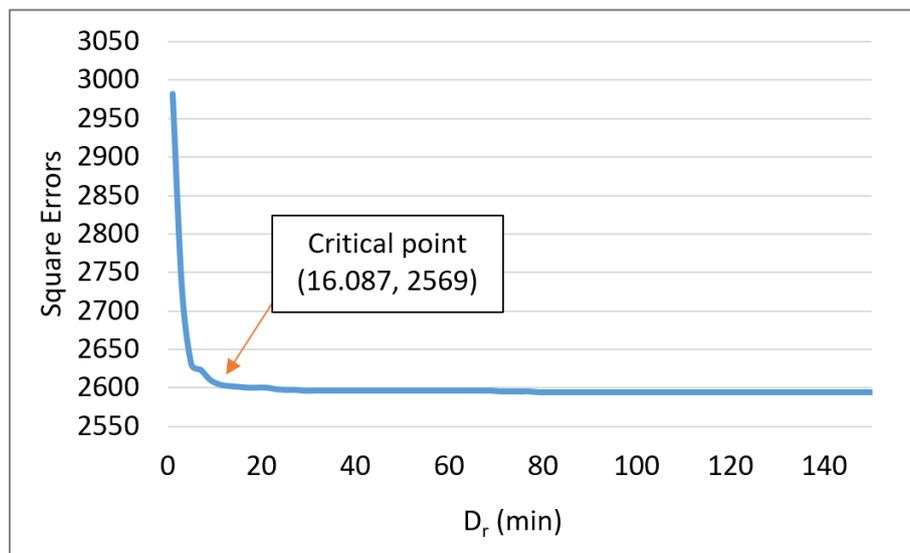


Figure 7.3. Error in the EDS model (Equation 7.1) as a function of  $D_r$

One of the objectives of the EDS model was to minimize the volume of the ellipsoid. Therefore, although the least squared error was achieved at  $D_r = 80$  min, the optimal  $D_r$  was selected as the critical point in Figure 7.3. The final EDS model is depicted in Equation 7.2.

$$\text{EDS} = \frac{x^2}{16.087^2} + \frac{y^2}{2.064^2} + \frac{z^2}{4.438^2} \begin{cases} \leq 1 & \text{indoor use} \\ > 1 & \text{outdoor use} \end{cases} \quad (7.2)$$

The best fit model for each of the three methods (SVM, RF and EDS) was deployed on the training data set to evaluate model performance. The UBL method was also applied to the training data set to predict event classes. This was done to create a realistic comparison of model performance between the EDS model and the UBL approach. The resulting classifications made by each of the four models are presented in a confusion matrix (Table 7.4). The confusion matrix provides details about the number of events in each class of the train data set as well as the number of events from the classification results.

Table 7.4. Confusion matrix on classification results (train set)

		Actual	
		Indoor	Outdoor
SVM Prediction	Indoor	109,909	26,001
	Outdoor	18,870	102,778
RF Prediction	Indoor	102,056	1,092
	Outdoor	26,723	127,687
EDS Prediction	Indoor	108,470	31,540
	Outdoor	20,309	97,239
UBL Prediction	Indoor	128,755	110,107
	Outdoor	24	18,672

Table 7.4 illustrates that out of the 128,779 indoor events in the WEAM data set, the UBL prediction method correctly identified 128,755 of the indoor events as indoor events. The UBL method thus performed the best in terms of correctly classifying indoor events as indoor event, and had the highest recall value of 0.9998. On the contrary, the UBL method performed the worst in correctly identifying outdoor events as outdoor events. The UBL method was only able to correctly identify 18,672 of the 128,779 outdoor events, resulting in a low specificity value of 0.1450. The specific performance metrics of each model are depicted in Table 7.5.

Table 7.5. Model performance on the training data set

Classification model	Recall	Specificity	AUC
SVM	0.8535	0.7981	0.8258
RF	0.9633	0.9935	0.9784
EDS	0.8423	0.7551	0.7987
UBL	0.9998	0.1450	0.5724

Table 7.5 shows that the RF model had a specificity value of 0.9935 and a recall value of 0.9633, implying that of the 128,779 indoor events in the training data set, 93.33% were correctly classified (recall), and of the 128,779 outdoor events in the training data set, 99.35% were correctly classified (specificity). A comparison of the ROC curves of all four classification models is depicted in Figure 7.4. The RF model performed the best, having the highest AUC value of 0.9784. Thus, for the purpose of this study, the RF algorithm was selected as the final WEAM model.

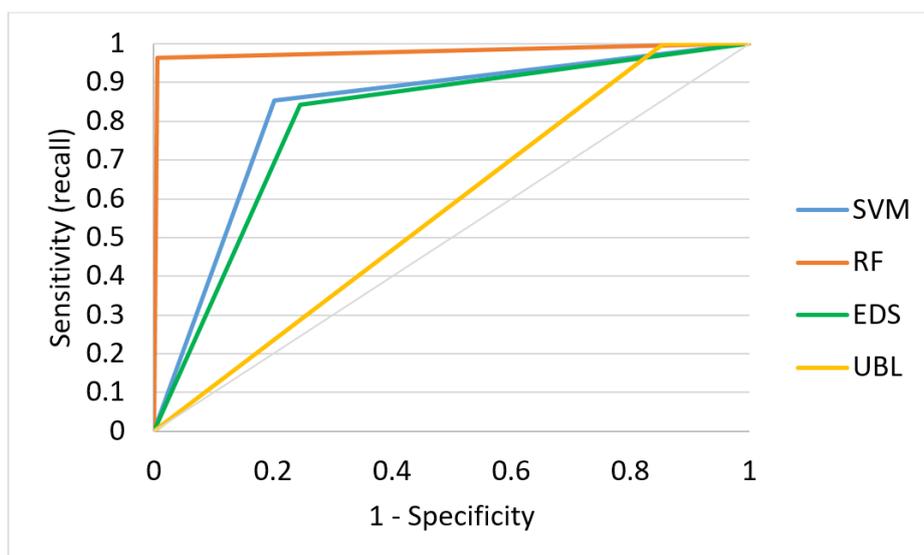


Figure 7.4. Comparing the ROC curves of the classification models

Table 7.4 and Table 7.5 also show that for imbalanced data sets, the EDS model performs much better than the UBL method in terms of correctly identifying event classes. Although the results show that the UBL has an almost perfect recall (0.9998), the model misclassified 85.50% of all outdoor events as indoor events. This high misclassification percentage suggests that the UBL method is not transferable to other locations or other time periods having different habits and household fixtures. The EDS model can thus be considered as an enhanced alternative to the UBL method, providing prediction results with high balanced accuracies. Utility managers could benefit from a more complex classification method, such as the EDS method, which has been calibrated with more recent end-use studies from two different regions.

### Model performance

The final WEAM model was employed on the test set to evaluate the model performance. A summary of the model performance is provided in Table 7.6. A confusion matrix, Table 7.6(a), as well as important statistical metrics, Table 7.6(b), are included. Utility managers are also interested in the feasibility aspects, such as the water demand for each class. Thus, the WEAM model performance on the event volumes were also evaluated, and is depicted in Table 7.6(c).

Off the 32,194 indoor data points, WEAM was able to classify 30,985 events correctly (96.25%). The model performed well and apportioned accurate volumes for each classification class. WEAM was able to apportion 82.99% of the indoor event volumes correctly, and 98.2% of the outdoor event volumes. The final WEAM model had an AUC value of 0.9140, which is considered excellent.

Table 7.6(a). WEAM confusion matrix

		Actual	
		Indoor	Outdoor
WEAM Prediction	Indoor	30,985	16
	Outdoor	1,209	103

Table 7.6(b). WEAM performance metrics on the test set

Performance metric	Value
Recall	0.9625
Specificity	0.8656
AUC	0.9140
Accuracy	0.9621

Table 7.6(c). WEAM performance on total event volume

	Performance metric	Performance
Indoor	Total indoor volume (gal)	41,161.50
	Indoor volume correctly classified (gal)	34,161.05
	Accuracy	82.99%
	Indoor volume misclassified as outdoor (gal)	7,000.45
	Misclassification rate	17.01%
Outdoor	Total outdoor volume (gal)	6,475.94
	Outdoor volume correctly classified (gal)	6,359.27
	Accuracy	98.2%
	Outdoor volume misclassified as indoor (gal)	116.67
	Misclassification rate	1.80%

### Model validation

Another model validation test was conducted to illustrate the performance and capabilities of the WEAM model. The WEAM model was employed on the remaining 20% of the WEAM data set, not used for model development (unseen data, not used for testing or training). The WEAM achieved an AUC value of 0.7846, classifying 96.25% of all indoor events correctly, and 60.67% of all outdoor events correctly. With respect to classifying end-use event volumes correctly, WEAM achieved a true positive rate (volume of indoor events correctly apportioned) of 81.94%, and a true negative rate (volume of outdoor water use correctly classified) of 97.64%.

Although the validation results show good model performance, the accuracy remained the same in terms of volume, when the model was employed on the unseen validation data set. The outdoor event count accuracy reduced when the model was employed on the unseen validation data set, but the event count was not considered to be as significant for practical application as the total volume. Future research could improve the model by adding additional training parameters to the data set, such as event start time, day of the week, socio economic factor, season, and so forth, to improve the model performance. It is important to note that only parameters that can be obtained from coarser end-use data sets should be added as input parameters. For instance, data sets with lower resolutions would not necessarily be able to extract both an event's peak intensity and most frequent intensity. Thus, thought should be given as to what input parameters are practical to obtain from coarser end-use data sets, before using the parameters to further calibrate WEAM.

## **CONCLUSION**

Understanding end-use water consumption at residential properties can improve the way municipalities and water authority managers monitor and manage water restriction interventions, especially during seasonal water scarcity. Past water end-use publications have focussed their studies on data obtained from smart meters producing high resolution data. Although these studies have made valuable contributions towards disaggregating individual household end-uses, consumption data are often only available at a reduced temporal or spatial resolution, especially in developing countries. Three different mathematical models were developed and compared for application on lower resolution data sets. The models were calibrated based on three input parameters (end-use events characteristics). The best performing classification model distinguishing between household indoor water use and outdoor water use, was termed the WEAM model. The random forest model outperformed the other models, and also showed the limitations to the current upper bound limit method.

The novel method presented in this paper now allows useful information to be extracted from relatively coarser end-use data sets – with specific reference to the classification of the water use event as being either indoor or outdoor. It is hoped that by applying this method on AMR meter data sets, water utilities from a range of socio-economic settings can have greater opportunities to improve water security through better informed demand management programs. Further research could explore the generalization of WEAM by incorporating socio-economic and climatological variables, including consideration for dwelling type (ex. single family homes, low cost housing, apartment blocks, etc.).

## **ACKNOWLEDGEMENTS**

The authors would like to acknowledge the detailed and helpful comments from two anonymous reviewers. We believe this is a much stronger paper as a result.

## DATA AVAILABILITY STATEMENT

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. (That includes the EDS algorithm, SVM code, RF code).

## REFERENCES

Arregui, F. 2015. "New software tool for water end-uses studies." Proceedings of 8th IWA International Conference on Water Efficiency and Performance Assessment of Water Services, Cincinnati, OH, USA, 20-24 April 2015: pp. 20-24.

Alcocer-Yamanaka, V., Tzatchkov, V.I., Arreguin-Cortes, F. 2012. "Modelling of drinking water distribution networks using stochastic demand." *Water Resour. Manage. EWRA*. 26(7): 1779-1792. doi: 10.1007/s11269-012-9979-2

Aquacraft 2010. Trace Wizard® software version 4.1. 1995-2010 Aquacraft, Inc. Boulder, CO, USA.

Athuraliya, A., Gan, K., Roberts, P. 2008. "Yarra Valley Water 2007 Appliance Stock and Usage Patterns Survey." Yarra Valley Water, Victoria.

Beal, C.D. and Stewart, R.A. 2013. "Identifying Residential Water End-Uses Underpinning Peak Day and Peak Hour Demand." *J. Water Resour. Plann. and Manage.*, 140(7): 04014008. doi:10.1061/(ASCE)WR.1943-5452.0000357

Beal, C.D. and Stewart, R.A. 2011. "South East Queensland residential end-use study: final report." Urban Water Security Research Alliance Technical Report No.47. ISSN 1836-5566 Gold Coast, Australia.

Beal, C.D., Jackson, M., Stewart, R.A., Miller, A., Rayment, C. 2018. "Identifying and understanding the drivers of high water consumption in remote Australian Aboriginal and Torres Strait Island communities." *J. of Cleaner Production*, 172: 2425-2434. doi:10.1016/j.jclepro.2017.11.168

Beal, C.D., Stewart, R.A., Fielding, K. 2011. "A novel mixed method smart metering approach to reconciling differences between perceived and actual residential end-use water consumption." *J. of Cleaner Prod.* 60(2013): 116-128. doi: 10.1016/j.jclepro.2011.09.007

Blokker, E.J.M., Vreeburg, J.H.G., van Dijk, J.C. 2010. "Simulating Residential water demand with a stochastic end-use model." *J. Water Resour. Plann. Manage.* 136(1): 19-26. doi: 10.1061/(ASCE)WR.1943-5452.0000002

Botha, B.E., Jacobs, H.E., Terblanche, U. 2018. "Clothes washing as household end-use: comparison of different appliance models in view of expected water savings." International Water Distribution System Analysis (WDSA) / Computing and Control for the Water Industry (CCWI) Joint Conference 2018, Kingston, Ontario, Canada, Vol 1(2018): 020.

Botha, B.E., Jacobs, H.E., Biggs, B., Ilemobade, A.A. 2017. "Analysis of shower water use and temperature at a South African university campus." International Computing and Control for the Water Industry (CCWI) Conference. doi:10.15131/shef.data.5364562.v1

Britton, T. C., Stewart, R. A., & O'Halloran, K. R. 2013. "Smart metering: enabler for rapid and effective post meter leakage identification and water loss management." Journal of Cleaner Production, 54: 166-176. doi: 10.1016/j.jclepro.2013.05.018

Buchberger, S.G. and Wells, G.J. 1996. "Intensity, duration, and frequency of residential water demands." J. Water Resour. Plann. Manage., 122(1): 11-19.  
doi: 10.1061/(ASCE)0733-9496(1996)122:1(11)

Buchberger, S.G. and Wu, L. 1995. "A model for instantaneous residential water demands." J. Hydr. Engrg., ASCE, 121(3), 232-246. doi: 10.1061/(ASCE)0733-9429(1995)121:3(232)

Buchberger, S.G., Carter, J.T., Lee, Y.H., Schade, T.G. 2003. "Random demands, travel times, and water quality in dead ends." Rep. No. 90963F, AWWARF, Denver.

Cole, G. and Stewart, R.A. 2013. "Smart meter enabled disaggregation of urban peak water demand: precursor to effective urban water planning." Urban Water J. 10(3): 174-194.  
doi: 10.1080/1573062X.2012.716446

Cominola, A., Giuliani, M., Castelletti, A., Rosenberg, D.E., Abdallah, A.M. 2018. "Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management." Environ. Modell. Software, 102(2018): 199-212.  
doi: 10.1016/j.envsoft.2017.11.022

Creaco, E., Farmani, R., Kapelan, Z., Vamvakieridou-Lyroudia, L., and Savic, D. 2015. "Considering the mutual dependence of pulse duration and intensity in models for generating residential water demand." J. Water Resour. Plann. Manage., ASCE, 141(11): 04015031.  
doi: 10.1061/(ASCE)WR.1943-5452.0000557

DeOreo, W.B. 2011. "Analysis of water use in new single-family homes." Aquacraft Inc. Boulder, CO.

DeOreo, W.B. and Mayer, P.W. 2012 “Insights into declining single-family residential water demands.” *J. American Water Works Association*. 104. doi:10.5942/jawwa.2012.104.0080

DeOreo, W.B., Mayer, P.W., Dziegielwski, B., Kiefer J.C. 2016. “Residential Uses of Water 2016.” Water Research Foundation. Denver, CO. ISBN 978-1-60573-236-7

Farquard, M.A.H. and Bose, I. 2012. “Preprocessing unbalanced data using support vector machine.” *Decision Support Systems*. 53 (2012): 226–233. doi:10.1016/j.dss.2012.01.016

Fisher-Jeffes, L., Gertse, G., Armitage, N.P. 2015. “Mitigating the impact of swimming pools on domestic water demand.” *Water SA*. 41(2): 238–246. doi: 10.4314/wsa.v41i2.09

Glenn, D.T, Endter-Wada, J., Kjelgren, R., Neale, C.M.U. 2015. “Tools for evaluating and monitoring effectiveness of urban landscape water conservation interventions and programs.” *Landscape and Urban Plan*. 139(2015): 82-93.  
doi: 10.1016/j.landurbplan.2015.03.002

Ilemobade, A., Jacobs, H.E. and Botha, B. E. 2018. “Sorting high resolution end-use data – A pleasurable headache.” 1st International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada – July 23-25.

Inman, D., Jeffrey, P. 2006. “A review of residential water conservation tool performance and influences on implementation effectiveness.” *Urban Water J*. 3(3):127–143.  
doi: 10.1080/15730620600961288

Jacobs, H.E., Fair, K., Geustyn, L.C., Daniels, J., Du Plessis, J.A., 2007. “Analysis of water savings: A case study during the 2004-2005 water restrictions in Cape Town.” *Journal of the South African Institution of Civil Engineers*, 49(3):16-26.

Jorgensen, B.S., Martin, J.F., Pearce, M.W., Willis, E.M. 2013. “Predicting household water consumption with individual-level variables.” *Env. Behaviour*, 46(7): 872-897.  
doi: 10.1177/0013916513482462

Makki, A., Stewart, R.A., Beal, C.D., Panuwatwanich, K. 2015. “Novel bottom-up urban water demand forecasting model: revealing the determinants, drivers and predictors of residential indoor end-use consumption.” *Res., Conservation and Recycling*, 95: 15-37.  
doi:10.1016/j.resconrec.2014.11.009

Makki, A.A., Stewart, R.A., Panuwatwanich, K., Beal, C. 2011. “Revealing the determinants of shower water end-use consumption: enabling better targeted urban water conservation strategies.” *J. Cleaner Production*, 60: 129-146. doi: 10.1016/j.jclepro.2011.08.007

Makwiza, C. and Jacobs, H.E. 2017. "Sound recording to characterize outdoor tap water use events." *J. Water Supply Res. Technol. AQUA*. 66(6): 392-402. doi: 10.2166/aqua.2017.120

Meyer, B.E. and Jacobs, H.E. 2019. "Garden irrigation as household end-use in the presence of supplementary groundwater supply." *Water SA* 45(3): 447-455.  
doi: 10.17159/wsa/2019.v45.i3.6741

Meyer, B.E., Jacobs, H.E., Ilemobade, A. 2020. "Extracting household water use event characteristics from rudimentary data." *J. Water Supply Res. Technol. AQUA*. 69(4): 387–397.  
doi: 10.2166/aqua.2020.153

Meyer, N., Jacobs, H.E., McKenzie, R. and Westman, T. 2018. "The effect of controlled pressure adjustment in an urban water distribution system on household demand." *J. Water Supply Res. Technol. AQUA*. 67(3):218-226. doi: 10.2166/aqua.2018.139

Nel, N. and Jacobs, H.E. 2019. "Investigation into untreated greywater reuse practices by suburban households under the threat of intermittent water supply." *Journal of Water, Sanitation and Hygiene for Development* 9(4):627-634. doi: 10.2166/washdev.2019.055

Nguyen, K.A., Stewart, A.R, Zhang, H., Sahin, O. 2018. "An adaptive model for the autonomous monitoring and management of water end-use." *Smart Water* 3(5).  
doi:10.1186/s40713-018-0012-7

Nguyen, K.A., Zhang, H., Stewart, A.R. 2013. "Development of an intelligent model to categorise residential water end-use events." *J. Hydro-environ. Res.* 7(2013): 182-201.  
doi:10.1016/j.jher.2013.02.004

Pastor-Jabaloyes, L., Arregui, F.J., Cobacho, R. 2018. "Water end-use disaggregation using soft computing techniques." *Water* 10(1):46. doi: 10.3390/w10010046

Pretorius, D., Crouch, M.L., Jacobs, H.E. 2019. "Diurnal water use patterns for low-cost houses with indigent water allocation: a South African case study." *J. Water Sanitation Hyg. for Develop.* 9 (3): 513-521. doi: 10.2166/washdev.2019.165

R Core Team 2020. "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Scheepers, H.M. and Jacobs, H.E. 2014. "Simulating residential indoor water demand by means of a probability based end-use model." *J. Water Supply Res. Technol. AQUA*. 63(6): 476-488. doi: 10.2166/aqua.2014.100

Sterne, L. N. 2019. "Characterisation of household water use events using a non-intrusive sensor." M.Eng. thesis, Stellenbosch, South Africa: Stellenbosch University.

Survis, F.D. and Root, T.L. 2012. "Evaluating the effectiveness of water restrictions: A case study from Southeast Florida." *J. Environ. Manage.* 112: 377-383.

doi: 10.1016/j.jenvman.2012.08.010.

Vitter, J.S., Webber, M. 2018. "A non-intrusive approach for classifying residential water events using coincident electricity data." *J Environ Model Softw* 100:302-313.

doi: 10.1016/j.envsoft.2017.11.029

Willis, R.M., Stewart, R.A., Giurco, D.P., Talebpour, M.R., Mousavinejad, A. 2013. "End-use water consumption in households: impact of socio-demographic factors and efficient devices." *J. Cleaner Prod.* 60: 107-115. doi: 10.1016/j.jclepro.2011.08.006

Willis, R.M., Stewart, R.A., Panuwatwanich, K., Williams, P.R. and Hollingsworth, A.L. 2011. "Quantifying the influence of environmental and water conservation attitudes on household end use water consumption." *J. Environ. Manage.* 92(8): 1996-2009.

doi: 10.1016/j.jenvman.2011.03.023

Wu, X. Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.s., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D. 2008. "Top 10 algorithms in data mining." *Knowledge and Information Systems.* 14(1):1-37.

doi: 10.1007/s10115.007.0114.2.

## Chapter 8.

### Classifying household water use into indoor and outdoor use from a rudimentary data set – A case study in Johannesburg, South Africa

Bettina Elizabeth Meyer<sup>1\*</sup>, Heinz Erasmus Jacobs<sup>1</sup> and Adeshola Ilemobade<sup>2</sup>

1. Department of Civil Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa, [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za), [hejacobs@sun.ac.za](mailto:hejacobs@sun.ac.za)

2. School of Civil and Environmental Engineering, University of the Witwatersrand Johannesburg, Private Bag 3, WITS, 2050, South Africa, [Adesola.Ilemobade@wits.ac.za](mailto:Adesola.Ilemobade@wits.ac.za)

\*Corresponding author: [bebotha@sun.ac.za](mailto:bebotha@sun.ac.za)

This paper was submitted for review and possible publication to the Journal of Water, Sanitation & Hygiene for Development (JWaSHDev)

#### **ABSTRACT**

Distinguishing between indoor use and outdoor use is becoming increasingly important, especially in water scarce regions, since outdoor use is typically targeted during water restrictions. Household water use is typically measured at a single water meter, and the resolution of the metered data is typically too coarse to employ on commercially available disaggregation software, such as flow trace analysis. This study is the first to classify end-use events from a rudimentary data set, into indoor use or outdoor use. This case study was conducted in Johannesburg, South Africa, and quantified the volume of water used indoors and outdoors at 63 residential properties over 217 days. A recently developed model for classifying water use events as either indoor or outdoor, based on relatively coarse water meter data, was employed in this study. A total of 212 060 single end-use events were classified as being either indoor or outdoor. The indoor and outdoor consumptions were compared to survey results and demand predictions made for the study area. It was found that 30% of all events were outdoor, based on the total volume.

**Keywords:** End-use events, Low resolution data, Residential water demand, Water classification models

## INTRODUCTION

### Household water consumption

Water demand continues to increase due to rapid rates of population growth (Vörosmary et al. 2005). Water utilities require the most detailed and accurate information regarding residential water consumption when developing water demand management (WDM) strategies. The effectiveness of applying water demand strategies is remarkably reduced because of the limited understanding of residential consumption (Sahin et al. 2014). Better knowledge and understanding of how and where households consume water allow for targeted and effective WDM strategies as well as economic incentives (Nguyen et al. 2013).

High resolution (sub-minute sampling) data have been used in the past to run water end-use disaggregation algorithms to provide detailed information on household end-use consumption behaviour. Household end-uses include the shower, washing machine, toilet, dishwasher, taps, and garden irrigation (Nguyen et al. 2013). Residential water consumption could fundamentally be classified as either indoor use or outdoor use. Table 8.1 summarises a range of water end-use studies reporting on indoor and outdoor water use as distinct components of total household water consumption. Studies conducted during periods with water restrictions enforced were not included in Table 8.1. The end-use studies presented in Table 8.1 were based on high resolution data (0.014 L/pulse – 0.1 L/pulse every 1 s – 10 s) and employed flow trace analysis software for end-use classification.

### Conventional and smart water meters

Smart meters record water consumption information and communicates this information on a real-time basis (Cole and Stewart 2013). Smart meters are regarded as water meters linked to loggers that record at high resolution frequencies, allowing for automated data measurement readings and real time monitoring (Giurco et al. 2008). The value derived from smart meter data is dependent on the meter resolution and the logging frequency. Smart meters are able to record high resolution data at volumetric measurements of 0.014 L/pulse (compared to the 0.5 L/pulse or 1.0 L/pulse measured by conventional mechanic meters), and at logging frequencies of 1 s, 5 s or 10 s (Nguyen et al. 2013, Beal and Stewart 2013, Mead and Aravinthan 2009, Willis et al. 2011, Kowalski and Marshallsay 2005, Roberts 2005). The high resolution time series data may be paired with advanced flow trace analysis software to disaggregate end-use events. Smart meters, however, are not common. The costs of smart water meters are relatively higher than regular water meters. Additionally, more data are required to be communicated, stored, and processed, which requires additional infrastructure and technical staff with the relevant expertise.

Table 8.1. Residential indoor and outdoor water consumption

End-use study	Location	Percentage of total water demand			Comment
		Indoor	Outdoor	Leaks	
Mayer and DeOreo (1999)	USA	35.8%	58.7%	5.5%	
Loh & Coghlan (2003)	Perth, Australia	45.0%	54.0%	1.0%	
Roberts (2005)	Yarra Valley, Australia	68.9%	25.4%	5.7%	Average annual contributions.
Heinrich (2007)	Auckland, New Zealand	88.0%	8.0%	4.0%	
Beal et al. (2011)	Brisbane, Australia	79.5%	7.2%	13.3%	Leaks, dishwasher, irrigation and bath water use were reported in some, but not all, of the homes. In homes where outdoor use was reported, outdoor use was reported to be 20.6% of the total consumption.
	Gold Coast, Australia	86.3%	9.4%	4.3%	
	Sunshine Coast, Australia	79.1%	6.8%	14.1%	
	Ipswich, Australia	95.4%	1.7%	2.9%	
Willis et al. (2009)	Gold Coast, Australia	91.0%	8.0%	1.0%	Sample group reported a high level of concern for water conservation.
		85.0%	14.0%	1.0%	Sample group reported a medium level of concern for water conservation.
Hussien et al. (2016)	Duhok city, Iraqi Kurdistan	96.0%	4.0%	0.0%	Medium to high income households. Study was conducted over winter months. Hussien et al. (2016) suggests outdoor consumption to be much higher over the summer period.
		92.4%	7.6%	0.0%	
		91.8%	8.2%	0.0%	

Water authorities often collect water use data manually on a monthly, quarterly, or biannually basis (Nguyen et al. 2013). This practice results in daily or sub-daily water demand being estimated as an average water use, which can lead to inaccuracies. Current water metering systems predominantly rely on mechanical water meters, which generate a pulse after a specified volume has passed through the water meter, say every 0.5 L, 1.0 L or 5.5 L (Roberts 2005, Cole and Stewart 2013), without being able to record the time of any particular event smaller than the meter pulse volume (Nguyen et al. 2013). Data recorded at such coarse resolutions are considered rudimentary data, as the resolutions are too low for commercially available end-use disaggregation software (Meyer et al. 2020). Subsequently, investigations into household end-use consumption have never been conducted despite some studies reporting on more regular recording frequencies of 15 min (Pretorius et al. 2019), or 1 h (Cardell-Oliver et al. 2016).

Knowledge regarding household water consumption at end-use level is essential for understanding residential water consumption behaviour (Stewart et al. 2010). Effective water monitoring methods become increasingly important in water scarce regions prone to water restrictions, which typically target outdoor use (Hemati et al. 2016). Cominola et al. (2018) reported that sub-minute metering frequencies are required for end-use disaggregation, however, the trade-off between the meter pulse volume and the extent of information gained from the metered data has yet to be explored. The financial benefits of investing in smart metering technology have not been extensively investigated, contributing to the reluctance by many utilities to invest in the technology. Implementing regular water meters is more economically viable compared to more expensive smart meters, especially over a large scale.

### **Description of study site**

Increasing drought and population growth in many South African communities have driven the need to understand household water consumption behaviour. During severe drought conditions in 2015, the National Department of Water and Sanitation (DWS) restricted water use and put in place a 15% curtailment on urban water use (DWS 2016). As a result, Johannesburg Water (JW) introduced level-2 water restrictions in November 2015 and water restriction tariffs in September 2016 (JW 2016). Johannesburg, located in South Africa, is serviced by JW. Under level 2 water restrictions, consumers are limited to only irrigate their gardens with hand held hosepipes or buckets, and garden irrigation is only permitted between 06h00 and 18h00 every day. Car washing and swimming pool filling were not permitted under level 2 restrictions. Johannesburg's rainfall is concentrated in the warm summer period. During winter, Johannesburg experiences dry seasons. The month with the lowest number of average rain days (2 days) is June (winter), and the highest number of average rain days (15 days) is January (summer).

Residential water use in Johannesburg is normally measured and billed monthly. JW commissioned this case study and set out to determine to what extent measured rudimentary data can be used to obtain water end-use information at a household level. The study site comprised 63 homes in the Lonehill suburb and was conducted from September 2016 to January 2018. The study sample was divided into 54 residential semi-detached town houses in a security complex and 9 stand-alone residential properties. The plot sizes range from approximately 150 m<sup>2</sup> to 250 m<sup>2</sup> within the security complex and from approximately 1 000 m<sup>2</sup> to 1 500 m<sup>2</sup> for the stand-alone properties. The people per household (PPH) ranged from 1 person to 4 people. Lonehill is a middle- to high-income suburb. It has a literacy rate of more than 92%, covers a land area of about 5 km<sup>2</sup>, and has an average household income more than double that of South Africa and Gauteng Province.

## Objectives

Specific objectives of the case study were to:

- determine outdoor and indoor water use expressed as a percentage of the total household water demand;
- better understand household water consumption within the case study site; and
- compare classification results with theoretical demand estimates and survey responses.

## METHODOLOGY

### Data collection

Sensus iPerl water meters were installed at the 63 properties and recorded water flow measurements at a resolution of 1 L/pulse (in line with common utility meter resolutions). The meters were combined with data loggers (recording at 15 s intervals), in order to investigate what level of household water consumption information can be obtained from a rudimentary data set. The meters were paired with loggers to allow for sub-minute recordings, which is required for end-use extraction. The study period (September 2016 to January 2018) was selected because of the availability of resources (e.g. students and research funds) and physical access to the meters within the security complex. The data measured by the water meter were transmitted and stored on a FTP server, 30 km from the study site. Smart meter data were missing during some days (or prolonged periods). While some vacancy of property is normal, other challenges regarding the infrastructure and software contributed to the zero consumption days, and was reported on by Ilemobade et al. (2018). The total number of days with recorded consumption was 217 days. Data from the JW billing system were also collected for the period June 2016 to May 2017.

Detailed information on the properties and their residents were gathered using questionnaires (surveys). The questionnaires were developed and administered to willing household respondents in 2017. Prior to administration, ethics clearance was applied for and obtained from the University of the Witwatersrand, Johannesburg. Roughly half of the study sample completed the surveys (32 out of the 63), of which 24 (68%) opted to remain anonymous. Of the 32 survey responses received, only 11 respondents indicated their physical address. Only 11 of the homes could thus be linked to corresponding water meter data. In addition to the surveys, meter verification exercises were conducted at six properties. The meter verification involved simultaneously taking smart meter and consumer meter readings at specific end-uses (i.e. toilet, bath, shower and basin). This exercise, while simple, provided valuable additional information about the validity of the smart meter and consumer meter readings. The meter verification exercises also allowed for on-site leak inspections, and no real leaks were reported.

## Theoretical demand estimates

Jacobs et al. (2017) provided theoretical estimates of water end-use consumption at homes located in Johannesburg, South Africa. The theoretical estimates were made using the Residential End-Use Model (REUM) (Jacobs and Haarhoff 2004). Predictions for indoor use were based on the PPH, ranging from 1 PPH to 4 PPH for middle income homes. The typical end-use event volume and frequency of the most notable end-uses were used to calibrate REUM, and the values were based on earlier studies. The outdoor predictions were based on plot sizes (500m<sup>2</sup> and 1500m<sup>2</sup>), rainfall, crop factor and evaporation. The rainfall predictions were determined using historical data, dating back  $\pm 90$  years. Table 8.2 summarises the theoretical estimates for 4 different occupancy values, and 2 different plot sizes.

The values presented in Table 8.2 are assumed to be typical of middle-income households in the Johannesburg area. The proportion of household demand contributing to indoor use is higher for homes with a larger number of occupants. In line with other findings, the proportion of outdoor water consumption, for larger properties with irrigated gardens, is notably higher in summer (December to February), when rainfall is prevalent. This large outdoor water requirement can be explained by the difference between the evaporation rate and rainfall in the summer months (Jacobs et al. 2017).

Table 8.2. Theoretical demand estimates for residences in the Johannesburg

Plot size		500 m <sup>2</sup>											
PPH	Class	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
1	Indoor (%)	58%	57%	51%	56%	59%	66%	77%	86%	89%	83%	77%	68%
	Outdoor (%)	42%	43%	49%	44%	41%	34%	23%	14%	11%	17%	23%	32%
2	Indoor (%)	75%	75%	69%	73%	76%	81%	88%	93%	95%	92%	88%	83%
	Outdoor (%)	25%	25%	31%	27%	24%	19%	12%	7%	5%	8%	12%	17%
3	Indoor (%)	82%	81%	77%	80%	82%	86%	91%	95%	96%	94%	91%	87%
	Outdoor (%)	18%	19%	23%	20%	18%	14%	9%	5%	4%	6%	9%	13%
4	Indoor (%)	85%	85%	81%	84%	86%	89%	93%	96%	97%	95%	93%	90%
	Outdoor (%)	15%	15%	19%	16%	14%	11%	7%	4%	3%	5%	7%	10%
Plot size		1,500 m <sup>2</sup>											
PPH	Class	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
1	Indoor (%)	17%	16%	13%	15%	17%	21%	33%	48%	56%	41%	32%	23%
	Outdoor (%)	83%	84%	87%	85%	83%	79%	67%	52%	44%	59%	68%	77%
2	Indoor (%)	31%	30%	24%	28%	31%	37%	52%	67%	73%	61%	51%	40%
	Outdoor (%)	69%	70%	76%	72%	69%	63%	48%	33%	27%	39%	49%	60%
3	Indoor (%)	39%	38%	32%	37%	40%	47%	61%	75%	80%	69%	60%	50%
	Outdoor (%)	61%	62%	68%	63%	60%	53%	39%	25%	20%	31%	40%	50%
4	Indoor (%)	46%	45%	38%	43%	46%	53%	67%	80%	84%	75%	66%	56%
	Outdoor (%)	54%	55%	62%	57%	54%	47%	33%	20%	16%	25%	34%	44%

## Data processing

In order to classify water use events, individual events first had to be extracted from the measured data. The raw data from the study site had durations of time when no data was recorded. The gaps in measured data presented challenges when cleaning the data set and preparing it for analysis. Ilemobade et al. (2018) discussed other factors that contributed to and exacerbated the anomalies in the data set, and also presented the process of cleaning the raw data set. Prior to data analysis, 9 homes were removed from the study sample due to poor data quality. Thus, 54 homes remained in the study sample.

Meyer et al. (2020) developed a Python End-use Extraction Tool (PEET), which could be employed to extract event characteristics (i.e. duration, volume, flow intensity) of individual end-uses from the cleaned time series data. PEET identified the start of an event when a pulse measurement (volume in L) was recorded. If no subsequent measurement was taken within 30 s, it constituted the end of an event. The difference between the initial meter volume measurement and the final volume measurement was the volume of the single event. The difference in the time stamp of the first recording and the recording after 30 s, was the duration of the event. PEET assumed an event to have a rectangular shape (Alcocer-Yamanaka et al. 2012, Buchberger and Wu 1995). Thus, the intensity of an event was determined by dividing the event volume by the event duration. Intensity was calculated in L/s.

Because of the rudimentary nature of the data (limited to 1 L/pulse), Meyer et al. (2020) could not distinguish between a genuine 1 L event over a 15 s interval, and multiple smaller events (such as filling a glass of water or rinsing a plate) that accumulated to 1 L over the period. These two instances would appear identical, as both would result in a 1 L measurement taken over 15 s, with an intensity of 0.067 L/s. Meyer et al. (2020) grouped all the events with intensities  $< 0.067$  L/s and categorised these events as minor events. All other extracted events were ascribed as major events. In order to classify end-use events, all minor events were removed from the data set, and were labelled as unknown events. The final data set thus only consisted of major events. Major events comprised 75.8% of all event consumption in the extracted data set, meaning 24.2% of the initial data set was filtered out and labelled as unknown events. The final data set presented by Meyer et al. (2020) consisted of 212 060 major end-use events.

## Classification model

PEET is able to extract individual end-uses from a rudimentary time series data set, but not able to classify the end-uses as being indoor or outdoor. As a result, Meyer et al. (submitted) developed a classification model, WEAM, which is able to categorise an end-use event as being indoor or outdoor, based on three event characteristics, namely event duration ( $D$ ), event volume ( $V$ ), and event intensity ( $I$ ). Similar to Trace Wizard (Aquacraft 2010) and Identiflow (Kowalski and Marshallsay 2005), WEAM categorises end-uses based on decision

trees. WEAM is intrinsically a random forest (RF) machine learning algorithm which makes classification predictions by evaluating the correlation between the event characteristics ( $D$ ,  $V$ ,  $I$ ). Meyer et al. (submitted) reported that WEAM is able to correctly classify between 60.7% and 96.2% of end-use events as being either indoor use or outdoor use. With respect to classifying end-use events correctly, WEAM correctly apportioned 81.94% of the total indoor consumption as indoor events, and 97.64% of the total outdoor events were correctly classified as outdoor events. One major benefit of WEAM is its applicability on rudimentary data sets.

WEAM was thus selected as the classification model for this case study, due to its high accuracy and applicability on rudimentary data sets. It is important to note that the event characteristic values extracted by PEET first have to be converted from SI units to Imperial units, in order to be classified by WEAM.

## **RESULTS AND DISCUSSION**

### **Final data set**

As part of the case study, questionnaires were administered to the residents to gather information regarding the number of people living in each home, end-use fixtures, end-use patterns, etc. Only 11 of the 63 administered questionnaires provided useful information, which is why the focus of this study subsequently shifted to the 11 properties. The properties were renumbered accordingly, Home H01 through H11, in line with ethical requirements. Home H11 was the lone single, stand-alone residential property, and the other 10 homes were semi-detached town houses in a security complex. The number of people in each of the homes were determined from the questionnaire responses.

There were several periods (months) over the study period with anomalies and measurement gaps. Potential reasons for these data gaps have been articulated earlier in the data processing section (i.e. infrastructure challenges). From May 2017 until September 2017, no meter data were recorded. Table 8.3 depicts the number of days in each month meter data were recorded. An assumption was made that days with measured data were an acceptable representation of the indoor and outdoor demand ratio for the particular month. In other words, even with data gaps, sufficient information was obtained from the recorded data to satisfactorily represent consumer behaviour in terms of outdoor use and indoor use. The only time this assumption was invalid was for April 2017, where 3 days of measured consumption was considered inadequate to represent the entire month's water use behaviour.

Table 8.3. Dates with reported water use from meter measurements

Month	Sep 2016	Oct 2016	Nov 2016	Dec 2016	Jan 2017	Feb 2017	Mar 2017	Apr 2017	Oct 2017	Nov 2017	Dec 2017	Jan 2018
Number of recorded days	24	31	30	31	31	28	31	18	11	30	31	31
Home Code	Number of days with readings											
H01	15	22	24	4	10	12	21	4	0	0	0	0
H02	23	23	25	14	13	21	24	3	0	0	0	0
H03	23	22	29	14	12	21	24	3	11	13	19	16
H04	23	16	22	12	12	21	24	3	11	13	17	16
H05	23	19	23	10	12	20	24	2	11	13	16	14
H06	23	22	26	13	12	20	24	3	11	13	18	6
H07	23	21	28	10	12	17	24	3	11	13	13	17
H08	23	21	29	12	13	22	24	3	11	13	18	17
H09	20	17	23	11	11	19	23	3	11	12	18	17
H10	23	23	27	12	12	19	23	3	11	14	19	18
H11	23	23	29	13	12	22	24	3	11	13	20	17

### Classification results

The initial data sample consisted of 63 homes, however, due to poor data quality at 9 homes, the study sample was reduced to 54 homes. PEET extracted end-use events and filtered out all minor events, which contributed to 24.2% of the total volume of the household demand. Subsequently, these minor events were categorised as “unknown” consumption, since it was unclear whether these minor events were indoor or outdoor low flow events or whether they were background leaks. The classification results obtained from employing WEAM on the data set are depicted in Table 8.4.

Further investigation only focussed on the 11 homes chosen based on information obtained from survey responses. The proportion of indoor use and outdoor use as a percentage of the total consumption is also summarised in Table 8.4. Table 8.4 shows that the 11 homes selected was a good representation of the entire data set in terms of apportioned indoor use, outdoor use, and unknown events as a percentage of the total demand.

Table 8.4. Classification of end-use events

Data set	Proportion of total demand (%)		
	Indoor use	Outdoor use	Unknown
Entire data set	45.48	30.30	24.22
11 homes	46.98	30.43	22.59

## Correlation between proportion of total water demand and factors influencing household water demand

The proportion of the total water demand classified as indoor and outdoor events, for each of the 11 homes over the total study period, are summarised in Table 8.5. The home specific information, such as PPH and property size are also included in Table 8.5.

Table 8.5. End-use event classifications and household information

Home Code	PPH	Property size (m <sup>2</sup> )	Proportion of total demand (%)			
			Indoor	Outdoor	Unknown	TOTAL
H01	4	201.9	65.3	30.0	4.7	100.0
H02	1	168.3	87.7	7.1	5.1	100.0
H03	4	207.5	59.0	18.8	22.2	100.0
H04	2	168.3	72.3	20.1	7.7	100.0
H05	3	237.9	40.6	40.1	19.3	100.0
H06	2	207.0	61.4	14.1	24.4	100.0
H07	2	167.5	60.1	10.3	29.6	100.0
H08	1	212.6	51.7	40.0	8.3	100.0
H09	1	167.9	6.7	4.3	88.9	100.0
H10	1	168.3	31.8	62.9	5.4	100.0
H11	3	1 141.8	39.3	46.8	13.9	100.0

Although restriction tariffs were introduced in September 2016, no water restrictions prohibited outdoor water use. Home H09 showed inadequate results, with over 88% of the household water consumption categorised as unknown use, and was thus not further considered for analysis. The theoretical estimates depicted in Table 8.2 showed a distinct correlation between PPH and the percentage of total demand attributed to indoor use. The indoor use proportion of total demand is higher for homes with higher occupants. This correlation is not so apparent in Table 8.5. Table 8.6 and Table 8.7 were subsequently generated to visually show any correlation between PPH and indoor use as a proportion of total household water demand, as well as any correlation between property size and outdoor use. Table 8.6 ranks the PPH from least (1) to most (4), and Table 8.7 ranks according to property sizes, starting with the smallest.

Table 8.6. Correlation between PPH and indoor use as proportion of total demand

Home Code	PPH	Property size (m <sup>2</sup> )	Proportion of total demand (%)			
			Indoor	Outdoor	Unknown	TOTAL
H02	1	168.34	87.7	7.1	5.1	100.0
H08	1	212.62	51.7	40.0	8.3	100.0
H10	1	168.28	31.8	62.9	5.4	100.0
H04	2	168.33	72.3	20.1	7.7	100.0
H06	2	206.97	61.4	14.1	24.4	100.0
H07	2	167.51	60.1	10.3	29.6	100.0
H05	3	237.94	40.6	40.1	19.3	100.0
H11	3	1 141.75	39.3	46.8	13.9	100.0
H01	4	201.92	65.3	30.0	4.7	100.0
H03	4	207.47	59.0	18.8	22.2	100.0

Table 8.7. Correlation between property size and outdoor use

Home Code	PPH	Property size (m <sup>2</sup> )	Proportion of total demand (%)			
			Indoor	Outdoor	Unknown	TOTAL
H07	2	167.51	60.1	10.3	29.6	100.0
H10	1	168.28	31.8	62.9	5.4	100.0
H04	2	168.33	72.3	20.1	7.7	100.0
H02	1	168.34	87.7	7.1	5.1	100.0
H01	4	201.92	65.3	30.0	4.7	100.0
H06	2	206.97	61.4	14.1	24.4	100.0
H03	4	207.47	59.0	18.8	22.2	100.0
H08	1	212.62	51.7	40.0	8.3	100.0
H05	3	237.94	40.6	40.1	19.3	100.0
H11	3	1 141.75	39.3	46.8	13.9	100.0

Table 8.6 suggests no observed correlation exists between PPH and indoor use as a proportion of total household demand, which is in contrast with the theoretical estimates shown in Table 8.2 and counter-intuitive. This does not mean that indoor use does not increase with an increase in PPH, since such a correlation has been reported on in numerous studies (Blokker et al. 2010, Mead and Aravinthan 2009, Bradley 2004, Liu et al. 2003, Martinez-Espineira 2002). It is impossible for both indoor use and outdoor use percentages to increase within a home since the total (100%) is fixed. Therefore, one reason the correlation between PPH and indoor use is possibly not shown in Table 8.6 is due to the smaller impact indoor events have on total demand. Indoor events typically have smaller volumes compared to outdoor event volumes.

The correlation between outdoor events and property size were also investigated, and is depicted in Table 8.7. With the exception of House H10, an increase in property size results in a larger proportion of the total demand being attributed to outdoor use. Previous studies have reported on a direct relationship between outdoor water use and property size (Fox et al. 2009, Gato 2006, Jacobs and Haarhoff 2007). Due to outdoor use typically being larger volume events compared to indoor events, the increase in outdoor water demand has a more notable impact on the total demand.

### **Comparison between metered results, billing data, and survey responses**

The average daily household water use extracted from consumer meters (billing data) were compared to the derived average daily water use recorded by the smart meters. For the purpose of this comparison, the water use was evaluated over the total recording period for each device. In other words, zero consumption days were removed from the recording period, in order to restrict the impact of zero consumption on the average daily use. Table 8.8 provides a summary of the results for the 10 homes with available consumer meter data linked to survey responses.

The meter verification exercise conducted as part of this study confirmed that the smart meters' errors are permissible. Thus, the high difference between the average per capita water use values for the mechanical meters (billing data) and the smart meters is most likely due to metering error of the mechanical meter. Past studies have reported meter errors as high as 53% due to meter aging (Mutikanga et al. 2011). Future research could possibly conduct field tests to evaluate the accuracy of the older mechanical meters, and determine whether newer meters should be installed. Accurate metering will result in accurate billing, which could potentially lead to an increased revenue for water service providers.

Survey responses from Home H08 and H11 indicated regular garden irrigation, which was also identified by the classification results. Figure 8.1 shows the high percentage of the total consumption classified as outdoor use for these two homes. The classification results also showed noticeable outdoor water consumption at home H03, however, the survey results reported no garden irrigation at the property. WEAM could thus be utilised to identify homes with garden irrigation events at properties who reportedly have no outdoor use. The application of WEAM could potentially prove very useful during times when water restrictions are in place, especially if outdoor use is not permitted.

Previous studies suggest outdoor use to be seasonal, driven by weather-related variables such as rainfall (Fisher-Jeffes et al. 2015). Seasonal variability can also be noticed in Figure 8.1, with less water being consumed over the wet period (high rainfall period). Similar to the trends suggested by the theoretical estimates in Table 8.2, months with the highest rainfall over the study period (December, January) showed lower outdoor use, and months with lower rainfall over the study period (September, October) showed higher outdoor use.

Table 8.8. Comparison between billing data and smart meter data

Home Code		H01	H02	H03	H04	H05	H06	H07	H08	H10	H11
PPH		4	1	4	2	3	2	2	1	1	3
Municipal consumer meter (billing data)	Total water use over recording period (kL)	80	81	300	205	97.5	167	67	160	52	533
	Recording period (days)	329	294	329	329	329	329	329	298	329	329
	Water use per dwelling unit (L/du/day)	244	274	911	623	297	507	204	535	158	1 619
	Average per capita water use (L/c/d)	61	274	228	311	99	254	102	535	158	540
Smart meters	Total water use over recording period (kL)	28	35	157	66	26.3	80	51	157	92	240
	Recording period (days)	112	146	207	190	187	191	192	206	204	210
	Water use per dwelling unit (L/du/day)	247	241	760	348	141	418	268	761	452	1 144
	Average per capita water use (L/c/d)	62	241	190	174	47	209	134	761	452	381
Difference in average per capita water use (%)		1.0	12.1	16.6	44	52.6	17.5	31	42	187	29

The implementation of WEAM on the coarser data, as presented in this study, suggests that lower resolution end-used data have more benefits than is currently being explored. Although only major end-use events were analysed in this paper, the results presented provide valuable insight into the proportion of monthly water consumption used for indoor use and outdoor use at the study site. Based on the results obtained, and the robustness of PEET and WEAM to analyse coarse data sets, implementation of water demand measures can now be investigated in future research.

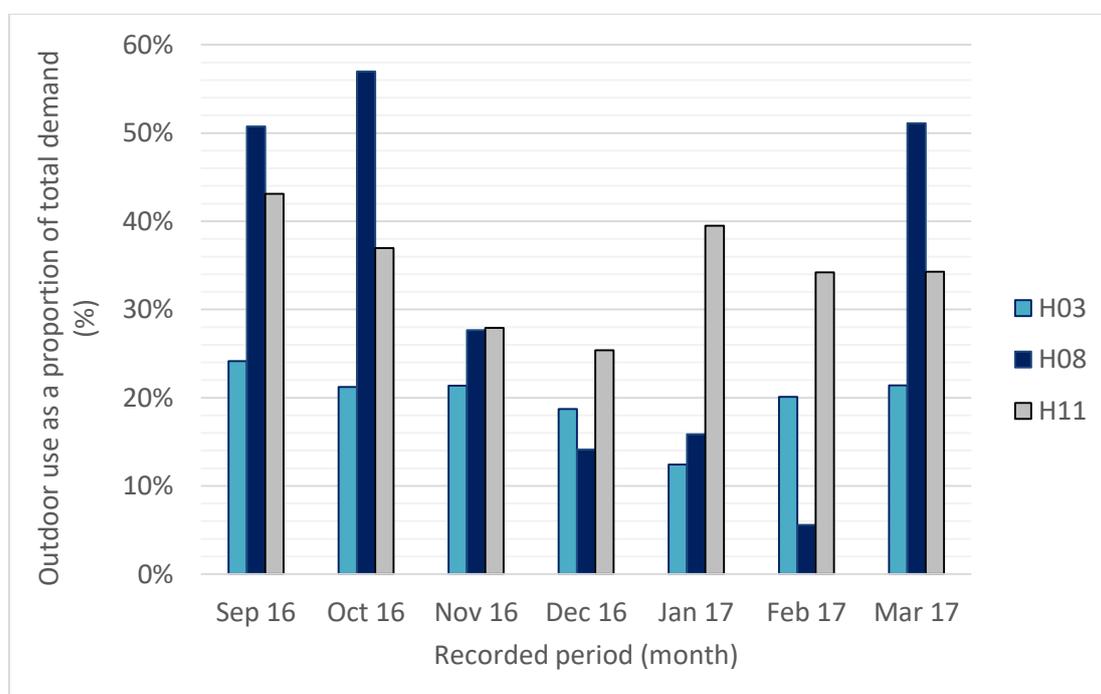


Figure 8.1. Monthly outdoor consumption at Homes H03, H08, H11

## CONCLUSION

Understanding household water demand at end-use level is important for effective WDM strategies. This paper presents a case study that was conducted in Johannesburg, South Africa. In the case study, household water demand was recorded with meter resolutions set to 1 L/pulse (rudimentary data), at 15 s frequencies. Specific objectives of this case study were to classify household water use events extracted from a rudimentary data set into indoor use and outdoor use, to better understand consumer consumption behaviour at the study site, and to compare theoretical consumption estimates with actual results. This study therefore addressed the problem of classifying indoor and outdoor water use events with limited and coarse end-use data. PEET (Meyer et al. 2020) was used to extract end-use events from a rudimentary data set while WEAM (Meyer et al. submitted) was utilised to classify the extracted end-uses into indoor or outdoor use. The results presented in this paper provide insight into the proportion of monthly water consumption used indoors and outdoors at the study site, expressed as a percentage of the total household water demand.

Outdoor use was identified at all 11 homes, even though some residents did not report any garden irrigation. Classification tools implemented in this case study could thus be useful to monitor whether homes adhere to water restrictions, especially if outdoor use is limited or prohibited. An average of 30% of the total water demand was classified as being outdoor use (neglecting unclassified events), and was seasonally driven, with higher outdoor consumption occurring over the dry months.

Although PEET was successful in extracting end-use events from a rudimentary data set, a large portion of the total water demand (24.2%) was not classified. Future research should investigate different meter resolutions (e.g. 0.5 L/pulse) to determine the optimal meter resolution to minimise the proportion of events classified as “unknown”. Future research could also assess the impact of implemented water demand management measures from low resolution household water use data sets, considering the valuable insights obtained using PEET and WEAM.

## ACKNOWLEDGEMENTS

The authors are grateful to Johannesburg Water for funding this project, contributing in part to the PhD bursary of Mrs Meyer.

**REFERENCES**

Alcocer-Yamanaka V., Tzatchkov V.I., Arreguin-Cortes F. 2012. Modelling of drinking water distribution networks using stochastic demand." *Water Resour. Manage., EWRA*. **26**(7): 1779-1792. doi: 10.1007/s11269-012-9979-2

Aquacraft 2010. Analysis of water use in new single family homes.

URL: <http://www.allianceforwaterefficiency.org/WorkArea/DownloadAsset.aspx?id=6020>

Beal C., Stewart R.A., Huang T. and Rey E. 2011. South East Queensland residential end use study. *J. Australian Water Association*. **38** (1): 80-84.

Beal C. and Stewart R.A. 2013. Identifying residential water end uses underpinning peak day and peak hour demand. *J. Water Resour. Plann. Manage., ASCE*, **136**: 19-26.  
doi:10.1061/(ASCE)WR.1943-5452.0000357

Blokker E.J.M. Vreeburg J.H.G. and van Dijk, J. C. 2010. Simulating Residential Water Demand with a Stochastic End-Use Model", *J. Water Resour. Plann. Manage.*, **136**: 19- 26.

Bradley R. 2004. Forecasting domestic water use in rapidly urbanizing areas in Asia. *J. Environ. Eng.* **130**(4): 465–471. doi:10.1061/(ASCE)0733-9372(2004)130:4(465)

Buchberger S.G. and Wu L. 1995. A model for instantaneous residential water demands. *J. Hydr. Engrg., ASCE*, **121**(3): 232-246. doi: 10.1061/(ASCE)0733-9429(1995)121:3(232)

Cardell-Oliver R., Wang J. and Gigney H. 2016 Smart meter analytics to pinpoint opportunities for reducing household water use. *J. Water Resour. Plann. Manage., ASCE*, **142**(6): 04016007.  
doi: 10.1061/(ASCE)WR.1943-5452.0000634

Cole G. and Stewart R.A. 2013. Smart meter enabled disaggregation of urban peak water demand: Precursor to effective urban water planning. *Urban Water Journal*, **10**(3): 174-194.  
doi: 10.1080/1573062X.2012.716446

Cominola A., Giuliani M., Castelletti A., Rosenberg D.E., Abdallah A.M. 2018. Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management. *Environmental, Modelling and Software* **102**(2018): 199-212.  
doi: 10.1016/j.envsoft.2017.11.022

Department of Water and Sanitation, DWS 2016. Integrated Vaal River System - Limiting the use of water in terms of Item 6 of Schedule 3 of The National Water Act of 1998 for Urban and Irrigation purposes in the catchment areas of the dams supplying The Integrated Vaal River System and from The System [Online],

<https://www.johannesburgwater.co.za/wp-content/uploads/2016/08/Gazetted-Level-3-Water-Restrictions-12-August-2016.pdf> (accessed 01 March 2017).

Fisher-Jeffes L., Gertse G. and Armitage N.P. 2015. Mitigating the impact of swimming pools on domestic water demand. *Water SA* **41**(2) 238-246. doi:10.4314/wsa.v41i2.9

Fox C., McIntosh B.S., Jeffry P. 2009. Classifying households for water demand forecasting using physical property characteristics. *Land use policy* **26**(2009): 558-568.  
doi: 10.1016/j.landusepol.2008.08.004

Gato S. 2006. Forecasting urban residential water demand. PhD thesis, RMIT University

Giurco D., Carrard N., McFallan S., Nalbantoglu M., Inman M., Thornton N., White S. 2008. Residential End use Measurement Guidebook: A Guide to Study Design, Sampling and Technology. Prepared by the Institute for Sustainable Futures, UTS and CSIRO for the Smart Water Fund, Victoria.

Heinrich M. 2007. Water end use and efficiency project (WEEP) – Final Report. BRANZ Study Report 159. Branz, Judgeford, New Zealand.

Hemati A., Rippy M., Grant S., Davis K. and Feldman D. 2016. Deconstructing demand: the anthropogenic and climatic drivers of urban water consumption. *Environ. Sci. Technol.*, **50** (23):12557-12566. doi:10.1021/acs.est.6b02938

Hussien W.A, Memon F.A., Savic D.A. 2016. Assessing and Modelling the Influence of Household Characteristics on Per Capita Water Consumption. *Water Resour. Manage.*, **30**(2016): 2931–2955. doi 10.1007/s11269-016-1314-x

Ilemobade A., Jacobs H.E. and Botha B. E. 2018. Sorting high resolution end-use data – A pleasurable headache. Proceedings. 1st International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada – July 23-25.

<https://ojs.library.queensu.ca/index.php/wdsa-ccw/article/view/12279> (accessed 24 October 2020).

Jacobs H.E. and Haarhoff J. 2004. Structure and data requirements of an end-use model or residential water demand and return flow. *Water SA* **30**(3): 293-304.  
doi:10.4314/wsa.v30i3.5077

Jacobs H.E., Ilemobade A.A. and Botha B.E. 2017. End-use estimates of water demand within certain residential categories in the city of Johannesburg. Computing and Control for the Water Industry (CCWI), Sheffield, 5th-7th September 2017. Journal contribution. doi:10.15131/shef.data.5364556.v1

Johannesburg Water, JW 2016. Water Restrictions for Johannesburg [Online], Available: <https://www.johannesburgwater.co.za/level-2-water-restrictions> (accessed 01 March 2017).

Kowalski M. and Marshallsay D. 2005. Using measured microcomponent data to model the impact of water conservation strategies on the diurnal consumption profile. *Water Science & Technology Water Supply* 5(3):145-150. doi: 10.2166/ws.2005.0094

Liu J., Savenije H. and Xu J., 2003. Forecast of water demand in Weinan City in China using WDF-ANN model. *Phys. Chem. Earth*, **28**: 219–224. doi: 10.1016/S1474-7065(03)00026-3

Loh M. and Coghlan P. 2003. Domestic water use study. Water Corporation, Perth. [https://www.water.wa.gov.au/\\_\\_data/assets/pdf\\_file/0016/5029/42338.pdf](https://www.water.wa.gov.au/__data/assets/pdf_file/0016/5029/42338.pdf)

Martinez-Espineira R. 2002. Residential water demand in the Northwest of Spain. *Environ. Resour. Econ.* **21**: 161–187. doi: 10.1023/A:1014547616408

Mayer P.W. and DeOreo W.B., 1999. Residential end uses of water. Aquacraft, Inc. Water Engineering and Management, Boulder, CO, USA.

Mead N. and Aravinthan V. 2009. Investigation of household water consumption using smart metering system. *Desalination Water Treat. J.* **11**(1-3): 115-123. doi: 10.5004/dwt.2009.850

Meyer B.E., Jacobs H.E., Ilemobade A.A. 2020. Extracting household water use event characteristics from rudimentary data. *J. Water Supply Res. Technol. AQUA.* **69**(4): 387–397. doi: 10.2166/aqua.2020.153

Meyer B.E., Nguyen K., Beal C.D., Jacobs H.E. and Buchberger S. (submitted). Household water end-use classification model for implementation on coarser data: toward improving the benefits of lower resolution end-use data sets. *Journal Water Resources Planning and Management*, WRENG-4842

Mutikanga H.E., Sharma S.K., Vairavamoorthy K. 2011. Investigating water meter performance in developing countries: A case study of Kampala, Uganda. *Water SA* **37**(4): 567-574. doi: 10.4314/wsa.v37i4.18

Nguyen K.A., Zhang H., Stewart A.R. 2013. Development of an intelligent model to categorise residential water end-use events. *J. Hydro-environ. Res.* **7**(2013): 182-201.  
doi:10.1016/j.jher.2013.02.004

Pretorius D., Crouch M.L., Jacobs H.E. 2019. Diurnal water use patterns for low-cost houses with indigent water allocation: a South African case study. *J. Water Sanitation Hyg. for Develop.* **9**(3): 513-521. doi: 10.2166/washdev.2019.165

Roberts P. 2005. Yarra Valley Water 2004 residential end use measurement study. Final report, June 2005. Yarra Valley Water. Melbourne, Australia.

Sahin O., Siems R.S., Stewart R.A., Porter M.G. 2016. Paradigm shift to enhanced water supply planning through augmented grids, scarcity pricing and adaptive factory water: a system dynamics approach. *Environ Model Softw* **75**(1):348-361. doi: 10.1016/j.envsoft.2014.05.018

Stewart R.A., Nguyen K., Beal C., Zhang H., Sahin O., Berton E., Vieira A.S., Castelletti A., Cominola A., Giulianie M., Giurco D., Blumenstein M., Turner A., Liu A., Kenway S., Savić D.A., Makropoulosk C., Kossieris P. 2018. Integrated Intelligent Water-Energy Metering Systems and Informatics: Visioning a Digital Multi-Utility Service Provider. *Environmental Modelling and Software* **105**(2018): 94–117. *Environmental Modelling and Software*.  
doi: 10.1016/j.envsoft.2018.03.006

Vörosmary C.J., Douglas, E.M., Green, P.A. and Revenga, C. 2005. Geospatial indicators of emerging water stress: an application to Africa. *Ambio*, **34**: 230–6.  
doi: 10.1579/0044-7447-34.3.230

Willis R.M., Stewart R.A., Panuwatwanich K., Capati B., Giurco D. 2009. Gold Coast domestic water end use study. *Journal of Australian Water Association*, **36**(6): 79-85.  
doi: 10.1016/j.jclepro.2011.08.006

Willis R.M., Stewart R.A., Panuwatwanich K., Williams P.R. and Hollingsworth A.L. 2011. Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. *J. Environ. Manage.* **92**(8): 1996-2009.  
doi: 10.1016/j.jenvman.2011.03.023

## Chapter 9.

### Discussion

#### 9.1 DEFINING DATA RESOLUTION

Household water demand is typically measured by water meters. Meter readings could be time-based or event-based. The resolution of water meter data can thus be viewed in two dimensions, namely temporal or volumetric. Time-based water meters record flow volume at different temporal scales, i.e. fixed intervals. These fixed intervals can be as frequent as 1 s (Kowalski and Marshallsay 2003), 5 s (Beal and Stewart 2011), 10 s (Mead and Aravinthan 2009), where end-use disaggregation is possible, or as infrequent as monthly, quarterly, or even yearly (Nguyen et al. 2013), where the only information gained is total consumption. Water demand studies have also been conducted at recording frequencies between these two extremes, at 15 min (Pretorius et al. 2019) or 1 h (Cardell-Oliver et al. 2016), reporting on anomalous events, such as peak hour demand.

Terminology used to distinguish between different meter resolutions is often ambiguous. Cominola et al. (2018) evaluated the trade-off between temporal resolution of water meters, and the level of information that can be gained from the data. Cominola et al. (2018) deemed sub-minute metering resolutions as high resolution data, because at this level of accuracy, end-uses can be disaggregated from the data set. The term 'high resolution data' has also been used freely in end-use studies disaggregating end-use events, recording flow volumes at a sub-minute temporal scale. Temporal data resolutions, ranging from 1 min to 1 hour, were referred to by Cominola et al. (2018) as intermediate metering, coarser resolutions, and medium resolution data.

There has been no study evaluating the trade-off between volumetric resolution and level of data – in terms of household demand at end-use level. The volumetric resolution of the 'high resolution' end-use studies, which were successful in disaggregating end-uses, ranged from 0.014 L/pulse (Beal and Stewart 2011) to 0.1 L/pulse (Pastor-Jabloyes et al. 2018). Typical residential water meters used in South Africa were found to have a volumetric resolution of 1.0 L/pulse or 0.5 L/pulse. This research study proposed a distinction between data resolution where end-use disaggregation is possible, and data resolutions where disaggregation is not possible, but still provide some level of information at a household level. Data obtained from meter readings with pulse volumes larger than 0.1 L/pulse (which was the case for this research), which is still able to provide information on end-use level (although it is not end-use disaggregation) had to be distinguished from data resolutions that are not able to provide any information about end-use consumer behaviours.

Figure 9.1 depicts the 4 distinct categories suggested for future application, in terms of data resolution:

1. High resolution data: end-use disaggregation is possible.
2. Rudimentary data (focus of this study): possible to extract some useful information regarding water use events (e.g. indoor versus outdoor classification), however, end-use disaggregation is not yet possible. Further investigation is needed to determine what end-uses, if any, can be identified with this data resolution. The term “coarser resolution data” can also be used to describe this type of data resolution.
3. Intermediate data: household information at end-use level is not possible. Household diurnal patterns can be developed, and provide information on peak hour or peak day flow rate.
4. Low resolution data: unable to extract information other than total consumption over the specified recording period.

Figure 9.1 was developed specifically for a single measurement point at a residential property with an accessible water meter. The realm of unmetered homes fall beyond the scope of this graph.

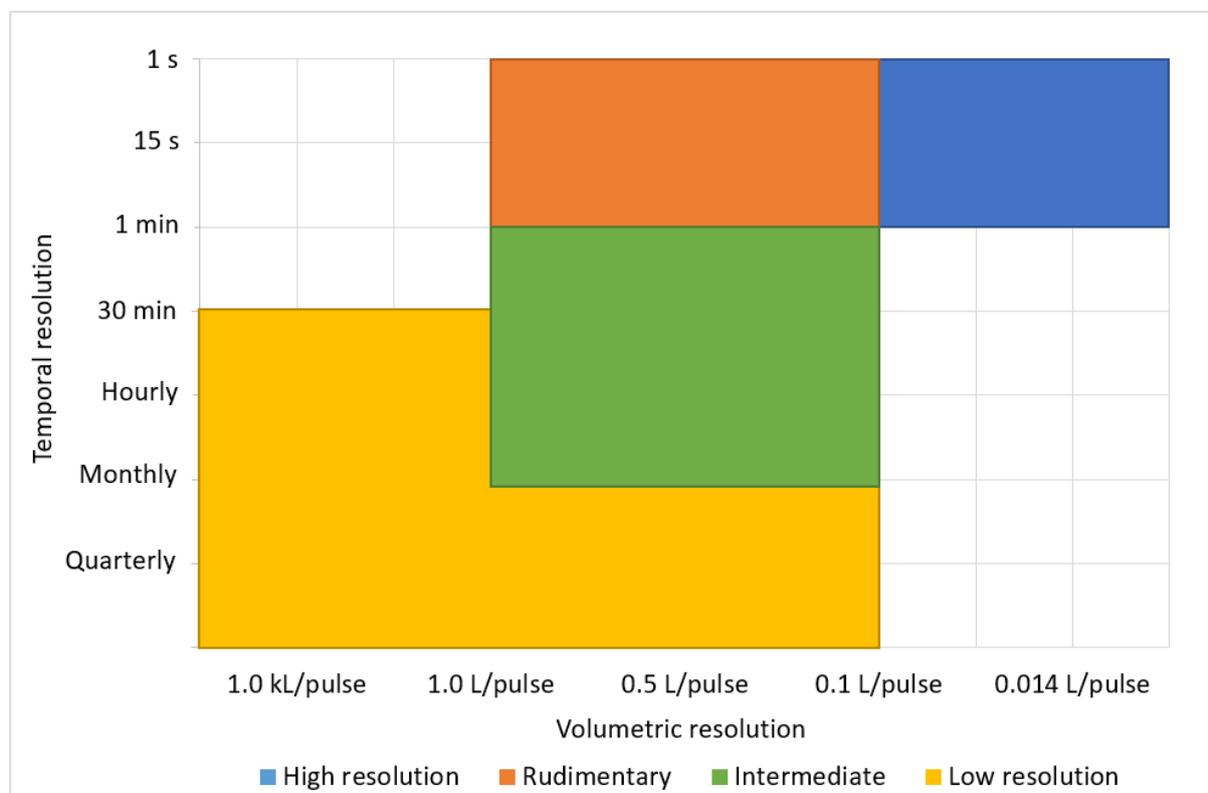


Figure 9.1. Proposed data resolution categories

High resolution data was thus referred to as data that are accurate enough to disaggregate end-uses. Rudimentary resolution data referred to data that are too coarse for end-use disaggregation, but could still provide valuable information at end-use level. Medium or intermediate resolution data referred to data that provide information on household consumption behaviour (such as peak hour demand), but were unable to provide information on household end-uses. High resolution data and rudimentary data thus had the same range of temporal resolution data (sub-minute), but differed in the volumetric resolution. Rudimentary data and medium or intermediate data had the same volumetric resolution, but differed in temporal resolution.

## **9.2 TYPICAL SITUATION OF LIMITED WATER END-USE DATA**

Residential water demand models are often used to assist water managers with making important decisions regarding WDM strategies. Understanding residential water consumption behaviour could also help establish new water use benchmarks. Where it is not practical to record actual household water demand, demand models are used. In most cases, end-use demand models are calibrated using stochastic estimates, historic billing data, survey responses, or appliance standards obtained from the manufacturers. Previous studies have reported on inaccuracies when relying on manufacturing standards and survey responses, due to residents being biased when reporting on their own water use (Mead and Aravinthan 2009). Stochastic estimates rely on a significant amount of input parameters to populate the model. The parameters also require information about end-use characteristics. Limited knowledge regarding end-use characteristics results in stochastic models relying largely on surveys and manufacturer standards, thus limiting the accuracy of a stochastic demand model.

One of the most accurate methods to determine household water demand is through actual measurements. Water consumption can be recorded at the point of use, or at a single point on the property (entry point). Utilities record household water use at the entry point of a house, using mechanical water meters. These water meters are typically programmed to record water flow at 1 L/pulse and measurements are manually taken every hour, month, quarter or half-year. Due to the low resolution of data obtained from mechanical meters, the information is typically used to determine monthly or yearly consumption, peak hour and peak day demand. Mechanical meters have, to this day, not been able to give any information regarding household water consumption at an end-use level.

In order to better understand household water consumption at end-use level, various previous studies recorded residential water use with smart meters, taking readings every 10 s at 0.014 L/pulse resolution. Commercially available flow trace analysis software could then be used to disaggregate the high resolution data into end-use components. These types of high resolution measurements are very uncommon, especially in developing countries.

Compared to the mechanical meters, smart meters are more expensive, require large data storage capacity, and need more resources (personnel) to analyse the data. Consequently, it is not practical for utilities, especially in developing countries, to implement smart meters at household level on a large scale. This research thus set out to address the issue of limited information regarding household water end-uses for utilities who only have access to coarser end-use data.

### 9.3 DIRECT AND INDIRECT MEASUREMENT APPROACHES

Various end-use studies were conducted as part of this research to gain information regarding specifics and characteristics of water end-uses. Multiple measuring approaches were considered and used, including direct and indirect flow sensing approaches. Table 9.1(a) and Table 9.1(b) summarise the direct and indirect methods implemented, respectively. The tables also specify whether the measurements were taken at a single entry point, or at the point of use. The data resolutions, advantages and disadvantages are also depicted in the tables, including general comments.

Table 9.1(a). Indirect measurement methods implemented in this study

Consumer Surveys	
Implementation	Chapters 2, 6, 8
Gathered information	Any information (within ethical constraints)
Targeted end-use(s)	All end-uses
Analysed information	Duration, frequency, time of use
Data resolution	N/A
Advantages	Flexibility, relatively simple to implement
Disadvantages	Lower accuracy, ethical restrictions, post-processing of data required
Comment	Surveys are good for end-use studies if combined with other measuring methods. Surveys should not be the sole method used to gather end-use information.
Temperature recorders (iButtons)	
Implementation	Chapters 2, 3
Gathered information	Time stamp, temperature
Targeted end-use(s)	Garden irrigation, shower
Analysed information	Duration, frequency, time of use
Data resolution	Temperature measurements at 1 min and 2 min intervals
Measurement location	Point of use
Advantages	Non-intrusive, relatively low cost, no plumbing changes needed, small, rugged, ground truth data
Disadvantages	Post-processing of data required
Comment	Only applicable on long duration events or events that use hot water. Not recommended for large scale studies.

Table 9.1(b). Direct measurement methods implemented in this study

<b>Mechanical water meter without data logger</b>	
Implementation	Chapters 4
Gathered information	Consumption, meter reading data
Targeted end-use(s)	Washing machine
Analysed information	Volume
Data resolution	Flow measurements per event; 1 L/pulse
Advantages	Accuracy, ground truth data
Measurement location	Point of use
Disadvantages	Manual readings, plumbing changes needed, on site personnel needed
Comment	Measurements were manually taken after every event. Not practical for large scale studies.
<b>Smart water meter with data logger (high resolution)</b>	
Implementation	Chapter 7
Gathered information	Consumption, meter reading data
Targeted end-use(s)	All end-uses
Analysed information	Duration, intensity, volume, frequency, time of use
Data resolution	Flow measurements every 5 s; 0.014 L/pulse
Measurement location	Single point at residential property
Advantages	Accuracy, automated readings
Disadvantages	High cost, plumbing changes needed
Comment	Measurements were not taken during this study, instead, the data recorded by the high resolution smart meters were received from an external source, and the data was used, in part, to develop WEAM.
<b>Smart water meter with data logger (lower resolution)</b>	
Implementation	Chapters 6, 8
Gathered information	Consumption, meter reading data
Targeted end-use(s)	All end-uses
Analysed information	Duration, intensity, volume, frequency, time of use
Data resolution	Flow measurements every 15 s; 1 L/pulse
Measurement location	Single point at residential property
Advantages	Accuracy, automated readings
Disadvantages	Plumbing changes needed
Comment	The resolution was purposefully set to 1 L/pulse, in line with common utility meter resolutions. A sub-minute recording interval was selected for end-use disaggregation.

#### **9.4 GARDEN IRRIGATION AS OUTDOOR END-USE**

Residential water consumption is primarily categorised into indoor use and outdoor use. Garden irrigation is the largest contributor to outdoor water use if a garden is present and is often the first end-use being restricted during drought periods. As a result, consumers may turn to alternative non-potable water sources, such as groundwater, during stringent restrictions. In most countries, South Africa included, groundwater abstraction at residential properties is not monitored or measured. Monitoring potable alternative resources, such as groundwater, is important for long term water security. Meyer and Jacobs (2019) measured actual groundwater use by private homeowner, and is the first end-use study conducted in South Africa reporting on point of use measurements at residential GAPs.

Temperature loggers were placed on the outflow pipes of groundwater abstraction points (GAPs), recording the change in temperature when water passed through the pipe (indicating an event). Analysis of the time series data presented valuable information regarding garden irrigation event duration and irrigation frequency at the study site. Monte Carlo analysis was deployed on the event characteristics (frequency, duration and intensity) to develop statistical distributions for each. A subsequent prediction model (DFI model) was developed to model groundwater garden irrigation in an unrestricted scenario. A sensitivity analysis suggested duration is the event characteristics with the largest effect on garden irrigation consumption.

It would be expected that residents with GAPs would irrigate more regularly and for longer durations compared to residents irrigating their gardens using water from the potable water distribution system. Therefore, the event characteristic results are not representative of a larger region, or consumers beyond the study site. However, the DFI model presented by Meyer and Jacobs (2019) is scalable over different regions and time periods, since the parameters of the distribution curves can be populated with site specific values. Utilising temperature loggers as indirect method for measuring water usage at privately owned GAPs proved useful. Following the same method proposed by Massuel et al. (2009), Meyer and Jacobs (2019) showed that data recorded with temperature loggers can successfully be used to obtain valuable information regarding garden irrigation events in an unrestricted scenario (water supplied by GAPs).

#### **9.5 SHOWER AND WASHING MACHINE END-USE CHARACTERISTICS**

Showers and washing machines were identified as the largest contributors to indoor water consumption. Botha et al. (2017, 2018) thus conducted end-use studies to gather information regarding the event characteristics of these end-uses. End-use measurements were taken at the point of use, using an indirect flow sensing approach (temperature loggers) and a direct flow sensing approach (mechanical water meters). The temperature loggers were placed on the shower heads, measuring the temperature variation over time.

A substantial increase in temperature indicated a shower use event. The temperature threshold indicating the start and end of a shower event was determined empirically. The temperature loggers could successfully identify the start and finish times of showers, thus providing valuable information regarding shower event identification, and event durations. The application of temperature loggers can thus be applied to hot water end-uses where temperature variation is expected to be relatively large. Although the type of shower head influenced the shower event volumes, a sensitivity analysis showed that duration contributed most to the total shower event volume.

Washing machine event volumes were manually recorded using mechanical water meters installed at the point of use, which is one of the most accurate means of recording water consumption. Manufacturer standards may not align with the true water use of washing machines. Chapter 4 reported that replacing top loaders with front loaders may hold water savings potential at the study site. Similar statements were made by Mead and Aravinthan (2009) and Roberts (2005). In order to confirm the water savings potential from changing appliance type, information regarding the load size is needed, and little data is available concerning the actual load size per wash cycle (Pakula and Stamminger 2009).

Both papers presented in Chapter 3 (Botha et al. 2017) and Chapter 4 (Botha et al. 2018) contribute to the understanding of water demand at individual end-use scale. Future work could extend the recorded data from these studies by generating additional synthetic water demand time series data, as described by Alvisi et al. (2014).

## **9.6 TOOL DEVELOPMENT FOR END-USE CHARACTERISTICS EXTRACTION**

The percentage of total household water consumption contributed to outdoor use varies significantly over regions, seasons and property types. Measuring consumption at the point of entry can give insight into household water use. A sub-minute logging frequency is needed to disaggregate end-use events and understand water use at an end-use level. Previous end-use studies recorded water consumption data at high resolutions (0.014 L/pulse) and short frequencies (1 s – 10 s) and were able to classify end-uses from the time series data. This high resolution setting for meters are uncommon, and utilities typically measure water consumption at coarser resolutions.

Although some studies, recording flow measurements every 15 minutes or 1 hour, have given valuable insights into anomalous events such as leaks and peak day demands, they have failed to provide information on household end-use events. To date, no study has reported on the metering resolution (L/pulse) needed for end-use disaggregation. The extent to which measured rudimentary data can be used to obtain water end-use demand information at a household level was thus explored.

As part of a first study to classify end-use events from a rudimentary data set, smart meters were installed at a study site in Johannesburg, measuring water flow every 15 s at 1 L/pulse intervals. The resolution was purposefully set to 1 L/pulse, in line with common utility meter resolutions and the sub-minute recording interval was selected based on the requirements for end-use disaggregation. Before end-use events can be classified, they first have to be extracted from a time series. No software was commercially available to extract end-use events from the recorded data set, due to the volume resolution of 1 L/pulse being too coarse. Consequently, an extraction tool was developed as part of this study. Following a similar logical approach incorporated in flow trace analysis software to extract end-use events, PEET, an automated end-use extraction tool developed by Meyer et al. (2020), was able to extract notable end-use events from a rudimentary data set. Due to the rudimentary nature of the data, low flow events and background leaks had similar event characteristics and PEET was unable to distinguish between them. Subsequently, PEET grouped and classified all these low flow events as minor events, with the intention that only notable events should be further analysed to obtain household end-use information. It is important to note that although the extraction tool was automated, the pre-processing of the data (cleaning) was time consuming, and should be taken into account when planning for end-use studies.

## **9.7 CLASSIFICATION MODEL DEVELOPMENT AND VALIDATION**

A novel apportionment model was developed to classify the extracted end-use events into indoor use and outdoor use. Three methods were evaluated to determine the best model to correctly classify indoor events and outdoor events from coarser end-use data, while minimising the classification error. Even though the model must be applicable on rudimentary data, high resolution data with known end-use events were used to train, calibrate, test and validate the model. The best performing model was a random forest machine learning algorithm, categorising end-use events as being either indoor or outdoor; with the three input parameters being the three event characteristics, namely event duration, event volume and event intensity. WEAM can thus be employed on the end-use events extracted from PEET, in order to provide insight into household water demand at end-use level.

The novel classification model proposed by Meyer et al. (submitted) was able to correctly classify between 60.7% and 96.2% of notable end-use events. Despite not being able to identify individual end-use components, WEAM is the first classification model that can be employed on coarser data in order to improve the benefits of lower resolution data sets. It is hoped that by applying this novel method on coarser data sets, water utilities from a range of socio-economic settings can have greater opportunities to improve water security through better informed demand management programmes.

## 9.8 MODEL APPLICATION

In order to evaluate the application of PEET and WEAM on measured data, a case study was conducted in Johannesburg, South Africa. This study addressed the problem of classifying indoor and outdoor water use events with limited information. As part of the case study, surveys were administered to the residents to gain insight into the perceived notion of their water consumption. Roughly half of the study sample completed the surveys, of which 68% opted to stay anonymous. Thus only 17% of the administered surveys provided useful information. The information from the surveys were, however, limited, as some surveys were only partially completed. Measured data are thus vital and important to understand actual household water use.

Smart meters were installed at 63 residential properties logging water flow every 15 seconds at a resolution of 1 L/pulse. The recorded time series data were considered rudimentary, due to the 1 L/pulse meter resolution. Prior to data analysis, 9 homes were removed from the study sample due to poor data quality. Notable household end-use events were extracted from the rudimentary time series data using PEET. Minor end-uses (most likely indoor events), were discarded from the data set prior to classification, since PEET was only able to extract notable end-uses. About 24% of the total volume of all events was attributed to minor events. WEAM was subsequently deployed and categorised each extracted event based on its physical characteristics, namely duration, volume, and intensity. Outdoor use and indoor use as a proportion of the total household consumption were reported on. The results were compared to theoretical estimates, as well as survey responses.

Residents indicating regular garden irrigation were noticed in the classification results and had a large percentage of their total consumption classified as outdoor use. However, outdoor demand was also noticed at homes which did not indicate any outdoor use. It is apparent that WDM strategies or water restrictions targeting outdoor use need to be monitored for effective implementation. Although only notable end-uses were apportioned, the results presented as part of this research provide valuable insight into the proportion of monthly water consumption used for indoor use and outdoor use at the study site.

## Chapter 10.

### Conclusion

#### 10.1 SUMMARY OF FINDINGS

Water utilities often have limited information regarding household water demand at end-use level, and typically only have access to rudimentary data. This research was conducted to determine the extent to which measured rudimentary data can be used to obtain water end-use demand information at a household level. The aim of the research was to identify and develop methods to evaluate and quantify household water demand at an end-use level, in the absence of high resolution data. This research paper investigated an indirect flow sensing approach at the point of use, measured actual end-use events, proposed a classification model, and applied the classification model in a case study, all in order to better understand household water consumption at end-use level.

Temperature loggers were investigated as an indirect flow measurement approach, measuring temperature variation at the point of use. The use of temperature loggers to measure temperature variation in a water pipe (indicating when water is flowing through the pipe) was not new, however, the application thereof at household level was novel. Temperature loggers were successfully deployed on groundwater extraction points (GAPs) and shower heads, and provided value information regarding irrigation and shower event characteristics. Chapter 2 (Meyer and Jacobs 2019) and Chapter 3 (Botha et al. 2017) suggest temperature loggers are able to identify the start and finish times of long duration events (i.e. garden irrigation) and household events that use hot water (i.e. shower events). Temperature loggers are thus limited in its applicability to be universally used at all end-use events at a home.

Significant efforts have been made in recent years to lower household water demand, including initiatives to educate communities on water security and the implementation of water wise appliances. As a result, household water consumption varies between regions, and over different time periods, and is highly dependent on the end-use consumption behaviour of consumers. Therefore, understanding household water use at the present time can be very important, because relying on historic data can possibly lead to inaccuracies in demand predictions. Previous studies identified the shower and washing machine as the largest contributors to household indoor demand, and suggests garden irrigation contributes most to outdoor water demand (if a garden is present). Subsequently, end-use studies were conducted, measuring event characteristics at the point of use, reporting on the ground truth consumption of these end-use events.

Chapter 2 (Meyer and Jacobs 2019), Chapter 3 (Botha et al. 2017) and Chapter 4 (Botha et al. 2018) reported on garden irrigation's, shower's and washing machine's event durations, frequencies, intensities, and volumes. Statistical distributions for event duration, event intensity, event frequency and event volume was proposed for garden irrigation, shower and washing machine household end-uses. These end-use event characteristics can be used to calibrate end-use demand models with suitable parameters, which could lead to more accurate demand predictions in the respective study areas.

The classification model, WEAM (Meyer et al. submitted), was proposed to better understand household water consumption at end-use level. WEAM is able to classify end-use events once the individual end-use events are extracted from a time series. At the time of this study, no commercially available extraction tools was applicable on rudimentary data sets. Consequently, an end-use extraction tool, PEET (Meyer et al. 2020), was developed to extract end-use event characteristics from a rudimentary data set. Due to the coarser data, PEET was not able to identify a difference in the event characteristics between some low flow events and minor leaks, and consequently grouped these events together and categorised them as minor events. All other events extracted were considered notable end-use events. WEAM was developed for application on rudimentary data, and was successful in categorising the extracted events as being indoor or outdoor, by evaluating the correlation between the physical event characteristics, namely duration, intensity and volume. PEET demonstrated that end-uses can be extracted from coarser data sets, and the novel model presented by Meyer et al. (submitted) showed that the extracted end-uses can successfully be classified as being either indoor water use or outdoor water use. Although these models are not able to identify specific end-uses, valuable information can now be extracted from utility water meters, measuring at volumetric resolutions of 1 L/pulse.

## **10.2 NOVEL CONTRIBUTIONS**

This research study demonstrated that rudimentary data can provide useful information regarding household indoor and outdoor water consumption. Temperature loggers coupled with temperature variation analysis can successfully be deployed at household level to determine the start and finish times of events (thus quantify event duration), as well as determine the frequency of use. Although the implementation of temperature loggers at GAPS was not new, the application thereof at household boreholes and well-points, and extended application thereof on hot water indoor end-uses (such as the shower) has been demonstrated. This research also found that the implementation of temperature loggers are limited to end-uses that use hot water, or end-uses that have long durations.

Notable end-use studies were conducted, measuring water consumption at the point of use. These studies were the first end-use studies conducted in Africa recording shower, washing machine and garden irrigation events (from GAPS) at the point of use. The studies reported on the physical characteristics of each of the respective end-uses, contributing to

understanding South African water end-uses. Statistical distributions for event duration, intensity, and frequency were also developed for garden irrigation events, shower events and washing machine events. Water demand models, especially for application in South Africa where the end-use studies were conducted, can be populated with the actual consumption data for more accurate predictions.

Comprehensive international end-use studies have been conducted in the past analysing high resolution data and reporting on end-use characteristics. The resources needed for such studies typically limit the application to developed countries with access to high resolution data. No commercially available tool existed to extract and classify water end-use events from more commonly available, coarser resolution water meters. An end-use extraction model, PEET, was thus developed to extract end-uses and their physical characteristics from rudimentary data. PEET can be universally deployed in developing countries and in regions with limited access to high resolution data.

In order to classify the extracted end-uses, a novel classification model, WEAM, was proposed. WEAM was developed for application on coarser resolution data. WEAM classifies an end-use event based on three event characteristics, namely duration, volume and intensity. Although WEAM was not able to classify events into specific end-uses, the model was able to successfully apportion end-use events into indoor use and outdoor use. The novel method presented in this paper now allows useful information to be extracted from relatively coarser end-use data sets – with specific reference to the classification of the water use events as being either indoor or outdoor. PEET and WEAM uses data that are readily available (coarser resolution data), which enables utilities to easily replicate these methods. The methods and models developed as part of this research could be applied by water utilities to develop informed WDM strategies, ultimately contributing to water security.

### **10.3 FUTURE RESEARCH**

Due to the high variability of water consumer behaviours around the world, more end-use studies, measuring water consumption at the point of use, should be conducted to better understand household end-use behaviour in the present time. With the rapid growth rate of technology, reliable flow metering devices could become cheaper in the near future, and should be considered as viable options for end-use studies. The results from such studies can also be used to calibrate current end-use demand models, to ultimately contribute to more accurate demand predictions. The DFI model presented in Chapter 2 could be expanded in the future to incorporate seasonal variability, different irrigation methods and also other types of supplementary household water supply, such as rainwater and greywater.

A change of washing machine appliance could hold water savings potential. Numerous studies have reported on the strong correlation between the number of clothes washing loads per week and PPH, however, little is known about the load size. Because load size plays a big part in whether washing machines are over loaded or under loaded, future research could investigate the trade-off between washing load size, type of appliance (front or top loader) and PPH, to better understand water savings potential. As a water savings initiative, consumers should also be encouraged to use their washing machines to its full capacity.

Previous studies reported that sub-minute metering frequencies are required for end-use disaggregation. Future research could investigate at what volumetric resolution end-use event can be extracted and classified. Chapter 5 briefly commented on the trade-off between the meter pulse volume and the extent of information gained from the metered data. Future research should explore this trade-off in more depth. A more comprehensive analysis can give insight into the lowest resolution of data required for end-use identification. Such findings could contribute to improved versions of PEET and WEAM.

Although WEAM showed good model performance, future research is required before the model can be used commercially. Future research could benefit by adding additional training parameters to the data set, such as event start time, day of the week, season, and so forth, to improve the model performance. All these proposed parameters can be extracted using PEET. Additionally, further research could explore the generalisation of WEAM by incorporating socio-economic and climatological variables.

## References

- Alvisi S., Ansaloni N. and Franchini M. 2014. Generation of synthetic water demand time series at different temporal and spatial aggregation levels. *Urban Water Journal*, **11**(4): 297-310. doi: 10.1080/1573062X.2013.801499
- Cahill R., Lund J.R., DeOreo B and Medellín-Azuara J. 2013. Household water use and conservation models using Monte Carlo techniques. *Hydrology and Earth System Sciences*, **17**(10): 3957-3967. doi: 10.5194/hess-17-3957-2013
- Cloete A.H. 2016 A domestic electrical water heater application for smart grid. M.Eng. thesis, Stellenbosch, South Africa: Stellenbosch University.
- Cody R.A., Tolson B.A. and Orchard J. 2020. Detecting Leaks in Water Distribution Pipes Using a Deep Autoencoder and Hydroacoustic Spectrograms. *J. Comput. Civ. Eng.*, 2020, **34**(2): 04020001 doi: 10.1061/(ASCE)CP.1943-5487.0000881
- Evans R.P., Blotter J.D. and Stephens A.G. 2004. Flow rate measurements using flow induced pipe vibration. *J. Fluids Eng.* **126**(2): 280–285. doi:10.1115/1.1667882
- Foekema H. and Engelsma O. 2001. A changing consumption pattern (water use at home 2001), TNS NIPO, Amsterdam (in Dutch).
- Fogarty J., Au C. and Hudson S.E. 2006. Sensing from the Basement: A Feasibility Study of Unobtrusive and Low-Cost Home Activity Recognition. In: *Proc. of UIST 2006*: 91–100.
- Froehlich J.E., Larson E., Campbell T., Haggerty C., Fogarty J. and Patel S.N. 2009. HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. *Proc., 11th Int. Conf. on Ubiquitous Computing, Orlando, FL*: 235–244.
- Froehlich J.E., Larson E., Saba E., Campbell T., Atlas L., Fogarty J. and Patel S.N. 2011. A longitudinal study of pressure sensing to infer real-world water usage events in the home. *Pervasive computing*, K. Lyons, J. Hightower, and E. Huang, eds., Springer, Berlin: 50–69.
- Garcia X., Llausas A. and Ribas A. 2014. Landscaping patterns and socio demographic profiles in suburban areas: Implications for water conservation along the Mediterranean coast. *Urban Water Journal*, **11**(1): 31-41. doi: 10.1080/1573062X.2012.758296
- Grafton R.Q., Ward M.B., To H. and Kompas T. 2011. Determinants of residential water consumption: Evidence and analysis from a 10-country household survey. *Water Resources Research*, Volume 47: W08537. doi.org/10.1029/2010WR009685

Gupta A., Mishra S., Bokde N.D. and Kulat K.D. 2016. Need of smart water systems in India. *International Journal of Applied Engineering Research*, **11**(4): 2216-2223.

Hand M., Shove E. and Southerton D. 2005. Explaining Showering: a Discussion of the Material, Conventional, and Temporal Dimensions of Practice. *Sociological Research Online*, **10**(2). doi: 10.5153/sro.1100

Hauber-Davis G. and Idris E. 2006. Smart Water Metering. *Water: Journal of the Australian Water Association*, **33**(3): 56-59.

Hendron R. and Burch J. 2008. Development of Standardized Domestic Hot Water Event Schedules for Residential Buildings. National Renewable Energy Laboratory. Energy Sustainability Conference, Long Beach, California, USA, 27-30 June, 2007.

Janik H. and Kupiec A. 2007. Trends in Modern Car Washing. *Polish Journal of Environmental Studies*, **16**(6): 927-931. <http://www.pjoes.com/Keyword-in+car+washing/163300>

Kim Y., Schmid T., Charbiwala Z.M., Friedman J. and Srivastava M.B. 2008. NAWMS: Non-intrusive autonomous water monitoring system. In *Proceedings of the 6<sup>th</sup> ACM Conference on Embedded Network Sensor Systems*, Raleigh, NC, USA, 4-7 November 2008: 309-322.

Kowalski M. and Marshallsay D. 2005. Using measured microcomponent data to model the impact of water conservation strategies on the diurnal consumption profile. *Water Science & Technology Water Supply* **5**(3): 145-150. doi: 10.2166/ws.2005.0094

Larson E., Froehlich J., Campbell T., Haggerty C., Atlas L., Fogarty J. and Patel S.N. 2012. Disaggregated water sensing from a single, pressure-based sensor: An extended analysis of HydroSense using staged experiments. *Pervasive and Mobile Computing*, **8**(1): 82-102. doi: 10.1016/j.pmcj.2010.08.008.

Liu A., Giurco D., Mukheibir P. 2016. Urban water conservation through customised water and end-use information. *Journal of Cleaner Production*, **112**(4): 3164-3175. doi: 10.1016/j.jclepro.2015.10.002

Nel P.J.C., Booysen M.J., Van der Merwe A.B. 2015a. Electric Water Heater Energy Consumption Determination Using Outlet Temperature and Volumetric Estimation. IEEE CIASG, December 2015, Cape Town, South Africa.

Nel P.J.C., Booysen M.J., Van der Merwe A.B. 2015b. Using thermal transients at the outlet of electrical water heaters to recognise consumption patterns for heating schedule optimisation. IEEE NTMS, July 2015, Paris, France.

Pakula C. and Stamminger R. 2010. Electricity and water consumption for laundry washing by washing machine worldwide. *Energy Efficiency* 3: 365-382. doi: 10.1007/s12053-009-9072-8

Pirow N.O., Louw T.M., and Booysen M.J. 2018. Non-Invasive Estimation of Domestic Hot Water Usage with Temperature and Vibration Sensors. *Flow Measurement and Instrumentation*, Volume 63: 1-7.

Rosenberg D.E. 2007. Probabilistic estimation of water conservation effectiveness. *J. Water Resour. Plann. Manage.*, **133**(1): 39–49. doi: 10.1061/(ASCE)0733-9496(2007)133:1(39)

Paulsen R.J., Smith C.F., O'Rourke D. and Wong T. 2001. Development and Evaluation of an Ultrasonic Ground Water Seepage Meter. *Groundwater* 39: 904–911. doi: 10.1111/j.1745-6584.2001.tb02478.x

Smith D. and Shilley H. 2009. Residential Car Wash Monitoring Study, Federal Way: City of Federal Way: Public Works Department, Surface Water Management Division.

Srinivasan V., Stankovic J., and Whitehouse K. 2011. WaterSense: water flow disaggregation using motion sensors. In *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. Association for Computing Machinery, New York, NY, USA, 19–24. doi:<https://doi.org/10.1145/2434020.2434026>

Van Zyl H.J., Haarhoff J. and Husselmann M.L. 2003. Potential application of end-use demand modelling in South Africa. *Journal of the South African Institution of Civil Engineering*, 45(2): 9-19. Available from: [http://www.scielo.org.za/scielo.php?script=sci\\_arttext&pid=S1021-20192014000100007&lng=en&nrm=iso](http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S1021-20192014000100007&lng=en&nrm=iso). ISSN 2309-8775.

Wong L.T. and Mui K.W. 2007. Stochastic modelling of water demand by domestic washrooms in residential tower blocks. *Water and Environment Journal*, **22**(2): 125-130. doi: 10.1111/j.1747-6593.2007.00087.x

Young H.D., Freedman R.A., Ford A.L. and Sears F.W. 2012. *Sears and Zemansky's University Physics: with Modern Physics*. San Francisco, Pearson Addison Wesley. ISBN-10: 0201603365

## Appendix A: Declarations of candidate and co-authors

### Declaration by the candidate (Chapter 2)

The nature and scope of my contribution to Chapter 2 of this thesis, the published paper "Garden irrigation as household end-use in the presence of supplementary groundwater supply", were as follows:

Nature of contribution	Extent of contribution
- Conceptualised and wrote the paper	95%
- Review of literature	
- Data collection	
- Carried out data and statistical analysis	

The following co-authors contributed to Chapter 2 of this thesis, "Garden irrigation as household end-use in the presence of supplementary groundwater supply":

Name	e-mail address	Nature of contribution	Extent of contribution
H. E. Jacobs	hejacobs@sun.ac.za	Assisted with editing and of the paper	5%

Signature of candidate: .....  
Date: 7 April 2020.....

### Declaration by co-authors:

The undersigned hereby confirm that

1. The declaration above accurately reflects the nature and extent of the contribution of the candidate and the co-authors to Chapter 2 of this thesis, "Garden irrigation as household end-use in the presence of supplementary groundwater supply",
2. No other author contributed to Chapter 2 of this thesis, "Garden irrigation as household end-use in the presence of supplementary groundwater supply", besides those specified above, and
3. Potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 2 of this thesis, "Garden irrigation as household end-use in the presence of supplementary groundwater supply".

Signature	Institutional affiliation	Date
	Stellenbosch University	2 May 2020

**Declaration by the candidate (Chapter 3)**

The nature and scope of my contribution to Chapter 3 of this thesis, the published paper "Analysis of shower water use and temperature at a South African university campus", were as follows:

Nature of contribution	Extent of contribution
- Conceptualised and wrote the paper	80%
- Review of literature	
- Data collection	
- Carried out data and statistical analysis	

The following co-authors contributed to Chapter 3 of this thesis, "Analysis of shower water use and temperature at a South African university campus":

Name	e-mail address	Nature of contribution	Extent of contribution
H. E. Jacobs	hejacobs@sun.ac.za	Assisted with compilation and writing of the paper	5%
B. Biggs	biggsb@jgafrika.co.za	Assisted with data collection	10%
A.A. Ilemobade	Adesola.Ilemobade@wits.ac.za	Assisted with editing	5%

Signature of candidate:   
 Date: 14 Aug 2019

**Declaration by co-authors:**

The undersigned hereby confirm that

4. The declaration above accurately reflects the nature and extent of the contribution of the candidate and the co-authors to Chapter 3 of this thesis, "Analysis of shower water use and temperature at a South African university campus",
5. No other author contributed to Chapter 3 of this thesis, "Analysis of shower water use and temperature at a South African university campus", besides those specified above, and
6. Potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 3 of this thesis, "Analysis of shower water use and temperature at a South African university campus".

Signature	Institutional affiliation	Date
	Stellenbosch University	2 May 2020
	JG Afrika Consulting Engineers	2 May 2020
	University of the Witwatersrand	08 April 2020

**Declaration by the candidate (Chapter 4)**

The nature and scope of my contribution to Chapter 4 of this thesis, the published paper “Clothes washing as household end-use: Comparison of different appliance models in view of expected water savings”, were as follows:

Nature of contribution	Extent of contribution
- Wrote the paper	85%
- Review of literature	
- Data and statistical analysis	

The following co-authors contributed to Chapter 4 of this thesis, “Clothes washing as household end-use: Comparison of different appliance models in view of expected water savings”:

Name	e-mail address	Nature of contribution	Extent of contribution
H. E. Jacobs	hejacobs@sun.ac.za	Assisted with writing and compilation of the paper	5%
U. Terblanche	ute@3e.eu	Data collection	10%

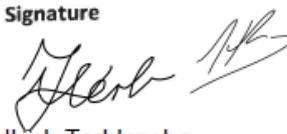
Signature of candidate:  .....

Date: 14 Aug 2019 .....

**Declaration by co-authors:**

The undersigned hereby confirm that

1. The declaration above accurately reflects the nature and extent of the contribution of the candidate and the co-authors to Chapter 4 of this thesis, “Clothes washing as household end-use: Comparison of different appliance models in view of expected water savings”,
2. No other author contributed to Chapter 4 of this thesis, “Clothes washing as household end-use: Comparison of different appliance models in view of expected water savings”, besides those specified above, and
3. Potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 4 of this thesis, “Clothes washing as household end-use: Comparison of different appliance models in view of expected water savings”.

<p><b>Signature</b></p>  <p>Ulrich Terblanche</p>	<p><b>Institutional affiliation</b></p> <p>Stellenbosch University Centre of Renewable and Sustainable Energy Studies, Stellenbosch University</p>	<p><b>Date</b></p> <p>2 May 2020 14/08/2019</p>
--	--	---

**Declaration by the candidate (Chapter 6)**

The nature and scope of my contribution to Chapter 6 of this thesis, the published paper "Extracting household water use event characteristics from rudimentary data", were as follows:

Nature of contribution	Extent of contribution
- Conceptualised and wrote the paper	85%
- Review of literature	
- Carried out data and statistical analysis	

The following co-authors contributed to Chapter 6 of this thesis, "Extracting household water use event characteristics from rudimentary data":

Name	e-mail address	Nature of contribution	Extent of contribution
H. E. Jacobs	hejacobs@sun.ac.za	Compilation and writing of the paper	5%
A.A. Ilemobade	Adesola.Ilemobade@wits.ac.za	Data collection and writing of the paper	10%

Signature of candidate:  .....

Date: 7 April 2020 .....

**Declaration by co-authors:**

The undersigned hereby confirm that

1. The declaration above accurately reflects the nature and extent of the contribution of the candidate and the co-authors to Chapter 6 of this thesis, "Extracting household water use event characteristics from rudimentary data",
2. No other author contributed to Chapter 6 of this thesis, "Extracting household water use event characteristics from rudimentary data", besides those specified above, and
3. Potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 6 of this thesis, "Extracting household water use event characteristics from rudimentary data".

Signature	Institutional affiliation	Date
	Stellenbosch University	2 May 2020
	University of the Witwatersrand	08 April 2020

**Declaration by the candidate (Chapter 7)**

The nature and scope of my contribution to Chapter 7 of this thesis, the published paper “Household water end-use classification model for implementation on coarser data: toward improving the benefits of lower resolution end-use data sets”, were as follows:

Nature of contribution	Extent of contribution
- Conceptualised and wrote the paper	75%
- Review of literature	
- Data collection	
- Carried out data and statistical analysis	
- Developed the model	

The following co-authors contributed to Chapter 7 of this thesis, “Household water end-use classification model for implementation on coarser data: toward improving the benefits of lower resolution end-use data sets”:

Name	e-mail address	Nature of contribution	Extent of contribution
H. E. Jacobs	hejacobs@sun.ac.za	Providing data, editing and review	5%
K. Nguyen	k.nguyen@griffith.edu.au	Providing data, writing, editing and review	10%
C. D. Beal	c.beal@griffith.edu.au	Providing data, editing and review	5%
S. Buchberger	buchbesg@ucmail.uc.edu	Providing data, editing and review	5%

Signature of candidate:  .....  
 Date: 21 October 2020 .....

**Declaration by co-authors:**

The undersigned hereby confirm that

1. The declaration above accurately reflects the nature and extent of the contribution of the candidate and the co-authors to Chapter 7 of this thesis, “Household water end-use classification model for implementation on coarser data: toward improving the benefits of lower resolution end-use data sets”,
2. No other author contributed to Chapter 7 of this thesis, “Household water end-use classification model for implementation on coarser data: toward improving the benefits of lower resolution end-use data sets”, besides those specified above, and
3. Potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 7 of this thesis, “Household water end-use classification model for implementation on coarser data: toward improving the benefits of lower resolution end-use data sets”.

Signature	Institutional affiliation	Date
	Stellenbosch University	21 October 2020
	Griffith University	21 October 2020
	Griffith University	21 October 2020
	University of Cincinnati	21 October 2020

Digitally signed by Steven Buchberger  
 DN: cn=Steven Buchberger, o=College of  
 Engineering and Applied Science,  
 ou=Civil Engineering Program,  
 email=steven.buchberger@uc.edu, c=US  
 Date: 2020.10.21 17:22:25 -0400

**Declaration by the candidate (Chapter 8)**

The nature and scope of my contribution to Chapter 8 of this thesis, the published paper “Classifying household water events extracted from a rudimentary end-use data set – A Case Study conducted in Johannesburg, South Africa”, were as follows:

<b>Nature of contribution</b>	<b>Extent of contribution</b>
- Conceptualised and wrote the paper	85%
- Review of literature	
- Carried out data and statistical analysis	

The following co-authors contributed to Chapter 8 of this thesis, “Classifying household water events extracted from a rudimentary end-use data set – A Case Study conducted in Johannesburg, South Africa”:

<b>Name</b>	<b>e-mail address</b>	<b>Nature of contribution</b>	<b>Extent of contribution</b>
H. E. Jacobs	hejacobs@sun.ac.za	Compilation and writing of the paper	5%
A.A. Ilemobade	Adesola.Ilemobade@wits.ac.za	Data collection and writing of the paper	10%

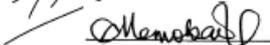
Signature of candidate:  .....

Date: 17 October 2020 .....

**Declaration by co-authors:**

The undersigned hereby confirm that

4. The declaration above accurately reflects the nature and extent of the contribution of the candidate and the co-authors to Chapter 8 of this thesis, “Classifying household water events extracted from a rudimentary end-use data set – A Case Study conducted in Johannesburg, South Africa”,
5. No other author contributed to Chapter 8 of this thesis, “Classifying household water events extracted from a rudimentary end-use data set – A Case Study conducted in Johannesburg, South Africa”, besides those specified above, and
6. Potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 8 of this thesis, “Classifying household water events extracted from a rudimentary end-use data set – A Case Study conducted in Johannesburg, South Africa”.

<b>Signature</b>	<b>Institutional affiliation</b>	<b>Date</b>
	Stellenbosch University	17 October 2020
	University of the Witwatersrand	19 October 2020