Artificial Intelligent Techniques in Residential Water
End-use Studies for Optimized Urban Water
Management

Mr Ao Yang
B. ENG (HONS)

School of Engineering and Built Environment
Griffith University

Submitted in fulfilment of the requirements of the degree of
Master of Philosophy

October 2018
Abstract

In the urban water planning and management industry, end-use water consumption monitoring is a primary tool for water demand management and source substitution. Numerous residential end-use consumption studies have been carried out worldwide in the last two decades. With the rapid development of intelligent technology, the traditional time-consuming process for water flow data disaggregation has been replaced by a smart water metering system with advanced analysis. However, the existing water flow trace analysis system cannot accurately disaggregate all categories of residential water end-use events. In response to this issue, this research focused on developing new techniques, which can improve the autonomous categorisation accuracy of the residential water flow disaggregation process. A rigorous research method was adopted to achieve the above-mentioned research objectives and included the following two stages: (1) review and testing of pattern recognition techniques; and (2) software development. This study employed the extensive South-east Queensland (SEQ) Residential Water End Use Study dataset to undertake the development of the intelligent and autonomous water end-use recognition technique.

Due to the array of objectives, methods, and results, this thesis has been structured around two refereed journal publications produced during the MPhil study. Two themes emerged from the research, namely: (1) development of hybrid intelligent model for mechanised water end-use analysis; and (2) optimising water end-use analysis process with Self-organising maps and K-Means clustering.

The application of many sophisticated intelligent techniques has been attempted in order to tackle this complex problem. In the first stage, the original application of Dynamic Time Warping (DTW) algorithm was found to be ineffective due to settings of the threshold value. Through further investigation into the existing database, Artificial Bee Colony (ABC) and K-Medoids algorithm were selected. In this stage, this technique was applied to assist in finding toilet events in an artificially mixed data. 95.71% accuracy for correctly classified mechanical events was achieved when tested on 136 mixed events from different categories. The performance of the selected algorithms have been compared against previously reported approaches, with the technique and accuracy comparisons presented in a refereed journal paper. While the ABC and K-Medoids approach to clustering flow data into water end-use categories was suitable for mechanical end-use categories, it was less effective for other behaviourally influenced
categories. Further exploration of various water flow data clustering techniques was required in order to discover a more suitable approach for the preliminary clustering of flow data into all of the water end-use categories. This prompted the undertaking of the research activities for the second journal paper described as follows.

The study continued with the development of a hybrid technique in the second stage. Self-organising maps (SOM) and K-means algorithms were applied to the existing software Autoflow through pre-grouping of water end-use events in order to improve the accuracy. The verification on two datasets (i.e., (1) over 100,000 single events, and (2) 30 independent homes), resulted in an improvement in water end-use categorisation accuracy, when compared to the original technique employed in Autoflow, for each residential end-use category. Accuracy improvements were particularly noticeable for the mechanical water end-use event categories (i.e., washing machine, toilet, and evaporative cooler).

The research outcomes have implications for researchers and the water industry. For researchers, the revised Autoflow v3.1 developed in this study is more accurate than previous versions reported in the literature. The novel hybrid pattern recognition approach and the associated algorithms employed in this latest Autoflow v3.1 version can be adapted for a range of pattern recognition problems. For the water industry, an accurate and autonomous water end-use analysis software tool has a range of implications, including, providing bottom-up data for demand forecasting and infrastructure planning, evidence-based water demand management, and end-use level customer feedback phone and web-based applications.
Statement of Originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

_________________________________________ Date: 26/09/2018
Ao Yang
# Table of Contents

ABSTRACT ..................................................................................................................... I

STATEMENT OF ORIGINALITY ............................................................................. III

TABLE OF CONTENTS ............................................................................................. IV

LIST OF PUBLICATIONS ...................................................................................... VIII

LIST OF FIGURES ...................................................................................................... IX

LIST OF TABLES ........................................................................................................ XI

LIST OF ACRONYMS ............................................................................................... XII

ACKNOWLEDGEMENTS ...................................................................................... XIII

ACKNOWLEDGEMENT OF PAPERS INCLUDED IN THIS THESIS ........ XIV

1. INTRODUCTION ............................................................................................. 1

1.1 Research Background ....................................................................................... 1

1.2 Research Objectives .......................................................................................... 3

1.3 Research Method Overview ............................................................................. 4

1.3.1 Phase 1: Knowledge acquisition................................................................. 5

1.3.2 Phase 2: Hybrid Intelligent Model for Mechanised Water End-use analysis..... 6

1.3.3 Phase 3: Optimising Water End-use Analysis Process with Self-Organising
Maps and K-Means Clustering.............................................................................. 6

1.4 Thesis Layout ..................................................................................................... 6

2. LITERATURE REVIEW ................................................................................. 9

2.1 Water End-use Study ....................................................................................... 9

2.1.1 Integrated Urban Water Management ......................................................... 9

2.1.2 Water Demand Management ..................................................................... 10

2.1.3 Residential Water End-use Study ............................................................... 10

2.2 Water Metering Process ............................................................................... 11

2.2.1 Traditional Water Metering Process.......................................................... 11

2.2.2 Advanced Smart Water Metering Process................................................. 12
Table of Contents

2.3 Existing Water Flow Trace Analysis Software ............................................. 17
  2.3.1 Trace Wizard .......................................................................................... 17
  2.3.2 Identiflow ............................................................................................... 18
  2.3.3 HydroSense ............................................................................................ 19
  2.3.4 Autoflow .................................................................................................. 21

2.4 Existing Water End-use Pattern Recognition Techniques .......................... 22
  2.4.1 Dynamic Time Warping Algorithm (DTW) ............................................. 22
  2.4.2 K-Medoids Clustering Algorithm ............................................................. 27
  2.4.3 Artificial Bee Colony (ABC) .................................................................... 29
  2.4.4 Self-organising Maps (SOM) .................................................................. 32
  2.4.5 K-means Clustering Algorithm ................................................................. 33

2.5 Proposed Techniques .................................................................................. 34

2.6 Chapter Summary ....................................................................................... 34

3. RESEARCH METHODOLOGY ................................................................... 37
  3.1 Data Collection and Analysis ..................................................................... 37
    3.1.1 Data collection and processing ................................................................. 37
    3.1.2 Residential water end-use pattern analysis ............................................. 38

3.2 Overview of the Research Method ............................................................. 42
  3.2.1 Phase 1: Knowledge acquisition ............................................................... 42
  3.2.2 Phase 2: Mechanical event analysis ......................................................... 43
  3.2.3 Phase 3: Single event analysis ................................................................. 49

3.3 Chapter Summary ....................................................................................... 51

4. HYBRID INTELLIGENT MODEL FOR MECHANISED WATER END-USE ANALYSIS ............................................................... 52
  4.1 Introduction ................................................................................................. 53
  4.2 Intelligent Techniques for Time Series Clustering ...................................... 54
    4.2.1 Overview of time series clustering techniques ....................................... 54
    4.2.2 Data clustering in water end use analysis .............................................. 55
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>Hybrid Pattern Recognition Techniques</td>
<td>56</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Dynamic time warping algorithm</td>
<td>56</td>
</tr>
<tr>
<td>4.3.2</td>
<td>K-Medoids clustering</td>
<td>57</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Artificial Bee Colony (ABC) algorithm</td>
<td>57</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Hybrid algorithm implementation to water end use analysis</td>
<td>59</td>
</tr>
<tr>
<td>4.4</td>
<td>Experimental Results and Discussion</td>
<td>60</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Hybrid algorithm implementation to water end use analysis</td>
<td>61</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Discussion</td>
<td>62</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusion and Future Work</td>
<td>62</td>
</tr>
<tr>
<td>5.</td>
<td>OPTIMISING WATER END-USE ANALYSIS PROCESS WITH SELF-ORGANISING MAPS AND K-MEANS CLUSTERING</td>
<td>64</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>65</td>
</tr>
<tr>
<td>5.2</td>
<td>Background</td>
<td>67</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Conventional water metering process</td>
<td>67</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Existing autonomous water end use classification models</td>
<td>68</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Water end-use classification process using <em>Autoflow</em></td>
<td>69</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Overview of applied techniques</td>
<td>71</td>
</tr>
<tr>
<td>5.3</td>
<td>Enhanced Water end-use Classification Model</td>
<td>72</td>
</tr>
<tr>
<td>5.3.1</td>
<td>SOM for water end-use pattern recognition</td>
<td>72</td>
</tr>
<tr>
<td>5.3.2</td>
<td>K-means for end-use study</td>
<td>75</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Hybrid SOM-K-means model for residential water end-use pattern</td>
<td>76</td>
</tr>
<tr>
<td>5.4</td>
<td>Model Validation</td>
<td>79</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Independent testing process</td>
<td>80</td>
</tr>
<tr>
<td>5.4.2</td>
<td><em>Autoflow</em> recognition accuracy using independent homes</td>
<td>81</td>
</tr>
<tr>
<td>5.5</td>
<td>Conclusion</td>
<td>83</td>
</tr>
<tr>
<td>6.</td>
<td>CONCLUSION, CONTRIBUTIONS AND IMPLICATIONS</td>
<td>86</td>
</tr>
<tr>
<td>6.1</td>
<td>Conclusion</td>
<td>86</td>
</tr>
<tr>
<td>6.2</td>
<td>Research Contributions</td>
<td>87</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>6.2.1 Contributions to the existing body of knowledge</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>6.3 Study Limitations and Future Research</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>6.4 Closure</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>90</td>
<td></td>
</tr>
</tbody>
</table>
List of Publications

The following papers were produced to disseminate concepts and results from the work undertaken by the author during the course of this MPhil research study.

Journal Publications


List of Figures

Figure 1-1 Research methods overview ................................................................. 5
Figure 2-1 Overview of reviewed literature topics .................................................. 9
Figure 2-2 General water end-use categories ......................................................... 11
Figure 2-3 Water consumption reading (Sydney water) .......................................... 12
Figure 2-4 Smart water metering system (Wills et al., 2009) .................................. 13
Figure 2-5 Decision tree algorithm in Identiflow (Kowalski & Marshallsay, 2005) .... 19
Figure 2-6 Verification of Identiflow (Kowalski & Marshallsay, 2005) ................. 19
Figure 2-7 A pressure stream with ground truth labels from deployment site H2  
  (Froehlich, 2011) ..................................................................................................... 20
Figure 2-8 A sample of instrumented fixtures from our ground truth deployments 
  (Froehlich, 2011) ..................................................................................................... 20
Figure 2-9 A structure of Autoflow (Nguyen et al., 2015) ...................................... 22
Figure 2-10 (a) Two completely different water end-use events (b) Reference event 
  (Nguyen et al., 2011) ............................................................................................... 26
Figure 2-11 Result of DTW algorithm ((a) Threshold value = 700, Identified events = 
  43 .................................................................................................................................. 27
Figure 2-12 A typical process of K-Medoids Algorithm (Amena, 2014) ............... 27
Figure 2-13 The Kohonen’s feature map (Kuo et al., 2002) .................................... 32
Figure 3-1 Format of information collected from data logger .............................. 38
Figure 3-2 A standard shower event ...................................................................... 39
Figure 3-3 Examples of tap event .......................................................................... 39
Figure 3-4 Example of bathtub event .................................................................... 40
Figure 3-5 Example of irrigation event ................................................................. 40
Figure 3-6 Example of clothes washer events ...................................................... 41
Figure 3-7 Example of dishwasher events ........................................................... 41
Figure 3-8 Example of toilet event ....................................................................... 42
Figure 3-9 Results with different threshold values ................................................ 44
Figure 3-10 Position of initial centers with different tests ..................................... 47
Figure 3-11 Sketch map of hybrid pattern recognition techniques ........................ 48
Figure 3-12 Sketch map of hybrid pattern recognition techniques ........................ 48
Figure 3-13 Overall water end-use event classifications procedures with SOM+K- 
  means ......................................................................................................................... 50
Figure 3-14 SOM-K-means clustering analysis procedure included in Autoflow v3.1 . 51
List of Figures

Figure 4-1 Result of DTW algorithm ((a) Threshold value=700, Identified events=43. (b) Threshold value=350, Identified events=56) ..................................................... 55

Figure 4-2 Three directions of grid node in warping path ............................................. 56

Figure 4-3 Improved technique flow chart .................................................................... 59

Figure 4-4 Changes of fitness values in testing dataset .................................................. 61

Figure 4-5 (a): Toilet events categorized using Dynamic Time Warping algorithm. (b): Toilet events categorized using proposed hybrid algorithm. ............................... 62

Figure 5-1 Overall event classification procedures .......................................................... 70

Figure 5-2 Water end-use network in SOM ................................................................ 73

Figure 5-3 Change of neighbor area with iterations ......................................................... 74

Figure 5-4 Overall water end-use event classification procedures with SOM+K-means ......................................................................................................................... 78

Figure 5-5 SOM-K-means clustering analysis procedure included in Autoflow v3.1 ... 79

Figure 5-6 Correctly classified water end use events using technique ............................... 80

Figure 5-7 Computational processing time of Autoflow (v3.1 versus v2.1) ...................... 82
List of Tables

**Table 2-1** Studies with water flow trace analysis software.............................................. 17

**Table 2-2** Summary of Temporal-Proximity-Based Clustering Approach (Rani and Sikka, 2012). .................................................................................................................... 36

**Table 3-1** The categorisation accuracy with different threshold values. ......................... 45

**Table 4-1** Comparison of the Performance between 2 Different Algorithms.................. 62

**Table 5-1** Water end-use event classification accuracy comparison (v3.1 versus v2.1) 81

**Table 5-2** Independent data verification comparison (v3.1 versus v2.1) ....................... 81
List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>Artificial Bee Colony</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>CSIRO</td>
<td>Commonwealth Scientific and Industrial Research Organisation</td>
</tr>
<tr>
<td>DTW</td>
<td>Dynamic Time Warping algorithm</td>
</tr>
<tr>
<td>ENSO</td>
<td>El Niño-Southern Oscillation</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>IUWM</td>
<td>Integrated Urban Water Management</td>
</tr>
<tr>
<td>NZ</td>
<td>New Zealand</td>
</tr>
<tr>
<td>PAM</td>
<td>Partitioning Around Medoids</td>
</tr>
<tr>
<td>REU</td>
<td>Residential End Uses</td>
</tr>
<tr>
<td>SEQ</td>
<td>South East Queensland</td>
</tr>
<tr>
<td>SEQREUS</td>
<td>South East Queensland Residential End Use Study</td>
</tr>
<tr>
<td>SOM</td>
<td>Self-Organising Maps</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>WSAA</td>
<td>Water Services Association of Australia</td>
</tr>
<tr>
<td>WRc</td>
<td>Water Research Centre</td>
</tr>
<tr>
<td>WBKMS</td>
<td>Web-Based Knowledge Management System</td>
</tr>
</tbody>
</table>

Measurement Units

<table>
<thead>
<tr>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>litres</td>
</tr>
<tr>
<td>L/pulse/s</td>
<td>litre per pulse per second</td>
</tr>
</tbody>
</table>
Acknowledgements

First, and foremost, I would like to thank Prof Hong Zhang, Prof Rodney Stewart and Dr Khoi Nguyen for being supportive supervisors. Right from the beginning, Prof Zhang and Prof Stewart has been very enthusiastic, energetic and consistent in providing me with guidance, encouragement, and unlimited mentoring support. I have been very fortunate to have them as my supervisors. I am also grateful for the generous support of Dr Nguyen who has provided invaluable knowledge and suggestions since the first day I met him and continuously throughout the course of my study. I also wish to extend my gratitude to Griffith University for providing a good learning environment and research support.

Last, but not least, I wish to thank my beloved family: my parents, Lijun Yang and Jingyun Du. I am grateful for their love, encouragement, and support, right from the minute I decided to pursue this study. I could not have completed this study without them.
Acknowledgement of Papers Included in this Thesis

Section 9.1 of the Griffith University Code for the Responsible Conduct of Research (“Criteria for Authorship”), in accordance with Section 5 of the Australian Code for the Responsible Conduct of Research, states:

To be named as an author, a researcher must have made a substantial scholarly contribution to the creative or scholarly work that constitutes the research output, and be able to take public responsibility for at least that part of the work they contributed. Attribution of authorship depends to some extent on the discipline and publisher policies, but in all cases, authorship must be based on substantial contributions in a combination of one or more of:

- conception and design of the research project
- analysis and interpretation of research data
- drafting or making significant parts of the creative or scholarly work or critically revising it so as to contribute significantly to the final output.

Section 9.3 of the Griffith University Code (“Responsibilities of Researchers”), in accordance with Section 5 of the Australian Code, states:

Researchers are expected to:

- Offer authorship to all people, including research trainees, who meet the criteria for authorship listed above, but only those people.
- accept or decline offers of authorship promptly in writing.
- Include in the list of authors only those who have accepted authorship
- Appoint one author to be the executive author to record authorship and manage correspondence about the work with the publisher and other interested parties.
- Acknowledge all those who have contributed to the research, facilities or materials but who do not qualify as authors, such as research assistants, technical staff, and advisors on cultural or community knowledge. Obtain written consent to name individuals.

Included in this thesis are papers in Chapters 4 and 5 which are co-authored with other researchers. My contribution to each co-authored paper is outlined at the front of the relevant chapter. The bibliographic details are:

Chapter 4:
Acknowledgement of Papers Included in this Thesis


Chapter 5:

Appropriate acknowledgements of those who contributed to the research but did not qualify as authors are included in each paper.

___________________________ Date: 26/09/2018
Ao Yang

___________________________ Date: 26/09/2018
Supervisor: Hong Zhang (Principal supervisor and co-author)

___________________________ Date: 26/09/2018
Supervisor: Rodney A. Stewart (Associate supervisor and co-author)

___________________________ Date: 26/09/2018
Supervisor: Khoi A. Nguyen (Associate supervisor and co-author)
1. Introduction

This thesis explores new approaches for disaggregating the residential water end-use consumption and develops two intelligent and autonomous models in *Autoflow* software, which is an intelligent water management system, to group similar water end-use events into different categories for single detached households. The models were developed using a prior dataset, which was collected in the South-east Queensland (SEQ) region of Australia. The region, which includes the interconnected urban area of Brisbane, Gold Coast, Ipswich and Sunshine Coast, is considered a rapidly developing area in Australia. The resulting developed and tested models will be able to accurately cluster single mechanical water end-use events. This chapter provides an introduction to the research through a brief description and an overview of the research method. The chapter concludes by providing details on the layout and structure of the thesis. The background information for the research is described in the following section.

1.1 Research Background

Water is a substance that can neither be created nor destroyed, and it is converted from one form to another (Bouwer, 2000). Of the Earth’s water, 96.5% is stored in seas or oceans, and 1.7% in groundwater and glaciers respectively (Gleick, 1993). Of the available freshwater, 98% is stored as groundwater, with only 2% being accessible in lakes and streams. Hence, freshwater is a limited resource that needs to be managed appropriately (Bouwer, 2000). In addition, rapid population growth has also become a challenge to the global water supply. The urban population of the world is expected to rise from the current value by 54%-66% in 2050, and the world’s population is projected to increase to 9.8 billion in 2050 and 11.2 billion by 2100 (Gerland et al., 2014; UN, 2017). The United Nations (UN) estimates that two-thirds of the world’s population will be experiencing water shortages by 2025 (Barlow, 2008; UN News, 2009). By 2050, approximately 1.6 to 2 billion people are expected to be experiencing severe water stress in the Asia-Pacific region (Satoh et al., 2017).

Australia is the world’s driest inhabited continent which has the most unpredictable rainfall patterns, therefore, protection of the nation’s already finite water supply is considered as the top priority (Birrell et al., 2005). The evidence shows that anthropogenic caused global climate change is increasingly affecting weather patterns (Bates et al., 2008; CSIRO, 2010). This change is predicted to result in the alteration of river runoff
Introduction

and water availability, and cause increased precipitation variability and intensity (CSIRO, 2007). These changes will decrease water supply in glaciers and snow cover, increase water pollution, instigate sea level rise, and alter water quantity and quality (Pittock, 2006; Solomon et al., 2007; Bates et al., 2008; Allison et al., 2009).

Urban water utilities made decisions based on finance and engineering in the traditional way; in the last decade, significant consideration has been placed on incorporating sustainability principles into the process (Mitchell, 2006). The integrated urban water resource management (IUWRM) process involves the planning and integration of supply, demand and source substitution options for the sustainable, secure and reliable supply of water to meet projected future water demands of cities or towns (White, 2001; Mitchell, 2006). The combination of socio-behavioural and technological strategies to promote water conservation is the focus (Corral-Verdugo et al., 2002). Some of the benefits of IUWRM include (White & Turner, 2003; Mitchell, 2006):

- Reduced potable demand, wastewater discharges and stormwater flows;
- Lower peak flows and flood damage;
- Enhanced water efficiency;
- Improved stormwater quality (load and concentration); and
- Incorporated green infrastructure, such as wetlands for wastewater treatment.

Water demand management (WDM) is an important part of IUWRM, which can assist in delaying the need for new supply infrastructure (White, 2001). It is a focus for Australia’s municipal and water utilities, and a range of projects which were developed to manage and reduce both residential and industrial water consumption demands (Turner et al., 2005). In addition, WDM includes pricing, metering and enforcement.

The climate of Australia is influenced by its surrounding oceans. In eastern Australia, the El Niño-Southern Oscillation (ENSO) phenomenon can result in floods and prolonged droughts. This exceptional phenomenon has caused extreme events that affect the water supply in Australia (McCarty et al., 2001). In Australia, water scarcity is a significant issue because of the extremely variable rainfall and common droughts. In 2005, Birrell et al. (2005) indicated that water demand will increase by an average of 37% in major cities between 2001 and 2031 by considering the relationship between demographic change and urban consolidation on domestic water use in Australia. Another report has estimated that
the demand for water is expected to grow by more than 40% in many parts of the world by 2050 (Ross, 2017). This research, which is based on water sources, recognises the urgency of water demand in relation to both residential users and industry, as well as the need for the design and application of a new and efficient water-management system.

Currently, many water demand-management schemes have been undertaken in Australia, including new and refurbished facilities in water-efficient design, the management of water infrastructure in such a way as to maximise efficiency, and the use of alternate water sources to replace potable water use (Britton et al., 2009). However, due to Australia’s inadequate metering process, these demand-management practices have resulted in ineffective activities. Facing the current water resource challenge, a knowledge management system that focuses on water consumption is an urgent requirement. The goal of such a system is to achieve a combination of smart metering, end-use water consumption data, wireless communication networks, and information management systems. The purpose of this system would be to provide, for both the customer and the water utilities, real-time information on how and when water is consumed (Stewart et al., 2010) and is significantly dependent on efficient smart meters. In recent years, some water-demand modellings that use smart meters, have been applied in some cities around the world. Despite available software applications for water end-use analysis, such as Trace Wizard (DeOreo et al., 1996), Identiflow (Kowalski & Marshallsay, 2003), HydroSense (Froehlich et al., 2011) and Autoflow (Nguyen et al., 2015), there is still the requirement to improve water end-use categorisation accuracy. Therefore, the key goal of the present study is to develop intelligent clustering algorithms for smart meters that would be able to autonomously categorise water flow trace data into particular water end-use event categories (clothes washer, dishwasher, toilet, etc.), thus serving more effective water-management systems.

1.2 Research Objectives

The study was executed to provide knowledge to address the above-mentioned research goal. The objectives of this MPhil project include but are not limited to:

- Understanding the characteristics of water usage data and patterns of each specific water end-use category and existing intelligent algorithms applied;
- Identifying both the strengths and weaknesses of existing approaches to residential water end-use pattern recognition;
• Developing a technique, which is a hybrid combination of existing advanced techniques (e.g., Artificial Bee Colony, Dynamic Time Warping algorithm, K-Medoids, etc.) for pattern recognition, based on distance measurement;
• Developing a hybrid technique, which is a hybrid combination of Self-organizing map (SOM) and K-means, for clustering similar events into groups based on distinct water-flow features.
• Validating the efficiency and accuracy of the developed algorithm.

The research was conducted within the confines of the following study scope:

• The study was limited to the water end-use data available for the context of the SEQ region of Australia.
• The research was restricted to end-use analysis of the residential sector only (i.e., it does not include the commercial or industrial sector).

1.3 Research Method Overview
There are three main phases of the research method: knowledge acquisition phase; hybrid intelligent model for mechanised water end-use analysis; and optimising water end-use analysis process with Self-organising maps and K-means clustering. Each of these phases includes numerous stages that contained unique methods and design. Each key phase and associated stage of the research method is presented in Figure1-1. A summary description for the key research phases is presented below.
1.3.1 Phase 1: Knowledge acquisition

Phase 1 involved the process of obtaining the required background knowledge and formulating research objectives. The result was the development of two separate modules, one to analyse mechanical events and the second to analyse all single-event classifications. The second module was combined into the previous flow trace analysis
system software *Autoflow*, which is able to update the existing pattern-recognition techniques and improve their accuracy.

### 1.3.2 Phase 2: Hybrid Intelligent Model for Mechanised Water End-use analysis

To promote the present study, the dataset from South East Queensland Residential End Use Study (SEQREUS) was utilised, which had been collected from high-resolution water meters (0.014 litres per pulse). In addition, data loggers were used to record water volume data versus time (i.e., 5- or 10-second steps). Collected data were then analysed and grouped into different end-use categories as the dataset for the development of modules. The key goal of Phase 2 was to develop algorithms for the classification of mechanical events from the recorded flow trace. For this phase, existing pattern recognition tools, such as Dynamic Time Warping (DTW), were adapted and combined with Artificial Bee Colony (ABC) and K-Medoids algorithms to provide a new hybrid technique for the above-mentioned issues.

### 1.3.3 Phase 3: Optimising Water End-use Analysis Process with Self-Organising Maps and K-Means Clustering

Phase 3 involved the development of an intelligent autonomous system, *Autoflow v2.1*. *Autoflow*, which served as a prototype tool, used to solve the problem of autonomously classifying residential water end-use events (Nguyen et al., 2015). The developed pattern-recognition technique was combined with the Self-organising Maps (SOM) and K-Means clustering algorithms, which had a 1.7% to 9% improvement in accuracy considering over 100,000 single events. In addition, a complete autonomous water end-use application, named *Autoflow v3.1*, was verified for 30 independent homes in Australia to confirm its effectiveness in flow trace disaggregation.

### 1.4 Thesis Layout

This research is presented using a hybrid thesis layout that includes both traditional thesis chapters and reformatted peer-reviewed publication chapters. This layout differs from a traditional thesis since it includes published peer-reviewed papers instead of the traditional data analysis, results and discussion chapters. However, the traditional chapters (i.e., introduction, literature review, and research method and data analysis) are presented. Published, accepted or submitted peer reviewed journal and conference publications constitute the remainder of the chapters. Each reformatted-paper chapter includes a literature review, methodology, results and discussion specific to the topic of
that particular paper. The last chapter summarises the overall research conclusions, contributions, limitations, and future research directions.

There are two distinct peer-reviewed publications within this thesis:

(i) The establishment of a hybrid technique to improve accuracy of clustering mechanical events (i.e., Journal paper 1).

(ii) The employment of SOM and K-means clustering algorithms to develop single-event classification modules and apply them to the water flow trace system, Autoflow (i.e., Journal paper 2).

In its entirety, the thesis includes six chapters. Chapter 1 outlines the background of water end-use, and a summarised overview of the research method. In addition, this chapter presents the objectives of the research according to the literature review and the established research gaps. Furthermore, the research objectives guide the design of the research methods and provide support for the results to meet the objectives.

Chapter 2 provides a detailed review of all literature relevant to water end-use studies, smart water metering, existing water end-use pattern-recognition techniques and water flow trace analysis systems. Moreover, according to the earlier research covering fields of water flow trace analysis, this chapter also presents the key gaps that currently exist in the body of knowledge.

Chapter 3 describes the research method, specifies the key theories according to the objectives, and discusses the techniques adopted in this study. Initially, this chapter presents the methodology applied to carry out the entire research project. Additionally, detailed methods for each particular phase and stage are introduced. Moreover, this chapter also includes data collection and analysis.

Chapter 4 presents the first reformatted peer-reviewed journal paper, which includes a detailed application of ABC, DTW and K-Medoids clustering for grouping mechanical events. It describes a clustering process that selects the maximum number of toilet events from a mixed dataset.

Chapter 5 presents the second reformatted journal paper. This chapter displays a hybrid technique of pattern recognition for categorising single events, using both the SOM and K-Means algorithms within the autonomous, Autoflow software. Two verification processes using different databases or independent homes were conducted to indicate the
Introduction

effectiveness of the suggested algorithm. The chapter also explores discussion of and prospects for the intelligent autonomous system in the future.

Finally, Chapter 6 summaries the key research outcomes, and limitations of the current research. This chapter also presents the implications of the research for the water industry as well as the pattern recognition field.
2. Literature Review

This chapter presents a review of the literature pertinent to the research. The topic is introduced as an overview of water end-use studies in recent years. A comparison between traditional water-metering processes and smart water metering is outlined. The chapter continues with a detailed description of the existing water end-use pattern-recognition techniques and their applications, which include different types of water flow trace analysis systems. A thorough overview of earlier studies on water flow trace systems around the world is presented, which informed the development of the new hybrid technique for similar problems in the software Autoflow. The chapter also presents the current gaps in the body of knowledge, and details the current research approach. An overview of the literature examined in this chapter is presented in Figure 2-1.

![Figure 2-1 Overview of reviewed literature topics](image)

2.1 Water End-use Study

2.1.1 Integrated Urban Water Management

Integrated Urban Water Management (IUWM) is a popular worldwide concept (Closas et al., 2012; Bahri, 2012), and particularly within Australia (Mitchell, 2006; Ferguson et al., 2013; Mukheibir et al., 2014). This process includes three services: water supply,
sewerage, and drainage (Marlow et al., 2013). There are other similar terms that can also refer to these concepts of IUWM, including Integrated Water Resource Management, Total Water Management, and Integrated Water Cycle Management (Furlong et al., 2015). However, all of these terms maintain that all water services should be considered holistically for water managers to achieve the best outcomes for the community (Guthrie et al., 2017). For water supply, the IUWM process involves the planning and integration of supply, demand and source substitution options for the sustainable, secure and reliable supply of water to meet the demands of the areas (White & Fane, 2001; Inman & Jeffrey, 2006; Mitchell, 2006).

2.1.2 Water Demand Management
Protection of the nation’s water supply is the top priority in Australia, which is the world’s driest inhabited continent (Birrell et al., 2005). However, global climate change is affecting weather patterns (CSIRO, 2010). Due to climate change, Australia increasingly experiences many climatic extremes, such as heatwaves, floods, droughts and frosts (Westra et al., 2016). Climate change and these extremes have affected water demand and water availability (Jaramillo & Nazemi, 2018). In addition to climate change, the changing human lifestyles and eating habits have also produced an increase in water demand (Koech et al., 2018). Both the population and climate change present challenges for water resource managers (McDonald et al., 2011), and water security is a major concern for water authorities, (Grafton et al., 2011, Yigzaw and Hossain, 2016), public and private industries, and all levels of government in Australia (Inman and Jeffrey, 2006). Water demand management has evolved as an important strategy to reduce water consumption (Stavenhagen et al., 2018) and is also a key element of IUWM, which assists in delaying the need for new supply infrastructure (White & Fane, 2001; Anderson, 2003).

2.1.3 Residential Water End-use Study
Appropriately planning to ensure a secure and sustainable water supply for future populations is a necessary measure for Australia. The water end-use study, which is a detailed prediction tool, is essential to support planning of water demand and security. Both the growth in the number of dwellings and increasing demand for water are two important factors that have triggered a focus on understanding, predicting, forecasting, measuring, and validating urban water consumption. For residential households, water end-use studies can provide information about essential water consumption in terms of when, where, how, and why residential users consume water in their homes (White & Fane, 2001; Giurco et al., 2008). Determining water consumption, which includes the
shower, tap, irrigation, clothes washer, dishwasher, toilet flushing and leaks, is the purpose of end-use studies (Gato, 2006).

![General water end-use categories](image)

**Figure 2-2** General water end-use categories

Another important purpose of a water end-use study is to improve urban water security. Drought and population demands lead to numerous instances that the water capacity of water supply reservoirs in South-East Queensland (SEQ) dropped below 20% over the last ten years. It has caused state or local governments to execute different water supply plans, coupled with a range of demand management interventions, in order to improve urban water security (Willis et al., 2013). Hence, such water end-use study data are pertinent for daily demand forecasting and for refinement in the planning and management of water demand and supply for the regions (Gato, 2006; Schlarfrig, 2008). Giurco et al. (2008) emphasise that several end-use studies have been undertaken in Australia. According to Schlarfrig (2008), there has been increasing recognition of the need for more comprehensive and frequent end-use studies. Giurco et al. (2008) and White et al. (2003) recommended that there should be more studies in the water end-use consumption field. However, the different regions of the studies will be determined by the per-capita consumption (Mayer and DeOreo, 1999).

### 2.2 Water Metering Process

#### 2.2.1 Traditional Water Metering Process

The traditional urban water-management system results in most of the water-metering systems in Australia being unable to provide real-time or high resolution water consumption data. Due to the inaccurate level of data resolution, conventional meters also cannot record when (i.e., time of day), and where (i.e., different types of water device) the consumption takes places. Normally, water consumption readings are recoded manually on monthly or half-yearly bases. Figure 2-3 shows an example of a water consumption reading, which uses only three data points (This Reading, Last Reading and
Consumption) to describe the water use states of a residential household for 3 months. Clearly, this traditional or existing water-metering system is unable to provide enough support to water planning or other water management processes. In addition, this metering process cannot meet the increasing requirements of water managers on utilisation for water resources management and does not assist society at large in reducing the pressure of water security issues associated with climate change (Stewart et al., 2009).

<table>
<thead>
<tr>
<th>Account for residential property</th>
<th>1 Sample Ave Sampletown</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Water meter details</strong></td>
<td></td>
</tr>
<tr>
<td>Meter Reading Period: 20 Oct 15 - 22 Jan 16</td>
<td></td>
</tr>
<tr>
<td>Meter No. ABCD1234</td>
<td>This Reading 1675</td>
</tr>
<tr>
<td>Total water used in 94 days was 17 kilolitres</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2-3** Water consumption reading (Sydney water)

### 2.2.2 Advanced Smart Water Metering Process

With rapid developments in smart and intelligent technology in recent years, such as pattern recognition, machine learning, and deep learning, traditional metering of residential water end use has been replaced by the smart water metering process. Household water, as one of the major sources of supply for indoor use or outdoor use, needs more intelligent equipment to measure the water flow. Water authorities can read the water use information in real time from the smart water-metering system (Montginoul & Vestier, 2018), and the data collected from smart water meters can develop various water-demand models to understand the factors contributing to peak water demand (Willis et al., 2010, Gurung et al., 2014, Gurung et al., 2016, Sahin et al., 2015, Sahin et al., 2017, Beal & Stewart, 2013, Savić et al., 2014).

Smart water-metering studies for water end-use have been carried out for nearly 20 years, beginning in 1998 (Beal et al., 2013), and the technology has become more advanced; that is, the type of data transport has changed from manual to automatic. In addition, the questions with which households are concerned regard accuracy and features. For the smart water-metering system, the major parts are categorised and different users can get valuable information. An example of an interested user is a government organisation,
which needs to know the residential water consumption in order to regulate water prices. However, there is not truly a perfect model to be published and used in real life. Until now, there has been some substantial progress in the development of intelligent models for the categorisation of residential water end-use events (e.g., Nguyen et al., 2013).

Compared to the existing metering process, the smart water-metering system is highly accurate for event categorisation. Stewart et al. (2009) emphasised that current water metering methods cannot provide real-time data of water consumption because the data are recorded manually on a monthly or even half-yearly basis. Therefore, this conventional metering approach is not adequate to meet the requirements of different types of users such as the government and the whole of society. However, Giurco et al. (2008) stated that smart meters should afford additional high-quality feedback of water usage such as end-use data because this function is useful for water utilities and policy makers. In addition, Neenan (2008) pointed out that there are three basic procedures of smart water meters. These are to capture, collect, and communicate, and all of these functions are focused on a real-time (or nearly real-time) data basis. To achieve these purposes, smart water meter equipment should be composed of a high-resolution water meter and a logger for water data. Thus, the water data can be recorded as an electronic signal for analysis (Britton et al., 2008; Stewart et al., 2009; Willis et al., 2013). However, Hauber-Davis and Idris (2006) provided an approach for analysing the electronic signals of smart water meters in which the water data can be transferred to computers or central data hubs via the GSM network. An example of a smart water-metering system is displayed in Figure 2-4.

![Smart Water Metering System](image)

**Figure 2-4** Smart water metering system (Wills et al., 2009)

Loh and Coghlan (2003) applied smart meters and loggers (unspecified) as a tool in data capture to analyse water end-use for both single-residential households and multi-
residential households from 1998 to 2001 that were located in Perth, Western Australia. Washing machines, showers, and toilets are the primary water end-use categories in this research, and it presents the more salient results of data analysis and summarises critical findings. Additionally, the results of this study are helpful to the Corporation for improving the forecasting of future demands and developing water use efficiency programmes.

In 2003, a collaborative research project by the Water Research Centre (WRc), including 13 water companies and the UK Environment Agency, yielded further information on the microcomponents of water use in domestic properties in the UK. The difference from others is the dwelling type and importantly software used for analysis. Moreover, investigating the uncertainty and diurnal variations of the data has allowed the investigation of demand reduction strategies at household and supply-area levels. This paper expands on the above issues and discusses how this approach can be used to help plan and manage the water supply in the future (Kowalski & Marshallsay, 2005).

The Tampa Water Department Residential Conservation Study was carried out in 2001 in Florida, USA, based on high-end users, which means an average daily per-capita use higher than 60 gallons per capita per day. This research measured the impact of a variety of indoor water-conservation measures on both aggregate and individual water use patterns. The results from this study make it clear that residential retrofits from the customer’s perspective can be a cost-effective tool for saving water and that customers are quite satisfied with the performance of the new high-efficiency toilets and clothes washers currently available. In addition, the results also provided compelling evidence of the effectiveness of internal water-conservation measures and justification for continued support of cost-effective programmes across the country (Mayer et al., 2004).

Roberts (2005) proposed the modelling of water end-use of Yarra Valley Water (Australia), which used high-resolution meters and data loggers to collect water usage data for two weeks in February 2004 (summer) and two weeks in August 2004 (winter). This report presents the results of the residential end-use measurement study as well as comparing where possible, the results of actual measurement with survey-based estimates. In addition, the study has also enabled more informed design and assessment of various demand-management programmes. It also provided a valuable data set from which to provide customers with informative usage data via their quarterly account statement.
The NZ Water End Use and efficiency project, located in the Auckland region, was carried out in 2007. It took six months and included summer and winter. The results from this study provided a useful insight into the ways water is used in single detached homes and ways to measure consumption of users’. By obtaining accurate end-use information, areas in which water can be used more efficiently, can be identified. The most substantial impact could be achieved by installing LFSHs, front-loading washing machines, and dual flush toilets. The installation of a rain tank or greywater systems could reduce mains water consumption from the mains even further. In addition, reduction in the amount of hot water used reduces the energy required to heat it and hence the direct costs involved. Water and wastewater treatments are energy intensive processes and a reduction in water demand translates into electricity savings. A reduction in energy use reduces the amount of greenhouse gases, mainly CO₂, which contribute to global warming (Heinrich, 2007).

A 2-year project that started in 2009 was carried out on the Gold Coast, Australia. A total of 151 households are involved in this study. The sampling regime was between winter 2008 and summer 2009. The advent of high-resolution smart meters and data loggers allowed for the disaggregation of water flow recordings into a registry of water end-use events (e.g., showers, washing machine, and taps). This study shows that resource consumption awareness devices, such as the one evaluated in this project, assist resource consumers in taking ownership or their usage and tackling their individual and/or society driven conservation goals, ultimately helping to reduce the ecological footprint of built environments (Willis et al., 2010).

In 2010, in order to estimate the household water end use in three localities in drought-prone Trincomalee in the dry zone of Sri Lanka, a study was implemented by Sivakumaran and Aramaki (2010). Most water in the research area derives from taps, reservoirs and dug wells. Questionnaires and interviews were the two primary ways to determine each end use and the monthly usage of tap water. Household water is used mostly for bathing, washing clothes, vehicles, and toilets. Despite the small sample size and the economic and social differences, the results correspond quite well to those found in other research.

With the increasing pressure from water scarcity due to global warming, a better water-management system needs to be applied in Australia (Savenije & Van Der Zaag 2002). However, the metering process is the most essential part of the existing water-management system. As with other measuring devices, a traditional water meter is not an
ideal measuring instrument and is not capable of registering the exact amount of water consumed by a user. The weakness of the current water-metering system in Australia is that it cannot provide real-time water consumption data or sufficient data points. Water consumption data is generally recorded manually on a monthly or a quarterly basis and the whole year’s data describes only total consumption in kilolitres of water. There is no further information that might describe when and where the consumption takes place (e.g., washing machine or toilet), in any water-metering instruments. The existing water-metering system cannot contribute to the water planning and management processes. A new water planning and management system, which is called the Web-Based Knowledge Management System (WBKMS), can provide real-time information on water consumption. This system integrates smart metering, end-use water consumption data, wireless communication networks and information management system (Stewart et al., 2010). As an essential part of the water-management system, smart water meters that use new technology, can capture and transmit water-use information when a water end-use event happens or almost as it happens. Smart water meters carry out three basic data functions: capture, collection and transmission (Neenan, 2008).

First, a smart water meter with high resolution (e.g., 72 pulses per litre), can capture the water consumption automatically and electronically. To get usable data, a data logger is linked to the smart meter. The water consumption can be transformed into an electronic signal. Hence, these signals can be read by the computer with data distribution technologies (Nguyen, 2011).

Mead (2008) applied smart meters in an investigation of domestic water end use located in Toowoomba, Australia. The sample size for this study was ten households, which were selected to be fitted with high-resolution water meters and data loggers. This study has indicated that smart meters can provide detailed data about water end use. These data can be used to reveal the effectiveness of water-conservation appliances and fixtures. This project also has provided insights into the variables that affect water consumption by appliances and fixtures. The results can be used to better educate the users about water conservation and how they can make a difference.

In summary, smart water metering is a useful technology that can be applied in Australia as an indispensable element in the new water-management system. The system can provide benefits for both consumers and water utilities for controlling water consumption to face the shortage of water resources.
2.3 Existing Water Flow Trace Analysis Software

As an important element of smart water metering, water flow trace analysis software has been developed in recent years. The target of this type of software is to categorise a raw water flow trace data into appropriate end-use categories (i.e., shower, tap, and toilet), which is the main concept of water end-use study. According to the methodology, there are two primary type of software for disaggregation: (1) decision tree algorithm-applied by Trace Wizard® (DeOreo et al., 1996) and Identiflow® (Kowalski & Marshallsay, 2003); (2) machine-learning algorithms-applied by HydroSense (Froehlich et al., 2011), BuntBrainForEndUses® (Arregui, 2015), REU2016 (Vitter & Webber, 2018), and Autoflow (Nguyen et al., 2015); Table 2-1 shows a summary of studies and the software they applied in the last two decades.

### Table 2-1 Studies with water flow trace analysis software

<table>
<thead>
<tr>
<th>Research</th>
<th>Location</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeOreo &amp; Mayer (2000)</td>
<td>USA</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>Mayer et al., (2000)</td>
<td>USA</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>Mayer et al., (2003)</td>
<td>USA</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>Mayer et al., (2004)</td>
<td>USA</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>DeOreo et al., (2011)</td>
<td>USA</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>DeOreo (2011)</td>
<td>USA</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>Heinrich (2007)</td>
<td>New Zealand</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>Willis et al., (2009)</td>
<td>Australia</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>Mead &amp; Aravinthan (2009)</td>
<td>Australia</td>
<td>Trace Wizard®</td>
</tr>
<tr>
<td>Kowalski &amp; Marshallsay (2003; 2005)</td>
<td>UK</td>
<td>Identiflow®</td>
</tr>
<tr>
<td>Froehlich et al., (2011)</td>
<td>N/A</td>
<td>HydroSense</td>
</tr>
<tr>
<td>Nguyen et al., (2015)</td>
<td>Australia</td>
<td>Autoflow</td>
</tr>
</tbody>
</table>

2.3.1 Trace Wizard

From 1994 to 1996, DeOreo & Mayer (1994) first proposed the concept of measuring the residential end use of water and collected the water consumption flow data using AquaCraft® in the USA. AquaCraft® developed software, named Trace Wizard®, which can automatically disaggregate flow data into household end uses (DeOreo et al., 1996). DeOreo & Mayer (2000) reported that the accuracy of end-use analysis could be improved through the independent analysis of data by several analysts. This process can also be done by undertaking a home visit or audit to assist in determining water use fixtures and behaviours of the home which are being analysed by Trace Wizard®.
Mayer et al. (1999) first applied manual loggers and Trace Wizard to analyse the water end use of single detached households in the USA and Canada for two weeks in summer and 2 weeks in winter. The selected results based on indoor water end-use analysis include the clothes washer, shower, and toilet. This study is an earlier analysis of water end-use areas in the single-family residential households and provided the predictive models for projecting residential water demands. However, there still are many other areas that are not covered in this research such as multi-family residential water use, commercial water use, and institutional water use.

The disaggregation process accords with the characterisation of the water flow trace (i.e., volume, duration, and peak flow rate). The analytical process can be summarised as follows:

- Identify the efficiency of each water device and determine the characteristics of each water end-use category (i.e., volume, duration, maximum flow rate, and most frequent flow rate);
- Create a template according to the diaries, which are completed by the household, and flow trace data;
- Apply an analytical process based on the above information and the analysts’ experience. All individual events will be grouped into different categories or identified as combined events.

The advantages of Trace Wizard® are two-fold: (1) due to the pattern recognition algorithm being simple, the disaggregation process is completed in a short period, and (2) the result depends on the input data, which must be collected and processed. However, the disadvantage is that the preparation process takes an extended period of time for one household. Another disadvantage of Trace Wizard® is that the accuracy is reduced when more than two events occur concurrently (Mayer & DeOreo, 1999; Nguyen et al., 2013).

### 2.3.2 Identiflow

Similar to Trace Wizard®, Identiflow® provides high-resolution data and can analyse it using a decision-tree algorithm to perform a disaggregation of the water consumption at the household level. Compared to the Trace Wizard®, the development of Identiflow® software focused on the ability to analyse the combined events. The active approach and verification are presented in Figure 2-5 and Figure 2-6.
Due to the application of the decision tree algorithm, the processing time of the technique can be reduced significantly. This is the most important advantages of Identiflow®. Compared to Trace Wizard®, simplification of inputting data and automation of uploading data is another advantage. However, the categorisation accuracy heavily depends on the physical features of a water device. Two water end-use events that occurs from two different types of water devices but have similar physical features (i.e., volume, flow rate, and duration), will be grouped into the same category; therefore, the accuracy is reduced when many water devices have similar water flow trace patterns.

2.3.3 HydroSense

HydroSense is software that relies on pressure sensors to disaggregate water end-use events. This system records the pressure waves, which are different between each water device, and identifies when the fixtures are opened or closed (Figure 2-7). For evaluating
the performance of the approach using real data, a sensing network was applied in residential households including three homes and two apartments (Figure 2-8).

![Figure 2-7 A pressure stream with ground truth labels from deployment site H2 (Froehlich, 2011)](image1)

![Figure 2-8 A sample of instrumented fixtures from our ground truth deployments (Froehlich, 2011)](image2)

For analysing and classifying the collected data, Bayesian estimation was used in this system. The specific process is presented as follows: (1) buffering and separating the incoming water-pressure stream into transient parts; (2) comparing each segmented pressure transient to the labelled templates using a set of algorithms; (3) a language model that uses bigram analysis, is applied to determine the likelihood of a sequence of valve signatures, such as valve opening and closing valve events, into paired tuples; (4) extracting features from paired tuples and comparing them with the smoothed probability distributions: and (5) the probabilities from the above steps were multiplied together for
each sequence. For a sequence with the highest probability, the validation showed that two sensors could improve the accuracy from 96% (single sensor) to 98% at the fixture category level.

HydroSense has a high accuracy due to the utilisation of a vast sensor network. Similar to Trace Wizard®, HydroSense is not affected significantly by the social and demographic features. This technique was designed to identify events when the valve was closed or opened, which was based on the collected sensor signal for classification purposes. However, the main drawback of this system is its requirement for many pressure sensors to be attached to water devices inside households in order to accurately identify end-use events (i.e., Froehlich et al. (2011) used 33 sensors in a single household). Additionally, aesthetics is another issue since sensors are connected directly to the water device and are highly visible in the household.

2.3.4 Autoflow

Nguyen et al. (2011) applied the DTW algorithm to select prototypes in single water end-use events that have a similar magnitude and length, such as clothes washers and dishwashers. The advantage of this technique is that both the processing time and memory consumption are reduced. In addition, the prototypes data can be applied to the pattern recognition of water consumption end-use events. The results of this research demonstrated that the DTW algorithm was able to categorise water end-use events that have similar patterns.

Nguyen et al. (2013) examined the intelligent model to categorise single residential water end-use event patterns, including the shower, tap, dishwasher, clothes washer, full-flush toilet, half-flush toilet, bathtub, irrigation, and leak. A robust hybrid analytical method, including the Hidden Markov Model, Dynamic Time Warping, and event probability, was applied in this model. Context data combined with the event category probability established from the HMM algorithm enabled more accurate predictions of inconclusive events.

Nguyen et al. (2013) applied gradient vector filtering to the previous intelligent model to disaggregate complex water-flow data into end-use combination events such as showers and toilet flushes occurring in the same period. This research confirmed that the flow trace could be analysed automatically using an intelligent algorithm in smart water metering. In addition, since combined events can account for a large proportion of
residential consumption, this disaggregation and recognition technique is an integral part of an automated pattern-recognition model in water-management system.

Then, Nguyen et al. (2015) proposed an intelligent application software Autoflow, which is serves as a prototype tool to solve the problem of autonomously categorising residential water consumption data into a registry of single and combined water end-use events. This software is combined with the Hidden Markov Model (HMM), Artificial Neural Network (ANN) and Dynamic Time Warping (DTW) techniques to examine both the shape pattern and physical characteristics of each event in order to identify to which water end-use categories it belongs. Figure 2-9 shows the overall procedure of Autoflow.

The specific recognition process was presented as follows: (1) using HMM to estimate likelihood based on the event flow-rate pattern; (2) using ANN to examine physical features similarities (e.g., rate of change of water flow); (3) combining HMM and ANN likelihood decisions to categorise water events; and (4) applying further decision support algorithms such as DTW and event probability tables if uncertainty still exists.

Version 2 of Autoflow achieved an overall water end-use recognition accuracy of just over 90%, which is near the target of 95% that is considered sufficient for a commercial application of the software. In addition, this system only relies on high-resolution water meters, and analysis processes depend on hybrid techniques. However, the existing self-learning function within this software needs to be developed to characterise better the distinct end-use characteristics of new residential houses.

2.4 Existing Water End-use Pattern Recognition Techniques

2.4.1 Dynamic Time Warping Algorithm (DTW)

Clustering technique was employed as a main classifier for the present study, which helped to disaggregate a raw flow trace data into particular end-use categories. The study
Literature Review

revealed that, by incorporating a technique called Dynamic Time Warping algorithm into the existing clustering algorithm background, the combined technique demonstrated a powerful tool in water end-use analysis, which overcame more challenging issues faced by the standalone clustering algorithm. In DTW algorithm, the threshold is a parameter to describe the similarity between water events, and its purpose is considered as a filter. Water events, which can satisfy the similarity requirement, can be grouped as one category. In contrast, the other water events cannot be grouped in the same group. Parameters of the DTW algorithm, such as window parameter size and threshold value are changed depending on different water end-use category. For example, when DTW applied to analyse events which have a long duration and high volume (e.g. shower, bathtub), the threshold will be set up a high value. Until now, there is still no guideline to introduce how to set parameters strictly. This is a task that requires researchers’ experience. However, DTW plays an important role in this study as it was applied to help classified toilet events. Aspects of DTW including processes, a combination of K-Medoids and ABC are discussed in Chapter 4.

In the current study, DTW was employed as a supplementary technique to help increase the overall disaggregation accuracy. In the mechanised water end-use analysis, it was used to measure the distance between water events. The basic calculation processes of K-means can be summarized in three steps:

*Step 1: Calculate the distance matrix of the two sequences between points.*

*Step 2: Looking for a path from the bottom left corner to the upper right corner of the matrix, making the sum of the elements on the path is minimum.*

*Step 3: Output the path*

However, looking for this particular path should to follow three guidelines:

- Boundary conditions;
- Continuity;
- Monotonic.

In this study, the main objective of DTW algorithm is to determine the distance between two different water end-use events and this distance will be as a benchmark in K-Medoids clustering.
Dynamic Time Warping is an algorithm for measuring two time or speed sequences that have similar features. For example, similarities in walking patterns would be detected, even if in one video the person was walking slowly and in another he or she were walking more quickly or even if there were accelerations and decelerations during one observation (Jeong et al., 1990). The DTW algorithm has been widely applied in many areas, including science, medicine, industry, and finance (Keogh & Ratanamahatana, 2005). A well-known application of the DTW algorithm is automatic speech recognition, which can cope with different speaking speeds (Bhagat, 2005).

In general, two given sequences (e.g., time series) are “warped” non-linearly in the time dimension to determine a measure of their similarity independent of specific non-linear variations in the time dimension. A computer can find an optimal match between these two sequences (WorldLingo, 2003). Hence, managing time distortions by realigning time series when comparing time series sequences is a major function of the DTW algorithm (Zhang et al., 2017). Comparing to other pattern matching algorithms, DTW is an algorithm particularly suited to matching sequences with missing information, provided there are long-enough segments for matching to occur (Myers & Rabiner, 1981).

For example, given two-time series sequences, $X$ and $Y$, the DTW algorithm can also calculate the distance between $X$ and $Y$. This function can classify a set of time series sequences because similarity or distance between two sequences is the major criteria (Hsu et al., 2015). In addition, Nguyen et al. (2011) applied the DTW algorithm as a tool for measuring the distance between two single events based on time series, to select prototype for the water end-use study. Dynamic time warping is also a much more robust algorithm for measuring distance and allowing similar shapes to match even if they are out of phase in the time axis (Ratanamahatana & Keogh, 2004).

Rath and Manmatha (2003) described DTW as a procedure that can be used to match handwritten words in noisy historical documents. The experiments showed that this algorithm is better and faster than shape context matching. In biotechnological application, Gollmer and Posten (1996) applied the DTW as a tool for supervision with particular reference to fault detection. While other techniques such as observer-based or statistical approaches, require either mathematical model descriptions or many historical data sets for training, the programming using DTW was proposed to be much more advantageous.
Galbally and Galbally (2015) proposed a pattern recognition approach based on the DTW algorithm for automatic transient identification in the energy field. The results show that there is a high accuracy of the present approach, which is combined with other features such as its low level of complexity and its lower requirements for training data. Johnstone et al. (2015) introduced a novel approach for discrete event simulation output analysis based on the DTW algorithm. The results show that the proposed technique can automatically identify critical factors that influence the overall dwell time of system entities. In addition, due to the new technique providing insight into system performance, it can also replace the traditional analysis techniques.

Several researchers have also been undertaken into the adaptation of the DTW technique to the present project; however, the achieved accuracy was not as high as expected. In general, DTW is a simplified approach in pattern recognition, especially when the sequences are based on time series. Nguyen et al. (2011) found that the shortcoming of the DTW algorithm for water end use is that two completely different events can be classified in the same group because they have similar distances to the reference, as illustrated in Figure 2-10:
Figure 2-10 (a) Two completely different water end-use events (b) Reference event (Nguyen et al., 2011)

In this example, distance from the first and second sequences (Figure 2-10 (a)) to the reference are 1295 and 1291 respectively. Hence, the limitation of this method is its heavy dependence on the setting of the threshold value, which defines the similarity between two objects. A larger threshold value would classify a group that includes two or more types of samples and the accuracy of the algorithm thus would be reduced. On the contrary, a smaller threshold value may assign two relatively similar samples into different respective clusters. The following Figure 2-11 illustrates an example of a clustering result with different threshold values.
Thus, a different approach is needed to be explored that could overcome the addressed difficulty and the utilisation of the K-medoids and Artificial Bee Colony (ABC) seemed to be ideal selections. An overview of the K-medoids and ABC techniques and their applications are presented in the following section.

2.4.2 K-Medoids Clustering Algorithm

K-Medoids is a clustering approach used in the program partitioning around medoids (PAM). This algorithm has the ability to measure the quality of different clusters of the same data set. The idea of using this method for cluster analysis was introduced by Vinod (1969) and was discussed by Rao (1971), Church (1978), and Mulvey and Cowder (1979). The principle is that the $k$ objects should be chosen as the centre of each cluster to minimise the total distance between an object and the representative object of the cluster to which it belongs (Kaufman & Rousseeuw, 1990).

Figure 2-12 shows a typical process of the K-Medoids clustering algorithm, and the specific processes are presented as follows: (1) randomly select $k$ samples as the initial cluster centres; (2) determine the distance (i.e., Euclidean distance, DTW distance) between each object and the centres; (3) assign objects to each cluster according to the minimum distance principle; (4) systematically select each object as a new cluster centre.

**PAM: A Typical K-Medoids Algorithm**

![Diagram of K-Medoids Algorithm](image)
from each cluster and find the optimal solution; and (5) repeat step 2 to step 4 until the cluster centres do not change.

Park and Jun (2009) proposed a new algorithm for K-Medoids clustering, which runs like the K-means algorithm. Comparing to with the original algorithm, the computing time with comparable performance against the partitioning around the centres has been reduced significantly. In addition, the initial selection of centres employed in the proposed approach performs quite well compared with other methods of naively selecting initial centres.

Arora and Varshney (2016) compared two of the most popular clustering algorithms K-Means and K-Medoids in evaluation on dataset transaction 10k of KEEL. The comparison results show that the K-Medoids is better than K-Means in both the time taken for cluster centre selection and the space complexity of overlapping clusters. In addition, K-Medoids is better regarding execution time, is non-sensitive to outliers and reduces noise compared with K-Means as it minimises the sum of dissimilarities of data objects.

Chen et al. (2017) proposed an approach that delineates urban functional areas with building-level social media data with DTW distance based on the K-Medoids method. This study provides an alternative way of characterising the intra-city urban spatial structure, and it is a successful combination of Dynamic Time Warping and K-Medoids.

Rai and Upadhyay (2017) applied K-Medoids as a tool in the assessment of bearing performance degradation. In this study, K-Medoids clustering and empirical mode decomposition (EMD) are encouraged. The results demonstrate that the potential application of K-Medoids clustering as a useful tool for performance degradation assessment of bearings.

Lacko et al. (2017) combined K-Medoids clustering and 3D anthropometry in product sizing methods for a human body. The results suggest that either shape-based or constrained K-Medoids clustering could meet the requirements of the exact shape and size fit for head-based products.

As a fast and straightforward technique for clustering, K-Medoids has an excellent ability for convergence and is robust in partial searches. However, it still can be improved in two areas that include initial centres selection and distance measurement. First, the shortcoming of K-Medoids is that the result depends on the selection of the initial value, but swarm intelligence is an adaptable technique for solving this problem. However,
swarm intelligence techniques are still more popular in soft computing or applied mathematics than engineering especially as utilised in water end-use studies.

2.4.3 Artificial Bee Colony (ABC)

All social insects, including ants, bees, and termites, have shown their impressive collective problem-solving capabilities. Properties associated with their group behaviour, such as self-organisation, robustness, and flexibility, are seen as characteristics that should be exhibited by artificial systems for optimisation, control or task execution. In the last decade, this has been caused many researchers to take social insects as examples; some algorithms have been developed, which are inspired by the strictly self-organised behaviour of these insects. These approaches can be subsumed under the concept of “Swarm Intelligence” (Bonabeau et al., 1999). However, not only insects but also a flock of birds or a school of fish or even an immune system can be seen as a swarm.

Self-organisation and division of labour are two necessary and sufficient properties of behaviour in an intelligent swarm. As a result, a model of honey bee swarms has been developed, which consists of three essential components: food sources, employed foragers, and unemployed foragers. Food sources depend on some factors, such as their proximity to the nest and the richness or concentration of their energy, which together can be represented by the “profitability”. Employed bees can carry and transmit information about a particular food source they are currently exploiting or by which they are “employed”. The unemployed foragers then focus on looking out for a food source to exploit (Karaboga, 2005).

In the ABC algorithm, the colony of artificial bees including three groups of bees: (1) employed bees, which are going to the food sources on their own; (2) onlookers, which are waiting on the dance area to make a decision about choosing a food source; and (3) scouts, which carry out random searches. The goal of the entire colony is to find the source of the most nectar. According to this theory, the ABC algorithm can be summarised into four steps:

Step 1- Initial population:

At the initial stage, a random bee colony is initialised, \( B = \{B_1, B_2, \ldots, B_i, \ldots B_{SN}\} \), which has \( SN \) solutions (food source positions). Each solution \( B_i = \{b_{i1}, b_{i2}, \ldots, b_{iD}\} \) is a vector with \( D \)-dimensions with \( D \) representing the number of parameters. The maximum cycle number is defined as \( N_{max} \). The food source is randomly generated by Eq. 2-1.
\[ B_i^j = B_{\text{min}}^j + \text{rand}(0,1)(B_{\text{max}}^j - B_{\text{min}}^j) \quad j \in \{1, 2, ..., D\} \]  

\[ \text{where } B_i^j \text{ is the } j\text{th dimension of the solution vector;} \]

\[ B_{\text{max}}^j \text{ is the maximum value of the } j\text{th dimension;} \]

\[ B_{\text{min}}^j \text{ is the minimum value of the } j\text{th dimension;} \]

\[ \text{rand}(0,1) \text{ is the uniform random number in the interval } (0,1). \]

**Step 2- Employed Phase:**

In this stage, employed bees randomly select a food source with a new position around the old one and the search process is written as Eq. 2-2.

\[ V_i^j = B_i^j + R_i^j(B_{k}^j - B_i^j) \quad j \in \{1, 2, ..., D\} \text{ and } k \in \{1, 2, ..., SN\} \]

\[ \text{where } R_i^j \text{ is a random number between } [-1, 1]. \]

After default greedy selection of the new food source position, employed bees will determine the fitness function. If the fitness function of the new source is higher than of the previous one, the new position \( V_i^j \) will replace the \( B_i^j \). Otherwise, bees will maintain the previous position.

**Step 3- Onlooker Phase:**

After all the employed bees complete the neighbour search, an onlooker bee will choose a food source depending on the probability value \( P_i \), which is written as Eq. 2-3.

\[ P_i = \frac{F_i}{\sum_{n=1}^{SN} F_n} \]

\[ \text{where SN is the number of food sources and } F_i \text{ is the fitness values of solution } i, \text{ which represents the amount of nectar in the food source. In addition, onlookers will also use greedy selection and follow Step 2 to choose a food source in the neighbour search.} \]

**Step 4- Scout Phase:**

In this stage, if the position of a food source cannot be updated within the time limits with sufficient parameters to abandon a food source, the food source is assumed to be abandoned and the employed bee of this food source will change into a scout. In addition,
the scout will search for a new food source according to Eq. 1 to replace the previous one. The above steps are repeated until the cycle number $N_{max}$ is met.

ABC is one of the more popular techniques in the field of applied mathematics and computation. Karaboga and Basturk (2007) presented the ABC algorithm, which was used for optimising multivariable functions to compare with other algorithms including the Genetic Algorithm (GA), Particle Swarm Algorithm (PSO) and Particle Swarm Inspired Evolutionary Algorithm (PS-EA) through five benchmark functions. The results indicated that the ABC algorithm can get out of a local minimum and can be efficiently used for multimodal function optimisation.

Apart from applications to function optimisation, ABC can also be applied in many other fields. Karaboga and Ozturk (2011) have utilised ABC in the clustering of the benchmark classification problems for classification. The simulation results, which are based on thirteen typical test data sets from the UCI Machine Learning Repository, show that the proposed algorithm can be applied successfully to clustering.

Basu (2013) has developed a promising approach for solving the Multi-Area Economic Dispatch (MAED) problem using the ABC algorithm. In this study, the ABC is used to calculate the minimum cost in different complex areas including generators and multi-fuel sources. The simulation results indicated that the proposed methodology has the ability to converge to a better-quality solution than other comparison algorithms including differential evolution (DE), evolutionary programming (EP) and the real-coded genetic algorithm (RCGA).

Klein et al. (2015) applied the ABC algorithm as a tool for modelling of truck engine powertrain components. The simulation results show that the artificial wavelet neural network approach can be useful and is a promising technique in powertrain components modelling. This proposed technique allows modelling the dynamical behaviour of powertrain components of a truck engine that is combined with the ABC approach.

Sabat et al. (2010) proposed a technique based on the artificial bee colony to extract the small signal equivalent circuit model parameters of a Gallium arsenide (GaAs) metal extended semiconductor field effect transistor device and compares its performance with the (PSO) algorithm. The simulation results show that this algorithm is able to extract successfully the small signal model parameters of metal semiconductor field effect transistor (MESFET). In addition, it also revealed that ABC is more robust than other
similar approaches. Comparing to the PSO algorithm, it has a less relative error between the measured and modelled S-parameters.

2.4.4 Self-organising Maps (SOM)

Apart from the above technique, two popular clustering techniques have been established and applied in the development of water flow trace analysis software, including SOM (Kohonen, 1982) and the K-means clustering algorithm (MacQueen, 1967).

The SOM is a type of artificial neural network with an unsupervised learning process as proposed by Kohonen (1982). This approach has behaviour similar to that of the brain under visual and memory simulation. For example, different cells in the retina can react according to the pattern of input-graphics. This behaviour assures that the neural network has more efficient reactions in separating different patterns. According to this resemblance, the self-organising map can be used to find patterns and perform cluster analysis.

![Figure 2-13 The Kohonen’s feature map (Kuo et al., 2002)](image)

The SOM network consists of an input layer and an output layer (Kohonen layer) as shown in Figure 2-13. The output layer is usually designed as a two-dimensional arrangement of neurons that have the ability to project a high dimensional input to the low-dimensional grid, maintaining the topological order. In this layer, each neuron represents a cluster. In addition, the input layer of nodes is fully connected to the output nodes and owns a weight vector. Euclidean distance and cosine distance are two major criteria between the weight vectors of the output nodes and input nodes.

Moreover, SOM has proven to be a robust approach for the analysis of data as applied to different fields such as engineering, biomedicine, finance, and ecology. Tasdemir and Merenyi (2011) stated that SOM has the ability to group multiple prototypes in each cluster, which is helpful in describing the complex cluster structures. Delgado et al.
(2017) applied SOM to the data-clustering analysis methodology. The experimental results demonstrate that the proposed technique can tackle limitations of parameter settings associated with the clustering methodology.

Kalteh et al. (2008) stated that SOM is a promising technique, which can be applied to investigate, model and control many types of water resources processes and systems. In addition, the number of applications based on the SOM technique has been increasing in water resources problem in recent years.

Mounce et al. (2014) introduced an application of the SOM technique to a water quality problem. The results show that SOM is a useful tool for visual correlation discovery. Blokker et al. (2016) applied SOM technique to investigate the relationship between water ages and various microbial parameters. Laspidou et al. (2015) utilised SOM as a classification approach to cluster consumers according to their water consumption. The results show that this technique can be promising for autonomous classification of water consumers based on urban water demand.

2.4.5 K-means Clustering Algorithm

K-means algorithm is an observational learning algorithm. It is an unsupervised learning that the category or even the number of categories are unknown before clustering. At present, K-means clustering is widely used in statistics analysis, biology, database management and marketing. The basic calculation processes of K-means can be summarized in five steps:

Step 1: Randomly select K elements as the initial cluster centre from all elements.

Step 2: Calculate distance between remaining elements to the centre of the K clusters separately, and classify the elements into clusters with the minimum distance.

Step 3: According to the results of clusters, recalculate the center of each cluster.

Step 4: Re-cluster all elements according to the new centers identified in Step 3.

Step 5: Repeat Step 4 until the clustering results converged and output the results.

As an unsupervised algorithm, the K-means algorithm belongs to a typical approach in clustering analysis. Lloyd (1982) stated the standard algorithm as an approach for pulse-code modulation, but the term “K-means” was proposed by MacQueen (1967). The central theory of the algorithm is to divide the data set into different categories through
an iterative process; furthermore, the similarity between objects within each generated category is high, and the similarity of objects between categories is low.

Jain (2010) reported that K-means is one of the most popular and straightforward clustering algorithms. Although it was proposed over 50 years ago and thousands of clustering algorithms have been created since then, K-means is still widely used in data clustering. However, the number of application based on the K-means clustering algorithm in water end-use studies is still small. Hence, K-means clustering was utilised as a techniques for enhancing *Autoflow v2.1* as described in Chapter 5.

2.5 Proposed Techniques

According to disadvantages of using clustering algorithms individually, the ABC swarm intelligence will be applied in this study to reduce the impact of the initial centres. Second, the original distance measuring in the K-Medoids algorithm is the Euclidean distance based on points in Cartesian coordinates. There has been some research that is based on the Temporal-Proximity-Based Clustering Approach, as shown in Table 2-2. However, due to the peculiarity of water end-use studies, the dataset always is flow trace data based on a time series. Therefore, the traditional distance measuring will be replaced by the DTW algorithm for measuring the distance between two water end-use events. In addition, The SOM algorithm has the advantage of producing natural results (Dorffner, 2001) but has two limitations related to its static architecture and capabilities to characterise hierarchical relations within the data (Rauber et al., 2002). Hence, two hybrid techniques including (1) Artificial Bee Colony, Dynamic Time Warping, and K-Medoids clustering and (2) SOM and K-means are proposed in this study for water end use clustering.

2.6 Chapter Summary

This chapter provides a review of the literature pertinent to the topics surrounding the water end-use study, water metering process, existing water flow trace analysis system and existing water end-use pattern-recognition techniques. For Australia, it is pertinent that a secure water supply be secured for growing populations and to protect against climate change through the management of water. While several water flow trace analysis systems have been undertaken, all of them still have room for improvement, as explained above. Therefore, the development of an autonomous flow trace analysis system based on pattern-recognition techniques, is needed. This literature review provides the required knowledge in all relevant areas for the development of such a system or technique. The
chapter concludes with a summary of the current gaps in different techniques to tackle the problem faced. The research method adopted to undertake this research is detailed in Chapter 3.
Table 2-2 Summary of Temporal-Proximity-Based Clustering Approach (Rani and Sikka, 2012).

<table>
<thead>
<tr>
<th>Clustering criteria</th>
<th>Clustering Algorithm</th>
<th>Authors</th>
<th>Journal (Year)</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Density based Subsequence Clustering</td>
<td>Hardy Kremer</td>
<td>ICDMW (2010)</td>
<td>Detecting climate change</td>
</tr>
<tr>
<td></td>
<td>Kernal DBScan</td>
<td>S. Chandrakala</td>
<td>IJCNN (2008)</td>
<td>Multivariate time series</td>
</tr>
<tr>
<td></td>
<td>Modified k-means</td>
<td>M. Vlachos</td>
<td>SIAM (2003)</td>
<td>Non-specific</td>
</tr>
<tr>
<td></td>
<td>Modified k-means</td>
<td>C. Guo</td>
<td>WiCOM (2008)</td>
<td>Real world stock</td>
</tr>
</tbody>
</table>
3. Research Methodology

This chapter presents the methodology adopted for the current study, specifically the research approach, design and analytical techniques adopted to meet the requirements of the research objectives. Two hybrid techniques, which have different applications (i.e., (1) for mechanised water end-use analysis, and (2) for optimisation of water end-use analysis process), were selected as the most suitable approaches to achieve the research objectives. The method framework highlighting the three project phases was presented in Figure 1-1. This chapter details the models and a discussion follows regarding the three primary research phases and the individual stages within each phase of the research.

3.1 Data Collection and Analysis

3.1.1 Data collection and processing

The flow trace data for developing the technique is collected from 252 residential households in the urban south-east corner of the State of Queensland (SEQ), Australia, which installed a new water-management system including smart meters and data loggers from 2010 to 2011. (Beal & Stewart, 2013). The smart water meters can provide a high resolution (0.014L/pulse/5s) of water-flow data to complete a water end-use or micro-component disaggregation process (i.e., tap, shower, and clothes washer) (Nguyen et al., 2013).

As the first step, artificial data including 2 to 5 different types of water end-use categories, needed to be fabricated. The following Figure 3-1 shows a typical format of the script from the data logger, which includes a code from the data logger to be used for different homes, data and times of recording, the intervals between two consecutive records, and the corresponding flow rates passing through the water meter (Nguyen et al., 2013).
In Figure 3-1, this flow rate means that the water meter has recorded the number of pulses in a duration of five seconds. The resolution of the water meter used in this study is 71.5 pulses per litre. Hence, letting $q$ represent the actual flow rate (L/min), $m$ is the resolution of the water meter (pulse/litre), and $s$ is the sampling interval (second), the value of $q$ can be written as Eq. 3-1.

$$q = \frac{60}{m \cdot s} \left( \frac{L}{min} \right)$$  \hspace{1cm} (3 - 1)$$

where $m$ and $s$ are 71.5 and 5 respectively. Therefore, 1 pulse is equal to 0.1678 litre per minute ($q=0.1678$ (L/min/pulse)). For example, in the fourth row of the table, the actual flow rate $q_4= 33 \times 0.1678=5.5374$ (L/min).

### 3.1.2 Residential water end-use pattern analysis
The primary objective of the end-use pattern-analysis step was to utilise the existing water flow trace analytical system to discover the distinct patterns of each end-use category, thus accumulating analytical experience for later research. Within this part of the chapter, typical examples of each category were displayed so that their main characteristics could be examined.

- Shower

<table>
<thead>
<tr>
<th>Code</th>
<th>Date</th>
<th>Time ($t_i$)</th>
<th>Duration</th>
<th>Flow rate ($a_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>07139S218,</td>
<td>14/06/10</td>
<td>19:14:10,</td>
<td>5,</td>
<td>0</td>
</tr>
<tr>
<td>07139S218,</td>
<td>14/06/10</td>
<td>19:14:15,</td>
<td>5,</td>
<td>1</td>
</tr>
<tr>
<td>07139S218,</td>
<td>14/06/10</td>
<td>19:14:25,</td>
<td>5,</td>
<td>5</td>
</tr>
<tr>
<td>07139S218,</td>
<td>14/06/10</td>
<td>19:14:45,</td>
<td>5,</td>
<td>33</td>
</tr>
<tr>
<td>07139S218,</td>
<td>14/06/10</td>
<td>19:14:50,</td>
<td>5,</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 3-2 A standard shower event

Figure 3-2 shows a standard shower stall event, which does not exhibit the high flow rate at the beginning or end of the shower event; also, it tends to have a relatively consistent flow rate through the duration of the shower event.

- Tap

Figure 3-3 Examples of tap event
Tap usage is the most common water end-use encountered when undertaking a residential water consumption analysis. Generally, there are two tap type events; those having a different flowrate, while most faucet duration is relatively brief (Figure 3-3).

- **Bathtub**

![Figure 3-4 Example of bathtub event](image)

Similar to shower and tap events, the bathtub is one of the water end-use, which heavily depend on human habits. The volume, duration, and flowrate of each bath event relies on water device that according to the individual requirements. Figure 3-4 only shows a general example of a bathtub event.

- **Irrigation**

![Figure 3-5 Example of irrigation event](image)

Due to the various patterns, irrigation is also a complicated category to identify. Both automatic and manual sprinkler can process a different pattern, which depends on human
behaviour. Figure 3-5 shows a standard irrigation event, which lasted more than 7 minutes, with a relative constant flowrate.

- Clothes washer

\[\text{Figure 3-6 Example of clothes washer events}\]

The example in Figure 3-6 shows a clothes washer with similar peak flowrates. Generally, a peak flowrate range of 14 to 18 litres per minutes in the present study. Due to the preset series of wash functions, a typical water usage model of clothes washer can be found in the study and this type model is identified as mechanical water end-use category.

- Dishwasher

\[\text{Figure 3-7 Example of dishwasher events}\]

Similar to clothes washer events, dishwasher events also have similar water use patterns, as displayed in Figure 3-7. Each dishwasher event typically uses between 24.6 and 45.4 litres to wash a series of dishes, and a peak flowrate range of 3.6 to 7.4 litres per minutes.
• Toilet

Figure 3-8 Example of toilet event

Due to a different type of toilet devices, toilet events analysis is a complicated issue. While the time of holding the flush button down is according to the user behaviour, the amount of water consumption is relatively constant. Therefore, toilet event can also be identified as a mechanical water end-use category, as presented in Figure 3-8.

3.2 Overview of the Research Method

This research section in the current study aims to identify and understand the existing techniques, including the strengths and weaknesses of each, used to analyse the water consumption study. The application development section then aims to describe all tasks to be finished, including the new hybrid techniques to improve the category accuracy of water end-use events, verification of the efficiency of the independent data, and evaluation of the techniques for progress in the future.

Figure 1-1 shows that the methodology and activities, which are separated into three distinct phases. The first phase was conducting the knowledge acquisition required for the present research. The second phase was a new proposed method, which applied to toilet events analysis, including ABC, DTW, and K-Medoids techniques. The last phase was combining the hybrid technique with Autoflow to improve the category accuracy based on two different verification processes.

3.2.1 Phase 1: Knowledge acquisition

The main task of Phase 1 was to start a review of the literature, to formulate the research objectives, and to design the most suitable methodology to tackle the problems, which relates to the present study.
For determining the current gaps in the field of research, an extensive review of the literature was carried out to accumulate knowledge, including water end-use studies, the water-metering process, existing water flow trace analytical system, and existing available techniques for their development. Chapter 2 details the topics explored in the literature, moreover, Chapters 4 and 5 provides concise literature reviews associated with the refereed publications. After determination of the current gaps, the research objectives were formulated according to the literature review. Then, additional literature was reviewed to design the most suitable methodology to meet the requirements of the research. The completion of each stage in Phase 1 was necessary before starting Phases 2 and 3.

3.2.2 Phase 2: Mechanical event analysis

Phase 2 involved the development of a hybrid technique in performing a particular single events from mixed residential water-flow data. The phase presented a study of the required techniques and methodology for the mechanical event classification, the outcome of which was tested and published in a journal paper.

A hybrid combination of the ABC, DTW and K-Medoids algorithm was proposed to tackle this issue. The accuracy of the overall technique was verified in the independent dataset, with an average accuracy of greater than 95.71%. The trials of the existing pattern-recognition techniques were designed purposely to apply the existing pattern-recognition techniques, including DTW and K-Medoids, to the present problem to evaluate their strengths and weaknesses.

- Dynamic Time Warping algorithm

DTW has been known for its power to assess the similarity of two sequences based on time series with different lengths and magnitude in pattern matching issues. However, it has a limitation, as in this study, with the results showing that the accuracy changed with the selection of threshold value.
In experiments that only utilised the DTW algorithm, the reference event is the same. However, the threshold value is changed from 100 to 1000 and picked number from mixed data is changed, as shown in Figure 3-9. Table 3-1 shows that this single technique has a low recognition accuracy of below 80%.
Table 3-1 The categorisation accuracy with different threshold values.

<table>
<thead>
<tr>
<th>Threshold value</th>
<th>Total events</th>
<th>Total toilet events</th>
<th>Picked toilet event number</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>136</td>
<td>70</td>
<td>8</td>
<td>11.43</td>
</tr>
<tr>
<td>200</td>
<td>136</td>
<td>70</td>
<td>15</td>
<td>21.43</td>
</tr>
<tr>
<td>400</td>
<td>136</td>
<td>70</td>
<td>35</td>
<td>50.00</td>
</tr>
<tr>
<td>500</td>
<td>136</td>
<td>70</td>
<td>47</td>
<td>67.14</td>
</tr>
<tr>
<td>750</td>
<td>136</td>
<td>70</td>
<td>51</td>
<td>72.86</td>
</tr>
<tr>
<td>1000</td>
<td>136</td>
<td>70</td>
<td>55</td>
<td>78.57</td>
</tr>
</tbody>
</table>

In this project, each water end-use category may have more than thousands of single events. In the same category, events may have different parameters according to the user’s behaviour. Duration is one of the most important parameters to a water end-use event. The purpose of the present study is to cluster events, which have a similar pattern but have different duration, and classified different categories. The Euclidean distance is normally used for measuring the distance between two points in two dimensions space. However, the distance between two water end-use events is considered as the distance between two sequences. Hence, DTW algorithm is more suitable as a distance measurement method to combine with K-Medoids clustering.

- K-Medoids clustering algorithm

Another trial of the existing pattern-recognition techniques is based on the K-Medoids algorithm. As a simple and fast technique for clustering, K-Medoids has a good ability for convergence and is robust in partial searches. However, the results can be changed due to the selection of the initial centres. There is an experimental result of testing that included 100 replicates with the number of initial centres equal to 5, showing that the positions of the centres are changed with the number of tests performed due to the random selection of initial centres within this technique, as shown in Figure 3-10.

In this project, each water end-use event is considered as a single point in a certain range (e.g. all water end-use events in the day). Every water end-use event should be grouped into different water end-use categories. Cluster analysis is a technique to divide similar objects into different groups or more subsets by static classification. Therefore, this project can be considered as a clustering model. Cluster analysis including hierarchical methods, partition-based methods, density-based methods, and model-based methods. K-medoids clustering is an algorithm of partition-based methods. The theory of this
algorithm is that a lot of points through the K-medoids clustering and the results are that
the points in the same group are close enough, and the points between groups are far
enough. For hierarchical methods, this type of cluster algorithm does not have a global
objective function like the partition-based methods and the clustering operation is
irreversible, this will lead to a high cost of computing storage. For density-based methods,
this type of algorithm is not sensitive to high dimensional data, in the project, all water
events are presented by high-dimensional data. For model-based methods, there are two
main models including generative model and neural network model. Gaussian Mixture
Models is a representative model of generative model and Self-organising Maps is a
model which based on the neural network model. In addition, Self-organising Maps is
also applied in the second hybrid technique and will be introduced in the next part.

In the combined technique, cluster number setting is only in K-Medoids clustering
algorithm. And each cluster represents a particular water end-use category. In general,
there are 8 water end-use categories in a household (e.g. clothes washer, dishwasher,
toilet, evaporative cooler, shower, tap, bathtub and irrigation). According to the input
dataset, there are different numbers of cluster in each experiment. However, if the number
of categories is unknown, we just set a maximum value and some clusters will not have
events in the results. In Figure 3-10, there is a result shows the limitation of using K-
Medoids clustering algorithm individually. The reason of choosing 5 initial clusters is
that this experiment only focuses on mechanical water end-use event categories (e.g.
clothes washer, dishwasher, half flush toilet, full flush toilet and evaporative cooler).

(a) Position of center 1
Figure 3-10 Position of initial centers with different tests
**Figure 3-11** Sketch map of hybrid pattern recognition techniques

**Figure 3-12** Sketch map of hybrid pattern recognition techniques
For solving these of problems, ABC was considered as the method with the greatest potential to reduce the influence of the weaknesses when each technique was used alone. Several attempts have been made to build a primary study for toilet events analysis, based on the ABC technique. Moreover, a hybrid combination of ABC, DTW, and K-Medoids resulted in an advanced model, as shown in Figure 3-11 and Figure 3-12. Its accuracy was confirmed by verification of the independent data, which was greater than 95%. However, the present technique still has limitations. When this hybrid technique is applied to other types of residential water end-use events, such as the shower, bathtub, and washing machines, the proposed technique had a relatively higher processing time due to the datasets having a time series structure. For developing a technique that has a good performance on most types of residential water end-use event, a second hybrid technique was proposed in Phase 3.

3.2.3 Phase 3: Single event analysis

For the intelligent water flow trace software to be widely accepted, the ability to disaggregate residential water flow trace into different single end-use categories plays an important role. In addition, the categorisation accuracy is the criteria of this type of system. This chapter presents a development of previous software Autoflow described in Chapter 2. It was tested on a larger number of water consumption data and independent household samples, with above 85% accuracy on each category (i.e., shower, clothes washer, and dishwasher). A brief introduction and discussion of the development of its technology are presented below.

Similar to the development process of the toilet event analysis module, the trial on the previous hybrid technique (ABC, DTW, and K-Medoids) shows that this technique has a good efficiency and accuracy for toilet events, one type of small mechanical water end-use; however, for others water end-use events, the efficiency and accuracy have reduced significantly. However, we applied this technique to other events, which have long duration and max flowrate, such as shower, bathtub, and irrigation (long duration), tap (high max flow rate). The efficiency has significantly reduced. Such as for analysing 1000 clothes washer events, the processing time is continued about 4-5 hours. Therefore, we tried algorithms such as SVM (support vector machine), and EM (Expectation Maximization). Finally, we choose the SOM and k-means algorithms and combined them in a new hybrid technique, which has a good efficiency and accuracy in all residential water end-use categories.
For improving the final accuracy of the *Autoflow* software, the major task is to improve the accuracy of the initial pre-grouping stage. The new module, named *Autoflow v3.1*, applied Self-organising maps (SOM) and K-means clustering algorithms to initially pre-group the water end-use events to replace the existing DTW algorithm. Once this stage is successfully developed, the final category accuracy can show an increase. The overall procedures with SOM+K-means in *Autoflow* is shown in Figure 3-13, and each analytical process is shown in Figure 3-14. Two verification processes, including a large amount data and thirty independent homes in Australia, were conducted. The outcomes are explained clearly in Chapter 5.

![Diagram](image_url)

**Figure 3-13** Overall water end-use event classifications procedures with SOM+K-means
3.3 Chapter Summary
This chapter presented the overarching research design and each aspect of the proposed technique to satisfy the research objectives. The proposed approaches were appropriate to strengthen the research design and to address all the objectives. This section includes the gain in knowledge of several pattern-recognition techniques, such as ABC, SOM, DTW, K-Medoids, and K-means while identifying the study approach, as well as an understanding of the past research conducted world-wide. Each refereed publication in Chapters 4 and 5 also provides a detailed description of the specific approach applied.
4. Hybrid Intelligent Model for Mechanised Water End-use Analysis

Statement of contribution to co-authored published paper
This chapter includes an abbreviated co-authored paper accepted for publication.

The bibliographic details of the paper, including all authors, are:


My contribution to the paper involved: literature review, development of hybrid technique using ABC, DTW, and K-Medoids algorithms, verification of the proposed method, figures and tables, writing and editing.

Date: 26/09/2018
Ao Yang

Date: 26/09/2018
Supervisor: Hong Zhang (Principal supervisor and co-author)

Date: 26/09/2018
Supervisor: Rodney A. Stewart (Associate supervisor and co-author)

Date: 26/09/2018
Supervisor: Khoi A. Nguyen (Associate supervisor and co-author)
Hybrid Intelligent Model for Mechanised Water End-use Analysis

Water end use clustering using hybrid pattern recognition techniques - Artificial Bee Colony, Dynamic Time Warping and K-Medoids Clustering

Abstract: The smart water meter collected data has made a great progress for the categorization of residential water end use events, the efficiency and accuracy still need to be improved. In this paper, an advanced algorithm is proposed for clustering the end-use category of a mechanical appliance. For this study, the database of end use events was collected using smart meters from over 200 households located in South-east Queensland (SEQ), Australia. Firstly, the raw data is pre-processed and physical characteristics (e.g., volume, duration, max flowrate, etc.) are extracted. Due to the type of the dataset is water end used flow data, which based on time series, a K-Medoids clustering algorithm based on the Dynamic Time Warping algorithm is used for clustering. In addition, a swarm intelligence which is named Artificial Bee Colony algorithm brings the whole system into equilibrium. Numerical experiments are based on toilet flushing events. Results indicate that the hybrid technique improves the clustering accuracy from 82.85% to 95.71%, and it can be implemented to other mechanical water end use events such as clothes washers and dish washers.

4.1 Introduction

Pattern recognition has been a popular research field over the past few decades, such as character recognition, speech recognition and medical applications. With the advanced development of smart water meter technology, extensive water end use data has been collected and employed for various studies since 1998 such as USE Investigation of domestic water end use, Yarra Valley Water Residential End Use study and WA Water Corporation Domestic Water Use study. Most recently, Nguyen et al., (2013) developed an integrated intelligent pattern recognition model to automate the categorisation of residential water end-use events.

For most of existing water metering systems, the collected data, usually in quarterly basis, cannot provide real-time information or other management information about water service. To overcome this issue, a smart water management system based on smart metering technology was proposed which would provide real-time information that showed the how, when and where water is consumed for both user and the water utility (Nguyen et al., 2013). Nguyen et al., (2011) applied Dynamic Time Warping (DTW) algorithm to categorise the water end use events. And one of the major parts of water end use classification task is water event clustering. In this part of research, if existing algorithms such as simple DTW algorithm, simple K-Medoids clustering and Artificial
Neural Network (ANN) were applied to clustering water events, the accuracy and the efficiency are not enough to meet the needs of both users and water utilities. Nguyen et al., (2015) introduced an intelligent autonomous system for residential water end use classification which combines Hidden Markov Model (HMM), Artificial Neural Networks (ANN) and the Dynamic Time Warping (DTW) algorithm.

The purpose of this study is to develop an improved intelligent technique to cluster similar water end use events together. This hybrid technique consists of Artificial Bee Colony (ABC), Dynamic Time Warping algorithm (DTW) and K-Medoids Clustering. Its advantages include: (i) the ABC algorithm combines with K-Medoids Clustering, which has strong local search capability, could improve the comprehensive performance of clustering algorithm, (ii) the K-Medoids clustering based on DTW distance can group the mixed water events into different clusters according to the pattern characteristics, and (iii) the hybrid algorithm has the ability of global searching and optimization.

4.2 Intelligent Techniques for Time Series Clustering

4.2.1 Overview of time series clustering techniques

Rani and Sikka (2012) stated that time-series clustering is one of the concepts of data mining and listed some techniques that are used in time series clustering and depend on distance measuring. For distance measuring, the Euclidean and DTW are two popular techniques. For times series clustering, K-Medoids is the most frequently occurring algorithm. However, these approaches have not yet been applied in water end use analysis. Han et al., (2001) found that clustering is a process of a set of objects moving into clusters and the characteristic is objects which in one cluster are similar but they are dissimilar to objects from other clusters. There are many algorithms for K-Medoids clustering but the partitioning around medoids (PAM) which was proposed by Kaufman and Rousseeuw (1990) is the most popular approach. However, Han et al., (2001) pointed out that the PAM algorithm needs a long computational time for a large data set, so the efficiency is low. However, Park and Jun (2009) reported a new algorithm of K-medoids clustering which is simple and fast and has been tested in many different type of data sets. K-Medoids clustering is one of the centroid-based clustering or rather, an improved algorithm of K-means. Its advantages include: (i) K-Medoids algorithm has the ability to process a large dataset. (ii) K-Medoids algorithm can reduce the effect of the outliers on the results. However, K-Medoids clustering also has advantages such as the influence of initial medoids being strong and the ability of global searching being poor.
4.2.2 Data clustering in water end use analysis

DTW algorithm is a method for measuring the similarity between two time series with different lengths. In this task, each water end-use event can be defined as a time sequence and the objective is to find the distance between two events. Nguyen et al., (2015) demonstrated an application of DTW algorithm for prototype selection and similar event clustering in the water end use classification task. However, the disadvantage of this technique was the heavy dependence on the use of threshold value, which defined the similarity between two samples. A large threshold value would allow two significantly different samples to be considered similar and clustered into the same group, while a small threshold value may assign two relatively similar samples into two different groups. As a consequence, different threshold value settings will result in different number of events in each cluster. The different number of events which are considered to be similar based on two different threshold value settings is shown in Figure 4-1.

![Figure 4-1 Result of DTW algorithm](image)

**Figure 4-1** Result of DTW algorithm (a) Threshold value=700, Identified events=43. (b) Threshold value=350, Identified events=56)
4.3 Hybrid Pattern Recognition Techniques

To overcome the disadvantage of DTW, a combination of K-Medoids, DTW algorithm and Artificial Bee Colony algorithm have been proposed where each approach will be in charge of a significant role in the whole algorithm. Presented below is a brief summary of each technique.

4.3.1 Dynamic time warping algorithm

The main application of DTW algorithm is to determine the distance between two events which have a similar pattern. The processes of this algorithm include following three steps:

**Step 1:** Suppose that there are two time sequences of the water flow data $Q = (q_1, q_2, ... q_i, ... q_m)$ and $C = (c_1, c_2, ... c_j, ... c_n)$ of length $m$ and $n$ respectively. Define $d(q_i, c_j) = |q_i - c_j|$, and $D(q_i, c_j)$ as the total DTW distance between $q(1:i)$ and $c(1:j)$ with warping path from $(1,1)$ to $(i,j)$, which follows $(1 \leq i \leq m)$ and $(1 \leq j \leq n)$.

**Step 2:** Accumulated distance using DTW algorithm is written as Eq. 4-1.

$$D(q_i, c_j) = d(q_i, c_j) + \min D$$

(4 - 1)

The $\min D$ represents the minimum distance between two nodes in warping path and is written as Eq. 4-2

$$\min D = \{D(q_{i-1}, c_j), D(q_{i-1}, c_{j-1}), D(q_i, c_{j-1})\} \text{ and } 1 \leq i \leq m \text{ and } 1 \leq j \leq n$$

(4 - 2)

The mapping path is from $(q_1, c_1)$ to $(q_m, c_n)$ and $D(q_1, c_1) = d(q_1, c_1)$ are the initial conditions of this algorithm. In addition, the Figure 4-2 as shown below presents 3 directions in warping path and explains the Eq. 4-2 clearly.

![Figure 4-2 Three directions of grid node in warping path](image)
Step 3: The final Dynamic Time Warping distance between \( Q \) and \( C \) is \( D(q_m, c_n) \).

In this study, the main objective of this algorithm is to determine the distance between two different water end-use events and this distance as a benchmark in K-Medoids clustering.

4.3.2 K-Medoids clustering

As an unsupervised learning, clustering plays an important role in pattern recognition. In this study, the dataset is the flowrate time series, and distance measurement is based on DTW distance. Here, the K-Medoids clustering is divided by five steps:

Step 1: Randomly select \( K \) samples as the initial cluster centroid from \( n \) objects which represents by \( C_1, C_2, \ldots C_j, \ldots C_k \).

Step 2: Determine the DTW distance between each object and medoids following the procedure in Section 3.1. Here, the equation of DTW distance is written as Eq. 4-3.

\[
d(ij) = dtw\{O(i), C(j)\} \quad 1 \leq i \leq n \text{ and } 1 \leq j \leq k
\]  

(4-3)

where \( d(ij) \) represents the DTW distance between object \( i \) and medoid \( j \).

Step 3: Assign objects to each cluster according to the minimum distance principle. Such as \( O(i) \in \text{Group}\{\min\{d(ij)\}\} \).

Step 4: Select each object as new cluster medoid \( C_i \) (\( 1 \leq i \leq n \)) orderly from each cluster and find the optimal solution. In this algorithm, the sum of DTW distance between objects and medoid is the criteria of the optimal result. The new cluster centroid \( C_i \) which have minimum result of cost function will replace the initial cluster centroid \( C_j \).

Step 5: Repeat step 2 to step 4 until the cluster centroid does not change.

4.3.3 Artificial Bee Colony (ABC) algorithm

ABC algorithm is an optimisation technique based on the foraging behaviour of honey bee in swarm intelligence. Karaboga (2005) proposed the ABC algorithm, and in the beginning, applied it in solving function. Then, Karaboga and Basturk (2007; 2008) pointed out that compared with other swarm intelligence such as particle swarm optimization (PSO) and evolutionary algorithm (EA), the ABC algorithm has higher efficiency. The specific description of this algorithm is: employed bees onto food source and sharing the information with onlookers, if there is a food source abandoned by employed bees or onlookers, the employed bee of this food source will turn into a scout.
The mission of scouts is to find the new food source. According to this theory, the ABC algorithm can be divided into 4 steps:

**Step 1: Initial Population:** A random bee colony is initialised, \( B = \{ B_1, B_2, ..., B_i, ..., B_{SN} \} \) which has SN solutions (food sources). Each solution \( B_i = \{ b_{i1}, b_{i2}, ..., b_{iD} \} \) is a vector with D-dimension and D represents the number of parameters. The food source is randomly generated by the Eq.4-4.

\[
B_i^j = B_{\min}^j + r \cdot (B_{\max}^j - B_{\min}^j) \quad j \in \{1, 2, ..., D\}
\]

where, \( B_i^j \) is the jth dimension of the solution vector; \( B_{\max}^j \) is the maximum value of the jth dimension; \( B_{\min}^j \) is the minimum value of the jth dimension; \( r \) is the uniform random number in interval (0,1).

In addition, the maximum cycle number is defined as \( N_{\max} \).

**Step 2: Employed Phase:** In this step, employed bees randomly select a new position of food source around the old one and the equation of search new food source is presented as Eq. 4-5.

\[
V_i^j = B_i^j + R_i^j (B_i^j - B_k^j) \quad j \in \{1, 2, ..., D\} \text{ and } k \in \{1, 2, ..., SN\}
\]

where, \( R_i^j \) is a random number between [-1, 1].

After greedy selection of new position of food source, employed bees will determine the fitness function, if the fitness function of the new source is higher than that of the previous one, the new position \( V_i^j \) will replace the \( B_i^j \). Otherwise, bees will keep the old position \( B_i^j \).

**Step 3: Onlooker Phase:** After all the employed bees complete the neighbour search, an onlooker bee will choose a food source depending on the probability value \( P_i \), written by the following equation:

\[
P_i = \frac{F_i}{\sum_{n=1}^{SN} F_n}
\]
where SN is the number of food sources and $F_i$ is the fitness value of solution $i$ which represents the nectar amount of the food source. In addition, onlookers will also use greedy selection and follow Step 2 to choose food source in neighbour search.

**Step 4: Scout Phase:** In this step, if a position of food source cannot be updated within the *limit* times which parameters in order to abandon food source, this food source is assumed to be abandoned and the employed bee of this food source will change into a scout. In addition, the scout bee will search a new food source according to Equation 4-4 to replace the previous one. Repeating the above steps until cycle numbers meet the $N_{max}$.

### 4.3.4 Hybrid algorithm implementation to water end use analysis

K-Medoids clustering is an algorithm which has advantages such as being simple, fast for convergence and robust for partly searching. However, it can be improved because the dependence on the initial medoid is strong and the ability of global research is poor. For this reason, a swarm intelligence needs to be used in this study to reduce the impact of the initial medoid on the combined algorithm. This section will explain how these three techniques have been integrated to improve the clustering efficiency. The basic steps of techniques are described below and the hybrid techniques operation flow chart is summarized in Figure 4-3.

![Figure 4-3 Improved technique flow chart](image-url)
Step 1: The first required step of the hybrid algorithm is to initialize parameters and select the initial medoids. The number of bees is \( N_B \), and the first half of the colony is employed bees and the second half consists of onlookers. The number of cluster is equal to \( M \). The maximum iteration number and control parameter are \( N_{\text{max}} \) and \( L \) respectively.

Step 2: Each bee is employed to search food source using the Equation (4), then the probability \( P_i \) is calculated when all employed bees finish their searching. Every onlooker will choose their employed bees according to the \( P_i \) and employed bees will find a new source again.

Step 3: If the fitness function a food source is not the best result for the whole colony after \( L \) iteration, the employed bee will change into a scout and find a new source using the Equation (3). In this study, the fitness function is written by:

\[
F_i = \sum_{j=1}^{M} \sum_{x \in C_j} DTW(x, C_j)
\]

where \( F_i \) represents the sum of DTW distance between objects, \( x \) and medoid of the \( j^{\text{th}} \) cluster, \( C_j \) in \( i^{\text{th}} \) iteration.

Step 4: Using the result of Step1 as the initial medoid and operate a K-Medoids clustering for the dataset. A new group of medoids will be calculated from this procedure and use these medoids to update the bee colony. If the number of iterations is less than \( N_{\text{max}} \), repeating the above steps. Otherwise, the hybrid algorithm will stop and get the optimal medoids.

Step 5: As the initial points, these optimal medoids are used in the K-Medoids clustering and each cluster can be determined.

4.4 Experimental Results and Discussion

The water end use datasets utilised for hybrid algorithm were from the South-east Queensland Residential End Use Study (SEQREUS). The data is recorded by high-resolution smart water meter, which collect 0.014L/pulse water consumption data every 5 seconds, from over 200 homes and analysed by Trace Wizard™ (Nguyen et al., 2013). The mixed water events consist of nine different water categories, including shower, tap, dishwasher, clothes water, bathtub, irrigation, leak, full-flush toilet and half-flush toilet. In this experiment, a dataset of 136 water end use events, including 70 toilet events mixed with 66 events from other categories, was used to test the hybrid algorithm. The overall
Hybrid Intelligent Model for Mechanised Water End-use Analysis

The objective is to group all toilet events together, and the method efficiency is evaluated based on the number of toilet events grouped together over the number of total toilets event present in the dataset. For the Artificial Bee Colony algorithm in proposed algorithm, the parameters are set as follows, which are the colony size $N_B = 100$, the maximum cycles number $N_{max} = 500$ and the $L$ is equal to 200.

4.4.1 Hybrid algorithm implementation to water end use analysis

The convergence graph of hybrid algorithm, which is based on the dataset, is shown in Figure 4-4. K-Medoids clustering makes the algorithm quickly reach the local extremum value and the Artificial Bee Colony allows the algorithm to escape the local maxima and reach the global optimal value. The results will not change in a large number of experiments. This result also shows that this hybrid algorithm has a good effect in convergence.

![Figure 4-4 Changes of fitness values in testing dataset](image)

The classified results of Dynamic Time Warping algorithm and hybrid techniques are shown in Figure 4-5.

![Figure 4-5](image)
Figure 4-5 (a): Toilet events categorized using Dynamic Time Warping algorithm. (b): Toilet events categorized using proposed hybrid algorithm.

Figure 4-5(a) shows that 58 toilet events were grouped together using the DTW algorithm alone as presented in (Nguyen et al., 2011), while the proposed algorithm has successfully clustered 67 toilet events. The overall testing results was presented in Table 4-1 below.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Identified events</th>
<th>Total toilet events</th>
<th>Testing Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTW algorithm</td>
<td>58</td>
<td>70</td>
<td>82.85%</td>
</tr>
<tr>
<td>Hybrid techniques</td>
<td>67</td>
<td>70</td>
<td>95.71%</td>
</tr>
</tbody>
</table>

4.4.2 Discussion

Compared to the Dynamic Time Warping algorithm, the hybrid algorithm is capable of maintaining the accuracy and also improving the efficiency. Due to combination of ABC algorithm, clustering results are relatively stable. When applying this technique, the distance from medoids to all other samples are required to be determined many times until the whole system is in equilibrium. The condition of equilibrium is the sum dynamic time warping distance of each cluster is the minimum, which represents the main theory of clustering.

4.5 Conclusion and Future Work

The study presented an improved pattern recognition technique in water end-use events analysis. This hybrid algorithm can also be used as a part of an improved technique for improving the smart water metering analysis system. In addition, an out-of-order dataset can be classified and better analysed through this hybrid technique. In the testing dataset,
the results of experiments show that this algorithm has high efficiency, accuracy, and practicability. However, the limitation of this method is that it still requires an initial selection of the number of clusters prior to running the algorithm, and if this number is set inappropriately, the final clustering result will be reduced. In conclusion, the proposed algorithm is recommended for all the problems which are based on time series analysis.

In the analysis of water end use, toilet event is a part of mechanical processes, clothes washer and dish washers are the same type end use as toilet flushing events and they have also a high degree of pattern and periodicity. The next task is to apply this hybrid algorithm in these two water end uses and other energy pattern analyses such electricity or natural gas.
5. Optimising Water End-use Analysis Process with Self-organising Maps and K-means Clustering

Statement of contribution to co-authored published paper

This chapter includes an abbreviated co-authored paper accepted for publication.

The bibliographic details of the paper, including all authors, are:


My contribution to the paper involved: literature review, development of a methodology to perform a single event pre-grouping process, verification of the proposed method, figures and tables, writing and editing.

Date: 26/09/2018
Ao Yang

Date: 26/09/2018
Supervisor: Hong Zhang (Principal supervisor and co-author)

Date: 26/09/2018
Supervisor: Rodney A. Stewart (Associate supervisor and co-author)

Date: 26/09/2018
Supervisor: Khoi A. Nguyen (Associate supervisor and co-author)
Abstract: The aim of residential water end-use studies is to disaggregate water consumption into different water end-use categories (i.e. shower, toilet, etc.). The authors previously developed a beta application software (i.e. *Autoflow v2.1*) that, provides an intelligent platform to autonomously categorise residential water consumption data and generate management analysis reports. However, *Autoflow v2.1* software water end use event recognition accuracy achieved was between 85-90%, leaving room for improvement. In the present study, a new module augmented to the existing procedure improved flow disaggregation accuracy, resulting in *Autoflow v3.1*. The new module applied Self-organising maps (SOM) and K-means clustering algorithms for undertaking an initial pre-grouping of water end-use events before the existing pattern recognition procedures were applied (i.e. ANN, HMM, etc.) For validation, a dataset consisting of over 100,000 events from 252 homes in Australia, were employed to verify accuracy improvements derived from augmenting the new hybrid SOM and K-means algorithm techniques into the existing *Autoflow v2.1* software. Water end use event categorisation accuracy ranged from 86% to 94.2% for the enhanced model (*Autoflow v3.1*), which was a 1.7% to 9% improvement on event categorisation.

5.1 Introduction
The urban population is expected to increase by current 54% to 66% in 2050 (Gerland et al., 2014). Population growth and higher per capita water usage presents a challenge to the future water demand (Curmi et al., 2013). Moreover, the number of mega-cities with more than 10 million inhabitants will grow to over 40 by 2030 (UNDESA, 2014) creating significant water demand in certain locations (Cosgrove & Cosgrove, 2012). The combination of growing urban development and climate change will have a significant impact on water security (Nguyen et al., 2015). Therefore, it is important to conduct systems modelling to examine whether water supply sources are adequate to meet future demand scenarios (Sahin et al., 2015).

Recently, a number of studies have demonstrated that water end-use analysis can be utilised to inform policy and practices for urban water systems (Nguyen et al., 2013). Specially, water end-use studies can provide essential water consumption information regarding when, where, how and why residential uses consume water in the home (Stewart et al., 2010; Willis et al., 2011). However, the accurate disaggregation of large
amounts of water flow data into different end-use categories is challenging; therefore, this is the focal point of this water end-use study.

Over the last few years, numerous smart water metering programs have been applied to a number of cities worldwide to monitor residential water consumption (Cominola et al., 2015). There are presently two main approaches being applied for undertaking the disaggregation of residential smart meter water data: (1) decision tree algorithm-applied by Trace Wizard® (DeOreo et al., 1996) and Identiflow® (Kowalski & Marshallsay, 2003); and (2) machine learning algorithms-applied by HydroSense (Froehlich et al., 2011), BuntBrainForEndUses® (Arregui, 2015), REU2016 (Vitter & Webber, 2018), and Autoflow (Nguyen et al., 2015).

Autoflow employs a hybrid combination of pattern recognition algorithms (i.e. Hidden Markov Model and Dynamic Time Warping algorithm), and data mining techniques (i.e. event probability analysis) to learn distinct flow signature patterns for each water end-use category (Nguyen et al., 2015; Nguyen et al., 2013a; Nguyen et al., 2013b; Nguyen et al., 2014). This approach overcomes the drawbacks of other existing models to address the water end-use classification problem. For Trace Wizard®, this software is strongly dependent on the experience of the analyst, and a two-week flow data requires two hours of work by the analyst (Nguyen et al., 2013a). In addition, the accuracy of Trace Wizard® is reduced when more than two events occur concurrently (Mayer et al., 1999). Identiflow®, has a higher reported accuracy than Trace Wizard® (Nguyen et al., 2013a), but its accuracy heavily depends on the physical features for each fixture/appliance being inputted into the software (Cominola et al., 2015). The drawback of HydroSense is that it needs a large number of obtrusive pressure sensors (i.e. 33 sensors in a single household) connected to each water device (Froehlich et al., 2011) to be effective.

The first version of Autoflow v1.1 provided recognition accuracy greater than 90% for a set of mechanised appliances (e.g. clothes washer, dishwasher, toilet, etc.), while behaviourally influenced water end-use categories (e.g. shower, bathtub and irrigation) was below 70%. It achieved an average overall pattern recognition accuracy across all categories of 85% (Nguyen et al., 2014). Autoflow v2.1 achieved an overall water end use pattern recognition accuracy of 90% and better handled combined event disaggregation (Nguyen et al., 2015). While Autoflow v2.1 accuracy is getting close to the targeted 95% recognition accuracy required for a commercial application, there are some deficiencies
Optimising Water End-use Analysis Process with Self-organising Maps and K-means Clustering

with this version related to the initial clustering of discrete events into the most likely water end-use category before more detailed pattern recognition algorithms are applied.

To address this deficiency in *Autoflow v2.1*, the present study developed a hybrid model combining Self-Organizing Maps (SOM) and K-means clustering techniques, in order to improve the existing data clustering process. SOM has been applied in water resources problems (Kalteh et al., 2008), water quality analysis (Blokker et al., 2016) and pattern analysis (Laspidou et al., 2015). Moreover, SOM and K-means are combined and applied in market segmentation (Kuo et al., 2006) and water distribution systems (Brentan et al., 2018). In this paper, water end-use pattern recognition accuracy improvements related to the addition of the newly developed SOM+K-means computational procedure has been evaluated (i.e. *Autoflow v3.1* compared to *Autoflow v2.1*). The overall clustering process was divided into 2 main stages: (1) weight matrix estimation using SOM based on the water end-use events features; and (2) clustering all of the unclassified water end-use events according to the weight matrix.

The following sections of the paper outline the study background, the enhanced classification model, the model validation and discussion, and finally a description on how this research is an integral part of the urban water industry. The Background (Section 2) briefly introduced the importance of conducting water end-use studies, the difference between the conventional water metering process and advanced process, the existing models for autonomous water end-use classification, an overview of the existing versions of Autoflow, and persuasion for the latest version of *Autoflow*. Section 3 provides a detailed application of SOM and K-means, which has been applied as pre-grouping process in the proposed version *Autoflow v3.1* to improve the recognition accuracy of single water end-use event analysis. Then, two independent verification processes, limitations, and future research directions are provided in the model validation section (Section 4). Finally, the Section 5 summarises the *Autoflow* software and outlines the importance of its widespread implementation for the urban water industry.

5.2 Background

5.2.1 Conventional water metering process

Water end-use studies have been increasingly used by water businesses globally in the past decade to understand water demand characteristics and trends for their customers. Moreover, water end use data underpins future demand forecasts for a city, and the design of effective demand management strategies. Reliable accounting and management of
water demand is now expected from customers in the current digital and sustainability conscious era. Smart metering and other sensor technologies now provides the opportunity to collect big data on water usage within each segment of an intelligent water network. However, useful analysis of this big data, especially that required to disaggregate residential water consumption into discrete water end use categories, previously required extensive human resources. This resource intensive water end use data analysis process needed to be automated to make it feasible for viable widespread implementation. The authors addressed this problem by designing an autonomous water end use analysis software tool *(Autoflow)* that will enable future smart meters to be equipped with this firmware or cloud linked software that will provide water end use information to each and every residential customer as well as urban water processonals. Such near real-time water end use information will significantly enhance the management of future urban water resources.

Water consumption data is recorded manually on a monthly or even half-yearly basis. Current water billing systems typically use only two to four data points to describe a whole year’s worth of water consumption data of a household. Kilolitre is often the current resolution of reports, which are counted by conventional water meters, which do not have the ability to record when (i.e. the time of day) and where (e.g. tap, toilet) the consumption takes place (Stewart et al., 2013). The current water metering system has many limitations as it does not provide real-time or continuous water flow data, and does not provide flow data of a sufficient resolution to allow for water end-use event disaggregation. Continuous water flow data (i.e. minute to hourly) is essential for real-time network modelling and optimised water infrastructure modelling and engineering (Gurung et al., 2015; Beal et al., 2013; Savic et al., 2016). Water end use data is useful for water demand management (Willis et al., 2013), enhanced water infrastructure planning (Gurung et al., 2015), managing water peak demand (Beal & Stewart, 2013), and understanding water-energy linkages (Siems et al., 2013; Stewart et al., 2018).

### 5.2.2 Existing autonomous water end use classification models

With the advent of smart meters providing greater volumes of high-resolution water flow data, researchers have begun to develop advanced models to autonomously or semi-autonomously classify collected real-time and continuous water consumption data into different water end-use categories. There are two main categories of models, namely, descriptive models and predictive models (Cominola et al., 2015). Descriptive models
analyse the observed water consumption behaviours of users and these models can focus on consumption patterns according to the resolution of water flow data (e.g. Gurung et al., 2015; Beal et al., 2013; Loh et al., 2003; SDU, 2011). Predictive models focus on solving the problem of water demand (e.g. Nguyen et al., 2015; Willis et al., 2013; Polebitski & Palmer, 2009; Makki et al., 2013). Nguyen et al., (2013b) provides a comprehensive critique of the strengths and weaknesses of each of the existing models. This critique suggested that the utilisation of smart water meters at the property boundary, supported by firmware which could autonomously disaggregate water flow data into discrete water end use categories through applying machine learning techniques, was the most feasible approach to realise a vision of end-use data being delivered to customers and utilities.

5.2.3 Water end-use classification process using Autoflow

Many applications have been developed to automate this costly process including HydroSense, Identiflow, and Trace Wizard, but these existing software have many limitations preventing their widespread implementation. Autoflow was developed to address the deficiencies evident in these applications. Autoflow software is now being utilised by five water utilities in Australia. Autoflow enabled the accurate and reliable use event disaggregation of water consumption data, which is useful for understanding when, where and how various categories of residential consumers use water in their homes. The primary functions of the software include, but are limited to, collecting and transferring water consumption data through a smart water metering system, and analysing the data and producing a wide range of water end use reports which can be accessed by a relevant stakeholders (i.e. consumers, water utilities, etc.). Anonymous summary reports can also be compiled for government departments to assist them to formulate a range of policies (e.g. incentives, outdoor water restrictions, behavioural marketing, etc.).

The Autoflow offers a robust pattern recognition procedure through the hybrid combination of existing popular techniques. Autoflow v1.1 software combined Hidden Markov Model (HMM), Artificial Neural Networks (ANN), and the Dynamic Time Warping (DTW) algorithm (Figure 5-1). The first version was suitable for mechanical water end-use category analysis, such as clothes washers, dishwashers, and toilets (Nguyen et al., 2013a; Nguyen et al., 2013b). Autoflow v2.1 included an additional search process using the Dynamic Time Warping algorithm. This enhanced version marginally
increased the classification accuracy of mechanised end uses categories (Nguyen et al., 2015).

Figure 5-1 Overall event classification procedures

The goals of the design of the Autoflow processing algorithms is to provide both accuracy and efficiency. The first version of Autoflow (i.e. v1.1) achieved an average overall pattern recognition accuracy across all categories of 85%. The overall accuracy of Autoflow v2.1 was over 90%, and with 4.9% and 8.0% improvement for single event and combined event categorisation, respectively. In terms of efficiency, the Autoflow v2.1 can pattern
recognize almost 100 events in one second, which is considerably faster than an expert using Trace Wizard. *Autoflow* v2.1 also provided an illustrative user interface offering user-defined water consumption web page to view daily, weekly and monthly consumption tables, as well as charts on consumers’ water demand across major end-use categories (e.g. clothes washer, shower, irrigation, etc.). The system also provides customer alerts as to whether they have leaks rather than waiting for the present slow feedback process from the traditional metering process (Nguyen et al., 2015; Nguyen et al., 2013a; Nguyen et al., 2013b).

Given that mechanised end use categories such as clothes washer, dishwasher, and toilet usually consist of similar and repeated patterns, one of the initial *Autoflow* analysis modules completes a grouping process where similar patterns are initially classified as belonging together before subsequent detailed pattern recognition procedures are completed. *Autoflow* v1.1 and v2.1 adopted DTW as the main technique to achieve this task. However, recent investigations indicated that DTW had some limitations since it grouped similar patterns based on a pre-determined distance threshold between the unclassified events and the reference event. Two events would be considered similar if their distances to a third event (i.e. reference event) were close. However, this distance value is just a dimensionless number which represents the shape difference between two given patterns, and in some cases two completely different events could have the same distance as the reference event. As a result, DTW could assign those two events into the same initial category although their water end use patterns are completely different. While this incorrect classification was often discovered in subsequent analysis, it increased data processing effort and resulted in higher rates of erroneous classification.

### 5.2.4 Overview of applied techniques

Apart from DTW, many other popular clustering techniques have been established and widely applied, including SOM (Kohonen, 1982), K-means clustering (MacQueen, 1967), and Artificial Neural Network (ANN) (Moon et al., 2009) to address various complex pattern matching problems, such as handwriting, speech recognition, fingerprint recognition, surface water level and seabed liquefaction predictions (Sannasiraj et al., 2004; Zhang et al., 2007). Through investigating many of the above mentioned models, it was determined that the SOM and K-means clustering were highly suitable approaches for clustering similar water end use events together as required in this study.
The SOM is a type of ANN with an unsupervised learning process (Kohonen, 1982). This approach attempts to replicate the function of the brain when it receives visual and memory stimulation, and is useful for efficiently identifying common patterns and clustering them together.

The SOM network consists of an input layer and an output layer (Kohonen layer). The output layer is designed as a two-dimensional arrangement of neurons that have an ability to project a high dimensional input to low-dimensional grid, maintaining the topological order. In this layer, each neuron represents a cluster. In addition, the input layer of nodes is connected to the output nodes and own a weight vector respectively. Euclidean distance and cosine distance are two major criteria between the weight vector of output nodes and input nodes.

The SOM algorithm has the advantage of producing intuitive results (Dorffner, 2001), but has two limitations related to its static architecture and capabilities to characterise hierarchical relations within the data (Rauber et al., 2002). The K-means algorithm is a well-established statistical technique for unsupervised iterative data cluster analysis. SOM and K-means clustering were the utilised techniques for enhancing Autoflow v2.1 as described in the next section.

5.3 Enhanced Water end-use Classification Model

*Autoflow* v2.1 adopted the DTW technique for grouping similar end-use events. However, DTW was heavily dependent on a threshold value, which defined the similarity between two samples. For *Autoflow* v3.1, DTW was replaced by the hybrid SOM+K-means clustering analysis module, in order to improve residential water end-use pattern recognition accuracy. The following sections describe how these techniques were applied to this pattern recognition problem.

5.3.1 SOM for water end-use pattern recognition

The formulated six-step procedure detailed below applied the SOM technique for the initial clustering of residential water end use events into different water end-use categories.

*Step 1-Input layer settings:*

As shown in Figure 5-2, the input layer are given vector $E_l (l = 1, 2, ..., P)$ of the water end-use events, which are then normalized as $\hat{E}_l (l = 1, 2, ..., P)$, where $P$ is the number
of unclassified events. In the present study, each water event has 3 features including volume, duration and maximum flowrate. Therefore, an additional subscript $x$ is used to present the features and input events are presented by $E_{xl} (x = 1,2,3; l = 1,2, ... P)$.

**Step 2-Output layer settings:**

A 2D network of $m$ rows and $n$ columns of water end-use categories is set up in output layer as shown in Figure 5-2. And the total number of categories $T$ is written as equation 5-1.

$$T = m \times n$$  \hspace{1cm} (5 \hspace{1cm} 1)

![Figure 5-2 Water end-use network in SOM](image)

**Step 3-Assign matrix vectors:**

Assign a random initial value to the weight matrix vector, due to the network is fully connected, each input node is connected to each category. Therefore, the weight matrix vector between input event and categories is $W_{lj} (l = 1,2, ... P; j = 1,2, ... , T)$, which can be normalised to equation 5-2.

$$\tilde{W}_{lj} = \frac{W_{lj}}{\|W_{lj}\|} \hspace{1cm} (5 \hspace{1cm} 2)$$

In output layer, the weight matrix vector is $\tilde{W}_{xtj} (x = 1,2,3; l = 1,2, ... P; j = 1,2, ... T)$.

**Step 4-Determine winner category and define the winner neighbourhood area:**

In this step, the distance between input vectors and all categories in output layer is calculated. The iterative method is used to determine the winner category $j^*$ according
Optimising Water End-use Analysis Process with Self-organising Maps and K-means Clustering to the minimum distance from input nodes to output nodes, and the weight matrix vector of winner category is $\mathcal{W}_{xtj^*}$, which can be normalised to equation 5-3.

$$
\|\mathcal{E} - \mathcal{W}_{xtj^*}\| = \min_{j \in \{1,2,\ldots,T\}} \{\|\mathcal{E} - \mathcal{W}_{xtj}\|\} \tag{5 - 3}
$$

Then, the initial neighbour area of winner category is defined as $N_{j^*}(t = 0)$, and it is a larger zone, which it decreases over iteration steps as shown in Figure 5-3. The maximum iteration is $I_{\text{max}}$.

![Figure 5-3 Change of neighbor area with iterations](image)

In addition, the neighbour area will update and the function is written as equation 5-4.

$$
N_{j^*}(t) = N_{j^*}(t = 0) * e^{-\frac{t \cdot \log N_{j^*}(t = 0)}{I_{\text{max}}}} \tag{5 - 4}
$$

**Step 5-Adjust winner weight matrix vectors and update neighbour area:**

According to the SOM network, the winner category has an activation value of ONE while other neurons have activations of ZERO, as equation 5-5.

$$
z_j(t + 1) = \begin{cases} 
1 & j = j^* \\
0 & j \neq j^* 
\end{cases} \tag{5 - 5}
$$

However, only the winner neuron can adjust the weight matrix vector. The adjustment function given by equation 5-6.

$$
\mathcal{W}_{xlij}(t + 1) = \mathcal{W}_{xlij}(t) + \eta(t)[\mathcal{E}_{xl} - \mathcal{W}_{xlij}] \quad x = 1,2,3; j = 1,2,\ldots,T; l = 1,2,\ldots,P \tag{5 - 6}
$$

where $0 < \eta(t) \leq 1$, and it is a function, which decreasing with the number of iterations, to ensure the convergence of the algorithm. In this present paper, the learning rate is given by equation 5-7.
Optimising Water End-use Analysis Process with Self-organising Maps and K-means Clustering

\[ \eta_j(t) = \eta_j(t = 0) \cdot e^{-\left(\frac{t}{I_{\text{max}}^j}\right)} \]  

(5 – 7)

Step 6 - End judgement:

When the learning rate \( \eta(t) \leq \eta_{\text{min}} \) or \( \eta(t) = 0 \), the training process is finished. When the stopping condition is not satisfied, the algorithm will go back to the Step 2 of the procedure until the maximum iteration number \( I_{\text{max}} \) is reached.

5.3.2 K-means for end-use study

The specific process of the K-means algorithm is as follows:

Step 1 - Initial centroids selection:

Randomly select \( k \) samples as the initial cluster centroid \( C_k \) from all winner weight matrix vectors \( \overline{W}_{xj} \). In this present paper, the winner weight matrix vectors are the input data in K-means algorithm in order to clustering matrix first and get the final weight matrix as the initial centroids in next grouping process. For input data with 3 features, each centroid can be represented as equation 5-8.

\[ C_k = [c_{k1}, c_{k2}, c_{k3}] \]  

(5 – 8)

Step 2 - Assign matrix vectors:

In this step, each vector will be grouped according to the similarity (distance) \( d_k \) to the centroids. The distance to each centroid is calculated for each input vector can be calculated by equation 5-9. Then, the object is assigned to the closest (minimum distance) centroid, which can be normalised as equation 5-10.

\[ d_k(w, c) = \sqrt{\sum_{i=1}^{m} (w_{ki} - c_{ki})^2} \quad m = 3 \]  

(5 – 9)

\[ w(j) \in \text{group}(\min(d)) \quad j = 1, 2, \ldots T \]  

(5 – 10)

where \( w_{ki} \) is an element of weight matrix vector.

Step 3 - Update the centroid:
When all matrix vectors have been grouped, the centroids position is updated. The new position of it according to the new means of each cluster. The position of group \( k \) is written as equation 5-11.

\[
P_k = \frac{\sum_{i=1}^{N_k} w_{ki}}{N_k}
\]  

(5 - 11)

where \( N_k \) is the number of elements belonging to group \( k \).

**Step 4-Iteration:**

Repeat Step 2 and Step 3 of this procedure until the centroid of the clusters no longer changes so far. At the end of the process, the distance between the clusters is maximal and distance between the centroids and their data is minimal. After this step, the clustering result is a new set of matrix vectors \( \bar{W}_{xlj} \), \( x = 1,2,3; \ j = 1,2, \ldots, k; \ l = 1,2, \ldots, P \).

**Step 5-2\textsuperscript{nd} K-means clustering for unclassified events:**

Input the previous result as the initial centroids. Then input the all unclassified events and grouped those using specific K-means, the process is same as Step 2. Then, the unclassified will be grouped into \( k \) groups.

### Hybrid SOM-K-means model for residential water end-use pattern

Application of the algorithms for the data clustering process described above was crucial before the operation of the prior developed pattern matching algorithms and associated analysis processes. The structure of the SOM network is robust but simple, and can handle outliers. In the situation where the mesh of neurons is larger, the number of clusters is larger. Reducing the number of clusters in the SOM algorithm can reduce the final number of water end-use groups. However, the quality or the accuracy of clustering will be decreased. In addition, the clustering result of K-means algorithm is greatly affected by the selection of the initial cluster centroid. If the initial centres are not properly selected, the result may fall into a local optimal solution, rather than global optimal result. For these reasons, this section presents the model operationalization process for this specific problem, including how these two techniques were combined and how they can improve the clustering process in Autoflow v3.1.
The existing water end-use model predominately applied only computing techniques (Pastor-Jabaloyes et al., 2018). The entire procedure for the hybrid technique includes two stages. Firstly, each water event is presented by extracted features and grouped through SOM (reaching a certain number of cycles), as well as a set of weight vectors are calculated. Secondly, these vectors that serve as the initial cluster centroids were applied in K-means clustering. Figure 5-1 presents the algorithms implemented in the previous version *Autoflow v2.1*. In this version, the initial pre-grouping process was undertaken by the Dynamic Time Warping algorithm before the decision making stage. Moreover, this pre-grouping process is a major component of the whole pattern recognition analysis process, and the accuracy of this stage can affect the final water category recognition accuracy directly. Hence, it is necessary to seek a method with high accuracy and efficiency in the pre-grouping process to improve the overall accuracy of Autoflow. In the developed version *Autoflow v3.1*, the hybrid technique combining SOM and K-means has replaced the DTW algorithm in the initial pre-grouping process, as shown in Figure 5-4. The specific subset of the entire procedure for the newly included SOM-K-means clustering process is detailed in Figure 5-5. Readers are referred to Nguyen (Nguyen et al., 2015; Nguyen et al., 2013a; Nguyen et al., 2013b; Nguyen et al., 2014) for the detailed analytical procedures for the other parts of the procedure described in Figure 5-4.

The advantages of this hybrid technique include: (i) maintains the characteristics of the self-organizing of SOM algorithm; (ii) maintains the high efficiency of K-means clustering; (iii) reduces the convergence time in SOM; and (iv) reduces the impact of the initial centroid selection in the K-means algorithm.
Figure 5-4 Overall water end-use event classification procedures with SOM+K-means
5.4 Model Validation

As described below, two stages of model testing and verification were undertaken in this study:

(i). Model testing utilised the herein recommended hybrid combination of SOM+K-means on a total of 9,200 random event samples.
(ii). Event model verification was conducted by applying the *Autoflow v3.1* software tool to 30 independent homes from Australia and comparing against using original *Autoflow v2.1*.

### 5.4.1 Independent testing process

A total of 9,200 samples were randomly selected from 104,271 disaggregated samples of 8 different categories for testing. These selected samples were mixed together, and the SOM+K-means analysis module was adopted to partition them into 8 different groups, where each group was expected to consist of similar events. The correctly classified water end use events are shown in Figure 5-6. The model accuracy is determined by dividing the number of correctly grouped events for each category by the number of events for testing of that category. For example, for the clothes washer event category, the number of testing samples was 1500 and the number of clothes washer events correctly grouped was 1398, which resulted in an accuracy of 93.2%. Table 1 details that the *Autoflow v3.1* water end use event recognition accuracy ranged from 86% for bathtub to 94.2% for dishwasher. It is interesting to note in Table 5-1 that the hybrid model improved the accuracy of both the more predictable mechanised water end-use event categories (i.e. clothes washer, dishwasher, evaporative cooler and toilet), as well as those more variable categories heavily dependent on human behaviours (i.e. tap, bathtub, irrigation and shower). Overall, the newly developed hybrid method that has been augmented in *Autoflow v3.1* has achieved a higher overall pattern recognition accuracy when compared to the prior utilised DTW technique applied in *Autoflow v2.1* for this similar clustering function.

![Table](image)

**Figure 5-6** Correctly classified water end use events using technique.
Optimising Water End-use Analysis Process with Self-organising Maps and K-means Clustering

Table 5-1 Water end-use event classification accuracy comparison (v3.1 versus v2.1)

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of total samples</th>
<th>No. of samples testing</th>
<th>DTW accuracy (v2.1) (%)</th>
<th>Hybrid technique accuracy (v3.1) (%)</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clothes washera</td>
<td>11,168</td>
<td>1,500</td>
<td>86.2</td>
<td>93.2</td>
<td>7</td>
</tr>
<tr>
<td>Dishwashera</td>
<td>3,458</td>
<td>1,500</td>
<td>91.8</td>
<td>94.2</td>
<td>2.4</td>
</tr>
<tr>
<td>Toilet</td>
<td>13,852</td>
<td>1,500</td>
<td>89.8</td>
<td>92.8</td>
<td>3</td>
</tr>
<tr>
<td>Evaporative cooler</td>
<td>23,542</td>
<td>1,500</td>
<td>86.9</td>
<td>88.6</td>
<td>1.7</td>
</tr>
<tr>
<td>Shower</td>
<td>5,743</td>
<td>1,500</td>
<td>80.4</td>
<td>89.4</td>
<td>9</td>
</tr>
<tr>
<td>Tap</td>
<td>45,959</td>
<td>1,500</td>
<td>86.2</td>
<td>88.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Bathtub</td>
<td>217</td>
<td>100</td>
<td>80</td>
<td>86</td>
<td>6</td>
</tr>
<tr>
<td>Irrigation</td>
<td>332</td>
<td>100</td>
<td>80</td>
<td>87</td>
<td>7</td>
</tr>
<tr>
<td>All categories</td>
<td>104,271</td>
<td>9,200</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: a Mechanical water end-use event categories

5.4.2 Autoflow recognition accuracy using independent homes

To determine the true accuracy of the Autoflow v3.1 model, 30 independent home datasets were collected. These 30 homes based in Australia were independent of the original training and testing datasets and each water end-use event was certified by home owners.

Table 5-2 Independent data verification comparison (v3.1 versus v2.1)

<table>
<thead>
<tr>
<th>Category</th>
<th>Applied model</th>
<th>Number of home</th>
<th>Average accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Accuracy</td>
<td>80–90%</td>
</tr>
<tr>
<td>Clothes washera</td>
<td>1b</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2c</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Dishwashera</td>
<td>1</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Toilet</td>
<td>1</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>Evaporative cooler</td>
<td>1</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Shower</td>
<td>1</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Tap</td>
<td>1</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>Bathtub</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Irrigation</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes: a Mechanical water end-use events; b1- Autoflow v3.1 and c2- Autoflow v2.1; dIrrigation water end-use events are fewer and sporadic so difficult to quantify accuracy in a small sample.
Table 5-2 compares the average accuracy achieved for each category between the *Autoflow v2.1* developed using DTW, HMM and ANN and the *Autoflow v3.1* (i.e. SOM + K-means + ANN + HMM) for the independent 30 homes sample. Water end use event pattern recognition accuracy for v3.1 ranged from 75.1% for irrigation to 90.1% for clothes washer. Irrigation event recognition accuracy should be treated with caution and is not considered reliable for small samples (i.e. N=30 homes) since this usage category has a smaller number of samples and is sporadic. As can be seen from this table, an average accuracy of above 85% has been achieved for mechanical categories, with the maximum of 90.1% for clothes washer and minimum of 85.5% for toilet. In comparison with the *Autoflow v2.1*, the new v3.1 model has provided greater overall accuracy, especially for the mechanised water end use categories. Notable version comparison improvements include: 87.6% compared to 82.1% for dishwasher; 86.8% compared to 83.4% for evaporative air cooler; and 85.5% compared to 83.1% for toilet. It is noticeable that the newly included SOM-K-means clustering algorithm has been more effective for mechanised water end use event categorisation.

![Figure 5-7](image.png)

*Figure 5-7 Computational processing time of Autoflow (v3.1 versus v2.1)*

Additionally to water end use pattern recognition accuracy improvements, the processing time to analyse the 30 independent residential homes water end-use events was determined, as shown in Figure 5-7. Figure 5-7 illustrates the processing time comparison between *Autoflow v3.1* and previous version *Autoflow v2.1*. *Autoflow v3.1* has reduced the flow data end-use disaggregation time compared with the previous version. As evident in Figure 5-7, processing time per event reduces further with *Autoflow v3.1* when there is
a greater number of unclassified events. In terms of number of events, although Autoflow v2.1 had a good efficiency, it still cannot achieve a processing speed of 100 events per second. In the developed Autoflow v3.1, a pattern recognition speed greater than 100 events per second could be achieved. Overall, a 20% improvement in processing time was achieved for Autoflow v3.1 compared to Autoflow v2.1.

Testing results have indicated that the replacement of DTW by the combination of SOM with K-means has improved recognition accuracy in Autoflow v3.1, especially for the mechanised water end use categories. The hybrid technique was less effective for behaviourally influenced water end use categories (e.g. shower) since there is a much greater variety in the patterns of these events; thereby making clustering more challenging.

However, Autoflow still can be improved in future research. The developed analytical approach was reliable for categorising most end-use categories with the exception of irrigation, tap and shower. In order to significantly improve recognition accuracy of these end-use event categories, which heavily depend on human behavior, further research on how to unobtrusively incorporate local contextual data (i.e. residential household descriptive and behavioural information) into the Autoflow analysis process is required. To achieve this, the development of self-learning algorithms for incorporation into the final analytical phase needs to be completed. Collecting and analysing each event feature of a much larger sample of events (e.g. different regions, dwelling type, etc.) will be the focus of future research. The future developed version of Autoflow, that includes self-learning functionality, should complete a more accurate disaggregation of water end-use data for new residential households having some different water using fixtures.

5.5 Conclusion

The Autoflow software, which is an integrated water management system, employs smart water metering and a series of intelligent algorithms to automate the disaggregation of high-resolution residential water flow data into discrete water end-use events. Autoflow v3.1 presented in this paper includes a new analysis module to the prior version that groups similar events using a hybrid combination of customised SOM and K-means algorithm. The verification process demonstrated that the enhanced version increased the recognition accuracy. Moreover, the study showed that the inclusion of the SOM+K-means method better clustering of mechanised water end-use events, such as clothes.
Optimising Water End-use Analysis Process with Self-organising Maps and K-means Clustering

washer, dishwasher, evaporative cooler and toilet. *Autoflow v3.1* has significant implications for the urban water industry.

*Autoflow* is a commercial software, which is currently being used by various water utilities in Australia (e.g. Yarra Valley Water, City West Water and South-east Water). This application was developed using MATLAB and it can be used both as both a desktop and web application. At this stage, the main features of the software include, but are limited to, collecting and transferring water consumption data through a smart water metering system, analysing the data and producing a wide range of reports which can be accessed by various users (i.e. consumers, water utilities and government organisations). However, work is underway to embed this novel software as firmware within smart meters and transfer processed water end-use data to consumers via an application on their phone or computer.

*Autoflow* has significant implications for consumers and water businesses, government departments, the metering and software industry, and the wider urban water industry. For consumers, *Autoflow* provides them with an easily accessible interface, which allows consumers to monitor where and when they are consuming water in their household. By consumers having water end use data readily available to them on a daily basis, they will be have better understanding on their water consumption and be able to target efficiency efforts (i.e. Understand that their shower consumption is high). For the water utility, compiled anonymised city-level water end use data can assist them with a range of operational functions including water demand management, distribution network infrastructure planning, post-meter leakage management, customer engagement, to name a few. Government departments and agencies charged with urban water policy can use water end use data for designing effective water efficiency policies (i.e. rebates, restrictions, etc.). The water metering industry are currently developing advanced digital meters that can record and transmit high resolution water usage data (i.e. 0.01L / 1 second); such metering technology facilitates widespread autonomous water end use analysis. *Autoflow* software could be embedded into future smart meters firmware, or alternatively housed in a server for cloud-based processing. Based on these arguments, there is a strong need for further research on developing intelligent water consumption analysis algorithms and associated software such as *Autoflow*. The creation of such software poses significant benefits for infrastructure planners, water demand managers,
Optimising Water End-use Analysis Process with Self-organising Maps and K-means Clustering

architects and developers, policy advisors, to name a few, who seek to better understand water consumption patterns.
6. Conclusion, Contributions and Implications

The final chapter presents a summary of the primary findings of this research. The contributions and limitations of the research are detailed for future research directions. The chapter begins with section 6.1 reiterating the research objectives and highlighting the key outcomes that satisfied these objectives. Section 6.2 identifies the theoretical and practical contributions made by this study. The limitations of the research and suggestions for future research are presented in Section 6.3. Section 6.4 concludes the chapter and thesis discourse.

6.1 Conclusion
The principal objectives of this research were (i) recognising the distinct patterns of each residential water end-use event category; (ii) identifying the limitations of existing pattern recognition approaches for a residential water end-use study; (iii) developing new approaches in water end-use disaggregation including the application of advanced techniques that have a higher accuracy; and (iv) calibrating and validating the developed model. More specifically, the research aimed to autonomously recognise the distinct patterns of each water end-use category through an analysis of the collected data. Existing water flow trace analytical systems, such as Trace Wizard, Identiflow, HydroSense, and Autoflow, were also examined to identify shortcomings that need to be developed through the proposed techniques. In addition, many pattern clustering and classification approaches, such as ABC, SOM, DTW, K-Medoids, and K-Means, have been investigated and applied to achieve two primary modules for mechanical-event analysis and single-event analysis. The final developed techniques were tested against datasets (i.e., (1) a complex dataset containing 70 toilet events to verify the mechanical-event recognition module accuracy, (2) more than 100,000 single-event from 8 categories, and (3) 30 independent homes in Australia) to validate its accuracy and efficiency in classifying single events.

The existing literature, including research publications, and academic and industry reports were critically reviewed as presented in Chapter 2 and the introductory sections of Chapter 4 and 5. This review focused on the advanced techniques and technologies for the disaggregation of water consumption using smart water meters in the last few decades. In this phase, a detailed description of the distinct patterns of each category was analysed to help develop hybrid techniques. Phase 2 involved the application of different pattern
recognition and data analysis techniques to develop a core module for mechanical-event classification. One journal article, containing pertinent literature, methodology, data, and results, was detailed in Chapter 4. Phase 3 introduced a hybrid combination of SOM and the K-Means algorithm applied to further develop the existing software *Autoflow*, which can increase the recognition accuracy for each water end-use category. Phase 3 results culminated into a second journal paper in Chapter 5.

### 6.2 Research Contributions

With the development of rapid computing technology, many analytical tools have been proposed to enable the water end-use disaggregation task. Despite the numerous techniques proposed, none of them has thoroughly solved the problem, and more studies are still necessary to develop the existing systems. This research study was conducted with the view to advance hybrid techniques for the development of an intelligent system for residential water end-use disaggregation. Contributions and implications for the urban water management are outlined in the following sections.

#### 6.2.1 Contributions to the existing body of knowledge

This research has been analysed in two stages, both of which have made a contribution to the hydro-informatics discipline, particularly in relation to the development of pattern-recognition techniques that can autonomously disaggregate water consumption data into useful end-use information for customers and water businesses.

- An accurate method was developed to tackle the classification problem in the water flow trace analysis field. The required method was developed by combining existing pattern-recognition techniques (i.e., DTW, ABC, and K-Medoids) for the particular category study. The method was applied for analysing mechanical events, which showed an improvement of more than 10% in classification accuracy (Chapter 4).

- A hybrid technique was developed that replaced the existing pre-grouping clustering module in the *Autoflow* software (i.e., Version 3.1 superseded Version 2.1). With this development, the model would be able to improve the disaggregation accuracy of single event recognition, especially for the mechanical type of events (i.e., clothes washer, dishwasher, and evaporative cooler) (Chapter 5).
Along with an array of theoretical contributions, this study will also provide significant insight into the development and effectiveness of urban water management at the development scale in the following manner.

- Developing the ability to monitor the effect of enforcement or restriction levels on residential water consumption;
- Providing real-time water consumption data to customers in order to increase the level of knowledge and understanding of residential water consumption;
- Providing to the water business an extensive database of anonymised city-level water end-use data to assist it with a range of operational functions including water demand management, distribution network infrastructure planning, post-meter leakage management, and customer engagement.

6.3 Study Limitations and Future Research
The present study used a variety of rigorous pattern-recognition techniques and analysis procedures to categorise the residential water end-use events, which produced an array of theoretical and practical results that is able to apply to urban water infrastructure planning and management. However, some limitations still exist and are summarised as follows:

- The applied parameter values used to classify toilet events in Chapter 4 were achieved through the analysis of an artificial mixed database. However, for identifying an appropriate value for the most accurate identification of other end-use categories, further research needs to be undertaken across a number of data to establish more rigorous criteria.
- To further improve the classification accuracy, in addition to SOM and K-means, an Independent Component Analysis technique can be incorporated to analyse the physical characteristics of each event (i.e., volume, duration, and flow rate), which are likely to have a significant impact on the overall classification process.
- Finally, the Autoflow software is considered as firmware to embed within smart meters. Consumers can monitor the water data via an application on their phone or computer. In addition, compiled anonymised water data can also be provided to the water utility business for optimised urban water management.

6.4 Closure
A hybrid combination of different pattern-recognition techniques was carried out to develop the existing residential water end-use disaggregation system. The first three chapters of the thesis set the scene for this MPhil project, namely, the introduction,
overarching literature review, and research methodology. These chapters constitute the basic framework for the research. Then, the thesis is predominantly composed of two peer-reviewed journal papers related to the various research objectives contained within the two distinct phases of the project, namely, mechanical-event and single-event analysis. Finally, this thesis concludes with a summary of the research contributions, implications, limitations, and proposed recommendations for future research in the field of optimising urban water management.
References


Dorffner, G. (2001). Limitations of the SOM and the GTM. Retrieved from https://pdfs.semanticscholar.org/c93a/e9ffeda90c9ea4cd951989a00a0afde8845b.pdf


**References**

https://pdfs.semanticscholar.org/015d/f4d97ed1f541752842c49d12e429a785460b.pdf


References


UN. (2009). *Majority of world population face water shortages unless action taken, warns Migiro.* UN News Centre. Online multimedia, New York, USA.


