Usage-Based Optimal Energy Control of Residential Water Heaters

by

MJ Ritchie

Dissertation presented in fulfilment of the requirements for the degree of Doctor of Philosophy in Engineering (Electronic) in the Faculty of Engineering at Stellenbosch University

Supervisors: Prof. M.J. Booysen,
Dr J.A.A. Engelbrecht

December 2021

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Plagiaatverklaring / *Plagiarism Declaration*

1 Plagiaat is die oorneem en gebruik van die idees, materiaal en ander intellektuele eiendom van ander persone asof dit jou eie werk is.
Plagiarism is the use of ideas, material and other intellectual property of another's work and to present is as my own.

2 Ek erken dat die pleeg van plagiaat 'n strafbare oortreding is aangesien dit 'n vorm van diefstal is.
I agree that plagiarism is a punishable offence because it constitutes theft.

3 Ek verstaan ook dat direkte vertalings plagiaat is.
I also understand that direct translations are plagiarism.

4 Dienooreenkomstig is alle aanhalings en bydraes vanuit enige bron (ingesluit die internet) volledig verwys (erken). Ek erken dat die woordelikse aanhaal van teks sonder aanhalingstekens (selfs al word die bron volledig erken) plagiaat is.
Accordingly all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.

5 Ek verklaar dat die werk in hierdie skryfstuk vervat my eie oorspronklike werk is en dat ek dit nie vantevore in die geheel of gedeeltelik ingehandig het vir bepunting in hierdie module/werkstuk of 'n ander module/werkstuk nie.
I declare that the work contained in this assignment is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

Studentenommer / *Student number*	**Handtekening / *Signature***
MJ Ritchie **Voorletters en van / *Initials and surname***	December 2021 **Datum / *Date***

Abstract

Water heating contributes up to 40% of a household's total electricity usage and places a substantial burden on the electricity grid due to high power ratings and users' largely simultaneous hot water usage. It is estimated that South Africa uses 38.5 GWh/day on water heating. Tanked water heaters can store thermal energy for long periods of time with a portion lost to the environment. Although many demand response strategies have been proposed to mitigate costs from time-of-use pricing of electricity, many developing countries have flat-rate pricing and are heavily reliant on the burning of fossil fuels for electricity. This thesis explores optimisation to develop a system that provides electric water heaters (EWHs) with the optimal heating schedule to minimise overall electrical energy usage without compromising the comfort of the customer.

The efficacy of energy management techniques that model water heaters and the accuracy of their simulation results rely on representative hot water usage profiles. A probabilistic data-driven model for modelling individualised hot water profiles and an accompanying hot water usage simulator is proposed in this thesis. The model is trained and evaluated using high quality data gathered from 77 residential households which is first refined by a data cleaning process. A hot water demand predictor is also developed on top of the probabilistic model to predict future hot water profiles that can be used for optimisation.

A novel dynamic programming (DP) approach is presented to achieve optimal control of a single node EWH. To accommodate the natural stratification that occurs, a novel A* search algorithm approach is also presented to achieve optimal control of an EWH with stratification. The DP and A* approaches produce the optimal plan for hot water usage profiles with perfect foreknowledge of water drawn from the EWH. The A* approach is further utilised to produce an optimal plan for the hot water usage profile when water drawn is predicted. All three conditions are tested with a simulator that is equipped with a novel temperature feedback controller to minimise the effects of mispredictions and model inaccuracies between the optimal planning and the execution, and a reactive water mixer to simulate the user behaviour.

Three strategies are explored for optimal control of domestic water heating that do not depend on the default thermostat control, namely: matching the delivery temperature in the hot water; matching the energy delivered in the hot water; and a variation of the second strategy which provides for *Legionella* sterilisation. For each of these strategies we examine the energy used in heating, the energy delivered at the tank outlet, and issues of convenience to the user to determine how much energy an EWH can save. It was concluded that the most energy savings is achieved when the energy matching strategy is used, where 21.9 % is saved when water usages are perfectly known and 9.6 % is saved when they are predicted.

Uittreksel

Waterverhitting dra by tot 40% van die totale elektrisiteitsverbruik van 'n huishouding en plaas 'n aansienlike las op die elektrisiteitsnetwerk as gevolg van hoë drywingsaanvraag en die grootliks gelyktydige gebruik van warm water deur gebruikers. Daar word beraam dat Suid-Afrika 38.5 GWh/dag op waterverhitting gebruik. Tenkverwarmers kan termiese energie vir lang tydperke stoor met 'n kleiner gedeelte wat in die omgewing verlore gaan. Alhoewel daar baie aanvraag-bestuur strategieë voorgestel word om die koste van die prys van tyd-van-gebruik tariewe te verminder, is daar baie vaste pryse in baie ontwikkelende lande en is dit sterk afhanklik van die verbranding van fossielbrandstowwe vir elektrisiteit. Hierdie tesis ondersoek die ontwikkeling van 'n stelsel wat elektriese waterverwarmers (EWHs) voorsien van die optimale skema om die totale elektriese energieverbruik te verminder sonder om die gemak van die gebruiker te benadeel.

Die doeltreffendheid van tegnieke vir energiebestuur wat waterverwarmers modelleer en die akkuraatheid van hul simulasie resultate is afhanklik van verteenwoordigende warmwatergebruikprofiele. 'n Waarksynlikheids- en datagedrewe model vir die modellering van geïndividualiseerde warmwaterprofiele en 'n gepaardgaande warmwatergebruiksimulator word in hierdie proefskrif voorgestel. Die model word opgelei en geëvalueer met behulp van hoë kwaliteit data wat versamel is uit 77 residensiële huishoudings wat eers deur 'n data-skoonmaakproses verfyn word. 'n Voorspeller vir die aanvraag na warm water word met die waarskynlikheidsmodel ontwikkel om toekomstige warmwaterprofiele te voorspel wat gebruik kan word vir optimering.

'n Nuwe dinamiese programmering (DP) benadering word aangebied om optimale beheer van 'n enkele node EWH te bewerkstellig. Om die natuurlike stratifikasie wat voorkom te akkommodeer, word 'n nuwe A* soekalgoritme benadering aangebied om optimale beheer van 'n EWH met stratifikasie te bewerkstellig. Die DP en A* benaderings lewer die optimale plan vir die warmwatergebruiksprofiel met perfekte voorkennis van die water wat uit die EWH getap gaan word. Die A* benadering word verder gebruik om 'n optimale plan vir die warmwatergebruiksprofiel op te stel op grond van 'n voorspelling van die water wat uit die EWH getap gaan word. Al drie die toestande word getoets met 'n simulator wat toegerus is met 'n nuwe temperatuurterugvoerbeheerder om die gevolge van verkeerde voorspellings en onakkuraathede tussen die optimale beplanning en die uitvoering te minimeer, en 'n reaktiewe watermenger wat die gedrag van die gebruiker simuleer.

Drie strategieë word ondersoek vir optimale beheer van huishoudelike waterverwarming wat nie afhang van die standaardtermostaatbeheer nie, naamlik: ooreenstemming met die afleweringstemperatuur in die warm water; ooreenstemming met die energie wat in die warm water gelewer word; en 'n variasie van die tweede strategie wat voorsiening maak vir *Legionella* sterilisasie. Vir elk van hierdie strategieë ondersoek ons die energie wat gebruik

word by die verhitting, die energie wat by die afvoer van die tenk gelewer word, en die verbruiker se gemak om te bepaal hoeveel energie 'n EWH kan bespaar. Daar is tot die gevolgtrekking gekom dat die meeste energiebesparings behaal word wanneer die strategie vir energie ooreenstemming gebruik word, waar 21.9 % bespaar word wanneer watergebruik heeltemal bekend is en 9.6 % bespaar word wanneer die watergebruik voorspel word.

Publications

This dissertation contains work that has been published which are listed as follows:

- M. J. Booysen, J. A. A. Engelbrecht, M. J. Ritchie, M. Apperley, and A. H. Cloete, "How much energy can optimal control of domestic water heating save?" *Energy for Sustainable Development*, vol. 51, pp. 73–85, 2019 [1]
- M. J. Ritchie, J. A. A. Engelbrecht, and M. J. Booysen, "A probabilistic hot water usage model and simulator for use in residential energy management," *Energy and Buildings*, 2020 [2]
- J. Engelbrecht, M. J. Ritchie, and M. Booysen, "Optimal schedule and temperature control of stratified water heaters," *Energy for Sustainable Development*, vol. 62, pp. 67–81, 2021 [3]
- M. J. Ritchie, J. A. Engelbrecht, and M. J. Booysen, "Practically-achievable energy savings with the optimal control of stratified water heaters with predicted usage," *Energies*, vol. 14, no. 7, p. 1963, 2021 [4]

Acknowledgements

Firstly, I wish to express my sincere gratitude to Prof. MJ (Thinus) Booysen and Dr Japie Engelbrecht for our most productive meetings and captivating discussions that always kept me motivated and passionate for my research. I honestly could not have asked for a more brilliant team. I would like to also thank my colleagues at the MTN Mobile Intelligence Lab as well as the coffee machine. I would also like to give a special thanks to MTN for their funding and support. I am extremely grateful to my friends, family, Block 7 and of course the boys for all the encouragement throughout the year. I also want to give a special thanks to IVY for believing in me as much as I believed in them. Finally, I want to acknowledge and thank the Water Research Commission for funding the Geasy project which provided a great source of data for this thesis.

Contents

List of Figures

List of Tables

Nomenclature

Acronyms and Abbreviations

ADHDP	Action Dependant Heuristic Dynamic Programming
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
BBS	Binary Backtracking Search
BIP	Binary Integer Programming
BLO	Binary Linear Optimisation
BPSO	Binary Particle Swarm Optimisation
BST	Binary Search Tree
CG	Column Generation
DDP	Deterministic Dynamic Programming
DP	Dynamic Programming
DSM	Demand-side Management
DTL	Data Translation Layer
EM	Schedule Control with Energy Matching
EML	Schedule Control with Energy Matching plus *Legionella* prevention
EWH	Electric Water Heater
GA	Genetic Algorithm
IoT	Internet of Things
MPM	Multi-Period Model
MQTT	Messaging Queuing Telemetry Transport
MSM	Minimal Simulation Model
OC	Optimal Control
PDR	Packet Delivery Rate
PSO	Particle Swarm Optimisation
QL	Q-Learning
RL	Reinforcement Learning
RLE	Run Length Encoding
RMS	Root Mean Squared
SC	Schedule Control
SDP	Stochastic Dynamic Programming
SEC	Smart EWH Controller
SPS	Shortest Path Search
STL	Seasonal and Trend Decomposition
TC	Thermostat Control
TM	Schedule Control with Temperature Matching
UTC	Universal Time Coordinated
WRC	Water Research Commission

List of Symbols

$\mathcal{U}$	set of admissible inputs
$\mathcal{X}$	set of admissible states
Δt	sampling period
γ	log normal distribution differing parameter
$\hat{h}$	specific enthalpy
$\mathcal{N}$	multivariate normal distribution
$\mathcal{N}_{\mathrm{i}}$	search tree node
$\mathcal{P}$	priority queue
$\mathscr{F}$	quantity being compared for either event start times or duration
μ	log normal distribution identical parameter
ρ	density of water
σ	standard deviation
σ^2	variance
τ	on-event duration time
A	tank area
c_{P}	constant pressure-specific heat capacity of water
c_{v}	constant volume-specific heat capacity of water
d	thermocline layer thickness
E	Energy of the tank
h	heuristic component
J	cost function
k	thermal conductivity of water
L	tank length
Q	volumetric flow rate
R	thermal resistance
r	tank radius
R^2_{diff}	normalised coefficient of determination
U	normal distribution
diff	differentiated data set
semidiff	semi-differentiated data set
undiff	undifferentiated data set

Chapter 1

Introduction

1.1 Background and motivation

The rise in population growth places strain on the supply capacity of electricity and the suppression of greenhouse gases all around the world [5]. Electricity generation in developing countries struggle the most to meet the needs of the electricity grid demand. According to the Global Energy Statistical Yearbook 2018, the global domestic electrical energy usage in 2017 was just over 22 000 TWh – a 67 % increase in usage from the year 2000 [6]. In Africa, it was 663 TWh – a 75 % increase from the year 2000 [1]. A major contributor of electricity consumption is the water heater, which contributes up to 40 % of a household's total energy usage [7]. In South Africa, which relies almost exclusively on electricity for energy, the approximately five million water heaters contribute an estimated 30 % to 50 % of the grid load at peak times, and 7 % of the country's daily energy requirements [7, 8]. South Africa also has a large dependence on fossil-fuels for generating electricity, with 88 % of the country's electricity generated from burning coal, making the environmental water heating footprint substantial [9, 2]. Moreover, approximately 38.5 GWh/day of electricity is used by domestic water heaters in South Africa, equating to a CO_2 release of 37 700 kg/day [1, 10, 11, 12, 13].

Demand-side management (DSM) aims to shift peak loads to reduce the intensity on the grid and prevent "blackouts". Many developing countries such as South Africa face challenges of monitoring the electricity usage of a community due to the lack of infrastructure and as a result, it makes DSM difficult to apply. This provides the possibility to take advantage of the EWH to reduce the energy demand and lift the burden off the grid. An EWH is a thermal storage device and therefore time insensitive, and allows for the implementation of control strategies and scheduling to reduce the overall electricity usage. Electrical energy is supplied to the EWH to increase the thermal energy inside the tank. Thermal energy that is lost to the surrounding environment of the tank is referred to as *thermal losses*, and it's rate is proportional to the temperature differential between the ambient temperature and that of the water inside the tank. Therefore, the overall electrical energy usage of the EWH can be reduced by providing it with the optimal control that finds a balance between minimising the tanks thermal losses as well as preventing the outlet water temperature from negatively affecting user satisfaction. Moreover, the maximum demand on the grid due to the EWH varies depending on the orientation of the EWH (i.e. vertical or horizontal) [14].

The design and implementation of optimal control for an EWH requires knowledge of the hot water demand profile of the household to produce energy-saving scheduling. The efficiency of the system depends on the accuracy of the water profile. This means considering the individual household's behavioural patterns that directly influence the profile shape. The other main determinant that must be accounted for is temporal variations, such as seasonal changes (e.g. thermal losses are influenced by inlet water temperature and ambient temperature), holidays, and day-to-day habits. Many existing control strategies for EWH control try to find the optimal time-of-use pricing for user benefits. Developing countries, however, use flat-pricing and neglect time-of-use pricing. Additionally, control strategies are important for reducing the carbon footprint and mitigating climate change. This highlights the importance of an energy-driven control strategy that does not disrupt the satisfaction of the users. Furthermore, determining the optimal EWH control requires a EWH model that accurately models the thermodynamics of the system with measurable outcomes as well as suitable optimisation techniques that are applicable for determining the optimal schedule for such a model.

1.2 Problem statement

The EWH requires a significant amount of electrical energy and it's impact on developing countries causes additional energy scarcity and greenhouse gasses due to coal dependence. To reduce their energy usage, control strategies and scheduling can be implemented to reduce thermal losses by ensuring that the water temperature is at a satisfactory level for the user only when it's necessary. To reduce the thermal losses to a minimum, the control strategy needs to be optimal. Additionally, the tank must be sufficiently heated at least once a day to prevent the growth of an infectious bacteria known as *Legionella pneumophila* [15].

The development of an optimal control strategy for practical use has its own challenges. The control strategy is more effective if its adaptive and the usage behaviour is modelled accurately. This means accounting for the variations in behaviour that is effected by the individual user and temporal factors such as seasonality. Furthermore, the physical characteristics of the water heater and the water inside, such as thermal resistance and stratification, must also be considered.

1.3 Research gap

Traditional thermostat control of residential EWHs is energy inefficient and does not take into account the user's hot water demand profile. Energy losses can be reduced if the water heating is optimised for the anticipated hot water demand profile. However, the hot water demand profile varies from user to user as well as the time of the week and year.

A literature study is performed in Chapter 3 on existing hot water usage forecasting models. The study found that no hot water usage model considers all the factors that influence a water profile. These are the individuality of a household, annual variations such as seasons and holidays, user behavioural patterns from day to day as well as high time resolution which is necessary for effective optimisation. There is also lack of these models available from research for user purposes and those that are available make many

assumptions such as labelling each water usage to specific end-use devices. Moreover, they lack the ability to generate hot water usage profiles with conservative water usages for stochastic optimal control techniques.

Optimal control techniques are used to optimise the energy usage of EWHs while satisfying the user's hot water demand profile. To enable optimal control, hot water usage models are required that can predict individual user behaviour with a sufficiently high temporal resolution, based on historically measured usage data. A suitable model of an EWH is also required to simulate optimal EWH control and obtain results for evaluating the optimisation approach. Water stratification should be accounted for in the model to improve the accuracy of the energy flow inside the tank as well as the accuracy of the results.

A literature study is performed in Chapter 4 on existing EWH models as well as related work on optimisation methods applicable to the EWH. Despite the plethora of EWH control optimisation research, no studies have applied the considered optimal control techniques using a EWH model that accounts for water stratification and none of the optimal control formulations account for all of the following: time-varying external disturbances, such as ambient temperature and cold water inlet temperature, and input constraints, such as scheduled electricity supply interruptions. None of the studies appear to have used high temporal resolution water data for their simulations (i.e. minutely data rather than hourly etc.). Previous optimisation methods typically focus on reducing time-of-use pricing. This is not applicable in developing countries which have flat-pricing (i.e. no time-of-use incentives) and rely on energy control. In these countries, households desire to reduce their overall energy demand, which will also reduce average grid demand. Taking South Africa as an example, power generation struggles to meet the energy demand. Additionally, reducing energy demand also reduces an individual's carbon footprint and consequently the effects of climate change. This highlights the need for an energy-driven optimisation strategy.

When evaluating the simulation results of an EWH with optimal control, it is important to define constraints for the system such that the user's satisfaction is not negatively affected by water usages when compared to a simulation where the EWH implements thermostat control. For example, studies do not ensure that the water temperature experienced by the user is equivalent to that of thermostat control during water usages. Therefore, the energy savings reported by these studies are not significant as the users satisfaction is jeopardised. Furthermore, not all studies make use of constraints that allows for the outlet water temperature of the tank to be lower than that of thermostat control during water usages, as long as an equivalent quantity of energy is used during the water usage. This is made possible by simulating the user adjusting the mixing of hot and cold water leaving the tap in reaction to the initial water usage temperature experienced.

Temperature feedback control for the EWH is required to ensure that the temperature inside the EWH follows the optimal reference temperature based on the user's hot water profile, and to compensate for deviations from the reference temperature caused by unanticipated usage events. Related works either do not mention how the temperature inside the EWH is corrected in the event of unanticipated water usage or the solution is vastly inefficient. There also exists no mention in related works on using the A* algorithm for

optimisation of an EWH.

Legionella pneumophila is a infectious bacteria that grows inside water tanks, making it a serious health risk to humans. However, this can be prevented if the whole tank is sufficiently heated at least once a day to ensure that the bacteria is sterilised [15]. No studies include optimisation that ensures the sterilisation of *Legionella*. Table 1.1 summarises the major optimisation research gaps in related work.

Table 1.1: Major optimisation research gaps in related works.

Ref	*Optimal* control	*Stratification*	*Temperature* matching	*Energy* matching	*A** optimisation	*Legionella* sterilisation
Sepulveda (2010) [16]	✓	✗	✗	✗	✗	✗
Kepplinger (2014) [17]	✓	✗	✗	✓	✗	✗
Kepplinger (2015) [18]	✓	✓	✗	✓	✗	✗
Kepplinger (2016) [19]	✓	✗	✗	✓	✗	✗
Al-Jabery (2016) [20]	✓	✗	✗	✗	✗	✗

1.4 Research questions

In the context presented here, and to address the problem statement above, this thesis will answer the following research questions:

1. How can residential electric water heater control be optimised to minimise the energy used while satisfying the individual user's hot water demand profile?
2. How can an individual user's hot water demand profile be predicted based on historically measured usage data, in a format that is useful for optimal control?
3. How much energy can theoretically be saved with optimal control for electric water heaters compared to traditional thermostat control, when perfect foreknowledge of the hot water usage profile is available?
4. How much energy can practically be saved with optimal control for electric water heaters compared to the traditional thermostat control as well as theoretical optimal control, when a predicted hot water usage profile based on historically measured data is used. The practical data acquisition and tank state estimation is not accounted for, which would be considered in a real-world implementation.

1.5 Research objectives

This thesis has the following objectives:

1. To develop optimal control techniques for residential electric water heaters that minimise the electrical energy used while satisfying the user's hot water demand profile and limiting the growth of *Legionella* bacteria. It will take into account time varying external disturbances, such as ambient temperature and cold water inlet temperature, and input constraints, such as scheduled supply-side interruption of the electricity supply.

2. To develop pattern recognition techniques to predict a user's hot water demand profile as a function of time, based on historically measured usage data, so that the switching control of the electric water heater can be optimised for the anticipated hot water demand profile.
3. To develop feedback control techniques that control the temperature inside the residential electric water heater to follow the planned optimal temperature trajectory, to reject disturbances such as unanticipated hot water usage, and to provide robustness to model uncertainty.
4. To develop temporal stochastic models for user's hot water usage behaviour based on the historically measured usage data, so that synthetic hot water draws can be generated for simulation purposes.
5. To determine how much energy can theoretically be saved with optimal control for electric water heaters compared to traditional thermostat control, when perfect foreknowledge of the hot water usage profile is available.
6. To determine how much energy can practically be saved with optimal control for electric water heaters compared to the traditional thermostat control, when a predicted hot water usage profile based on historically measured data is used.

1.6 Original contributions

The work presented in this thesis makes the following original contributions:

1. The residential electric water heater control problem was mathematically formulated in a novel way as an optimal control problem with the objective to find the heating element switching signal and the optimal temperature state trajectory to minimise the energy used while satisfying an anticipated hot water usage profile.
2. A novel dynamic programming-based technique was developed that solves the optimal control problem using a one-node, lumped-parameter model of the electric water heater.
3. A novel A*-based technique was developed that solves the optimal control problem using a two-node lumped parameter model of the electric water heater that takes stratification into account.
4. A novel feedback control technique was developed that controls the temperature inside the electric water heater to follow the planned optimal temperature trajectory, rejecting disturbances such as unanticipated hot water usage and providing robustness to model uncertainty.
5. A novel hot water usage model was developed that uses clustering and statistical analysis to model the user's temporal hot water usage behaviour based on historically measured usage data. The hot water usage model is used to predict the anticipated hot water usage profile for the optimal control algorithm.
6. A reactive hot water usage simulation model was developed that generates a synthetic hot water usage profile with random variations based on the clustering and statistical properties of the historical usage data. The model simulates the fact that the user will adjust the ratio of cold water and hot water mixing based on the temperature of the hot water.

7. The study determined how much energy can theoretically and practically be saved with optimal control for electric water heaters compared to the traditional thermostat control.

1.7 System overview

The goal of the system is to minimise the electrical energy used by the EWH while preventing the user from experiencing cold water temperatures for any given hot water usage profile. An optimal schedule planner is developed which produces the optimal heating schedule for the EWH heating element to achieve the goal of the system for any given household water usage profile. Either measured or predicted hot water usage data is used by the optimal planner. The overall electrical energy savings are determined by a simulator which simulates the EWH with the optimal plan and corresponding hot water usage data. The simulator is aided with a user water mixer and temperature feedback controller to minimise deviations between the optimal planning and simulation. An overview of the final system developed in this thesis is shown by the block diagram in Figure 1.1 and indicates in red the chapter that contributes to the development of the respective component block.

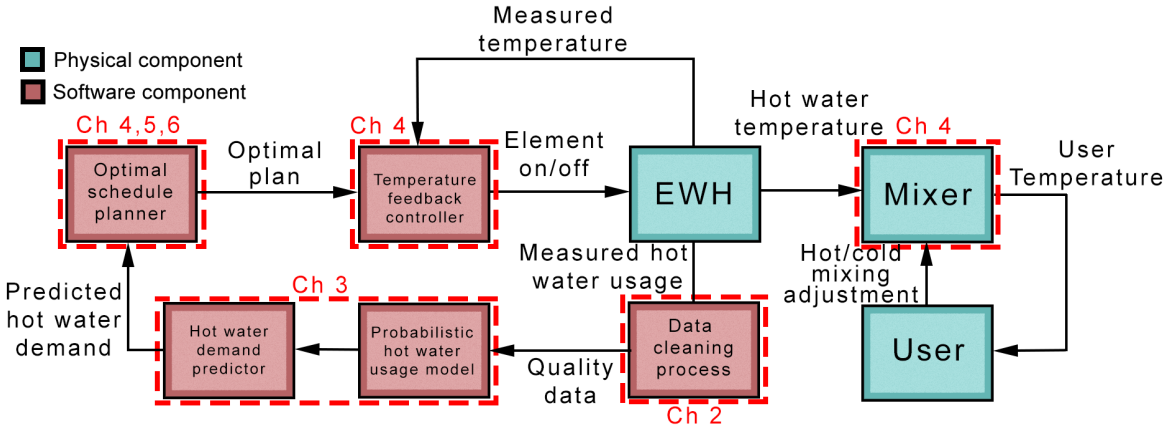


Figure 1.1: Block diagram of optimal EWH control determined by an optimal temperature planner. Components are indicated as a physical or software component of the system. The component contributed by each chapter are indicated in red.

The EWH is controlled by a heating element that can be switched either on or off. The optimal control sequence for the heating element and corresponding optimal temperature trajectory is determined by an optimal schedule planner. The optimal schedule planner could be implemented using either a model that does or does not account for water stratification. If the EWH model that does not account for water stratification is used, then a Dynamic Programming (DP) algorithm is used to solve the optimal control problem. If the EWH model that does account for water stratification is used, then the optimal control problem is solved using an A* search algorithm. The optimal schedule planner uses a predicted hot water demand profile to plan the optimal EWH temperature trajectory. The predicted hot water demand profile is provided by the hot water demand predictor, which in turn uses a probabilistic hot water usage model developed in Chapter 3. The probabilistic hot water usage model is obtained by fitting a probabilistic model to historical measured hot water usage data obtained from the EWH temperature and flow

rate sensors. The measured EWH temperature and flow rate data is cleaned by a data cleaning process developed in Chapter 2 before being used by the hot water usage model. The DP approach is developed in Chapter 5 to produce the optimal planning for water usage profiles that are perfectly know. The A* approach is developed further in Chapter 6 to also produce the optimal plan for when water usages are predicted.

The impact of optimal EWH control is evaluated by simulating the EWH with the optimal schedule and corresponding water usage profile and obtaining the results. For these simulations, it is assumed that the ambient and inlet water temperature for the EWH is assumed to be 20 °C. The simulator is aided with two control mechanisms which are developed in Chapter 4 to ensure all the system goals and constraints are satisfied. The first is a temperature feedback controller developed to provide the simulation with the ability to follow the optimal temperature trajectory. If the EWH deviates from the optimal temperature, as a result of unanticipated water usages or model uncertainties, the feedback control system controls the water temperature inside the EWH to follow the optimal temperature by actuating the heating element based on feedback from the temperature sensor. The second is a mixer developed to provide energy matching when the temperature deviates from the optimal plan. The user is simulated to experience the hot water temperature at the EWH outlet and use a mixer to adjust the hot and cold mixing of water at the tap to achieve their desired user temperature. The simulator can also differentiate between intentional and unintentional water usages. Unintentional water usages refers to water drawn from the tank that is not effectively used. This is due warm water remaining in the piping network between the EWH and the point of use that will contribute to energy losses since the water will cool down.

The heating control strategies, EWH models and optimisation techniques are developed in Chapters 4, 5 and 6 and all the results of these developed systems are evaluated in Chapter 7. Moreover, this thesis also evaluates: the impact of the *usage losses* that result when hot water is drawn unintentionally; the impact of simulating an EWH that models stratification with a optimal schedule which was produced without accounting for it; and the impact of optimal control that uses predicted water usages opposed to having perfect foreknowledge of them.

1.8 Thesis overview

The research performed for this thesis will be presented as follows:

- Chapter 2 presents a data cleaning process where anomalies are identified in raw EWH measurement data and high-quality data is produced for the next chapter.
- Chapter 3 presents the development of the probabilistic hot water usage model. First, a literature study is performed on existing hot water usage models. Then, a probabilistic hot water usage model is developed that uses historical hot water usage data measured by temperature and flow rate sensors in the EWH. A hot water usage simulator is also developed that implements the model to produce random variations of representative hot water usage profiles for individual households. Finally, a hot water demand predictor is developed to produce conservative profiles from the probabilistic model.

- Chapter 4 presents the development of the EWH thermodynamic models, as well as the formulation of the optimal control problem. First, a literature study is performed on existing types of EWH models, methods for EWH control and different optimisation techniques. Then, heating control strategies are defined which will be considered for optimisation. The one-node and two-node EWH thermodynamic models are then developed and the optimal control problem is finally formulated.
- Chapter 5 presents the development of the dynamic programming algorithm that solves the optimal control problem using the one-node EWH model.
- Chapter 6 presents the development of the A* search algorithm that solves the optimal control problem using the two-node EWH model that includes the effects of stratification.
- Chapter 7 presents the results of the simulation tests that were performed to evaluate the performance of the EWH optimal control system, and to compare the ideal performance when the predicted hot water usage is perfectly known and the practical performance when the hot water usage is predicted with the hot water usage model.
- Chapter 8 provides a summary of the work performed, presents the key findings, and gives recommendations for future research.

Chapter 2

Data cleaning

This chapter explains the data cleaning process which was used to obtain reliable data for the system developed in this thesis. It should be noted that data cleaning is not part of the final system, but is rather a "one-time" necessity for the Geasy data that was acquired for the purpose of this study. Data cleaning refers to eliminating any fault in the data, such that what was supposed to be measured is not reflected in the data captured in the database. This includes data errors, inconsistencies, and object identification problems [21]. The process involves identifying anomalies within an imperfect data set, often referred to as "dirty data", by applying a sequence of operations to eliminate anomalies and verifying the effect of the newly processed data.

Figure 2.1 shows a block diagram that summarises the system developed in this chapter and the process that was followed to obtain high-quality, clean EWH data profiles from raw data measurements. The sections that cover each component are indicated in parenthesis. In Section 2.2, the acquisition of the data is explained. Following this, in Section 2.1, the concept of data quality is explained first followed by different types of data anomalies and quality criteria which measure the *correctness* of the data. A system is developed in Section 2.3 which first identifies existing anomalies and the extent of quality for the raw data. Tools are then developed to clean the data and satisfy the quality criteria. Lastly, in Section 2.4, the data is analysed to verify that it meets a satisfactory level of quality and is compiled into a high-quality database.

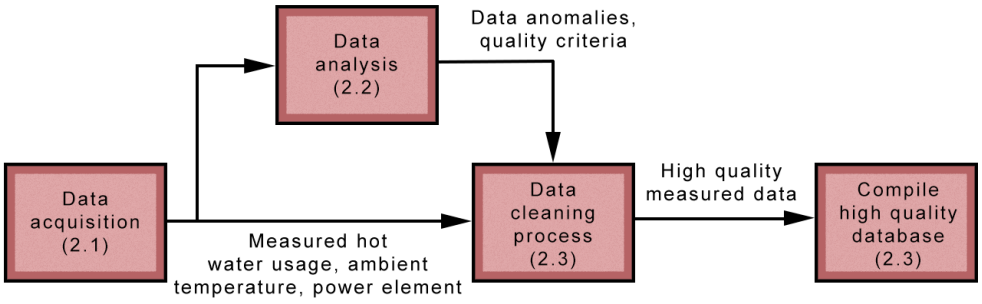


Figure 2.1: Block diagram of data cleaning system developed in this chapter.

2.1 Data acquisition

The data for this project came from the Geasy Smart EWH Controller project, which was coordinated by BridgIoT and funded by the Water Research Commission (WRC). BridgIoT is a company that integrate and develop solutions for internet of things (IoT) systems for consumers [22]. The aim of the project was to install Smart EWH controllers (SECs) on EWHs to monitor and control water and electricity usage in households. The project aimed at installing these for over 400 households.

2.1.1 Geasy SEC

The Geasy Smart EWH Controller is a device developed by BridgIoT and aims to reduce electrical energy by temperature set-point control as well as reduce water usage. The main features of the Geasy are temperature control, leak detection and notification, water metering, energy measurement and remote water-inlet shutoff. The Geasy also allows for manual and automatic heating scheduling. It claims to save up to 30 % energy consumption of a geyser and multiple EWH installations contributing to peak demand management [22]. The idea is to replace the Geasy's current scheduling function with the usage-based optimal energy control algorithms proposed in this thesis. However, this chapter focuses on the Geasy's data collection function.

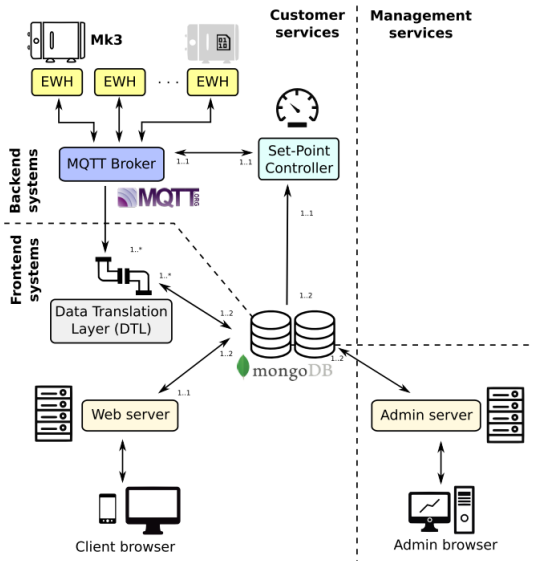


Figure 2.2: Architecture of system showing interface of Mk3 sensors with MQTT Broker and database [23].

Figure 2.3: Mk3 EWH device [23].

2.1.1.1 SECs

The SEC used for the Geasy project was developed in the Master's project of Cloete (2017) [23] which contributed to the work discussed in this section. The author uses this work to provide insight on the device used to collect data.

SEC architecture:

Figure 2.2 shows the overall architecture of the system. A broker receives measured data from the controllers and sends set-point data back to the Mk3 controller. A translation layer reads all data sent to the broker and sends this data to be stored in a database.

MQTT broker: The Message Queuing Telemetry Transport (MQTT) is a publish-subscribe protocol [24]. The *broker* is a central server where clients can connect and send messages to the broker under a specified topic which is then sent back to other clients subscribed to this topic.

The EWH controllers publish data samples periodically along with an IMEI number unique to each device to the broker. The controllers also subscribed to the set-point topic which allows for the set-point controller to publish set-points to the different EWHs.

Data Translation Layer (DTL): This is an application that provides an interface that connects the broker and the database. It is subscribed to all EWH topics so that it can read all time-stamps published by EWHs and store them in a database.

MongoDB: MongoDB is an online document database with scalability and flexibility. All data that is read by the DTL is sent and stored in this database. The data used in this project comes from this database.

Sensors: The EWH controller is comprised of sensory hardware developed by Brown (2016) [25]. Temperature sensors are installed externally to measure specific sections of the EWH. Measurements have an accuracy of $1°C$. The temperature is measured at four different sections: On the surface of the pipe at the outlet, 500 mm along the outlet pipe, 500 mm along the inlet pipe, and in the air within a 1m radius of the tank.

A flow-meter at the inlet uses a Hall effect sensor to measure the volume of water entering the EWH at a resolution of 0.5L. The data is recorded with the unit L/minute or L/sample. The power meter comprises of an inductive loop which measures the current delivered to the element with a voltage of 230V RMS. Power is calculated as a product of V_{RMS} and I_{RMS} and has an accuracy of 5 %. The unit for recorded power is kW. Figure 2.3 shows the physical layout of the Geasy.

Actuator: A latching relay is placed in series with the thermostat to switch the state of the element of the EWH. Utilising the latch allows for the state of the relay to be retained during the event of a power failure.

Collected data performance: Data is sent from the EWH controller once per minute. The packet delivery rate (PDR) represents the portion of the packets successfully received of the expected amount. The results of experiments in [23] were analysed and revealed that the average PDR per day was 78 %. A number of factors contributed to data loss:

- Signal strength
- Network provider
- Clients turning power off
- Mobile data depletion

Data samples: The database acquired from Geasy contains minutely EWH measurements for numerous households. The first field in each data sample of each profile is a time stamp. Each time stamp gives an accuracy up to the correct second of the time the data was recorded. Each time stamp is followed by "+02:00" which refers to the UTC offset and indicates the time zone used in South Africa. The attributes associated with each time stamp are power, the four temperature measurements and water flow.

The motivation for data cleaning is that the raw database contains profiles where anomalies are observed, caused by technical issues, and the time domain covered by each profile can vary from less than a month to over a year. The final system needs profiles that cover all four seasons. Therefore it is important that the profiles used contain sufficient and accurate data.

Run Length Encoding(RLE): This is a form of data compression where sequences of repeated values within a specific data element are converted into a count of the occurrence of that consecutive value followed by the value. For example, a given data sequence "XXXXGGJJJ" would become "4X2G3J" after a RLE algorithm is applied. RLE is applied to the EWH data before it is stored in the database. The type of data used in this process includes numerous periods of time where water flow remains zero and temperatures do not change. The number of required data samples is therefore greatly reduced by RLE which only leaves data samples in the database where any of the measurement fields changes from the previous time stamp.

2.2 Data Quality

Different studies show a variety of definitions for the term *data quality*. An easy way to summarise it is simply defining it as the ability of data to meet the requirements of the user, i.e. fitness for use [26]. When looking at the data of any real world application, high data quality is vital to achieve success. Without this, the source data and the obtained results are both unreliable. Redman (2002) says that unless extreme efforts are made, the error rate of data is approximately 5 % [27]. To get a better picture of the importance of clean data, a 2001 New York survey from Price Waterhouse Coopers concluded that of 599 companies, 75 % of them had economic losses due to data quality problems. 37 % of these data-driven companies were actually very confident in the quality of their own data, and 15 % were very confident in their trading partners' quality of data [26, 28].

The definitions of data anomalies and quality criteria in the following two sections are based on the paper *Problems, Methods and Challenges in Comprehensive Data cleaning* by Muller *et al.* (2003) [29].

2.2.1 Data anomalies

Data anomalies refer to entities in data which degrade the data effectiveness, processing performance and hinders conclusions that can be made from interpreting the data. Data anomalies are what give databases properties of bad quality data. Once the different types of anomalies are identified, one can then determine different criteria to score the quality of the data within a database.

Databases: The material for this project is provided by a database. Databases are made up of files known as data sets. Each data set represents the raw measured data for an individual EWH. A data set consists of data objects known as tuples/data samples/instances (representing the rows) and are described by features known as attributes (representing the columns). Tuples refer to a set of related values from a finite set of domains. In a database, data anomalies can occur and are classified into 3 categories.

2.2.1.1 Syntactical anomalies

Syntactical anomalies refer to anomalies which are present in the format of the data.

Lexical Errors: These errors occur when data does not match the specified structure intended. If a data tuple is expected to have 4 values attached to it and appears that the entry has less than expected, this results in a lexical error.

Domain Format Errors: These errors occur when an attribute of a data object does not conform with the specified format. This can cause data which contains correct information to be regarded as unreadable by a data process.

Irregularities: These errors occur when there are inconsistencies in the values for an attribute. This can be caused when an attribute only accepts a value without a unit of measurement, then the value could be representing a unit which is not assumed for that field. Irregularities can also be caused by different uses of abbreviations.

2.2.1.2 Semantic anomalies

Semantic anomalies refer to anomalies which degrade the accuracy of representing the *real world* and anomalies which cause redundancy.

Integrity constraint violations: This refers to the value of an attribute falling outside the domain of acceptable values and therefore being incorrect. Constraints should be enforced so that only acceptable values are used to accurately describe the *real* world.

Duplicates: This refers to multiple tuples which describe the same instance. These tuples are not necessarily identical, and in the case of being non-identical it becomes a challenge to determine how to deal with the duplicates.

Invalid tuples: These anomalies are the hardest types of anomalies to find as they refer to anomalies which are not defined by any of the other types. This type of anomaly can be unknown and can sometimes be impossible to identify from the data.

2.2.1.3 Coverage anomalies

Coverage anomalies decreases the amount of data that is available to describe the *real world*.

Missing values: This refers to missing attributes within the tuples. Although the value could be NULL, sometimes a NULL value can be allowed. This depends on if the integrity constraints allow for NULL values. In these cases, one would need to investigate if the

value should exist to identify an anomaly.

Missing tuples: This refers to tuples which are missing from the database which describe a unique instance in the *real world*.

2.2.1.4 Relevant anomalies

The following table summarises the relevant anomalies that were identified in the raw EWH data set.

Table 2.1: Table showing relevance of all anomalies for the raw data.

*A*nomalies	Relevant			Irrelevant
	Very	Moderately	Unlikely	
Missing Values	✓			
Missing Tuples	✓			
Integrity Constraint Violations		✓		
Duplicates		✓		
Lexical Errors				✓
Domain Format Errors				✓
Irregularities				✓
Invalid Tuples				✓

Connection Dropouts: A major cause of anomalies in the data provided for this thesis is dropped connections between the EWH and the server, which stores data in the database. This causes various anomalies to appear throughout the data. The most frequent anomaly to occur is *missing tuples*. The EWH controller sending the data is equipped with a buffer to mitigate connection dropouts. The buffer accumulates the amount of water used during the duration of the dropout. When the connection is re-established, the accumulated water usage is sent to the database as the usage for a single time instant. This means that misleading measurements are sent directly after a connection dropout which could result in misinterpretation of the actual events. This causes *integrity constraint violations* due to the anomaly.

2.2.2 Quality Criteria

A set of criteria must be satisfied to confirm a good quality of data. The quality criteria are a set of different scores given to data to rate it on its performance. The criteria are directly related to anomalies and give a means of summarising the extent of damage that the anomalies cause. A score is defined such that it describes a specific criterion effectively and is used both for determining if data cleaning is required as well as analysing the effect of the data cleaning process on the data.

2.2.2.1 Completeness

Completeness is defined as the portion of actual data samples available out of the overall expected amount required to completely describe the mini world. A mini world is a

representation of a certain part of the real world. Completeness is related to coverage anomalies and the formula for scoring completeness is defined as follows:

$$\text{Completeness} = \frac{\#\text{ of actual data samples}}{\mathbb{R}} \tag{2.1}$$

where $\mathbb{R}$ is the expected amount of data samples required to completely describe the system.

2.2.2.2 Validity

Validity is defined as the portion of data samples in $\mathbb{R}$ without semantic anomalies that violate the integrity constraints imposed. The formula for scoring validity is defined as follows:

$$\text{Validity} = \frac{\#\text{ of actual valid data samples}}{\mathbb{R}} \tag{2.2}$$

A data sample is only valid if it satisfies all the integrity constraints imposed on any of the attributes. In context with the type of data given, the constraint is imposed on measurements such as water flow, temperatures, and power to ensure all their values are expected.

2.2.2.3 Density

Density is defined as the portion of data points in $\mathbb{R}$ which do not contain missing values within the tuples. The formula for scoring density is defined as follows:

$$\text{Density} = \frac{\#\text{ of actual data samples not missing values}}{\mathbb{R}} \tag{2.3}$$

2.2.2.4 Accuracy

Accuracy is defined as the portion of correct data samples in $\mathbb{R}$. It describes the combination of completeness, validity and density for the given data and the formula for scoring accuracy is defined as follows:

$$\text{Accuracy} = \text{Completeness} - \text{Validity}' - \text{Density}' \tag{2.4}$$

where Validity$'$ and Density$'$ refers to their respective complementary percentages. Equations 2.1, 2.2 and 2.3 all share a similar denominator which means that each criterion can be subtracted from one another to maintain a similar score. Furthermore, it was later discovered that density anomalies cannot be identified in the data because a data sample only includes a value for a particular attribute if the value changed from that of the previous sample. This means that there is always an uncertainty whether values are actually missing for any data sample. Equation 2.4 is modified as follows:

$$\text{Accuracy} = \text{Completeness} - \text{Validity}' \tag{2.5}$$

2.2.2.5 Uniqueness

Uniqueness is defined as the portion of unique data samples in $\mathbb{R}$. The formula for scoring uniqueness is defined as follows:

$$\text{Uniqueness} = \frac{\#\text{ of unique data samples}}{\mathbb{R}} \tag{2.6}$$

This criterion describes the portion of data that is not regarded as duplicates. The presence of duplicate data samples is ignored when calculating completeness as they increase the size of databases without giving any new useful information, whereas completeness finds the portion of useful data. Although duplicates exist in the database, they are not considered for the data cleaning process as they do not reduce the quality of data.

2.2.2.6 Quality

Quality summarises the resulting score of all the other criteria. The formula for scoring quality is defined as follows:

$$\text{Quality} = \text{Accuracy} = \text{Completeness}' - \text{Validity}' \tag{2.7}$$

This shows that quality accounts for all anomalies and is the final calculation for assessing data quality. Figure 2.4 shows the relationship of the different criteria that makes up data quality and shows which ones are effective. The figure shows that data quality is equal to data accuracy.

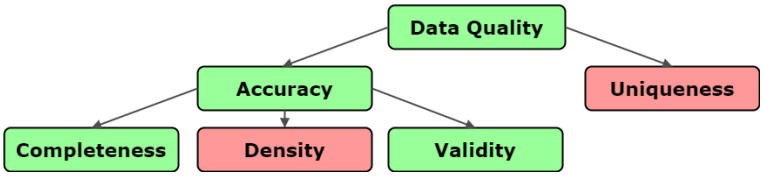


Figure 2.4: Relationships between the effective (green) and disregarded (red) criteria.

2.3 Data cleaning process

Data cleaning involves data going through a sequence of stages. These include identifying anomalies which cause the reduction in data quality, deciding on and applying operations to the collected data and finally analysing the results. The last step shows the effects of the operations on the data to determine the improved quality, to identify new anomalies and anomalies that were not completely reduced and to determine if the data should undergo the whole process again in a continuous loop until the data quality is at a satisfactory level.

2.3.1 Overview of process

This chapter follows a data cleaning process as defined by Muller *et al.* (2003) [29]. A flow diagram is represented in Figure 2.5 with the respective subsections that describe each step. Following this process produces data that is free from anomalies and is the most accurate representation of the mini-world.

The first step is **data auditing** which is identifying existing anomalies in the database. This step involves statistical and parsing methods to detect whether anomalies occur and what characteristics they possess. Section 2.1.1 summarises anomalies existing in the data; Section 2.3.3 discusses how connection dropouts can be identified; and Section 2.3.5 shows how anomalies are identified. This sometimes includes a statistical approach.

The next step is **specification** which is the process of detection and elimination of anomalies. A process is designed which produces clean data using a flow of operations. The proceeding step - **execution** - is executing the process designed in the previous step. Section 2.3.4 covers both these steps.

The final step is **post-processing** which is investigating the results to determine if the goal of data cleaning has been obtained. This step usually branches back to the first step in the scenario where data must be cleaned further. However, this is not the case in this thesis as the process is designed such that a satisfactory level of quality is obtained after one iteration. This is covered in Section 2.3.6.

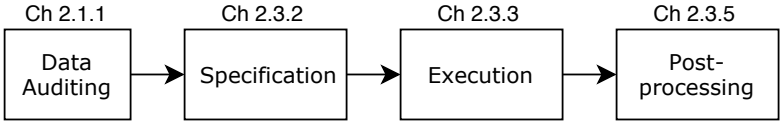


Figure 2.5: Overview of data cleaning process, specifying sections relevant for the steps.

2.3.2 Data Requirements

The final results of data cleaning produce good quality profiles for the monitored data for a number of EWHs which is the input data for the following chapters. Each profile must consist of minutely time-stamps without any missing samples. Moreover, associated with each should be the sample water flow rate (L/minute), ambient temperature (°C) and power (W) with consistent units. From analysis of the database performed later in this chapter, the time period that the monitored data was gathered started mid 2017 and ended in December 2018. Only seasons and holidays within these years therefore need to be considered for the process.

Seasons: Seasonality is important for this thesis as it affects the forecasting of a user's water usage profile because of their seasonal behaviour changes due to the weather. Significant weather factors are temperature and precipitation [30, 31].

Holidays: Holidays are avoided for the desired profiles as they include unexpected behaviour patterns and periods of useless data when the occupants are away. A constraint that prevents the inclusion of holidays must thus be implemented in the reconstruction of profiles. The consideration of seasons and holidays in the data cleaning process require knowledge of the start and end dates of these time periods. Table 2.2 shows the dates of each season and major holidays (December/January and June/July) for 2017 and 2018.

Day of week: Considering the day of the week when constructing each profile is important as this also impacts the users' behavioural patterns. There can be multiple reasons

Table 2.2: Dates of interest for cleaning process in a "MM-DD" format for each year considered. Subscripts h1 and h2 refer to the two major holiday periods.

Symbol	Description	2017	2018
Seasons			
T_{summer}	Starting date of the Summer season.	12-01	12-01
T_{autumn}	Starting date of the Autumn season.	03-01	03-01
T_{winter}	Starting date of the Winter season.	06-01	06-01
T_{spring}	Starting date of the Spring season.	09-01	09-01
Holidays			
$T_{h1\|start}$	Starting date of the December/January holiday.	12-06	12-13
$T_{h1\|end}$	Ending date of the December/January holiday.	01-17	01-08
$T_{h2\|start}$	Starting date of the June/July holiday.	06-30	06-23
$T_{h2\|end}$	Ending date of the June/July holiday.	07-24	07-16

for this, such as a household doing washing on the same day each week or less frequent water usages on weekends [30, 31].

A month (4 weeks) is constructed for each season for the resulting clean profiles. Each month is extracted from weeks that occur consecutively after one another and the correct day of the week structure is retained. For a fair comparison of each season and profile, a constraint is declared that all profiles must start on the same day of week (i.e. Monday).

2.3.3 Challenges

Due to the RLE format of the data, missing samples cannot be identified easily but rather require identifying the occurrence of connection dropouts. Connection dropouts are important to identify as they are the cause of most anomalies existing in the data. A difficult challenge arises in the cleaning process where characteristics of dropouts must be identified through careful observation. A value of zero is sent for the power and water flow attributes when a dropout begins and ends. These can occur during periods of no water usage, periods where water usage starts or ends, and over multiple days. The solution is an algorithm that can identify each of the dropout scenarios in the data.

2.3.4 Data tools

This section describes the full process of data cleaning operations that takes in the raw data and produces the final set profiles.

2.3.4.1 Python

The Python programming language was used in conjunction with *Jupyter Notebook*. The *Numpy* and *Panda* packages were used for reading, writing and manipulation of data and the *MATLAB Plots* package was used for graphing. These packages were also used for the implementation of the processes discussed in further chapters.

2.3.4.2 Constraints

Constraints are defined for the range of acceptable values for the different types of measurements. Any measurements outside the specified constraints will result in invalid data

and contribute to a profile's validity.

Water flow: The maximum possible water flow rate at the EWH outlet pipe that is accepted as valid depends on the maximum possible water flow of the physical piping. The outlet of an EWH uses a valve and pipe which both have a diameter of typically 22mm. Modern EWHs use a flow pressure of 400-600kPa [32]. The rated water flow for these pipes from 400 to 600kPa is approximately 90 to 110L/min [33]. Since the exact pressure of each individual EWH is not known, the constraint for maximum water flow, $W_{\max}$, is 110L/min.

Ambient temperature: The temperature range expected for South Africa is between 0 and 50°C. However, since an EWH is typically found inside the ceiling of a household, temperatures can assume to increase by up to 10°C. Therefore $T_{\text{amb(max)}}$ is equal to 60°C.

Power: The rated power P_{rated} of the element for an EWH is identified for each EWH by looking at where the majority of non-zero measurements are centralised. Power measuring sensors are known to have a 5 % inaccuracy and the power measurement constraints are defined as follows:

$$P_{\max} = P_{\text{rated}} \times 1.05 \tag{2.8}$$

$$P_{\min} = P_{\text{rated}} \times 9.95 \tag{2.9}$$

2.3.4.3 Best Quality Profiles

The main goal of data cleaning is to produce profiles reconstructed from the best quality data available from monitored EWH data. The first step is finding the quality of each day due to anomalies as well as days missing. The data reconstruction process is divided into two phases. The first phase is identifying the best window of potentially good data that throws out the least bad quality profiles. The second phase assesses the quality of the identified window and reconstructs the profile by replacing the worst days with better quality days and finally interpolating the profiles.

2.3.4.4 Dropout Detection

Connection dropouts are first identified to examine the anomalies in the raw data set. The occurring dropouts are identified and the missing amount of samples during the dropout are recorded. A set of connection dropouts D is declared as follows:

$$D_{\text{i}} = \left\{ d_{\text{i}}[1], d_{\text{i}}[2], ..., d_{\text{i}}[N_{\text{i}}] \right\} \tag{2.10}$$

where i indicates the EWH and N_{i} is the last day of the EWH raw profile. Each entry d[n] represents the total number of dropout samples per day. The summation of the dropout samples for a single EWH is calculated as follows:

$$D_{\text{i}}^{*} = \sum_{n=1}^{N_{\text{i}}} d_{\text{i}}[n] \tag{2.11}$$

Equations 2.10 and 2.11 are used next to identify and obtain scores for the whole EWH profile and score for each day within it, respectively.

2.3.4.5 Quality per day

Quality is the score obtained for the reconstruction phases. A set of quality scores for each day is declared as follows:

$$Q_{\mathrm{i}} = \left\{ q_{\mathrm{i}}[1], q_{\mathrm{i}}[2], ..., q_{\mathrm{i}}[N_{\mathrm{i}}] \right\} \tag{2.12}$$

$$q_{\mathrm{i}}[n] \in [0,\ 100] \tag{2.13}$$

$$Q_{\mathrm{i}}^{*} \in [0,\ 100] \tag{2.14}$$

where each quality entry $q_{\mathrm{i}}[n]$ and overall EWH quality Q_{i}^{*} represents a percentage of data quality. The score of completeness and validity is required to obtain quality.

Completeness: A set of completeness scores for each day is declared as follows:

$$C_{\mathrm{i}} = \left\{ c_{\mathrm{i}}[1], c_{\mathrm{i}}[2], ..., c_{\mathrm{i}}[N_{\mathrm{i}}] \right\} \tag{2.15}$$

$$c_{\mathrm{i}}[n] \in [0,\ 100] \tag{2.16}$$

$$C_{\mathrm{i}}^{*} \in [0,\ 100] \tag{2.17}$$

where each completeness entry $c_{\mathrm{i}}[n]$ and overall EWH completeness C_{i}^{*} represents a percentage of data completeness. The per day completeness $c_{\mathrm{i}}[n]$ is calculated as follows:

$$c_{\mathrm{i}}[n] = \frac{1440 - d_{\mathrm{i}}[n]}{1440} \times 100 \tag{2.18}$$

The value of 1440 is used as this is the number of expected samples for a single day assuming minute intervals. The overall EWH completeness is calculated as follows:

$$C_{\mathrm{i}}^{*} = \frac{1440 \times N_{\mathrm{i}} - D_{\mathrm{i}}^{*}}{1440 \times N_{\mathrm{i}}} \times 100 \tag{2.19}$$

Validity: A data sample is invalid if any of the considered measurements do not satisfy the defined constraints. The constraints are implemented as follows:

$$W \leq W_{\max} \tag{2.20}$$

$$T_{\mathrm{amb}} \leq T_{\mathrm{amb(max)}} \tag{2.21}$$

$$P_{\min} \leq P \leq P_{\max} \tag{2.22}$$

A set of validity scores for each day is declared as follows:

$$V_{\mathrm{i}} = \left\{ v_{\mathrm{i}}[1], v_{\mathrm{i}}[2], ..., v_{\mathrm{i}}[N_{\mathrm{i}}] \right\} \tag{2.23}$$

$$v_{\mathrm{i}}[n] \in [0,\ 100] \tag{2.24}$$

$$V_{\mathrm{i}}^{*} \in [0,\, 100] \tag{2.25}$$

where each validity entry $v_{\mathrm{i}}[n]$ and overall EWH validity V_{i}^{*} represents a percentage of data validity. The per day validity $v_{\mathrm{i}}[n]$ is calculated as follows:

$$v_{\mathrm{i}}[n] = \frac{\sum_{k=1}^{1440} x_{\mathrm{i}}[k]}{1440} \times 100 \tag{2.26}$$

where $x_{\mathrm{i}}[k]$ is equal to one if Equation 2.20, 2.21 and 2.22 are satisfied for all measurements for the k^{th} sample of the day. The overall EWH validity is calculated as follows:

$$V_{\mathrm{i}}^{*} = \frac{\sum_{n=1}^{N_{\mathrm{i}}} v_{\mathrm{i}}[n]}{N_{\mathrm{i}}} \tag{2.27}$$

Quality: Each per day quality entry and overall EWH quality is calculated as follows:

$$q_{\mathrm{i}}[n] = c_{\mathrm{i}}[n] - v_{\mathrm{i}}'[n] \tag{2.28}$$

$$Q_{\mathrm{i}}^{*} = C_{\mathrm{i}}^{*} - V_{\mathrm{i}}'^{*} \tag{2.29}$$

where $v_{\mathrm{i}}'[n]$ and $V_{\mathrm{i}}'^{*}$ are the percentage complements. The value of Q_{i}^{*} allows for EWH profiles to be discarded from the following phases if they do not satisfy a profile quality threshold Q_{thresh}.

2.3.4.6 Phase One: Window Searching

The description of these phases of profile reconstruction refer to building up a 28-day profile for one season and then repeating the process for all seasons. The first phase is to select the month for each season that has the best potential for good data quality. Determining this potential is done by first defining a count window with a width of 28 days. The window looks at which EWH profiles contain missing days during the same duration of the window. A missing day is easily assumed for EWH i on day n if the value for $c_{\mathrm{i}}[n]$ is 0 %. The window counts the number of profiles that contain at least one missing day inside the window rather than the total number of missing days of all the profiles. This is because the purpose of the window is to keep only profiles containing no missing days for a specific period and therefore minimise the number of discarded profiles.

Each month window of the considered season is denoted by $m_{\mathrm{i}}[k]$, where i refers to the EWH and k refers to the index of the week that the month starts on. The first day of the window is always Monday and each iteration will move the window forward by a week until the last week of the window is the last week of the season. A value of one is given $m_{\mathrm{i}}[k]$ if at least one of its days is missing. The overall window score for missing days is calculated as follows:

$$M[k] = \sum_{i=1}^{I} m_{\mathrm{i}}[k] \tag{2.30}$$

where I is the number of EWHs considered. This gives a count of profiles that have at least one missing day in the given period. A diagram of the process for phase one is shown in Figure 2.6.

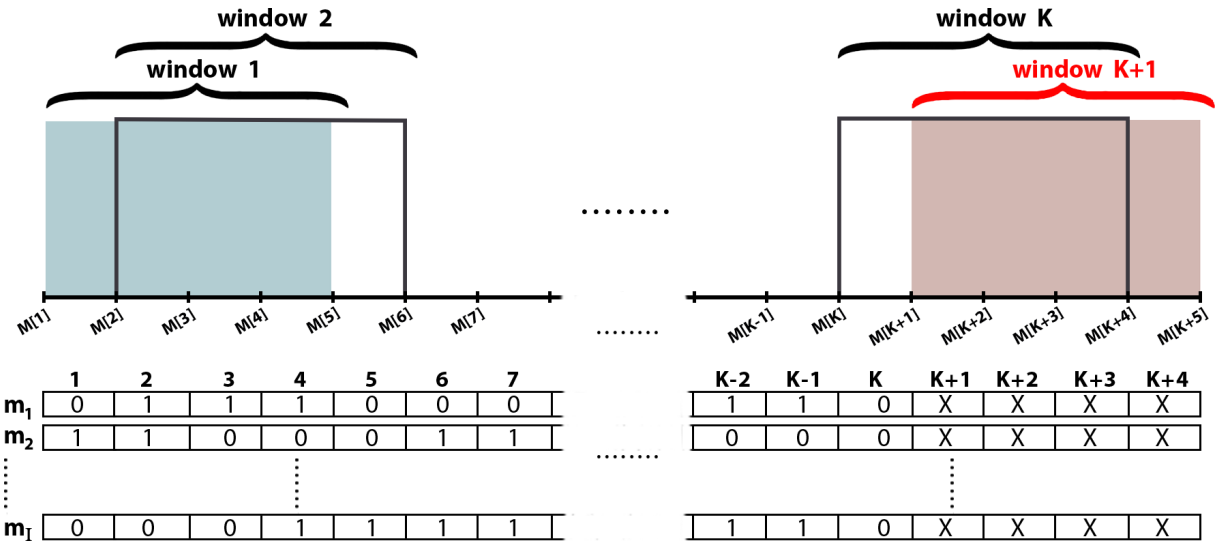


Figure 2.6: Diagram showing process of the count window. Iterations for windows start at the first day of $M[1]$ and finish at the K^{th} window, where all following windows are invalid as they fall outside of the season bounds.

Once all the iterations of the window are complete, the optimal month is declared as M^* and is calculated as follows:

$$M^* = \min(M) \tag{2.31}$$

Holiday filter: During any iteration of the phase, any values of $M_{\text{i}}[k]$ that contain any public or school holiday dates are discarded. This is modelled as a black box that filters out holiday periods from the windowing selection.

2.3.4.7 Phase Two: Month Reconstruction

The optimal month allows for each EWH profile to extract the 28-day period from the raw data with the associated daily quality scores from Q_{i}. These days extend from $q[1]$ to $q[28]$. The prevention of missing days in the remaining profiles still does not ensure that each day has sufficient data quality. Each profile undergoes a month reconstruction process where days not satisfying a daily quality threshold q_{thresh} are replaced with days from the same EWH that do satisfy it.

It must be ensured that days are replaced with the same day type (e.g. Monday must be replaced with a Monday). Therefore, the daily quality of all-day types are separated into sets, where the Monday set is declared as follows:

$$Q_{\text{mon, i}} = \left\{ q_{\text{mon, i}}[1], q_{\text{mon, i}}[2], ... \right\} \tag{2.32}$$

This set is also sorted in descending order of quality; the first item on the list is the first pick for replacing a bad quality day. This allows for bad days to be replaced with the best possible quality days that exist in the EWH profile for that season. This step is completed when all the scores for the 28 days satisfy q_{thresh}. Figure 2.7 shows a diagram of this process. The associated raw data for each of these day scores are extracted and connected in the correct order to create profile P_{i}^*.

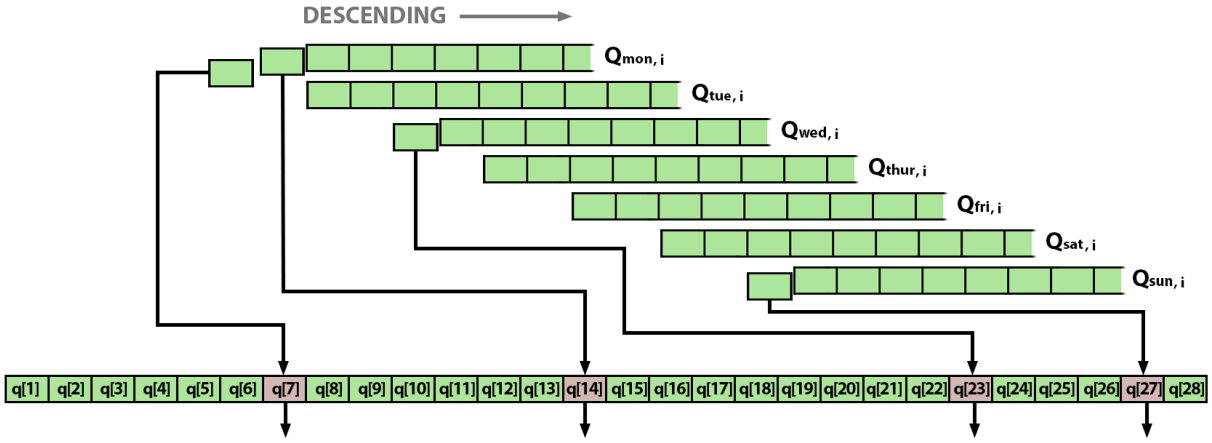


Figure 2.7: Diagram showing process of the month reconstruction. Green blocks represent days with data quality satisfying q_{thresh} and red blocks represent days that do not.

Run Length Encoding: Although the raw data presents the data in an RLE format, the ideal final profiles of the cleaning process is to include a data sample for all samples within the time period. The unrepresented data samples are filled in and measurements are given duplicated values of the previous existing data sample. Data samples are also filled in where data is missing due to dropped connections. The measurements for these data samples are left empty as they are chosen when the profile is interpolated.

Duplication: When a duplicated data sample is found in the raw data, the second data sample is ignored as it holds identical information to the first and is redundant.

Irregular sampling: Due to the irregular sampling, data samples will sometimes skip a sample or create a duplicate sample. When sample n is skipped due to irregular sampling, a data sample is inserted with its measurements taken as the average of the previous and next sample and is shown in equation 2.33. This equation is also used to calculate the data sample that replaces the two duplicated data samples.

$$x[n] = \frac{x[n-1] + x[n+1]}{2} \tag{2.33}$$

2.3.4.8 Interpolation

The final step to producing clean EWH profiles P_{i} is interpolating missing data measurements.

Ambient temperature: Linear interpolation is used to complete the ambient temperature profile because it is a good approximation of the expected temperature gradient to fill in the period between the temperatures before and after the missing samples.

Water flow: The water flow profile uses linear interpolation similarly to the ambient temperature case but only in specific situations. When a dropout occurs before the start of water usage and the connection is re-established during water flow, there significant uncertainty when the water usage would actually have started. Therefore, the missing

samples before partial water usage are interpolated with zero water flow. This also happens in the situation when a dropout occurs during a water usage and re-establishes when the water usage has finished. The only situation where linear interpolation occurs is when the dropout occurs between samples of positive water flow and there are no more than three consecutive samples missing. This is because it is likely that water flow occurred during these samples.

Power: Since power can only switch between on (approximately rated power) and off (zero) value, linear interpolation only occurs in the situation similar to water flow where up to three samples are missing between two on measurements. Zeroes are interpolated everywhere else for this profile.

2.3.4.9 Seasons

The process described is repeated for each season. Data is saved to an excel spreadsheet and each timestamp has an associated measurement for water flow rate, ambient temperature, and power for each season. Figure 2.8 shows a flow diagram summarising the entire data cleaning process.

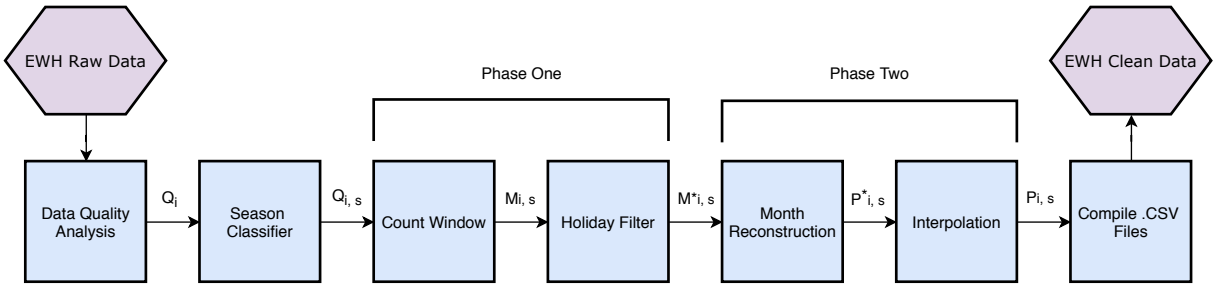


Figure 2.8: Flow diagram of data cleaning process and indicating input data for the next step (subscript s indicates season). Purple blocks represent a database and blue blocks represent operations in the process.

2.3.5 Data analysis

This section first performs an analysis of the raw database to discard invalid EWH profiles. Following this is an analysis of the different criteria leading up to the data quality of each raw profile. Lastly, a decision of the best time period for the data cleaning process is aided by a timeline analysis of all the existing data.

2.3.5.1 Invalid data profiles

An initial analysis of the raw database is completed to discard EWH profiles which will most likely be discarded during the data cleaning process due to lack of data and thus prevent wasting resources. An EWH profile is discarded if it does not satisfy one of the following conditions:

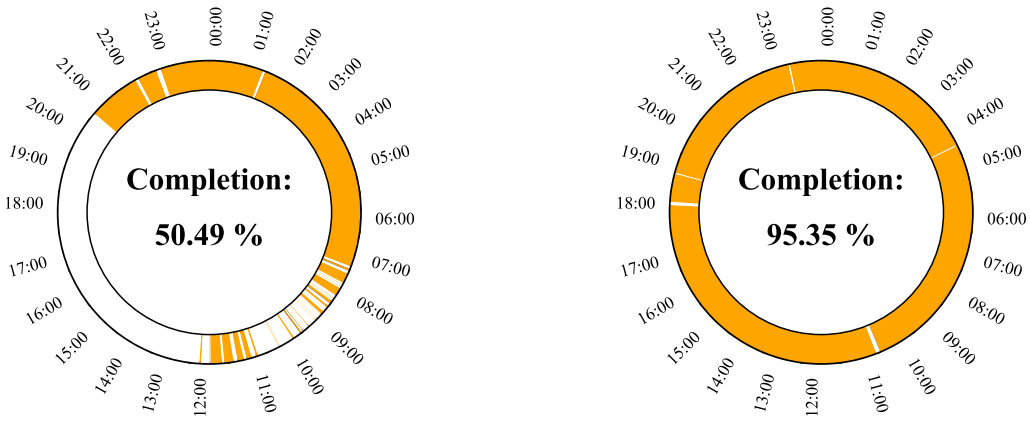
1. **File size too small:** If the profile file size is < 1Mb.
2. **Date shortage:** If the monitored period does not cover at least four seasons.

3. **Overall volume shortage:** If the total water usage volume is $<$ 100L. It is assumed that either the monitored period is too short, or the sensors were broken.

The number of profiles within the raw database reduced from 245 to 145 EWH profiles after the database was filtered due to the above-mentioned conditions.

2.3.5.2 Completeness

Completeness is the largest and most important anomaly that exists in the profiles: it is crucial to identify the portion of missing samples. Not only should the amount of missing samples be found, but it is important to investigate their distribution during the day to confirm whether a significant amount of important information has been lost. Figure 2.9a and 2.9b represent the distribution of existing (orange) and missing (white) data samples. Figure 2.9a shows a day with bad data completeness and Figure 2.9b shows a day with good data completeness. The middle of each circle plot shows the completeness for the day.



(a) Completion for a bad day.

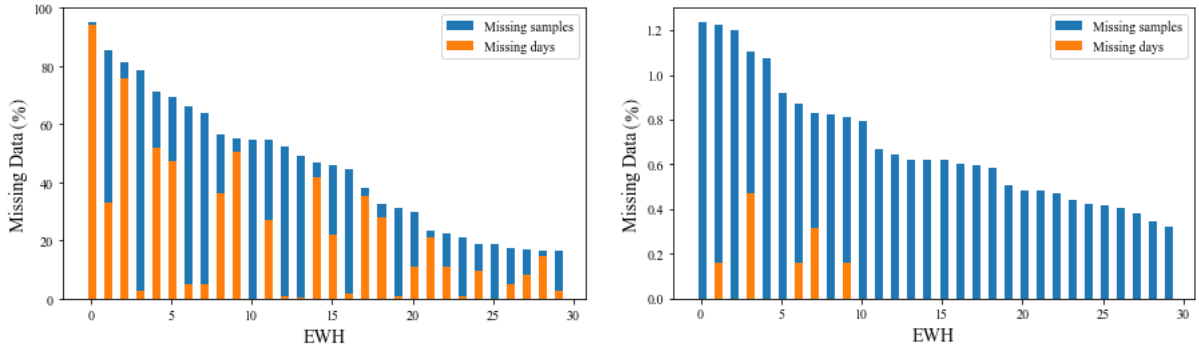
(b) Completion for a good day.

Figure 2.9: Circle plots of bad completeness (left) and good completeness (right) days with existing samples (orange) and missing samples (white).

The influence of missing days on data completeness of each profile is analysed in Figure 2.10. The blue represents the percentage of data missing for each profile and the orange represents the portion of missing data due to missing days. All the profiles are sorted in descending order of missing data. Figure 2.10a shows 30 profiles with the most missing data and Figure 2.10b shows 30 profiles with the least missing data. These figures show that profiles with bad completeness generally are largely affected by missing days whereas profiles with good completeness are rarely affected by missing days. This highlights the importance of discarding profiles not satisfying Q_{thresh} before phase one as they will make finding the optimal month more difficult.

2.3.5.3 Validity

Validity is inspected for the different fields of attributes to estimate if constraints should be imposed on the respective data. Knowledge of the types of invalid data that exist can be used to score a criterion which contributes to the quality of the data.

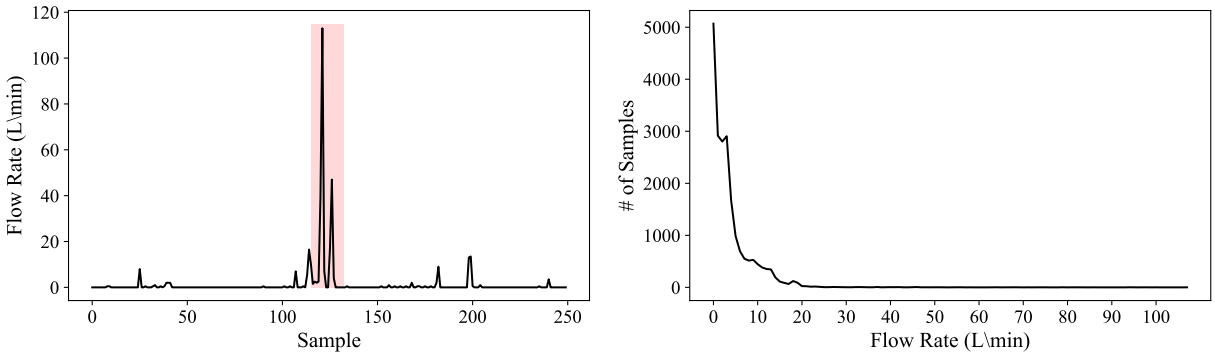


(a) **(b)**

Figure 2.10: Percentages of missing data (blue) and missing days (orange) for different EWH profiles. a) Bar graph of the 30 profiles with the most missing data in descending order of missing data with the y-axis scale ranging from 0 to 100%. b) Bar graph of the 30 profiles with the least missing data in descending order of missing data with the y-axis scale ranging from 0 to 1.3%.

Water flow: Figure 2.11a shows a water profile taken from the raw data. A large spike in water flow highlighted in red shows that invalid data exists. Figure 2.11b shows a tally of different water flow rates for all the raw profiles to identify invalid outliers. The x-axis represents the domain of identified flow rates which can be more than 100 L/min. However, due to the rare existence of invalid flow, a decision to apply action to invalid flow rates should only be considered if the final profiles contain them.

The ambient temperature and power data measurements are not used for any of the components developed in this thesis. Therefore, the analysis of validity for these attributes are shown in Appendix A.1.



(a) Water profile containing an invalid spike.

(b) Sample count for each water flow rate value.

Figure 2.11: Representations of valid and invalid water flow rate data.

2.3.5.4 Quality

The quality of the profiles is analysed to show the portion of profiles that satisfy the quality threshold and that will be used for the window searching phase. Equation 2.29 is used to acquire the quality of the profiles which are shown in Figure 2.12. A red line

plotted on the graph represents the quality threshold Q_{thresh} of 80 % to give an indication of profiles that are likely to satisfy this criterion. The red line indicated that 84 % of the profiles are likely to satisfy this criterion, which is 126 of the remaining 145 profiles.

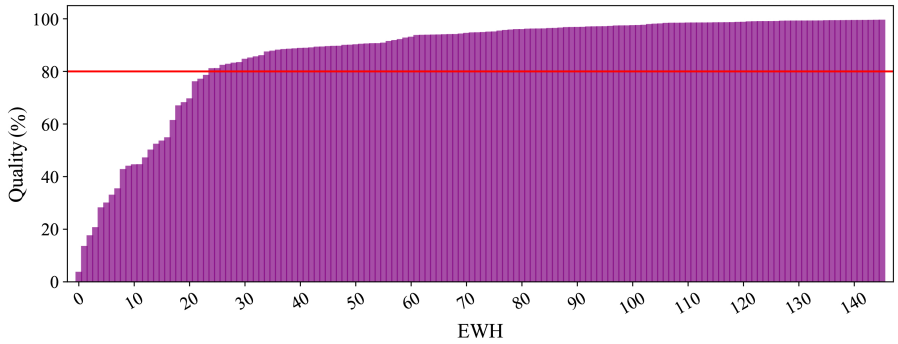


Figure 2.12: Bar graph of quality of the valid profiles sorted in ascending order of quality. The red line represents $Q_{\text{thresh}} = 80\,\%$.

With the per day quality calculated for all profiles, the next step is determining the time period to consider for the window searching phase. Figure 2.13a shows a distribution of dates that correspond to the first and last time-stamp that is present for each profile and the seasonal periods are shown by the shaded areas. From the figure, it is concluded that the time period considered for phase one is the period between the upper and lower quartile of the start and stop date, respectively. Coincidentally this period covers all four of the seasons as well.

Figure 2.13b shows both the average per day quality and the total number of missing days across all the EWH profiles over the time period selected as a result of Figure 2.13a. The shaded areas represent major holiday periods. The figure shows that the average per day quality gradually lowers and the total number of missing days spikes upwards during the second holiday period.

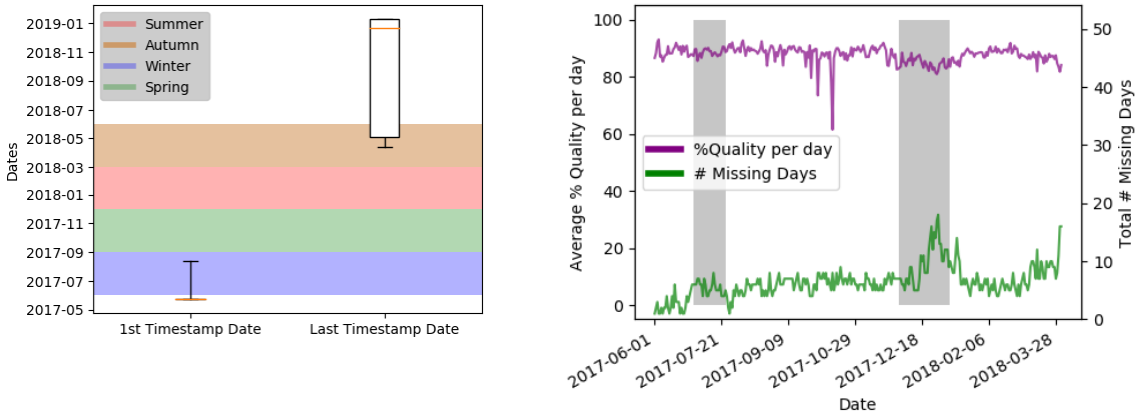
2.3.6 Results

This section describes the results of the data cleaning process and the analysis thereof. The results of interpolation that occurs during the second phase of the data cleaning process were analysed first. Following this, the quality of the EWH profiles after each of the two phases were evaluated to be of a satisfactory level of quality. Finally, the water flow rate versus time profiles were examined qualitatively and quantitatively.

2.3.6.1 Interpolation

The final step of the month reconstruction phase is analysed to confirm that the missing data points were effectively interpolated with appropriate values.

Water flow: Figure 2.14a and 2.14b show the water profile time-sample versus flow rate for an original water profile shown in blue and the sections filled in by interpolation indicated in red. This confirms that the profiles were successfully reconstructed.

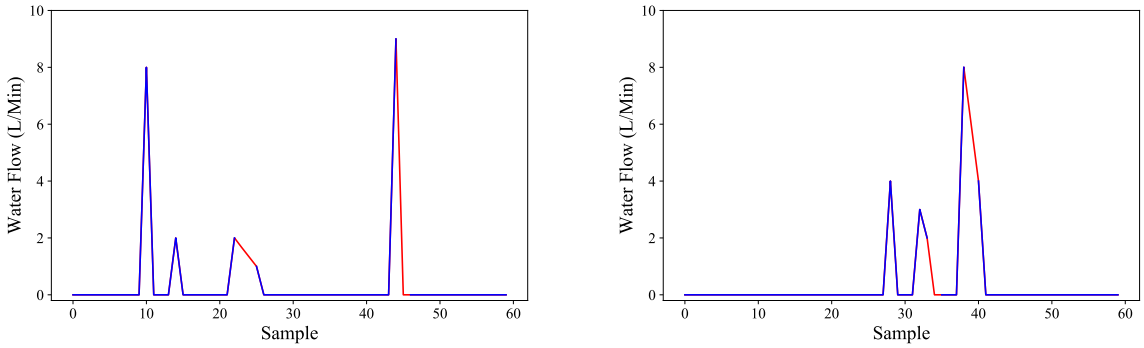


(a) Box and whiskers showing the distribution of start and stop dates of the different profiles.

(b) Average quality per days and total number missing days for all profiles. Shaded areas represent major holiday periods.

Figure 2.13: Distribution of profile start and stop dates and average quality and missing days for all profiles.

For the same reason described in Section 2.3.5.3, the interpolation results for the ambient temperature and power are shown in Appendix A.2.

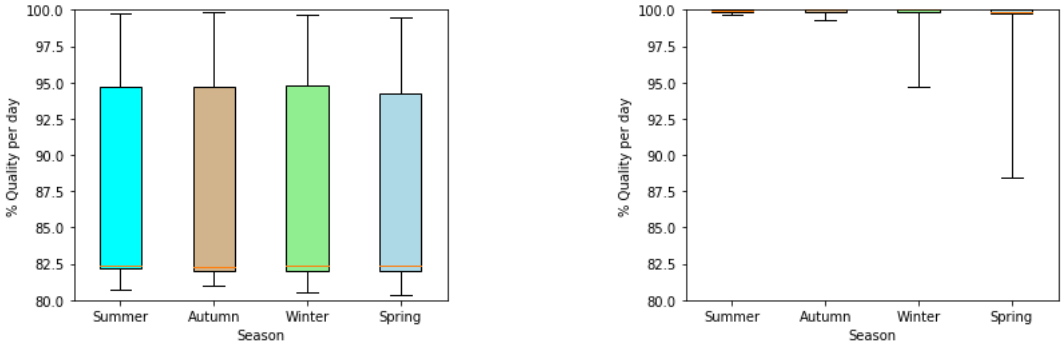


(a) **(b)**

Figure 2.14: Time-sample versus flow rate of original water profile (blue) and portions reconstructed by interpolation (red).

2.3.6.2 Quality analysis

There was a total of 126 profiles before the first phase was initiated. After the first phase was completed, there were 77 profiles remaining and therefore 77 profiles after the second phase was completed. Figure 2.15a shows box-and-whisker plots of the distribution of per day quality of all 77 profiles for each season after phase one was completed. Figure 2.15b shows the same distribution, but after the second phase was completed. Comparing these figures, it shows that the median per day quality of the profiles increases from 82.5% to about 100%. This confirms that the outcome of the data cleaning process is high-quality data and provides a high level of confidence for an accurate representation of EWH water profiles which are used in the remainder of this thesis.



(a) After phase one completed.

(b) After phase two completed.

Figure 2.15: Box and whiskers of the per day quality of the remaining 77 EWH profiles and are separated into seasons.

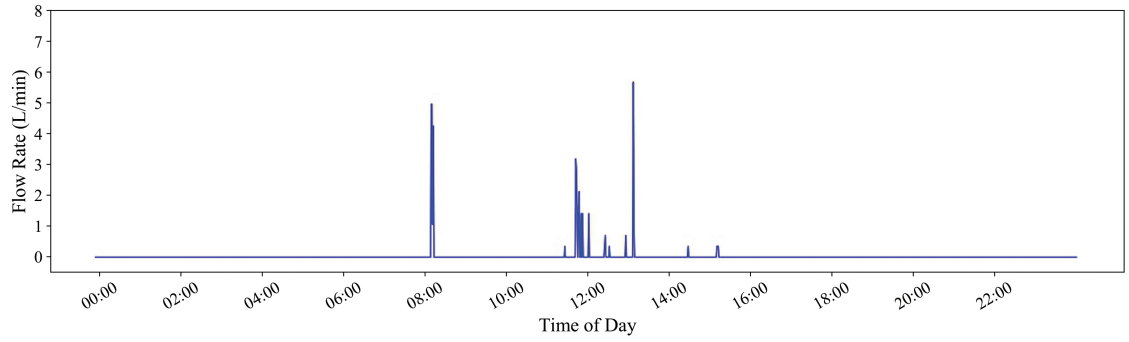
2.3.6.3 High-quality water profiles

The high-quality water profiles are qualitatively shown for a 24-hour cycle for a single EWH in Figure 2.16a and 2.16b. The EWH was picked specifically because it has the water profile that uses the largest volume of water over all four months. The figures show the days that used the least and most volume of water, respectively. Appendix B shows a box and whisker for each one of the 77 EWHs distribution of the total daily volume of water used. The figure is repeated four times and each one represents the distributions for one of the seasons.

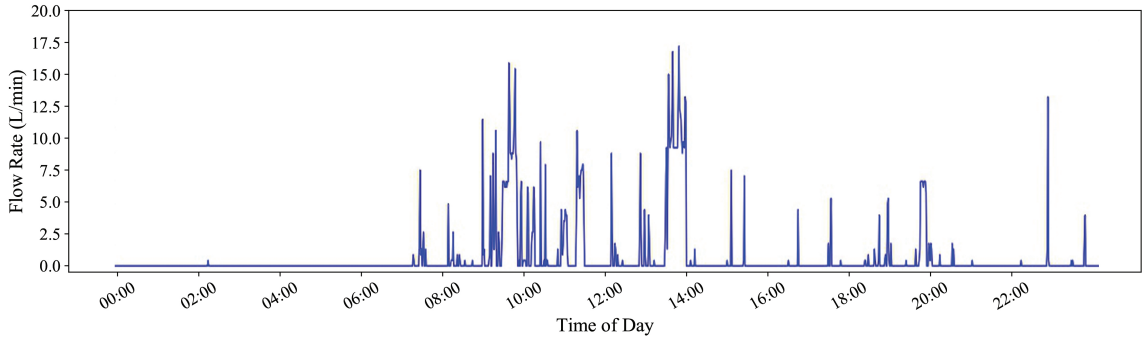
A nominal water profile for each season is shown in Figure 2.17. The water profile is produced by taking the median total volume of water that is used over 30-minute intervals over all days and EWH profiles, and plotting them in their respective 30-minute time intervals. The water profile typically has two peaks, one in the morning (06:00 to 07:00) and one in the afternoon (18:00 to 20:00). There is generally activity during the day, and no activity during the early hours of the morning. The figure shows how seasonality affects the water profile, with winter having significantly larger peaks than the other seasons during the peak hours and autumn having the smallest peaks.

2.4 Summary

The results produced in this chapter show that the final database of water profiles are high-quality and reliable. Each EWH has a water profile with a duration of one month for each season. It is important to consider the water profiles for all the seasons to ensure that the final results of the system are not skewed. All of the clean EWH data is used in the next chapter to develop the probabilistic hot water usage model. It is also used to determine the optimal plan and used during the EWH simulations of the final system for the corresponding water profiles, as described in Chapter 7. However, only one week is used per season for each EWH in this chapter.



(a) Day with the least volume of water drawn.



(b) Day with the most volume of water drawn.

Figure 2.16: 24-hour water profile versus flow-rate for the EWH with largest volume used over the four months.

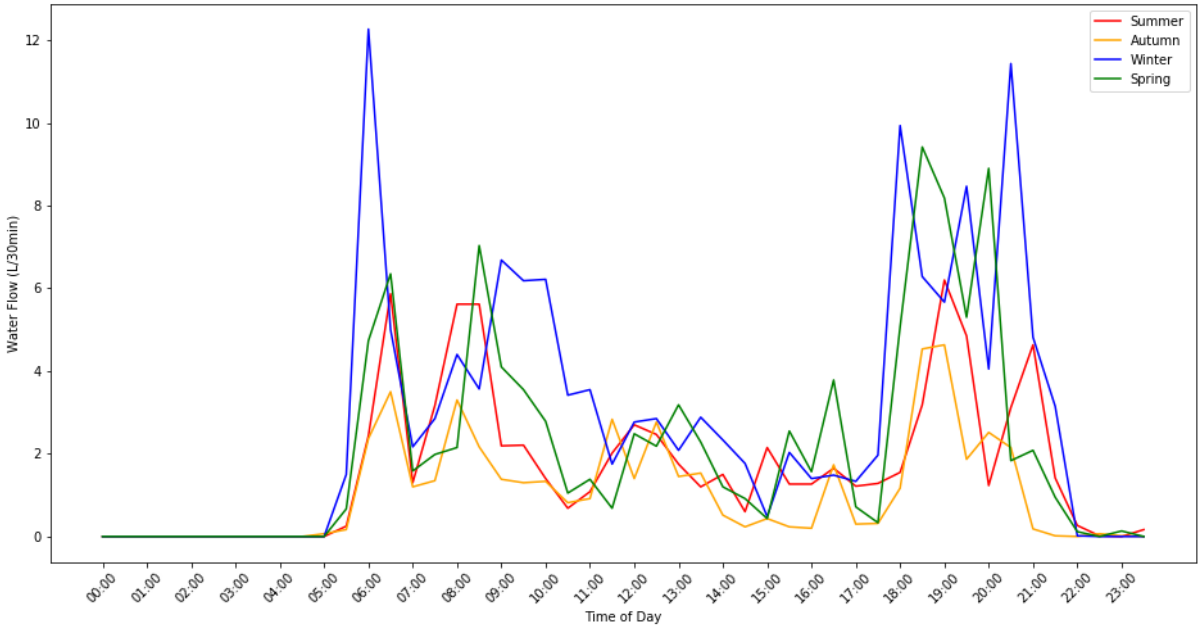


Figure 2.17: Comparison of the nominal water profile for each season of the final high-quality water profiles.

Chapter 3

Probabilistic hot water usage model

Water heating contributes up to 40% of a household's total electricity usage and places a substantial burden on the electricity grid due to high power ratings and users' largely simultaneous hot water usage. The main determinants of its electricity draw are physical properties such as set temperature, insulation, and plumbing configuration; environmental conditions such as ambient temperature and inlet temperature; and the hot water usage profiles. These profiles include the usage volumes, the times of usage and the outlet temperatures. The efficacy of energy management techniques that model water heaters and the accuracy of their simulation results therefore rely on representative hot water usage profiles.

This chapter proposes a probabilistic data-driven model for modelling individualised hot water profiles. The model is accompanied by a hot water usage simulator and hot water demand predictor that includes factors such as seasonality, day of the week and household individuality. Figure 3.1 shows a block diagram that summarises the relationship between the probabilistic hot water usage model, the hot water usage simulator, and the hot water demand predictor. The sections in which the components are discussed are indicated between the parenthesis. A literature study is conducted in Section 3.1 to understand the determining factors that influence a hot water usage profile. This is followed by a summary of existing water usage models from related work. Lastly, techniques that are considered for the proposed model are reviewed. Section 3.2 explains the development of the probabilistic model, the hot water usage simulator, and the hot water demand predictor. Section 3.3 covers a statistical analysis of generated hot water usages to verify the model and simulator. Section 3.4 shows the results of a single EWH's hot water usage profile as well as the results of many EWHs.

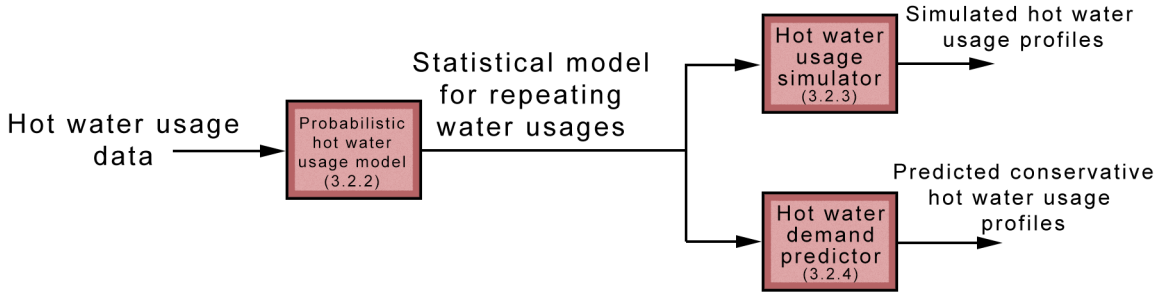


Figure 3.1: Block diagram of system developed in this chapter.

3.1 Literature study

Water heaters with tanks are reasonably good at storing thermal energy. This makes them well-suited to demand-side management (DSM) – an approach that reduces the impact of heavy loads on a constrained grid. However, besides managing demand from water heaters, effective DSM strategies should balance two additional, competing requirements: managing the total energy used and ensuring user satisfaction [34]. To achieve this balance requires large-scale simulation of water heaters using accurate thermal models. The main determinants of such a model are environmental (e.g. inlet water temperature and ambient temperature); the physical properties of the water heater (e.g. thermal losses specification, element rating, volume); installation and configuration parameters (e.g. plumbing setup, retrofitted insulation blankets, orientation); and the hot water usage profile [35, 36, 18, 37].

After a DSM-ready smart water heater is installed, the main determinant of variable use is the user's water profile. This directly affects the energy profile [36]. The accuracy of a water usage profile has a substantial and direct influence on the accuracy of modelling a tanked water heater [39, 38]. For example, Wang *et al.* (2007) modelled and simulated a gas-powered water storage tank and found deviations of between 8 % and 15 % for different water profiles [40]. Domestic hot water and the resulting energy profiles typically have two peaks daily, one in the morning and one in the afternoon [42, 44, 43, 41]. Accordingly, when assessing energy-saving initiatives it is important to characterise heater profiles accurately [45].

3.1.1 Factors affecting water profiles

Individual differentiation and characterisation of water profiles reduces the uncertainties for predictive analyses and removes assumptions based on over-sizing and generalisation [46]. Forecasting household water usage patterns for individual heaters in DSM strategies could improve the reliability of the response to energy surpluses and shortages [47].

Almost all the existing research uses individual measurements of point-of-use devices in a household to determine the end-use cases, which are used to develop the household's total water profile. Such models are generally complex and expensive to install. A cheaper alternative is load modelling [48]. These models provide individual differentiation for all water used, but a household's usage varies over time: weekly patterns vary for days of the week and annual patterns vary for seasons, regions and climate [49, 44]. The day of the week and the season have distinct effects on the water demand, the significant factors being temperature and precipitation [50, 30, 31]. Gato *et al.* (2007) developed a time-series forecasting model of water usage where the correlation increased by 17% when seasonality was considered [30]. Other factors to consider are holidays and once-off special events [31].

The building type and functionality substantially influence the water usage profile. An office space or warehouse would generally consume less water than a residential household [51]. The amount of occupants and size of building also determine the water usage profile. This research only looks at residential households where no household information is provided. Therefore, the effects of building type and functionality will be neglected.

The City of Cape Town in South Africa recently experienced one of the worst droughts since 1904. The term "Day Zero" was given to the dreaded expectations of one's household taps going dry. During the time period - from January 2015 until January 2018 - it was reported that the total daily household water consumption dropped from 540 to a staggering 280L per day [49].

3.1.2 Modelling water profiles

Given the importance of getting accurate usage profiles, various models have been proposed in the literature. The characteristics of the main models (categorised according to Fuentes *et al.* (2018) [52]) are discussed here and summarised in Table 3.1.

Table 3.1: Work and models for simulated water usage profiles. An "S" in the "Day" column indicates that days were differentiated only by workdays and weekends and a tick indicates that the differentiation extended to the day of the week. A dash indicates that the source did not specify the information. The "Unique" column indicates if the model accounted for household individuality.

Ref	Category	Hot water	Season	Day	Unique	Res.
ASHRAE Standard 90.2 [53]	Technical standards	✓	-	-	✗	1 hr
HVAC Handbook [41]	Technical standards	✓	-	-	✗	1 hr
Edwards *et al.* [54]	Data collected	✓	-	-	✓	5 min
Widen *et al.* [48]	Stochastic	✓	-	S	✓	5 min
Hendron *et al.* [55]	Stochastic	✓	✓	S	✗	<1 min
Jordan *et al.* [56]	Stochastic	✓	✓	✓	✗	1 min
Buchberger *et al.* [57]	Stochastic	✓	✓	S	✓	1 min
Blokker *et al.* [58]	Stochastic	✓	-	S	-	1 sec
Scheepers *et al.* [59]	Stochastic	✓	✗	-	-	-
Richard [60]	Stochastic	✓	✓	✗	-	1 min
Yao *et al.* [61]	Stochastic	✗	✓	-	-	1 min
Negnevitsky *et al.* [62]	Stochastic	✓	✗	✗	✗	5 min
Aki *et al.* [63]	Stochastic	✓	✗	✗	✓	10 min
Bakker *et al.* [64]	Data-learning	✗	✗	✓	✓	1 hr
Pflugradt [65]	Behavioural	✗	✓	✓	✓	-
Popescu *et al.* [66]	Time-series forecasting	✓	-	✓	✓	-
Gelavzanskas [47]	Time-series forecasting	✓	-	✓	✓	1 hr
Roux [67]	Time-series forecasting	✓	-	✓	✓	1 min
Jack *et al.* [42]	Data-driven	✗	✗	✗	✓	1 min
This PhD	Data-driven	✓	✓	✓	✓	1 min

3.1.2.1 Technical Standards

Existing standards provide typical daily hot water usage profiles that are available for design purposes. Marini *et al.* (2015) evaluated simulation tools based on various standards to compare simulated results with monitored results for modelling hot water usage in buildings [68]. They found that the simulated total hot water demand was underestimated by up to 30 % and overestimated by up to 40 %. The energy efficiency of the simulated model was also overestimated by between 14 % and 22 % compared to the actual hot water heater.

The ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) standard is the hot water draw model that is possibly the most frequently used.

It has two versions. The first - and the older and more widely used of the two - comes from the ASHRAE Standard 90.2 [53] which does not specify the source of data. The second, from the HVAC Applications Handbook, 2003, has data collected from Canadian residences and provides a profile for a typical household with two adults, two children, a clothes washer and a dishwasher [69]. The ASHRAE standard has been used to perform large simulations of grids comprising water heaters, but it does not differentiate between users. Moreover, it uses a coarse time segmentation of one hour and, like other technical standards, does not consider weekly and seasonal variations, thus water profiles based on technical standards are typically generalised. Gholizadeh *et al.* (2016) proposed a model to reduce the electricity costs of energy usage by adjusting the heaters' set-point temperatures and simulating water profiles based on the ASHRAE Standard 90.2 [70].

3.1.2.2 Data-collected models

Data-collected models stray from using probabilistic models and advance towards a technique that focuses on numerous participants and large amounts of data. The technique uses long-term databases of water measurements and simulations to produce a water profile based on the average. Edwards *et al.* (2015) used data measurements from 73 Canadian households to produce annual water draw profiles [54]. Of the total households, 12 were selected to produce five-minute resolution profiles based on four usage levels and three discrete temporal usage patterns for each level. These usage levels corresponded to the mean, median, 20 % and 80 % percentiles of usage for the sample of households. The temporal usage patterns were made up of predominantly morning usage, predominantly evening usage, and usage constant throughout the day. The research revealed that different water usage patterns can lead to significant changes when it comes to predicting water usage profiles.

3.1.2.3 Stochastic models

Stochastic methods model the probability distribution of household water usage using time-of-use data. Widen *et al.* (2009) used time-of-use data to generate electricity and hot water profiles [48]. Individuals in Sweden used diaries to track the sequence and time of electrical or hot water activities during one weekday and one weekend day. They then tried to match the activities to end-use categories and used information from surveys and product tests to estimate appliance parameters. The results showed a promising correlation between simulated and measured data. Hendron *et al.* (2010) determined hot water events for different activities using clustering and probability distributions [55]. The event characteristics were approximated by fitting normal and exponential functions. Jordan *et al.* (2001) used a similar approach with the exclusion of clustering [56]. Yao *et al.* (2005) produced domestic energy demand profiles from aggregated random appliance profiles [61]. They clustered the energy profiles according to the factors that influence a household's profile, such as the number of occupants, the times of the first and last appliance usage of each day, and the occupancy patterns, determined by the periods of the day when the house was unoccupied [62, 63].

Buchberger *et al.* (1996) based their event probability distributions on the assumption that water demand follows a homogeneous Poisson rectangular pulse process [57]. This means that events can be approximated by rectangular shapes with the rate of occurrences being time dependent. However, the results showed that this process cannot model the

variance in occurrence well enough and the data came from a total of only four households. Scheepers *et al.* (2014) devised a method for generating a residential indoor cold water profile based on probabilistic end-use models [59]. The profiles consisted of high resolution measurements with the temporal sample difference being as small as one second. Residential water demand profiles were simulated using SIMDEUM, which was developed by Blokker *et al.* [58]. As in the model by [57], water events were approximated by rectangular pulses for six end-use cases and aggregated water demands were obtained using a Monte Carlo method. However, they used information on total water usage and could not distinguish between hot and cold water. Richard (2016) improved on this by using volume balancing techniques to approximate the portion of hot water from the total [60]. The resulting daily water demand profiles were similar to popular existing patterns such as the ASHRAE Standard 90.2.

3.1.2.4 Data-learning models

Neural networks are computational models consisting of "neurons" which have the ability to learn, generalise and categorise data. The "network" is made up of simple one-task processing units that are connected by signals which are weighted. A neuron receives signals from its surrounding neighbours to determine an output signal. The neural network can be divided into input, output, and hidden neurons. Input data from outside is fed into the input neurons and output neurons send out the output data while all the neurons between these layers are "hidden" as they are not seen from the outside. Data is trained to calculate the weights that best fits the data. Training data includes the desired outputs so that the weights are adjusted until the errors between the actual output and the expected output is minimised[64].

Bakker *et al.* (2008) obtained the heat demand of a household to calculate the electricity demand and used them in a multi-layer feed-forward neural network [64]. The heat demand for each following day was predicted with knowledge of the expected temperature and the previous day's heat profile. The output data was generated with an hourly resolution, as not enough training data was available to make more accurate predictions. The results were good, despite an hourly mean absolute deviation of approximately 36 kWm. The authors suggested that the results could be improved if more factors were considered in the model.

3.1.2.5 Behavioural models

Behavioural models are based on factors that influence load profiles such as work-days and holidays. Pflugradt (2016) produced a psychological desire model to avoid calculating probability distributions [65]. He used a program called a behaviour-based load profile generator which makes it possible to simulate an individual's behaviour, to obtain energy and water usage data. He concluded that it works well for individual households; however, the quality of the results is strongly determined by the input. It is difficult to gather information about residents and more assumptions must be made for larger areas of households.

3.1.2.6 Time series models

Time series models forecast a system's future behaviour by analysing historical data to make statistical conclusions based on time correlation. The original data is decomposed into trend, seasonality, and residual components. Autocorrelation functions and partial autocorrelation functions are used to compare a time series with a time-displaced version of itself. A residual component showing zero correlation means that the components were successfully separated. A strict requirement for time series modelling is that the data must be stationary, meaning that it must have a time-invariant mean, variance, and autocorrelation. Many tests are available to check if time series data is stationary. A variety of forecasting models can then use this information to predict future data. More information on time series forecasting can be found in [71].

To forecast the hot water demand for residential households, Gelavzanskas *et al.* (2015) tested various time series forecasting techniques and used different forecasting models [47]. As the data showed strong daily and weekly correlation, the best results were obtained using a Loess (locally estimated scatterplot smoothing) seasonal decomposition of a time series with an ARIMA (autoregressive integrated moving average) model and an STL (seasonal and trend decomposition) exponential smoothing model. The study did not, however, account for the time of year, which contributes to both seasonality and holidays. Popescu *et al.* (2008) explored a similar approach using an ARMA (autoregressive moving average) model to simulate the water demand for a block of apartments [66].

The existing time series forecasting models are suitable for forecasting the profile for an individual household, but they fail to capture seasonal effects. Time series is closely related to machine learning, where a great deal of data is required.

3.1.2.7 Data-driven models

These models forecast water usage by analysing the relationships and pattern of measured data without matching water usages to specific household devices. Jack *et al.* (2018) proposed a minimal simulation model (MSM) that simulates smart control of a water heater to assess its impact on service [42]. This model used the monitored electricity demand of a domestic water heater to generate water profiles that are simulated from a hot water tank model. The parameters for generating water profiles were adjusted manually until the simulated and monitored electricity demand profiles matched – a process that had to be repeated for each individual heater.

3.2 Development

The probabilistic model presented here statistically models the daily hot water usage profile for individual EWHs based on historically measured flow rate data. This model can be used to model hot water demand and the resulting residential energy demand. A hot water usage simulator is created that implements the probabilistic model. This simulator generates simulated hot water usage profiles (flow rate versus time) based on the probabilistic model that was fitted to the historically measured usage data. The hot water usage simulator can be used as a component of a larger EWH simulator. A hot water demand predictor is also created that uses the probabilistic model to generate hot

water usage predictions. The parameters are selected conservatively to produce water profiles suitable for optimal planning of unanticipated water usages in Chapter 7. The probabilistic model is described in Section 3.2.1, followed by the usage simulator in Section 3.2.2 and lastly the demand predictor in Section 3.2.3.

3.2.1 Probabilistic hot water usage model

The hot water usage model statistically models the hot water usage behaviour for a given EWH by identifying repeating daily hot water usage events, and determining the statistical distributions of each usage event's time of day, volume of water used, and average flow rate. To do this, the model groups the measured flow rate data into clusters of time, volume, and flow rates, to which it fits distributions and determines their probabilities of occurrence, as shown in Figure 3.2.

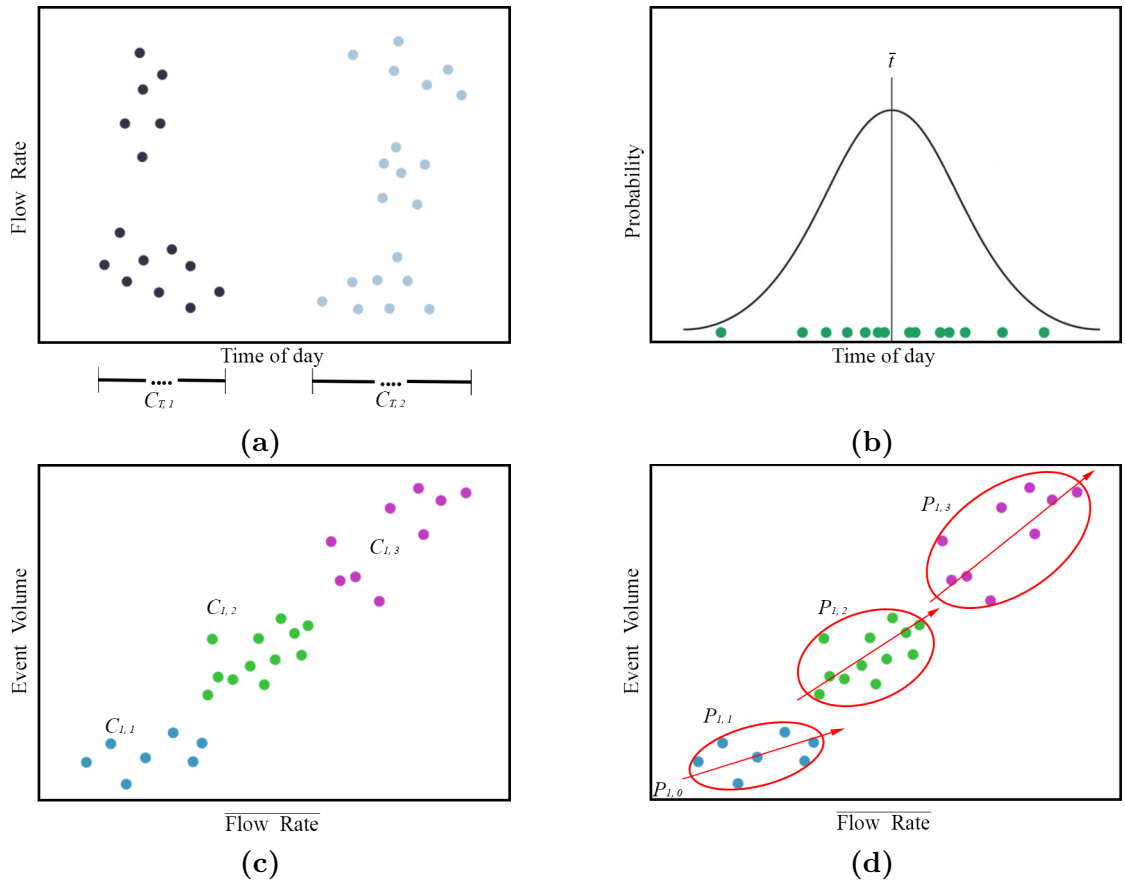


Figure 3.2: (a) The flow rate data points are clustered according to the time of day to produce time clusters $C_{T,1}$ and $C_{T,2}$. (b) Start times measured for a time cluster (green) are fitted with a Gaussian distribution where t' represents the mean middle time of the time cluster. (c) Water event parameters recorded for each time cluster are clustered according to volume. (d) Volume clusters are fitted with 2D Gaussian distributions (red ovals) with correlation of volume and flow rate shown with a red arrow. Each cluster is associated with its own probability of occurrence, $P_{\mathrm{n,m}}$.

The model handles hot water usage variations due to seasonality by separating the measured data for a given EWH into four seasonal sets and determining different model

parameters for each season (summer, autumn, winter, spring). The model handles variations in the usage pattern due to the day of the week by further separating the seasonal data set into three subsets: the undifferentiated data set, the semi-differentiated data set, and the differentiated data set. The undifferentiated data set assumes that the usage behaviour does not differ based on the day of the week, and that there is only one usage pattern which is the same for all seven days of the week. The semi-differentiated data set assumes that the usage behaviour differs between weekdays and weekends, and that there are two different usage patterns: one for weekdays and one for weekends. The differentiated data set assumes that the usage behaviour differs based on the day of the week, and that there are seven different usage patterns, one for each day of the week (Mondays, Tuesdays ... Sundays). For each season, we therefore derive three variations of our probabilistic hot water usage model, by applying the model to each of the three different data subsets: undifferentiated, semi-differentiated, and differentiated. We now present the procedure for deriving the model from the measured flow rate data. The same procedure is followed to derive the model parameters for the four seasons, and for the three model variations. The only difference is the data set that is used to derive the model parameters.

To identify repeating usage events, the daily measured flow rate data for all the days in a given data set are superimposed on a single day cycle. The flow rate data are grouped into time clusters according to time of the day using k-means clustering [72], as shown in Figure 3.2a. We determine the optimal number of time clusters using the elbow method [73]. Time clusters are represented by $C_{T,p}$, where the subscript p is the index of the time cluster. Each time cluster now represents a repeating usage event, and the flow rate data points in the same time cluster on the same day are considered to represent a single usage event. We determine the probabilistic distribution of the time-of-day of each repeating event by fitting a Gaussian probability density function to the time-of-day component of all the flow rate data points that belong to each time cluster, as shown in Figure 3.2b. The Gaussian probability density function is determined for time cluster p by calculating the time-of-day mean $\bar{t}_p$ and variance σ^2_{tp} for the flow rate data points within the time cluster, and is represented as follows:

$$\mathcal{N}(\bar{t}_p, \sigma^2_{tp}) \tag{3.1}$$

Next, we analyse the distribution of the volume and the average flow rate of the daily usage events within each time cluster. For a given time cluster, the volume and the average flow rate of the usage event on each day are calculated and plotted on a scatter plot, as shown in Figure 3.2b. To identify different types of usage events that may occur within the same time slot, the (volume, flow rate) data points are subclustered according to volume, as shown in Figure 3.2c. The volume subclustering within each time cluster is again performed using k-means clustering and the elbow method. Volume subclusters are represented by $C_{p,q}$, where subscript q is the index of the volume subcluster within time cluster p. Each event volume is captured with its associated average flow rate to retain correlation between the two variables.

The volume subclusters within a time cluster represent different hot water usage events that may occur in the same time slot. However, on a given day, only one hot water usage event can occur in each time slot. The volume subclusters within the same time cluster therefore represent different possible realisations of a single hot water usage event within the given time slot. Every day, each possible realisation of the hot water usage event has

a certain probability of occurring. For each time cluster, we therefore need to calculate the probability of each volume subcluster being realised. The probability $P_{p,q}$ of volume subcluster q being realised in time cluster p, is calculated by dividing the number of days that the given subcluster occurs by the total number of days in the data set. (Note that this approach also accommodates the probability that no usage event occurs in the time slot on certain days, since there will also be a volume cluster for zero volume events).

Finally, we characterise the probabilistic distribution of the volume and average flow rate for each volume subcluster. Since the volume and the average flow rate of a usage event may be statistically dependent, we fit a Gaussian joint probability density function to the (volume, flow rate) data points in each volume subcluster, as shown in Figure 3.2d. Each volume subcluster is therefore characterised by a joint (volume, average flow rate) mean represented by $(\bar{V}_{p,q}, \bar{Q}_{p,q})$ and a joint (volume, average flow rate) variance represented by the volume variance $\sigma^2_{vp,q}$, the average flow rate variance $\sigma^2_{Qp,q}$, and the correlation coefficient $\rho_{p,q}$, and the Gaussian joint probability density function is represented as follows:

$$\mathcal{N}(\bar{V}_{p,q}, \bar{Q}_{p,q}, \sigma^2_{vp,q}, \sigma^2_{Qp,q}, \rho_{p,q}) \tag{3.2}$$

To summarise, the probabilistic hot water usage model models the number of expected hot water usage events each day (time clusters); the probabilistic distribution of the time of day for each usage event (Gaussian probability density function); a number of different types of usage events that may occur in the same time slot (volume subclusters); the probability of each type of usage event occurring; and the probabilistic distribution of the volume and average flow rate of each type of usage event (using the Gaussian joint probability distribution function for each volume subcluster).

3.2.1.1 Clustering overview

This section covers the clustering method considered as well as methods for determining the number of clusters. This leads to the clustering design choices that are used as part of the development of the probabilistic hot water usage model.

Data analysis is used in many applications. An important attribute for many procedures is the grouping or classification of patterns based on similarity. More specifically, patterns can be observations, items or features and are unsupervised (unlabelled) [72]. Clustering analysis produces labels for each data point and shows to which group it is most likely to belong. Clustering methods have two different approaches for assigning patterns to clusters: **hard** clustering assigns an observation to a single cluster whereas **fuzzy** clustering assigns different degrees of ownership of multiple clusters to an observation [72]. This section explores a widely used **hard** clustering method which is applicable for developing the model in this chapter: partitional clustering .

Partitional clustering

These methods have the advantage of being computationally cheap. The algorithm usually performs multiple iterations with different starting positions and the iteration with the best clusters it picked [72]. The squared error algorithm is a commonly used criterion function which calculates a score e^2 for the current status of K clusters and has the

following form:

$$e^2 = \sum_{j=1}^{K} \sum_{i=1}^{N_j} ||x_i^j - c_j||^2 \tag{3.3}$$

where N_j refers to the number of data points x_i^j assigned to cluster j and c_j is the corresponding cluster centre. K-means is the most used method which makes use of the squared error criterion. Each cluster centre starts in a random position and data points are assigned to the clusters. This is followed by reassigning the centres and the process repeats until a some stopping criteria is met [72]. The k-means algorithm is described by the following steps:

1. Place k cluster centres which are assigned to random positions or data points within the represented space.
2. Assign each data point to the closest cluster centre.
3. Recalculate the cluster centre based on the current data points assigned to that cluster.
4. Repeat steps 2 and 3 until cluster centres do not change [73].

An example of the above steps is shown in Figure 3.3. The steps are repeated until one event of two possible scenarios occurs. The first is if the clustering converges to a point where the new score of e^2 does not change significantly. The second is when the first criteria has not been met and the stated maximum iterations has occurred [72].

3.2.1.2 Methods for determining desired number of clusters

A problem that arises with cluster analysis is determining the amount of clusters to be obtained in the final clusters. Most clustering methods require the number of clusters to be defined beforehand in order to proceed. The easiest method is simply to visually inspect it; however, this is not efficient. Therefore, methods of determining an appropriate amount of clusters are investigated in the following sections.

Elbow method

The elbow method is the oldest method and is so named because it looks like the "elbow" in the graph. The basics of the method is to start with cluster amount $K = 2$, and increments this value to a desired maximum value. At each iteration of K, a cost function is calculated for the given data collection based on the predefined number of clusters. These costs are plotted to form a line graph. The appropriate value for K is the point on the graph where the decreasing gradient transitions from steep to plateau. All points after this have small cost changes, indicating that further clustering is not needed. The elbow method is demonstrated in Figure 3.4, where the "elbow" on the graph at $K = 4$ is the desired number of clusters [73].

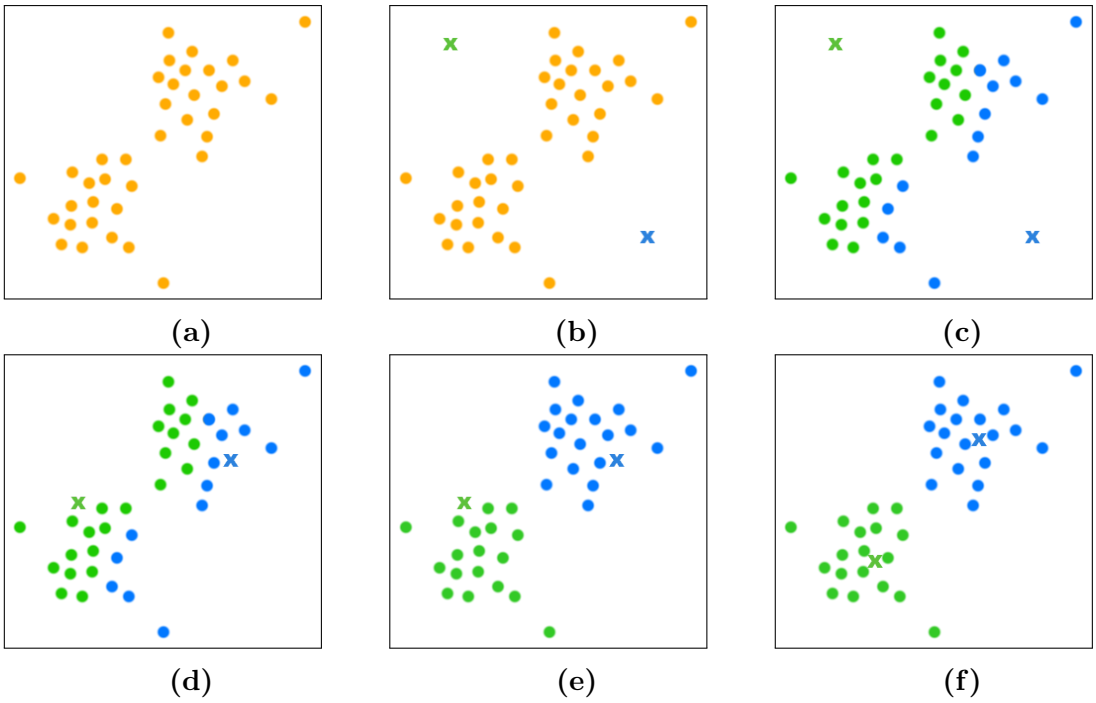


Figure 3.3: (a) Representation of data points. (b) Two cluster centres are randomly assigned and shown by a cross symbol. (c) Data points are assigned labels to their closest cluster centre. (d) Cluster centres are recalculated based on new clusters. (e) Data points are re-assigned labels to their closest cluster centre. (f) Cluster centres are recalculated based on new clusters.

Silhouette method

The silhouette method focuses on comparing the "within cluster" distances with the "between cluster" distance. A silhouette score for data point i in cluster j is calculated with the following formula:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{3.4}$$

where *a(i)* is the average distance between data point i and all other elements within cluster j and *b(i)* is the minimum of the average distances between data point i and all other elements within all other clusters. The silhouette score is calculated for a range of different values for cluster amount K, and is calculated as the average silhouette score for all data points. The score can range between -1 and 1, where -1 means the data point is classified wrong and 1 means perfect clustering [73].

Cross-validation

Cross-validation is a method that estimates the number of clusters based on clustering stability. The data is separated into multiple parts, or *folds*, where one part is the testing set and the remaining parts are the training sets. The training set is passed to a clustering method to produce partitions. New data points are appointed to the clusters by calculating the minimum distance to all the cluster centres. This method provides a means to calculate a measure of difference between the probability of occurrence for the

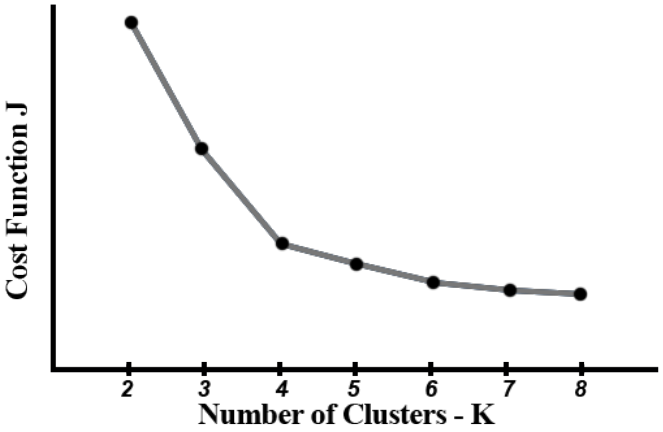


Figure 3.4: Graph shows the results of the elbow method, where the elbow in the graph is at $K = 4$.

clusters between the training and testing sets [74].

A problem with this approach is that clustering is only stable if the cost function has a global minimum. Clustering stability will also not necessarily determine the optimal number of clusters [73].

3.2.1.3 Design choices

The goal of clustering is to distinguish water usage patterns within a household in terms of time of usage and intensity of the usage. The clustering technique chosen was k-means because of its low computation complexity and simple implementation in Python. The implementation uses the *Scikit Learn K-means* package to perform the necessary clustering.

Before clustering is performed, the best number of clusters K is determined. The elbow method was chosen as it produced the best values for K for the type of data used and allows for the desired number of clusters to be determined autonomously by identifying the "elbow". Once all the scores are acquired, the "elbow" of the graph is determined by defining an elbow convergence threshold, E_{thresh}, and detecting the first occurrence of where the gradient between two points is less than this threshold.

3.2.2 Hot water usage simulator

Algorithm 1 shows the procedure used to generate simulated usage profiles from a water heater's probabilistic usage model.

Assume that we wish to generate the hot water usage profile for a number of days. To do this, we iterate through all the days, and for each day, we iterate through all the time clusters and generate a random hot water usage event for each time cluster. To generate the hot water usage event for a specific time cluster, we first determine the time of the event by drawing a random sample from its Gaussian time of day distribution, as shown in Figure 3.2b. Next, we determine which volume subcluster is realised within the time cluster by randomly sampling the subcluster index q based on the probability distribution

Algorithm 1 Hot water usage simulator algorithm

procedure GENERATE SIMULATED USAGE DATA
  for `each day` **do**
    for `each time cluster p` **do**
      $x \leftarrow$ *Randomly sample number between 0 and 1* // where x is a random variable
      $q, i \leftarrow 0$
      while $i < Q_p$ *and* $q = 0$ **do** // where Q_p is the optimal number of volume subclusters
        $i \leftarrow i + 1$
        if $x < \sum_{n=1}^{i} P_{p,i}$ **then** // where P is the volume subcluster occurrence probability
          $q \leftarrow i$ // assign picked volume subcluster to q
        end if
      end while
      if $q > 0$ **then** // event occurred
        $V_{p,q}, Q_{p,q} \leftarrow \mathcal{N}(\bar{V}_{p,q}, \bar{Q}_{p,q}, \sigma^2_{vp,q}, \sigma^2_{Qp,q}, \rho_{p,q})$ // generate one sample
        $T_{\mathrm{dur}p,\,q} \leftarrow \frac{V_{p,q}}{Q_{p,q}}$
        $t_{mp,q} \leftarrow \mathcal{N}(\bar{t}_p, \sigma^2_{tp})$ // generate one sample
        $t_{\mathrm{start}p,\,q} \leftarrow t_{mp,q} - \frac{1}{2} \times T_{\mathrm{dur}p,\,q}$
      end if
    end for
    for `each water usage event` **do** // construct events
      flow rate$(t) \leftarrow Q \quad t_{\mathrm{start}} \le t \le t_{\mathrm{start}} + T_{\mathrm{dur}}$
      $\leftarrow 0 \quad t < t_{\mathrm{start}}$ *and* $t > t_{\mathrm{start}} + T_{\mathrm{dur}}$
    end for
    Superimpose all water usage events to obtain simulated hot water usage profile for day
  end for
end procedure

of all the volume subclusters within time cluster p. Once the volume subcluster has been determined, we determine the volume and the average flow rate of the hot water usage event by drawing a random sample from the Gaussian joint (volume, flow rate) distribution for the given volume subcluster, as shown in Figure 3.2d. Given the randomly sampled time, volume, and average flow rate of the event, the simulated flow rate versus time is constructed. First, the duration of the usage event is calculated by dividing the volume by the average flow rate, using

$$T_{\mathrm{dur}} = \frac{V}{\bar{Q}} \tag{3.5}$$

where T_{dur} is the duration of the event, and V and $\bar{Q}$ are the randomly sampled volume and average flow rate. The start time of the event is then calculated using

$$t_{\mathrm{start}} = t_{\mathrm{m}} - \frac{T_{\mathrm{dur}}}{2} \tag{3.6}$$

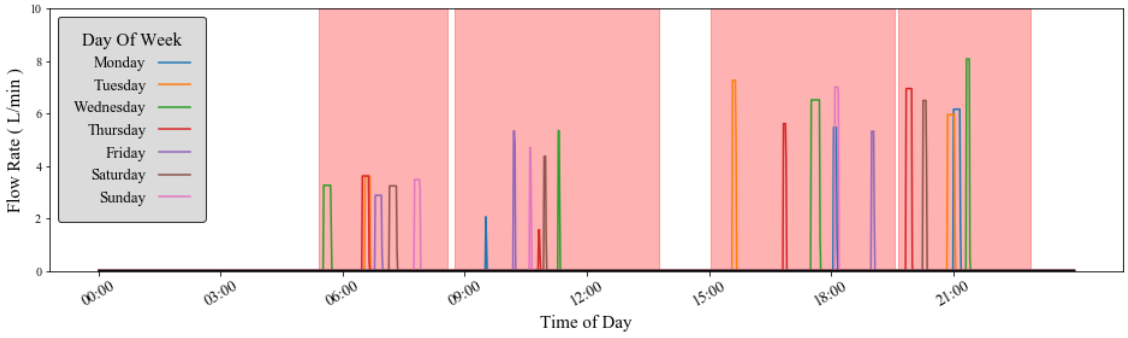
where t_{start} is the start time of the event, t_{m} is the randomly sampled time of the event, and T_{dur} is the duration of the event. (The time t_{m} is the time in the middle of the event, and therefore half of the duration T_{dur} must subtracted to obtain the start time of the event t_{start}. The simulated usage event is then constructed using the following equation

$$Q(t) = \begin{cases} \bar{Q}, & t_{\mathrm{start}} \le t \le t_{\mathrm{start}} + T_{\mathrm{dur}} \\ 0, & \text{otherwise} \end{cases} \tag{3.7}$$

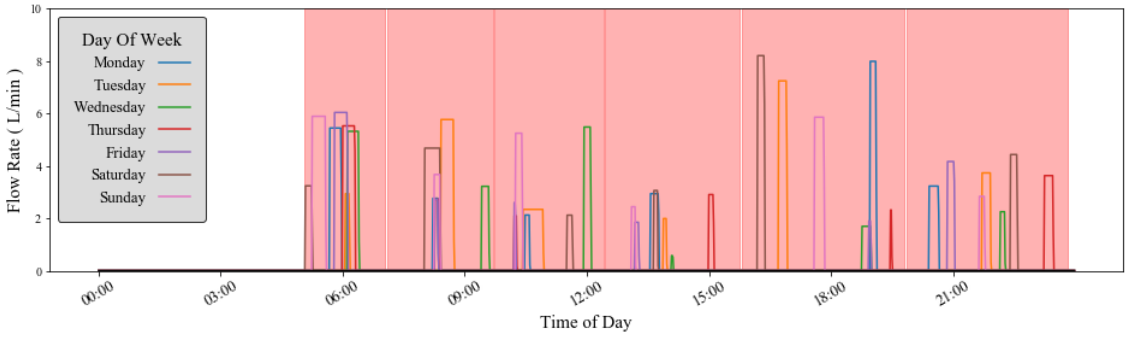
where $Q(t)$ is the instantaneous hot water flow rate as a function of time. A hot water usage event is therefore simulated as a constant hot water flow rate between the start

time and the end time of the event and has a zero-flow rate otherwise. The constant flow rate equals the randomly sampled average flow rate of the event, and the end time equals the randomly sampled start time plus the duration of the event.

The simulated hot water usage profile for the entire day is constructed by superimposing the simulated hot water usage events of all the time clusters. This hot water usage model can generate simulated profiles for a specific day (e.g. Monday, Tuesday, etc.) over a desired period of time. Using the differentiated model, we generated a week's profile by producing a one-day water profile from each of seven hot water usage models, each representing a day of the week. Figure 3.5 shows the simulated profiles generated by the undifferentiated model for seven days against the backdrop of the identified time clusters in red. It shows the profile for one week on a 24-hour cycle, for a heater with light usage and one with heavy usage.



(a)



(b)

Figure 3.5: Simulated hot water profiles generated in the identified time clusters (shaded red region) for (a) a light usage EWH and (b) a heavy usage EWH. The undifferentiated model was used.

3.2.3 Hot water demand predictor

The hot water usage predictor generates conservative usage profiles from a water heater's probabilistic usage model. Similar to the hot water usage simulator, the conservative profile is generated for a number of days for a specific season by iterating through all the days of the week, and iterating through all the time clusters and generating a hot water usage for each time cluster. To generate a conservative usage event for each time cluster, the time of the event is determined by picking a sample t' from its Gaussian time-of-day

distribution that represents the 5^{th} percentile of the distribution, as indicated with the red line in Figure 3.6a. Next, the volume and average flow rate of the hot water usage event is determined by picking a sample V' and $\bar{Q}'$ from its Gaussian joint distribution that represents the 95^{th} and median of the distribution respectively, as indicated with the red lines in Figure 3.6b.

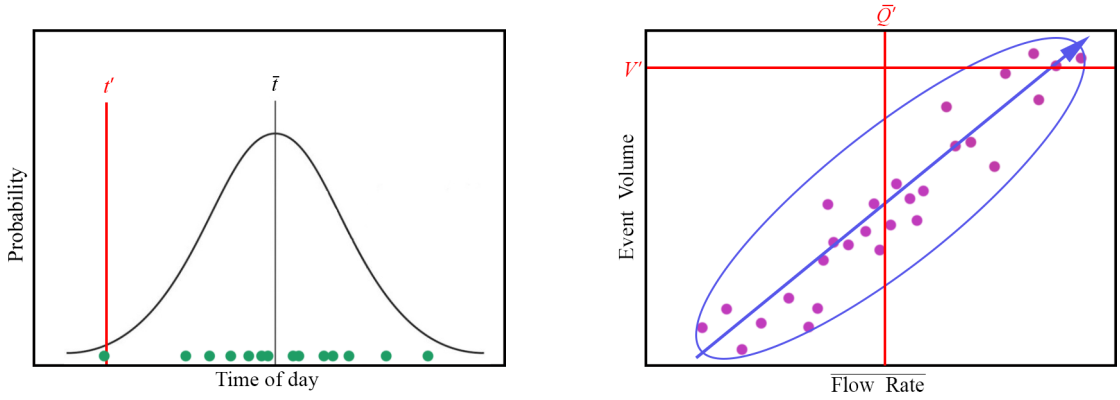


Figure 3.6: (a) Start times measured for a time cluster (green) are fitted with a Gaussian distribution where $\bar{t}$ represents the mean middle time of the time cluster. The conservative start time t' selected by the predictor is indicated in red. (b) The data points of the whole time cluster is fitted with a 2D Gaussian distributions (blue oval) with correlation of volume and flow rate shown with a blue arrow. The conservative volume V' and average flow rate $\bar{Q}'$ selected by the predictor is indicated in red.

3.3 Model and simulator verification

The simulator was verified by comparing the statistical distribution of the hot water usage events generated by the simulator with the statistical distribution of the measured usage events identified from the measured data. The verification consisted of four evaluations:

1. The statistical distributions of the simulated hot water usage events were compared with the statistical distributions of the real hot water usage events that were identified in the measured flow rate data. The comparison was done for two example EWHs - one with a light usage profile, and one with a heavy usage profile - and with their probabilistic models derived using the undifferentiated data set.
2. A large number of simulated usage profiles were generated for an EWH with an average usage profile. Histograms of the time of day of usage events were plotted and compared against the Gaussian probability density functions for the time of day identified from the measured flow rate data.
3. The frequency of the simulated usage events that belong to each volume subcluster within the same time cluster was calculated and compared with the probability of the volume subcluster occurring, as identified from the measured flow rate data.
4. Scatter plots of the volume and average flow rates of the daily usage event were plotted and compared with the confidence ellipses of the Gaussian joint probability functions for the volume and flow rate identified from the measured flow rate data.

The generation of volume clusters for a profile consisting of five time clusters for the undifferentiated model is shown in Figure 3.7a. The water events were obtained for 200 weeks' worth of Monte Carlo realisations of a simulated hot water profile and were plotted against the volumes of each water event. The time clusters are colour-coded and the volume clusters are differentiated by symbols to make it clear to which clusters the water events belong. This profile specifically shows that the user has larger water events occurring in the morning and evening hours and smaller water events occurring in-between. Table 3.2 shows the expected probabilities for each volume cluster and the percentage of volume clusters that actually occurred in the following columns. The smallness of the differences between the two columns shows that the model generates water events very close to the expected probabilities.

Table 3.2: Table of event occurrences of each volume cluster within each time cluster for the 200 weeks of measured and simulated water events taken from Figure 3.7a.

	Volume cluster 1		Volume cluster 2	
Time cluster	Measured	Simulated	Measured	Simulated
1	50 %	49 %	18 %	15 %
2	25 %	24 %	36 %	34 %
3	25 %	22 %	11 %	7 %
4	61 %	57 %	18 %	16 %
5	57 %	54 %	11 %	12 %

The rest of Figure 3.7 shows the modelled distributions (based on measured data) and the simulated statistical distributions for an EWH that has three different time clusters. Approximately 2000 water events were generated for each time cluster to produce these distributions of the data that accurately reflect the cluster's statistical properties. Figures 3.7b, 3.7c and 3.7d show the modelled normal distribution (orange) of the time of day based on the measured data and a histogram of the simulated even times (green) generated by the usage simulator. The legend of each figure shows the mean and standard deviations of the measured data and the simulated events. The statistical properties of the simulated start times agree with those of the statistical properties of the measured data.

Figures 3.7e, 3.7f and 3.7g show scatter plots of the volume and flow rate parameters for simulated water events for the same three time clusters as the start times. From each time cluster three volume clusters were identified and colour-coded accordingly. A confidence ellipse is also shown for the measured data that was used to identify each volume cluster. Each ellipse has a confidence interval of 95% to show that 95% of the simulated profile's water events fall within these regions. These figures show that the simulated profiles retain the mean and standard deviation for both volume and flow rate as well as the correlation between them.

3.4 Results - comparison with existing models

We firstly compared the ability of the three variations of our model to model hot water demand against that of two popular models from the literature. We also used a validated heater model to compare the heaters' simulated energy profiles with their modelled profiles. Finally, we compared the results with the "notch test", an experiment that was

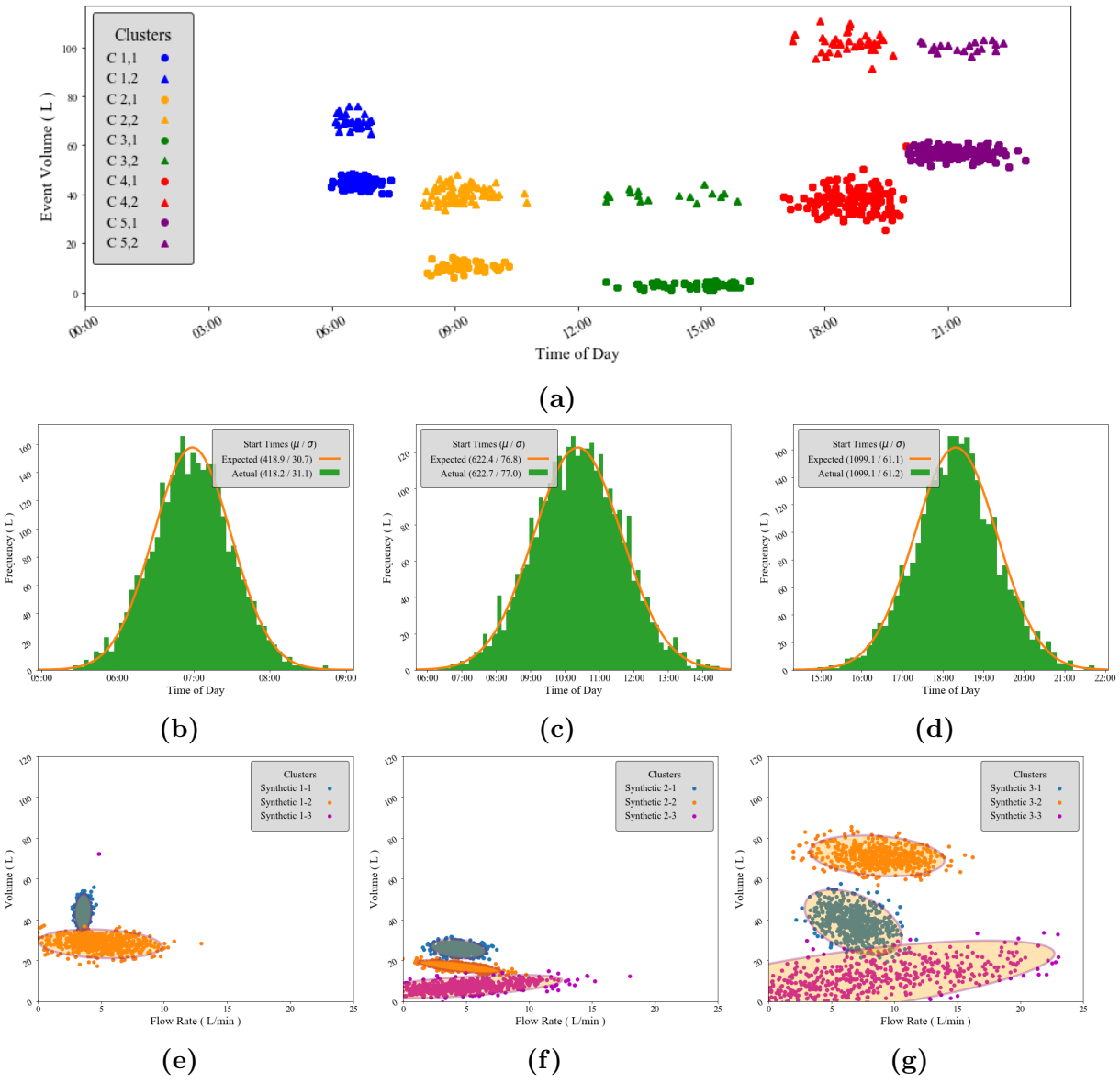


Figure 3.7: a) Scatter plot of water events generated for many realisations of a single EWH's simulated hot water profile for the undifferentiated model. The occurrence frequencies are shown in Table 3.2. b-d) Start times of measured data distributions (orange) and generated simulated water events (green) for three time clusters. Statistical properties of each are shown in the legend. e-g) Volume and flow rate scatter plots of the simulated water events generated for each time cluster. Also shown are 95% confidence ellipses produced by the measured data for each volume cluster.

done on the South African national electricity grid using an estimated 120 000 heaters, to determine the aggregate electrical profiles of residential households' water heaters [43].

For comparison, hot water usage profiles were generated from the minimal simulation model (MSM) devised by Jack *et al.* (2018) [42] and the ASHRAE 90.2 implementation by Gholizadeh *et al.* (2016) [70].

3.4.1 MSM model

Jack *et al.* (2018) proposed a model that generated the electricity demand of a domestic water heater using monitored electricity data of the heating element [42]. A water profile was generated using the model and simulated on a water heater model with thermostat control. The parameters that determined the water profile were adjusted until the simulated electricity demand matched the monitored data. For the comparison, the model was modified to use hot water data as an input and obtain a simulated hot water profile that matches the monitored data. The method is summarised in the following steps:

1. The monitored data is analysed to obtain information of the events within it. The start times and duration of all the events are accumulated and tallied into 1 hour and 15-minute bins, respectively. This is used as a basis to measure the similarity of the synthetic results.
2. Up to three major events can be generated, each with their own probability $\mathrm{P_k}^{\mathrm{major}} > 0$. The start times are random and normally distributed around a fix time of the day with standard deviation ${\sigma_\mathrm{k}}^{\mathrm{major}}$. The duration of each event ${\tau_\mathrm{k}}^{\mathrm{major}}$ is random and is determined by a log normal distribution as follows:

$$P({\tau_\mathrm{k}}^{\mathrm{major}}) = \exp\{[\ln(\tau^{\mathrm{major}}) - {\mu_\mathrm{k}}^{\mathrm{major}}]^2/2({\gamma_\mathrm{k}}^{\mathrm{major}})^2\} \tag{3.8}$$

   With parameters ${\mu_\mathrm{k}}^{\mathrm{major}}$ and ${\gamma_\mathrm{k}}^{\mathrm{major}}$. It is assumed that the amount of major events and fixed major event times were determined by inspection for each profile. However, this is extremely inefficient when simulating a lot more profiles. This step was therefore made generic by using a peak finding algorithm to determine the greatest peaks and obtain the fixed times of these peaks.
3. Minor events are also generated for each day and are assumed to have an equal probability $\mathrm{P_{minor}} > 0$ of occurring at any time during a single interval over a 24-hour period. The duration of the minor events is also determined using the log normal distribution by equation 3.8 with parameters μ_{minor} and γ_{minor}. However, using a single interval makes this step constrained and inflexible as minor event occur multiple times in a day. The probability was therefore adjusted to be more realistic and generic by multiplying the fixed probability $\mathrm{P_{minor}}$ by the monitored data start time tallies, depending on the hourly time slot.
4. All the parameters defined in the previous steps are varied until a good fit is obtained with the monitored data. A normalised coefficient of determination is used to measure the fit between simulated and monitored data and is defined as follows:

$${R^2}_{\mathrm{diff}} = 1 - \frac{\sum_i ({\mathscr{F}_\mathrm{i}}^{\mathrm{m}} - {\mathscr{F}_\mathrm{i}}^{\mathrm{s}})^2}{\sum_i ({\mathscr{F}_\mathrm{i}}^{\mathrm{m}} - {\bar{\mathscr{F}}_\mathrm{i}}^{\mathrm{m}})^2} \tag{3.9}$$

Where $\mathscr{F}$ is the quantity being compared for both data sets, which is either start times or duration, and subscripts m and s refer to monitored and simulated data. The score of R^2_{diff} can vary from 1 (perfect fit) to 0.

The original MSM model was inconvenient for a large number of profiles. Because the parameters would have to be manually adjusted to fit each unique profile, it would take many hours to do. We therefore adapted this model to fit the data by autonomously iterating the parameters to allow the model to work with larger data-sets. Setting up the model requires manual intervention to assess the quality of the match between the measured and modelled profiles. We automated this assessment using a normalised coefficient of determination, R^2_{diff}. A threshold is defined for the R^2_{diff} of the start times and duration quantities, R_{start} and $R_{duration}$.

The modified MSM method successfully produced results for all profiles to meet the threshold. Jack *et al.* (2018) obtained R^2_{diff} scores from about 0.6 to 0.99 for the 17 household profiles used [42]. The more generic method was able to produce results above the threshold for all profiles when thresholds R_{start} and $R_{duration}$ are both set to 0.85. Figure 3.8a shows a tally comparison of simulated (black bars) and monitored (grey bars) for the start times and duration and the R^2_{diff} scores both above 0.8 for a household. The results of a household for this thesis shows the same graphs in Figure 3.8b with both scores above 0.9.

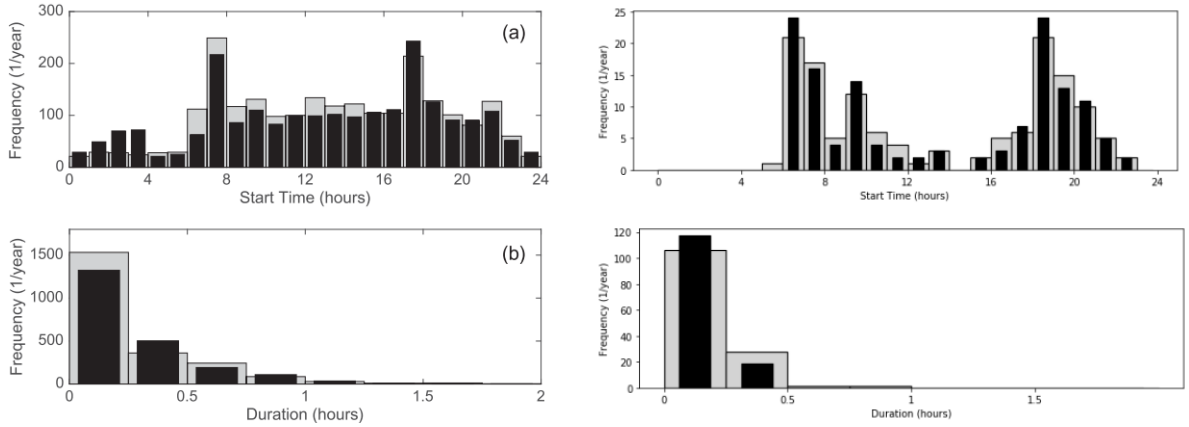


Figure 3.8: Tally comparing monitored (grey) and simulated (black) results of event start times and duration for generated profiles using the MSM method. The results of MSM (left) [42] obtained R^2_{diff} scores > 0.80 for both cases and the results for this thesis (right) obtained R^2_{diff} scores > 0.90 for both cases.

3.4.2 ASHRAE model

The ASHRAE model is a deterministic profile with no random variations. To implement the model, we added random noise to the ASHRAE 90.2 profile, similar to Gholizadeh *et al.* (2016) [70]. We used a normal Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 0.06$. To ensure a fair comparison, the ASHRAE profiles were scaled to match the aggregated median weekly water usage volume of the measured data. Each water usage

event was spread to occur over the hour with a per-minute water flow rate, so that it fitted a normal distribution with deviations at the hour boundaries of

$$T^{\text{hour}} = \text{Average Usage} + U(-500\,\text{sec}, 500\,\text{sec}) \tag{3.10}$$

3.4.3 Evaluation of water usage modelling performance and comparison with other models

We compared the ability of our hot water usage model to model hot water usage demand (differentiated, semi-differentiated and undifferentiated versions) against two prominent existing models for all 77 household water heaters. We evaluated the models' hourly and weekly volume modelling and then compared the error levels of the modelled profiles and the measured profiles. The hourly modelling error is the difference between the modelled volume of water usage over the next hour and the measured volume of water usage over the same hour and is calculated as follows:

$$e_{\text{model|h}}[n] = |V_{\text{meas}}[n] - V_{\text{model}}[n]| \tag{3.11}$$

where n refers to the minute time sample and

$$V_{\text{meas}}[n] = \sum_{n}^{n+60} Q[n] \tag{3.12}$$

The hourly modelling error was calculated using a moving window that calculates the hourly volume modelling error for each minute of the day. For a given EWH, we can compare the hourly modelling error of our model with the hourly modelling errors of the other models (MSM and ASHRAE) by comparing their box and whisker plots. The differences between the resulting hourly modelling errors of one of our models, such as the differentiated model, and other model, in this case ASHRAE, are calculated as follows:

$$e_{\text{res|h}}[n] = e_{\text{diff|h}}[n] - e_{\text{ASH|h}}[n] \tag{3.13}$$

The weekly volume modelling is calculated by finding the difference between the volumes of water used for each week between the measured and the other model's generated profiles, as follows:

$$e_{\text{model|w}}[m] = |V_{\text{meas}}[m] - V_{\text{model}}[m]| \tag{3.14}$$

where m refers to the week of a profile and

$$V_{\text{meas}}[m] = \sum_{m\times 60\times 24\times 7}^{(m+1)\times 60\times 24\times 7} Q[m] \tag{3.15}$$

The resulting weekly modelling error is calculated as follows:

$$e_{\text{res|w}}[m] = e_{\text{diff|w}}[m] - e_{\text{ASH|w}}[m] \tag{3.16}$$

The EWH-specific relative volume modelling error is the average volume modelling error of our model for a specific EWH minus the average volume modelling error of the other model for the same heater. To compare the general modelling performance of our

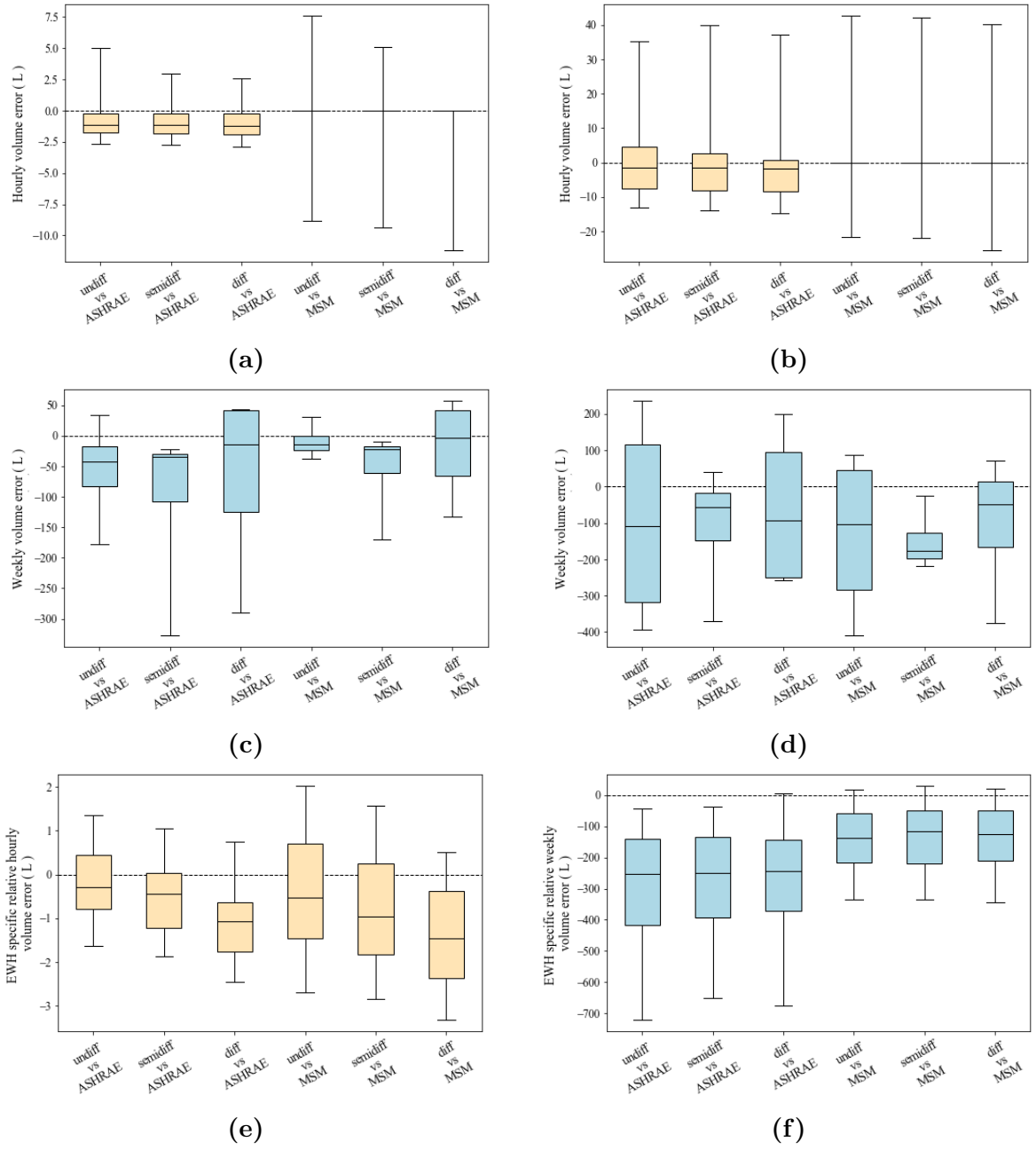


Figure 3.9: a-d) Comparison of the water volume modelling errors for the nominal simulated hot water profiles of our model for all data sets and other existing models. Results are shown for hourly volume modelling errors (Equation 3.13) for a) a light water usage household and b) a heavy water usage household, and results are also shown for weekly volume modelling errors (Equation 3.16) for c) a light water usage EWH and d) a heavy water usage EWH. A negative value indicates that our model had a smaller modelling error, compared to the measured data, than the model against which it is compared. e-f) Comparison of the EWH-specific e) hourly and f) weekly water volume modelling errors (Equations 3.17 and 3.18) for the nominal simulated hot water profiles of our model for all data sets and other existing models for 77 households.

model (differentiated, semi-differentiated and undifferentiated data sets) against the general modelling performance of the other model for all 77 heaters, we plot the distribution of the relative heater-specific modelling error for all the heaters. The hourly and weekly heater-specific relative modelling error for EWH i for our model, such as the differentiated model, and other models, such as ASHRAE, is calculated as follows:

$$\bar{e}_{\text{rel|h}}[i] = \bar{e}_{\text{diff|h}}[i] - \bar{e}_{\text{ASH|h}}[i] \tag{3.17}$$

$$\bar{e}_{\text{rel|w}}[i] = \bar{e}_{\text{diff|w}}[i] - \bar{e}_{\text{ASH|w}}[i] \tag{3.18}$$

The results also compare the three versions of our model for the three different data sets used. The following notations are given for our model: *diff*, the differentiated data set; *semidiff*, the semi-differentiated data set; and *undiff*, the undifferentiated data set.

Nominal water profiles were generated for a heater with a light usage (EWH A) and one with a heavy usage (EWH B). Figures 3.9a and 3.9b show box and whisker plots of the hourly volume modelling errors (calculated using Equation 3.13) of the different models for EWH A and B, respectively. Looking at the subtraction order of this equation, we see that a negative value indicates that our model had a smaller modelling error when compared to the measured data than the two other models did. A positive value indicates that our model's modelling error was larger. Figure 3.9a shows that our model's hourly modelling errors were mostly smaller than those of the ASHRAE model for the light profile. Comparing our model to the MSM model, we see that the performance of our model improves as we move from the undifferentiated to the differentiated version. The absence of a box for the MSM plots is because the flow is mostly zero during the day - a phenomenon captured in the measured data and both our and the MSM models - whereas the ASHRAE model has a water event for every hour. The outcome of this is that most hourly volume errors remain zero for both the MSM models and ours. For the heavy profile of EWH B in Figure 3.9b our model does not perform as well as the ASHRAE model. However, its performance improves as we move from the undifferentiated to the differentiated version.

Figures 3.9c and 3.9d show box and whisker plots of the weekly volume modelling errors (calculated using Equation 3.16) of the various models for EWH A and EWH B, respectively. These figures also show that in most cases all three versions of our model have smaller modelling errors than the other models. We generated a weekly nominal simulated water profile of 77 households to produce the results in this section. Table 3.3 shows the mean of the hourly volume modelling errors in litres for each model compared with those of the measured data for the seven days of the week and the mean of all the days. The mean errors for the ASHRAE and MSM models are 5.9 and 6.0 L. Our model dramatically improves on these figures, from the undifferentiated version's 5.6 L error, to the semi-differentiated version's 5.3 L error, to the differentiated version's modest 4.7 L error. The table not only shows that our model performs better; it also highlights the importance of differentiating between days of the week to forecast water usages with only small errors.

Figures 3.9e and 3.9f show box and whisker plots similar to those in Figure 3.9, but instead show the heater-specific hourly and weekly volume modelling errors calculated according to Equations 3.17 and 3.18, respectively. The negative bias of each plot in both figures

Table 3.3: Table showing the mean hourly volume modelling error (measured in litres) for each model and are differentiated by day of the week. The last column shows the mean error of a full week.

Model	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mean
ASHRAE	5.9	6.0	6.1	5.9	5.8	5.7	5.9	5.9
MSM	6.2	5.8	6.2	5.8	5.7	6.0	6.2	6.0
undiff	5.4	5.1	5.4	5.2	5.4	6.4	6.4	5.6
semidiff	5.3	4.9	5.2	5.0	5.2	5.8	5.7	5.3
diff	4.7	4.1	4.6	4.3	4.5	5.3	5.0	4.7

shows that our model will on average produce smaller modelling errors than the ASHRAE and MSM models. In Figure 3.9e, it is clear that when we use the more differentiated data sets, the modelling errors of our model become even smaller than those of the ASHRAE and MSM models. Figure 3.9f does not show any significant changes when we use a different data set for our model, showing that the weekly volume variations average out over the week.

3.4.4 Individual EWH energy usage modelling performance

The energy profiles for all heaters were simulated using a validated thermal model[75] of an EWH, and using both the measured hot water usage data and the simulated hot water usage events. The heater-specific energy usage modelling error was determined by comparing the simulated energy usage when using the simulated hot water usage events to the simulated energy usage when using the measured hot water usage data. The daily energy usage modelling error (in kWh) for an individual EWH is calculated as follows:

$$e_{\text{model|kWh}}[d] = |E_{\text{meas}}[d] - E_{\text{model}}[d]| \tag{3.19}$$

where d refers to the day in the profile and E refers to the total daily energy usage in kWh. We compared the energy modelling performance of our probabilistic model against the energy modelling performance of the other two models by comparing our average daily energy modelling error against their average daily energy modelling error for each individual EHW, and then plotting the distribution of the EWH-specific *relative* daily energy modelling error over all 77 EWHs. The relative daily energy modelling error of our model relative to another model, e.g. the ASHRAE model, is calculated by taking the difference between our average daily energy modelling error and their modelling average daily energy modelling error, as follows:

$$\bar{e}_{\text{res|kWh}}[i] = \bar{e}_{\text{diff|kWh}}[i] - \bar{e}_{\text{ASH|kWh}}[i] \tag{3.20}$$

where i is the EWH index, $\bar{e}_{\text{rel|kWh}}$ is our relative daily energy modelling error, $\bar{e}_{\text{diff|kWh}}$ is the average daily energy modelling error of our probabilistic model using the differentiated data set, and $\bar{e}_{\text{ASH|kWh}}$ is the average daily energy modelling error of the ASHRAE model. The measured and simulated water profiles for all the heaters in autumn were simulated on an EWH model [76] to produce the energy profile for each household. Figure 3.10 shows box and whisker plots comparing the daily modelling errors of the ASHRAE and MSM models with all three versions of our model (calculated using Equation 3.20). The figure shows that our model will on average model the daily energy load 1kWh/day better than the other models, as shown by the negative median of each box and whisker plot, which is 17% of the median heater's 5.9 kWh/day.

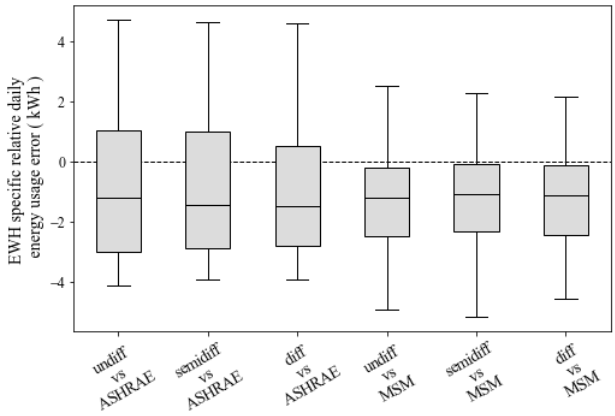


Figure 3.10: Comparison of EWH-specific mean daily energy modelling errors (Equation 3.20) for the nominal simulated hot water profiles of our model for all data sets and other existing models for 77 households.

3.4.5 Aggregated EWH energy usage modelling performance compared to Eskom load data

Finally, we compared a normalised 24-hour energy profile of our 77-household grid, using each heater's simulated water profile, with the "notch test" from Eskom (the South African national electricity grid) [43]. Eskom's "notch test" was a DSM initiative to establish the electricity consumption in South Africa in 1997 [44]. The comparison considered only workdays in autumn, to match the data from Eskom. Figure 3.11 shows the normalised electricity demand over a 24-hour cycle for the measured data, comparing all the models as well as the results obtained from "notch test" for all 77 heaters. The ASHRAE model deviates the most from the measured profile, performing worst at the morning peak, and has early morning demand spikes misaligned with the measured profile. The MSM model and our model both model the early morning spike, but the MSM model overestimates it substantially. The MSM model's profile generally follows the trend of the measured profile, but again overestimates the demand in the period between the morning and evening peaks. Our model's profile accurately follows the measured profile, except for the evening peak where the load is slightly lower than the measured one. This could be because our sample has regional and external variations from the Eskom sample. The root mean squared (RMS) errors of the models' demand profiles are 0.10 for the ASHRAE model, 0.08 for the MSM model, and 0.04 for the differentiated version of our model. Expressed as a percentage of the median heater, these are 16.9% for the ASHRAE model, 13.6% for the MSM model, and 6.8% for our differentiated model.

3.5 Summary

The probabilistic hot water usage model and simulator developed in this chapter was verified to generated synthetic hot water usages that retain the statistical properties of the measured data. Moreover, the performance of the water usage model was compared with two other existing models (ASHRAE and MSM) and showed that our model had the smallest hourly volume modelling errors with that of the measured data. The mean hourly volume modelling error in litres was 6.0 L for MSM, 5.9 L for ASHRAE, 5.6 L for the undifferentiated data set, 5.3 L for the semi-differentiated data set and 4.7 L for the

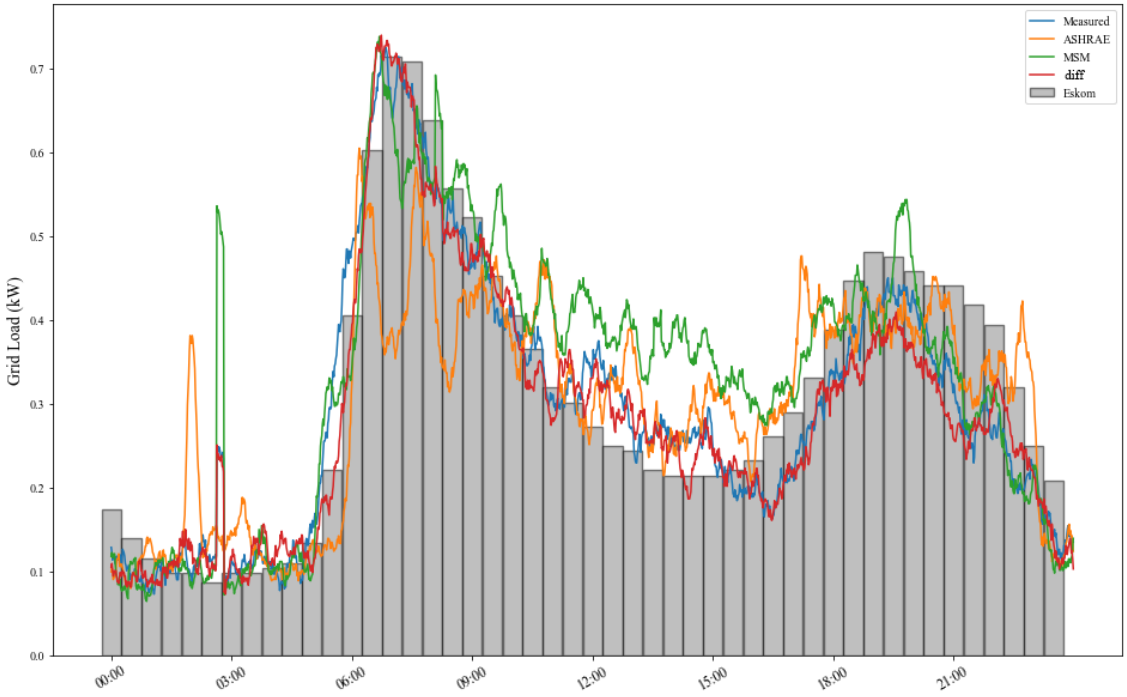


Figure 3.11: Normalised electrical demand plot over 24 hours for 77 EWHs for weekdays in autumn. The Eskom "notch test" electrical demand is represented by a bar plot [43].

differentiated data set of our model. Furthermore, our model is able to model the daily energy load 1kWh/day better than the other models on average for an individual EWH, and the RMS error of the aggregated EWH energy usage was 6.8% from that of the measured profile, showing that our model is ideal for energy optimisation systems.

The developed system can generate synthetic hot water usages for any of the four seasons. If the model is to be used in future work and applied to a real-world scenario, a "blending" strategy could be implemented that accounts for the gradual change from one season to the next. This can improve the accuracy of modelling a households hot water usage profile for practical application.

Chapter 4

EWH modelling and the optimal control problem

In this chapter, an optimal control problem is defined for the heat schedule of an EWH. This chapter explores different types of existing EWH models and a suitable model is chosen that accurately describes the dynamics of the system and has measurable outcomes for which the optimal control problem can be designed. A one- and two-node model were both selected for which the optimal control problem is solved. Four different heating control strategies are defined in this chapter to represent four different possible objectives for the optimal control problem. The four heating control strategies are: thermostat control, optimal temperature matching, optimal energy matching, and optimal energy matching with *Legionella* prevention.

Figure 4.1 shows a block diagram that summarises the system for which the optimal control problem is designed and the sections where each components is discussed is indicated in parenthesis. In Section 4.1, a literature review is performed to investigate existing EWH models, methods of EWH control and optimisation techniques. Section 4.2 defines the heating control strategies that will be considered for optimisation. In Section 4.3, the dynamics of the one- and two-node EWH models are defined. In Section 4.4, the optimal control problem is defined and is solved for in Chapter 5 and 6. Lastly, Section 4.5 describes the development of two EWH feedback mechanisms which are implemented in the final system.

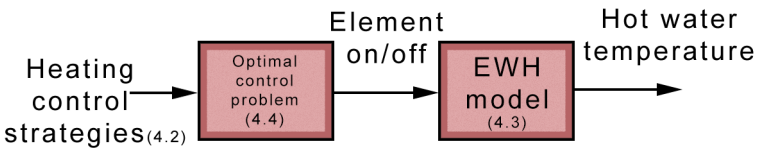


Figure 4.1: Block diagram of components defined in this chapter.

4.1 Literature study

A literature review is performed in this chapter which has the following structure: Section 4.1.1 reviews different types of EWH models. Section 4.1.2 discusses different methods for EWH control. Section 4.1.3 briefly discusses the optimal control problem, reviews different optimisation techniques that have been used in related work and lastly discusses shortest path search algorithms. These reviews are used for the development of the EWH model and optimal control problem which is discussed in the remainder of the chapter.

4.1.1 EWH modelling

Domestic water heaters are suitable for the implementation of DSM methods for energy management techniques and optimisation. Different EWH models have been proposed in literature to assist with DSM algorithms and many studies use simple physical models. However, if one is to model the future thermal behaviour of the tank, a model that incorporates accurate state estimations is required [77].

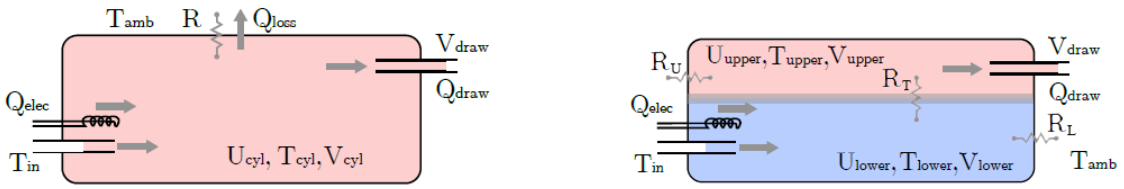


Figure 4.2: Diagram of (a) a one-node and (b) a two-node EWH model [23].

4.1.1.1 One-node model

The one-node model is shown in Figure 4.2a. It assumes that the water inside the tank is at a single uniform average temperature and that the volume of water inside the tank remains constant [37]. The temperature of the water drawn through the outlet pipe is therefore assumed to be at this temperature and cold water simultaneously mixes with the remaining water inside the tank. The one-node model accurately models the temperature profile of an EWH *as long as* water draws do not occur. The accuracy decreases more when the water usage is larger. In reality, drawing a volume of water from the tank causes stratification between the bodies of water at different temperatures, and this is not accounted for in the case of the one-node model [75].

4.1.1.2 Two-mass composite model

Diao *et al.* (2012) presents a two-mass composite EWH model which transitions between one and two bodies of water at separate temperatures as a result of water draws [37]. The height of each mass of water is tracked in the two-mass state to account for heat losses to the environment. A disadvantage of this approach is that heat transfer between the two masses is neglected [37, 77]. It provides insufficient accuracy to simulate over long time periods due to the error accumulation [78].

4.1.1.3 Two-node model

The two-node EWH model is similar to that of the two-mass composite model but with the addition of accounting for heat transfer between the two water masses, or nodes [77]. The two-node EWH model is shown in Figure 4.2b and differs from the one-node model by accounting for stratification. The interface between the hot and cold water nodes creates an inversion layer where heat transfer occurs due to natural stratification and the vertical height of this layer is constantly tracked [75].

4.1.1.4 Multi-node model

The multi-node model comprises multiple temperature layers stacked above each other, and each layer is modelled as an open system energy balance. Figure 4.3 shows the EWH model, where T_{i} refers to the temperature at node i of a total of n nodes. The thermal energy of a single node is affected by the electrical energy, water usage, standing losses and conduction between layers [77].

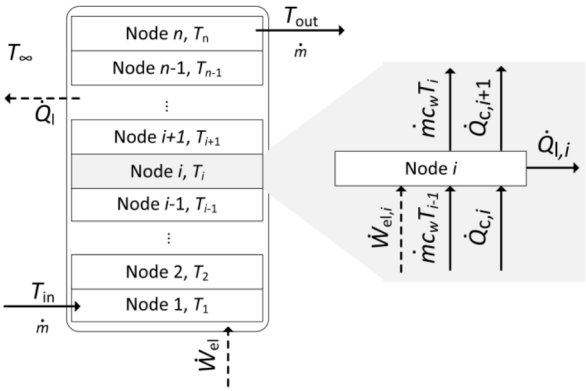


Figure 4.3: Multi-node model of an EWH and energy balance for a single node [17].

This model mimics the temperature gradient that occurs in the tank when there is a temperature inversion due to water leaving the tank. Kepplinger *et al.* (2019) evaluated that the multi-node model accurately models the average temperature of the tank and outlet temperature, but requires 50 times more computational effort than that of the one-node model [77].

4.1.2 EWH control

This section looks at different methods of controlling a residential EWH. The first is traditional control, followed by schedule control and lastly, optimal control. This section discusses the optimal control problem, followed by different optimisation techniques and related work.

4.1.2.1 Traditional control

An EWH is typically controlled by a thermostat which switches the heating element on or off. The thermostat keeps the temperature of the water inside the tank at a specific set-point. When the temperature falls below the set-point, the thermostat switches the

heating element on and turns it off when the temperature is above the set-point [79]. Although this is the traditional control method, it is inefficient because a higher water temperature means higher thermal losses. Therefore it is ideal for the temperature of the water to not be at the desired set-point temperature when water is not used. The heating element also continuously replaces the energy lost to the environment to keep the temperature at the set-point.

4.1.2.2 Schedule control

Schedule control (SC) aims to produce a simple control strategy to improve the efficiency of the EWH from traditional control. Schedule control analyses the water demand of a household and creates a heating schedule that turns on the thermostat for a fixed period of time before any significant water usage [80]. Schedule control can either be controlled by the user, based on their water usage behaviour or for DSM strategies, or automated to produce a heating schedule based on the water usage patterns.

4.1.2.3 Optimal control

Optimal control (OC) of an EWH means determining the control function that minimises the costs defined for the system while supplying sufficient hot water [17]. The next section investigates existing optimisation techniques applicable for the optimal control of EWHs.

4.1.3 The optimal control problem

An optimal control problem is a problem where the solution lies in determining the inputs to a dynamic system that optimises (minimises or maximises) the performance while satisfying all constraints imposed on the system. The subject has existed as far back as 1950, and the problems are solved numerically due to the high complexity of most applications. More recently, the complexity of methods and the variety of applications have increased drastically [81].

4.1.3.1 Optimisation techniques

This section reviews related work that makes use of different optimisation techniques to achieve optimal control. Table 4.1 summarises the related work discussed in this section in chronological order and shows: the optimisation method, the type of scheduling that the method is applied to and, in the case of hot water, the type of EWH model used, the time resolution used, and the reported energy savings if given.

Table 4.1: Table summarising all related work and methods for optimisation.

Ref.	Optimisation Method	Schedule type	EWH model	Res.	Reported Energy Savings
Laurent (1995) [82]	CG[1]	Hot Water	One	5 min	-
Sepulveda (2010) [16]	BPSO[2]	Hot Water	One	15 min	-
Tischer (2011) [83]	DP	Electricity	-	15 min	-
Kepplinger (2014) [17]	DP	Hot Water	Multi	1 hr	12%
Kepplinger (2015) [18]	BIP[3]	Hot Water	One	1 hr	4-12%
Heleno (2015) [84]	MPM[4]	Hot Water	One	1 hr	-
Kepplinger (2016) [19]	BLO[5]	Hot Water	Multi	1 hr	12%
Al-Jabery (2016) [20]	QL[6] and ADHDP[7]	Hot Water	One	1 hr	-
Roy (2017) [85]	GA[8] and DP	Electricity	-	1 min	17-36%
Ahmed (2017) [86]	BBS[9]	Electricity	-	1 hr	21-16%

[1] *CG: Column Generation.* [2] *BPSO: Binary Particle Swarm Optimisation.*
[3] *BIP: Binary Integer Programming.* [4] *MPM: Multi-period model.*
[5] *BLO: Binary Linear Optimisation.* [6] *QL: Q-learning.*
[7] *ADHDP: Action Dependent Heuristic Dynamic Programming.*
[8] *GA: Genetic Algorithm.* [9] *BBS: Binary Backtracking Search.*

Dynamic Programming (DP) is a computer programming technique that is also a tool for optimisation. The problem is broken down into a multistage decision process for which the globally optimal path can be found. A cost function defines the variables of the system to minimise or maximise. The optimal control sequence is obtained to produce the optimal trajectory of the system states while keeping within the bounds of admissible states that are represented by constraints [87]. DP produces a look-up table for the optimal state trajectory and optimal input sequence for any allowable initial state. A drawback of DP is the high computational complexity and large amount of memory required to store the look-up table. It also suffers from the "curse of dimensionality". This means that high dimensional systems exponentially increase the computational complexity and table sizes. This makes DP ideal only for low dimensional systems [88].

DP branches into two sub-categories: Deterministic Dynamic Programming (DDP) and Stochastic Dynamic Programming (SDP). The difference between these is that DDP refers to DP where there are no uncertainties, and SDP is DP where decisions are made with uncertainty. Kepplinger *et al.* (2014) proposed an optimisation method that used dynamic programming to optimise cost or energy use [17]. The EWH model used was multi-nodal and the method used hourly optimisation to control the EWH on an hourly basis [1]. In later work, Kepplinger *et al.* (2015) used a one-node model of an EWH to optimise both price and energy savings using a technique called Binary Integer Programming (BIP) [18]. This optimised a pseudo cost function relating to EWH usage with energy savings of 4 to 12%. The constraint on the system to satisfy cold events could not be fully accomplished. The methods used included matching the energy to that of thermostat control. However, they did not include temperature matching, which matches the start temperature of each water usage to that of thermostat control [1].

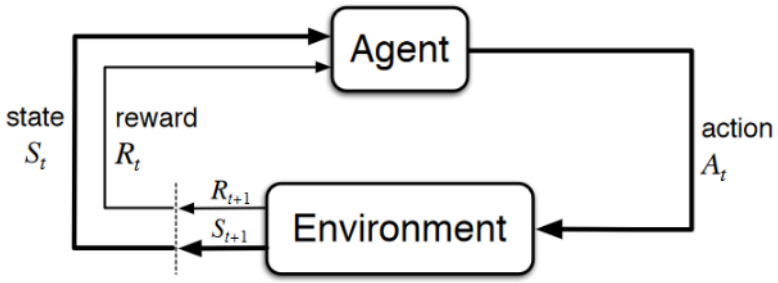


Figure 4.4: The agent-environment interaction in Q-learning [91].

Reinforcement learning (RL) provides the perspective of animal behaviour and how this can optimally control the surrounding environment [89]. More information on RL can be found in [89]. Q-learning is a form of RL which provides learning capabilities for optimal functionality. Q-learning learns a policy that communicates to agents the best action to perform at the current state in the environment. The best action is the action that results in the biggest reward, and the concept is summarised in Figure 4.4. Due to the step-by-step nature of obtaining the optimal sequence, the method is labelled as incremental DP [90]. Action Dependent Heuristic Dynamic Programming (ADHDP) is a generalisation of Q-learning that uses SDP and adaptive controllers to train a critic that tries to estimate a cost function to minimise the cost of the next time step [20]. Al-jabery *et al.* (2016) used Q-learning and ADHDP techniques to optimise the energy used by EWHs using a one-node model [20]. Q-learning was demonstrated to be the better method and could save 26% while ADHDP saved 21%. These savings were reported as cost and not energy.

An evolutionary optimisation technique known as particle swarm optimisation (PSO) is based on the concept of intelligent swarms. These are individuals interacting to optimise a goal by adapting to the environment, such an example is a flock of birds. The process begins by initialising a population with randomly assigned positions and velocities. Each individual keeps track of its personal best position as well as the population's best position. Each iteration assigns a random acceleration to each individual towards its personal best and population best positions. This process occurs until the population has converged on the global minimum or maximum [92]. Sepulveda *et al.* (2010) applied a binary particle swarm optimisation (BPSO) technique to represent a one-node EWH as particles that move through the specified space [16]. The position of the particles is evaluated by a fitness function that has the objective of minimising peak loads while maximising the EWH water temperature. Results showed that load peaks reduced between 13 and 16%.

4.1.3.2 Shortest path search

The shortest path search (SPS) is a common algorithm that finds the shortest path from a specified node (or position) to a destination node [93]. The shortest path problem (SPP) is used commonly in network related applications such as highway systems, railroads, and communication networks [94]. The next two sections look at two popular SPS methods known as the Dijkstra and A* algorithm.

4.1.3.2.1 Dijkstra algorithm

The Dijkstra algorithm is a well-known SPP technique which finds the optimal path from the starting node to destination. This algorithm searches for the minimum cost path by beginning with the starting node and iteratively searches all sub paths until the destination node is found [95]. This method tends to make paths expand concentrically from the starting node and has the disadvantage of poor search efficiency [94]. This process can make use of a priority queue to prioritise paths with lower "cost-to-arrive" over others to improve computational time. The "cost-to-go" is the accumulated cost to reach the current node from the starting node.

4.1.3.2.2 A* Algorithm

The A* algorithm is an extension of the Dijkstra algorithm which adds a heuristic component to the process. The A* algorithm will produce an identical optimal path to that of Dijkstra, however it will greatly reduce the number of iterations to achieve this. The evaluation function $f()$ keeps track of the cost of a specific path by calculating the path's current cost. The current cost is a combination of the "cost-to-arrive" $g()$ and the "cost-to-go" $h()$. The evaluation function is defined as follows:

$$f(n) = g(n) + h(n) \tag{4.1}$$

where n is the index of the current node considered. The "cost-to-go" is an estimated cost of reaching the destination node from the current node. Because this cost cannot be calculated without first finding the complete path, heuristics are used to obtain the best estimate of the cost which results in a faster algorithm.

4.1.3.2.3 SPS heuristics

SPS algorithms such as A* use heuristics to guide the search process from the current position closer to the shortest path to the target position. The heuristic component *h()* of the evaluation function is an estimated cost from the current to the target position. An important requirement of *h()* is that it must always be admissible. This means that the estimated cost must be lower than or equal to the actual cost between the current position and the target position as well as never being negative. In general, the heuristics reduce the number of nodes required to reach the target position and the extent of the reduction depends on how the heuristics are defined for the problem at hand [95].

There are two distance metrics that are commonly used to calculate SPS heuristics. The first is the **Euclidean** distance which is determined as the absolute distance from the current to the target position and is calculated using Equation 4.2, where symbols X and Y refer to the x and y Cartesian coordinates of the current or target position, respectively. The second distance metric, and also the less computationally expensive of the two, is the **Manhattan** distance which is determined by separating the distances of the two positions in the x and y-planes, and is calculated using Equation 4.3 [96].

$$h = \sqrt{(X_{\text{cur}} - X_{\text{tar}})^2 + (Y_{\text{cur}} - Y_{\text{tar}})^2} \tag{4.2}$$

$$h = |X_{\text{cur}} - X_{\text{tar}}| + |Y_{\text{cur}} - Y_{\text{tar}}| \tag{4.3}$$

4.2 Heating control strategies

In this thesis, thermostat control is taken as the baseline heating control strategy and three alternative heating control strategies, namely temperature matching, energy matching, and energy matching with *Legionella* prevention, are evaluated in terms of energy savings and number of cold events (cold events are usage events where the water temperature is lower than the temperature desired by the user).

4.2.1 (Baseline) Thermostat control (TC)

This is the mode in which water heaters are designed to be used and how most people use them. The thermostat strives to maintain the water at a target temperature, normally set between 65 °C and 75 °C, with a small hysteresis band around the set temperature. This heating strategy is wasteful, as it maintains a high temperature between draw events that may be far apart in time, which leads to the most thermal losses of the different strategies. For this heating strategy and set temperature, water is drawn from the water heater at a temperature that is higher than people require for most uses. To achieve a convenient temperature, the user normally must regulate the temperature by mixing with cold water to achieve a nominal temperature of approximately 40 °C [79, 97, 77].

4.2.2 Scheduled control with temperature matching (TM)

In an attempt to save on bills and reduce loss of energy, financially sensitive and environmentally aware users resort to turning off their heaters for extensive periods between the times when they need hot water [80]. The timing applied in this strategy is individually motivated and could vary significantly between users, but the optimal scenario (for a water heater with energy storage) is to switch the heater on just sufficiently in advance of the time the hot water will be needed [98, 80]. When applying optimisation techniques, as we have done in this study, optimisation constraints can be set to ensure that the same volume of water is drawn at the same temperatures as under thermostat control, but the thermal losses to the environment are minimised. This approach assumes that the user requires water at these high temperatures, intending to mix the water with cold water to achieve a desired temperature. For this control strategy, the heater will deliver the same amount of useful output energy as in the baseline thermostat control scenario.

4.2.3 Scheduled control with energy matching (EM)

An alternative approach to the optimised schedule control with temperature matching is to assume that the user does not require the water at the high temperatures, but rather is satisfied with a lower, more directly usable temperature. In this case, there is less need, or no need, to add cold water, as the water is already at or just above the desired temperature for use. A lower target temperature during water draw-offs, of say 38 °C, could be used. To ensure a fair comparison, we increased the volume drawn from the heater to ensure that the same amount of energy was delivered in the water drawn as under thermostat control [79, 97, 67].

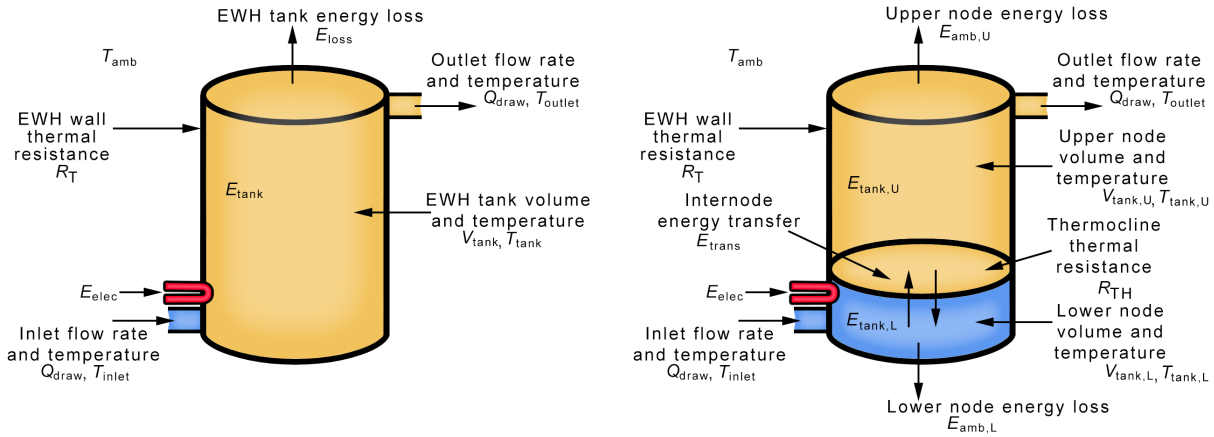
4.2.4 Scheduled control with energy matching plus *Legionella* prevention (EML)

Although optimising the heating schedule has energy benefits, there are health risks when the water is maintained and delivered at low temperatures. *Legionella pneumophila* thrives at temperatures between 32 °C and 42 °C and has been found in water heaters [79, 99]. According to Stout *et al.* (1986), to sterilise the bacteria, the heater must spend 11 min at 60 °C or 3 min at 70 °C, at least once a day [15].

We therefore define a *Legionella*-driven control strategy that ensures that the heater temperature reaches 60 °C for at least 11 min at least once before the largest water usage event of the day.

4.3 EWH modelling

The EWH thermodynamics can be modelled using a one-node, two-node, or multi-node model. As mentioned in Section 4.1.1, the one-node model does not model stratification, whereas the two-node and multi-node model do. The EWH tank can be oriented vertically or horizontally. In this thesis, only the vertical orientation of the tank is considered. The results can vary for a model that considers stratification if a horizontal orientation is used, but the computational complexity increases considerably. More specifically, the complexity is introduced because a numerical solution is required to calculate the tank area that holds the body of water for one of the nodes. The EWH is modelled according to an energy balance equation to track the energy inside the tank as a result of energy flow inside the system. Figure 4.5 shows diagrams to distinguish between the one- and two-node EWH model and are used to aid the derivation of the thermal dynamics of each model in the following sections.



(a) One-node EWH. **(b)** Two-node EWH with stratification.

Figure 4.5: Diagram of energy flow, thermal resistance, flow rate, temperature, and volume in a) one-node and b) two-node EWH.

4.3.1 One-node dynamics

In the one-node model, the body of water inside the tank is assumed to be at a single uniform temperature and is shown in Figure 4.5a. Thermal energy flows into the tank due to a heating element which consumes electrical power that is determined by the rated power of the heating element. When water is drawn from the tank at temperatures higher than that of the inlet pipe, there is a reduction in thermal energy at a volumetric flow rate. When there is a temperature difference between the water inside the tank and the ambient temperature of the environment, thermal energy is lost from the tank at a rate determined by the thermal resistance of the tank.

The thermal dynamics of the one-node lumped-parameter model is expressed as follows:

$$\dot{E}_{\text{tank}}(t) = P_{\text{elec}}(t) - P_{\text{draw}}(t) - P_{\text{loss}}(t) \tag{4.4}$$

where E_{tank} is the thermal energy in the tank, P_{elec} is the power delivered by the heating element,P_{draw} is the power leaving the tank when hot water is drawn, and P_{loss} is the power leaving the tank due to losses to the environment. The equation shows that the rate of change of thermal energy is directly influenced by P_{elec}, P_{draw} and P_{loss}.

4.3.1.1 Heating element

The heating element supplies electrical power at P_{rated} when the switch is on and zero when off and is shown as follows:

$$P_{\text{elec}}(t) \in \{0, P_{\text{rated}}\} \tag{4.5}$$

The energy supplied by the element is given as follows:

$$E_{\text{elec}} = \int_{t_0}^{t_f} P_{\text{elec}}(t)dt \tag{4.6}$$

4.3.1.2 Water usage

The power lost during hot water usage is given as follows:

$$P_{\text{draw}}(t) = \rho Q_{\text{draw}}(t)[\hat{h}_{\text{outlet}}(t) - \hat{h}_{\text{inlet}}(t))] \tag{4.7}$$

where $Q_{\text{draw}}(t)$ is the hot water outlet volumetric flow rate, ρ is the density of water and $\hat{h}_{\text{inlet}}$ and $\hat{h}_{\text{outlet}}$ are the specific enthalpy entering and leaving the water heater, respectively. Under conditions of constant pressure and constant specific heat capacity, this can be approximated by

$$P_{\text{draw}}(t) \approx c_P \rho Q_{\text{draw}}(t)[T_{\text{outlet}}(t) - T_{\text{inlet}}(t)] \tag{4.8}$$

where c_P is the constant pressure-specific heat capacity of the water, T_{outlet} is the hot water outlet temperature, and T_{inlet} is the cold water inlet temperature. The hot water outlet flow rate $Q_{\text{draw}}(t)$ is the superposition of the flow rate due to intentional usage $Q_{\text{usage}}(t)$ and the flow rate due to unintentional usage $Q_{\text{unintentional}}(t)$, defined below.

$$Q_{\text{draw}}(t) = Q_{\text{usage}}(t) + Q_{\text{unintentional}}(t) \tag{4.9}$$

Unintentional water usages refers to the volume of water that is drawn from the tank and is not effectively used by the user. At the end of a water usage, there remains warm water inside the piping network between the tank and the point of use. The water will cool down and the energy lost from the water will contribute to usage losses. The power output due to hot water draw $P_{\text{draw}}(t)$ is therefore also the superposition of the power output due to intentional usage $P_{\text{usage}}(t)$ and the power output due to unintentional usage $P_{\text{unintentional}}(t)$

$$P_{\text{draw}}(t) = P_{\text{usage}}(t) + P_{\text{unintentional}}(t) \tag{4.10}$$

with

$$P_{\text{usage}}(t) = c\rho Q_{\text{usage}}(t)[T_{\text{outlet}}(t) - T_{\text{inlet}}(t)] \tag{4.11}$$
$$P_{\text{unintentional}}(t) = c\rho Q_{\text{unintentional}}(t)[T_{\text{outlet}}(t) - T_{\text{inlet}}(t)] \tag{4.12}$$

Similarly, the energy output due to hot water draw $E_{\text{draw}}(t)$ is the superposition of the energy output due to intentional usage $E_{\text{usage}}(t)$ and the energy output due to unintentional usage $E_{\text{unintentional}}(t)$

$$E_{\text{draw}}(t) = E_{\text{usage}}(t) + E_{\text{unintentional}}(t) \tag{4.13}$$

The distinction between usage and unintentional usage is important because we give the usage profile to the optimal control algorithm as an objective to satisfy, while we treat the unintentional usage as a disturbance and an energy loss.

4.3.1.3 Standing losses

Thermal energy is lost to the environment due to the temperature difference of the water inside the tank and surrounding ambient temperature. The power heat loss due to the environment is calculated as follows:

$$P_{\text{loss}}(t) = \frac{1}{R_{\text{T}}}[T_{\text{tank}}(t) - T_{\text{amb}}(t)] \tag{4.14}$$

where R_{T} is the thermal resistance of the wall of the tank, T_{tank} is the temperature of the water inside the tank and T_{amb} is the ambient temperature.

4.3.1.4 Tank enthalpy and temperature

The enthalpy of the tank can be calculated using the following equation:

$$E_{\text{tank}}(t) = c_V \rho V_{\text{tank}} T_{\text{tank}}(t) \tag{4.15}$$

where c_V is the constant volume-specific heat capacity and V_{tank} and T_{tank} are the volume and temperature of the water inside the tank.

The temperature of the water inside the tank is assumed to be equal to the outlet temperature T_{outlet}. The relationship between the EWH enthalpy E_{tank} and the EWH water

temperature T_{tank} relative to a reference temperature where we define energy to be zero, is given as follows:

$$T_{\text{tank}}(t) = \frac{E_{\text{tank}}(t)}{c_V \rho V_{\text{tank}}} \approx \frac{E_{\text{tank}}(t)}{c_P \rho V_{\text{tank}}} \tag{4.16}$$

where c_V is approximately equal to c_P for water, and henceforth denoted as c.

4.3.2 Two-node dynamics

The two-node EWH in Figure 4.5b models stratification by introducing a thermocline that divides the tank into an upper and lower node which represents the hot and cold water, respectively. Water leaving the outlet pipe is at the temperature of the upper node and water entering the inlet pipe is at the temperature of the lower node. Inter-node energy transfer occurs due to the temperature difference at the thermocline between the two bodies of water at a rate determined by the thermal resistance of the thermocline. The heating element is situated at the bottom of the tank and provides thermal energy to the lower node.

When the tank is in a one-node state and water is drawn from the tank at a higher temperature than the inlet water temperature, the tank transitions into a two-node state. When all the hot water is drawn from the tank, the EWH transitions back to a one-node state and the temperature of the whole tank is that of the lower node. The EWH also transitions to a one-node state when the lower node temperature reaches the temperature of the upper node. The nodes are referred to as the upper and lower node and are designated by subscripts U and L respectively. The thermal dynamics of the two-node lumped-parameter model is a superposition of the upper and lower node and is expressed as follows:

$$\dot{E}_{\text{tank}}(t) = \dot{E}_{\text{tank,U}}(t) + \dot{E}_{\text{tank,L}}(t) \tag{4.17}$$

The upper node's rate of change of thermal energy is influenced by $P_{\text{draw,U}}$, $P_{\text{loss,U}}$ and the power of inter-node energy transfer $P_{\text{trans,U}}$ and is expressed as follows:

$$\dot{E}_{\text{tank,U}}(t) = -P_{\text{draw,U}}(t) - P_{\text{loss,U}}(t) - P_{\text{trans,U}}(t) \tag{4.18}$$

The lower node is supplied with power P_{elec} from the heating element and the lower node's rate of change of thermal energy is expressed as follows:

$$\dot{E}_{\text{tank,L}}(t) = P_{\text{elec}}(t) - P_{\text{inlet,L}}(t) - P_{\text{loss,L}}(t) - P_{\text{trans,L}}(t) \tag{4.19}$$

where $P_{\text{inlet,L}}(t)$ is the power entering the lower node due to the thermal energy in the cold water.

4.3.2.1 Water usage

The power lost from the upper node during water usage $P_{\text{draw,U}}(t)$ is influenced by the volumetric flow rate of water leaving the tank:

$$P_{\text{draw,U}}(t) \approx c_P \rho Q_{\text{draw}}(t) T_{\text{tank,U}}(t) \tag{4.20}$$

where $T_{\text{tank,U}}$ is the temperature of the upper node. The rate of change of the upper node volume is expressed as follows:

$$\dot{V}_{\text{tank,U}}(t) = -Q_{\text{draw}}(t) \tag{4.21}$$

where $V_{\text{tank,U}}$ is the volume of the upper node. The relationship between the upper and lower node volumes are expressed as follows:

$$V_{\text{tank,U}}(t) = V_{\text{tank}} - V_{\text{tank,L}}(t) \tag{4.22}$$

where $V_{\text{tank,L}}$ is the volume of water in the lower node. The volume ratio is calculated as follows:

$$V_{\text{r}}(t) = \frac{V_{\text{tank,U}}(t)}{V_{\text{tank}}} \tag{4.23}$$

Unintentional water usage: It can be noticed that the one-node EWH dynamics differentiated between intended and unintended water usages when water is drawn from the tank. This thesis explores the effect of unintentional water usages on the thermal energy drawn from only a simulated one-node EWH model. The two-node EWH model does not differentiate between the two types of usages and $P_{\text{draw,U}}$ and $P_{\text{draw,L}}$ is assumed to be a combination of them.

4.3.2.2 Standing losses and inter-node energy transfer

The power loss of each node is dependent on its temperature. The power loss of the upper node due to the environment is calculated as follows:

$$P_{\text{loss,U}}(t) = \frac{1}{R_{\text{tank,U}}}[T_{\text{tank,U}}(t) - T_{\text{amb}}(t)] \tag{4.24}$$

where $R_{\text{tank,U}}$ is the thermal resistance of the tank at the upper node. The power transfer of the upper node due to the thermal conductivity between the nodes is calculated as follows:

$$P_{\text{trans,U}}(t) = \frac{1}{R_{\text{TH}}}[T_{\text{tank,U}}(t) - T_{\text{tank,L}}(t)] \tag{4.25}$$

where R_{TH} is the thermal resistance of the thermocline. The power heat loss of the lower node due to the environment is calculated as follows:

$$P_{\text{loss,L}}(t) = \frac{1}{R_{\text{tank,L}}}[T_{\text{tank,L}}(t) - T_{\text{amb}}(t)] \tag{4.26}$$

where $R_{\text{tank,L}}$ is the thermal resistance of the tank at the lower node. The power transfer of the lower node is calculated as follows:

$$P_{\text{trans,L}}(t) = \frac{1}{R_{\text{TH}}}[T_{\text{tank,L}}(t) - T_{\text{tank,U}}(t)] \tag{4.27}$$

4.3.2.3 Thermal resistance

The thermal resistance $R_{\text{tank,U}}$ and $R_{\text{tank,L}}$ is calculated as follows:

$$R_{\text{tank,U}} = R\frac{A_{\text{tank}}}{A_{\text{tank,U}}} \tag{4.28}$$

$$R_{\text{tank,L}} = R\frac{A_{\text{tank}}}{A_{\text{tank,L}}} \tag{4.29}$$

where A_{tank}, $A_{\text{tank,U}}$ and $A_{\text{tank,L}}$ are the areas of the tank, upper node, and lower node, respectively. The areas are calculated as follows:

$$A_{\text{tank}} = 2\pi r^2 + 2\pi rL \tag{4.30}$$

$$A_{\text{tank,U}} = \pi r^2 + 2\pi L\frac{V_{\text{tank,U}}}{V_{\text{tank}}} \tag{4.31}$$

$$A_{\text{tank,L}} = \pi r^2 + 2\pi L\frac{V_{\text{tank,L}}}{V_{\text{tank}}} \tag{4.32}$$

where r is the tank radius and L is the length of the tank. The thermocline is assumed to have zero mass and only provides an interface of energy transfer due to the temperature difference. The thermal resistance of the thermocline is calculated as follows:

$$R_{\text{TH}} = \frac{d}{kA_{\text{TH}}} \tag{4.33}$$

where d is the layer thickness, k is the thermal conductivity of water and A_{TH} is the area of the thermocline between the nodes. This area of the thermocline is calculated as follows:

$$A_{\text{TH}} = \pi r^2 \tag{4.34}$$

4.4 The optimal control problem formulation

In this section we briefly explain optimal control theory, using Kirk (2012) [87] and Engelbrecht (2016) [88] as our primary source, and then formulate the EWH control as an optimal control problem.

4.4.1 Optimal control theory

Optimal control theory is concerned with finding a control law for a given system so that a specified optimality criterion is achieved. This criterion is typically specified as a cost function to be minimised. The problem is therefore to find a control law that produces the optimal control input signals and resulting optimal state trajectories that together minimise a given cost function, subject to a set of dynamic constraints, terminal state constraints, and state variable and control input inequality constraints. This can be expressed mathematically as follows

$$\mathbf{u}^*(t) = \arg\min_{\mathbf{u}(t)} J(\mathbf{x}(t), \mathbf{u}(t), t) \tag{4.35}$$

$$= \arg\min_{\mathbf{u}(t)} \int_{t_0}^{t_f - 1} g(\mathbf{x}(t), \mathbf{u}(t), t) \tag{4.36}$$

subject to the dynamic constraint

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \tag{4.37}$$

the terminal state constraint

$$\mathbf{x}(t_f) \in \mathcal{X}_f \tag{4.38}$$

and the state space and input space constraints

$$\mathbf{x}(t) \in \mathcal{X} \tag{4.39}$$

$$\mathbf{u}(t) \in \mathcal{U} \tag{4.40}$$

In these equations, J is the cost function, $\mathbf{x}(t)$ is a candidate state trajectory, $\mathbf{u}(t)$ is a candidate control input signal, t is time, $\mathbf{f}()$ represents the set of nonlinear, time-variant differential equations of the system, $g()$ is the state transition cost function, $\mathcal{X}_\mathrm{f}$ is the set of admissible final states, and $\mathcal{X}$ and $\mathcal{U}$ are the sets of admissible states and admissible inputs. $\mathbf{x}^*(t)$ and $\mathbf{u}^*(t)$ are the optimal state trajectory and optimal control input signal that minimise the cost function subject to the constraints.

4.4.2 Formulation of EWH thermal control as an optimal control problem

Given a hot water usage profile in terms of flow rate $Q_{\mathrm{usage}}(t)$ and desired minimum hot water temperature $T_{\mathrm{usage}}(t)$ as a function of time t, the cold water inlet temperature $T_{\mathrm{in}}(t)$ and the ambient temperature $T_{\mathrm{amb}}(t)$ as time-varying disturbance signals, and scheduled supply-side interruption of the electricity supply to the EWH $P_{\mathrm{max}}(t)$, we wish to determine the optimal EWH control signal $P^*_{\mathrm{elec}}(t)$ that will satisfy the hot water usage profile, while minimising the total energy used.

4.4.2.1 System dynamics

The system dynamics are defined as the nonlinear differential equations describing the EWH thermal dynamics for both cases of one and two-node as described above, and specifically by equations (4.4), (4.5), (4.8), and (4.14) for one-node and equations (4.17), (4.20), (4.24), (4.25), (4.22), (4.23), (4.26) and (4.27) for two-node dynamics.

4.4.2.2 System State

The state variable for the system is the thermal energy E of the water inside the EWH

$$x(t) = E(t) \tag{4.41}$$

4.4.2.3 Control Input

The control input for the system is the power input P_{elec} delivered by the heating element

$$u(t) = P_{\text{elec}}(t) \tag{4.42}$$

4.4.2.4 Disturbance

The water drawn from the tank acts as a disturbance to the system where it causes energy to leave the tank, and the control input must operate to maintain the energy of the tank between the limitations. The power lost due to water usage is defined for the one-node model as follows:

$$\mathbf{w} = P_{\text{draw}} \tag{4.43}$$

and for the two-node model as follows:

$$\mathbf{w} = P_{\text{draw,U}} \tag{4.44}$$

4.4.2.5 State constraints

The physical limitations on the thermal energy inside the EWH are specified by defining the following set of admissible states

$$E(t) \in [E_{\min}, E_{\max}] \tag{4.45}$$

where the minimum energy $E_{\min}$ and the maximum energy $E_{\max}$ correspond to the minimum temperature $T_{\min}$ and the maximum temperature $T_{\max}$ specified for the EWH

$$E_{\min} = c\rho V_{\text{tank}} T_{\min} \tag{4.46}$$
$$E_{\max} = c\rho V_{\text{tank}} T_{\max} \tag{4.47}$$

4.4.2.6 Input constraints

The control input constraints are specified by defining the following set of admissible inputs

$$P_{\text{elec}}(t) \in \{0, P_{\text{rated}}\} \tag{4.48}$$

The heating element is either off or on, which means that the power input P_{elec} delivered by the element is either zero or its rated power P_{rated}.

4.4.2.7 Boundary conditions

The initial and terminal boundaries conditions are specified by defining the following temperatures

$$T(t_{\text{i}}) = T_{\text{start}} \tag{4.49}$$
$$T(t_{\text{f}}) = T_{\text{end}} \tag{4.50}$$

where t_i is the initial time and t_f is the final time. The temperatures T_{start} and T_{end} are the desired initial and final temperatures. The boundary conditions are selected such that the optimal path begins and ends at the corresponding desired temperatures.

4.4.2.8 Cost function

The objective to minimise the total energy used is represented by the following cost function

$$J = \int_{t_i}^{t_f} P_{\text{elec}}(t)dt \tag{4.51}$$

Note that this a specific implementation of the general cost function

$$J = \int_{t_i}^{t_f} g(\mathbf{x}(t), \mathbf{u}(t), t)dt \tag{4.52}$$

with the state transition cost defined as

$$g(\mathbf{x}(t), \mathbf{u}(t), t) = P_{\text{elec}}(t) \tag{4.53}$$

4.4.2.9 Temperature profile constraints (usage and *Legionella* prevention)

The objective of satisfying the hot water usage profile is represented by the following time-varying state inequality constraints on the temperature of the water inside the one-node EWH

$$T_{\text{tank}}(t) \geq T_{\text{profile}}(t) \tag{4.54}$$
$$\tag{4.55}$$

and the temperature of the water for both nodes inside the two-node EWH

$$T_{\text{tank,U}}(t) \geq T_{\text{profile,U}}(t) \tag{4.56}$$
$$T_{\text{tank,L}}(t) \geq T_{\text{profile,L}}(t) \tag{4.57}$$

The need to increase the temperature once per day to prevent the growth of *Legionella* bacteria can also be included in this time-varying state inequality constraints.

For the **one-node model**, the profile temperature $T_{\text{profile}}(t)$ is set to the desired usage temperature T_{usage} for times that correspond to draw. The profile temperature $T_{\text{profile}}(t)$ is set to the minimum *Legionella* prevention temperature $T_{\text{Legionella}}$ to ensure that even the entire tank prevents bacterial growth. Both profiles are set to the minimum EWH temperature T_{min} for all other times.

$$T_{\text{profile}}(t) = \left\{ \begin{array}{ll} T_{\text{Legionella}} & \text{once per day to prevent } \textit{Legionella} \text{ growth} \\ T_{\text{usage}} & \text{when intentional draw causes } Q_{\text{draw}}(t) > 0 \\ T_{\text{min}} & \text{otherwise} \end{array} \right\} \tag{4.58}$$

For the **two-node model**, the upper node profile temperature $T_{\text{profile,U}}(t)$ is constrained to the desired usage temperature T_{usage} and the lower node profile temperature $T_{\text{profile,L}}(t)$ is constrained to the minimum *Legionella* prevention temperature $T_{\text{Legionella}}$ to ensure that the whole tank is sterilised. Both nodes are constrained to the minimum EWH temperature T_{min}.

$$T_{\text{profile,U}}(t) = \left\{ \begin{array}{ll} T_{\text{usage}} & \text{when intentional draw causes } Q_{\text{draw}}(t) > 0 \\ T_{\text{min}} & \text{otherwise} \end{array} \right\} \tag{4.59}$$

$$T_{\text{profile,L}}(t) = \left\{ \begin{array}{ll} T_{\text{Legionella}} & \text{once per day to prevent } \textit{Legionella} \text{ growth} \\ T_{\text{min}} & \text{otherwise} \end{array} \right\} \tag{4.60}$$

Cold Events: Cold events can occur as a result of a poor control strategy as well as the unreasonable expectation from a user. A water usage is "unreasonable" if too much water is drawn from the tank during an event for it to be possible to prevent the outlet temperature from falling below T_{usage} during water usage. The desired usage temperature T_{usage} is modified to a lower desired usage temperature T_{cold} during the event to compensate for the occurrence of a definite cold event. A cold event contributes to the negative comfort of users if, for example, they draw all the hot water from the tank and then expect more hot water shortly thereafter.

4.4.2.10 Electricity supply constraints

The constraint represented by the scheduled supply-side interruption of the electricity supply to the EWH is represented by the following time-varying input constraint

$$P_{\text{elec}}(t) \leq P_{\text{max}}(t) \tag{4.61}$$

The maximum power input is set to P_{rated} when the supply-side electricity is available, and to zero when the supply-side electricity is interrupted, with

$$P_{\text{max}}(t) = \left\{ \begin{array}{cl} 0 & \text{when the electricity is interrupted} \\ P_{\text{rated}} & \text{when the electricity is available} \end{array} \right\} \tag{4.62}$$

4.4.2.11 Constructing the temperature profile constraint

We construct the temperature profile constraint $T_{\text{profile}}(t)$ differently for temperature matching, energy matching, and energy matching with *Legionella* prevention. In constructing the constraint, we take into account "unreasonable" hot water usage profiles where it is impossible to prevent cold events from occurring, even if the heating element is switched on permanently. The temperature profile constraint is firstly constructed for the one-node model and then modified for the two-node model.

One-node construction:

Optimal temperature matching (TM): We construct the temperature profile constraint for temperature matching so that the required EWH temperature at the start of each usage event matches the EWH temperature achieved by thermostat control at the start of the same event.

Optimal energy matching (EM): We construct the temperature profile constraint for energy matching so that the EWH temperature remains above 40 °C for the duration of each usage event. However, we increase the outlet flow rate so that the energy in the volume of hot water delivered at 40 °C matches the energy in the volume of hot water delivered

at the thermostat control temperature.

Optimal energy matching with Legionella *prevention (EML):* We construct the temperature profile constraint for energy matching with *Legionella* prevention in exactly the same way as for energy matching (EM), except that once per day we increase the EWH temperature to 60 °C for 11 minutes to prevent the growth of *Legionella*. We schedule this to be just before the largest usage event for the day.

"Unreasonable" hot water usage profiles: To accommodate these profiles, we run a one-node forward simulation for the entire usage pattern, assuming that the heating element is always switched on, to determine cold events. Once a cold event is determined, the temperature profile constraint is modified to be the minimum temperature reached during the water usage.

Two-node construction:

Optimal temperature matching (TM): The constraint is modified onto the lower node temperature profile to ensure that the entire tank is at the required temperature. This prevents the possibility of the upper node temperature from dropping to the temperature of the lower node if all the water in the upper node is used during an event and the EWH transitions to a one-node state.

Optimal energy matching (EM): The constraint is modified onto the upper node temperature profile to ensure the upper node temperature remains above 40 °C for the duration of each usage event.

Optimal energy matching with Legionella *prevention (EML):* The constraint is modified onto the lower node temperature profile to ensure the temperature of the entire tank is sufficient to prevent *Legionella*.

"Unreasonable" hot water usage profiles: The two-node EWH model is used when running the forward simulation for the usage pattern. Once a cold event is determined, the upper node temperature profile is modified to the minimum upper node temperature reached during the water usage.

4.5 EWH feedback

EWH feedback is an important component to achieve optimisation and is required for correcting deviations which are expected to be caused by unanticipated water usages and model uncertainties. The following two sections respectively describe a temperature feedback controller and reactive water mixer which are an essential part of the optimal control system.

4.5.1 Temperature feedback control

The optimal temperature plan produced by the optimal planner is passed to the temperature feedback controller before determining the input of the EWH at any given time instance. The controller compares the measured temperature of the EWH with the desired

temperature set-point of the optimal plan. The controller will override the optimal input for the EWH so that the temperature of the EWH follows the optimal temperature (with hysteresis). The temperature feedback control corrects the EWH temperature when it deviates from the optimal plan. Temperature deviations are caused by unexpected water usages and model inaccuracies.

4.5.2 User water mixer

The user experiences the hot water temperature of the EWH when a usage event is intended. If the initial temperature experienced by user is not the desired temperature, the user will adjust the ratio of hot and cold water using a water mixer to reach such a desired temperature. The mixer will change the flow rate of the usage event to obtain the correct amount of energy desired by the user. The water mixer is a model of reality and is used to perform and evaluate simulation tests. This mechanism will also contribute to the *energy matching* heating control strategies where, if the optimal temperature is different to the thermostat set-point temperature for a usage event, the adjustment of the hot water flow rate will provide a matched energy usage from the EWH.

4.6 Summary

This chapter begins with a literature study on existing EWH models, different types of control strategies and optimisation techniques used in related work. The heating control strategies that are considered in this thesis are then defined as TC (baseline strategy), TM, EM and EML (optimal control strategies). Following this, the thermodynamics of the one- and two-node EWH model are derived. The derived EWH models and defined heating control strategies are used to aid with the formulation of the optimal control problem. The optimal control problem is solved using the two optimisation algorithms developed in Chapters 5 and 6.

Chapter 5

Dynamic programming optimisation

In this chapter, the optimal control problem formulated in Chapter 4 is solved using a dynamic programming (DP) approach that uses a one-node model of the EWH. DP is a search algorithm that models an optimal control problem as a multi-stage decision process. It uses the principle of optimality to find the optimal state trajectories and control inputs to minimise the cost function from all initial states. It uses the principle of optimality to drastically reduce the number of calculations required to determine the optimal control law. For systems with a large number of dimensions, the memory requirements of the DP algorithm become prohibitive. This is called "the curse of dimensionality". Fortunately, the one-node EWH system can be modelled as a first-order system, which makes DP a very suitable approach to solving the optimal control problem.

Figure 5.1 shows the control signal inputs (the water usage profile, heating control strategy and one-node EWH model) which are used by the DP optimisation algorithm to determine the outputs (the optimal heating schedule and temperature planning). Section 5.1 discretises the optimal control problem for the one-node EWH into a form for DP to be applied. Section 5.2 describes how the DP algorithm is initialised, executed, and obtains the optimal state trajectory and control inputs. Finally, Section 5.3 describes challenges that were overcome during the development of the algorithm.

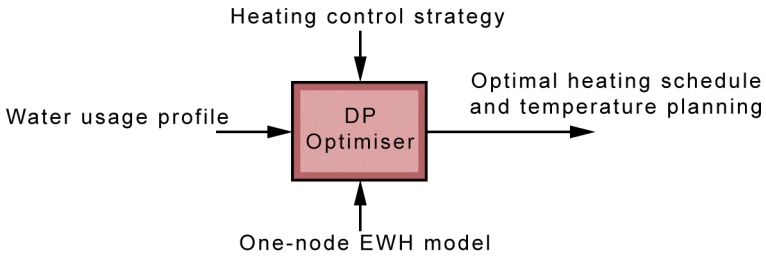


Figure 5.1: Control signal inputs and outputs of the DP optimiser described in this chapter.

The following sections describe the development of the DP algorithm that is used for optimising the heat schedule for an EWH using a similar approach performed by Engelbrecht (2016) [88] for flight envelope recovery.

5.1 Discretisation

To apply DP to an optimal control problem, the problem has to be discretised in time to represent different decision stages and discretised in state to represent a finite number of decisions to be made at each decision stage. The DP algorithm starts at the terminal stage and works backward in time through intermediate stages until it finds the optimal admissible path from the initial state to a terminal state.

5.1.1 Discrete-time dynamic model

The continuous-time differential equations describing the system dynamics are discretised to produce discrete-time difference equations that describe the state transition from one discrete time instant to the next.

$$E(k+1) \quad = \quad E(k) + [P_{\text{elec}}(k) - P_{\text{usage}}(k) - P_{\text{losses}}(k)]\Delta t \tag{5.1}$$

where Δt is the sampling period of the discrete time step.

5.1.2 Quantised state vector array

The continuous set of admissible states $\mathcal{X}$ is discretised to create a finite set of state values for the DP algorithm. We therefore create an array $\mathbf{X}_q$ of quantised state vector values $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n$ that span the set of admissible states $\mathcal{X}$ where subscript M is the last admissible state.

$$\mathbf{X}_q \quad = \quad \{E_1, E_2, \ldots, E_i, \ldots, E_M\} \tag{5.2}$$

5.1.3 Control inputs

The continuous set of admissible inputs $\mathcal{U}$ must be discretised to create a finite set of input values. We therefore create an input array $\mathbf{U}_q$ of quantised input vector values $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_n$ that span the set of admissible inputs $\mathcal{U}$.

$$\mathbf{U}_q \quad = \quad \{0, P_{\text{rated}}\} \tag{5.3}$$

5.1.4 Incremental state transition cost function

The continuous-time cost function is discretised by expressing the total path cost J_{ij} from the current state $\mathbf{x}_i(k)$ via the next state $\mathbf{x}_j(k+1)$ as the sum of the incremental cost ΔJ_{ij} of transitioning from the current state to the next state and the total path cost of the next state J_j to a final state.

$$J_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \Delta J_{ij}(\mathbf{x}_i, \mathbf{x}_j) + J_j(\mathbf{x}_j) \tag{5.4}$$

The incremental state transition cost ΔJ_{ij} is obtained by discretising the state transition cost function $g(\mathbf{x}(t), \mathbf{u}(t), t)$ as follows

$$\Delta J_{ij}(\mathbf{x}_i(k), \mathbf{x}_j(k+1)) \approx g(\mathbf{x}_i(k), \mathbf{u}_{ij}(k), k)\Delta t \tag{5.5}$$

where $\mathbf{u}_{ij}$ is the admissible control input that transitions the system from state $\mathbf{x}_i$ at time k to state $\mathbf{x}_j$ at time $k+1$. The incremental transition cost ΔJ is therefore obtained as

$$\Delta J_{ij}(\mathbf{x}_i(k), \mathbf{x}_j(k+1)) = P_{\text{elec}}(k)\Delta t \tag{5.6}$$

Figure 5.2 shows a table with the assignments of the state transition cost $g(\mathbf{x}_i(k))$ for each control input from the next state $j = x_2$ at time $t = 2$ to the current state $i = x_2$ and $i = x_1$. The path indicated in green represents the total path cost J_j for the considered scenario. The green blocks show the total path cost J_j that determined the next state being considered. The figure also shows that time moves forward from left to right and the algorithm execution goes in the opposite direction.

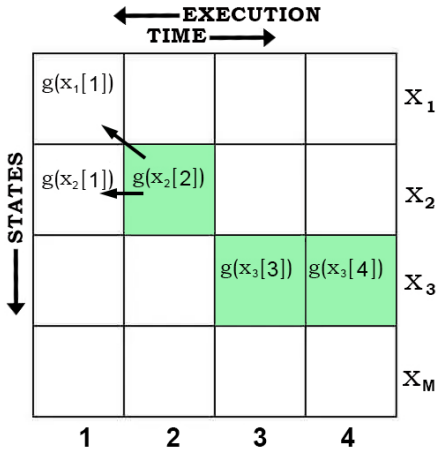


Figure 5.2: State transition cost assignments for $g(\mathbf{x}_i(k))$ and corresponding total path cost J_j (green).

5.1.5 Terminal state

The terminal state cost is the cost of the final state that all paths originate from and is defined as follows:

$$J_{\text{j}}(\mathbf{x}_{\text{j}}[N]) = \left\{ \begin{array}{ll} 0 & \text{if } \mathbf{x}_{\text{j}} \geq E_{\text{end}} \\ \infty & \text{otherwise} \end{array} \right\} \tag{5.7}$$

where N is the last time instance, and E_{end} indicates the energy in the tank that corresponds to the terminal boundary temperature T_{end}. This means that no optimal path produced by the algorithm will terminate below the desired terminal temperature.

5.1.6 Optimal path

The optimal state trajectory path is represented by $\mathbf{x}_\pi$ where subscript π refers to the path that minimises the electrical energy from the time instance $t = k$ to the final time sample $t = N$. It is obtained as follows:

$$\mathbf{x}_\pi = \mathbf{x}_{\text{i}_0}[k] \tag{5.8}$$

where i_0 is the initial state for the optimal path. This is the outcome of the algorithm with which the corresponding optimal input sequence $\mathbf{u}_\pi$ can also be found. The algorithm is presented in the next section which describes how the optimal state trajectory and control input sequence is obtained.

5.2 Algorithm

5.2.1 Initialisation

1. Create tables to store the quantised states $\mathbf{X}_q$, and the optimal path cost $\mathbf{J}^*$ from a given state to a final state, the optimal next state index $\mathbf{j}^*$, and the optimal control input $\mathbf{U}^*$, as follows:

$$\begin{aligned} \mathbf{X}_q &= \{E_{ik}\} && (5.9)\\ \mathbf{J}^* &= \{{J^*}_{ik}\} && (5.10)\\ \mathbf{j}^* &= \{{j^*}_{ik}\} && (5.11)\\ \mathbf{U}^* &= \{P^*_{\text{elec}ik}\} && (5.12) \end{aligned}$$

   where i is the quantisation index of the current state j is the quantisation index of the next state, and k is the index of the current time instant.

2. Populate the final column of the optimal cost table with zeroes to assign termination costs of zero to all final states.

$${J^*}_{iN} \leftarrow 0 \quad \forall i \tag{5.13}$$

3. Populate the rest of the optimal cost table with infinity values so that if a finite cost is calculated for a state, it will be lower than infinity and will replace the initial cost as the new lowest path cost.

$${J^*}_{ik} \leftarrow \infty \quad \forall i \quad \forall k \in [1, N-1] \tag{5.14}$$

4. Populate the final column of the optimal next state table so that the optimal next state (index j^*) for all final states is the state itself (index i).

$${j^*}_{iN} \leftarrow i \quad \forall i \tag{5.15}$$

5.2.2 Execution

1. Set the index j of the next state $\mathbf{x}_j(k)$ to the first index of the state vector array

$$j \leftarrow 1 \tag{5.16}$$

2. For the given index j of the next state $\mathbf{x}_j(k+1)$, calculate the previous state $\mathbf{x}(k)$ for all admissible values of the control input $\mathbf{u}(k)$.

   For each

$$\begin{aligned} P_{\text{elec}}(k) &\in \{0, P_{\text{rated}}\} && (5.17)\\ \text{and } P_{\text{elec}}(k) &\leq P_{\text{max}}(k) && (5.18) \end{aligned}$$

calculate

$$E(k) = E(k+1) - [P_{\text{elec}}(k) - P_{\text{usage}}(k) - P_{\text{losses}}(k)]\Delta t \tag{5.19}$$

with

$$P_{\text{usage}}(k) = c\rho Q_{\text{usage}}(k)[T(k) - T_{\text{in}}(k)] \tag{5.20}$$

$$P_{\text{losses}}(k) = \frac{1}{R_{TH}}[T(k) - T_{\text{ambient}}(k)] \tag{5.21}$$

where

$$T(k) = \frac{E(k)}{c\rho V_{\text{tank}}} \tag{5.22}$$

and the inlet water temperature $T_{\text{in}}(k)$ and the ambient temperature $T_{\text{ambient}}(k)$ are known external disturbances at time instant k.

3. Check whether the previous state $\mathbf{x}(k)$ is an admissible state.

$$E(k) \geq c\rho V_{\text{tank}} T_{\text{profile}}(k) \tag{5.23}$$

If the previous state is an admissible state, continue to step 4. Otherwise, continue iterating through the possible values of the control input.

4. Find the index i of the quantised previous state $\mathbf{x}_i(k)$ in the state vector array that is closest to the calculated previous state $\mathbf{x}(k)$.

5. Calculate the cost from the current state $\mathbf{x}_i(k)$ to a final state through this next state $\mathbf{x}_j(k+1)$

$$J_i \leftarrow \Delta J_{ij} + J*_j \tag{5.24}$$

where

$$\Delta J_{ij} = P_{\text{elec}}(k)\Delta t \tag{5.25}$$

6. If the new cost is lower than the lowest cost so far, it becomes the new lowest cost. Also, the control input $\mathbf{u}_i(k)$ and the next state index j become the new best control input $\mathbf{u}_i^*(k)$ and the new best next state j_i^* for the current state $\mathbf{x}_i(k)$. In other words, if $J_i < J_i^*$ then

$$\begin{aligned} J_i^* &\leftarrow J_i \\ \mathbf{u}_i^* &\leftarrow \mathbf{u}_i(k) \\ j_i^* &\leftarrow j \end{aligned}$$

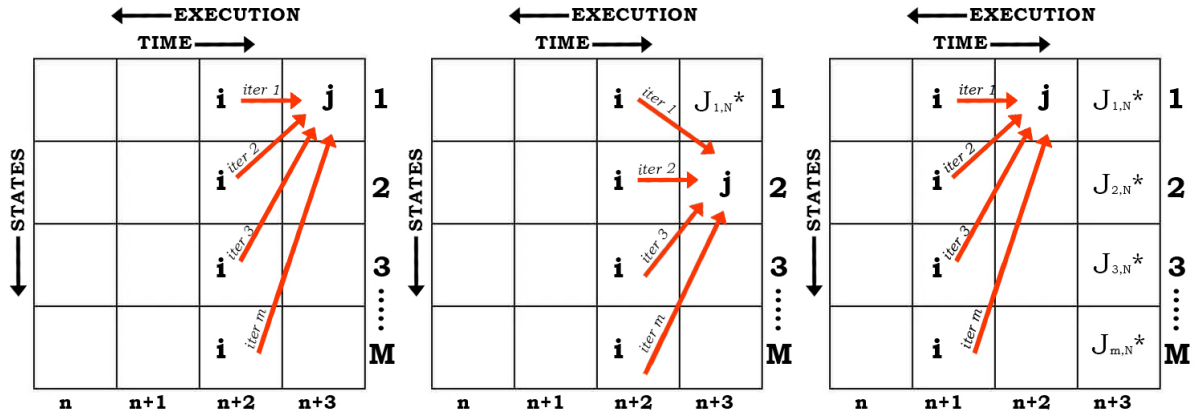
7. If the next state index j was not the last index, increment the next state index j and return to step 2.

$$j \leftarrow j + 1 \tag{5.26}$$

Otherwise, if the next state index j was the last index, and the time instant was not the first time instant, then step one time instant k backward and return to step 1.

$$k \leftarrow k - 1 \tag{5.27}$$

Or else, terminate the execution. Figure 5.3 shows the order of iterations for populating the various tables which occurs when the algorithm is executed. Each arrow is labelled with the iteration index for assigning values for each considered current state i to the next state j.



(a) $j = 1$ and $k = n + 3$ **(b)** $j = 2$ and $k = n + 3$ **(c)** $j = 1$ and $k = n + 2$

Figure 5.3: Iteration order for populating the look-up table. The red arrows represent different iterations of current state i.

5.2.3 Lookup table navigation

The DP algorithm produces a look-up table of the optimal state trajectories and optimal control sequences for EWH control from all time instants and admissible initial states. Given a time instant k and an initial state $\mathbf{x}_i(k)$, the optimal state trajectory $\{\mathbf{x}_\pi(k) : k = k, k+1, \ldots, N\}$ and the optimal control sequence $\{\mathbf{u}_\pi(k) : k = k, k+1, \ldots, N\}$ can be obtained by starting at the column index k and the row index i of the initial state in the lookup table and iteratively navigating through the lookup table by following the optimal next state indexes j^*. Figure 5.4 shows how the lookup table selects the optimal state trajectory $\mathbf{x}_\pi$ from an initial to final state and how the corresponding optimal input sequence $\mathbf{u}_\pi$ is produced.

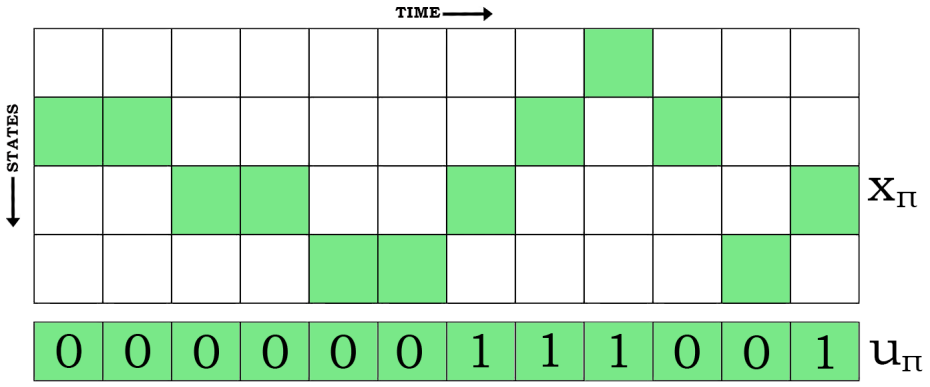


Figure 5.4: Look-up table for the optimal state trajectory sequence $\mathbf{x}_\pi$ (represented by the green blocks) and the path is determined by navigating the optimal next state table $\mathbf{j}^*$. Below the table shows how the corresponding optimal input sequence $\mathbf{u}_\pi$ is also similarly obtained from the optimal control input table $\mathbf{U}^*$. A 0 refers to $u = 0$ and a 1 refers to $u = P_{\text{rated}}$ for the duration of that time instance.

Although the lookup table produced from DP can find the optimal path from any desired state and time instance to a terminal state, only one optimal path is necessary from this

table. This is the path that extends from the desired initial boundary condition to the terminal boundary condition. By picking the initial state to be equivalent to the energy in the tank that corresponds to the desired initial temperature, T_{start}, the optimal path can be obtained to the desired terminal temperature, T_{end}. Since the lowest admissible terminal state was designed to correspond to T_{end}, the optimal path will end at this state to minimise the cost of the path. If it is impossible for the optimal path to terminate at this state, the algorithm will end at the lowest admissible terminal state possible. The lookup table is discarded because this desired optimal path is used as a reference temperature for which the temperature feedback controller is implemented to ensure that the reference temperature is followed by the measured temperature.

5.2.4 Further algorithm optimisation

DP is time consuming. Reducing algorithm execution times, where possible, is valuable. For each possible control input, either on or off, there is only one output. This means that there are only two current states that can branch from the previous time sample to the considered next state j. A significant reduction in iterations was made when the next state would first calculate which two current states were admissible, then it would only consider those two iterations for the current states that are iterated through for the next state j. Figure 5.5 shows that for the considered iteration of the next state j, the number of current state iterations for i was reduced to two, significantly reducing algorithm executions times.

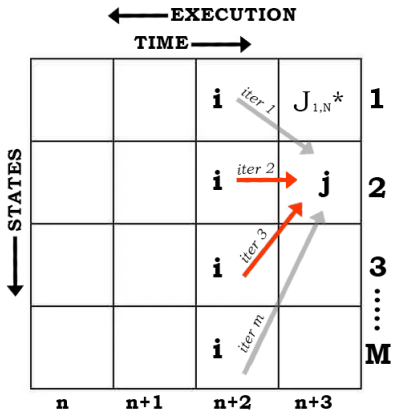


Figure 5.5: Diagram showing how the number of iterations for each next state j is greatly reduced.

Further time was reduced for the algorithm execution by reducing the number of iterations for next state j that are performed. This was achieved by imposing a constraint on the algorithm that only allows for an iteration of next state j if that state satisfied the constraints at that time sample.

5.3 Challenges

Two significant challenges were overcome in the design of the optimisation algorithm. The first was determining whether a cold event was expected to occur for the optimal

control planning and how the constraints would be modified to deal with it. The second was encountered when determining the optimal energy matching for the corresponding heating control strategy. The solutions were designed during the development of the DP algorithm, but were similarly implemented for the A* algorithm.

5.3.1 Cold event challenges

The constraints imposed on the optimal control problem imply that the energy in the tank must be within the maximum and minimum bounds defined for the tank for the instance of a considered optimal path to be admissible. Additionally, a path can also only be admissible if the outlet temperature does not fall below the desired usage temperature T_{usage} when water is drawn to prevent cold events. A problem arises when a cold event cannot be prevented, even when the heating element is "always on" over a period of time. This means that an "unreasonable" volume of water is drawn from the tank such that no admissible paths will exist for the water event with the currently defined constraints. This is solved by performing a forward simulation before the DP algorithm executes on all the water events separately with the energy of the tank set to E_{max} at the first-time instance. The heating element is set to "always on" for the whole simulation. If the outlet temperature falls below T_{usage} at any time instance, the constraint for T_{usage} at that instance is changed to the cold event temperature reached T_{cold} at that instance.

Even with this solution, there was still another challenge to face. Water events define their start and stop times depending on when water flow remains zero for five minutes before and after the event. This means that a forward simulation of individual events will not necessarily detect cold events that will occur due to the state of the tank after the previous event. The previous event could lower the tank to an energy level that cannot possibly reach Emax for the start of the next event's forward simulation. For profiles where this might occur, events are further categorised into *batch* events for forward simulations. The start and stop times depend on when water flow remains zero for the number of time samples calculated as follows:

$$n_b = \frac{E_{\text{max}} - E_{\text{min}}}{E_{\text{elec}}} \tag{5.28}$$

Where n_{b} is the minimum number of time samples required between batch events to ensure the forward simulation will be guaranteed to start at E_{max} if it is necessary. Figure 5.6 shows how batch events are categorised for the water profiles.

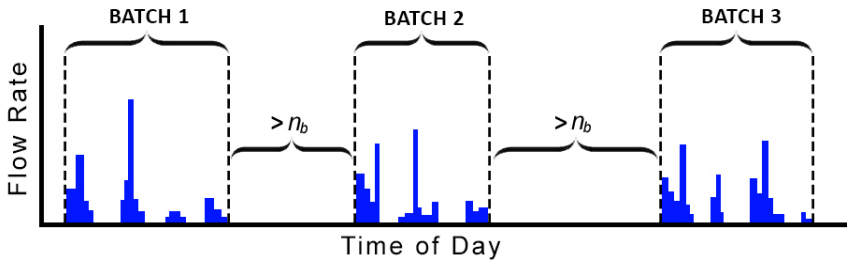


Figure 5.6: Diagram showing how water events (blue) are categorised into batch events and showing time intervals larger than n_b.

5.3.2 Energy matching challenges

Another challenge faced was developing the algorithm to generate the optimal planning for energy matching heating control strategies. When the EWH is simulated with the generated optimal plan, the user is simulated to adjust the mixing of hot and cold water in reaction to the water temperature that is initially experienced. This means that the outlet flow rate of hot water will change to ensure that an equivalent amount of thermal energy is delivered during the event to that of a simulation that uses TC for the same event. The optimisation algorithm is therefore modified to account for the change in flow rate that is expected as a results of the water mixer.

Despite the lower temperatures during water events for energy matching heating control strategy to that of TC, an equivalent delivery of energy is achieved by adjusting the flow rate during the water event to increase the total volume of water drawn. This is explained by the following equation that shows the relationship between thermal energy drawn from the tank, the outlet water temperature and volume:

$$E_{\text{draw}} = c\rho T_{\text{outlet}} V_{\text{draw}} \tag{5.29}$$

Before the DP algorithm executes, the thermal energy delivered by each event during a simulation controlled by TC is determined for the water profile. The energy delivered for event n is represented by $E_{\text{TC,n}}$. Energy matching is achieved by first tracking the thermal energy delivered for event n with optimal control (OC) and is represented by $E_{\text{OC,n}}$. The energy difference between that delivered by TC and OC is determined as follows:

$$\Delta E_{\text{n}} = E_{\text{TC,n}} - E_{\text{OC,n}} \tag{5.30}$$

where ΔE_{n} is a positive quantity if the water delivered was at a lower temperature for OC. By increasing the flow rate of the event for OC, and therefore delivering a larger volume of water, the energy delivery can match that of TC and the additional volume of water is calculated as follows:

$$V_{\text{add,n}} = \frac{\Delta E_{\text{n}}}{c\rho T_{\text{outlet}}} \tag{5.31}$$

where $V_{\text{add,n}}$ is the volume of additional water used for event n to achieve energy matching and T_{outlet} is the outlet water temperature. However, during the DP algorithm execution there are multiple paths that cross the event at different temperatures simultaneously. This means that for each admissible path, it is required to track its own thermal energy delivery and determine $V_{\text{add,n}}$ to satisfy the paths unique ΔE_{n}. This is solved by creating two additional tables during the algorithm initialisation as follows:

$$\mathbf{V}^*_{\text{add}} = \{V^*_{\text{add}ik}\} \tag{5.32}$$
$$\mathbf{\Delta E}^* = \{\Delta E^*{}_{ik}\} \tag{5.33}$$

where $\mathbf{V}^*_{\text{add}}$ stores the optimal volume of additional water for current state i at time instance k and $\mathbf{\Delta E}^*$ stores the optimal thermal energy delivery difference from that of TC. The entries $\Delta E^*{}_{ik}$ are used by Equations 5.31 to determine the entries of $V^*_{\text{add}ik}$ to achieve energy matching for any admissible optimal path.

Figure 5.7 shows a diagram that conveys how energy matching is achieved in the optimisation algorithm by increasing the flow rate based on how low the usage temperatures are from the set-point temperature. The figure shows a water profile (bar graph) where time is represented by the x-axis and the states of the EWH are represented by the y-axis (the height of the water usages shown indicates the flow rate which is not scaled by the y-axis). The black arrow shows the state trajectory that TC would expect to follow at approximately the set-point temperature. The blue portions of the water profile represents the original water profile. The red and green arrows represent two different optimal state trajectories that correspond to higher and lower temperatures, respectively. The temperatures of the red arrow are shown in between the set-point temperature and the green arrow trajectory. When energy matching is desired, additional water is drawn (shown by the red portion of the bar graph). The green arrow has the lowest temperature trajectory and even more additional water is drawn (shown by the combination of the red and green portions of the bar graph).

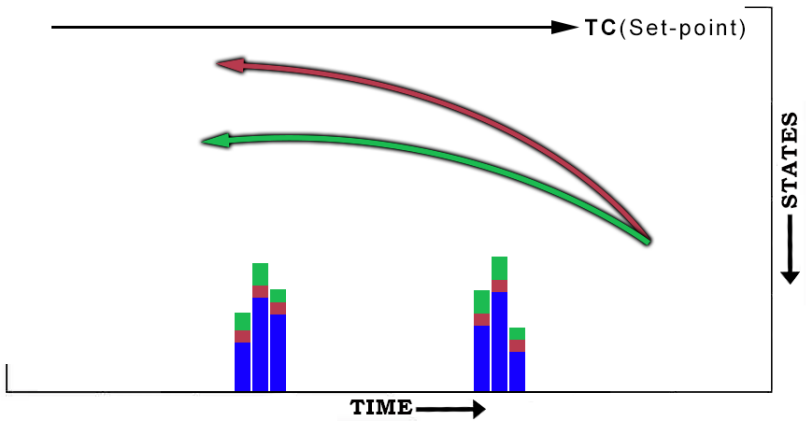


Figure 5.7: Diagram of a water profile where time is represented by the x-axis and the states of the EWH are represented by the y-axis. Water usages are shown by the bar graph and the height of each usage represents the flow rate. Two optimal state trajectories are shown by the red and green arrows. For energy matching, decreasing the temperatures along the trajectory during water usages will increase the flow rate. The black arrow represents the TC state trajectory.

5.4 Summary

The DP algorithm described in this chapter is used to produce the optimal heating schedule and optimal temperature trajectory for a given water profile. In Chapter 7, a EWH simulator uses these to control the heating element and determine the results of the optimal schedule control of a one-node EWH for the corresponding water profile.

Chapter 6

A* optimisation

In this chapter, the optimal control problem formulated in Chapter 4 is solved using an A* approach that uses a two-node model of the EWH. A* is a *shortest path* search algorithm that can be used to model an optimal control problem as a node-based data structure navigation process to find the optimal state trajectory and control inputs to minimise the cost function from a single initial state to a destination. The algorithm optimises its search time by introducing a heuristic component that estimates the path to a terminal state. However, the efficiency of the algorithm depends on the quality of the heuristic function chosen. In Chapter 5, the optimal state trajectory and control inputs were produced for only a first-order system one-node EWH model using DP. This chapter explores optimisation using the two-node model which is not a first-order system and makes A* more applicable. The A* algorithm builds a binary search tree with a size determined by the number of possible actions at each node of the tree. This approach is developed to handle both the one- and two-node EWH models because a two-node EWH transitions between the models when certain conditions are met.

Figure 6.1 shows the control signal inputs (the water usage profile, heating control strategy and two-node EWH model) which are used by the A* optimisation algorithm to determine the outputs (the optimal heating schedule and temperature planning). Similarly to the previous chapter, Section 6.1 discretises the optimal control problem for the two-node EWH into a form for A* to be applied. Following this, Section 6.2 describes how the A* algorithm is initialised, executed, and obtains the optimal state trajectory and control inputs.

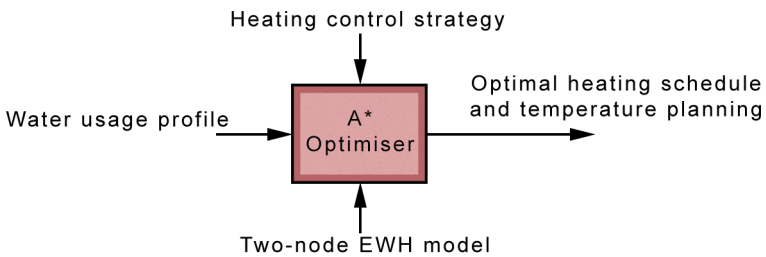


Figure 6.1: Control signal inputs and outputs of the A* optimiser described in this chapter.

The following sections follow a similar approach to that of the DP algorithm presented in the previous chapter.

6.1 Discretisation

To apply A* to an optimal control problem, the problem must be discretised in time to represent different decision stages, and discretised in state to represent a finite number of decisions to be made at each decision stage. The A* algorithm finds the optimal path by starting at the initial stage and then works through intermediate stages until it finds the *first* admissible path from the initial state to a desired terminal state. The first admissible path is also the optimal path due to the way that the paths are sorted in a priority queue.

6.1.1 Discrete-time dynamic model

The continuous-time differential equations describing the system dynamics are discretised to produce discrete-time difference equations that describe the state transition from one discrete time instant to the next. For the one-node case, the state transition is described as

$$E(k+1) = E(k) + [P_{\text{elec}}(k) - P_{\text{usage}}(k) - P_{\text{losses}}(k)]\Delta t \tag{6.1}$$

where Δt is the sampling period of the discrete time step. For the two-node case, the state transition is described as

$$E(k+1) = E(k) + [\dot{E}_{\text{U}}(k) + \dot{E}_{\text{L}}(k)]\Delta t \tag{6.2}$$

where $\dot{E}_{\text{U}}(k)$ and $\dot{E}_{\text{L}}(k)$ are given, respectively.

$$\dot{E}_{\text{U}}(k) = -P_{\text{usage,U}}(k) - P_{\text{losses,U}}(k) - P_{\text{trans,U}}(k) \tag{6.3}$$
$$\dot{E}_{\text{L}}(k) = P_{\text{elec}}(k) - P_{\text{usage,L}}(k) - P_{\text{losses,L}}(k) - P_{\text{trans,L}}(k) \tag{6.4}$$

and the discretised volume differential equations are

$$V_{\text{tank,U}}(k+1) = V_{\text{tank,U}}(k) - Q_{\text{draw}}(k)\Delta t \tag{6.5}$$
$$V_{\text{tank,L}}(k+1) = V_{\text{tank,L}}(k) + Q_{\text{draw}}(k)\Delta t \tag{6.6}$$

where Q_{draw} is the hot water outlet volumetric flow rate. The quantised state vector array, control inputs and incremental state transition cost function are discretised similarly to that of DP, and can be referred to in Section 5.1.2, 5.1.3 and 5.1.4, respectively.

6.1.2 Volume ratio

To reduce the number of variables that will be tracked by the A* algorithm, the volume ratio of the tank is calculated, expressed as the ratio of the volume of the upper node and the tank, as follows:

$$V_{\text{r}}(k) = \frac{V_{\text{tank,U}}(k)}{V_{\text{tank}}} \tag{6.7}$$

Since the volume of the tank is a constant, the volume of the upper and lower node can be obtained by this volume ratio.

6.1.3 Initial state

The initial state is specified at the first time sample $t = 1$ and is expressed as follows:

$$\mathbf{x}_{\mathrm{s}} = E_{\mathrm{start}} \tag{6.8}$$

where s indicates the starting state of the algorithm and E_{start} is equivalent to the energy of the tank at the desired starting temperature T_{start}. The EWH is also assumed to be in a one-node state at the first time sample.

6.1.4 Goal state

The goal of the algorithm is to find a path from the initial state $\mathbf{x}_{\mathrm{s}}$ to a goal state. The goal state is specified at the last time sample $t = N$ and is expressed as follows:

$$\mathbf{x}_{\mathrm{f}} = E_{\mathrm{end}} \tag{6.9}$$

where f indicates the goal state of the algorithm and E_{end} is equivalent to the energy of the tank at the desired final temperature T_{end}. The EWH is also assumed to be in a one-node state at the last time sample.

6.1.5 EWH model transition

The EWH current model mode is represented by

$$n_{\mathrm{s}} = \left\{ \begin{array}{ll} 1 & \text{if one-node EWH mode} \\ 2 & \text{if two-node EWH mode} \end{array} \right\} \tag{6.10}$$

and transitions between one- and two-nodes depending on when certain conditions are met. The EWH mode transitions can be represented as a state machine that begins in a one-node mode. Figure 6.2 shows a diagram of the state machine and it is used to aid with explaining the following scenarios where the EWH model transitions between the modes.

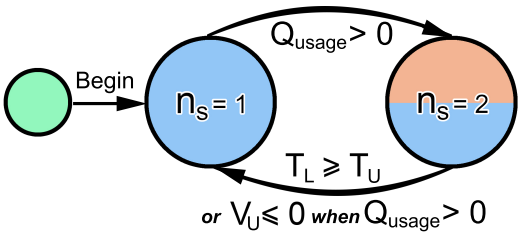


Figure 6.2: State machine diagram of the EWH mode transitions.

When the EWH model is in the one-node mode, it will transition to the two-node mode when the outlet flow rate is greater than zero:

$$Q_{\mathrm{usage}} > 0 \tag{6.11}$$

When the single node splits into two nodes, the lower node is assigned the same temperature as the inlet temperature, the volume of the upper node is reduced by the usage volume, and the volume of the lower node is increased by the usage volume:

$$T_L = T_{\text{in}} \tag{6.12}$$
$$V_U = V_{\text{tank}} - V_{\text{usage}} \tag{6.13}$$
$$V_L = V_{\text{usage}} \tag{6.14}$$

When the EWH model is in two-node mode, it will transition back to the one-node mode, if one of two conditions are met:

1. If the temperature of the lower node equals or would exceed the temperature of the lower node, or
2. The volume of the upper node is reduced to zero while the usage flow rate is greater than zero.

If the temperature of the lower node reaches the temperature of the upper node:

$$T_L \geq T_U \tag{6.15}$$

then the upper and lower nodes are merged into a single node represented by the upper node: (The temperature of the lower node assigned the temperature of the upper node, the volume of the lower node is set to zero, and the volume of the upper node is set to the volume of the tank.)

$$T_L = T_U \tag{6.16}$$
$$V_U = V_{\text{tank}} \tag{6.17}$$
$$V_L = 0 \tag{6.18}$$

If the volume of the upper node is reduced to zero:

$$V_U \leq 0 \quad \textit{when } Q_{\text{usage}} > 0 \tag{6.19}$$

then the upper and lower nodes are merged into a single node represented by the lower node: (The temperature of the upper node assigned the temperature of the lower node, the volume of the upper node is set to zero, and the volume of the lower node is set to the volume of the tank.)

$$T_U = T_L \tag{6.20}$$
$$V_U = 0 \tag{6.21}$$
$$V_L = V_{\text{tank}} \tag{6.22}$$

6.1.5.1 Heuristic search

In Section 5.1.4, the continuous-time cost function was discretised to be expressed as the sum of the incremental cost ΔJ_{ij} of transitioning from the current state to the next state and the total path cost of an initial state to the current state J_i, and was obtained as

$$J_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \Delta J_{ij}(\mathbf{x}_i, \mathbf{x}_j) + J_i(\mathbf{x}_i) \tag{6.23}$$

A heuristic cost function is introduced to the total path cost J_{ij} to accelerate the search algorithm by prioritising the most optimal path as the next iteration of the algorithm execution. This is accomplished by heuristics: estimating the path cost from the next state $\mathbf{x}_j(k+1)$ to the terminal state. Due to the setup of the optimal control problem as well as its lower computational complexity, the distance metric used to determine the heuristic cost is the Manhattan Distance. This metric considers the *horizontal* distance, or time, and the *vertical* distance, or energy state, from the next state to the terminal state at the last time sample. Equation 6.23 is modified as follows

$$J_{ij}^{\text{arr}}(\mathbf{x}_i, \mathbf{x}_j) \approx \Delta J_{ij}(\mathbf{x}_i(k), \mathbf{x}_j(k+1)) + J_i(\mathbf{x}_i(k)) \tag{6.24}$$

where J_{ij}^{arr} is the *cost-to-arrive* function and the total path cost is modified to introduce a heuristic cost and is expressed as

$$J_{ij}(\mathbf{x}_i, \mathbf{x}_j) = J_{ij}^{\text{arr}}(\mathbf{x}_i, \mathbf{x}_j) + J_{jf}^{\text{go}}(\mathbf{x}_j, \mathbf{x}_f) \tag{6.25}$$

where f refers to the goal state of the optimal state trajectory and $J_{ij}^{\text{go}}(\mathbf{x}_i, \mathbf{x}_j)$ is the *cost-to-go*.

6.1.5.2 Cost-to-go

At any time instance, the EWH must reach the terminal state after the result of thermal energy that is anticipated to still leave the tank. The *cost-to-go* estimates both the minimum amount of energy that must still be supplied to the tank to reach the goal state as well as how much thermal energy will leave the tank during the remaining water usages from the considered time instance.

Figure 6.3 shows a water profile where time is represented by the x-axis and the states of the EWH are represented by the y-axis (the height of water usages shown in blue indicates the flow rate which is not scaled by the y-axis). The goal state is indicated by the black circle at the final time instance. The red, green and orange circles are positioned in different areas to convey how the *cost-to-go* differs at each position. The blue arrows indicate the thermal energy that will leave the tank as a results of each water usage (i.e. the horizontal component of the heuristic cost due to it's distribution along the x-axis) and the red arrows indicate the energy that must be supplied to reach the goal state from the current state of each coloured circle (i.e. the vertical component of the heuristic cost due to the state difference along the y-axis). The *cost-to-go* for the red circle is the largest of the three because it is expecting to use the most thermal energy as a result of the water usages that remain before reaching the final time instance. The *cost-to-go* of the orange circle is smaller than that of red because there are less water usages remaining. The *cost-to-go* of the green circle is the least of all three positions because it requires less energy to reach the goal state.

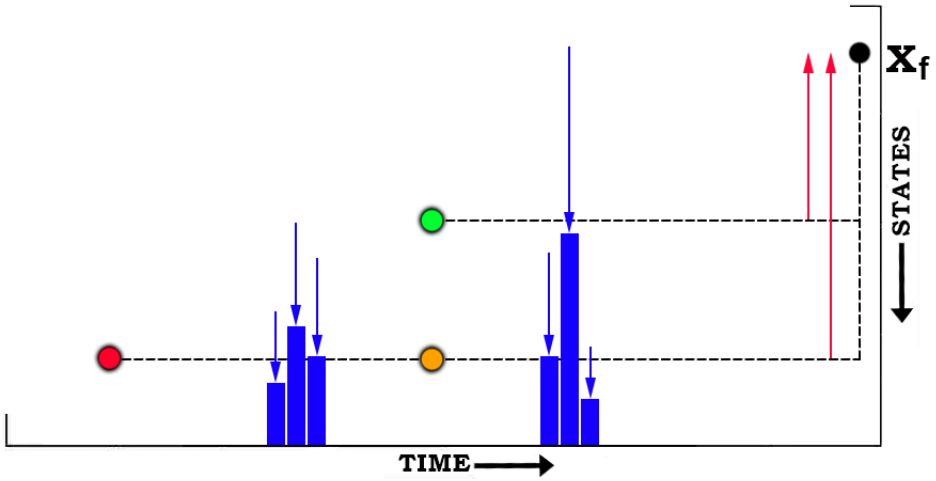


Figure 6.3: Diagram of a water profile where time is represented by the x-axis and the states of the EWH represented by the y-axis. Water usages are shown in blue and the height of each usage represents the flow-rate of the usage. The blue arrows indicate the horizontal component and the red arrows indicate the vertical component of the *cost-to-go* from each of the three different positions (red, orange and green circles) to the goal state (black circle).

The calculation to estimate $J_{jf}^{\text{go}}(\mathbf{x}_j, \mathbf{x}_f)$ from the next state $\mathbf{x}_j(k+1)$ to the final state f is

$$J_{jf}^{\text{go}}(\mathbf{x}_j, \mathbf{x}_f) \approx h(\mathbf{x}_j(k+1), \mathbf{x}_f(N), k+1, N)\Delta t \tag{6.26}$$

where N refers to the last time instance. The heuristics h is expressed as

$$h(\mathbf{x}_j(k+1), \mathbf{x}_f(N), k+1, N) = h_1(\mathbf{x}_j(k+1), \mathbf{x}_f(N)) + h_2(k+1, N) \tag{6.27}$$

where h_1 is the *vertical* component that is calculated by the state energy difference of the next state $\mathbf{x}_j(k+1)$ and the final state $\mathbf{x}_f(N)$ and is expressed by

$$h_1(\mathbf{x}_j(k+1), \mathbf{x}_f(N)) = |\mathbf{x}_j(k+1) - \mathbf{x}_f(N)| \tag{6.28}$$

and h_2 is the *horizontal* component that is calculated by an estimate of thermal energy that will still leave the tank from time sample $k+1$ to the final sample N. Due to the fact that standing losses only contribute to a portion of the thermal energy that leaves the tank, only a thermal energy decrease caused by water usages is considered and calculated by

$$h_2(k+1, N) = \sum_{n=k+1}^{N} E_{\text{usage,h}}[n] \tag{6.29}$$

where $E_{\text{usage,h}}$ is a pre-calculated list of estimated thermal energy that is drawn from the tank at each time instant and is expressed as

$$E_{\text{usage,h}} = [E_{\text{usage},1}, E_{\text{usage},2}, \ldots, E_{\text{usage}}, \ldots, E_{\text{usage},N}] \tag{6.30}$$

The list entries are calculated by a forward simulation of the water profile which acts as a disturbance to the EWH. The simulation is performed such that each water *event* ends with the outlet temperature remaining above T_{usage}. Both h_1 and h_2 provide a resulting estimate of thermal energy required to reach the goal state from the next state.

6.1.6 Binary search tree (BST)

This section describes how a BST data structure is used to aid the A* path searching algorithm. Although the navigation of the data structure in this thesis is modified from the traditional search process, more information on the BST data structure can be found in *Optimal Binary Search Trees* [100].

A significant difference between DP and A* implementations is the data structure that is initialised and executed for each approach. Whereas the DP approach generates a lookup table, the A* approach creates a BST comprised of internal nodes that are interconnected to one another by the next state j of the proceeding node to branch multiple search paths. The first series of nodes that reaches the terminal state node from the initial state node is also the optimal path. A single node object is represented by symbol $\mathcal{N}$ and is expressed as

$$\mathcal{N}_{ik} = \mathcal{N}_i(\mathbf{x}_\mathrm{i}[k]) \tag{6.31}$$

where $\mathcal{N}_{ik}$ is a node at current state i and time k. The node represents multiple *keys* of information that are required to track the necessary variables of the EWH at that instance of the path. A single node tracks the quantised state $\mathbf{x}_i$, the optimal total path cost $\mathbf{J}^*_{ik}$, the optimal next state $\mathbf{j}^*_{ik}$ as a result of $\mathbf{u} = 0$ and $\mathbf{u} = P_\mathrm{elec}$, the optimal upper node temperature $\mathbf{T}^*_{Uik}$, the optimal lower node temperature $\mathbf{T}^*_{Lik}$ and the optimal volume ratio $\mathbf{V}^*_{rik}$.

6.1.6.1 Initial state node

The initial state node is the first node initialised by the BST for the desired starting state s at the first-time sample $t = 1$ and is defined as follows:

$$\mathcal{N}_{s1} = \mathcal{N}_s(\mathbf{x}_\mathrm{s}[1]) \tag{6.32}$$

It is assigned the following optimal cost

$$\mathbf{J}^*_{s1} \quad = \quad 0 \tag{6.33}$$

This node is assigned a cost of zero to indicate that it is the origin of all succeeding paths that will branch from it. Assignments are only made for the optimal next state $\mathbf{j}^*_{s1}$ for each possible control input $\mathbf{u}_{s1}$ if an admissible node is reached as a result of the control input during the algorithm execution.

6.1.6.2 Terminal state node

A terminal state node is a unique modification of the predefined node object. For the goal state f at the last time samples $t = N$, the terminal node state is defined as follows:

$$\mathcal{N}_{fN} = \mathcal{N}_f(\mathbf{x}_\mathrm{f}[N]) \tag{6.34}$$

This node is unique as it only needs to save the goal state $\mathbf{x}_f$ and the desired final outlet temperature $\mathbf{T}^*_{UfN}$.

Figure 6.4 shows the execution of the A* BST. The green and red node (represented by circles) represent the initial and desired terminal state node, respectively. The nodes for

a single search path are shown (white circles) and the black arrows represent the next state for each interconnected node. The node at $t = 4$ determines the next state at $t = 5$ for each of the possible control inputs of u. For each possible input, the *cost-to-arrive* (green arrow) represents the cost from s to the considered node and the *horizontal* and *vertical* components of the *cost-to-go* (red arrows) represent the cost from the considered next state to f.

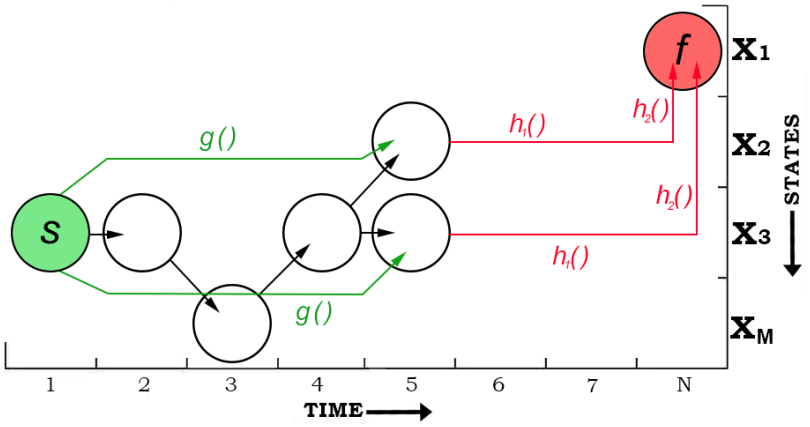


Figure 6.4: Diagram of A* search tree produced during the algorithm execution, showing cost assignments of cost-to-go (green arrows) and cost-to-arrive components (red arrows) to determine the cost of the considered next state nodes at $t = 5$ from node s to f.

6.1.6.3 Optimal path

The optimal state trajectory path is represented by $\mathbf{x}_\pi$ where subscript π refers to the path that minimises the electrical energy from the first-time sample $t = 1$ to the final time-sample $t = N$. It is obtained as follows:

$$\mathbf{x}_\pi = \mathbf{x}_\mathrm{s}[1] \tag{6.35}$$

where s is always the starting state for the optimal path. This is the outcome of the algorithm with which the corresponding optimal input sequence $\mathbf{u}_\pi$ can also be found. The algorithm is presented in Section 6.2 which describes how the optimal state trajectory and control input sequence is obtained.

6.1.7 Priority queue

Unlike the DP algorithm, the order of iterations for the algorithm execution does not move forwards or backwards in time, but instead jumps to the time instance depending on which search path has a higher *priority* than all the other search paths. A priority queue is defined as follows:

$$\mathcal{P} = \{\mathcal{N}_{ik}, \dots\} \tag{6.36}$$

where each entry is a node corresponding to any admissible search path ending. However, the queue organises the entries in ascending order based on each node's total path cost

J_{ik}. The queue guides the algorithm execution with the order of iterations for extending the search paths simultaneously. Each iteration begins by removing the first, and highest priority, node from the queue and is expressed as

$$\mathcal{P}^* = \mathcal{N}_{ik} \tag{6.37}$$

This node with the highest priority is used to determine if any admissible state is reachable for each possible control input, if that input were applied at the node's current state. This is the reason why the first search path to reach the terminal state node is also the optimal path.

6.2 Algorithm

6.2.1 Initialisation

1. Initialise a BST for the A* algorithm which will contain the initial state node, the terminal state node and all interconnecting nodes that search all admissible paths from getting from the starting state s to the final state f. Initialise node $\mathcal{N}_{s1}$ at starting state s and time $t = 1$ and assign the following

$$\mathbf{x}_{s1} = E_{s1} \tag{6.38}$$
$$\mathbf{J}^*_{s1} = 0 \tag{6.39}$$
$$\mathbf{j}^*_{s1} = \begin{cases} \infty, & \text{for } \mathbf{u} = 0. \\ \infty, & \text{for } \mathbf{u} = P_{\text{rated}}. \end{cases} \tag{6.40}$$

   where a value of ∞ for $\mathbf{j}^*_{s1}$ indicates that an admissible path has not been found to a node at the following time instance for the corresponding control input. The initial variables of the EWH are assigned for the starting node as follows

$$\mathbf{T}^*_{Ls1} = 0 \tag{6.41}$$
$$\mathbf{T}^*_{Us1} = T_{\text{start}} \tag{6.42}$$
$$\mathbf{V}^*_{rs1} = 1 \tag{6.43}$$

   where T_{start} is the desired starting outlet temperature and the EWH is initialised in a one-node model mode since $\mathbf{V}^*_{rs1} = 1$.

2. Initialise node $\mathcal{N}_{fN}$ at goal state f and time $t = N$ and assign the following

$$\mathbf{x}_i = E_{fN} \tag{6.44}$$
$$\mathbf{T}^*_{UfN} = T_{\text{finish}} \tag{6.45}$$

   where T_{end} is the desired final outlet temperature.

3. Initialise a priority queue with one entry corresponding to the starting node as follows:

$$\mathcal{P} = \{\mathcal{N}_{s1}\} \tag{6.46}$$

6.2.2 Execution

1. The highest prioritised node $\mathcal{N}_{ik}$ of the priority queue is removed from $\mathcal{P}$.

$$\mathcal{N}_{ik} \leftarrow \mathcal{P}^* \tag{6.47}$$

and the optimal total cost path J^*, the index of the current state i, and time sample k are assigned from the node as follows

$$J^*, i, k \leftarrow \mathcal{N}_{ik} \tag{6.48}$$

2. The temperature and volume of the upper and lower EWH nodes are also assigned the node entries corresponding to state index i and sample k. The volumes of the upper and lower node are extracted from the volume ratio $\mathbf{V}^*_{\text{rik}}$.

$$\begin{aligned} T_{\text{U}} &\leftarrow \mathbf{T}^*_{\text{Uik}} && (6.49)\\ T_{\text{L}} &\leftarrow \mathbf{T}^*_{\text{Lik}} && (6.50)\\ V_{\text{U}} &\leftarrow V_{\text{tank}} \times \mathbf{V}^*_{\text{rik}} && (6.51)\\ V_{\text{L}} &\leftarrow (1 - V_{\text{tank}}) \times \mathbf{V}^*_{\text{rik}} && (6.52) \end{aligned}$$

The EWH nodal mode n_{s} is assigned based on knowledge of volume ratio $\mathbf{V}^*_{\text{rik}}$.

$$n_{\text{s}} = \left\{ \begin{array}{ll} 1 & \text{if } V_{\text{rik}} = 1 \text{ or } 0 \\ 2 & \text{otherwise} \end{array} \right\} \tag{6.53}$$

3. For the given index i of the current state $\mathbf{x}(k)$, calculate the next state $\mathbf{x}_j(k+1)$ for all admissible values of the control input $\mathbf{u}(k)$.

For each

$$\begin{aligned} P_{\text{elec}}(k) &\in \{0, P_{\text{rated}}\} && (6.54)\\ \text{and } P_{\text{elec}}(k) &\leq P_{\text{max}}(k) && (6.55) \end{aligned}$$

If in *one-node* mode, calculate

$$E(k) = E(k+1) - [P_{\text{elec}}(k) - P_{\text{usage}}(k) - P_{\text{losses}}(k)]\Delta t \tag{6.56}$$

with

$$\begin{aligned} P_{\text{usage}}(k) &= c\rho Q_{\text{usage}}(k)[T(k) - T_{\text{in}}(k)] && (6.57)\\ P_{\text{losses}}(k) &= \frac{1}{R_{TH}}[T(k) - T_{\text{ambient}}(k)] && (6.58) \end{aligned}$$

where

$$T(k) = \frac{E(k)}{c\rho V_{\text{tank}}} \tag{6.59}$$

If in *two-node* mode, calculate

$$E(k+1) = E(k) + [\dot{E}_{\text{U}}(k) + \dot{E}_{\text{L}}(k)]\Delta t \tag{6.60}$$

where

$$\dot{E}_{\mathrm{U}}(k) = -P_{\mathrm{usage,U}}(k) - P_{\mathrm{losses,U}}(k) - P_{\mathrm{trans,U}}(k) \tag{6.61}$$
$$\dot{E}_{\mathrm{L}}(k) = P_{\mathrm{elec}}(k) - P_{\mathrm{usage,L}}(k) - P_{\mathrm{losses,L}}(k) - P_{\mathrm{trans,L}}(k) \tag{6.62}$$

with

$$P_{\mathrm{usage,U}}(k) = c_P \rho Q_{\mathrm{usage}}(k) T_{\mathrm{U}}(k) \tag{6.63}$$
$$P_{\mathrm{usage,L}}(k) = c_P \rho \Big(\frac{\dot{V}_{\mathrm{L}} - Q_{\mathrm{usage}}(k)}{\dot{V}_{\mathrm{L}}} \Big) [T_{\mathrm{L}}(k) - T_{\mathrm{in}}(k)] \tag{6.64}$$
$$P_{\mathrm{losses,U}}(k) = \frac{1}{R_U} [T_{\mathrm{U}}(k) - T_{\mathrm{amb}}(k)] \tag{6.65}$$
$$P_{\mathrm{losses,L}}(k) = \frac{1}{R_L} [T_{\mathrm{L}}(k) - T_{\mathrm{amb}}(k)] \tag{6.66}$$
$$P_{\mathrm{trans,U}}(k) = \frac{1}{R_{TH}} [T_{\mathrm{U}}(k) - T_{\mathrm{L}}(k)] \tag{6.67}$$
$$P_{\mathrm{trans,L}}(k) = \frac{1}{R_{TH}} [T_{\mathrm{L}}(k) - T_{\mathrm{U}}(k)] \tag{6.68}$$

where the thermal resistances R_U, R_L and R_{TH} are updated based on the volumes of the lower and upper nodes, which result in different surface areas.

4. Check if any of the conditions are met for state transitions

$$Q_{\mathrm{usage}} > 0 \tag{6.69}$$
$$T_L \geq T_U \tag{6.70}$$
$$V_U \leq 0 \quad when\ Q_{\mathrm{usage}} > 0 \tag{6.71}$$

If so, reassign T_U, T_L, V_U and V_L accordingly and update n_s.

5. Check whether the next state $\mathbf{x}(k+1)$ is an admissible state.

$$T_{\mathrm{tank,U}}(k) \geq T_{\mathrm{profile,U}}(k) \tag{6.72}$$
$$T_{\mathrm{tank,L}}(k) \geq T_{\mathrm{profile,L}}(k) \tag{6.73}$$

If the next state is an admissible state, continue to Step 6. Otherwise, continue iterating through the possible values of the control input.

6. Find the index j of the quantised next state $\mathbf{x}_j(k+1)$ in the state vector array that is closest to the calculated next state $\mathbf{x}(k+1)$.

7. Calculate the *cost-to-arrive* from the starting state $\mathbf{x}_{k_s}(1)$ to this next state $\mathbf{x}_j(k+1)$ through the current state $\mathbf{x}_i(k)$

$$J_{ij}^{\mathrm{arr}} \leftarrow \Delta J_{ij} + J_i \tag{6.74}$$

where

$$\Delta J_{ij} = P_{\mathrm{elec}}(k) \Delta t \tag{6.75}$$

Next, calculate the *cost-to-go* from this next state $\mathbf{x}_j(k+1)$ to the goal state $\mathbf{x}_f(N)$

$$J_{jf}^{\mathrm{go}} \leftarrow |\mathbf{x}_j(k+1) - \mathbf{x}_f(N)| + \sum_{n=k+1}^{N} E_{\mathrm{usage,h}}[n] \tag{6.76}$$

and finally calculate the total path cost as

$$J_{ij} \leftarrow J_{ij}^{\text{arr}} + J_{jf}^{\text{go}} \tag{6.77}$$

8. A new node $\mathcal{N}_{jk+1}$ at next state j and time $k+1$ is generated with the following new total cost assigned:

$$\mathbf{J}^* \leftarrow J_{ij}$$

and the new calculated values for the EWH as a result of the admissible control input is assigned as follows:

$$\mathbf{T}^*_{Ujk+1} \leftarrow T_U \tag{6.78}$$
$$\mathbf{T}^*_{Ljk+1} \leftarrow T_L \tag{6.79}$$
$$\mathbf{V}^*_{rjk+1} \leftarrow V_r \tag{6.80}$$

where the new volume ratio V_r is calculated as

$$V_r = \frac{V_U}{V_{\text{tank}}} \tag{6.81}$$

9. If a node does not exist for the next state j at time instance $k+1$, the new node $\mathcal{N}_{jk+1}$ branches out from the currently considered node $\mathcal{N}_{ik}$ with the new total path cost. Otherwise, if a node already exists at this position, the total path cost of the existing node and $\mathcal{N}_{jk+1}$ is compared. If $\mathcal{N}_{jk+1}$ has a lower total path cost, it then replaces the existing node, otherwise it is thrown out.

   If $\mathcal{N}_{jk+1}$ is not thrown out, the next state j for the currently considered node $\mathcal{N}_{ik}$ is assigned to the current state of the new node for the corresponding control input. The new $\mathcal{N}_{jk+1}$ is inserted into the priority queue as follows:

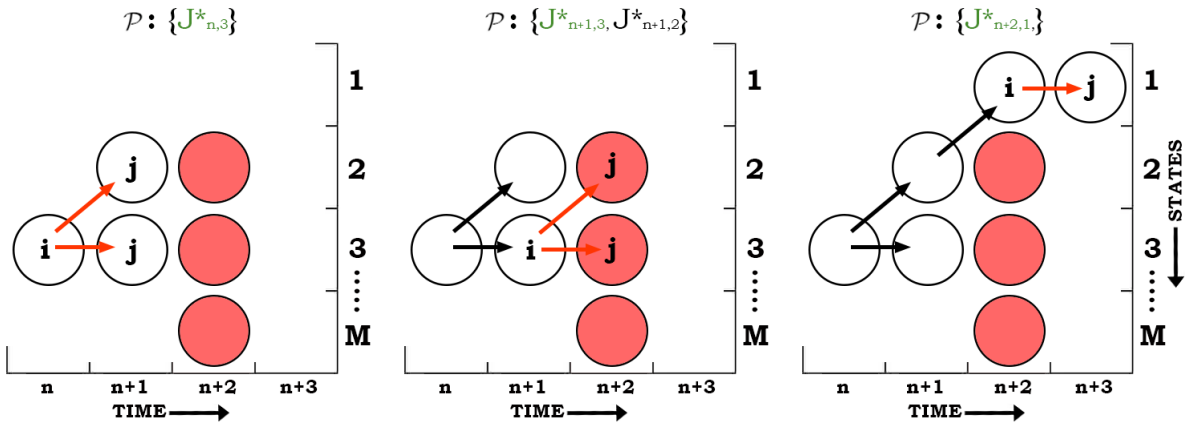
$$\mathcal{P} \leftarrow \mathcal{N}_{jk+1} \tag{6.82}$$

10. If the next state is equal to the goal state f at $k = N$, terminate the execution, otherwise, return to Step 1. Figure 6.5 shows how the order of iterations occur for navigating the optimal path from an initial state at $t = n$. The priority queue is shown above each step to show the order of iterating the search paths.

6.2.3 Optimal path navigation

The A* algorithm produces a BST of the optimal state trajectories and optimal control sequences for EWH control from a single starting state s to the goal state f. Given the initial state $\mathbf{x}_s(1)$, the optimal state trajectory $\{\mathbf{x}_\pi(k) : k = 1, 2, \ldots, N\}$ and the optimal control sequence $\{\mathbf{u}_\pi(k) : k = 1, 2, \ldots, N\}$ can be obtained by starting at the initial state node and iteratively navigating through the BST by following the optimal next state indexes j^*.

Figure 6.6 shows how the BST selects the optimal state trajectory $\mathbf{x}_\pi$ (green nodes) from the initial and goal state and how the corresponding optimal input sequence $\mathbf{u}_\pi$ is produced. The white nodes show search paths that did not continue as they either were not the optimal path or were cut off due to encountering a node that was not admissible (red). The initial and goal states are picked to be equivalent to the energy in the tank that corresponds to the desired boundary condition temperatures T_{start} and T_{end}, respectively.



(a) $i = 3$ and $k = n$ **(b)** $i = 3$ and $k = n + 1$ **(c)** $i = 1$ and $k = n + 2$

Figure 6.5: A* algorithm execution iteration order for finding the optimal path with the node considered from $\mathcal{P}^*$ shown as the first entry of $\mathcal{P}$ (green) for different stages of current state i and time instance k. Nodes are not admissible (red) at time $k = 2$ due to constraints imposed on the system and no entries are added to the $\mathcal{P}$ for those next state paths.

6.2.4 Safety measures

Despite the implementation of the temperature feedback controller and water mixer to minimise the temperature deviation from the optimal plan, additional safety measures are put into place to further reduce the risk of cold events.

Energy matching: For EM control, the optimal planner will generate the optimal temperature trajectory such that it just does not fall below 40°C during water usage. When water usages are predicted, there is always an uncertainty that the actual water usage will start earlier than what was predicted, or use a larger volume of water. This would certainly result in a cold event. Safety measures are implemented when water usages are predicted for EM control. A 10°C buffer is added to the usage constraint T_{usage} to provide a safety margin and greatly reduce the risk of cold events. However, the expected cost of this safety margin is a substantial reduction in energy savings for EM control.

Temperature matching: For TM control, predicting water usages in the optimal planning means that the outlet temperature must be equal to that of TC before each predicted water usage to ensure that the same amount of thermal energy is used. Safety measures are provided to TM by generating a conservative amount of predicted water usages equalling close to the upper bound of the number of water usages that happen per a day. This will ensure that the outlet temperature is matched to that of TC frequently enough in a day that if the expected maximum amount of water usages occurs for the actual water profile, the outlet temperature will most likely satisfy the constraint. In addition, the water mixer will adjust the outlet flow rate to draw the same amount of thermal energy as TC if the temperature is not perfectly matched. Similar to the previous safety measure, the expected cost of this is an increase in electrical energy usage used each day and therefore reduced energy savings.

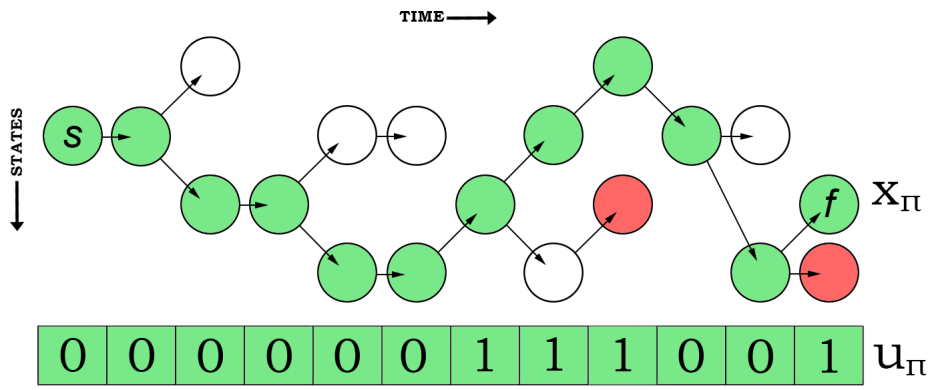


Figure 6.6: BST navigation for the optimal state trajectory sequence $\mathbf{x}_\pi$ (represented by the green nodes) and the path is determined by navigating the optimal next state table $\mathbf{j}^*$. Below the table shows how the corresponding optimal input sequence $\mathbf{u}_\pi$ is also similarly obtained by the control inputs of each node along the path. A 0 refers to $u = 0$ and a 1 refers to $u = P_{\text{rated}}$ for the duration of that time instance.

***Legionella* control:** For EML, when the hot water usage profile is perfectly known by the optimiser, the *Legionella* constraint is imposed so that the temperature of the tank is increased to at least 60 °C for 11 minutes before the start of the largest event on any given day. This prevents *Legionella* growth and also reduces the additional electrical energy required to satisfy the constraint. This becomes more complicated when the optimiser is forced to predict hot water usages and therefore has no certainty when the largest usage, or any usage, will occur in the day. The tank is thus heated instead at the first hour of each day to ensure that the simulated temperatures will prevent the growth of *Legionella* without being interrupted by unexpected water usages.

6.3 Summary

The A* algorithm described in this chapter is used to produce the optimal heating schedule and optimal temperature trajectory for a given predicted or measured water profile. In Chapter 7, a EWH simulator uses these to control the heating element and determine the results of the optimal schedule control of a two-node EWH for the corresponding water profile.

Chapter 7

Optimisation Results

The EWH optimal energy control system that was developed in the previous chapters is tested using a simulator that was developed for this purpose. Figure 7.1 shows a flow diagram that summarises how all the system components are connected. The sections within this chapter that use a specific component are shown in parenthesis. The EWH simulator and optimiser is executed either for a profile of actual water usage taken from the water draw data or for predicted hot water usage generated from the hot water demand predictor. Both the EWH simulator and optimiser will execute based on the selected heating control strategy and EWH model specified. The simulator simulates the EWH with parameters defined in the simulation setup. The simulator uses the optimal heating plan produced by the optimiser to determine the on and off state of the heating element. When predicted hot water usage profiles are simulated or the EWH is simulated with a model different from the optimal planning, the temperature feedback controller is implemented to compare the EWH measured temperature and optimal temperature for that instance of time and determine the heating element state for the next time instance. The optimiser will optimise the heating schedule for a hot water usage profile based on the optimisation algorithm that corresponds to the EWH model selected. The optimiser executes the optimisation algorithm with parameters defined in the simulation setup. The mixer and user are added to the simulation which also enables energy matching if it is desired for the selected heating control strategy.

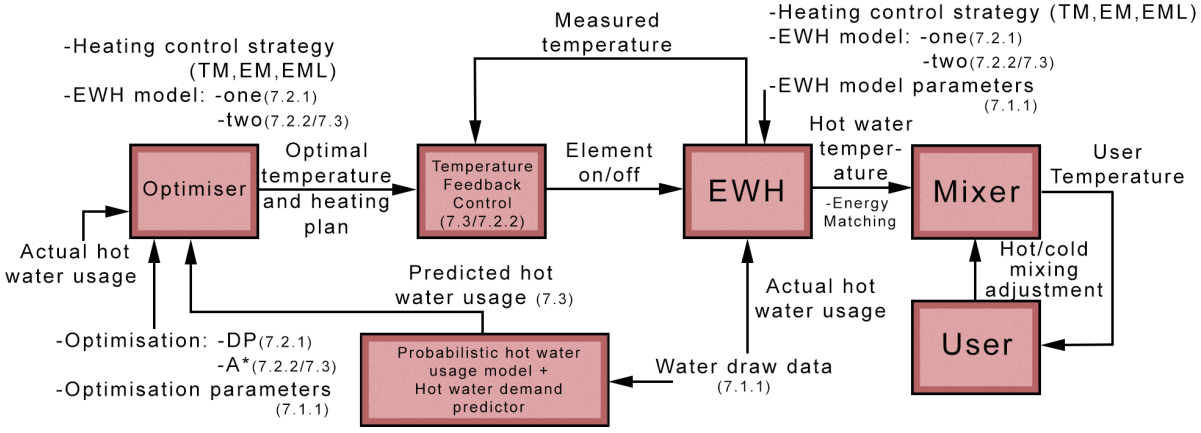


Figure 7.1: Block diagram of the simulator and system components used to obtain the results for this chapter.

The results in the following sections compare the following heating control strategies: thermostat control (TC), optimal temperature matching (TM), optimal energy matching (EM), and optimal energy matching with Legionella prevention (EML). In Section 7.1, the simulation setup is described which defines the parameters used for the system as well as the metrics for evaluating the results of the EWH simulation. Section 7.2 presents the simulation results for the optimal heating control strategies when they have perfect foreknowledge of the user's hot water usage. The results of the dynamic programming approach that uses the one-node EWH model and the A* approach that uses the two-node EWH model are compared. The benefit of including the effects of stratification in the EWH model when planning the optimal heating schedule is also evaluated.Section 7.3 presents the simulation results for the optimal heating control strategies when they do not have perfect foreknowledge, but instead use the predicted hot water usage profiles obtained from the probabilistic hot water usage model fitted to historical usage data. The results of the optimal heating control strategies when using the (imperfectly) predicted hot water usage are compared to the results when using perfect foreknowledge.

7.1 Simulation Setup

This section describes the layout of the simulation setup performed in this chapter. The data cleaning process, probabilistic hot water usage model, DP and A* optimiser and EWH simulator were all implemented with *Python* in *Jupyter Notebook*, an open-source web application, to obtain the results.

7.1.1 Simulation Parameters

Table 7.1 lists the dataset properties, parameters and constants used in the optimisation and simulation processes. The ambient temperature T_{amb} is assumed to be 20°C throughout the simulation. Furthermore, the inlet temperature T_{inlet} is chosen to be 20°C because it is assumed that the temperature of the water entering the tank is the same as the ambient temperature. The constraint for usage temperature T_{usage} is 40°C and any water drawn at a lower temperature is assumed to result in a *cold* event. The energy equivalent for the optimisation parameters are $\text{E}_{\text{tank(max)}}$, $\text{E}_{\text{tank(min)}}$ and $\text{E}_{\text{tank(usage)}}$. The initial and terminal boundary condition temperatures T_{start} and T_{end} are equivalent to the set-point temperature used for TC to ensure a fair comparison of results.

7.1.2 Metrics for evaluating the results of the EWH model

Water draw is aggregated into what we term usage *events*. An event starts when a tap is opened and stops when it is closed; or, put differently, an event starts when a non-zero volume is sampled one sample after a zero volume sample, and stops when the flow returns to zero after the non-zero sample. The event definition quantifies the water usage into bundles, providing a more convenient way of referring to sections of water usage patterns, and providing a metric for counting the number of times a user experiences an undesired temperature.

Given the plumbing between the water heater outlet and the point of use (i.e. tap), events with a volume of less than 2 L are unlikely to result in the hot water reaching the point

Table 7.1: Parameters used for simulations and optimisation.

Symbol	Description	Value	Unit
	EWH model parameters		
R_{TH}	Thermal resistance of EWH	0.4807	$\frac{K \cdot day}{kWh}$
c	Specific heat capacity of water	4184	$\frac{J}{kg \cdot K}$
ρ	Water density	1000	$\frac{kg}{m^3}$
T_{set}	Target temperature	68.5	°C
T_{hyst}	Hysteresis (deadband)	±1.5	°C
T_{amb}	Ambient temperature	20	°C
T_{inlet}	Inlet temperature of EWH	20	°C
V_{tank}	Tank volume of EWH	150	L
P_{rated}	Power rating of element	3	kW
	Optimisation parameters		
$T_{\text{tank(max)}}$	Maximum temperature of EWH	70	°C
$T_{\text{tank(min)}}$	Minimum temperature of EWH	20	°C
T_{usage}	Minimum target usage temperature	40	°C
T_{start}	Initial boundary condition of EWH	68.5	°C
T_{end}	Terminal boundary condition of EWH	68.5	°C
n_s	Node state of EWH model	1 or 2	
	Descriptive statistics of water draw dataset		
D	Duration	7	days
Δt	Sampling period	1	min
Resolution		0.5	L
Number of water heaters		77	
Number of seasons		4	
Average number of events per EWH per day		7.5	

of use. This can be shown to be true with the normal pipe diameter of 22 mm and a conservative pipe length of 5 m. This results in a volume of 1.9 L hot water drawn into the piping before the point of use. These events that draw hot water that does not reach the point-of-use are therefore considered unintended events, for which our algorithm does not instruct the element to heat, and which are excluded when counting the number of cold events. These events are likely to occur when the user uses a mixer tap with a position between hot and cold, or unwittingly opens the hot tap for a quick cold draw.

Table 7.2 lists the metrics we use to assess the performance of the various strategies. The energy metrics comprise the electrical energy, effective energy in the drawn water, energy lost to the environment, and energy lost due to unintentional usage. The effect of unintentional usage on the thermal energy drawn is only evaluated in the results of the one-node EWH model, for which the remaining sections include these events with the effective thermal energy drawn for all water usages. We use the change in electrical energy as a percentage reduction and as an absolute change to determine the achieved savings. We also report the average event temperatures, with the number of cold events, to establish user comfort and convenience.

7.2 Evaluation of the optimal EWH control using *perfect foreknowledge* of hot water usage

This section presents the simulations that were performed to determine how much energy can theoretically be saved with optimal control for electric water heaters compared to traditional TC, when perfect foreknowledge of the hot water usage profile is available.

Table 7.2: Metrics used for performance assessment.

Metric	Description	Unit
E_{elec}	Daily average electrical energy used per day for water heating.	kWh
E_{draw}	Daily average thermal energy in hot water drawn from tank (per day). Indicates effective energy used, but includes energy lost due to unintentional use.	kWh
E_{loss}	Daily average energy lost to environment through tank (per day).	kWh
$E_{\text{unintentional}}$	Daily average thermal energy lost due to unintentional use (per day).	kWh
$\Delta E(kwh)$	Reduction in electrical energy per day compared to thermostat control.	kWh
$\Delta E(\%)$	Reduction in electrical energy per day compared to thermostat control.	%
$\overline{T}_{\text{usage}}$	Average event temperature of the upper node (excludes unintentional use)	°C
Cold events	Number of events with any sample $T < 40$°C.	

The DP approach, that uses a one node EWH model to perform the planning, and the A* approach, that uses a two-node model to perform the planning, are both simulated and evaluated. In both cases, a two node model is used to simulate the real EWH that is being controlled. The energy savings and cold events when planning with a two node model are compared to the energy savings and cold events when planning with a one node model, to evaluate the benefit of taking stratification into account when performing the optimal planning. Section 7.2.1 evaluates the results of optimal planning for a one-node EWH model using DP and simulating it on a one-node EWH model. Section 7.2.2 evaluates the results of optimal planning for a two-node EWH model using A* and simulating it on a two-node EWH model. The section also compares the simulation results of a two-node EWH model that uses the optimal planning produced for a one and two-node EWH model followed by a discussion on the effect of accounting for stratification.

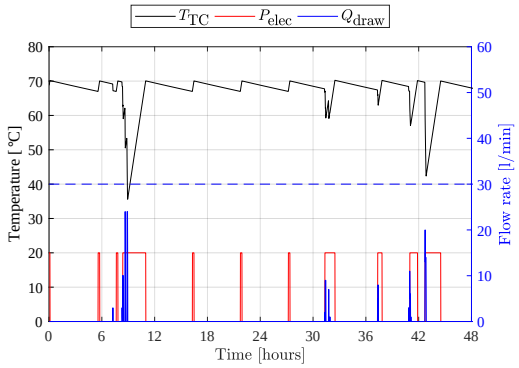
7.2.1 One-node EWH model with DP optimal planning

Illustrative example

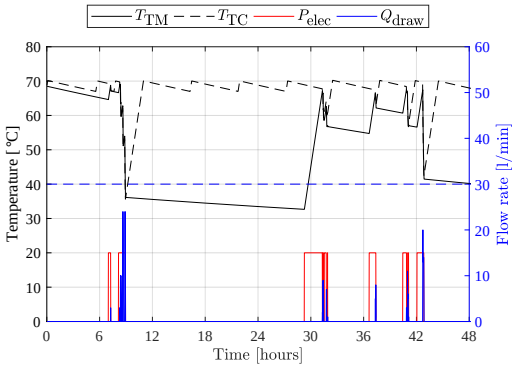
An illustrative example is discussed for a single EWH to convey the characteristics of the heating schedule, temperature trajectory and outlet flow rate for the one-node EWH. Figure 7.2 shows simulation results for the EWH, comparing TC, TM, EM and EML. For each heating control strategy, the EWH temperature, outlet flow rate, and heating element control signal are plotted as a function of time for a period of two days (48 hours). The results are shown for the same hot water draw profile, with all simulations starting from the same initial EWH temperature.

Thermostat control (TC): The TC keeps the temperature at about a set point of 68.5 °C (allowing for 1.5 °C hysteresis). When the temperature drops below 67 °C, the heating element switches on; when the temperature rises above 70 °C, it switches off. During usage events, when the outlet flow rate is non-zero, the temperature drops significantly. A single cold event is observed at time $t = 9$ hours, when the temperature dropped below 40 °C. This cold event was caused by an unusually large usage event. It could not be

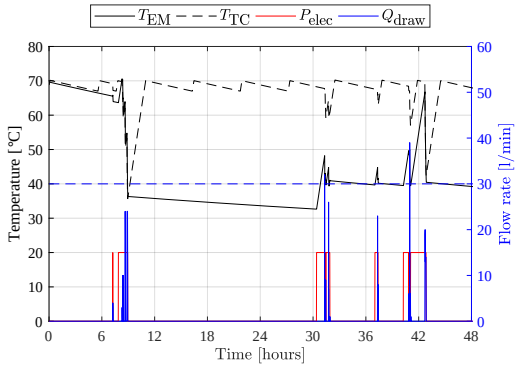
prevented even with the temperature at maximum at the start of the event, and with the heating element switched on for the full duration of the event. The electrical power consumption for thermostat control was 23.85 kWh over the two days.



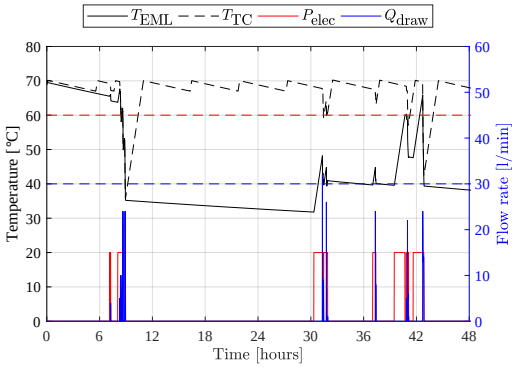
(a) Thermostat control



(b) Temperature matching



(c) Energy matching



(d) Energy matching with *Legionella* control

Figure 7.2: Simulation results of a one-node EWH model for TC, TM, EM, and EML, respectively. The plots show the EWH temperature, the outlet flow rate, and the element state. The EWH temperature for thermostat control is repeated on all plots for comparison. The cold event threshold (40 °C) is indicate with a dashed blue line. The *Legionella* threshold (60 °C) is indicated with a dashed red line.

Optimal temperature matching (TM): The TM control ensures that the temperature matches the corresponding temperature for thermostat control, but only during usage events. Between usage events, the TM control allows the temperature to drop, and the EWH only starts heating again just before the next usage event. The temperature between the usage events is therefore lower for TM than for TC. We observed the same cold event at time $t = 9$ hours, due to the unusually large usage event. Since TC was not able to prevent the cold event, starting at maximum temperature and with the heating element switched fully on, we did not expect that TM would be able to prevent the cold event either. The electrical power consumption for optimal temperature matching was 16.7 kWh over the two days.

Optimal energy matching (EM): The EM control ensures that the temperature remains above 40 °C during usage events. We increased the outlet flow rate to represent the fact that all the water that reached the user was drawn from the EWH, and that the hot water

from the EWH was not mixed with cold water before reaching the user. Between usage events, the EM control allows the temperature to drop, and only starts heating again just before the next usage event. The temperature both during and between the usage events is therefore lower for EM than for TC and TM. The same cold event is observed at time $t = 9$ hours, due to the overlarge usage event. The electrical power consumption for optimal energy matching was 16.55 kWh over the two days.

Optimal energy matching with Legionella prevention (EML): The results for EML look almost exactly the same as the results for EM control without this prevention, except that we increased the EWH temperature to the *Legionella* prevention threshold of 60 °C at time $t = 41$ hours and then held it there for 11 min. This increase in temperature was scheduled to be just before the largest usage event for the day, excluding cold events. We observed the same cold event at time $t = 9$ hours, due to the unusually large usage event. The temperature both during and between usage events was therefore almost the same for EM and EML, except that the latter had a higher temperature during the scheduled *Legionella* prevention heating. We therefore expected that EML would have a slightly higher power consumption. The electrical power consumption for optimal energy matching with *Legionella* prevention was 16.05 kWh over the two days.

Statistical results

The statistical results from all of the simulations performed for all 77 EWHs are discussed in this section. The distributions of the electrical energy used per day, thermal energy drawn per day, average outlet temperature during usage events, and thermal energy losses per day over all EWHs are shown in sub-figures (a) to (d) for Figure 7.3. A metric is defined to asses the performance of the electrical energy used by an EWH during the simulation. The average electrical energy used per day by an individual EWH is calculated using

$$\overline{P}_{\text{elec}|h} = \frac{\sum_{k=1}^{N_h} P_{\text{elec}|h}(k)\Delta t}{D} \quad \text{kWh/day} \tag{7.1}$$

where $\overline{P}_{\text{elec}|h}$ is the average electrical energy used by heater h per day, $P_{\text{elec}|h}(k)$ is the instantaneous electrical power used by heater h at sampling instant k, Δt is the sampling period, N_h is the total number of samples, and D is the total number of days in the data set. A metric is defined to asses the performance of the thermal energy in hot water drawn and the energy lost to the environment by an EWH during the simulation. The average thermal energy drawn per day $\overline{P}_{\text{draw}|h}$ and the average thermal energy losses per day $\overline{P}_{\text{loss}|h}$ for an individual EWH are calculated using

$$\overline{P}_{\text{draw}|h} = \frac{\sum_{k=1}^{N_h} P_{\text{draw}|h}(k)\Delta t}{D} \quad \text{kWh/day} \tag{7.2}$$

$$\overline{P}_{\text{loss}|h} = \frac{\sum_{k=1}^{N_h} P_{\text{loss}|h}(k)\Delta t}{D} \quad \text{kWh/day} \tag{7.3}$$

Two metrics are defined to asses the performance of electrical energy reduction compared to TC. The distributions of the electrical energy savings per day for each strategy (TM, EM, and EML), expressed both as a reduction in kWh per day and as a percentage reduction, are shown in sub-figures (e) and (f) for Figure 7.3. We calculate the distribution

of the energy savings for a particular strategy by first calculating the individual savings for each EWH, and then plotting the distribution of the individual savings over all EWHs. For example, we calculated the energy savings per day for the EM strategy compared to the baseline TC strategy for each individual EWH using the following formulas, and then plotted the distribution of savings for the EM strategy over all EWHs:

$$\Delta\overline{P}_{\text{elec}|h,\text{EM}}(\text{kWh/d}) = \overline{P}_{\text{elec}|h,\text{TC}} - \overline{P}_{\text{elec}|h,\text{EM}} \quad \text{kWh/day} \tag{7.4}$$

$$\Delta\overline{P}_{\text{elec}|h,\text{EM}}(\%) = \frac{\overline{P}_{\text{elec}|h,\text{TC}} - \overline{P}_{\text{elec}|h,\text{EM}}}{\overline{P}_{\text{elec}|h,\text{TC}}} \times 100\% \tag{7.5}$$

The effects of the different control strategies on energy and temperature are summarised in Table 7.3 and shown in Figure 7.3. Figure 7.3 (a) shows that all the other strategies used less electrical energy than the baseline TC strategy. Since hot water was not needed every day, the minimums were zero for the TM and EM strategies, but not zero for the EML method because of the *Legionella* control.

Appendix C shows a 20-day simulation of a water profile which compares the one-node EWH outlet temperature in Figure C.1 and C.2 for optimal control (OC) (with and without temperature-matching respectively) and TC. The plots also show water usage and the power supplied by the optimal heating schedule for TC and OC.

Temperature-matched optimisation

The median electrical energy used for TM was 6.2 kWh/day, 0.6 kWh/day (8.8 %) less than the 6.8 kWh/day median for TC. These reductions are presented as distributions in Figures 7.3 (e) and (f), as daily energy (kWh/day) reduction and percentage reduction, respectively.

Despite the reduction in electrical energy used to heat the water, Figures 7.3 (b) and (c) show that the thermal energy drawn and the outlet temperature during events, were the same for TM as for TC, respectively. The median event temperatures for TC and TM were both approximately the set point temperature of 68 °C. The number of cold events, a metric of negative user satisfaction, did not increase for the TM control, and actually decreased by 4 from 115 out of a total of 15 581 events. From the user's perspective, the TC and TM strategies will therefore have no adverse effect on perceived temperatures, despite the energy savings.

Looking at the distribution of savings for the water heaters, we found the percentage reduction for TM, given as [25^{th} *percentile*, **median**, 75^{th} *percentile*], was [*0.4*, **0.6**, *1.0*] kWh/day, or [*4.4*, **7.6**, *16.7*] %. This result is in line with the savings reported by [101, 102], and [23], indicating that their approaches employed heating schedules that provided sufficient heating time to effectively achieve results that resemble temperature matching. Figure 7.3 (d) shows that the median thermal losses decreased from 2.4 kWh/day to 2.1 kWh/day – a 0.3 kWh/day reduction.

Table 7.3: Energy, temperature, volume, and cold event results for one-node EWHs.

		TC	TM	EM	EML
$\overline{V}_{\text{hot}}$	(L/day)	119	119	196	169
E_{elec}	(kWh/day)	6.8	6.2	5.5	5.7
E_{draw}	(kWh/day)	4.3	4.3	4.3	4.3
$E_{\text{unintentional}}$	(kWh/day)	0.06	0.05	0.02	0.02
E_{loss}	(kWh/day)	2.4	2.1	1.0	1.4
$\overline{T}_{\text{usage}}$	(°C)	67.5	67.3	40.3	48.4
ΔE_{elec}(kWh)	(kWh/day)	–	*0.4*, **0.6**, *1.0*	*1.4*, **1.5**, *1.6*	*0.7*, **1.0**, *1.3*
ΔE_{elec}(%)	%	–	*4.4*, **7.6**, *16.7*	*13.9*, **21.7**, *33.1*	*12.3*, **16.4**, *18.5*
Cold events*		*0*, **0**, *0*	*0*, **0**, *0*	*0*, **0**, *0*	*0*, **0**, *0*

Note: The distributions are reported as 25^{th} *percentile,* ***median****,* 75^{th} *percentile*
Cold events are taken as the total cold events for the EWH that uses the* 25^{th} *percentile,* ***median**, and* 75^{th} *percentile of the total volume of water used.*
**All these cold events were generated by nine heavy users out of the 77, which means the median and* 75^{th} *percentile EWHs had zero cold events. The number of cold events did not increase from that of TC.*

Energy-matched optimisation

The average outlet temperature for EM is significantly lower than for TC and TM, as intended, with a median temperature of 40 °C. The implication is that the user will have to mix in less or no cold water to reach the desired temperature of 40 °C or below, and this was achieved in the optimal planning. Despite this lower temperature, the thermal energy drawn was the same as TC – 4.3 kWh/day – owing to a larger volume draw of 119 L vs. 196 L. The number of cold events for the EM strategy did not increase out of 15 581 events.

Figures 7.3 (a) and (b) show that the electrical energy used in the EM strategy was substantially less than either in TC or TM. The median electrical energy used for EM was 5.5 kWh/day, 1.3 kWh/day (19.1 %) less than the 6.8 kWh/day median for TC. The reduction is shown in Figures 7.3 (e) and (f), with respective energy and percentage reductions from the TC baseline of [*1.4*, **1.5**, *1.6*] kWh/day and [*13.9*, **21.7**, *33.1*] %. These results are in line with the simulated and field-measured and/or energy-matched results reported by [98, 18, 19] and [36]. The median standing loss for this strategy was a mere 1.0 kWh/day, a reduction of 1.4 kWh/day (58.3 %) from TC's 2.4 kWh/day.

Legionella control

The results for the *Legionella* strategy lie between the TM and EM strategies' results, as expected. The median electrical energy usage for *Legionella* control was 5.7 kWh/day, 16.2 % less than the median for TC. The results in Table 7.3 show that the respective percentage and absolute reductions from TC are [*12.3*, **16.4**, *18.5*] % and [*0.7*, **1.0**, *1.3*] kWh/day. The median standing losses for this strategy was 1.4 kWh/day, a reduction of 1.0 kWh/day from that of TC. The median event temperature for this strategy was 48.4 °C, while the median number of cold events was the same as for EM.

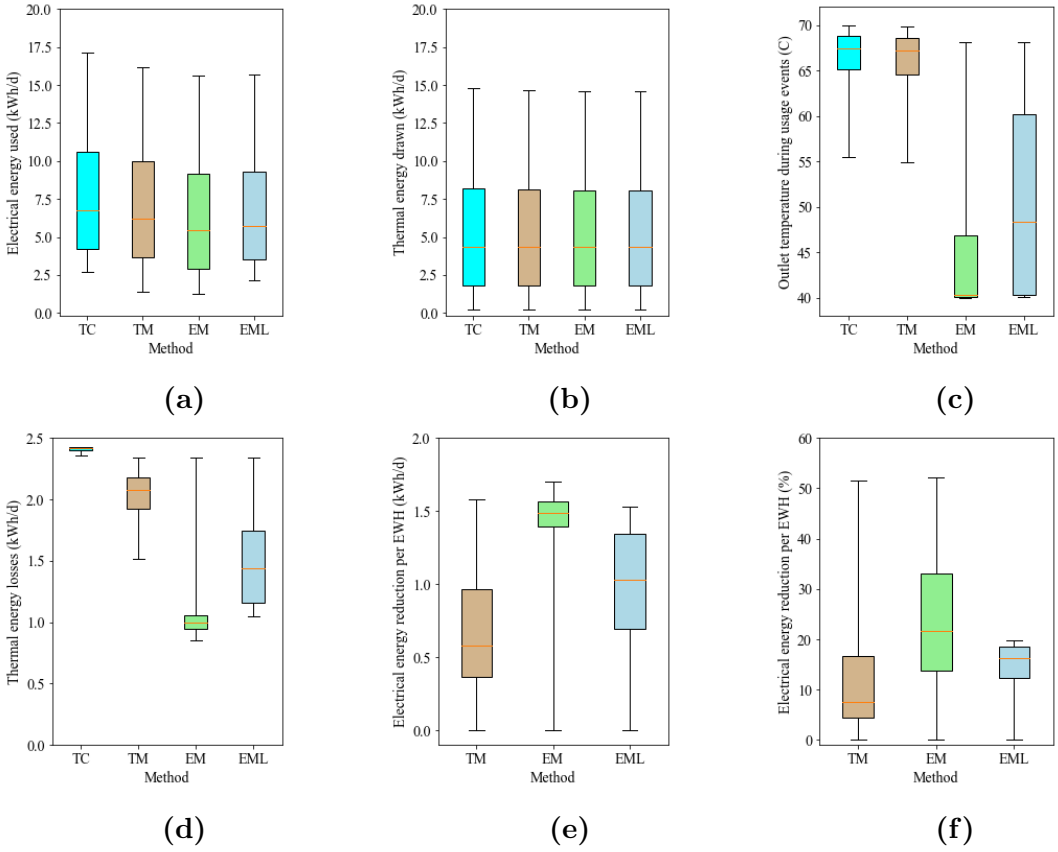


Figure 7.3: Energy and temperature results of the different control strategies represented as distributions for all water heaters using a **one-node** model. (a) depicts electrical energy used per EWH per day, (b) depicts thermal energy drawn per EWH per day, (c) depicts outlet temperatures during usage events, (d) depicts thermal losses per EWH per day, and (e) and (f) depict the savings achieved in electrical energy per EWH as a reduction in kWh per day and percentage of total used, respectively.

Unintentional usage loss

The unintentional usage loss, albeit very small, was 0.06 kWh/day for TC and 0.05 kWh/day for TM. This indicates that TM did not entirely counter the unintended usage losses, and that most of the savings were due to a reduction in standing losses as a result of the temperature-matched optimised scheduling.

The unintentional energy usage for EM was only 0.02 kWh/day, compared to the 0.06 kWh/day for TC. This result is important, because it demonstrates that EM control limits the detrimental effect of unintentional usage on the electrical energy used by essentially applying a lower temperature control. This result also makes intuitive sense – the amount of energy lost due to unintentional usage loss will be less when the water is at a lower temperature, which is also true for simply operating at a lower set point temperature.

It further shows that more savings can be achieved by OC than by merely reducing standing losses. Although these numbers are quite small, it stands to reason that the unintended usage is underestimated. This may explain the significantly higher savings reported by some recent field trials [23, 103, 35]. Consider, for example, the warm water

that remains in the piping network between the heater and the point of use after an event. The energy in this water also constitutes usage losses since that water will cool down if not drawn shortly after the initial event.

Assuming a conservative volume of 2 L in the piping between the heater and the point of use, each event will in fact comprise legitimate usage and 2 L unintentional usage loss. If the residual warm water in the piping is at a lower temperature than the TC set point, the net losses will be significantly less: the average volume used per user was 76 L/day with an average of 7.2 events/day. This means as much as 19 % of usage would have resulted in usage losses. As an example of unintended use, we have anecdotal evidence of users of a smart EWH controller who became aware that their casual workers were unnecessarily washing cars or cleaning floors with hot water – behaviour that was quickly remedied [44, 49].

It may be argued that this control is essentially the same as setting the thermostat to a lower temperature, but the EM control does better than just supplying a lower temperature on occasion: it ensures that hot water is delivered when needed. However, these savings will be highly dependent on the specifics of each user's behaviour, each installation's plumbing setup - including the spouts - and warrants further investigation.

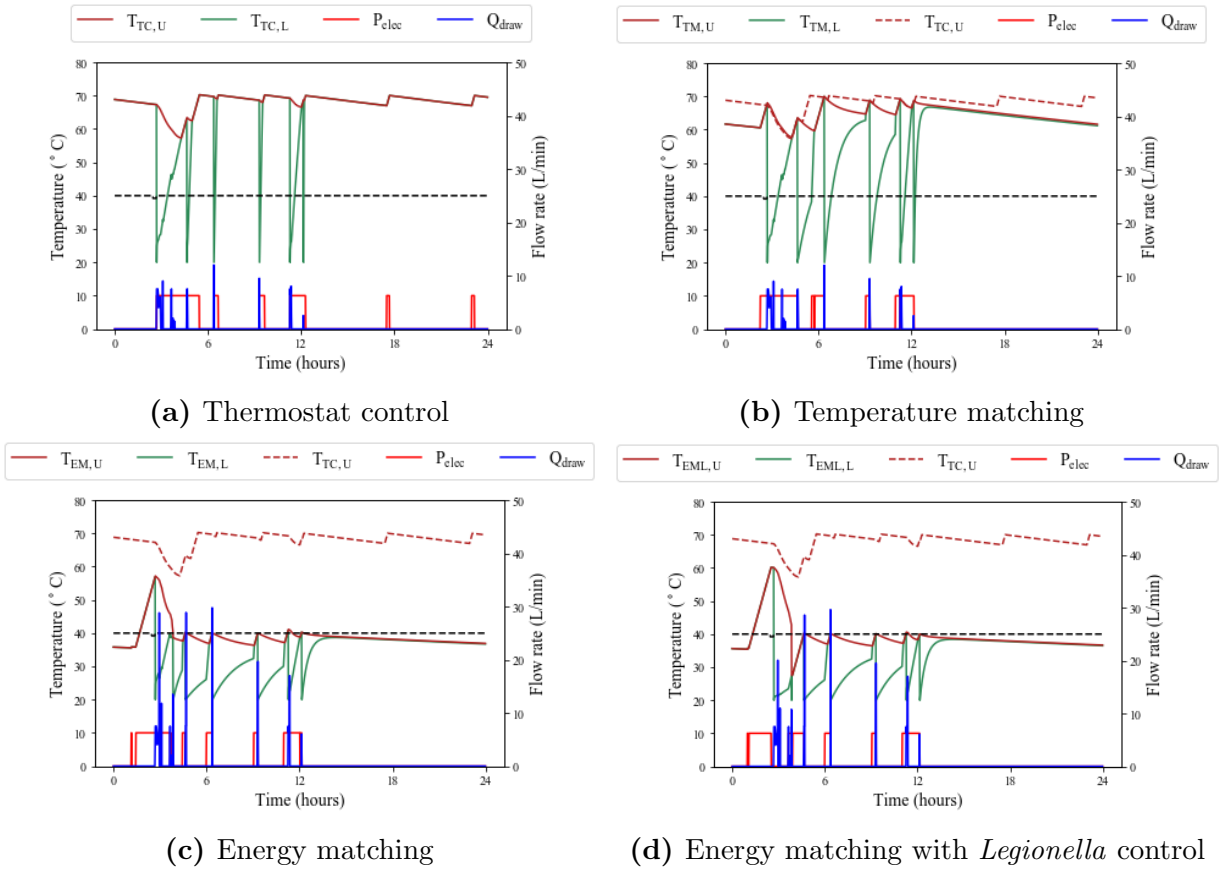
7.2.2 Two-node EWH model with A* optimal planning

Illustrative example

This section obtains results similar to the previous section for TC, TM, EM and EML, with the exception that a two-node EWH model is simulated, and the optimal planning is produced using A* that uses this two-node model. Figure 7.4 shows example simulation results for a single two-node EWH. For each control technique, the EWH temperatures, outlet flow rate , and heating element control signal are plotted as a function of time for a 24-hour period. The upper node thermostat control temperature $T_{\mathrm{TC,U}}$ is repeated in Figures 7.4 (a) to (d) as a reference. The results are shown for the same hot water usage profile, with all simulations starting from the same initial EWH temperature of 68.5 °C for both upper and lower nodes. The plots in Figure 7.4 are not shown for the first day and the observed starting temperature at $t = 0$ varies depending on the state of the tank at the beginning of the considered day.

Thermostat control (TC): TC for the two-node EWH exhibits similar behaviour to that of the previous one-node EWH shown in Figure 7.2. The difference between the models is observable when the outlet flow rate is non-zero. When the first usage of the day occurs at time $t = 3$ hours, the EWH transitions to a two-node state with the lower node temperature immediately dropping to 20 °C and the upper node temperature rapidly decreasing as a result of water stratification caused by the increasing volume of cold water. The heating element switches on when the upper node temperature drops below 67 °C. The heating element supplies power to the lower node which will therefore first need to reach the temperature of the upper node and return to a one-node state before the upper node temperature can increase to the set point temperature.

Optimal temperature matching (TM): The TM control ensures that the outlet temperature matches the corresponding temperature for TC, but only during usage events. Between



(a) Thermostat control

(b) Temperature matching

(c) Energy matching

(d) Energy matching with *Legionella* control

Figure 7.4: Simulation results for TC, TM, EM and EML, respectively. The plots show the EWH upper and lower node temperature, the outlet flow rate, and the element state. The EWH upper node temperature for TC is repeated on all plots for comparison. The cold event threshold (40 °C) is indicate with a dashed black line.

usage events, the TM control allows the temperature to drop, and the EWH only starts heating again just before the next usage event. Exponential responses are observed in the upper node and lower node temperatures when the heating element is off, but the two nodes are at different temperatures. These exponential responses represent their temperatures striving to one another due to the inter-node energy transfer.

Optimal energy matching (EM): The EM control ensures that the upper node temperature remains above 40 °C during usage events and the outlet flow rate increases so that the thermal energy drawn is equivalent to that of TC. The first usage of the day at $t = 3$ is large and the temperature of the tank is increased to 57 °C just before the event starts to ensure that the upper node temperature will remain above 40 °C throughout the event.

Optimal energy matching with Legionella prevention (EML): The results of EML are similar to the results for EM control but with *Legionella* prevention. The temperature of the whole tank is increased to the *Legionella* prevention threshold of 60 °C at time $t = 3$ hours and then held there for 11 min. The prevention of *Legionella* growth occurred at this time because this was the largest water event of the day and the temperature had to only increase by 3 °C from 57 °C of the optimal temperature observed for EM. This shows the effectiveness of imposing the *Legionella* prevention constraint just before the largest

event of the day to minimise additional electrical energy requirements.

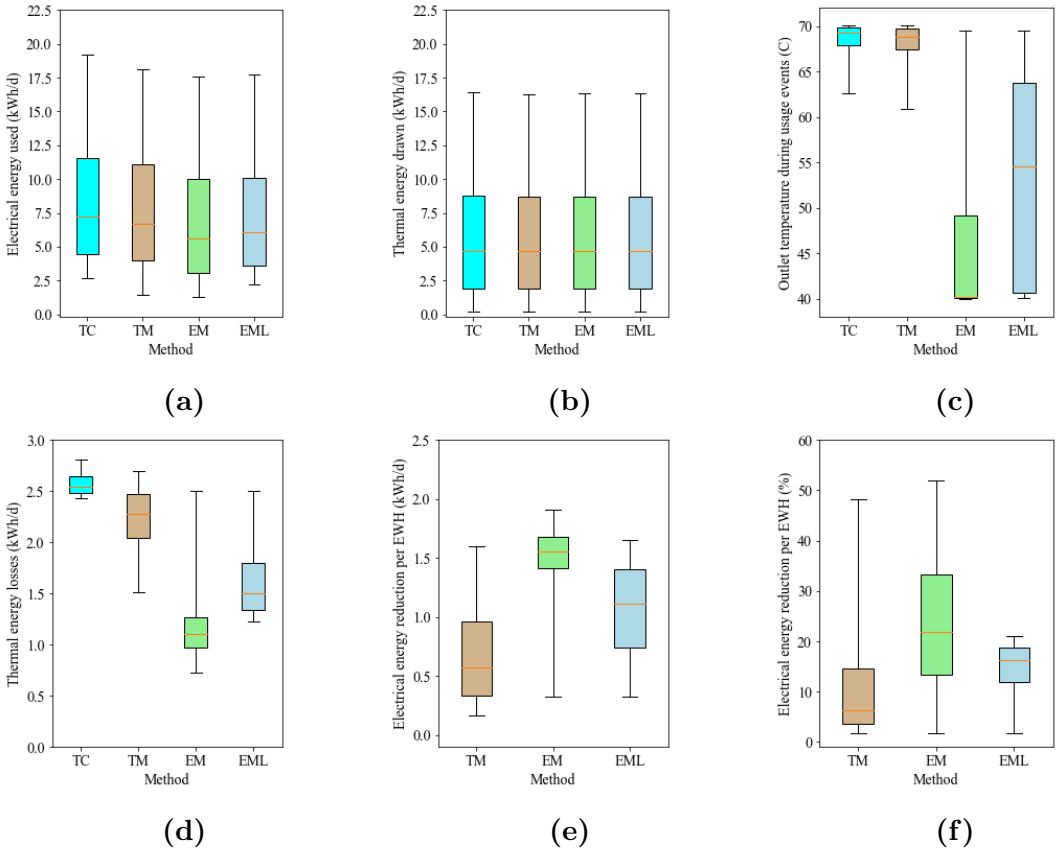


Figure 7.5: Energy and temperature results of the different control strategies represented as distributions for all water heaters using a **two-node** model and A* optimisation. (a) depicts electrical energy used per EWH per day, (b) depicts thermal energy drawn per EWH per day, (c) depicts outlet temperatures during usage events, (d) depicts thermal losses per EWH per day, and (e) and (f) depict the savings achieved in electrical energy per EWH as a reduction in kWh per day and percentage of total used, respectively.

Table 7.4: Energy, temperature, volume, and cold event results for two-node water heaters.

		TC	TM	EM	EML
$\overline{V}_{\text{hot}}$	(L/day)	119	119	231	192
E_{elec}	(kWh/day)	7.3	6.7	5.6	6.0
E_{draw}	(kWh/day)	4.7	4.7	4.7	4.7
E_{loss}	(kWh/day)	2.5	2.3	1.1	1.5
$\overline{T}_{\text{usage}}$	(°C)	69.3	68.9	40.2	54.5
ΔE_{elec}(kWh)	(kWh/day)	–	*0.3*, **0.6**, *1.0*	*1.4*, **1.6**, *1.7*	*0.9*, **1.2**, *1.4*
ΔE_{elec}(%)	%	–	*3.7*, **6.3**, *14.7*	*13.3*, **21.9**, *33.2*	*11.9*, **16.2**, *18.8*
Cold events*		*0*, **0**, *0*	*0*, **0**, *0*	*0*, **0**, *0*	*0*, **0**, *0*

Note: The distributions are reported as 25^{th} *percentile,* ***median****,* 75^{th} *percentile*
Cold events are taken as the total cold events for the EWH that uses the* 25^{th} *percentile,* ***median**, and* 75^{th} *percentile of the total volume of water used.*

Statistical results

The statistical results from all of the simulations performed for all 77 EWHs are discussed in this section. The metrics used to asses the performance of the simulations are also calculated using Equation 7.1, 7.2, 7.3, 7.4 and 7.5. The effects of the different control strategies on energy and temperature for water heaters using a two-node model are summarised in Table 7.4 and shown in Figure 7.5. As expected, 7.5 (a) shows that all the other strategies used less electrical energy than TC, even with Figure 7.5 (b) showing that all the control strategies had equivalent thermal energy usage.

Temperature-matched optimisation

The median electrical energy used for TM was 6.7 kWh/day and is (8.2 %) less than 7.3 kWh/day for TC. Figures 7.5 (c) shows the outlet temperature during events were approximately the same for TM and TC, despite the reduction in electrical energy used, and the median temperature difference was only 0.4 °C. The number of cold events did not increase for TM control and stayed at 115 out of a total of 15 581 events. Of the 77 EWHs, only nine had at least one cold event and the EWH with the most cold events was 61 of 521 events (11.7 %). The prevention of these cold events was inevitable since they also occurred for TC. Looking at the electrical energy reduction in Figures 7.5 (e) and (f), it is observed that the reduction for TM was [*0.3*, **0.6**, *1.0*] kWh/day ([*3.7*, **6.3**, *14.7*] %). Figure 7.5 (d) shows that the median thermal losses decreased from 2.5 kWh/day of TC to 2.3 kWh/day.

Energy-matched optimisation

As expected, the average outlet temperature for EM and EML was considerably lower than the other control strategies. The median for both these control strategies were 40.2 °C and 54.5 °C, respectively. These lower temperatures lead to the average volume draw increasing for EM and EML from 119 L to 231 L and 192 L, respectively. These lower temperatures also contribute to the median standing losses decreasing by 1.4 kWh/day (56 %) and 1.0 kWh/day (40 %) for EM and EML, respectively, from 2.5 kWh/day of TC. This explains the substantially less electrical energy used for EM, where the median electrical energy used for EM was 5.6 kWh/day. This was 1.7 kWh/day (23.3 %) less than the 7.3 kWh/day median for TC and this produced an energy reduction of [*1.4*, **1.6**, *1.7*] kWh/day and [*13.3*, **21.9**, *33.2*] %. The number of cold events did not increase for the EM control.

Legionella control

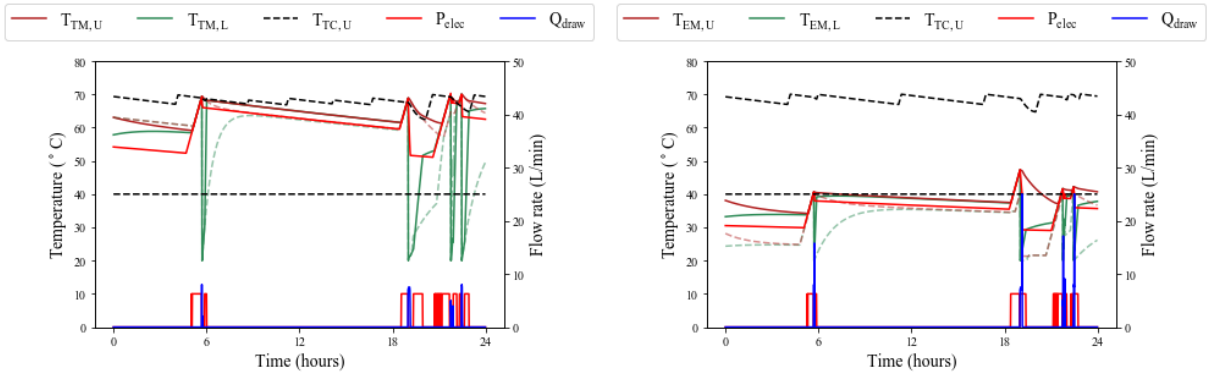
The median electrical energy used for *Legionella* control was 6.0 kWh/day, 17.8 % less than TC. This produced an energy reduction of [*11.9*, **16.2**, *18.8*] % and [*0.9*, **1.2**, *1.4*] kWh/day. The number of cold events did not increase for EML.

7.2.3 Impact of accounting for stratification in the planning

Illustrative example

The effect of accounting for stratification is determined by the using the optimal heating schedule planned for a one-node EWH model and simulating it with the corresponding

hot water profile on a two-node EWH model. This section compares the simulation results for the optimal planning produced for the one-node EWH model in Section 7.2.1 and the two-node EWH model in Section 7.2.2. Both optimal plans are simulated on a two-node EWH and the differences in results are compared. The two-node EWH will expect to encounter model inaccuracies when simulating the EWH with a one-node optimal heating plan. The temperature feedback controller will come into effect to ensure that the temperature of the EWH follows the one-node optimal temperature trajectory. The mixer and user are also simulated to ensure that thermal energy drawn for events are equivalent to that of TC simulated for a two-node EWH model.



(a) Temperature matching **(b)** Energy matching

Figure 7.6: Simulation results for TM and EM. The plots show the target temperature trajectory produced with the DP-based method that plans for a one-node EWH (red line), the temperatures achieved by the two-node EWH simulation (solid lines), and the temperatures achieved with the A*-based method that uses the two-node model (dashed lines). $T_{\mathrm{TC,U}}$ is repeated on all plots with a dashed black line. T_{usage} is indicated with a dashed black line at 40 °C.

Figure 7.6 shows example simulation results for a single two-node EWH over a 24-hour time period for TM and EM control. For each figure, the simulated two-node EWH temperatures (solid and dashed lines show one- and two-node planning, respectively), outlet flow rate and heating element control signal are plotted for the one-node planning. The upper node TC temperature is also indicated. The red temperature trajectory shown is the optimal temperature planned for a one-node EWH model and is used as the reference signal for the temperature feedback controller. The one-node EWH assumes that the outlet temperature is equal to that of the whole tank. The feedback controller will therefore ensure that the temperature of the whole tank follows the optimal temperatures.

Temperature Matching (TM): The figure shows that the outlet temperature of the tank is equal to that of TC for both the one- and two-node planning just before the water event at $t = 6$. After the event, the lower node temperature for the EWH that uses the one-node planning is immediately heated back up to the optimal temperature trajectory to ensure that the temperature of the whole tank follows the optimal temperature. This differs from the temperature for the EWH that uses two-node planning, as it is only required that the outlet temperature, which is the upper node temperature, follows its optimal temperature plan. This means that the simulations of the one-node plan will generally have a slightly higher tank temperature and the difference would be observable in the standing losses.

The electrical power consumption for TM control was 8.2 kWh and 8.0 kWh over the 24 hours for simulations that used one- and two-node model planning, respectively.

Energy Matching (EM): The figure shows that the outlet temperature of the tank is raised notably higher for the one-node planning than that of the two-node just before the water event at $t = 19$. Similar to TM, the feedback controller ensures that the whole tank is at a temperature that would avoid a cold event for the one-node planning. This results in a higher electrical energy usage than simulations that use the two-node planning and only ensure that the upper node temperature avoids cold temperatures during usage. The electrical power consumption for EM control was 9.45 kWh and 8.2 kWh over the 24-hours for simulations that used one- and two-node model planning, respectively.

Statistical results

The statistical results from all of the simulations performed for all 77 EWHs are discussed in this section. A metric is defined to asses the performance of the relative EWH-specific changes during the simulation. The change in electrical energy used per day when the optimal heating schedule and temperatures plan changes from a one- to two-node EWH model and then both simulated on a two-node EWH, the differences are calculated using

$$P^{\Delta N_s}_{\text{elec}|h} = P^{2}_{\text{elec}|h} - P^{1}_{\text{elec}|h} \quad \text{kWh/day} \tag{7.6}$$

where the superscript ΔN_s refers to the difference between the types of planning and superscripts 1 and 2 refer to the results of simulations using one or two-node model planning, respectively. Similar modifications are applied to the formulas that calculate the difference in thermal energy drawn per day, average outlet temperature during usage events, thermal energy losses per day and energy savings per day (kWh and percentage reduction). The simulation results for TM, EM and EML obtained in Figure 7.5 for two-node EWHs is compared with those obtained from the one-node planning that was used to obtain the results in Figure 7.3. The relative EWH-specific changes in results are are summarised in Figure 7.7 (a) to (f). These distributions indicate the effect of accounting for stratification when determining the optimal planning for an EWH.

Figure 7.7 (b) shows that the thermal energy drawn did not change between the two types of planning, and therefore both were equal to that of TC. Figure 7.7 (a) shows that the electrical energy used increased when stratification was not accounted for and the median increase was 0.08 kWh/day for TM; 0.4 kWh/day for EM; and 0.4 kWh/day for EML. For TM, the contribution for the increased energy usage is shown in Figure 7.7 (d), where the slight median increase of 0.03 kWh/day in standing losses is a result of the temperature feedback controller ensuring that the whole tank follows the outlet temperatures for the one-node optimal planning. For EM and EML, the electrical energy increase can be explained by a combination of the increased standing losses of 0.2 kWh/day for both control strategies and the sizeable outlet temperature increase during usage events shown in Figure 7.7 (c) of up to 8 °C and 7.6 °C for EM and EML, respectively.

As Figure 7.7 (e) and (f) shows, the electrical energy reduction relative to the baseline TC was decreased by [*0.01*, **0.08**, *0.17*] kWh/day, or [*0.1*, **1.2**, *2.5*] percentage points for the TM control; [*0.2*, **0.4**, *0.6*] kWh/day, or [*4.0*, **5.4**, *7.0*] percentage points for the EM

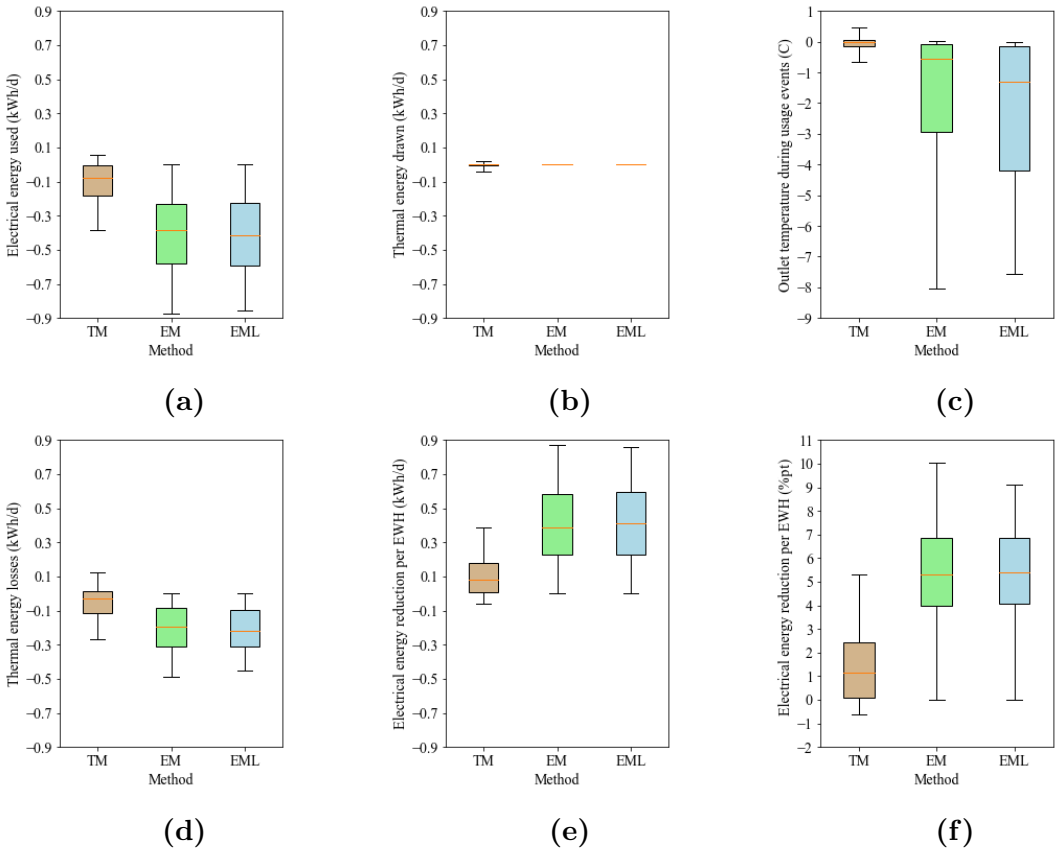


Figure 7.7: The change in energy and temperature results of the different control strategies represented as distributions for all water heaters, where the results from the one-node EWH planning is subtracted from that of the corresponding two-node results using Equation 7.6. (a) depicts difference in electrical energy used per EWH per day, (b) depicts difference in thermal energy drawn per EWH per day, (c) depicts difference in outlet temperatures during usage events, (d) depicts difference in thermal losses per EWH per day, and (e) and (f) depict the difference in savings achieved in electrical energy per EWH as a reduction in kWh per day and percentage points of total used, respectively.

control; and [*0.2*, **0.4**, *0.6*] kWh/day, or [*4.0*, **5.5**, *7.0*] percentage points for EML when stratification was not accounted for in the optimal planning. Furthermore, simulation results showed that the number of cold events did not change between the different types of optimal planning for any of the control strategies

Discussion of the effects of accounting for stratification in the planning

When evaluating the simulation results that compare the one- and two-node EWH model planning, it is clear that the effect of accounting for stratification does not have a big impact on TM, where the electrical energy savings increased by 1.2 percentage points when stratification was accounted for. However, the impact is substantially larger for EM and EML. This is largely due to the one-node plan increasing the temperature before usage events to a higher temperature to ensure that the whole tank does not fall below 40 °C. When stratification is not accounted for in the planning, the electrical energy savings drop by 5.4 and 5.5 percentage points for EM and EML, respectively. Additionally, the

temperature feedback controller ensured that the number of cold events did not change between the simulations that used the different types of optimal planning for all the heating control strategies.

7.3 Evaluation of the optimal EWH control using *predicted* hot water usage

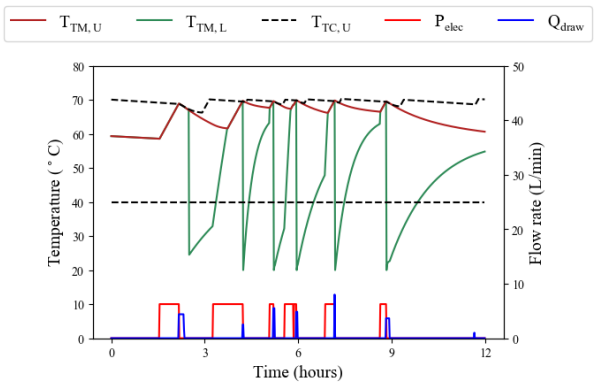
This section presents the simulations that were performed to determine how much energy can practically be saved with optimal control for electric water heaters compared to traditional thermostat control, when the hot water usage is predicted using the probabilistic hot water usage model fitted to past hot water usage measurements. The energy savings and number of cold events when using the predicted water usage are compared to the energy savings and number of cold events when using perfect foreknowledge of the hot water usage. Section 7.3.1 evaluates the results of optimal planning for a two-node EWH model using A* and simulating it on a two-node EWH model with the implementation of EWH feedback. Section 7.3.2 compares the simulation results of a two-node EWH model that uses the optimal planning for predicted and perfectly known hot water usage profiles. The effect of this on EWH optimisation is evaluated and discussed. The predicted water usages that are used by the optimiser to produce the optimal plan is generated by the hot water demand predictor developed in Chapter 3. The predictor predicts water usages for each day of a full week using the probabilistic hot water usage models produced from three weeks of actual hot water usage. The EWH simulation uses a week of unforeseen actual hot water usage which was not used in the development of the probabilistic hot water usage models for the same EWH.

7.3.1 Two-node EWH model with A* optimal planning

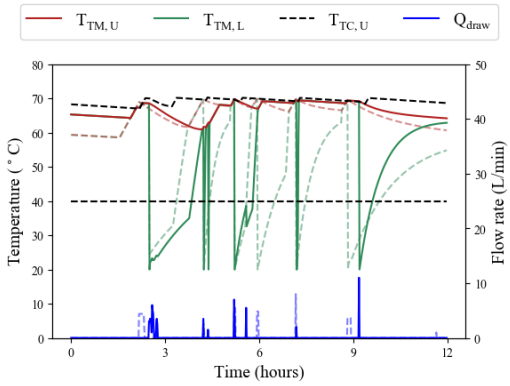
Illustrative example

Figure 7.8 shows example results for a single EWH where the optimal planning is first produced for predicted water usages and the EWH simulates it on the actual hot water usages. Figures 7.8 (a) and (c) shows the optimal heating schedule and temperature trajectory for a predicted water usage profile for TM and EM, respectively. Figure 7.8 (b) and (d) shows the corresponding simulated EWH temperatures for the actual water usages (solid line temperatures and water events) as a result of the temperature feedback controller and water mixer which try to guide the simulation along the optimal plan for the predicted water usages (dashed line temperatures and water events). All the figures show the upper node temperature for TC (dashed black line), which is simulated for the predicted water profile in (a) and (c) and the actual water profile in (b) and (d).

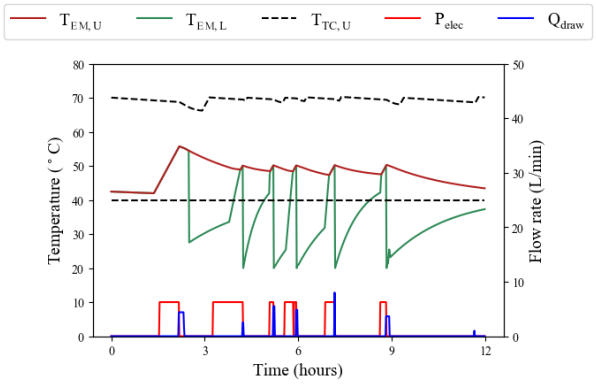
Optimal temperature matching (TM): During the simulation, the simulated TC temperatures differ from those used in the optimal planning due to the predicted water usages differing from the unforeseen actual usages. This means that the simulated temperatures will not necessarily match perfectly with those of TC before water usages. The effect of temperature mismatching is reduced by generating a predicted hot water usage profile with conservative start times so that the simulated temperature is at the set point temperature before the actual water usage can be expected and also by predicting a con-



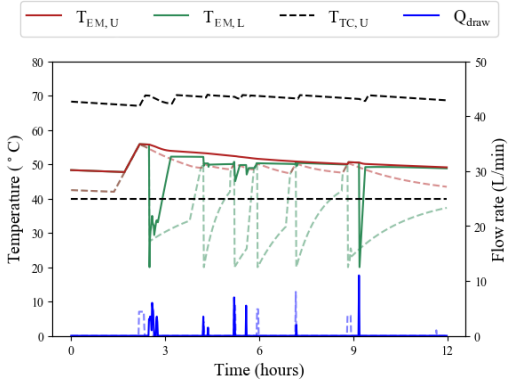
(a) Temperature matching optimal planning



(b) Temperature matching simulation



(c) Energy matching optimal planning



(d) Energy matching simulation

Figure 7.8: Simulation results for TC, TM, and EM, respectively. The plots show the EWH upper and lower node temperature and the outlet flow rate. (a) and (c) show the optimal planning for predicted water usages for TM and EM, respectively. (b) and (d) show the simulated temperatures and actual water usages (solid lines) and the optimal temperature trajectory and predicted water usage repeated from the previous plot (dashed lines) for TM and EM, respectively. The EWH temperature for thermostat control is repeated on all plots for comparison. The cold event threshold (40 °C) is indicated with a horizontally dashed black line.

servative amount of water usages for each day so that the temperature is prepared for the worst case of actual water usages. The water mixer will adjust the outlet flow rate so that a correct amount of thermal energy is delivered if the simulated temperature did not reach that of TC before usage. These safety measures are expected to reduce the electrical energy savings of the control strategy.

Optimal energy matching (EM): Unlike the optimal planning used in the previous sections for EM, a 10 °C safety margin results in the optimal temperature not falling below 50 °C during water usage. The purpose of this safety measure is to minimise the risk of cold events. The reduced risk is evident by the simulation temperatures which show that the outlet temperature can drop by a large extent from an unexpected water usage if a cold event were to occur.

Optimal energy matching with Legionella prevention (EML): Although EML control is not shown in the plots, it produces similar results to EM except for the tank heating to

60 °C at the start of the day.

Statistical results

The statistical results from all of the simulations performed for all 77 EWHs are discussed in this section. The effects of the different control strategies on energy and temperature for water heaters using a two-node model for optimal planning of predicted water usages are summarised in Table 7.5 and shown in Figure 7.9. Figure 7.9 (b) shows that all the control strategies delivered the same amount of thermal energy as TC. This confirms that the water mixer was successful in adjusting the outlet flow rate for perfect energy matching for all the strategies.

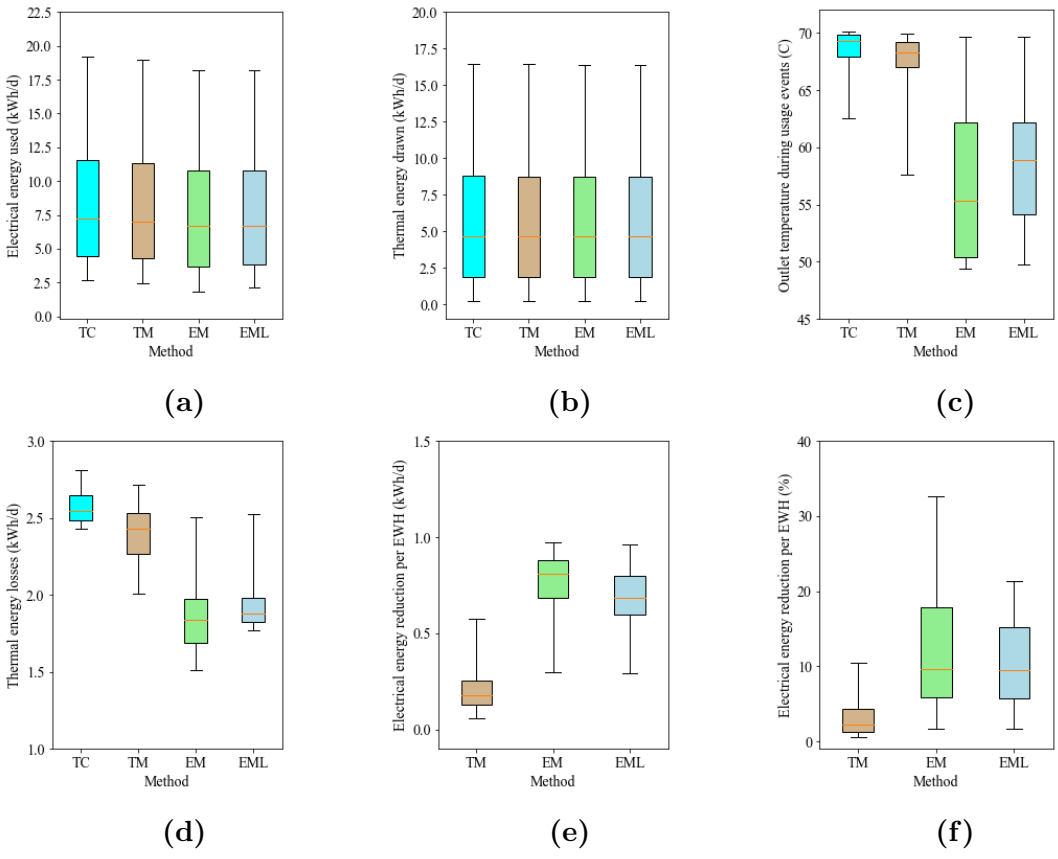


Figure 7.9: Energy and temperature results of the different control strategies represented as distributions for all water heaters. (a) depicts electrical energy used per EWH per day, (b) depicts thermal energy drawn per EWH per day, (c) depicts outlet temperatures during usage events, (d) depicts thermal losses per EWH per day, and (e) and (f) depict the savings achieved in electrical energy per EWH as a reduction in kWh per day and percentage of total used, respectively.

Temperature-matched optimisation

Figure 7.9 (a) shows that the median electrical energy used for TM was 7.0 kWh/day, 0.3 kWh/day (4.1 %) less than the 7.3 kWh/day median for TC. The median outlet temperature during events for TM dropped from 69.3 °C for TC to 68.4 °C, as shown in Figure

Table 7.5: Energy, temperature, volume, and cold event results for two-node water heaters with planning based on predicted water usages.

		TC	TM	EM	EML
$\overline{V}_{\text{hot}}$	(L/day)	119	123	150	142
E_{elec}	(kWh/day)	7.3	7.0	6.7	6.7
E_{draw}	(kWh/day)	4.7	4.7	4.7	4.7
E_{loss}	(kWh/day)	2.6	2.4	1.8	2.0
$\overline{T}_{\text{usage}}$	(°C)	69.3	68.4	55.3	58.9
ΔE_{elec}(kWh)	(kWh/day)	–	*0.1*, **0.2**, *0.3*	*0.7*, **0.8**, *0.9*	*0.6*, **0.7**, *0.8*
ΔE_{elec}(%)	%	–	*1.3*, **2.2**, *4.3*	*5.9*, **9.6**, *17.9*	*5.9*, **9.6**, *15.2*
Cold events*		*0*, **0**, *0*	*0*, **0**, *0*	*0*, **0**, *0*	*0*, **0**, *0*

Note: The distributions are reported as 25^{th} *percentile,* ***median****,* 75^{th} *percentile*
Cold events are taken as the total cold events for the EWH that uses the* 25^{th} *percentile,* ***median**, and* 75^{th} *percentile of the total volume of water used.*

7.9 (c). The difference is not significant but it is caused by the lack of perfect temperature matching as a result of water usage predictions in the optimal planning. The temperature mismatching is also evident in the table, where the average volume of water drawn increased to that of TC from 119 L to 123 L as a result of the mixer. Figure 7.9 (d) shows that the median thermal losses for TM was 2.4 kWh/day (0.2 kWh/day less than TC). Unlike the previous sections, the number of cold events increased from that of TC. Although the increase was not serious, the number of cold events increased from 115 of TC to 141 of the 15 581 events. However, the additional cold events only occurred for EWHs that already had cold events: only five of the 77 EWHs had a cold event increase. The EWH with the biggest increase in cold events grew from 19 to 26 cold events of the 284 events (a 2.46% increase of the total events for the EWH). Looking at the distribution of savings in Figure 7.9 (e) and (f), the electrical energy reduction for TM was [*0.1*, **0.2**, *0.3*] kWh/day, or [*1.3*, **2.2**, *4.3*] %.

Energy-matched optimisation

The average outlet temperature gap between TC and EM is significantly smaller because of the 10 °C safety margin imposed on the constraints. This explains why the outlet temperature does not drop below 50 °C and the median outlet temperature for EM is 55.3 °C. These increased outlet temperatures also cause the water mixer to adjust the outlet flow rate less to obtain the correct delivery of thermal energy. The average volume draw is 150 L, a substantial decrease from 231 L of the previous section for EM when water usages were provided with perfect foreknowledge. The median electrical energy used for EM was 6.7 kWh/day and, even though the electrical energy savings are dampened by the safety margins, a promising reduction in energy from TC is [*0.69*, **0.81**, *0.88*] kWh/day and [*5.9*, **9.6**, *17.9*] % and the best savings is as much as 34.0%. These savings are contributed to by the median standing losses of 1.8 kWh/day which is 0.8 kWh/day less than that of TC. The number of cold events increased from 115 of TC to 145, showing similar outcomes to TM with only four additional cold events.

***Legionella* control**

The results of EML shows small differences from that of EM. This can be explained by the safety margin of 50 °C which largely reduces the significance of the additional electrical energy required for the tank to reach 60 °C each day. The median outlet temperature during usage increased to 58.9 °C (a 3.6 °C increase from EM). The electrical energy used for this control strategy is approximately equal to that for EM. The electrical energy reduction was [*5.9*, **9.6**, *15.2*] % and [*0.6*, **0.7**, *0.8*] kWh/day, where only the 75$^{\text{th}}$ percentile was different from EM. The number of cold events was the same as for EM.

7.3.2 Effects of predicting water usages for optimisation

The effect of predicting water usages is determined by comparing the results that used the A* optimal planning which used perfect foreknowledge with the optimal planning that used predicted hot water usage profiles. The statistical results from all of the simulations performed for all 77 EWHs are discussed in this section. A metric is defined to asses the performance of the relative EWH-specific changes during the simulation. The change in electrical energy used per day is calculated using

$$P^{\Delta N_p}_{\text{elec}|h} = P^{P}_{\text{elec}|h} - P^{A}_{\text{elec}|h} \quad \text{kWh/day} \tag{7.7}$$

where the superscript ΔN_p refers to the difference between the types of planning and superscripts P and A refer to the simulation results that plan for predicted and actual water usage profiles, respectively. Similar modifications are applied to the formulas that calculate difference in thermal energy drawn per day, average outlet temperature during usage events, thermal energy losses per day and energy savings per day (kWh and percentage reduction). These results indicate the effect of predicting hot water usages when determining the optimal plan for an EWH. The simulation results for TM, EM and EML obtained in Figure 7.5 for the optimal planning produced for a two-node EWH model with perfect foreknowledge of the water usages is compared with those obtained in Figure 7.9 for the optimal planning of predicted water usages.

Figure 7.10 (b) shows that the median EWH delivered the same amount of thermal energy as TC for all the control strategies, with small deviations between the two types of planning, and shows the success of the mixer. Figure 7.10 (c) shows that most of the outlet temperatures for TM with predicted water usage planning were perfectly matched to that of the results for simulations that had perfect water usage knowledge, which can also assume to be matched to TC. The small changes of the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of 2.0 °C and 0.8 °C respectively show the extent of the temperature mismatching to that of TC. For EM and EML, the median outlet temperature increases of 10.2 °C and 4.1 °C, respectively, when water usages were predicted in the planning shows the impact of the safety margin is much harsher on EM. This is shown in Figure 7.10 (d) where the median standing losses increase for EM and EML were 0.72 kWh/day and 0.33 kWh/day, respectively.

Figure 7.10 (a) shows the median increase in electrical energy usage for TM, EM and EML were 0.35 kWh/day (4.5%), 0.74 kWh/day (19.6%) and 0.4 kWh/day (11.6%), respectively, when planning the predicted water usages. This results in the final electrical energy saving decreases relative to TC which are shown in Figure 7.10 (e) and (f). The

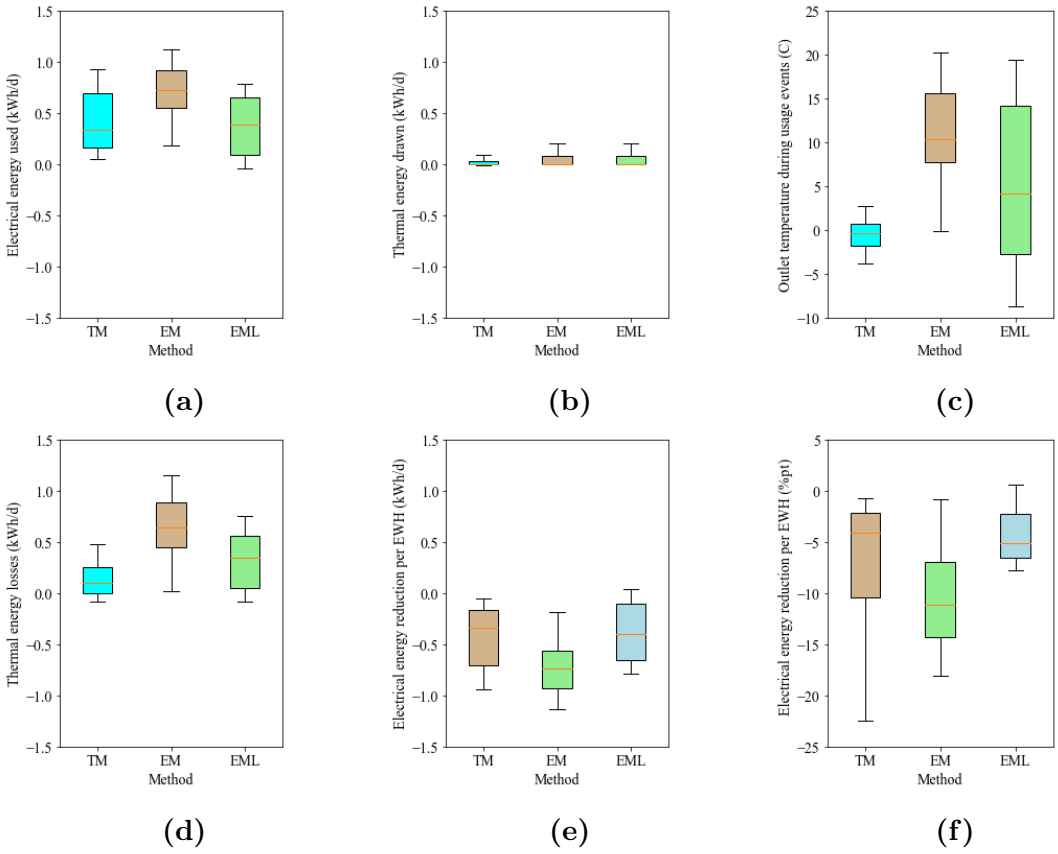


Figure 7.10: The change in energy and temperature results of the different control strategies represented as distributions for all water heaters, where the results from the perfectly known water usage planning is subtracted from that of the corresponding predicted results using Equation 7.7. (a) depicts difference in electrical energy used per EWH per day, (b) depicts difference in thermal energy drawn per EWH per day, (c) depicts difference in outlet temperatures during usage events, (d) depicts difference in thermal losses per EWH per day, and (e) and (f) depict the difference in savings achieved in electrical energy per EWH as a reduction in kWh per day and percentage points of total used, respectively.

decreases in electrical energy savings due to the optimal planning predicting water usages is [*0.22*, **0.31**, *0.72*] kWh/day, or [*2.30*, **4.11**, *10.31*] percentage points for TM; [*0.57*, **0.72**, *0.90*] kWh/day, or [*7.1*, **11.0**, *14.0*] percentage points for EM; and lastly, [*0.09*, **0.37**, *0.65*] kWh/day, or [*2.33*, **5.11**, *6.31*] percentage points for EML.

Discussion of the effects of predicting hot water usages in the planning

An evaluation of the results showed that the electrical energy usage increased for all the control strategies when the optimal planning considers predicted water usages as opposed to that of the actual hot water profile. For TM, the median electrical energy savings decreased from 0.6kWh/day (6.3 percentage points) when perfect foreknowledge was used to 0.2kWh/day (2.2 percentage points) when predictions were used relative to the baseline TC strategy. This shows that TM does not save a lot when the water usage profile is predicted, as a result of safety measures which are implemented to minimise the risk of cold events. For EM and EML, the median electrical energy usage savings

decreased from 1.6kWh/day (21.9 percentage points) and 1.2kWh/day (16.2 percentage points) when perfect foreknowledge was used to 0.8kWh/day (9.6 percentage points) and 0.7kWh/day (9.6 percentage points) when predictions were used. This ultimately shows that even though there were huge energy saving reductions when the safety measures were imposed on the system, the energy matching strategies can still save close to 10 percentage points of electrical energy each day as well as preventing the growth of *Legionella*.

The number of cold events that occurred did increase for the three control strategies, however of the total 15 581 events, the increase was 26 for TM and 30 for EM and EML. This result is insignificant as they only occurred for five of the 77 EWHs and those EWHs already showed evidence of cold events occurring due to their heavy usages in TC. Because the outlet temperatures for TM did not deviate far from that of TC, it shows that these heavy water usage profiles were on the verge of having more cold events than what was counted. Although these results represent the practical savings, they do not account for data acquisition and state estimation that would be required for a real-world implementation.

7.4 Summary

This section summarises how much energy can be saved, compared to the standard TC, from optimal heating strategies performed on 77 EWHs. The results of simulations for a one-node EWH where the optimal planning was produced using DP can **ideally** save

- 0.6kWh/day (7.6%) for TM,
- 1.5kWh/day (21.7%) for EM and
- 1.0kWh/day (16.4%) for EML

and the results for a two-node EWH where the optimal planning was produced using A* can **ideally** save

- 0.6kWh/day (6.3%) for TM,
- 1.6kWh/day (21.9%) for EM and
- 1.2kWh/day (16.2%) for EML

with perfect foreknowledge of the hot water usage. When comparing the results for a two-node EWH that performed simulations with the optimal planning produced for both a one- and two-node EWH, accounting for stratification can **ideally** increase the savings by

- 0.08kWh/day (1.2 percentage points) for TM,
- 0.4kWh/day (5.4 percentage points) for EM and
- 0.4kWh/day (5.5 percentage points) for EML.

The results of simulations for a two-node EWH where the optimal planning was produced using A* can **practically** save

- 0.2kWh/day (2.2%) for TM,
- 0.8kWh/day (9.6%) for EM and
- 0.7kWh/day (9.6%) for EML

with predictions of the hot water usage. Based on the results, the best performing optimal heating strategy in terms of energy savings was EM for all the simulations. The number of cold events that occurred only increased from that of TC for simulations where the water usages were predicted. From these results, TM had the smallest increase in cold events from TC. However, this is negligible as there were only four cold events less than EM and EML.

Chapter 8

Conclusion

Existing work related to the schedule optimisation of an EWH does not account for stratification and *Legionella* sterilisation, they use data with low temporal resolution and the results are not evaluated with fair comparisons to traditional EWH control. Moreover, if the optimal planning is to be used practically (i.e. this requires water usages to be predicted), there is no existing hot water usage model that can generate water profiles that accounts for all the factors that influence it. This thesis investigated solutions that fills these gaps in research and evaluated the potential electrical energy savings that can be achieved. Furthermore, optimal (schedule and temperature) control is applied to 77 energy-storing EWHs over a period of four weeks (one week for each season).

The probabilistic model we present in this thesis is an original contribution that statistically models the daily hot water usage profile for individual EWHs based on historically-measured flow rate data. This model can be used to model hot water demand and the resulting residential energy demand. We also present a hot water usage simulator that uses the probabilistic model to generate simulated hot water usage profiles (flow rate versus time). Our model can generate high-resolution water profiles for specific applications and is fully automated. It excels in its ability to characterise a household's water profile, taking into consideration differences between users and temporal variations such as the season of the year and the day of week. Our results confirm that the model produces simulated hot water profiles that closely match the measured data and are representative enough to be used in energy management strategies for modelling EWHs. Furthermore, the model's results improve on those of existing hot water usage models for representing the energy drawn from the grid.

The hot water usage simulator can be used as a component of a larger EWH simulator to improve the efficiency of energy systems such as water heaters and strategies such as demand-side management. The efficiencies can be further improved if more measured water flow data is available to increase the accuracy of forecasting. The availability of more measured data could also provide enough information to fit more accurate probability density functions to the clustered water usages. The model's unique visualisation of the data offers ideas that could be productive for further research in the field.

This thesis describes our novel formulation of the EWH control problem with the objectives to minimise the electrical energy used. This allowed for two novel optimisation algorithms, DP and A*, to be applied to the water heater that does and does not include

stratification, respectively. Because of model uncertainties and predicted water usages, a novel feedback control technique was developed to prevent the EWH temperature from straying away from the optimal plan. We used as our baseline TC (always on), which is the default mode of water heaters, and compared this strategy with three optimal control strategies. Our three strategies, using DP and A* to ensure optimal heating, ensure comparable delivery at the outlet of the water heater. The first strategy ensures a temperature-matched output with equal volume; the second provides an energy-matched output with decreased but still-hot temperature and increased volume; and the third adapts the second method to ensure that *Legionella* is sterilised despite the lower temperatures. A reactive hot water usage simulation model that can generate synthetic hot water usage profiles with random variations was developed to obtain the results. The model also simulates the user who will adjust the mixing of hot and cold water in reaction to the temperature.

When simulations used optimal planning for a one-node EWH model produced from DP with perfect foreknowledge of water usages, it was found that temperature-matching gave a median energy saving of 6.7 %, without adversely affecting the temperature at which water is delivered; the energy-matching method gave a median saving of 26.4 %, with a reduction in energy lost to the environment as well as a reduction in energy lost due to unintentional use; and the energy-matching method with daily *Legionella* sterilisation gave a median saving of 19.6 %. For all three strategies, the number of cold events did not increase from the baseline strategy.

When simulations used optimal planning for a two-node EWH model produced from A* with perfect foreknowledge of water usages, it was found that temperature-matching gave a median saving of 6.3 %, without adversely affecting the temperature at which water is delivered; the energy-matching method gave a median saving of 21.9 %, with a reduction in energy lost to the environment; and the energy-matching method with daily *Legionella* sterilisation gave a median saving of 16.2 %. For all three strategies, the number of cold events did not increase from the baseline strategy.

We also compared the results of the three heating strategies when the optimal schedule was planned with and without stratification taken into account. We found that there was no substantial difference for temperature-matching, where the median saving was 1.2%. However, the energy-matching method gave a median saving decrease of 5.4% when stratification was not considered in the planning; and the energy-matching method with daily *Legionella* sterilisation gave a median saving decrease of 5.5%. This was due to the fact that the planning without stratification assumes the temperature of the whole tank is equal to the outlet temperature, and therefore the temperature feedback controller will force the whole tank to satisfy the usage temperature constraints opposed to just the hot water portion of the tank.

The final and most important results were obtained when simulations used optimal planning produced from A* for predicted water usages produced from the hot water demand predictor. It was found that temperature-matching gave a median saving of 2.2 %, without adversely affecting the temperature at which water is delivered; the energy-matching method gave a median saving of 9.6 %; and the energy-matching method with daily *Legionella* sterilisation gave a median saving of 9.6 %. For all three strategies, the number

of cold events did increase from the baseline strategy, but the increase can be considered as negligible. The energy-matching method proved to be to best heating control strategy as it significantly saves the most energy for all the simulations.

In the future, this system could be implemented in a real world scenario to be practically used by consumers to achieve actual energy savings. The probabilistic hot water usage model and EWH optimiser could be implemented as part of a SEC directly attached to the EWH. Sensor measurements could be locally measured to generate, and continuously improve the hot water usage model of the tank. The optimiser can produce the optimal heating schedule to directly control the heating element switching state. Alternatively, the SEC could communicate with an external server, where a PC can provide more computational power for solving the necessary calculations. The simulations of the final system could also use real ambient and inlet water temperature measurements instead of making the assumption that it is constant.

By using both real world and synthesised hot water usage profiles, we determined the absolute best energy savings, with the fewest cold events, that can be achieved with scheduling and temperature control while causing the minimum of inconvenience to the consumer.

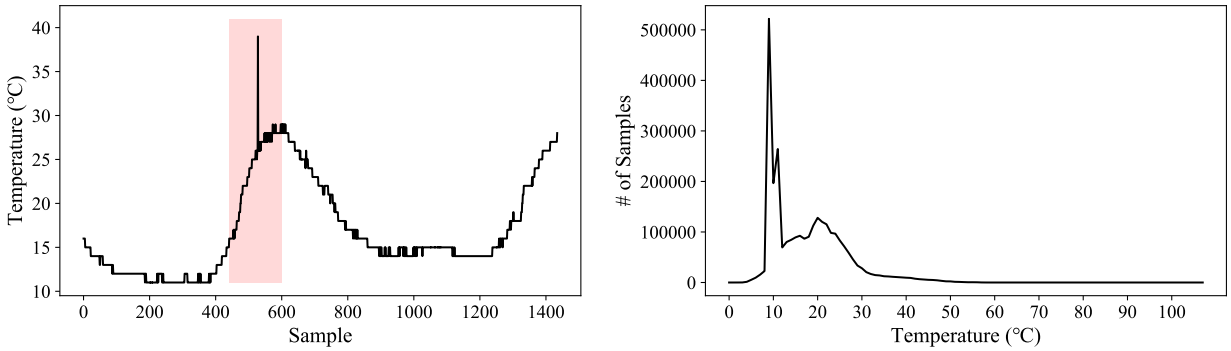
Appendices

Appendix A

Data cleaning

A.1 Data analysis

Ambient temperature: Figure A.1a shows an ambient temperature profile for raw data. Ambient temperature does not change instantaneously and follows a gradual cyclic motion of high and low temperature. The region marked in red, however, indicates an instantaneous change in temperature of over 10°C. Figure A.1b shows the tally of ambient temperatures contained in all raw data profiles. Considering every possible sample, the data ranges from 3 to 60°C which is within the expected range defined previously. The massive spike between 9 and 13°C is due to the fact that if the SEC cannot read an ambient temperature, it will send a reading within this range.

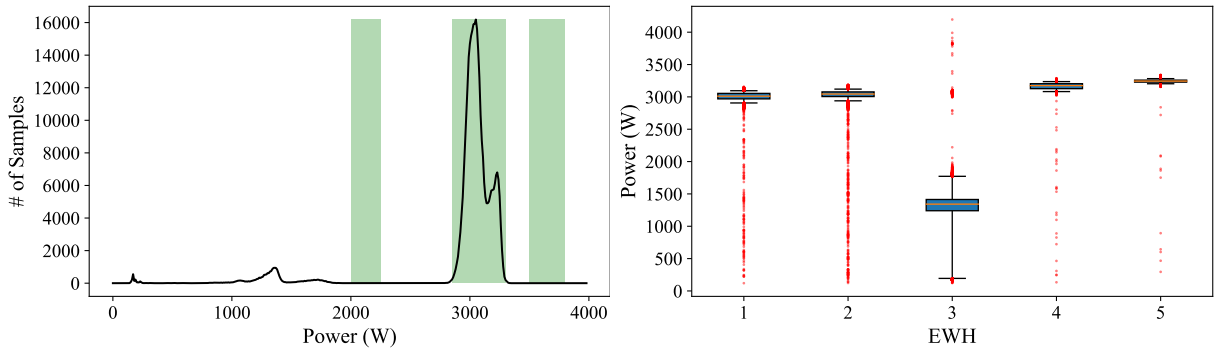


(a) Ambient temperature profile containing an invalid spike.

(b) Sample count for each ambient temperature value.

Figure A.1: Representations of valid and invalid ambient temperature data.

Power: A data sample for power can only be one of two values, either the rated power of the EWH element or zero. Figure A.2a shows a tally of different power measurements for all the raw profiles to identify all the heating element power ratings that are used by the EWHs. It is shown that there are power ratings of 2.1kW, 3.1kW and 3.6kW. Figure A.2b shows box-and-whisker plots of the distributions of the power measurements for 10 EWHs. The red dots represent outlier measurements. Outliers above the median are regarded as an invalid measurement. Outliers below the median are not invalid, as these values are obtainable if the heating element was not turned on for the full minute. EWH 3 shows a profile with a wide spread of invalid measurements and has a distinct power rating of 1.3kW.



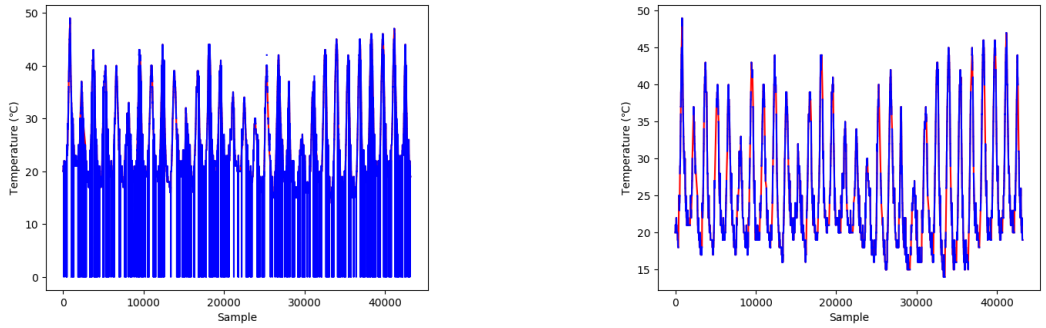
(a) Sample count for each power value.

(b) Distribution of power measurements for different EWHs.

Figure A.2: Representations of power measurements and power ratings.

A.2 Results

Ambient temperature: Figures A.3a and A.3b show the ambient temperature profile time-sample versus temperature for the original and interpolated sections. It can be seen from Figure A.3a that there is a frequent occurrence of zeroes in the profile and Figure A.3b shows the profile after zeroes are replaced with interpolated values.



(a) Including zero measurements.

(b) Not including zero measurements.

Figure A.3: Time-sample versus temperature of original ambient temperature profile (blue) and portions reconstructed by interpolation (red).

Power: Figure A.4 shows the results of interpolation for a power profile. This figure shows both how zeros are filled where missing as well as how interpolation occurs when data is missing in periods where the element is on. The maximum power $P_{elec(max)}$ is indicated in the figure with a black line.

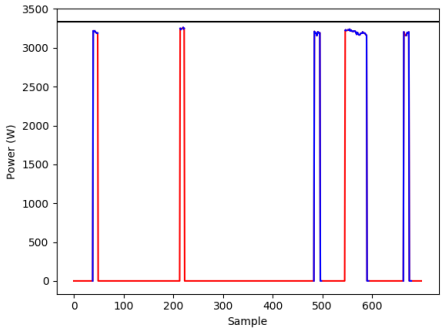


Figure A.4: Time-sample versus power of original power profile (blue) and portions reconstructed by interpolation (red).

Appendix B

Vol/Day per EWH for each season

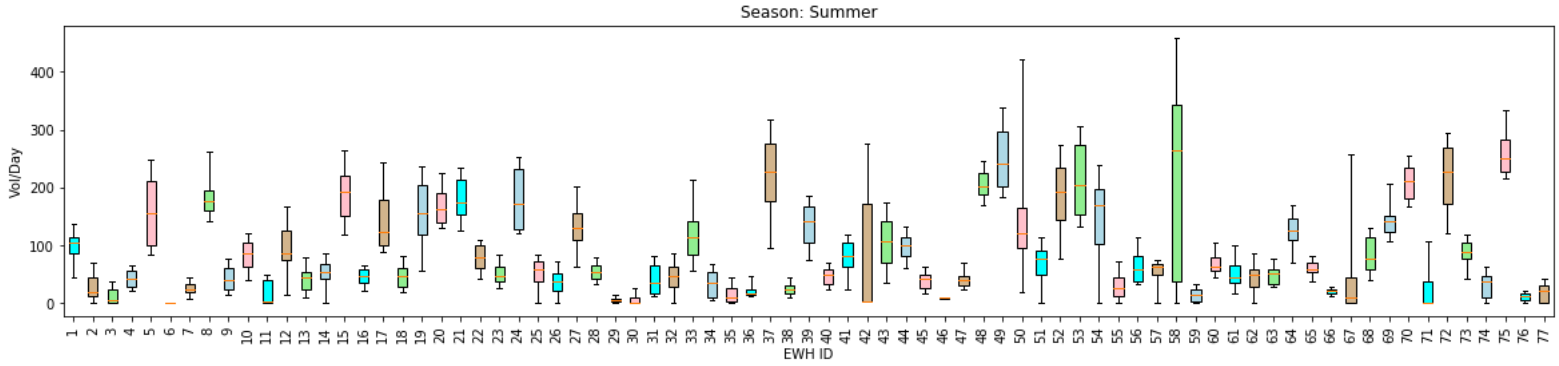


Figure B.1: Spread of volume used each day per EWH for Summer.

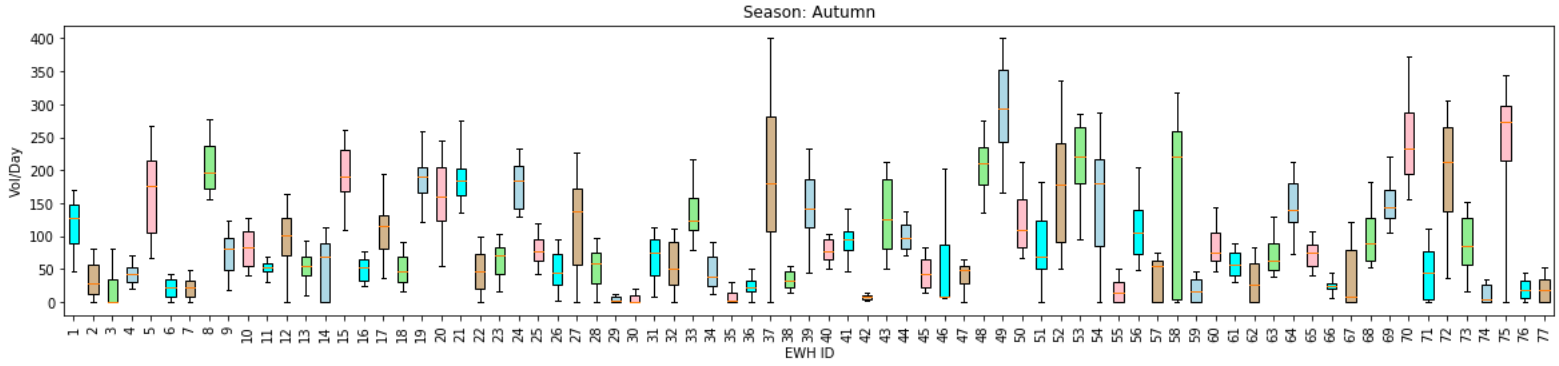


Figure B.2: Spread of volume used each day per EWH for Autumn.

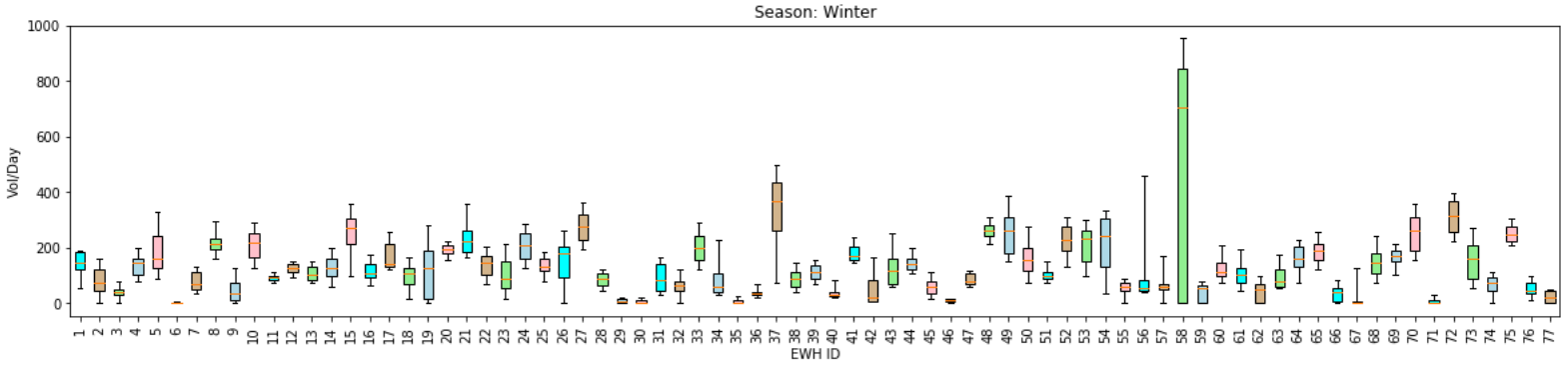


Figure B.3: Spread of volume used each day per EWH for Winter.

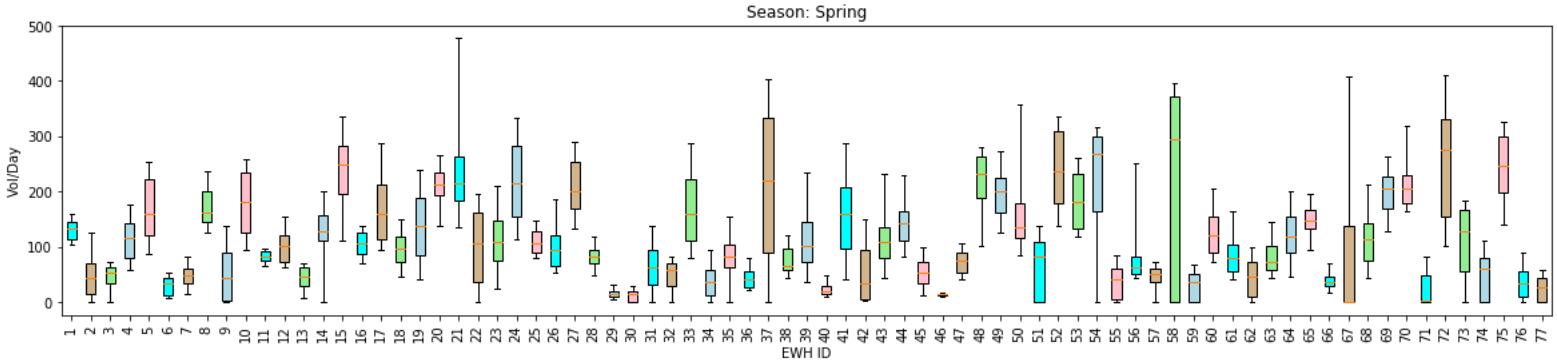


Figure B.4: Spread of volume used each day per EWH for Spring.

Appendix C

EWH Optimised Temperature Profile

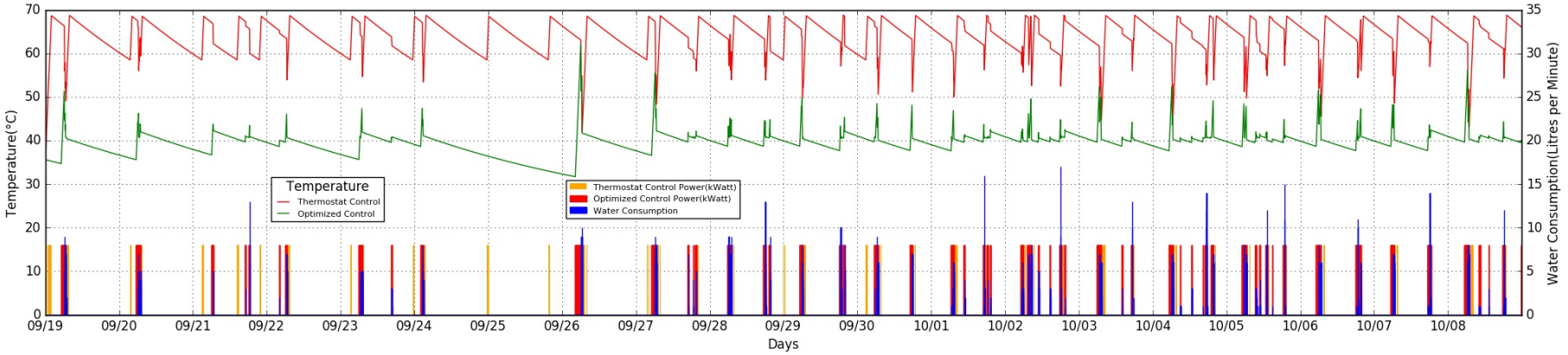


Figure C.1: Temperature and heating element (rated at 3kWh) plot of TC and OC for simulation of a hot water usage profile.

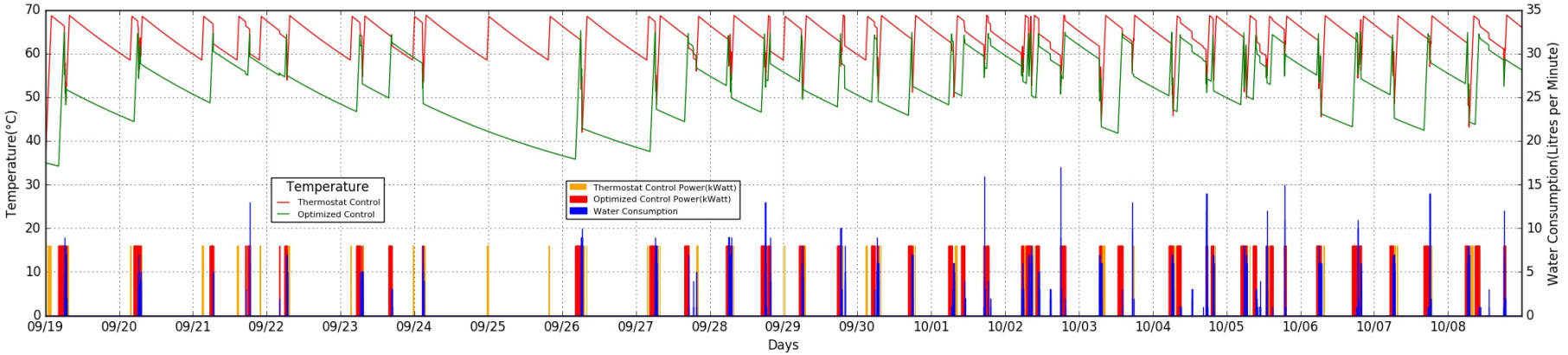


Figure C.2: Temperature and heating element (rated at 3kWh) plot of TC and TM for simulation of a hot water usage profile.

References

[1] M. J. Booysen, J. A. A. Engelbrecht, M. J. Ritchie, M. Apperley, and A. H. Cloete, "How much energy can optimal control of domestic water heating save?" *Energy for Sustainable Development*, vol. 51, pp. 73–85, 2019.

[2] M. J. Ritchie, J. A. A. Engelbrecht, and M. J. Booysen, "A probabilistic hot water usage model and simulator for use in residential energy management," *Energy and Buildings*, 2020.

[3] J. Engelbrecht, M. J. Ritchie, and M. Booysen, "Optimal schedule and temperature control of stratified water heaters," *Energy for Sustainable Development*, vol. 62, pp. 67–81, 2021.

[4] M. J. Ritchie, J. A. Engelbrecht, and M. J. Booysen, "Practically-achievable energy savings with the optimal control of stratified water heaters with predicted usage," *Energies*, vol. 14, no. 7, p. 1963, 2021.

[5] N. K. Arora and I. Mishra, *United nations sustainable development goals 2030 and environmental sustainability: Race against time*, 2019.

[6] *Global Energy Statistical Yearbook 2018*, https://yearbook.enerdata.net/#electricity-domestic-consumption-data-by-region.html, Accessed: Feb 2019, 2018.

[7] P. Hohne, K. Kusakana, and B. Numbi, "A review of water heating technologies: An application to the South African context," *Energy Reports*, vol. 5, pp. 1–19, 2019. DOI: 10.1016/j.egyr.2018.10.013.

[8] T. Skinner *et al.*, "An overview of energy efficiency and demand side management in South Africa," in *Presentation to the world bank/IFC workshop on appropriate incentives to deploy renewable energy and energy efficiency, Washington, DC. Available online: www. eskom. co. za*, 2012. [Online]. Available: https://www.esi-africa.com/wp-content/uploads/Tom_Skinner.pdf.

[9] K. Mathu, "Cleaning South Africa's coal supply chain," *Journal of Business Diversity*, vol. 17, no. 3, 2017. DOI: 10.33423/jbd.v17i3.1236.

[10] Nersa, *Nersa*, http://www.nersa.org.za, Accessed: January 2019, 2018.

[11] C. Forlee, "Water heating in South Africa facts and figures from the 1997" notch testing" program," in *5th International Domestic Use of Electrical Energy Conference, Cape Town*, 1998, pp. 265–270.

[12] N. Beute and G. Delport, "An historic overview of controlling domestic water heating," in *Proceedings of the 14th Domestic Use of Energy Conference*, Citeseer, 2006, pp. 41–46.

[13] *Eskom Geyser fact sheet 2013*, http://www.eskom.co.za/sites/IDM/Documents/1317_geyser_fact_sheet_no_rmr.pdf, Accessed: May 2018.

[14] Y.-C. J. Yen, K. J. Nixon, and W. A. Cronjé, "Maximum demand comparison of aggregated load profiles for vertical and horizontal ewhs due to a fixed draw event," *SAIEE Africa Research Journal*, vol. 110, no. 3, pp. 167–174, 2019.

[15] J. E. Stout, M. G. Best, and V. L. Yu, "Susceptibility of members of the family legionellaceae to thermal stress: Implications for heat eradication methods in water distribution systems," *Applied and environmental microbiology*, vol. 52, no. 2, pp. 396–399, 1986. DOI: 10.1016/j.semcdb.2010.11.003.

[16] A. Sepulveda, L. Paull, W. G. Morsi, H. Li, C. Diduch, and L. Chang, "A novel demand side management program using water heaters and particle swarm optimization," in *2010 IEEE Electrical Power & Energy Conference*, IEEE, 2010, pp. 1–5.

[17] P. Kepplinger, G. Huber, and J. Petrasch, "Demand side management via autonomous control-optimization and unidirectional communication with application to resistive hot water heaters," *ENOVA: Eisenstadt, Austria*, vol. 8, 2014.

[18] ——, "Autonomous optimal control for demand side management with resistive domestic hot water heaters using linear optimization," *Energy and Buildings*, vol. 100, pp. 50–55, 2015.

[19] ——, "Field testing of demand side management via autonomous optimal control of a domestic hot water heater," *Energy and Buildings*, vol. 127, pp. 730–735, 2016.

[20] K. Al-Jabery, Z. Xu, W. Yu, D. C. Wunsch, J. Xiong, and Y. Shi, "Demand-side management of domestic electric water heaters using approximate dynamic programming," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 36, no. 5, pp. 775–788, 2016.

[21] H. Galhardas, D. Florescu, D. Shasha, E. Simon, H. Galhardas, D. Florescu, D. Shasha, E. Simon, and A. Extensible, "An Extensible Framework for Data Cleaning," 2006.

[22] *Geasy - a smart geasy controller by bridgiot*, https://www.bridgiot.co.za/solutions/geasy-2/, Accessed: 2019-09-17.

[23] A. H. Cloete, "A domestic electric water heater application for smart grid," PhD thesis, Stellenbosch: Stellenbosch University, 2017.

[24] *Mqtt faq*, http://mqtt.org/faq, Accessed: 2019-09-17.

[25] J. W. K. Brown, "Design, and implementation of an intelligent water heater control module for feedback demand side management," PhD thesis, Stellenbosch: Stellenbosch University, 2016.

[26] L. Li, "Data quality and data cleaning in database applications," no. September, p. 242, 2012. [Online]. Available: https://pdfs.semanticscholar.org/f551/918d57f83f4092e291378f567b25bc614e29.pdf.

[27] T. C. Redman, "The impact of poor data quality on the typical enterprise," *Communications of the ACM*, vol. 41, no. 2, pp. 79–82, 2002. DOI: 10.1145/269012.269025.

[28] *A Progress Report From MIT Information Quality Conference — TDAN.com*, Accessed: 2019-03-08. [Online]. Available: http://tdan.com/a-progress-report-from-mit-information-quality-conference/5143.

[29] H. Müller and J. Freytag, "Problems, Methods, and Challenges in Comprehensive Data Cleansing," *Challenges*, no. HUB-IB-164, pp. 1–23, 2003. [Online]. Available: http://www.dbis.informatik.hu-berlin.de/fileadmin/research/papers/techreports/2003-hub%7B%5C_%7Dib%7B%5C_%7D164-mueller.pdf.

[30] S. Gato, N. Jayasuriya, and P. Roberts, "Forecasting residential water demand: Case study," *Journal of Water Resources Planning and Management*, vol. 133, no. 4, pp. 309–319, 2007.

[31] S. Zhou, T. McMahon, A. Walton, and J. Lewis, "Forecasting operational demand for an urban water supply zone," *Journal of hydrology*, vol. 259, no. 1-4, pp. 189–202, 2002.

[32] *Housecheck geyser installation*, https://www.housecheck.co.za/hot-water-geysers/, Accessed: 2019-09-17.

[33] *Pressure control valves*, https://www.specifile.co.za/wp-content/uploads/2017/06/COBR029-FEB13-PDF.pdf, Accessed: 2019-09-17.

[34] M. Roux, M. Apperley, and M. J. Booysen, "Comfort, peak load and energy: Centralised control of water heaters for demand-driven prioritisation," *Energy for Sustainable Development*, vol. 44, pp. 78–86, 2018. DOI: 10.1016/j.esd.2018.03.006.

[35] C. Matos, I. Bentes, S. Pereira, D. Faria, and A. Briga-Sa, "Energy consumption, CO2 emissions and costs related to baths water consumption depending on the temperature and the use of flow reducing valves," *Science of The Total Environment*, vol. 646, pp. 280–289, 2019. DOI: 10.1016/j.scitotenv.2018.07.290.

[36] P. J. C. Nel, M. J. Booysen, and B. Van der Merwe, "Saving on household electric water heating: What works best and by how much?" In *2017 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*, IEEE, 2017, pp. 1–6. DOI: 10.1109/ISGT-Asia.2017.8378439.

[37] R. Diao, S. Lu, M. Elizondo, E. Mayhorn, Y. Zhang, and N. Samaan, "Electric water heater modeling and control strategies for demand response," in *2012 IEEE power and energy society general meeting*, IEEE, 2012, pp. 1–8. DOI: 10.1109/PESGM.2012.6345632.

[38] R. Spur, D. Fiala, D. Nevrala, and D. Probert, "Influence of the domestic hot-water daily draw-off profile on the performance of a hot-water store," *Applied Energy*, vol. 83, no. 7, pp. 749–773, 2006.

[39] M. Dongellini, S. Falcioni, and G. L. Morini, "Dynamic simulation of solar thermal collectors for domestic hot water production," *Energy Procedia*, vol. 82, pp. 630–636, 2015.

[40] W. Wang, I. Beausoleil-Morrison, M. Thomas, and A. Ferguson, "Validation of a fully mixed model for simulating gas fired water storage tanks," in *Proc. Building Simulation*, 2007.

[41] H. ASHRAE, "Applications handbook, 2003," *American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., USA*, 1791.

[42] M. W. Jack, K. Suomalainen, J. J. W. Dew, and D. Eyers, "A minimal simulation of the electricity demand of a domestic hot water cylinder for smart control," *Applied Energy*, vol. 211, pp. 104–112, 2018.

[43] C. Forlee, "Water heating in South Africa facts and figures from the 1997 "notch testing" program. Presented at 5th International Domestic Use of Electrical Energy Conference, Cape Town.," Available at https://drive.google.com/open?id=0B6M3SFQa2Dy_Q3ZnZk1RdTJMa00, 1998.

[44] M. Roux and M. J. Booysen, "Use of smart grid technology to compare regions and days of the week in household water heating," in *2017 International Conference on the Domestic Use of Energy (DUE)*, IEEE, 2017, pp. 276–283. DOI: 10.23919/DUE.2017.7931855.

[45] A. Bertrand, A. Mastrucci, N. Schüler, R. Aggoune, and F. Maréchal, "Characterisation of domestic hot water end-uses for integrated urban thermal energy assessment and optimisation," *Applied Energy*, vol. 186, pp. 152–166, 2017.

[46] F. Iglesias and P. Palensky, "Profile-based control for central domestic hot water distribution," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 697–705, 2013.

[47] L. Gelažanskas and K. Gamage, "Forecasting hot water consumption in residential houses," *Energies*, vol. 8, no. 11, pp. 12 702–12 717, 2015.

[48] J. Widén, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegård, and E. Wäckelgård, "Constructing load profiles for household electricity and hot water from time-use data—modelling approach and validation," *Energy and Buildings*, vol. 41, no. 7, pp. 753–768, 2009.

[49] M. J. Booysen, M. Visser, and R. Burger, "Temporal case study of household behavioural response to Cape Town's "Day Zero" using smart meter data," *Water Research*, vol. 149, pp. 414–420, 2019.

[50] O. Gerin, B. Bleys, and K. De Cuyper, "Seasonal variation of hot and cold water consumption in apartment buildings," in *Proceedings of CIB W062, 40th International Symposium on Water Supply and Drainage for Building (Sao Paulo, Brazil,)*, 2014, pp. 1–9. [Online]. Available: http://www.sisconev.com.br/Uploads/CIB2014/Trab012800001523201407 0_000000.docx.

[51] B. Howard, L. Parshall, J. Thompson, S. Hammer, J. Dickinson, and V. Modi, "Spatial distribution of urban building energy consumption by end use," *Energy and Buildings*, vol. 45, pp. 141–151, 2012.

[52] E. Fuentes, L. Arce, and J. Salom, "A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1530–1547, 2018.

[53] ASHRAE, "Ansi/ashrae standard 90.2, 1993," *American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.*, 1993.

[54] S. Edwards, I. Beausoleil-Morrison, and A. Laperrière, "Representative hot water draw profiles at high temporal resolution for simulating the performance of solar thermal systems," *Solar Energy*, vol. 111, pp. 43–52, 2015.

[55] B. Hendron, J. Burch, and G. Barker, "Tool for generating realistic residential hot water event schedules," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2010.

[56] U. Jordan and K. Vajen, "Realistic domestic hot-water profiles in different time scales," *Report for IEA-SHC Task*, vol. 26, 2001.

[57] S. G. Buchberger and G. J. Wells, "Intensity, duration, and frequency of residential water demands," *Journal of Water Resources Planning and Management*, vol. 122, no. 1, pp. 11–19, 1996.

[58] E. Blokker, J. Vreeburg, and J. Van Dijk, "Simulating residential water demand with a stochastic end-use model," *Journal of Water Resources Planning and Management*, vol. 136, no. 1, pp. 19–26, 2009.

[59] H. Scheepers and H. Jacobs, "Simulating residential indoor water demand by means of a probability based end-use model," *Journal of Water Supply: Research and Technology-AQUA*, vol. 63, no. 6, pp. 476–488, 2014.

[60] M. Richard, "Simulating domestic hot water demand by means of a stochastic end use model," PhD thesis, Stellenbosch: Stellenbosch University, 2016.

[61] R. Yao and K. Steemers, "A method of formulating energy load profile for domestic buildings in the UK," *Energy and buildings*, vol. 37, no. 6, pp. 663–671, 2005.

[62] M. Negnevitsky and K. Wong, "Demand-side management evaluation tool," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 212–222, 2014.

[63] H. Aki, T. Wakui, and R. Yokoyama, "Development of a domestic hot water demand prediction model based on a bottom-up approach for residential energy management systems," *Applied Thermal Engineering*, vol. 108, pp. 697–708, 2016.

[64] V. Bakker, A. Molderink, J. L. Hurink, and G. J. Smit, "Domestic heat demand prediction using neural networks," in *19th International Conference on Systems Engineering*, IEEE, 2008, pp. 189–194.

[65] N. D. Pflugradt, "Modellierung von wasser und energieverbräuchen in haushalten," 2016.

[66] D. Popescu and E. Serban, "Simulation of domestic hot-water consumption using time-series models," in *Proceedings of the 6th IASME/WSEAS International Conference on Heat Transfer, Thermal Engineering and Environment, Rhodes, Greece*, 2008, pp. 20–22.

[67] M. Roux, "Load management of electric water heaters in a smart grid through forecasting and intelligent centralised control," Master's thesis, Stellenbosch: Stellenbosch University, 2018.

[68] D. Marini, R. A. Buswell, and C. J. Hopfe, "A critical software review-how is hot water modelled in current building simulation?," 2015.

[69] P. Fairey and D. Parker, "A review of hot water draw profiles used in performance analysis of residential domestic hot water systems," *Florida Solar Energy Center*, no. 2, 2004.

[70] A. Gholizadeh and V. Aravinthan, "Benefit assessment of water-heater management on residential demand response: An event driven approach," in *2016 North American Power Symposium (NAPS)*, IEEE, 2016, pp. 1–6. DOI: 10.1109/NAPS.2016.7747831.

[71] R. H. Shumway and D. S. Stoffer, "Time series analysis and its applications," *Studies In Informatics And Control*, vol. 9, no. 4, pp. 375–376, 2000. DOI: 10.1007/978-3-319-52452-8.

[72] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999, ISSN: 0360-0300. DOI: 10.1145/331499.331504. [Online]. Available: http://doi.acm.org/10.1145/331499.331504.

[73] T. M. Kodinariya and P. R. Makwana, "Review on determining number of clusters in k-means clustering," *International Journal*, vol. 1, no. 6, pp. 90–95, 2013.

[74] A. N. Tarekegn, K. Michalak, and M. Giacobini, "Cross-validation approach to evaluate clustering algorithms: An experimental study using multi-label datasets," *SN Computer Science*, vol. 1, no. 5, pp. 1–9, 2020.

[75] P. J. C. Nel, "Rethinking electrical water heaters," Master's thesis, Stellenbosch: Stellenbosch University, 2015.

[76] P. J. C. Nel, M. J. Booysen, and B. van der Merwe, "A computationally inexpensive energy model for horizontal electric water heaters with scheduling," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 48–56, 2016. DOI: 10.1109/TSG.2016.2544882.

[77] P. Kepplinger, G. Huber, M. Preißinger, and J. Petrasch, "State estimation of resistive domestic hot water heaters in arbitrary operation modes for demand side management," *Thermal Science and Engineering Progress*, vol. 9, pp. 94–109, 2019.

[78] Z. Xu, R. Diao, S. Lu, J. Lian, and Y. Zhang, "Modeling of electric water heaters for demand response: A baseline pde model," *IEEE Transactions on Smart Grid*, vol. 5, no. 5, pp. 2203–2210, 2014.

[79] P. M. Armstrong, M. Uapipatanakul, I. Thompson, D. Ager, and M. McCulloch, "Thermal and sanitary performance of domestic hot water cylinders: Conflicting requirements," *Applied Energy*, vol. 131, pp. 171–179, 2014, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2014.06.021.

[80] P. J. C. Nel, M. J. Booysen, and A. B. van der Merwe, "Energy perceptions in South Africa: An analysis of behaviour and understanding of electric water heaters," *Energy for Sustainable Development*, vol. 32, pp. 62–70, 2016. DOI: 10.1016/j.esd.2016.03.006.

[81] A. V. Rao, "A survey of numerical methods for optimal control," *Advances in the Astronautical Sciences*, vol. 135, no. 1, pp. 497–528, 2009.

[82] J.-C. Laurent, G. Desaulniers, R. P. Malhame, and F. Soumis, "A column generation method for optimal load management via control of electric water heaters," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1389–1400, 1995.

[83] H. Tischer and G. Verbic, "Towards a smart home energy management system a dynamic programming approach," in *2011 IEEE PES Innovative Smart Grid Technologies*, IEEE, 2011, pp. 1–7.

[84] M. Heleno, D. Rua, C. Gouveia, A. Madureira, M. Matos, J. P. Lopes, N. Silva, and S. Salustio, "Optimizing pv self-consumption through electric water heater modeling and scheduling," in *2015 IEEE Eindhoven PowerTech*, IEEE, 2015, pp. 1–6.

[85] T. Roy, A. Das, and Z. Ni, "Optimization in load scheduling of a residential community using dynamic pricing," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, IEEE, 2017, pp. 1–5.

[86] M. S. Ahmed, A. Mohamed, T. Khatib, H. Shareef, R. Z. Homod, and J. A. Ali, "Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm," *Energy and Buildings*, vol. 138, pp. 215–227, 2017.

[87] D. E. Kirk, *Optimal Control Theory: An Introduction*. Courier Corporation, 2012.

[88] J. A. A. Engelbrecht, "Automatic flight envelope recovery for large transport aircraft," PhD thesis, Stellenbosch: Stellenbosch University, 2016.

[89] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

[90] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[91] R. Sutton, *Reinforcement Learning: An Introduction Second Edition*. MIT Press, 2017.

[92] R. Eberhart and J. Kennedy, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, Citeseer, vol. 4, 1995, pp. 1942–1948.

[93] Y. Deng, Y. Chen, Y. Zhang, and S. Mahadevan, "Fuzzy dijkstra algorithm for shortest path problem under uncertain environment," *Applied Soft Computing*, vol. 12, no. 3, pp. 1231–1237, 2012.

[94] M. Noto and H. Sato, "A method for the shortest path search by extended dijkstra algorithm," in *Smc 2000 conference proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics.'Cybernetics Evolving to Systems, Humans, Organizations, and their Complex Interactions'(cat. no. 0*, IEEE, vol. 3, 2000, pp. 2316–2320.

[95] O. Wink, W. J. Niessen, and M. A. Viergever, "Minimum cost path determination using a simple heuristic function," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, IEEE, vol. 3, 2000, pp. 998–1001.

[96] S. K. Sharma and S. Kumar, "Comparative analysis of Manhattan and Euclidean distance metrics using A* algorithm," *J. Res. Eng. Appl. Sci*, vol. 1, no. 4, pp. 196–198, 2016.

[97] H. E. Jacobs, B. E. Botha, and M. Blokker, "Household Hot Water Temperature – An Analysis at End-Use Level," in *International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada – July 23-25*, 2018, pp. 1–14.

[98] M. J. Booysen, J. A. A. Engelbrecht, and A. Molinaro, "Proof of concept: Large-scale monitor and control of household water heating in near real-time," in *Proc. of International Conference on Applied Energy (ICAE)*, URL: http://hdl.handle.net/10019.1/95703, Pretoria, South Africa, Jul. 2013, pp. 1–8.

[99] W. Stone, T. M. Louw, G. K. Gakingo, M. J. Nieuwoudt, and M. J. Booysen, "A potential source of undiagnosed legionellosis: Legionella growth in domestic water heating systems in South Africa," *Energy for Sustainable Development*, vol. 48, pp. 130–138, 2019, ISSN: 0973-0826. DOI: 10.1016/j.esd.2018.12.001.

[100] S. Nagaraj, "Optimal binary search trees," *Theoretical Computer Science*, vol. 188, no. 1-2, pp. 1–44, 1997.

[101] A. Fanney and B. Dougherty, "The thermal performance of residential electric water heaters subjected to various off-peak schedules.," *ASME. J. Sol. Energy Eng.*, vol. 118, no. 2, pp. 73–80, 1996. DOI: https://doi.org/10.1115/1.2848010.

[102] C. H. K. Goh and J. Apt, "Consumer strategies for controlling electric water heaters under dynamic pricing," in *Working paper CEIC-04-02*, 2004, pp. 1–8.

[103] M. J. Booysen and A. H. Cloete, "Sustainability through intelligent scheduling of electric water heaters in a smart grid," in *2016 IEEE 2nd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress*, Aug. 2016, pp. 848–855. DOI: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.145.

[104] Rui Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, May 2005, ISSN: 1045-9227. DOI: 10.1109/TNN.2005.845141.

[105] A. Bhagat, N. Kshirsagar, P. Khodke, K. Dongre, and S. Ali, "Penalty parameter selection for hierarchical data stream clustering," *Procedia Computer Science*, vol. 79, pp. 24–31, 2016, Proceedings of International Conference on Communication, Computing and Virtualization (ICCCV) 2016, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2016.03.005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050916001368.

[106] *Eskom geyser fact sheet 2013*, http://www.eskom.co.za/sites/IDM/Documents/1317_geyser_fact_sheet_no_rmr.pdf, Accessed: 2019-09-13.

[107] G. Ladd and J. Harrison, "Electric water heating for single-family residences: Group load research and analysis. final report," Gilbert Associates, Inc., Reading, PA (USA), Tech. Rep., 1985.

[108] J. R. Koza, "Genetic programming," 1997.

[109] D. George, N. S. Pearre, and L. G. Swan, "High resolution measured domestic hot water consumption of Canadian homes," *Energy and Buildings*, vol. 109, pp. 304–315, 2015.

[110] P. Z. Peebles, *Probability, random variables, and random signal principles*. McGraw-Hill New York, 2001, vol. 3.