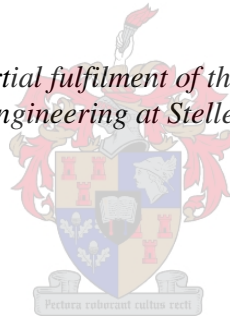# Probabilistic Load Modelling of Electrical Demand of Residential Water Heating

by
Graeme John Urban

Thesis presented in partial fulfilment of the requirements for the degree
Master of Engineering at Stellenbosch University

Supervisor: Professor H.J. Vermeulen
Department of Electrical and Electronic Engineering

March 2012

Declaration

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

G.J Urban

March 2012

Abstract

Energy efficiency and the move to renewable energy resources are of vital importance in growing profitable and sustainable economies. In recent years, greater emphasis has been placed on institutions, companies and individuals to reduce their electrical energy demand through energy management. In an attempt to reduce the demand, the electrical power utility in South Africa, Eskom, has introduced Demand Side Management programs and substantial increases in electricity tariffs. In addition to these, tax incentives have been offered to help off-set the capital costs associated with the investments made in replacing old electrical equipment with new electrically efficient equipment. Thus the need for accurate Measurement and Verification of electrical energy demand reduction, to substantiate fiscal claims, has become imperative. The main purpose of Measurement and Verification is to investigate the actual monetary performance of an energy savings project. Energy savings assessments, based on purely deterministic baseline demand, do not adequately represent the statistical nature of the savings impacts of many practical load systems, as disclosed in a reporting period. This thesis presents the development of a generic probabilistic methodology to determine the demand profiles of pre- and post-Energy Conservation Measures (ECMs) for practical load systems. The difference between the simulated demand of the pre- and post-ECMs for a particular set of variables represent the electrical demand impact. The electrical demand of the pre- and post-ECMs is defined in terms of Probability Density Functions, and derived using a multivariate kernel density estimation algorithm. The approach is tested using a simulation model of a water-heating geyser implemented in MATLAB. Three different ECMs are simulated using the geyser model and demand density estimation. The results of the demand impacts of the ECMs are presented and evaluated. With regards to possible future research this methodology could be applied to the evaluation of the demand impacts of heat pump technologies and solar water heaters.

ii

Opsomming

Energiedoeltreffendheid en die skuif na hernubare energiebronne is van deurslaggewende belang vir die ontwikkeling van winsgewende en volhoubare ekonomieë. Onlangs is meer klem geplaas op instansies, maatskappye en individue om hul aanvraag na energie te verminder met behulp van energiebestuur. In 'n poging om die aanvraag te verlaag, het Eskom, Suid-Afrika se elektrisiteitsverskaffer, aansienlike elektrisiteitstariefverhogings ingelyf en Aanvraagbestuursprogramme van stapel gestuur. Bykomend hiertoe is belastingaansporings ook aangebied, waarteen kapitale kostes, geassosieer met die vervanging van ou elektriese toerusting met nuwe elektries doeltreffende toerusting, afgeset kan word. Derhalwe het die behoefte aan akkurate Meting en Verifikasie van elektriese energie aanvraagvermindering, om finansiële eise te staaf, noodsaaklik geword. Die hoofdoel van Meting en Verifikasie is om die werklike finansiële prestasie van 'n energiebesparingsprojek te ondersoek soos bekend gemaak word tydens 'n verslagdoeningstydperk. Energiebesparingassesserings wat slegs gebaseer word op die suiwer deterministiese basislyn aanvraag na elektrisiteit, verteenwoordig nie die werklike statistiese aard van die besparingsimpakte van baie praktiese lasstelsels nie. Hierdie tesis stel die ontwikkeling van 'n generiese waarskynlikheids-metodologie voor, om die voor- en na-Energiebesparings-maatreëls se aanvraagprofiele vir sulke praktiese lasstelsels, vas te stel. Die verskil tussen die gesimuleerde aanvraag van die voor- en na- Energiebesparings-maatreëls vir 'n spesifieke stel veranderlikes verteenwoordig die elektriese aanvraag impak. Die voor- en na-Energiebesparings-maatreëls van die energieverbruik profieldata word gedefinieer in terme van Waarskynlikheidsdigtheidsfunksies en afgelei deur gebruik te maak van 'n meerveranderlike kerndigtheidafskattingsalgoritme. Die benadering is getoets deur gebruik te maak van 'n simuleringsmodel van 'n warmwaterstelsel geïmplimenteer in MATLAB. Drie verskillende voor- en na- Energiebesparings-maatreëls is gesimuleer met behulp van die warmwaterstelsel-model en aanvraag digtheidafskatting. Die resultate van die elektriese aanvraag impakte van die voor- en na- Energiebesparings-maatreëls word vervolgens bespreek en geëvalueer. Met betrekking tot moontlike toekomstige navorsing kan hierdie metodologie toegepas word om die aanvraag impakte van hittepomp- en sonwaterverwarmingstegnologieë te evalueer.

Acknowledgements

I would like to express my sincere gratitude to Professor H.J. Vermeulen for his assistance and advice during the duration of this project. To my fiancée Charlene, I would like to thank her for her support during the many hours that were invested in this thesis. Finally, I would like thank my family and friends that assisted me in one way or another to reach this milestone in my life.

Ask and he shall receive

Have faith and he shall receive more than he asked for.

Table of contents

List of figures

List of tables

Nomenclature

AMISE   Asymptotic Mean Integrated Squared Error

ANOVA   Analysis of variance

CDF(s)   Calmative Distribution Function(s)

DSM   Demand Side Management

ECM(s)   Energy Conservation Measure(s)

EPA   Environmental Protection Agency

ESCO   Energy Services Company

EVO   Efficiency Valuation Organisation

IPMVP   International Performance Measurement and Verification Protocol

kVA   kilo volt ampere

kW   kilo watt

kWh   kilo watt hour

M&V   Measurement and Verification

MISE   Mean Integrated Squared Error

MW   Megawatt

PDF(s)   Probability Density Function(s)

ROI   Return on investment

SABS   South African Bureau of Standards

Chapter 1   **Probabilistic methodology for demand characterization**

1.1   Introduction

The world continues to face substantial energy challenges. One of the most significant challenges is proving to be the delivery of secure and affordable energy. As a result of population growth and expanding economies, there will be an ever increasing competition for diminishing resources [1]. This will naturally lead to increased energy costs and the strategic importance of energy. Economies that are more reliant on fossil fuels are more exposed to fluctuation and increased energy costs [2]. Energy efficiency and the move to the use of renewable energy resources are of vital importance in growing profitable and sustainable economies. Another significant energy challenge stems from the greenhouse gas emissions that result from the production and use of energy [1]. The increased levels of these gases in recent times have lead to climate change regionally and globally. This has the potential risk of adversely effecting agriculture production and the increased frequency of natural disasters associated with abnormal weather conditions. These pose substantial risks to the future global economy.

It is evident that the desperate drive to move way from non-renewable fossil fuels as a source of energy, is gaining momentum. The world now looks to renewable energy resources to feed its ever increasing need for energy. In the past, the focus of energy management has been on energy load forecasting, so that forward planning could be done to secure more energy in the future by advanced purchasing of additional oil and coal [3]. This planning also included the building of more coal fired power stations. Nowadays the focus has shifted. Energy management is placing more emphasis on increasing the efficient use of energy. Electricity remains the preferred medium for the efficient transport of energy. Dispersed generation from renewable energy sources seems to be the logical way forward in securing reliable and cost effective electrical energy. With the imminent advent of electrically powered cars in our cities we will become more dependent on electricity and the electricity networks of the world [1]. The focus of electrical energy efficiency must not only be limited to the end-use technologies and

usage practices but must also include the efficient generation transmission and distribution of electricity [1].

The reality of the increasing importance of energy efficiency in everyday life is well phrased in the following excerpt from Time Magazine and it is by no means only applicable to the Unites States of America:

"*This may sound too good to be true, but the U.S. has a renewable-energy resource that is perfectly clean, remarkably cheap, surprisingly abundant and immediately available. It has astounding potential to reduce the carbon emissions that threaten our planet, the dependence on foreign oil that threatens our security and the energy costs that threaten our wallets. Unlike coal and petroleum, it doesn't pollute; unlike solar and wind, it doesn't depend on the weather; unlike ethanol, it doesn't accelerate deforestation or inflate food prices; unlike nuclear plants, it doesn't raise uncomfortable questions about meltdowns or terrorist attacks or radioactive-waste storage, and it doesn't take a decade to build. It isn't what-if like hydrogen, clean coal and tidal power; it's already proven to be workable, scalable and cost-effective. And we don't need to import it. This miracle juice goes by the distinctly boring name of energy efficiency, and it's often ignored in the hubbub over alternative fuels, the nuclear renaissance, T. Boone Pickens and the green-tech economy. Clearly, it needs an agent. But it's a simple concept: wasting less energy*"
[4].

1.2   Project motivation

The purpose of energy management is to ensure the sustainability of energy resources and assisting in ensuring the sustainability of the natural environment. The evaluation of the efficiency of energy consumers is at the heart of energy management. In recent years, greater emphasis has been placed on institutions, companies and individuals to reduce their energy demand needs through energy management. The electrical power producer in South Africa, Eskom has done this by way of substantial price increases in electricity tariffs and Demand Side Management (DSM) programs. In addition to these, tax incentives have been offered to help off-set the capital costs associated with the investments made in replacing old electrical equipment with new electrically efficient equipment.   This has resulted in the need for accurate Measurement and Verification (M&V) of electrical load demand reduction to substantiate the

fiscal claims. The main purpose of M&V is to present the actual monetary performance of an energy savings project. The M&V plan forms part of an energy-performance contract between the energy user and the energy savings company (ESCO), who implements the energy savings project. The results of the M&V report is measured against the ESCO's guaranteed level of performance [5].

Demand Side Management (DSM) interventions make use of Energy Conservation Measures (ECMs) to drive energy savings through increased efficient use of energy. Two examples of DSM initiatives that the South African electricity provider, Eskom, has being rolling out are: the compact fluorescent lighting program and the solar water heater replacement program. Eskom's Standard Product program encourages DSM initiatives by providing rebates for technology replacements that result in greater efficient use of energy [6]. Baseline energy profiles of electrical consumers form the basis of calculations to determine the impact that DSM initiatives and energy savings projects have on the energy demand. To determine the improvement in energy efficiency of an energy consumer, a reference or baseline consumption profile must be established against which a future consumption profile is to be evaluated. The baseline consumption profile can also be used for short-term load forecasting.

In the South African context and possibly many other parts of the world, baseline model determination has almost exclusively taken on a deterministic approach. This approach lacks the confidence factors associated with the numerical values of the predicted load. The International Performance Measurement and Verification Protocol (IPMVP) promotes the inclusion of confidence factors when reporting demand impacts [5]. This would give the energy manager the relative certainty that an energy load will be in a particular range of energy values. Although confidence factors are an important part of the information reported concerning a demand impact, the IPMVP fails to present a preferred method on how to calculate this information. It is believed that this may be the reason why M&V reporting has in general lacked this necessary information. The reason why this information is necessary stems from the fact that almost all energy conservation measurements require a financial investment. The main factors that play a role in the decision on whether or not to make the investment are: the return on investment (ROI) and the risk associated with realising the ROI. To calculate the risk, confidence factors of achieving the expected demand impact are required. The degree of complexity in calculating the

confidence factors increases with the degree of complexity of the energy load [5]. Loads that are influenced by multiple independent controlling variables pose the greatest challenge to quantify their demand impact that result from ECMs. It is imperative that in the case of complex load systems, the relationship between load and the independent controlling variables be presented when energy load impacts are evaluated. It is suggested in Chapter six that statistical methods should be used for the evaluation and interpretation of complex load profiles. These load profiles should rather be represented in the form of multiple Probability Density Functions (PDFs) which depict the dependency of the load in relation to its controlling independent variables. From these PDFs, load value ranges and their respective probabilities can be determined.

A conclusion may be drawn at this stage that energy management is of ever increasing importance to the built and natural environments and at the heart of energy management is the prediction of demand impacts of ECMs. Such prediction is no easy feat. As the Nobel Laureate Physicist, Neils Bohr, put it

"Prediction is a very difficult art, especially when it involves the future"

1.3   Project description

In South Africa the largest consumer of electricity in the residential sector is the heating of hot water. A conventional hot water heater constitutes 40 to 50 percent of a household's electricity demand, as quoted in the South African Energy Policy Discussion Document [7]. Eskom DSM initiatives have been promoting the replacement of conventional hot water heaters with solar water heaters and heat pumps. In 2007, the Eskom board approved a two billion rand investment available over five years to subsidise the installation of solar hot water heaters. Eskom hopes to subsidise the installation of 925 000 solar water heaters and in doing so reduce the demand on the grid by 578 MW [8].

The amount of electrical energy used by a geyser is largely dependent on ambient temperature and hot water demand. It will be seen in Chapter two that hot water demand is also largely dependent on ambient temperature. For this reason, hot water heating is considered a complex load with ambient temperature being the stochastic independent variable and electric load the dependent variable. To elevate the performance of DSM initiatives such as the installation of

geyser blankets or the replacement of a geyser with solar water heater or heat pump, baseline and performance period electrical consumption profiles need to be determined. As a result of the current need to substantiate the performance of DSM initiatives related to hot water heating, it was decided to set out the following objectives as per the following chapters of this thesis:

Chapter two

- investigation of the M&V methodology in detail;
- selection of an appropriate method of statistical inference to calculate the confidence levels of predicted savings;
- investigation of the purpose of load forecasting and load modelling;
- investigation of hot water consumption rates and presentation of technologies that are employed in hot water heating, and
- investigation of a method to predict the inlet water temperature of a geyser.

A comprehensive investigation is needed to determine the present methodology employed by M&V to determine demand impacts that result from DSM interventions. This investigation highlighted the lack of detail that was given by an M&V process on how to determine confidence levels of predicted savings impacts. This resulted in a need to investigate which method of statistical inference would be the most appropriate to interpret the geyser demand. The literature study in Chapter two then takes a broader view and investigates the purpose of load forecasting and load modelling. The focus shifts back to hot water heating by investigating the input variables to hot water heating technologies and the technologies themselves. The input variables that were investigated were hot water consumption rates and inlet water temperature.

Chapter three covers the following objectives:

- development of a methodology to model and quantify the electrical demand of a geyser;
- development of a geyser model, and
- development of a method for the statistical interpretation of the electrical demand.

As a result of the knowledge gained in the literature study, it was decided to develop a methodology to predict demand impacts that result from DSM initiatives. To do this, a so-called "sterile" environment was created to evaluate the methodology used to develop the geyser model

and statistically interpret its simulated demand. The geyser model and the code used to analyse its simulated demand was developed in MATLAB and SIMULINK. This model topology created a sterile simulation environment where the effectiveness of the proposed methodology could be evaluated. A description is given of the geyser model and the software that was written to simulate and analyse the input and output variables of the model.

Chapter four focuses on the statistical analysis of the input variables to the geyser model and includes the following objectives:

- investigation of the effect data sampling rate has on the simulated electrical demand;
- statistical analysis of the ambient temperature variable of the geyser model;
- statistical analysis of the inlet water temperature variable of the geyser model, and
- statistical analysis of the hot water consumption rate variable of the geyser model.

Input data to the geyser model may be available at different sampling rates. The influence of these different sampling rates on the simulated data was investigated through statistical methods. This was followed by an in-depth statistical analysis of the input variables to the geyser model. This was done to determine the effect the geyser model has on the input variables in relation to the electrical demand output variable.

The content presented in Chapter five is two-fold:

- analysis of the simulated demand data produced by the geyser model, and
- investigation of the dependency between independent and dependent variables of the geyser model as a result of ECMs.

The results of the simulations using different input variable sampling periods are presented. A group of simulation results of possible geyser ECMs were then statistically analysed and the relationship between dependent and independent variables investigated.

In closing, Chapter six provides the following achievements and conclusions in respect of the set objectives as well as recommendations for future work:

- conclusions on the appropriateness of the methodology developed to model a geyser;

- conclusions on the appropriateness of the methodology to quantify demand impacts as a result of intervention, and
- suggestions on the future scope of work to further develop the load modelling methodology.

Chapter 2   **Literature study**

2.1 **Overview**

The literature review begins with an explanation of the concept of M&V and the important role it plays in the evaluation and planning of ECMs. A description of the M&V system and process are followed by an account of the main purposes of M&V. The IPMVP is introduced as the foremost international protocol for M&V. The four measurement methodologies advocated by the IPMVP are then briefly described. A method to calculate the cost saving as a result of energy saving of an ECM is presented. A good overall description of M&V is necessary to see how the load modelling methodology described in Chapter three compliments the M&V system and process. It is in the author's opinion that the IPMVP still falls short in its recommended approach to the statistical analysis of an energy load and the predicted energy savings of an ECM. It is in this section of M&V where further development is needed.

The next section of the literature review investigates statistical inference that would prove to be a useful tool in the M&V process. Parametric and non-parametric interference is briefly described with reference to their respective advantages and disadvantages. It was decided to further investigate non-parametric univariate and bivariate density estimation. Kernel density estimation is a modern technique used in probability density estimation and its strengths and weaknesses are discussed. The importance and methods used to determine bandwidth selection for kernel density estimation is highlighted. The statistical techniques discussed in this section of the literature review are used in the statistical interpretation of the predicted impacts of the geyser ECMs in Chapter five.

Load forecasting is discussed in the next section of the review. This discussion sets the stage in order to highlight the importance of long and short-term load forecasting. This forecasting is not only demand savings but also a forecast of aggregated electrical load from a group to a network level of electrical consumers. Reference is made to a particularly relevant piece of work that was done on density estimation which was used to conduct short-term load forecasting.

The discussion on load forecasting is followed by a section on hot water consumption and hot water heating in a residential context. It was previously mentioned in Chapter one that hot water

heating accounts for the largest electrical consumer in the residential sector. A hot water supply device such as a geyser represents a complex electrical load because the electrical consumption of the hot water produced is dependent on independent variables such as ambient temperature. Hot water consumption is one of the input variables to the geyser model that is developed in Chapter three. A brief presentation of the most common water heating devices follows the hot water consumption discussion.

The other input variable to the geyser model is the inlet water temperature. In conclusion a method is discussed on how to predict the inlet water temperature for different burial depths of the water supply pipe at different times of the year.

2.2   Energy savings in terms of measurement and verification

2.2.1   Overview

M&V is defined in the IPMVP as the process of calculating energy savings of a facility that result from the implementation of an energy management program [5]. In a document published by the South African Bureau of Standards (SABS), M&V is described as "the process of quantifying energy savings or the impact by the determination of actual consumption and relevant energy-governing factors, and to develop baseline and baseline adjustments" [9]. Energy savings cannot be measured directly because it represents the absence of energy use. The calculation involves comparing the measured energy use before and after the implementation of a project, accounting for the adjustments as a result of changes in conditions [5]. The Australian Energy Performance Contracting Association describes M&V methods as "Methods used to measure and verify, in a defined and disciplined, rigorous and transparent way, the energy savings resulting from implementation of ECMs which have been planned and designed to improve the energy performance of a specific facility or group of specific facilities." [10]. The Association of Edison Illuminating Companies Load Research Committee introduces the requirement of statistical analysis when calculating an energy savings value when it defines M&V as follows: "M&V refers to the application of appropriate statistical and load research techniques to measure and verify the load reduction impact that result from the utilization of a Demand Response Program." [11]. The purpose of the statistical analysis is to contribute a level of confidence to the calculated energy saving value [11].

The goal of M&V, according to a report produced for the U.S Department of Energy Federal Energy Management Program is, " … to reduce the risk to agencies by providing a mechanism to evaluate the performance of a project through the term of the contract." [12].   The New Brunswick Industrial Program claims the objective of M&V is to quantify energy and financial savings as a result of ECMs.   This can be a challenging objective to achieve when the independent variables that influence the energy used during the performance period vary in relation to the baseline period.   This variance can significantly affect the processes to calculate energy use [13].

M&V and an M&V plan form the basis of the commercial contract between the ESCO and the agency, when ECMs are undertaken [14].   An appropriately designed M&V plan reduces the risk to which the agency is exposed when investing in an ECM.   A further challenge for M&V is to balance the M&V costs with the cost of improving the certainty of the quantity of energy saving as a result of the ECM [12].

The IPMVP is one of the most important tools used in a M&V process and was developed by the Efficiency Valuation Organisation (EVO) in an effort to overcome the challenges associated with the M&V process.   One of the most important reasons to manage the M&V challenges, stems from the fact that M&V forms the basis of commercial contracts involving ECMs.   IPMVP is used by professionals as a framework for preparing savings reports.   EVO promotes and encourages the efficient use of natural resources and end use efficiency as an alternative to new energy supply [5].   There are four options provided by IPMVP for determining savings and this will be discussed in greater detail in section 2.2.5.   The IPMVP does not cover the following aspects in any detail [15]:

- the design of metering systems;
- the cost estimation of M&V activities;
- energy engineering, and
- statistical analysis.

2.2.1.1   M&V system

The baseline, performance and savings determination are the three M&V process components that work together to form a complete system.   The M&V system deals with the situation that

10

once an ECM has been implemented, the original system is altered to such an extent that it is often impossible to conduct M&V retroactively. It is therefore a system requirement that M&V be initiated at the planning and feasibility stages of a project [13]. The energy use profile prior to the implementation of an ECM is referred to as the baseline energy use and the energy use profile after the implementation is the energy profile of the performance period or reporting period. The M&V system is depicted in Figure 2.2.1.
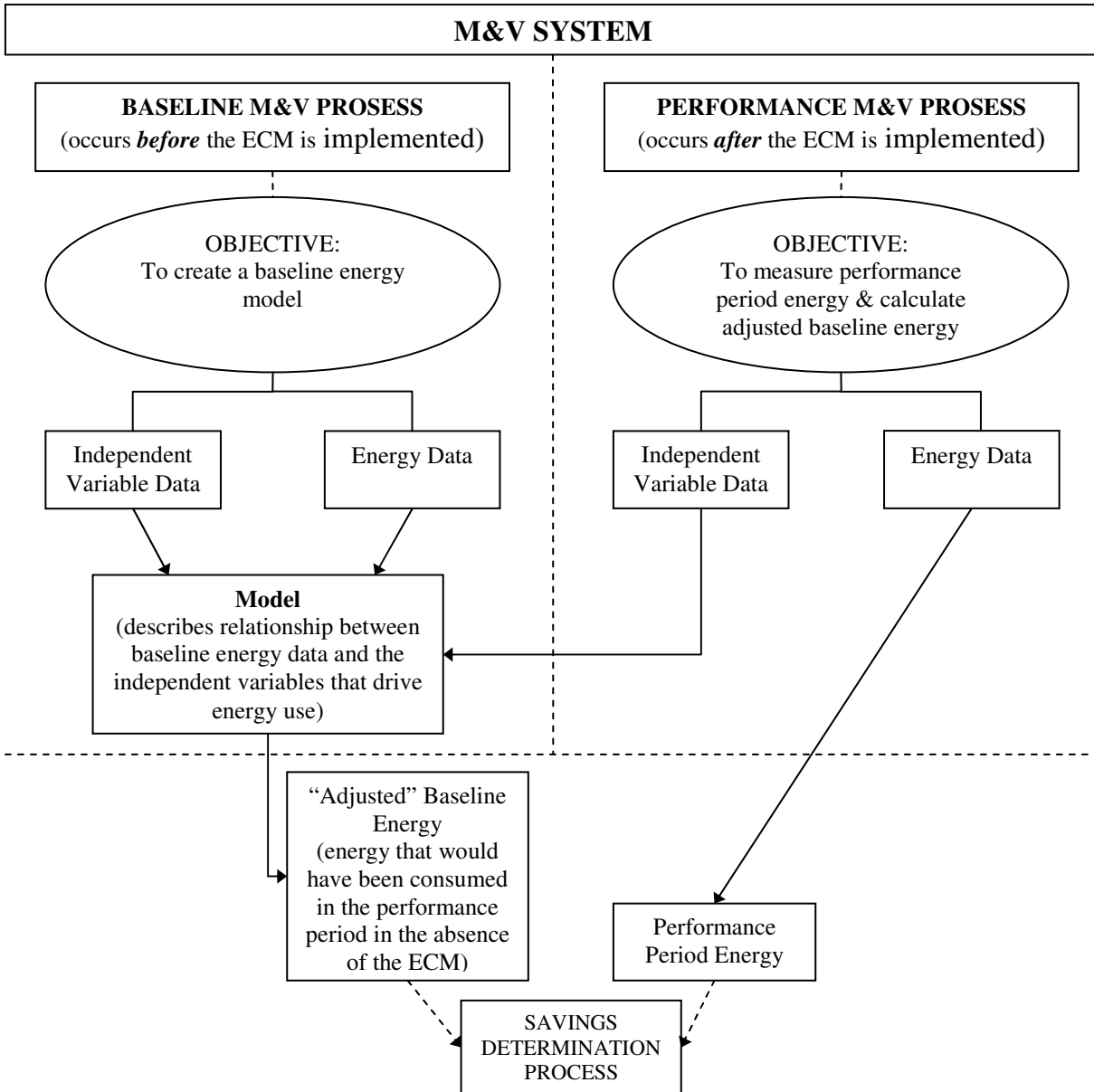
Figure 2.2.1: Process flow diagram of the M&V system used to quantify energy savings in production environments [13].

11

From this figure it can be seen that the baseline model must be created before the ECM is implemented. The baseline model is intended to describe the mathematical relationship between the baseline energy data and the independent variables that drive the energy usage [13]. In simple energy consumption systems there may not be any significant independent variables that drive baseline energy use. In these situations a baseline model is not required and the energy savings are simply calculated by subtracting the performance period's energy usage from the baseline energy usage, provided the two energy periods are of equal length. In more complex systems where independent variables drive the energy usage, a baseline model is a prerequisite in determining the impact the ECM has on energy usage. Once an ECM has been implemented, a baseline model is used to represent the baseline system. The performance period's data, consisting of independent variables, is then fed into the baseline model. The output of the model represents the calculated adjusted baseline energy usage or, stated differently, the energy that would have been used by the energy consumer during the performance period had the ECM *not* been implemented. The energy saved as a result of the implementation of the ECM is then calculated by subtracting the recorded energy usage during the performance period from the adjusted baseline energy usage [13]. The IPMVP uses equation 2.2.1 to calculate savings. The *Adjustments* accounts for the changes in conditions between the baseline and reporting periods.

$$Savings = \left(Baseline\text{-}Period\ Use\ or\ Demand - Reporting\text{-}Period\ Use\ or\ Demand \right) \pm Adjustments$$

<div align="right">**2.2.1**</div>

M&V may not be necessary if there is high certainty of the results of a project and if these results do not have to be substantiated. Verification of the impact of the project would, however, still need to be performed to make sure the project goals are met.

2.2.1.2   M&V process

The following activities form part of the M&V process [5]:

- installation, calibration and maintenance of metering equipment;
- collecting and screening of data;
- development of a method to calculate and estimate project impacts;
- calculations involving measured data, and
- third party verification of reports and quality assurance of energy management programs.

The New Brunswick Industrial Program maintains that the M&V process has the following structure [13]:

- required data and measurement boundary are specified to substantiate the intent of the ECM;
- measurement hardware is selected and installed;
- project data collection method is specified (This data is used in the baseline process to produce a baseline model. In the performance process the performance period data is used to calculate the adjusted baseline data.), and
- in the final step of the process, the various M&V tasks to be undertaken are assigned to designated people.

The Best Practice Guide to Measurement and Verification of Energy Savings, advocates the following order of approach to basic M&V processes [10]:

Before an ECM implementation:

- M&V plan must be developed and a method selected for each of the ECMs based on the objectives of the project;
- relevant energy and operating data from the baseline period must be recorded;
- computational methods and data to be used to calculate the baseline energy must be ascertained, and
- any special measurement equipment as specified by the M&V plan must be designed, installed and tested.

After the ECM implementation:

- M&V implementation must be verified and be consistent with the M&V plan;
- relevant energy and operating data from the performance period must be recorded;
- energy and cost savings are calculated and reported, and
- report's compliance with the M&V plan is then verified.

In the design and reporting process of M&V the needs of the user must be considered. Different methods are used depending on whether the energy user needs the whole facility to be evaluated

or just part thereof. This is referred to as setting the measurement boundary. A decision needs to be made on what energy-governing factors need to be taken into consideration when determining the baseline energy use. The length of the baseline and reporting period must be decided on [5]. The M&V process can itself introduce uncertainties such as:

- measurement and instrument errors;
- modelling errors;
- sampling errors, and
- planned and unplanned assumptions.

These uncertainties need to be quantified in a statistically meaningful way by taking in to account both accuracy and confidence level [10]. The level of uncertainty of the calculated savings needs to be appropriately set for a specific project. The cost of M&V increases along with the need for increased certainty with which the energy savings are reported. The reasons for this increase in cost are as follows [10]:

- measurement and instrument errors can be reduced by making use of more precise instruments that are normally more expensive to purchase;
- sampling errors can be mitigated by increasing sampling size which would again be more costly, and
- data errors and risk can be reduced by measuring, and not assuming, a quantity from a manufacturer (This additional measurement requires more time and money.).

The cost of the level of certainty of the calculated saving must be balanced with the value of the savings and be appropriate for the level of complexity of the ECM that was implemented [10].

2.2.1.3   Main purposes of M&V

The main purposes of an M&V process are to increase energy savings through a better understanding of the extent and nature of the impacts resulting from an energy management program. A well-developed M&V plan often forms the basis for a performance-based financial contract between the energy user, the ESCO and the financier. This plan also encourages better project engineering [10].   M&V also helps reveal and reduce maintenance and operational problems. M&V techniques help account for variances in the budget resulting from changes in

the facility's operating conditions. An M&V plan helps improve emission reduction reports which provide additional value to efficiency projects [5]. The M&V of ECMs helps national and industrial organisations promote and achieve resource efficiency and environmental objectives [10]. The ability to finance ECMs is based on sound prediction, measurement and verification of savings. These predictions and verification of savings are not only used to calculate the return on investment but also the risk associated in achieving the predicted results. The accuracy and confidence by which savings can be forecast play an important role in the financial risk management of the project [10].

A summary of reasons according to the U.S. Federal Report for using M&V strategies are [12]:

- necessary level of saving accuracy for an energy management project;
- risk allocation to the relevant parties;
- uncertainties are reduced to reasonable levels;
- performance monitoring of retrofitted energy savings equipment;
- finding of additional savings;
- improved operations and maintenance of the plant, and
- cost savings guarantee is verified.

2.2.2   Baseline characterization

Before ECMs are implemented, a system has a characteristic energy usage. If the ECM implementation is successful there would be a reduction in energy usage. To be able to compare the energy usage before and after the ECM implementation, a baseline, characterising the energy usage, needs to be determined [5]. This baseline takes into account input variables or energy-governing factors. Stated differently, the adjusted-baseline energy is representative of the amount of energy the plant would have used had there been no ECM. In Figure 2.2.2, the baseline and reporting periods are depicted for a manufacturing facility. It can be seen that there was an increase in production during the reporting period. The adjusted baseline takes this change in the production volume variable into account. If this was not the case, the energy savings would be under estimated. The difference between the adjusted baseline energy and the measured energy represents the savings or the avoided energy use in the reporting period [5].

15
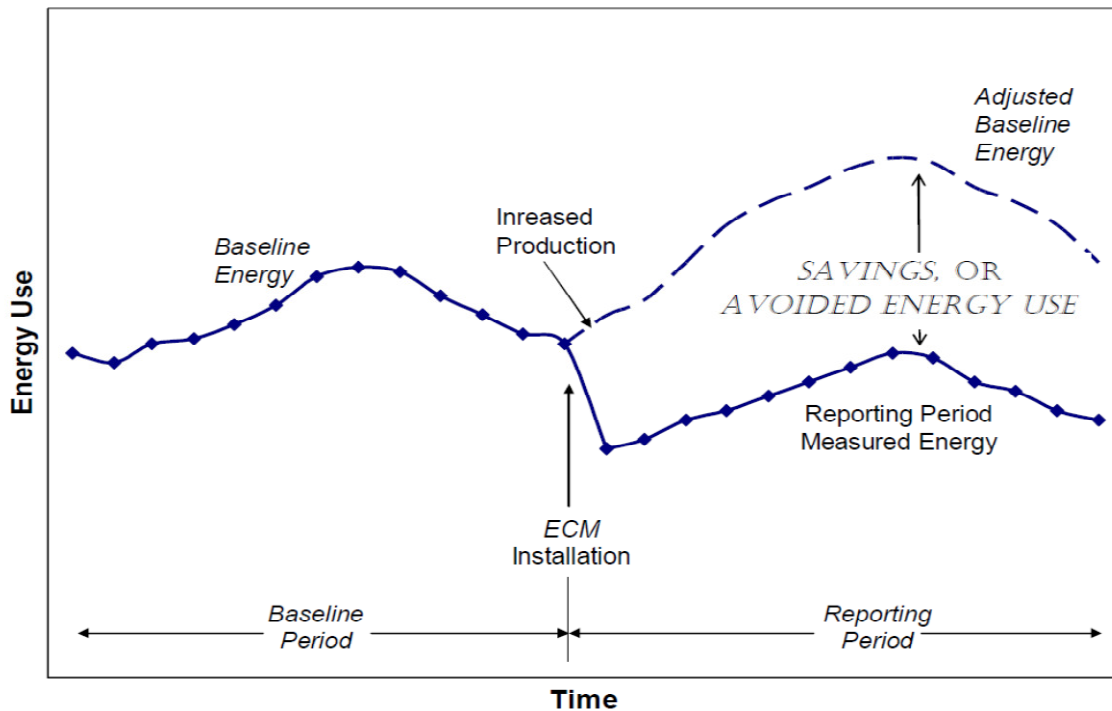
Figure 2.2.2: Energy used in a production facility during the baseline and reporting periods [5].

2.2.3 Baseline and reporting periods

Care needs to be taken when selecting a period for baseline and reporting periods. The baseline and reporting periods should include a full operating cycle from maximum energy use to minimum. An example of this is the energy used in a building. Energy use in a building is significantly affected by weather conditions and it is therefore necessary to include a whole year's seasonally dependent baseline data in the baseline period to define a full operating cycle. The baseline period should only include time periods in which fixed and variable energy-governing facts are known for the facility [5].

2.2.4 Determination of energy savings

Performance and usage are the two main factors that drive energy savings [12]. Performance refers to the rate of energy used during the performance of a specific task, -and usage refers to the time taken to perform the task. Both of the factors need to be known to calculate the energy savings of an ECM. This is graphically represented in Figure 2.2.3.
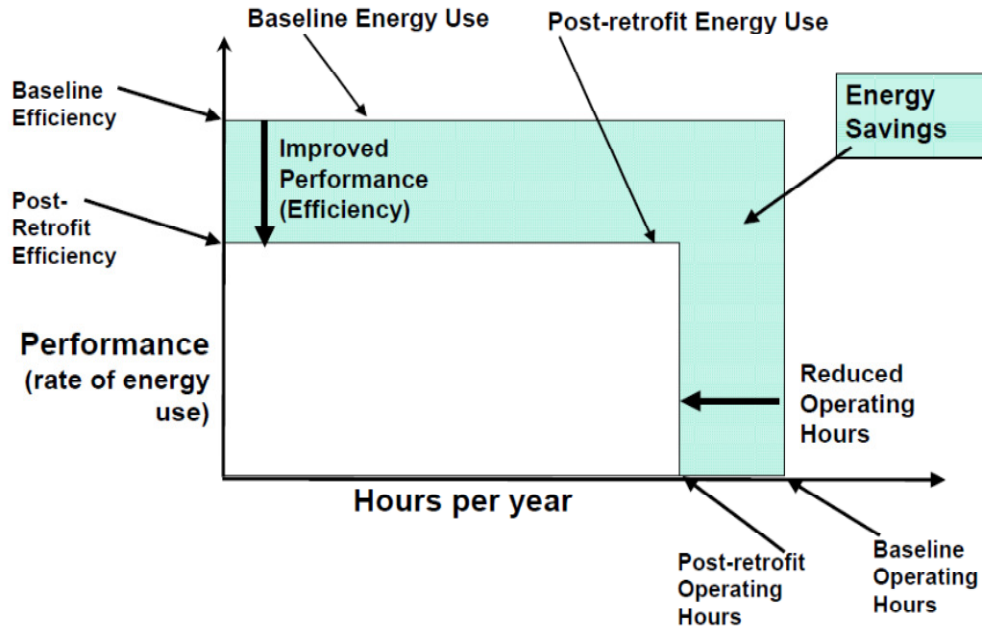
Figure 2.2.3: Energy performance and usage before and after the implementation of an ECM [12].

An ECM will normally aim to increase the energy efficiency and or reduce the time taken to perform the task. The shaded area in Figure 2.2.3 represents the energy saving or the avoided energy use [12].

Energy savings are calculated by obtaining the difference between the adjusted baseline and the actual energy usage. According to the SABS, energy savings are calculated by the following equation [9]:

$$E_s = B_{peu} - R_{peu} \pm A$$

2.2.2

where

E_s denotes the energy savings;

B_{peu} denotes the baseline period energy use;

R_{peu} denotes the reporting period energy use and

A accounts for adjustments.

17

The adjustments in the calculation allow for any changes in operating conditions that may have occurred in the baseline period in comparison to the reporting period. Two types of adjustments are possible, namely routine and non-routine adjustments. Routine adjustments allow for energy-governing factors to change routinely during the reporting period. Examples of these are weather conditions and production volume. Non-routine adjustments are allocated to energy-governing factors that are not usually expected to change, for example, facility size and number of weekly production shifts [9]. Equation 2.2.3 allows for the calculation of energy savings under different operating conditions.

$$E_s = B_{peu} - R_{peu} \pm A_r \pm A_n \qquad \textbf{2.2.3}$$

where

E_s denotes the energy savings;

B_{peu} denotes the baseline period energy use;

R_{peu} denotes the reporting period energy use;

A_r accounts for routine adjustments and

A_n for non-routine adjustments.

The calculation method of the adjustment depends on whether the savings are to be reported on the energy-governing factors of the reporting period or whether the savings were to be normalized according to another set of factors. Avoided energy use for a reporting period is the savings that are reported under the conditions of the reporting period. This represents the energy that would have been used if the ECM had not been instituted. If savings are reported under reporting period conditions, the energy in the baseline period needs to be adjusted to the reporting period conditions [9]. These energy savings are calculated by the following equation:

$$E_s = B_a - R_{peu} \pm A_n \qquad \textbf{2.2.4}$$

where

E_s denotes the energy savings;

18

B_a denotes the adjusted-baseline energy;

R_{peu} denotes the reporting period energy use and

A_n represents non-routine adjustments.

The adjusted baseline energy is calculated by developing a mathematical model that correlates the baseline energy data with the independent variables. The reporting period's independent variables are then inserted into the model so that the adjusted-baseline energy can be calculated.

A fixed-conditions basis, or normalized savings, are calculated using the following equation:

$$N_s = \left(B_e \pm A_r \pm A_n \right) - \left(R_{peu} \pm A_r \pm A_n \right)$$ **2.2.5**

where

N_s denotes the normalized savings;

B_e denotes the baseline energy;

R_{peu} denotes the reporting period energy use;

A_r represents routine adjustments and

A_n represents non-routine adjustments.

The savings calculated on this basis are referred to as "normalized savings". The input variables or conditions can represent an average or "normal set of conditions" rather than those of the reporting period or baseline period [9].

The following factors need to be considered when choosing between avoided energy used and normalized savings:

- Avoided energy used style of savings are dependent on the operating conditions of the reporting period. The savings for the reporting period cannot be directly compared with that of the savings predicted under baseline conditions [5].

19

- Normalized saving are not affected by the reporting periods conditions. A set of fixed conditions have been applied to the savings and can be directly compared with the predicted savings under the same set of fixed conditions. The savings can only be reported once a full cycle of reporting period energy use is completed. This is necessary to establish the mathematical correlation between the energy recorded during the reporting period and the recorded conditions [5].

2.2.5   Measurement methodology

The IPMVP advocates four options for determining savings [5]. If the performance of only the ECM must be determined, option A or B is more suitable, which are retrofit-isolation methods. However, if savings at facility level need to be determined option C or D is more appropriate. Other criteria for choosing the most appropriate option are [12]:

- M&V rigor required to achieve the required level of accuracy;
- complexity of the ECM;
- potential for changes in the performance of the ECM;
- ECM's savings value, and
- project's allocation risk.

The following factors influence the appropriate level of M&V and therefore the associated cost [16]:

- value of the projected savings;
- complexity of the retrofitted efficiency equipment;
- amount of efficiency equipment;
- number of interactive effects related to the energy consuming system;
- required level of certainty of the savings;
- risk associated with achieving the savings for the agency and ESCO, and
- availability and compatibility of an existing energy management system.

Past experience has shown that each of the IPMVP options has the following associated cost as a percentage of the total project cost [16]:

- Option A = 1-5 percent

- Option B = 3-10 percent

- Option C = 1-3 percent (if metering is already installed)

- Option D = 3-10 percent

M&V approaches are divided into two main groups, referred to as retrofit-isolation and whole-facility methods respectively. The retrofit-isolation methods consider the equipment in isolation with regards to the rest of the facility. In contrast, whole-facility methods consider the total energy used and de-emphasize specific equipment performance [12]. This grouping is depicted in Figure 2.2.4 which graphically represents where the boundary of an ECM is drawn for the different approach options. It can be seen in the Figure that for options A and B the energy consumer is considered in isolation to the rest of the house, where as for options C and D the whole house's energy consumption is considered.

Options A, B
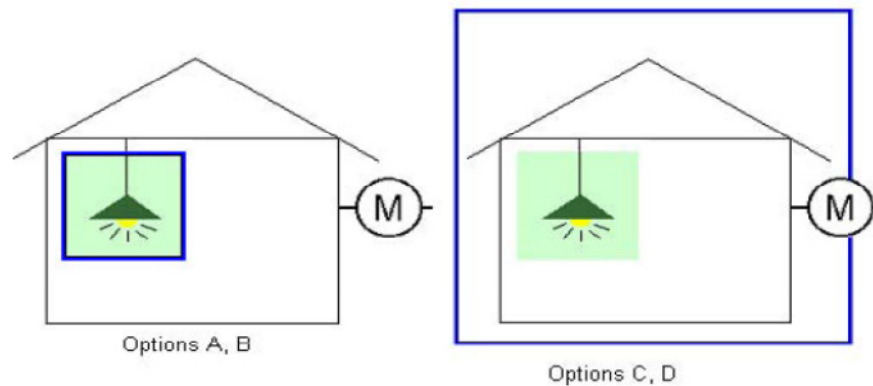
Options C, D

Figure 2.2.4: Representation of the two main methods of M&V namely, Retrofit Isolation and Whole Facility [12].

The four options are briefly described in the next four subsections.

2.2.5.1   Option A

Option A is a Retrofit in Isolation to the rest of the facility, where key parameters are measured and other parameters are estimated [9]. The measurement period ranges from short-term to continuous. This is dependent on the reporting period and the expected variation in the measured parameter. Parameters which are not measured are estimated and are based on historical data, manufacturer's specifications or engineering judgement. The savings are determined from the calculation of the baseline and reporting period energy while taking routine and non-routine

adjustments into account. An example of this option would be the determination of the savings made by the retrofitting of lights in a facility [12]. The power draw of the lights would be the key parameter measured with the operating hours estimated based on building schedules and occupancy behaviour. Such estimates should only be used where it can be shown that the combined uncertainty of the estimates does not significantly affect the overall reported savings [9]. This option allows for the narrowing of the measurement boundary which reduces the cost of monitoring by reducing the number of independent variables and static factors which would have an influence on the retrofit. This is possible because the measurement boundary only includes a section of the facility [5]. This option is appropriate for end-use technologies. The accuracy of this method is generally inversely proportional to the complexity [10]. The savings using this option are calculated with equations 2.2.6, 2.2.7 and 2.2.8 which allow for the measurement of at least one parameter and the estimation of the others for the baseline and reporting periods [9].

$$E_s = B_{peu} - R_{peu} \pm A \qquad\qquad 2.2.6$$

$$B_{peu} = f\left(V_{est1}, V_{est2}, ..., V_{msd1}, V_{msd2}\right) \qquad\qquad 2.2.7$$

$$R_{peu} = f\left(V_{est1}, V_{est2}, ..., V_{msd1}, V_{msd2}\right) \qquad\qquad 2.2.8$$

where

E_s denotes the energy savings;

B_{peu} denotes the baseline period energy use;

R_{peu} denotes the reporting period energy use;

A accounts for the adjustments;

V_{est1} denotes the first estimated value;

V_{est2} denotes the second estimated value;

V_{msd1} denotes the first measured value and

V_{msd2} denotes the second measured value.

2.2.5.2   Option B

Option B is a Retrofit in Isolation where all parameters are measured, which means that all parameters must be measured to compute the energy saved. Most ECMs can be calculated with this method. However, this option is more costly when there are complexities involved in the measurement of the variables. This method produces more accurate results when the savings patterns are more variable [9]. The savings, using this option, are calculated using the following equations:

$$E_s = B_{peu} - R_{peu} \pm A \tag{2.2.9}$$

$$B_{peu} = f\left(V_{msd1}, V_{msd2}\right) \tag{2.2.10}$$

$$R_{peu} = f\left(V_{msd1}, V_{msd2}\right) \tag{2.2.11}$$

where

E_s denotes the energy savings;

B_{peu} denotes the baseline period energy use;

R_{peu} denotes the reporting period energy use;

A accounts for the adjustments;

V_{msd1} denotes the first measured value and

V_{msd2} denotes the second measured value.

No estimates are allowed for in this option. There may also be no need for adjustments, but this depends on the length of the reporting period or the time frame between the reporting and baseline periods [9]. This approach is appropriate for savings evaluations of retrofits with performance and operation factors at a component or system level where verification of long term performance is required [12]. Compared to option A, this option is generally more complicated and costly to determine savings [5].

2.2.5.3   Option C

Option C allows for the whole facility or major sections of the facility to be evaluated for collective savings. Utility meters and sub meters are used to assess the energy impact of the ECM [12]. Energy savings for this option are estimated by using statistical models that are representative of the whole facility's energy use [10]. The expected energy savings must be greater than the unexplained energy variation that would normally occur at the whole facility. Energy savings should exceed the baseline energy by at least 10 percent when the reporting period is shorter than two years so that the savings are distinguishable from the baseline energy. The most challenging aspect of this option is identifying the changes in the facility that would require the inclusions of non-routine adjustments. This is especially true when the measurement periods are long [5]. A mathematical model is developed for each meter's energy usage pattern; this is then used to calculate the routine adjustments for the energy savings calculation. The model often correlates the energy usage to one or more independent variables such as weather, production volume and/or occupancy. An independent variable is a variable that changes regularly and has a direct influence on the energy consumption of the facility. This option should also use data for a complete year in the baseline and reporting periods, otherwise it could create a bias due to seasonal changes. Regression analysis can determine the number of independent variables that is necessary to include in the baseline data [5]. In general this option should be used with complex equipment replacement and for projects where the predicted savings are relatively large. This option is also useful when the ECM has a high degree of interaction with the rest of the facility [10].

2.2.5.4   Option D

Option D makes use of a calibrated simulation and simulation software to predict the energy usage for the baseline and or reporting period based on actual metered data. Accurate modelling and calibration, to measure energy data, pose the greatest challenges to this option. Savings are based on one or more energy usage estimate [9]. This option is especially useful when a baseline energy usage profile does not exist, as in the case of a new facility or a facility that has undergone a major retrofit [12]. The accuracy of the predicted saving depends on how well the simulation is calibrated to the metered energy data. The calibration data should include measured energy data, independent variables and static factors. After the calibration of the model, energy savings are calculated with two versions of equation 2.2.12. The first version

includes the ECM; the other not [9]. The operating conditions must be the same for both versions. The energy savings equation is as follows:

$$E_s = B_{peu} - R_{peu}$$
<div align="right">**2.2.12**</div>

where

E_s denotes the energy savings;

B_{peu} denotes the baseline period energy use and

R_{peu} denotes the reporting period energy use;

The calibration model may be replaced by actual energy data for either the baseline or reporting periods. To calculate the savings using calibration data from the reporting period, the equation is as follows [9]:

$$E_s = B_{peu} - C_{peu} \pm C_e$$
<div align="right">**2.2.13**</div>

where

E_s denotes the energy savings;

B_{peu} denotes the baseline period energy use;

C_{peu} denotes the actual calibration period energy and

C_e denotes the calibration error.

This option allows for the assessment of the estimated savings of each individual ECM in a multiple ECM project, whereas option C only allows for a single ECM. In an instance where an individual ECM must be evaluated and the metering used to collect the data is too difficult or too costly to install, this option may constitute a more feasible approach compared to options A and B [5]. When on-going savings reports are required, option D can be used for the first year. The data collected during this year can then be used to calibrate the simulation and to arrange the baseline data for option C [9].

The selection of the most appropriate IPMVP option for a particular project is made by the designer of the M&V plan. Figure 2.2.5 maps out the logical decisions that need to be made for an M&V option selection.
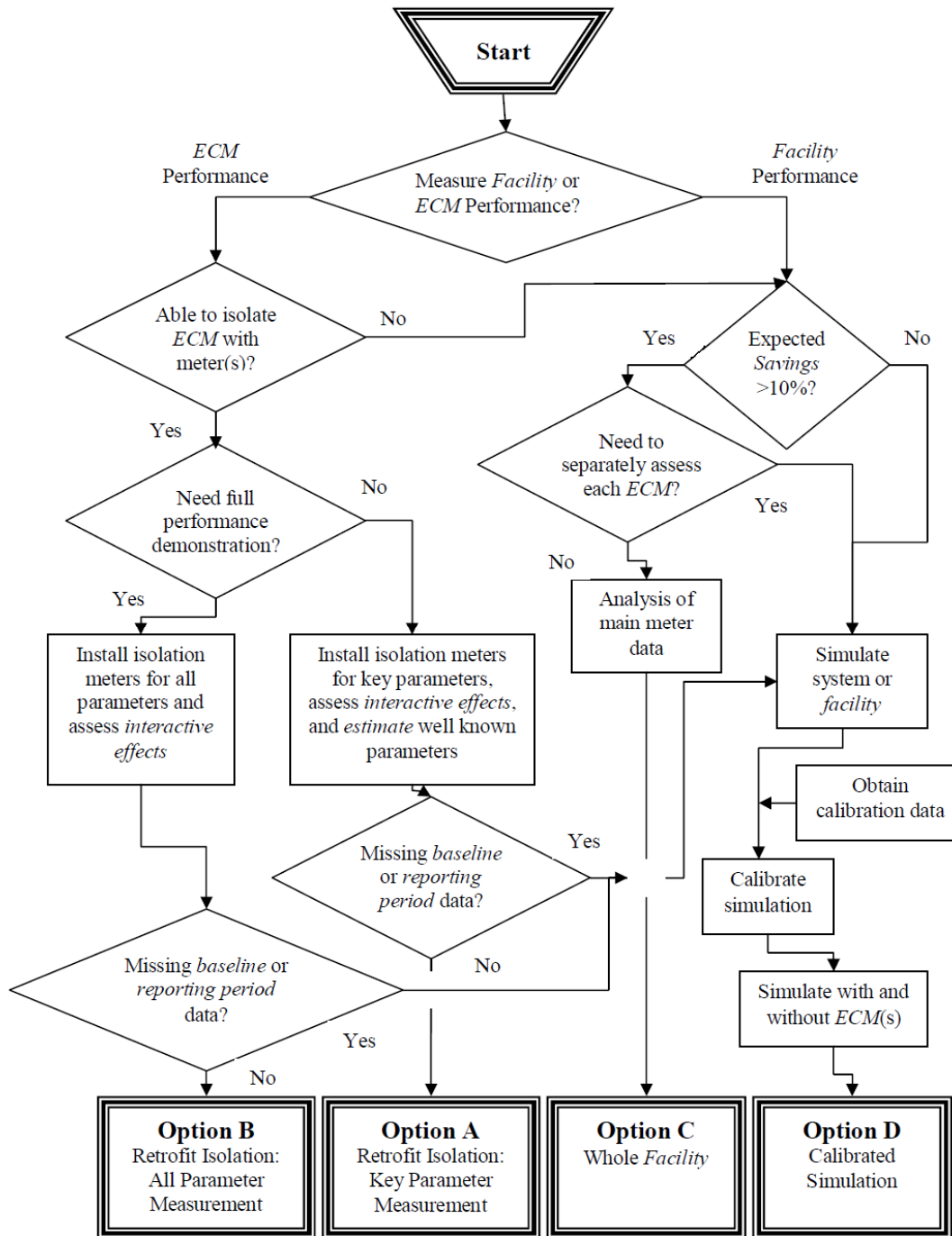
Figure 2.2.5: Logic diagram for the M&V option selection process [5].

26

2.2.6 Associated cost savings

Cost savings are calculated by applying a price to the associated energy saving. This is done by the application of the following equation:

$$C_s = C_b - C_r \qquad\qquad\qquad\qquad\qquad \textbf{2.2.14}$$

where

C_s denotes the cost saving as a result of avoided energy use;

C_b denotes the cost of the energy used during the baseline period plus any adjustments and

C_r denotes the cost of the energy used in the reporting period plus any adjustments.

The same pricing schedule must be applied to C_b and C_r when calculating cost savings provided that the pricing schedule remains the same for both periods [5]. The M&V process advocates the practice of stating confidence levels with all estimated saving values but to date, the IPMVP fails to give a detailed description of how this should be done. The following section investigates methods of statistical inference that would prove to be a useful tool in the M&V process.

2.3   Statistical inference

2.3.1   Parametric and non-parametric inference

"Statistics is the science that deals with the collection, summarization, presentation and interpretation of data." [17]. The focal point of statistics is data; from this we can make inferences. Statistical inference is the method of using data to infer the distribution that generated the data [18]. A statistical model $\Im$ is a set of distributions. A parametric model is a set of $\Im$ that can be parameterized by a finite number of parameters whereas a non-parametric model cannot. The focus of the two disciplines is different. In a parametric density family $f(\cdot | \theta)$, such as the two parameter Normal family $N(\mu, \sigma^2)$ where $\theta = (\mu, \sigma^2)$, the focus is on obtaining the best estimator $\hat{\theta}$ of θ. In the non-parametric case the focus is on obtaining $\hat{f}(\cdot)$, a good estimate of the density function $f(\cdot)$ itself [19]. Parametric models and curves rely on

27

prior knowledge of the equations that describe the data. Non-parametric models and curves are driven by the nature of the data and thus they allow the data to speak for itself. Non-parametric methods negate the need for model specifications and in doing so reduce the risk of misinterpretation of the data. In the parametric approach, strong assumptions are made on the formal structure of the data. This is not only the strength but also the weakness of parametric modelling. From a strength point of view, great gains in efficiency are possible by specifying a specific model. However, when the model is wrongly specified the inferences can be worse than useless. The classical approach to summarising the nature of a data set is to assume a parametric model. Smoothing methods can also be described as a compromise between parametric and non-parametric inference. The purpose of smoothing methods such as kernel and spline smoothing is to control the variance error, making it easier to see the underlying structure in the data [20].

2.3.2   Introduction to univariate density estimation

Density estimation is a very important concept in statistics and can be described as the construction of an estimate for the unknown density function of the recorded data. Given that a random variable X has a probability density function f, the probability of X is mathematically described as:

$$P(a < X < b) = \int_{a}^{b} f(x)dx \qquad \text{for all } a < b \qquad\qquad \textbf{2.3.1}$$

Density estimation can be divided into two groups, namely parametric and non-parametric density estimation. In parametric density estimation, a rigid assumption is made that the data comes from a known parametric family of distributions. In contrast to this, the goal of non-parametric density estimation is to estimate the density function with as few assumptions of the true function as possible. Non-parametric estimation can uncover features such skewness, multimodality and relationships between variables in the multivariate case which parametric estimation might not reveal. The different types of non-parametric density estimation include histograms, kernel, frequency polygons, spline estimators, orthogonal series and penalised likelihood. A detailed description of these estimators is discussed in Silverman, Scott and Simonoff [21, 19, and 17]. Kernel-based estimators are the most popular non-parametric estimators because in comparison, they are easy to interpret and implement.

2.3.3   Kernel density estimation

The univariate kernel density estimation will be discussed first. Following this discussion, an account of multivariate density estimation will be given, of which the bivariate case is of particular importance to the work described in this thesis. The bivariate density estimate draws a relationship between the dependent and independent variables.

A kernel is a smoothing or weight assigning function and satisfies the following condition [17]:

$$\int_{-\infty}^{\infty} K(x)dx = 1 \qquad\qquad 2.3.2$$

where

$\quad K \qquad\qquad$ denotes the kernel function.

In most of the applications, $-K$ will be a unimodal, symmetric probability density function.

A univariate kernel density estimator for a random sample $X_1,\ldots, X_n$ is described mathematically as follows:

$$\hat{f}(x;h) = n^{-1}\sum_{i=1}^{n} K_h(x - X_i) \qquad\qquad 2.3.3$$

where

$\quad K_h \qquad\qquad$ denotes the scaled kernel function;

$\quad n \qquad\qquad$ denotes the number of samples and

$\quad h \qquad\qquad$ denotes the bandwidth which is a positive, non-random number.

The scaled and unscaled kernels are related by

$$K_h(x) = h^{-1}K\left(h^{-1}x\right) \qquad\qquad 2.3.4$$

This definition could be described in simpler terms as the sum of the bumps at each sample point. K determines the shape of the bumps and h determines their width. Under mild conditions, the kernel estimate converges to the true density function. In Figure 2.3.1, scaled

density functions are placed at each of the samples which are represented by the red dots. The red line represents $\hat{f}$ and is the summation of the individual density functions at that sample point. Samples which are closer in value result in a higher overall density value in that region of samples [22].
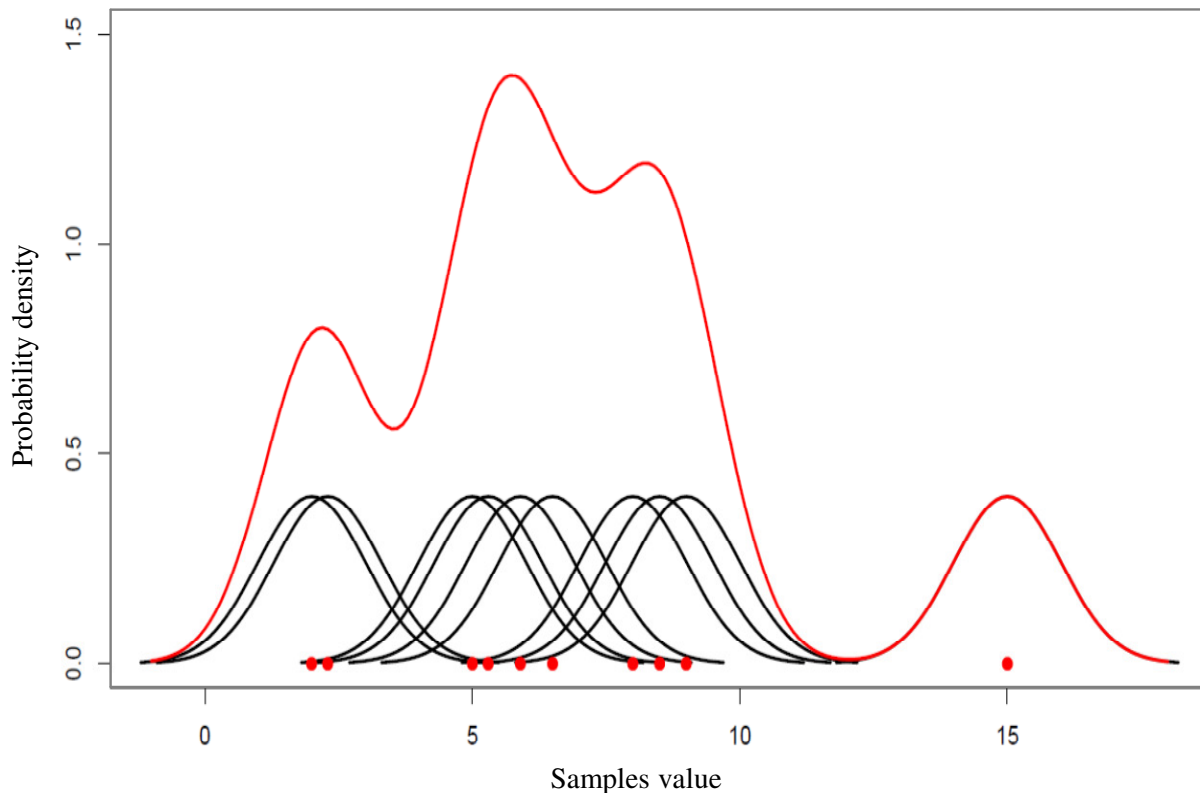
Figure 2.3.1: Estimated density function using kernel density estimation [22].

In kernel density estimation, two specifications need to be decided on. The first is the kernel; the second is the bandwidth. The performance of a kernel is measured by its asymptotic mean integrated squared error (AMISE) which is a global error criteria of how closely the estimated density function matches the true density function. The Epanechnikov kernel was designed to minimise AMISE. For that reason the Epanechnikov kernel is optimal and the other kernels efficiencies are measured relative to it [21]. Figure 2.3.2 depicts plots of different types of symmetrical kernel functions.
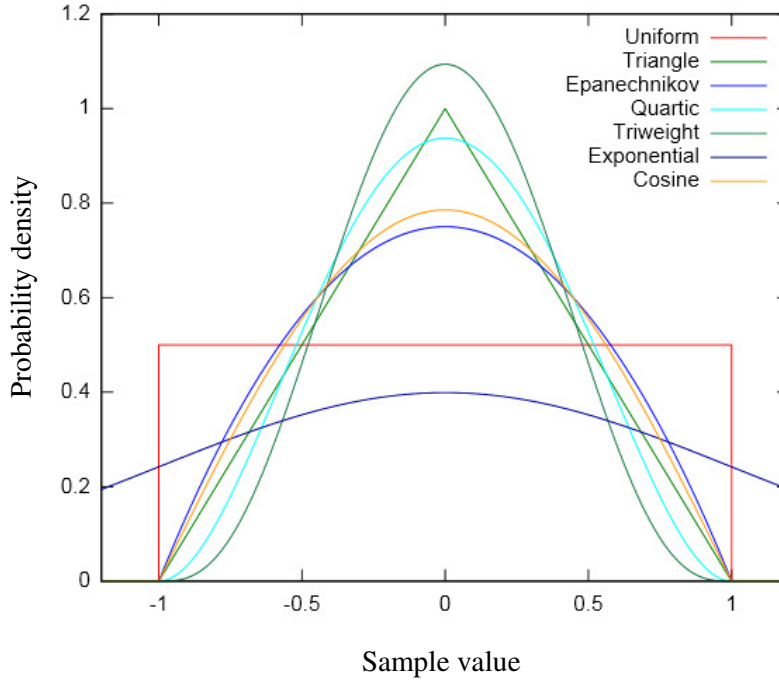
Figure 2.3.2: Different types of symmetrical unimodal kernel density functions [23].

A list of kernel functions is summarised in Table 2.3.1 along with their corresponding equation and efficiency.

Table 2.3.1: Equations of the different density functions and their relative efficiency [17].

Kernel	$K(x)$	For	Otherwise	Efficiency
Epanechnikov	$\dfrac{\frac{3}{4}\left(1-\frac{1}{5}x^2\right)}{\sqrt{5}}$	$|x|<\sqrt{5}$	0	1
Biweight	$\dfrac{15}{16}\left(1-x^2\right)^2$	$|x|<1$	0	0.9939
Triangular	$1-|x|$	$|x|<1$	0	0.9859
Gaussian	$\dfrac{1}{\sqrt{2\pi}}e^{-\left(\frac{1}{2}\right)x^2}$			0.9512
Rectangular	$\dfrac{1}{2}$	$|x|<1$	0	0.9295

31

The choice of kernel does not nearly have the same degree of impact on the estimate as the choice in bandwidth. There is a danger of under- or over-smoothing when choosing the bandwidth. It can be seen from Figures 2.3.3 to 2.3.6 that a variation in the bandwidth has a significant effect on the density estimation. The estimated density functions range from over- to under-smoothed. It is therefore necessary to use an automatic data-driven bandwidth selector [24]. AMISE and automatic bandwidth selectors are discussed in the next section.
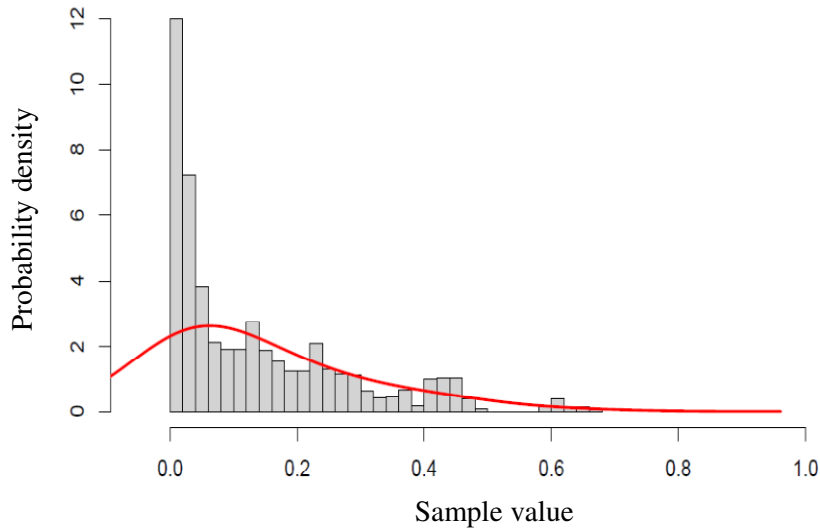
Figure 2.3.3: Estimated density function with a kernel bandwidth of 0.1 (over smoothing) [22].
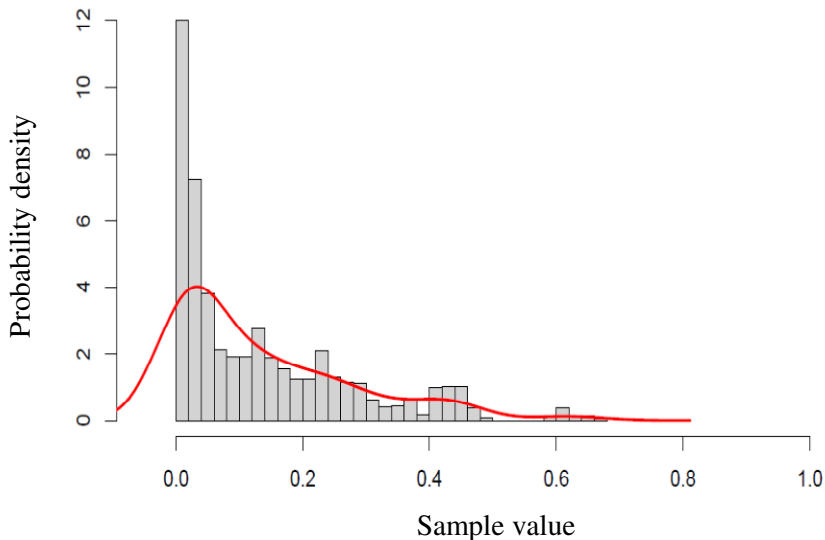
Figure 2.3.4: Estimated density function with kernel bandwidth of 0.05 [22].

32
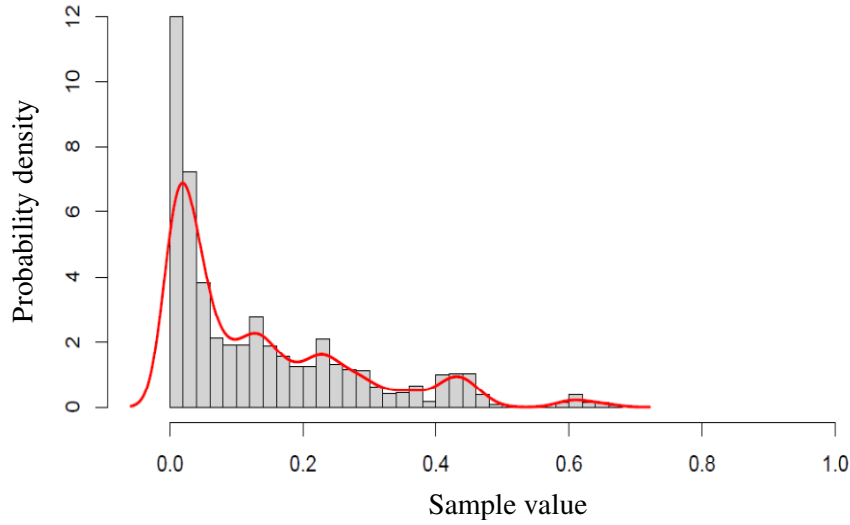
Figure 2.3.5: Estimated density function with a kernel bandwidth of 0.02 [22].
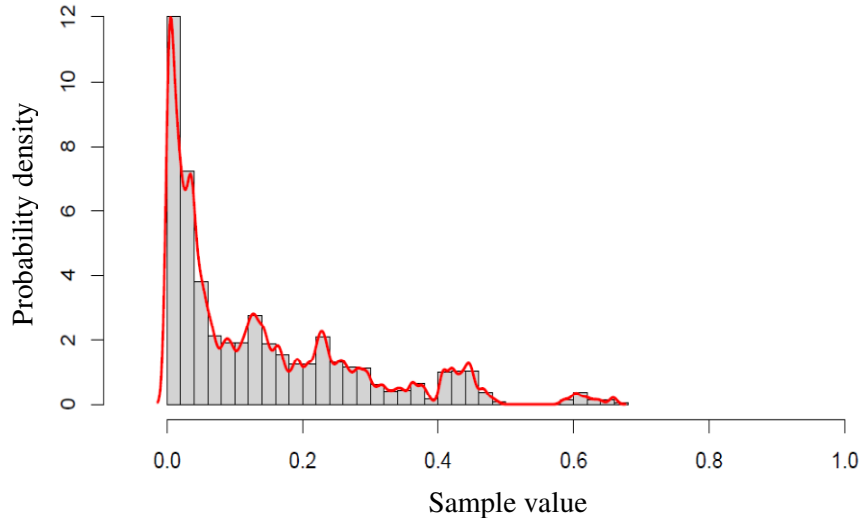
Figure 2.3.6: Estimated density function with a kernel bandwidth of 0.005 (under smoothing) [22].

The mathematical description of a d - dimensional multivariate kernel density estimate from a random sample $X_1, \ldots, X_n$ is as follows [19]:

$$\hat{f}(x; H) = n^{-1} \sum_{i=1}^{n} K_H(x - X_i)$$ **2.3.5**

where

    x denotes $x_1, \ldots, x_d$;

    X_i denotes $X_{i1}, \ldots, X_{id}$; $i = 1, \ldots, n$

33

K_H denotes the scaled kernel function;

n denotes the number of samples and

H denotes the $d \times d$ bandwidth matrix which is symmetric and positive definite.

The scaled and unscaled kernels are related by:

$$K_H(x) = |H|^{-\frac{1}{2}} K\left(H^{-\frac{1}{2}}x\right)$$

                                                                                         2.3.6

In the bivariate case, the same principle of summing density estimates is applied, as was done in the univariate case. Figure 2.3.7 is produced from the sample data set:

$$X_1 = (7,3), X_2 = (2,4), X_3 = (4,4), X_4 = (5,2), X_5(5,5)$$

and has a bandwidth matrix of [25]:

$$H = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$$

In Figure 2.3.7, the left plane the individual kernels are placed at the sample data points. In the right plane the individual kernels are summated to give a kernel density estimate.
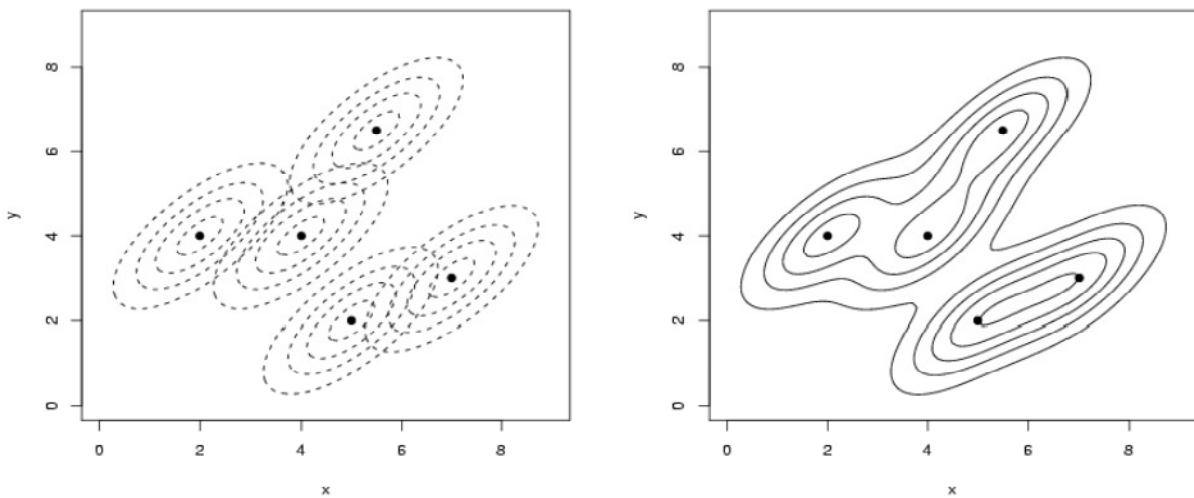
Figure 2.3.7: In the left plane individual kernels are placed at the sample points and in the right plane the individual kernels are summated to give a kernel density estimate [25].

34

2.3.4   Error criteria for kernel density estimates

There are methods to measure the discrepancy between the estimated density function and true density function. This measure is also referred to as the performance of the density estimate. The bandwidth selector plays a very important part in the performance of the density estimate. The most commonly used, and most mathematically tractable, global performance measure over the whole real line is the mean integrated square error (MISE) [26]:

$$MISE \equiv MISE\hat{f}\left(\cdot;H\right) = \mathrm{E}ISE\hat{f}\left(\cdot;H\right) = \mathrm{E}\int_{R^d}[\hat{f}\left(x;H\right) - f\left(x\right)]^2 dx \qquad \textbf{2.3.7}$$

where

$MISE$ denotes the mean integrated square error;

EISE denotes the expected value of the integrated square error and

H denotes the positive definite $d \times d$ bandwidth matrix.

The objective is to find H that minimizes the MISE, which can be written as follows:

$$H_{MISE} = \arg\min MSE\hat{f}\left(\cdot;H\right) \qquad \textbf{2.3.8}$$

where

H_{MISE} denotes the bandwidth relative to the mean integrated square error and

arg min   denotes the argument minimum.

It is very difficult to calculate H_{MISE} because MISE does not have a closed form solution. The exception to this is if f is a normal mixture and K is a normal kernel density [24]. This exception will be used to calculate the initial optimum bandwidths for the kernel density estimates of the impacts resulting from the simulated geyser ECMs. In other words, it will be assumed for the initial density estimates that the true density functions will have normal distributions. A Gaussian distribution is used as the initial kernel. The equation for this optimal bandwidth will be described in the next section. The next step is to find a tractable estimate of the MISE to overcome the problem of it not having a closed form solution [25].

The MISE can be rewritten in the following form:

$$MISE\hat{f}\left(\cdot;H\right)=\int_{R^d} MSE\hat{f}\left(x;H\right)dx$$

$$=\int_{R^d} Var\hat{f}\left(x;H\right)dx+\int_{R^d} Bias^2\hat{f}\left(x;H\right)dx$$

2.3.9

where

Var denotes the variance and

$Bias$ denotes the bias.

The expected value of the kernel density estimated is:

$$E\hat{f}\left(x;H\right)=EK_H\left(x-X\right)=\int_{R^d} K_H(x-y)f(y)dy=(K_H*f)(x)$$

2.3.10

where

$*$ denotes the convolution operator.

The bias is

$$Bias\hat{f}(x;H)=(K_H*f)(x)-f(x)$$

2.3.11

The variance is

$$Var\hat{f}(x;H)=n^{-1}[(K_H^2*f)(x)-(K_H*f)(x)^2]$$

2.3.12

Now MISE can be expressed in terms of the bias and variance as follows [25]:

$$MISE\hat{f}(x;H)=n^{-1}\int_{R^d}[(K_H^2*f)(x)-(K_H*f)(x)^2]dx+\int_{R^d}[(K_H*f)(x)-f(x)]^2 dx$$

$$=n^{-1}R(K)\left|H\right|^{-\frac{1}{2}}+(1-n^{-1})\int_{R^d}(K_n*f)(x)^2 dx-2\int_{R^d}(K_H*f)(x)f(x)dx+R(f)$$

2.3.13

where

$$R(g)=\int_{R^d} g(x)^2 dx$$

2.3.14

36

In this form, an asymptotic approximation can be calculated for MISE, which is known as AMISE and it is tractable so H_{MISE} can be calculated. To be able to derive the asymptotic approximation of MISE, the multivariate Taylor expansion must be applied to equation 2.3.13. A more detailed explanation of the manipulation of MISE and AMISE can be found in Wand and Jones for the univariate and multivariate cases respectively [24].

2.3.5 Bandwidth selection

One way to select a bandwidth is by visual interpretation. This can be done by selecting a large bandwidth and selectively decreasing it until the fluctuation in the density function seems more random than structural. This method assumes that there is a prior knowledge of the true structure. This method is time consuming when many separate density functions need to be evaluated. In most cases it is more beneficial to have automatic bandwidth selection. In the univariate case, a constant bandwidth is described by a scalar value whereas in the multivariate version it is described by a matrix. This bandwidth matrix is responsible for the degree and direction of smoothing of the density estimate. The bandwidth selection will initially be constrained to the diagonals of the bandwidth matrix. This has been the most common practice in the past. Duong developed a method to estimate the off-diagonal elements of the bandwidth matrix for multivariate density estimations [26]. Based on the assumption that the density estimate will be Gaussian in form, the following mathematical equations were used to select the diagonals of bandwidth matrix [27]:

$$h_p = \sigma_p (1-\rho^2)^{\frac{5}{12}} (1+\frac{\rho^2}{2})^{-\frac{1}{6}} n^{-\frac{1}{6}}$$ 2.3.15

and

$$h_T = \sigma_T (1-\rho^2)^{\frac{5}{12}} (1+\frac{\rho^2}{2})^{-\frac{1}{6}} n^{-\frac{1}{6}}$$ 2.3.16

where

h_p denotes the bandwidth for the demand variable;

h_T denotes the bandwidth for the ambient temperature variable;

σ_p denotes the standard deviation of demand;

σ_T denotes the standard deviation of ambient temperature and

ρ denotes the correlation coefficient.

These bandwidth values minimise the MISE of $\hat{f}$ but only if f and K_H are Gaussian [24].

2.3.6   Limitations in kernel density estimation

One of practical limitations of the kernel density estimator is when it is used on data from long tailed distributions. A fixed bandwidth implemented in this case tends to cause spurious noise in the tails of the estimate. If the bandwidth is increased enough to smooth the tail, essential detail is lost in the main part of the distribution [21]. There are techniques to overcome this problem of which two are the nearest neighbour method and the variable kernel method. In density estimations that are bound, conventional symmetrical kernel causes excessive bias near the boundary or boundaries [21]. A method has been developed to alleviate this problem by making use of a non-negative kernel such as a gamma or beta distribution near the boundaries. More limitations exist in kernel density estimation, but these two cases would potentially have a noteworthy impact on this work.

2.4   Load forecasting

2.4.1   Overview

In past decades, load forecasting was a much simpler task involving only straight-line extrapolations of historical energy consumptions. The required accuracy lay between a high and low estimate of predicted future load and the difference between the two increased as the forecasted period increased. According to Meetamehra there is an urgent need for precision in demand forecasting [28]. If a load forecast in the past was under estimated, turbine generating power plants were built to take care of the shortfall. The gestation period for such a plant is at most three years where as a coal fired plant is seven to twelve years. These turbine power plants run on oil which cost approximately $20 a barrel in the 1990s [29]. One of the ways that Eskom, as referred to in Chapter one, solved their immediate power shortfall crisis was to build gas turbine peaking plants that ran on fuels that cost approximately $90 a barrel in 2008 [29]. The

situation was so severe before the economic recession that these plants at times had to be operated as base loads plants which significantly contributed to the questionability of Eskom's solvency in early 2008 [30]. An electrical network operating at under capacity can lead to brown outs and even blackouts which can put the whole supply network at risk. In the other extreme, if over capacity was built the power station was moth balled and future growth would utilise this capacity. Presently, base load power stations are enormously expensive to build and finance. Environmentally compatible technologies are developing at an unprecedented rate, which means that power stations need to be built and commissioned at the exact time that they are required to take advantage of this technology and be as environmentally friendly as possible. Medium term forecasts indicate when these power stations must be built. Medium term forecasts make assumptions about technologies as well as social and economic systems locally and globally [3]. Short-term demand forecasts have the important role of setting the tariff structure. The nature of forecasting has changed in recent times. It is no longer adequate to only predict the peak demand and total energy used. It is now necessary to predict the whole load shape on an hour to hour basis [28]. DSM is used to alter this load shape. DSM engineers must strike a balance between maximum energy conservation and avoiding undue demand restrictions. Demand forecasting models are also used to evaluate the impact which new technologies will have on energy demand and to what effect rising energy prices will stimulate energy efficiencies. Demand forecasts are also done for each customer category and voltage level. This is done to aid social reforms whereby larger commercial and industrial customers subsidise low income customers [28].

Spatial load forecasts yield information on electrical demand that include location, magnitude and timing of the anticipated demand [3]. Table 2.4.1 provides a list of the different levels of equipment and their associated planning time frames for short and long range.

Table 2.4.1: Typical lead times for transmission and distribution system short and long range planning [3].

Level	Short range forecasting (years ahead)	Long range forecasting (years ahead)
Large generation (> 150 MVA)	13	30
Small generation (< 50 MVA)	5	20
Transmission (138 kV and above)	9	25

Level	Short range forecasting (years ahead)	Long range forecasting (years ahead)
Sub-transmission (34 kV – 138 kV)	7	20
Distribution substations	5	20
Primary three-phase feeders	3	15
Single-phase laterals	0.5	n.a.
Service transformers	0.2	n.a.

The most important aspect of short range planning and subsequent project development is the accurate predicting of future load growth. Typically the planning will only occur at times equal to lead times of the planned project. In contrast to short range planning, long range planning is not concerned with system additions being installed in time to meet needs, but rather that system additions will have lasting investment and performance value. For those in charge of planning there is always an uncertainty about future development and resulting load growth. On the one hand, they do not want to commit any resources or facilities for load growth that may not happen but on the other hand they cannot ignore that there are lead times required to build these facilities [3].

2.4.2 Density estimation in short-term electrical load forecasting

In a paper presented by Charytoniuk et al., a statistical model of residential load demand was generated by using non-parametric probability density estimation [27]. This method used customer demand data which was also used to determine customer classes. When calculating the PDFs for a group of customers, it is important that they all belong to the same customer class. It is then necessary to determine which variables are statistically influential on electricity demand. Examples of possible independent variables are: ambient temperature, wind speed, light intensity levels and humidity. The normalized demand and variable data is then divided into time intervals such as 30 minutes. The data is then further grouped together into characteristic day type of either weekday or weekend day. Finally the day type is grouped together into the seasons of the year namely: summer, winter, spring and autumn [27].

The PDF for a particular 30 minute period t, within a 24-hour day was generated with a PDF estimator with Gaussian kernel function [27]. The PDF is generated using the following equation:

$$f_t(p,T) = \frac{\sum_{i=1}^{n} \exp-\left[\frac{(p-p_i(t))^2}{2h_p^2(t)} + \frac{(T-T_i(t))^2}{2h_T^2(t)}\right]}{2\pi n h_p(t) h_T(t)} \qquad \textbf{2.4.1}$$

where

p denotes the normalized demand;

T denotes the ambient temperature in degrees Celsius;

n denotes the number of samples;

h_p denotes the bandwidth for the demand data and

h_T denotes the bandwidth for ambient temperature data.

The smoothing parameters h_p and h_T are calculated with the following equations respectively [27]:

$$h_p(t) = \sigma_p(t)(1-\rho^2(t))^{\frac{5}{12}}\left(1+\frac{\rho^2(t)}{2}\right)^{-\frac{1}{6}} n^{-\frac{1}{6}} \qquad \textbf{2.4.2}$$

$$h_T(t) = \sigma_T(t)(1-\rho^2(t))^{\frac{5}{12}}\left(1+\frac{\rho^2(t)}{2}\right)^{-\frac{1}{6}} n^{-\frac{1}{6}} \qquad \textbf{2.4.3}$$

where

$\sigma_p(t)$ denotes the standard deviation of normalized demand data;

$\sigma_T(t)$ denotes the standard deviation of temperature data and

$\rho(t)$ denotes the correlation coefficient.

A deterministic value of the normalized demand can be calculated for the 30 minute period from the resultant PDF by taking the conditional expected value of it [27].

The equation is as follows:

$$\hat{p} = E\left(p \,|\, T_t\right) = \int p \cdot f_{wt}\left(p \,|\, T_t\right) dp \qquad \textbf{2.4.4}$$

The expected demand in kVA for that period is calculated as follows

$$P = \hat{p} \cdot P_{avg} \qquad \textbf{2.4.5}$$

where

P_{avg} denotes the monthly average demand.

This result does not give an indication of probability. From the PDF it can be seen for which values of normalized demand the probability density is a maximum. To find the probability for those values of normalized demand, the area under the graph between these values has to be calculated. For this reason it may be more intuitive to view the Calmative Distribution Function (CDF) of the particular PDF because the probability can be read directly off the graph for the related normalized demand. These normalized demand values are then multiplied by the average demand to get a probability for a particular range of demand values. The CDF is the integral of the PDF and the following equation is used to calculate it:

$$F_W\left(w\right) = \int_{-\infty}^{w} f_W\left(\xi\right) d\xi \qquad \textbf{2.4.6}$$

where

F_W denotes the cumulative density function and

f_W denotes the probability density function.

Temperature and demand data was obtained for a group of residential electrical consumers. The method described by Charytoniuk et al. was followed and a probability density plot was generated as described in the above paragraph [27]. Figure 2.4.1 shows a typical three

dimensional surface plot of demand versus time profile for a residential group of customers during weekdays in November. The two daily demand peaks are indicated by the red arrows. The morning peak is larger in magnitude with a shorter duration in comparison to the evening peak.
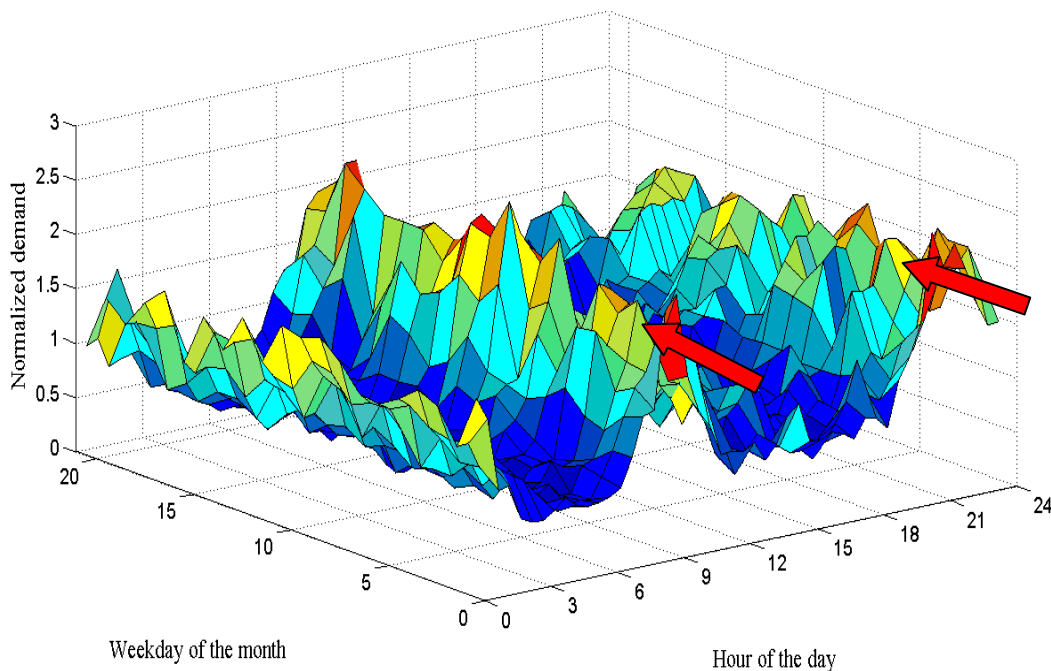
Figure 2.4.1: Three dimensional surface plot of demand versus time profile for a residential group of customers during weekdays in November.

Ambient temperature was chosen as the independent variable and demand would then be the dependent variable. Figure 2.4.2 depicts a three dimensional plot of ambient temperature versus time of day and day of the year. The cooler temperatures of the winter months can be seen in the plot. From this demand and temperature data a bivariate probability density function was generated for a single 30 minute period of a weekday in November. Figure 2.4.3 depicts this PDF which relates the probability density to ambient temperature and normalized demand. The statistical analysis of the demand for a customer or customer class compares many of these PDFs which relate all of the statistically significant variables to probability density and normalized demand for specific period of the day, day type and season of the year.
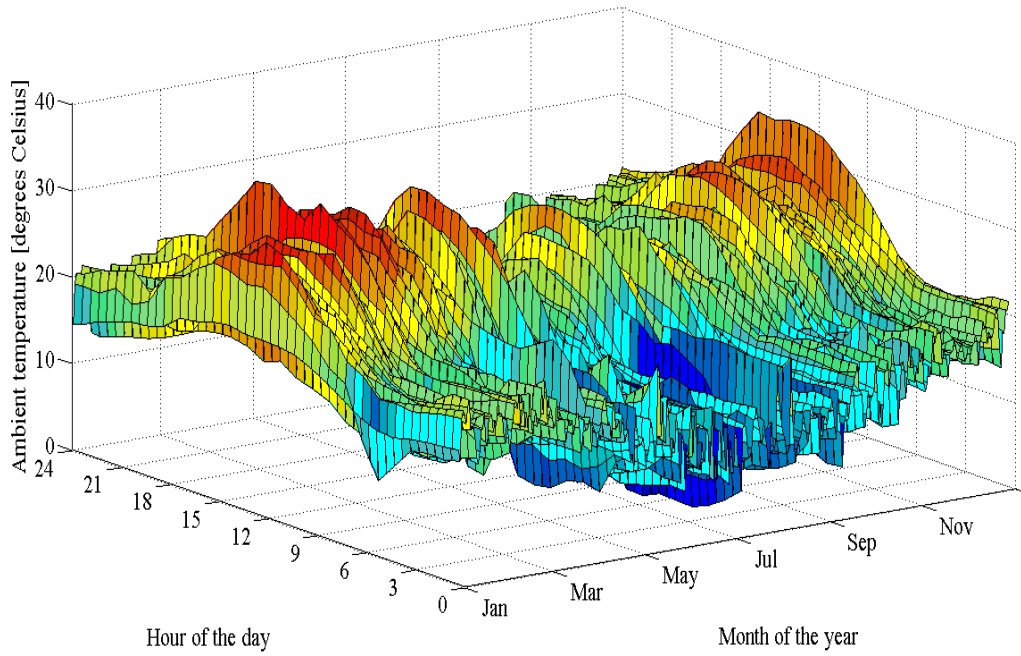
Figure 2.4.2: Three dimensional plot of the ambient temperature data that was recorded in close vicinity to the demand data of the energy consumers.
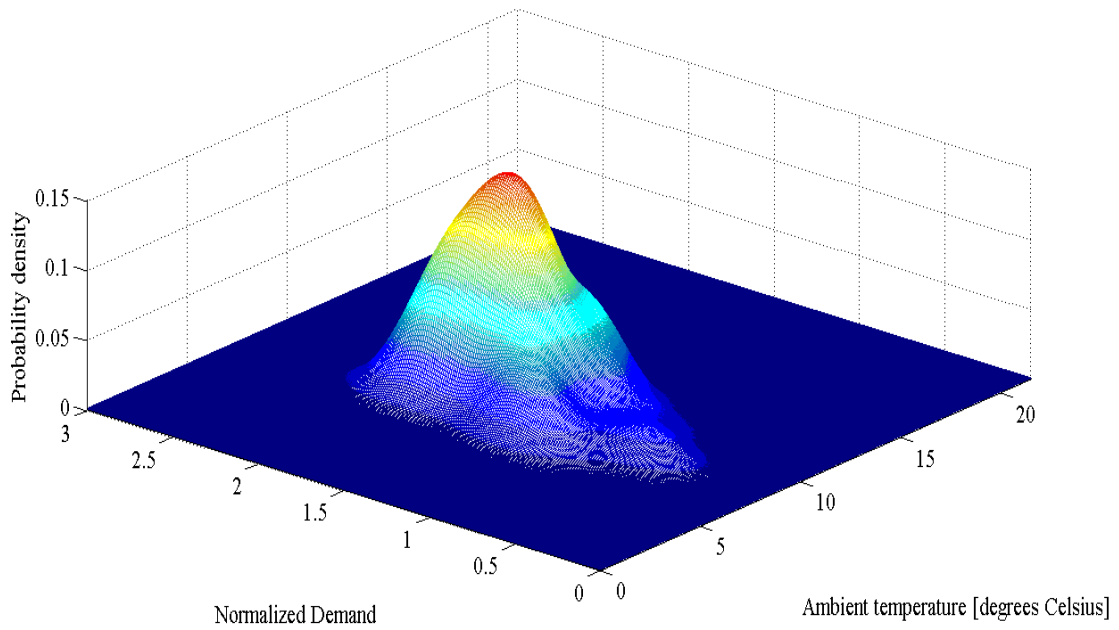
Figure 2.4.3: Bivariate probability density plot of normalized demand versus ambient temperature for a residential group of customers at 08h30 on a weekday in November.

2.5 Hot water consumption and water heaters

2.5.1 Overview

For many countries in the world domestic hot water heating is a significant component of the energy consumed by the residential sector. In South Africa, hot water heating is the largest energy consumer in the residential sector [31]. Figures as high as 40 to 50 percent of the monthly electricity account for middle to upper income households as a result of hot water heating have been reported [7]. It is a known fact that different income and housing density sectors have different hot water consumption patterns. This contributes to the different baseline electrical demand profiles that exist for the different housing and income sectors. South Africa has diverse climatic regions ranging from subtropical, temperate, semi desert to desert. The climatic regions have different annual ambient temperature profiles. Ambient temperature has an influence on hot water consumption patterns within the residential sectors which is evident in the hot water consumption profiles presented later in this section where the lower winter ambient temperature is associated with an increase in hot water demand. The hot water consumption drives the electrical consumption of the geyser. Ambient temperature also influences the standing losses and inlet water temperature to a geyser. The climatic regions in South Africa have very different solar radiation profiles which have a great impact on the performance of solar water heaters. Solar water heaters make use of a backup electrical element to heat the water when the solar radiation is inadequate to do so. Varying levels of solar radiation result in different baseline electrical demand profiles for hot water heaters based on the location within the country. In Chapter three a model of a conventional hot water heater is presented that was developed in SIMULINK to determine the extent which hot water demand and temperature have on baseline electrical demand.

2.5.2 Hot water consumption

The Central Statistical Service in South Africa maintains that 66.7 percent of dwellings in South Africa are defined as houses or parts of houses [32]. In a study that was conducted by Meyer et al., the domestic hot water consumption of 300 houses in Johannesburg, South Africa, was evaluated for a period of one year. Each third of the 300 houses represented samples from high-, medium- and low-density housing respectively [31]. Medium- and low-density housing are indicative of developed communities while high-density housing is indicative of developing

communities. The numbers of occupants for each associated housing density are given in Table 2.5.1.

Table 2.5.1: Number of occupants per house and the number of houses per square km for high-, medium- and low-density housing [31].

Category of house	Density (houses per square km)	Occupants/house
Low-density	845	3.1
Medium-density	2029	3.8
High-density	3857	6.2

In the Meyer et al. study, hot water consumption profiles were grouped together in four sets. The summer and winter profiles catered for the seasonal consumption changes. Weekday and weekend profiles catered for day of the week changes in consumption. The study concluded that there was an increase of 70 percent in hot water consumption per person per day for the winter months compared to the summer months. It was also concluded that people occupying low-density housing used 3.6 times more hot water than people that occupied high-density housing [31]. This is illustrated in Figure 2.5.1.
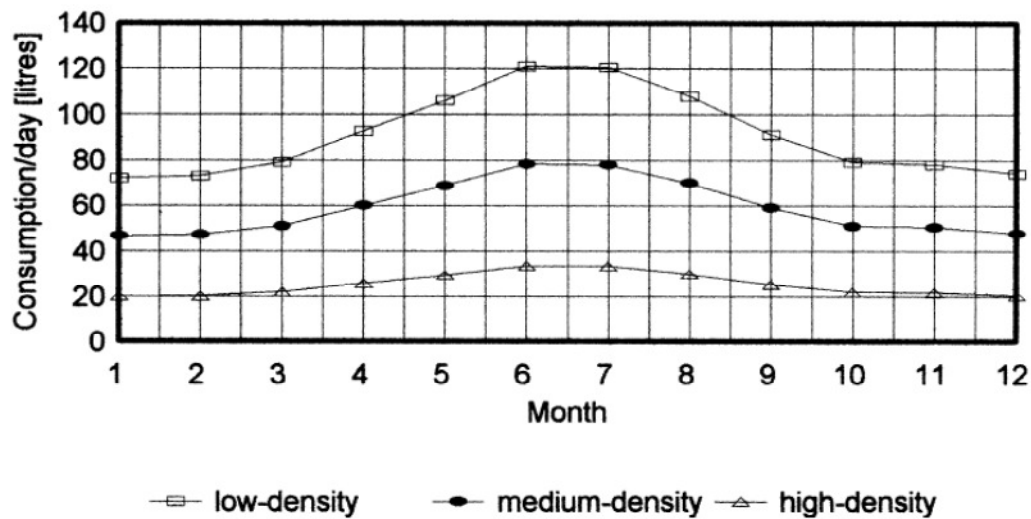
Figure 2.5.1: Hot water consumption per person per day for low-, medium- and high-density houses [31].

46

The maximum standard deviation from the recorded monthly average values of hot water consumption per person per day was ± 11 percent for summer and ± 22 percent for winter.

Thirty of these houses were fitted with digital flow meters with logging capabilities just upstream of their geyser inlet pipe. There were ten houses from each density category and the logging interval was 60 minutes. The maximum standard deviation of the averages for the 60 minute data was ± 17 percent during summer and ± 30 percent for winter [31]. These values of maximum standard deviation are used in Chapter three to generate the hot water consumption data. The average litres of hot water used per person for each hour of the day for each characteristic day is given in Figures 2.5.2 to 2.5.5. It can be seen in Figure 2.5.2 that there were two peak consumption periods, namely 06h00 to 08h00 and 19h00 to 21h00, depending on the density category.
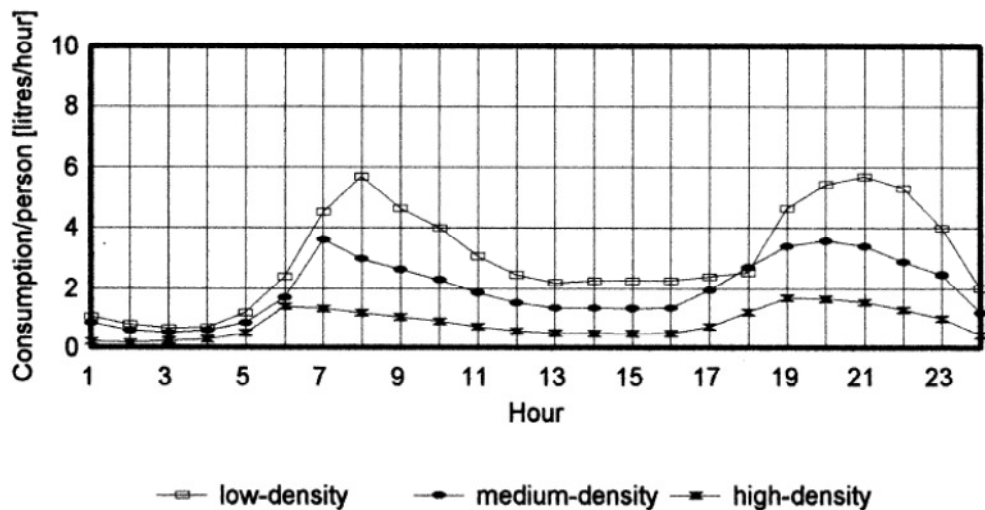
Figure 2.5.2: 60 minute hot water consumption data, per person per day during weekdays in summer [31].

The morning peak has a shorter duration than the evening peak. The time difference between the different peaks for the different housing densities is attributed to the time taken to get to work from the different housing density locations. It was found that people in low-density housing lived closer to their place of work where as people in high-density housing had to make use of public transport. As a result of this, the people living in high-density housing had to wake up earlier and go to bed earlier to get to their places of work [31].

47

Table 2.5.2 depicts the different average commuting times for people living in different housing densities.

Table 2.5.2: Average commute time from home to place of work for people that live in high-, medium- and low-density housing [31].

Housing density	Time to get to work (minutes)
Low	21
Medium	34
High	85

Figure 2.5.3 shows that the morning peak moved to 09h00 and the evening peak to between 21h00 and 22h00 on a weekend. This is attributed to people not having to get up as early during a weekend day to get to work [31].
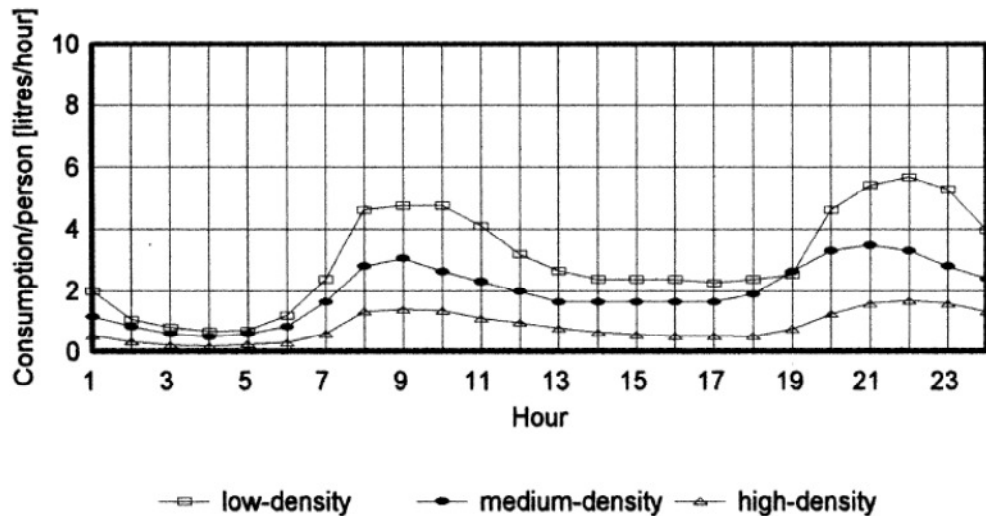
Figure 2.5.3: 60 minute hot water consumption data, per person per day during weekend days in summer [31].

In Figure 2.5.4 it can be seen that the average consumption increased substantially during the winter months. The morning peak occurred at the same time as the summer peak. The winter evening peak moved an hour earlier compared to the summer peak for the different housing densities [31].
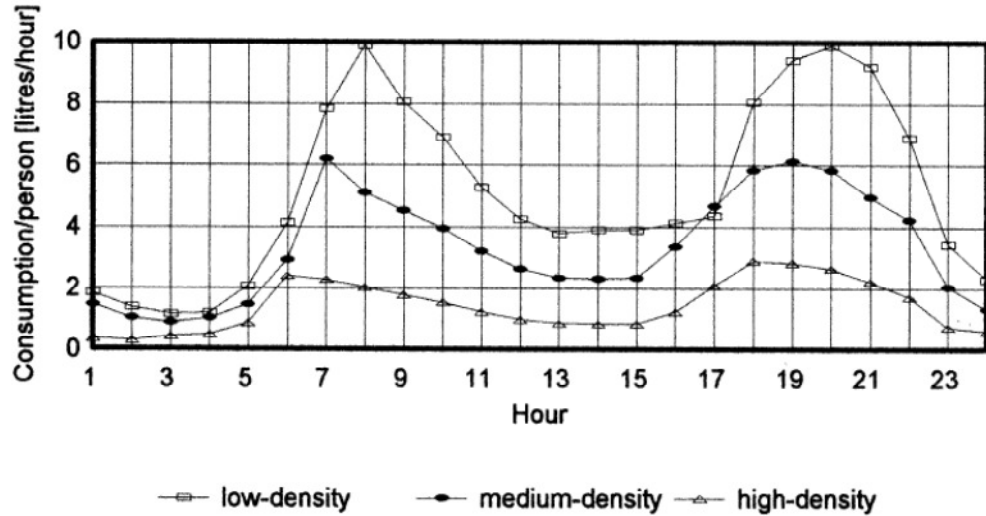
Figure 2.5.4: 60 minute hot water consumption data, per person per day during weekdays in winter [31].

In Figure 2.5.5, the morning peak moved to 09h00, which is two hours later when compared to the winter weekday consumption profile and the evening peak also moved later, to 21h00 [31].
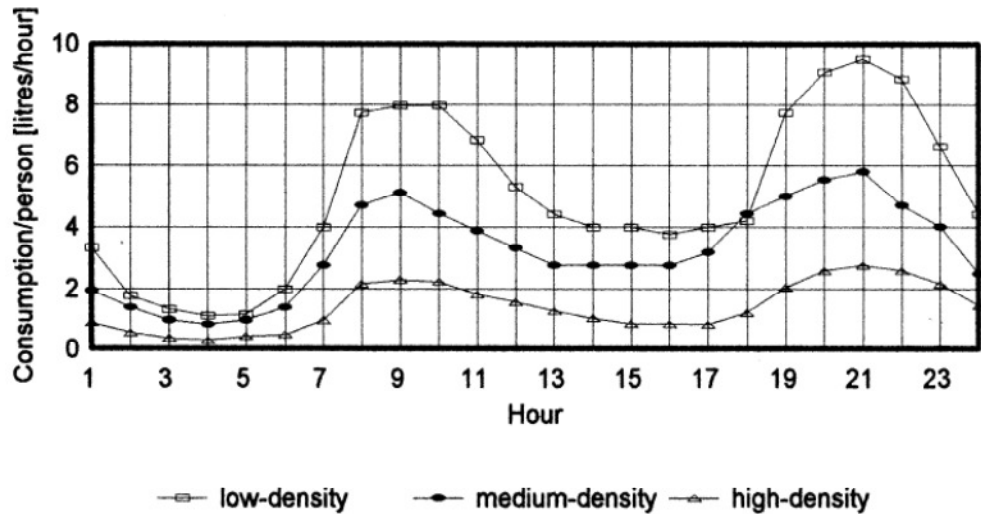
Figure 2.5.5: 60 minute hot water consumption data, per person per day during weekend days in winter [31].

DeOreo et al. concluded from flow trace analysis that was conducted, that on average the shower accounted for 25.1 percent of hot water used for low-density housing. This was only exceeded by the faucet category which was 34.3 percent of hot water consumed [33]. Household hot water

49

consuming facilities are listed in Table 2.5.3 with their percentage contribution to the total hot water demand.

Table 2.5.3: Household hot water used per person per day [33].

Category	Per Capita (gcd)	Household Use (gal per day)	Percent of Total Hot Water Use	Percent of Overall Water Use that is Hot Water
Bath	4.2	10.9	16.7	78.2
Clothes Washer	3.9	10.1	15.5	27.8
Dishwasher	0.9	2.3	3.6	100
Faucet	8.6	22.4	34.3	72.7
Leak	1.2	3.1	4.8	26.8
Shower	6.3	16.4	25.1	73.1
Toilet	0.0	0	0	0.0
Other	0.01	.03	0	35.1
Indoor Total	25.1	65.3	100%	39.6%
Sample size	*10*			
Avg. # of residents	*2.6*			

The data presented in Table 2.5.3 was obtained from the end-use statistics in Table 2.5.4. It is interesting to note that even though the faucet has a relatively short duration of flow, its total volume of hot water demand exceeds that of the shower use. It can be seen that the number of times the faucet was used was considerable at 4329 events. When one takes into account that the hot water faucet is purged of cold water almost each time it is used one might conclude that a substantial amount of hot water energy is lost during the cooling down of the hot water pipe to the faucet. This would be dependent on the hot water pipe volume leading to the faucet. An energy efficient solution to this energy loss may improve the overall hot water system efficiency.

Table 2.5.4: Hot water end use statistics for ten Seattle homes over a 14 day period [33].

Hot Water Use Events		Peak Flow (gpm)		Volume (gal)		Duration (min)		Avg. Flow (gpm)	
	n	Mean	Conf Int (95%)	Mean	Conf Int (95%)	Mean	Conf Int (95%)	Mean	Conf Int (95%)
Baths	101	3.16	0.34	9.48	1.42	4.60	0.73	2.23	0.28
Dishwashers	164	1.67	0.07	1.47	0.11	1.19	0.08	1.23	0.05
Showers	181	2.23	0.12	11.62	1.15	7.40	0.59	1.51	0.06
Faucets	4329	1.24	0.02	0.59	0.02	0.72	0.03	0.91	0.02
Clothes washers	168	2.43	0.14	7.28	0.88	3.17	0.33	2.10	0.13

Figure 2.5.6 shows the 60 minute use of hot water by end-use category as a percentage of the total hot water consumption based on the DeOreo et al. study [33].
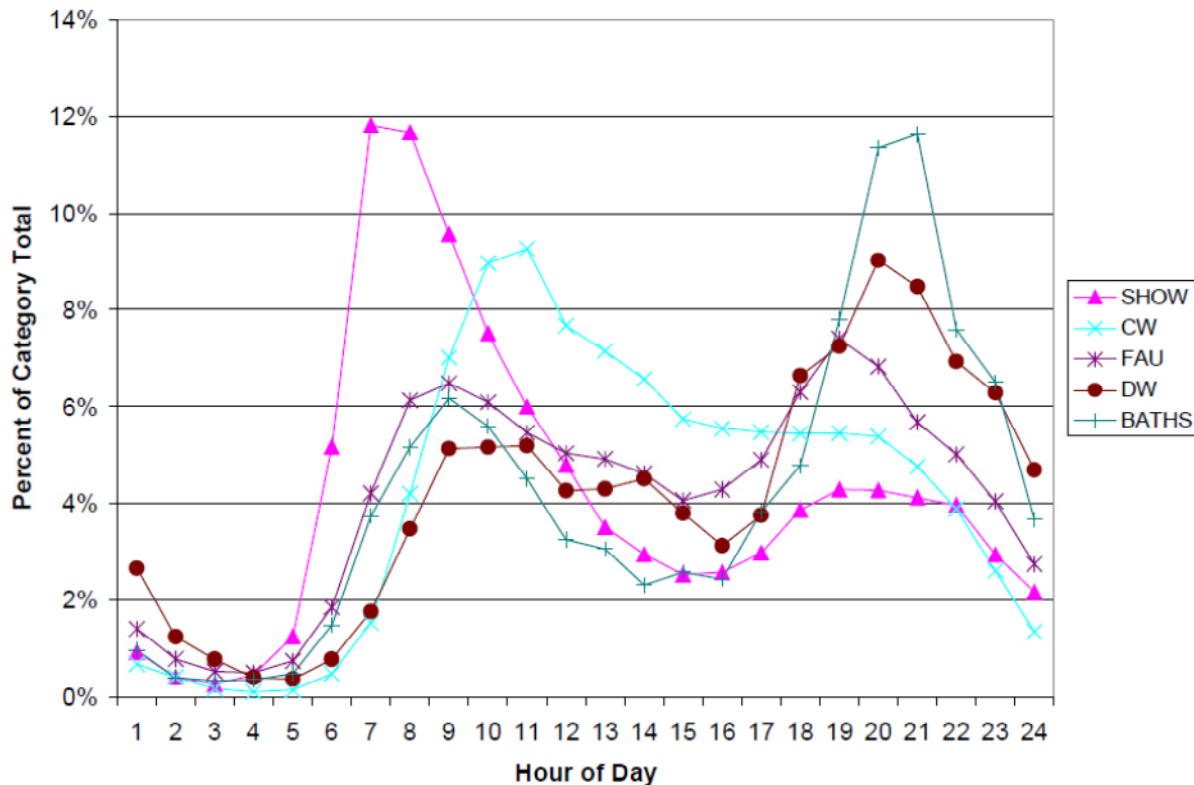
Figure 2.5.6: 60 minute use of hot water by end use category [33].

In a document produced by the Environmental Protection Agency of the United States (EPA), entitled: "Quantification of Exposure-Related Water Uses for Various U.S. Subpopulations", the statistical distributions related to water end-use categories for households were given. The information in this document would prove to be invaluable for the modelling of energy savings initiatives. An example of such an initiative would be the evaluation of energy saving when low flow shower heads are retrofitted. The document indicates that the distribution of shower period duration and flow rate should be regarded as lognormal. The time of use at which a shower period occurs is best represented by a normal distribution [34]. Based on this information, shower periods can be simulated as part of the total hot water consumption for a household. The standard shower's hot water consumption could be modelled and from this the lower flow shower head's hot water consumption could be subtracted. These models could then be

calibrated to the recorded data following standard M&V procedures to produce calculated energy savings.

2.5.3   Types of hot water heaters

In South Africa, the most prolific method to supplying hot water is with a conventional geyser. Through DSM projects, Eskom has promoted alternatives to hot water heating by means of geysers, such as solar water heating and to lesser extent heat pumps. In the USA, many more types of water heaters are used compared to South Africa. Table 2.5.5 provides a summary of the different types of water heaters, their cost to purchase, their cost to run and their life expectancy [35].

Table 2.5.5: U.S. Life cycle costs of various types of water heaters [35].

Water Heater Type	Avg. Cost US$	Expected Life	Annual Energy Cost US$	Cost Over 13 years US$
Gas Conventional Tank	$ 425	13	$165	$2,544
Gas High Efficiency Tank	$ 500	13	$145	$2,385
Gas Demand	$ 650	20	$140	$2,243
Oil Conventional Tank	$1100	8	$230	$4,777
Electric Conventional Tank	$ 425	13	$500	$6,925
Electric High Efficiency Tank	$ 500	13	$480	$6,740
Electric Demand	$ 600	20	$510	$7,020
Heat Pump	$1200	20	$190	$3,670
Indirect with Boiler	$ 700	30	$150	$2,253
Solar with Electric Back-up	$2500	20	$140	$3,445

It is evident from Table 2.5.5 that the conventional electric geyser is by far the most costly way to supply hot water. The solar water heater with backup electrical element is the most costly to purchase compared to the other hot water heating methods.

2.5.3.1   Conventional geyser

A conventional geyser consists of an insulated hot water cylinder and an electric heater element. As hot water is drawn from the cylinder, cold water enters it. This cold water is at a temperature close to soil temperature at which the water supply pipe is buried. Inlet water temperature will be discussed in the last section of this chapter. The geyser element can range in size from two to

five kW depending on the water capacity of the cylinder. The element is controlled by a hysteresis-type relay. The water in the cylinder is kept between a certain preset temperature range by this relay which switches the element on and off. This range is normally set between 55 and 65 degrees Celsius. A study conducted by Bosman et al. evaluated the electrical saving achieved by installing additional insulation around the cylinder of a geyser. An average annual saving of approximately 18 percent was realised [36]. The standing losses that were quoted in the study were used in the modelling of the geyser in Chapter three.

2.5.3.2 Demand or instantaneous water heaters

Demand or instantaneous water heaters do not heat and store hot water like a geyser does. When the faucet is turned on the electric element heats the water directly, thereby eliminating the standing losses associated with the conventional geyser. The one drawback of these water heaters is that they have a limited flow rate of between 7 to 15 litres per minute [35].

2.5.3.3 Solar water heaters

A solar water heater captures the solar radiation which is converted to thermal energy which in turn heats up the water stored in an insulated cylinder. In the past, solar water heaters have not been a financially feasible option because of their high capital costs, relatively poor performance and low electricity tariffs. This situation has changed. Eskom has a DSM program in place that subsidises a part of the capital cost of the heater. Electricity tariffs have increased substantially over the past couple of years and this has resulted in a substantial reduction in pay-back periods for solar water heaters. The majority of the solar water heaters in South Africa are equipped with an electrical element and function similar to a conventional geyser in the way that the electric element is off when the water is within the preset range. In this way, hot water is always available provided there is a supply of electricity even if there has been poor solar radiation during a day. Current solar thermal systems can utilise 30 to 60 percent of the solar radiation that falls on their collectors. As a result of this, solar water heaters have improved substantially in the past few years and now provide a low maintenance alternative to a conventional geyser. There are three main types of solar water heaters namely: vacuum tube, evacuated tube and flat plate collectors. The tube type is able to passively track the sun whereas the flat plate type is sensitive to sun position [37]. Figure 2.5.7 depicts an evacuated tube type solar water heater.
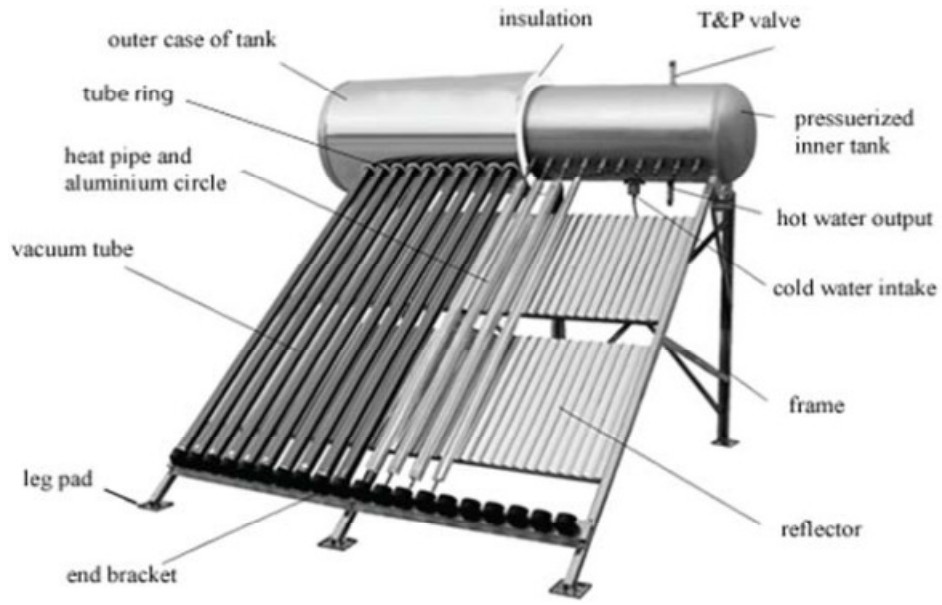
Figure 2.5.7: Evacuated tube collector type of solar water heater [38].

2.5.3.4   Heat pumps

Heat pumps work on the same principle as an air conditioning unit but just in reverse. The heat energy is transferred from a low temperature heat source to a higher temperature heat sink. The heat pump does not generate the heat directly; instead electricity is used to drive the compressor. Figure 2.5.8 shows the flow diagram of an air-type hot water heat pump.
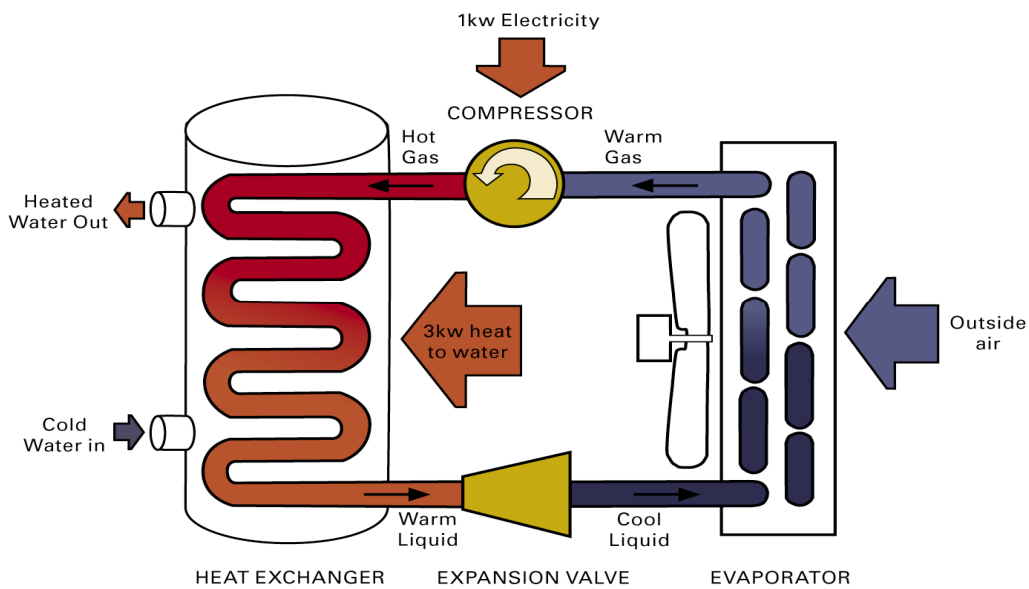
Figure 2.5.8: Flow diagram of an air-type hot water heat pump [39].

54

For an equivalent amount of energy, a heat pump water heater can heat two to three times more water than a conventional geyser [35]. They can also offer the added benefit of space cooling because the heat energy is extracted from a living space and transferred to the hot water cylinder. Many of these types of heaters have a backup electric element to heat the water when the air temperature is too low for efficient operation. The installation costs of these heaters are higher than that of a geyser but the running costs are considerably lower. Geothermal water heat pumps can operate in colder climates than the previously mentioned air heat pump water heaters. They extract heat energy from the ground rather than the air which is colder in winter than the ground. The geothermal heat pumps are more costly to install than the air heat pumps water heaters [35].

2.6 Geyser inlet water temperature

2.6.1 Overview

Inlet water temperature represents a significant source of energy for the energy system of the geyser. Understanding what affects the inlet water temperature and the degree of influence it has on the electrical energy consumption for a geyser will possibly assist in the improved design of hot water heating systems and DSM programs that are designed to manage their energy consumption. Water enters most residential geysers from the utilities' underground network of water supply pipes. It is common knowledge that climatic conditions are one of the controlling variables that affect the temperature of water stored below ground level. It will be assumed that the inlet water temperature will be in equilibrium with the soil temperature at the depth that the water supply pipe is buried below the ground surface. This assumption is based on the fact that the water entering a geyser is relatively small in comparison to the water stored in the underground supply pipe. This means that the temperature of the water in the supply pipe would have enough time to be in equilibrium with the surrounding ground temperature. Based on this assumption the soil temperature at the depth of the burial depth of the water supply pipe must be predicted to be able to predict the geyser inlet water temperature.

2.6.2 Prediction of soil temperature as a function of depth and time of year

Hillel advocates that the simplest mathematical representation of the fluctuation of the thermal profile in the soil at a particular depth is to assume that soil temperature oscillates as a pure sinusoidal function of time around an average value [40]. At each succeeding depth, the peak

minimum and maximum soil temperature values are dampened and shifted progressively in time. The degree of dampening of the soil is dependent on the frequencies of the fluctuations and the thermal properties of the soil [40]. Soil temperature varies in response to exchange processes that take place near the soil surface. The heat transfer through the soil's profile is influenced by the specific heat capacity, thermal conductivity and thermal diffusivity [41]. The thermal conductivity of soil is related to the soil moisture content. Water is a better thermal conductor than air and as a result of this, the thermal conductivity of soil increases as its moisture content increases [42].

The sinusoidal function that describes the annual variation of daily soil temperature at different depths is as follows [40]:

$$T(z,t) = T_a + A_0 e^{-\frac{z}{d}} \sin\left[\frac{2\pi(t-t_0)}{365} - \frac{z}{d} - \frac{\pi}{2} \right] \qquad \textbf{2.6.1}$$

where

$T(z,t)$ denotes the soil temperature at time t (day of the year) and depth z in meters;

T_a denotes the average soil temperature in degrees Celsius;

A_0 denotes the annual amplitude of the soil temperature in degrees Celsius;

d denotes the damping depth in meters of annual fluctuation and

t_0 denotes the time lag in days from an arbitrary starting date (1 January) to the occurrence of the years minimum temperature.

The dampening depth is calculated by using the following equation [40]:

$$d = \left(\frac{2D_h}{\omega} \right)^{\frac{1}{2}} \qquad \textbf{2.6.2}$$

where

D_h denotes the thermal diffusivity in m^2 per second and

56

$\omega = \dfrac{2\pi}{365}$ denotes the angular frequency of the temperature oscillation in radians per day.

The sinusoidal temperature equation was derived by solving the following partial differential equation [40]:

$$\frac{\partial T(z,t)}{\partial t} = D_h \frac{\partial^2 T(z,t)}{\partial z^2}$$ 2.6.3

2.6.3   Predicted soil temperatures versus recorded data

Nofziger used equation 2.6.1 to predict the soil temperature at different depths, this was then compared to measured data [43]. The results are depicted in Figure 2.6.1.
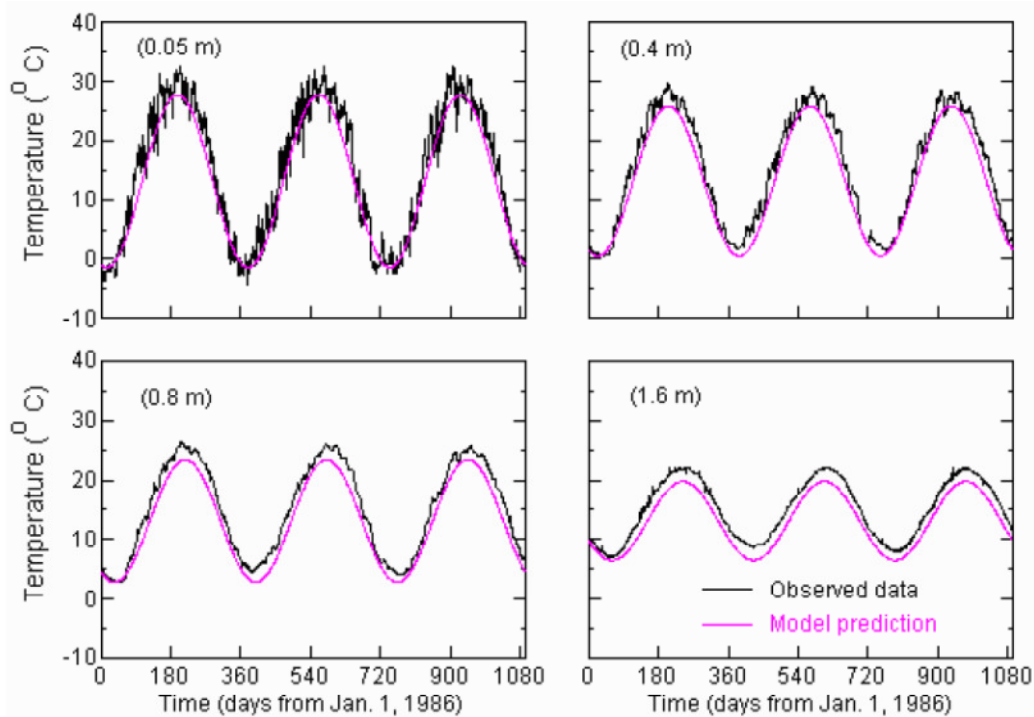
Figure 2.6.1: Comparison of measured and predicted soil temperatures at different soil depths based on ambient temperature [43].

From Figure 2.6.1 it can be seen that at a depth of 0.05 meters there is a higher frequency temperature variation imposed on a larger lower frequency sinusoid. The higher frequency variation becomes less significant with an increase in soil depth. It would seem that the soil acts as a high pass filter for increasing soil depths. The low frequency sinusoid represents the temperature variation of the soil as a result of the seasonal temperature fluctuation. The recorded

57

higher frequency fluctuation in the shallower soil depths is as a result of the day and night temperature variations. Ambient temperature was used to predict the soil temperatures. The predicted temperature values were consistently lower by two degrees Celsius than the observed values [43]. It is evident from the above figures that for an increase in soil depth there is a decrease in amplitude variation and an increase in phase shift of the soil temperature.

2.7   Conclusion

This final section concludes the literature review. The review was comprehensive and endeavoured to cover all aspects of relevance to the thesis topic. The review outlined a logical progression of topics which each proved to be pertinent in developing a thorough research methodology.

Chapter 3   **Load modelling of a residential geyser**

3.1 Overview

It was mentioned in Chapter one that the largest consumer of electricity in the residential sector of South Africa, is hot water heating. DSM initiatives have been promoting ECMs that reduce electrical demand of geysers or their replacement with heat pumps or solar water heaters. It is important for DSM initiatives to be able to predict and verify the change in electrical demand when replacing one technology with another. The impact would naturally be evaluated using the M&V process which advocates the need to calculate the confidence factor associated with the quoted magnitude of energy saving but the IPMVP does not describe a preferred method on how this should be achieved.

For these reasons it was decided to develop a model of a hot geyser in MATLAB's SIMULINK, which would form part of the overall analysis in which electrical savings and their associated confidence levels could be calculated. This model topology created a sterile simulation environment where the effectiveness of the proposed methodology could be evaluated. The geyser model is considered a complex electrical load for which the independent input variables are ambient temperature, hot water consumption and inlet water temperature. Electrical load demand is the dependent output variable calculated by the model. The electrical demand and ambient temperature are the input variables to the kernel density function. The output of this function is the bivariate probability density which indicates the statistical dependency between the independent variable, ambient temperature, and the dependent variable, electrical demand. The non-parametric method of the kernel density estimation as described by Charytoniuk et al. was used to generate the PDFs for the statistical interpretation of electrical demand versus ambient temperature [27]. The statistical nature of the input variables was also investigated and the results are presented in Chapter four. The flow diagram for simulating and analysing the demand of the geyser is depicted in Figure 3.1.1. In this figure it can be seen that the model of the geyser receives ambient temperature, hot water demand and inlet water temperature as input variables.

```
                    ┌──────────────────────┐
                    │ Inlet water temperature│
                    │        data          │
                    └──────────────────────┘
                              │
                              ▼
┌──────────────────┐   ┌──────────────┐   ┌──────────────────────┐
│ Ambient temperature│──▶│ Geyser model │◀──│ Hot water consumption│
│       data        │   └──────────────┘   │        data          │
└──────────────────┘          │            └──────────────────────┘
        │                     ▼
        │             ┌──────────────────┐
        │             │ Simulated electrical│
        │             │   demand data    │
        │             └──────────────────┘
        │                     │
        ▼                     ▼
              ┌──────────────────┐
      ──────▶ │  Data selection  │
              └──────────────────┘
                     │
                     ▼
             ┌──────────────────┐
             │  Kernel function │
             └──────────────────┘
                     │
                     ▼
          ┌──────────────────────┐
          │ Statistical interpretation│
          │  of simulated demand │
          └──────────────────────┘
```
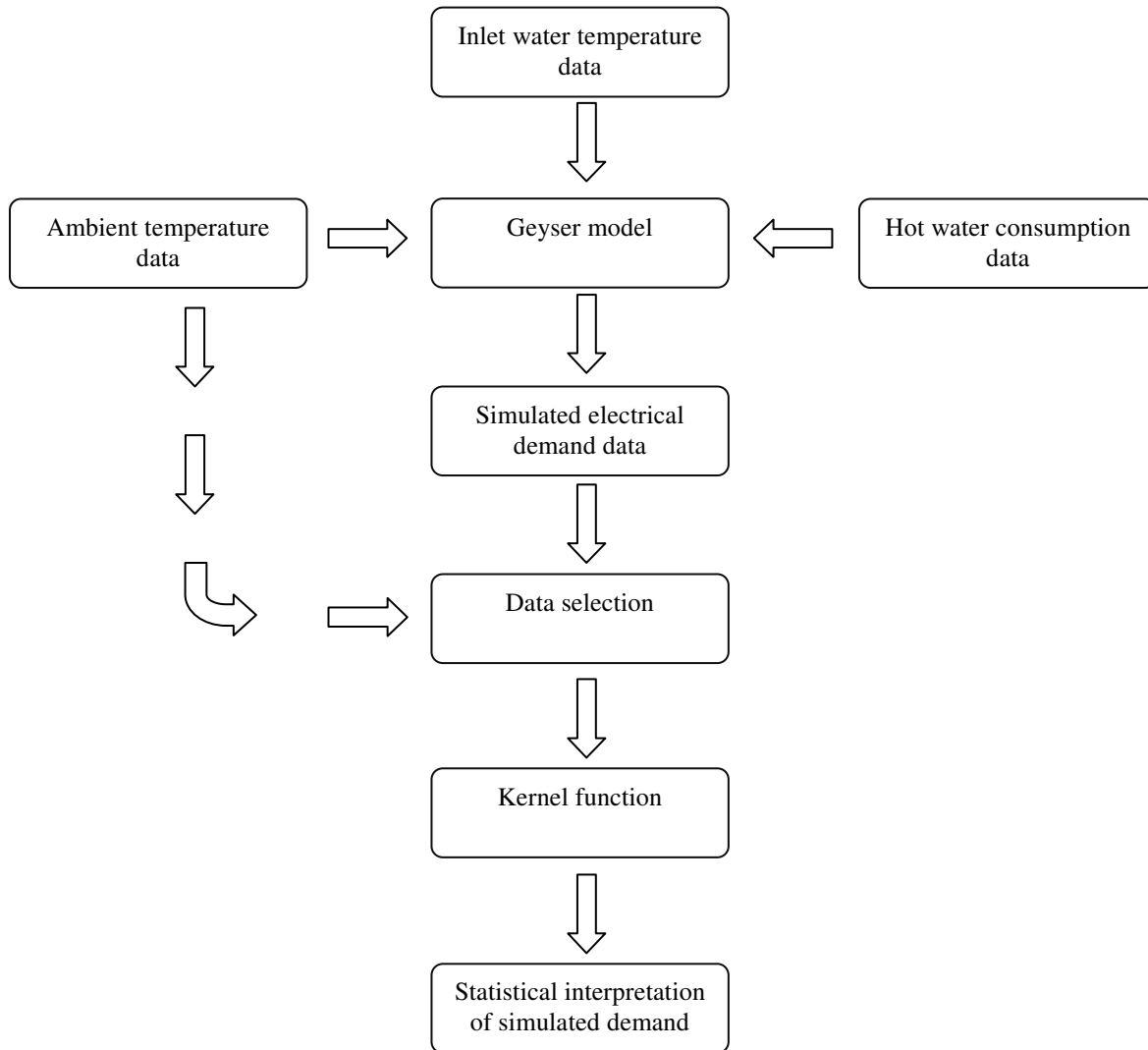
Figure 3.1.1: Flow diagram of the methodology used to model and statistically interpret the simulated electrical demand of the geyser.

The data for these input variables was either sourced or generated based on availability and is described as follows:

- Nine years' worth of ambient temperature was sourced from the South African Weather Bureau.

- The average hot water demand results from the Meyer et al. study were used to generate nine years of hot water consumption data [31]. This was done because no data source for a nine-year period was available.

60

- The inlet water temperature was generated from a function described by Nofziger [43]. This was done because no data source for a nine-year period was available.

It is envisaged that these three variables along with solar radiation will be recorded as a data set for a particular installation and form part of a future project to supply the much needed data for the evaluated hot water ECMs. The geyser model, rather than a solar water heater model, was chosen to evaluate the methodology because it did not need the additional solar radiation variable for which there was also limited data available. The author believes that the true extent of the dependency of electrical demand on ambient temperature will only be realised when a data set becomes available as the one described above.

The geyser model takes the input variables and performs a continuous simulation equivalent to a period of nine years to produce a simulated electrical demand data set. From this data set and that of the ambient temperature, groups of data were selected by a MATLAB function that was written for this purpose. These groups of data were selected on the bases of characteristic season of the year, day of the week, time of day and housing density. These groups of data formed the input data for the MATLAB kernel function which used for the statistical interpretation of the case study material. The probability density estimates produced by the kernel function can be used for two purposes. The first is to predict an electrical load range value with its associated probability at a particular ambient temperature for a particular period of the day. This can be used for short-term load forecasting. The second purpose is to evaluate the anticipated impact of an ECM.

The objective of this chapter is to describe the development of the geyser model and MATLAB code which was used for the statistical analysis of the case study. The MATLAB code used in this project is documented in Appendix B. As part of future development research, the geyser model could be replaced with a model of solar water heater or heat pump.

3.2 Input variables to geyser model

The geyser model has as input variables: ambient temperature in degrees Celsius, inlet water temperature in degrees Celsius and hot water demand in litres per second. All three variables have sampling periods of 1 or 30 minutes. The model is able to run simulations with different input variable sampling rates so that the effect on the simulated output data can be evaluated.

3.2.1   Ambient temperature data

Ambient temperature is the stochastic input variable to the geyser model and the kernel function. Nine years of ambient temperature data for Cape Town International Airport (CTIA), was sourced from the South African Weather Bureau [44]. This data was imported into the MATLAB work space by the *txt2mat* function [45]. The function reads the data in a text format and places it in a matrix. This matrix has temperature values for days of the years in rows and the hours of the day in columns. A function called *inpaintn* was used to search for, and interpolate, missing data values [45]. The data matrix was then reshaped to an array where each data value was assigned a number based on the time it was sampled at. This was done by making use of the MATLAB function *datenum.* The temperature data supplied by the South African Weather Bureau has a sampling period of 60 minutes. The data array was then linearly interpolated respectively to 1 and 30 minute sampling periods by the *resample* function. The *hhGeyserMod* function converted the temperature data arrays to a time series arrays which was then ready for input into the geyser model.

Figure 3.2.1 depicts the CTIA data during 2001. In this figure the diurnal (daily) and seasonal temperature range and fluctuations can be seen.
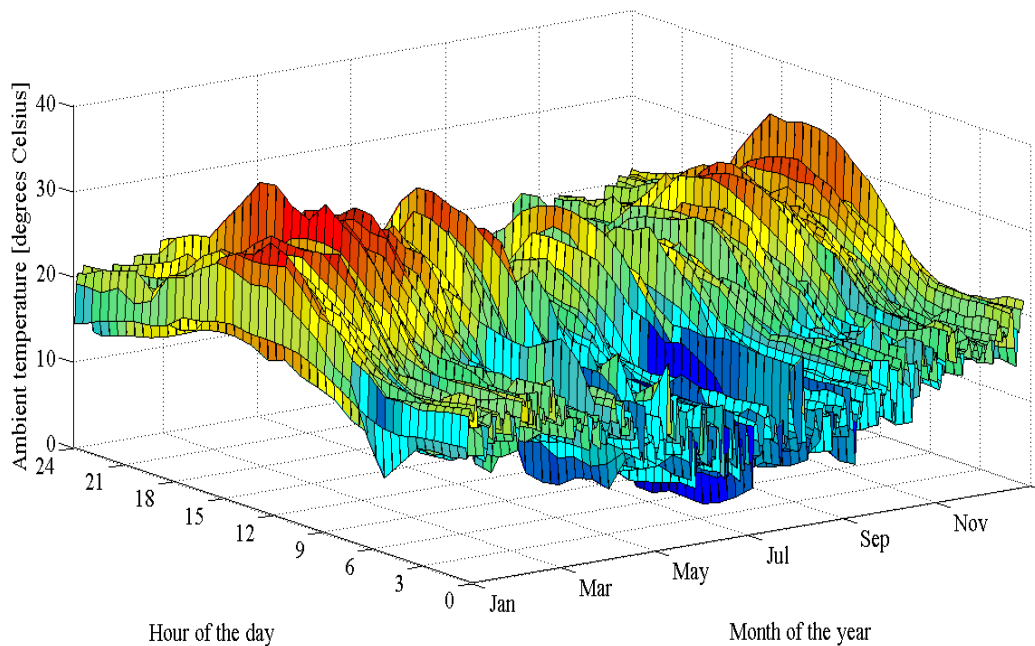
Figure 3.2.1: Ambient temperature data for Cape Town International Airport during 2001.

3.2.2   Hot water consumption data

Hot water consumption data was generated based on the average hot water consumption profiles that were recorded by the Meyer et al. study. This was done because there was a lack of readily available domestic hot water consumption data. The function *HotWgenData* was written to generate a hot water consumption data set. This was done by generating a set of random numbers for the summer and winter average recorded values. These two sets of random numbers had Gaussian distributions. These summer and winter random number sets' standard deviation was set to 17 and 30 percent respectively, which was consistent with the standard deviations observed in the recorded 60 minute data of Meyer et al study [31]. The random numbers were then added to the average 60 minute values of the consumption profiles of the Meyer et al. study. The function then allocated a corresponding date-time number to each of the hot water consumption rate values and then linearly interpolated the data from 60 minute respectively to 1 minute and 30 minute periods. The *hhGeyserMod* function converted the hot water consumption data arrays to time series arrays which were then ready for input into the geyser model. This was done for each of the three housing density hot water consumption profiles. The generated hot water consumption for a period of one year for different housing densities is depicted in Figures 3.2.2 to 3.2.4. The difference in seasonal and daily profiles of hot water consumption can be seen.
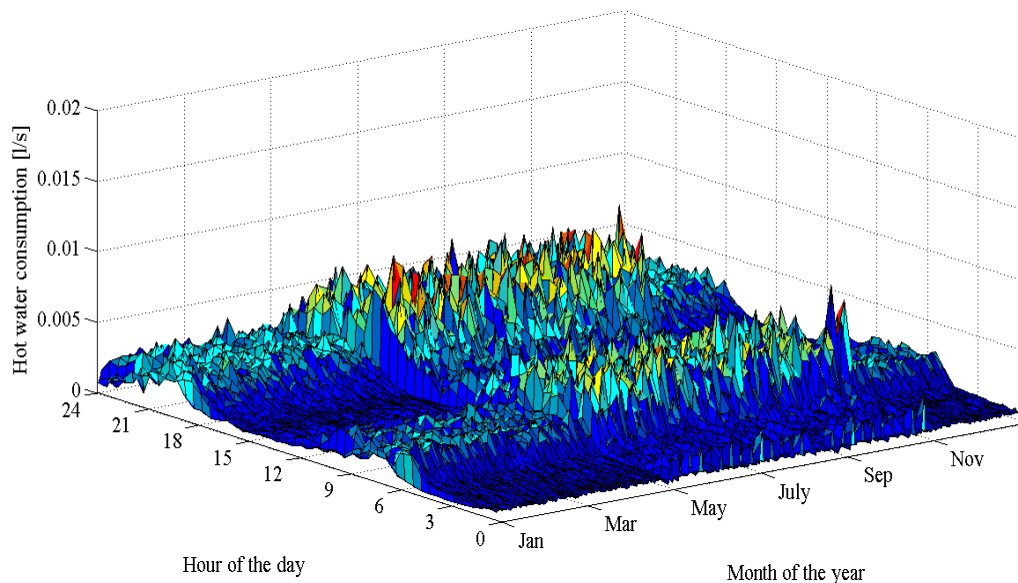
Figure 3.2.2: Hot water consumption generated for high-density housing for a period of one year.

63
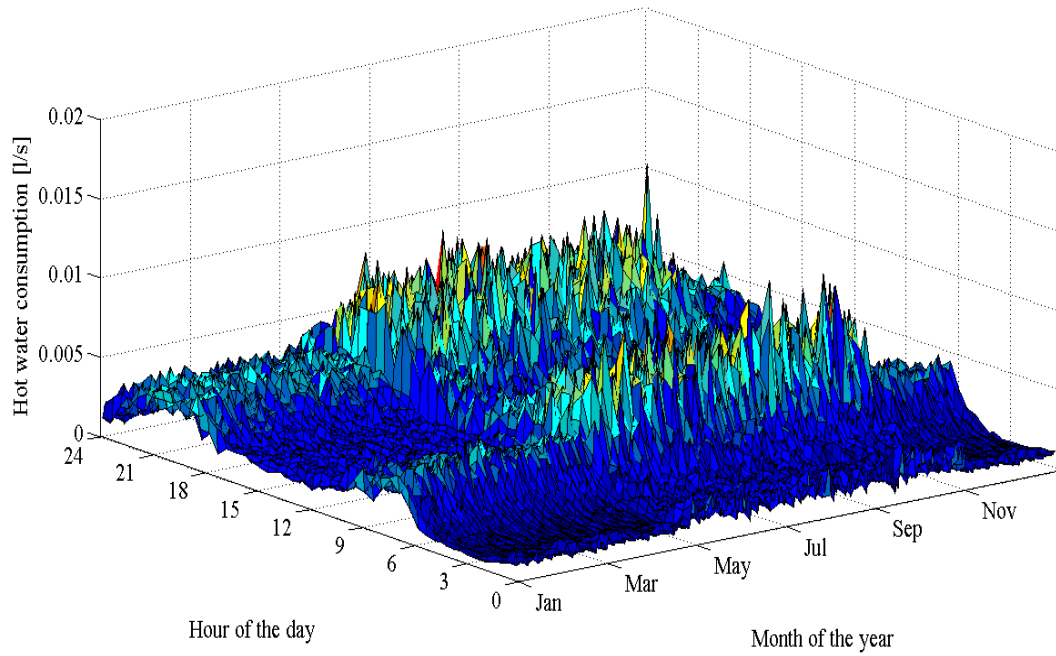
Figure 3.2.3: Hot water consumption generated for medium-density housing for a period of one year.
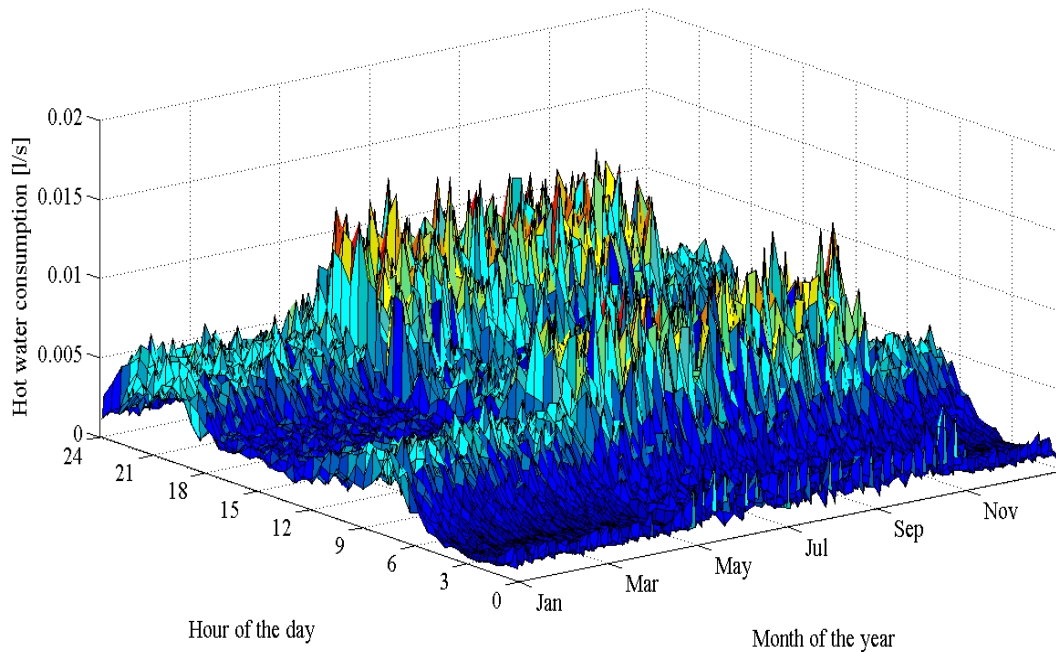
Figure 3.2.4: Hot water consumption generated for low-density housing for the period of one year.

From the hot water consumption figures the following observations can be made: there is a definite change in consumption profile during the seasonal transition. This is as a result of the averaging of the data in the Meyer et al. study. In reality this hot water consumption profile will have a smoother seasonal data transition. It was assumed the hot water consumption data had a Gaussian distribution, which may prove to be different once such data is recorded and analysed. However, it is none the less possible to investigate the effect different hot water consumption profiles has on electrical demand by specifying different parameters to statistical distributions using the *HotWgenData* function. In future work, this function could easily be upgraded so that monthly statistical parameters and distribution could be specified which would then smooth out the seasonal consumption profile. The generated hot water consumption rate increased from high-density housing to low-density housing, which is consistent with the Meyer et al. findings.

3.2.3   Geyser inlet water temperature data

The inlet water temperature data was generated based on the equation described by Hille [40]. This was done because of the lack of readily available geyser inlet water temperature data. The function *InWaterTemp* that was written used the equation described by Hille to generate the inlet water temperature for a specified simulation period. The variables that needed to be specified for this function were:

- depth at which the water supply pipe is buried;
- dampening depth of the soil;
- time period for which the data must be generated, and
- ambient temperature data set for the specified time period.

The *resample* function linearly interpolated the generated data respectively to 30 minute and 1 minute sampling periods. The *hhGeyserMod* function then converted the inlet water data arrays to time series arrays which were then ready for input into the geyser model. The generated inlet water temperature data, for a water supply depth of respectively one and seven meters, is shown in Figure 3.2.5. The soil dampen depth was 2.093 meters for these sets of generated data. It can be seen in Figure 3.2.5 that for an increase supply water pipe burial depth the amplitude of the seasonal inlet water temperature fluctuations decreases. The time phase shift also increases for an increase in burial depth.
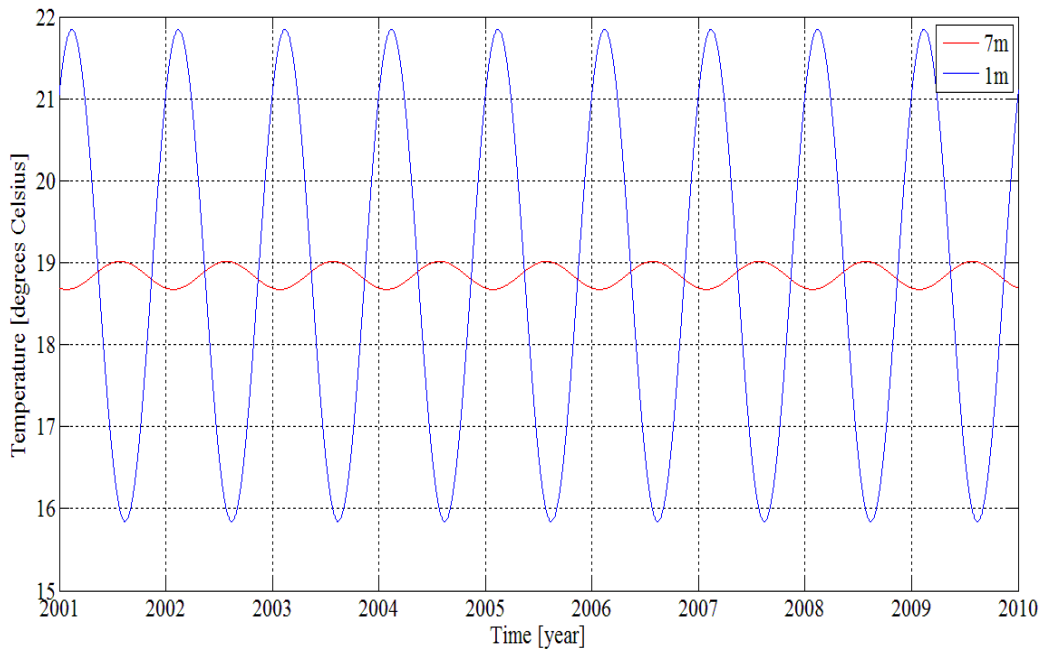
Figure 3.2.5: Generated geyser inlet water temperature data for a water supply pipe buried at a depth of one and seven meters for the years 2001 to 2009.

3.3   Geyser model

The geyser model was developed in SIMULINK and is based on the law of energy conservation which states that the total amount of energy in an isolated system remains constant. This is represented by the following equation:

$$E_{in} - E_{out} = 0 \qquad\qquad\qquad\qquad \textbf{3.3.1}$$

where

E_{in} denotes the energy entering the geyser in joules and

E_{out} denotes the energy leaving the geyser in joules.

This equation can be expanded to yield the following:

$$q_{elec} + q_{cold} - q_{hot} - q_{loss} = 0 \qquad\qquad\qquad\qquad \textbf{3.3.2}$$

66

where

q_{elec} denotes the rate of energy entering the geyser as a result of the electric heater element in joules;

q_{cold} denotes the cold water energy rate entering the geyser in joules;

q_{hot} denotes the hot water energy rate leaving the geyser as a result of hot water demand in joules and

q_{loss} denotes the rate of energy decrease as a result of standing losses in joules.

This energy balance is depicted in Figure 3.3.1.
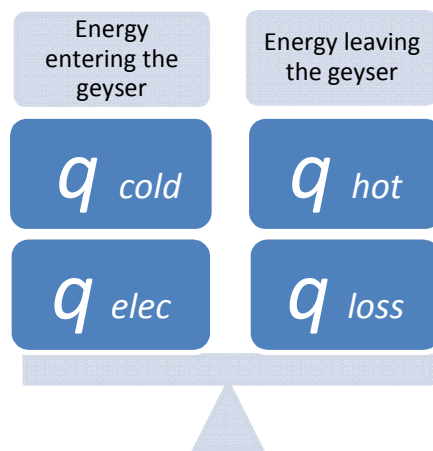
Figure 3.3.1: Heat energy transfer balance for the geyser model.

The simulation period of the model is one second and the model generates 30 minute electrical demand data. For different simulations, constants such as thermostat set point values, volume of geyser standing losses and power rating of the heater element can be adjusted. These adjustments allow for the simulation of a great variety of different geyser scenarios.

The heater element in the model is controlled by a hysteresis control type relay. This in turn controls a thermostat that is set between 55 and 65 degrees Celsius. The energy introduced by the heater element causes the temperature of the water to rise, until the upper temperature set point of the thermostat is reached. The thermostat then drops the signal to the relay at which point the heater element is switched off. Energy is lost from the geyser continuously through

standing losses. The rate of the energy loss is dependent on the temperature difference between ambient temperature and the temperature of the water inside the geyser. The energy inside the geyser is also reduced through hot water demand. The hot water removed from the geyser is replaced by an equal amount of cold water. In this way the temperature of the water in the geyser declines until it reaches the lower temperature setting on the thermostat at which point a signal is sent to the relay which in turn switches *on* the heater element. As a result, the temperature of the water begins to rise again. The energy in the geyser therefore fluctuates as the temperature of the water in the geyser fluctuates between its high and low set points. The total rate of change of energy inside the geyser, q_t, is calculated with the following equation:

$$q_t = q_{elec} + q_{cold} - q_{hot} - q_{loss} \qquad \qquad \textbf{3.3.3}$$

where q_t is the rate of total energy change in the geyser. Integrating q_t yields the change of energy inside the geyser. The change in energy in the geyser is therefore given as follows:

$$\Delta E_t = \int_{t_0}^{t} q_t dt \qquad \qquad \textbf{3.3.4}$$

where

ΔE_t \qquad denotes the change in energy in the geyser over a time period t in joules;

t_0 \qquad denotes the initial time from which the change is measured and

t \qquad denotes the time measured in seconds.

The instantaneous energy value is then divided by the volume of water in the geyser and the specific heat capacity of water which gives the instantaneous temperature. The temperature inside the geyser is calculated with the following equation:

$$T_{geyser} = \frac{\left(E_{initial} + \Delta E_t \right)}{C \times V_{geyser}} \qquad \qquad \textbf{3.3.5}$$

where

T_{geyser} \qquad denotes the temperature inside the geyser in degrees Kelvin;

C denotes the specific heat capacity of water in joules per kilogram per degrees Kelvin;

V_{geyser} denotes the volume of the geyser in litres;

$E_{initial}$ denotes the initial energy inside the geyser in joules and

ΔE_t denotes change in energy inside the geyser over a time period t, in joules.

The four energy transfers will now be described in the next subsections.

3.3.1 Electrical energy

The heater element in the model is set at 3 kW. The energy rate introduced into the geyser by the heater element, q_{elec}, is given by the following equation:

$$q_{elec} = 3kW = 3 \times 1000 = 3kJ / s \qquad\qquad 3.3.6$$

The model topology of the geyser element is depicted in Figure 3.3.2
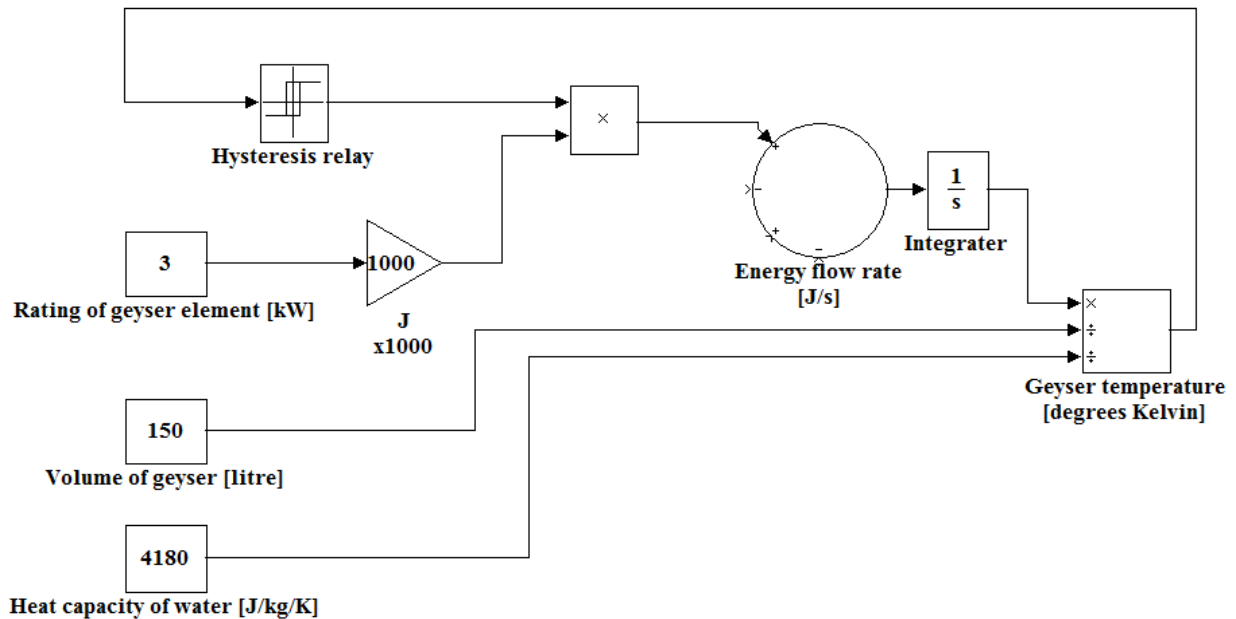
Figure 3.3.2: Model topology of the geyser element and its control.

3.3.2 Cold water energy

The energy flow rate of the cold water that replaced the hot water when the hot water is drawn out of the geyser is given by the following equation:

$$q_{cold} = C \times (T_{cold} + 273) \times V_{hot}$$ 3.3.7

where

q_{cold} denotes the energy flow rate of the cold water;

C denotes the specific heat capacity of water in joules per kilogram per degrees Kelvin;

T_{cold} denotes temperature of the cold water entering the geyser in degrees Celsius and

V_{hot} denotes the volume of hot water leaving the geyser in litres per second.

The model topology for calculating the energy introduced into the geyser when the cold water enters is depicted in Figure 3.3.3.
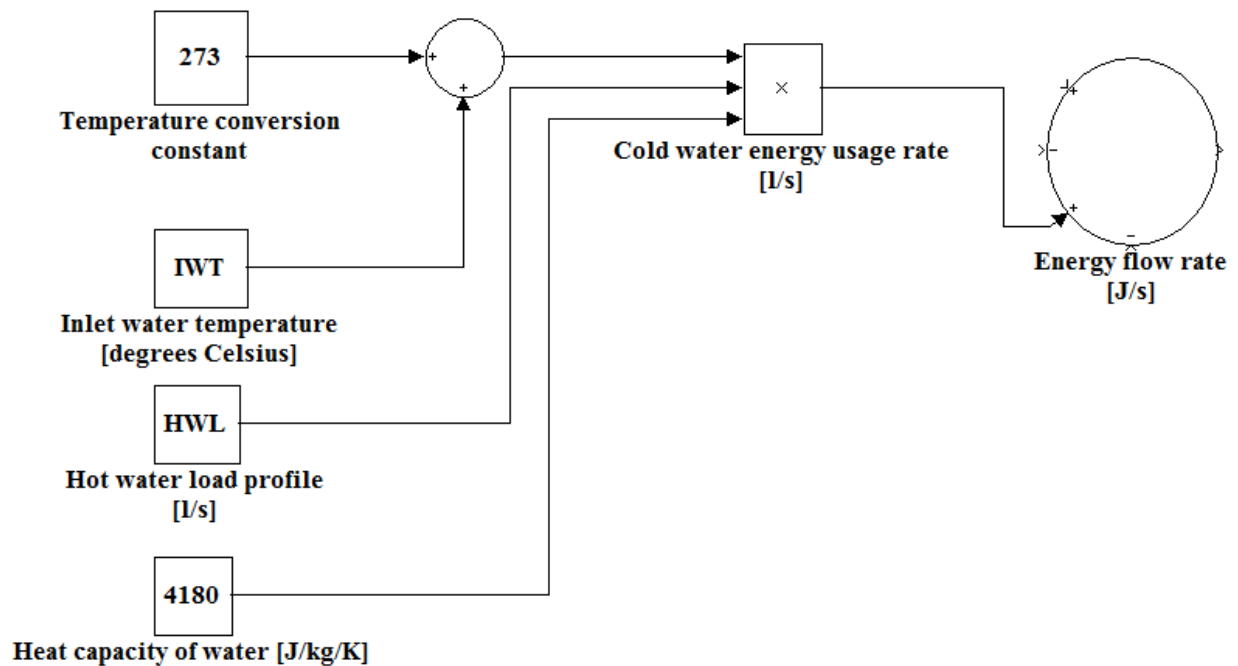
Figure 3.3.3: Model topology for the calculation of cold water energy flow rate.

70

3.3.3   Hot water energy

The energy flow rate of the hot water demand, q_{hot}, is given by the following equation:

$$q_{hot} = C \times T_{hot} \times V_{hot}$$
<div align="right">3.3.8</div>

where

C denotes the specific heat capacity of water in joules per kilogram per degrees Kelvin;

T_{hot} denotes the temperature of the water inside the geyser in degrees Kelvin and

V_{hot} denotes the flow rate of hot water demand in litres per second.

The model topology for calculating the energy withdrawn from the geyser is depicted in Figure 3.3.4.
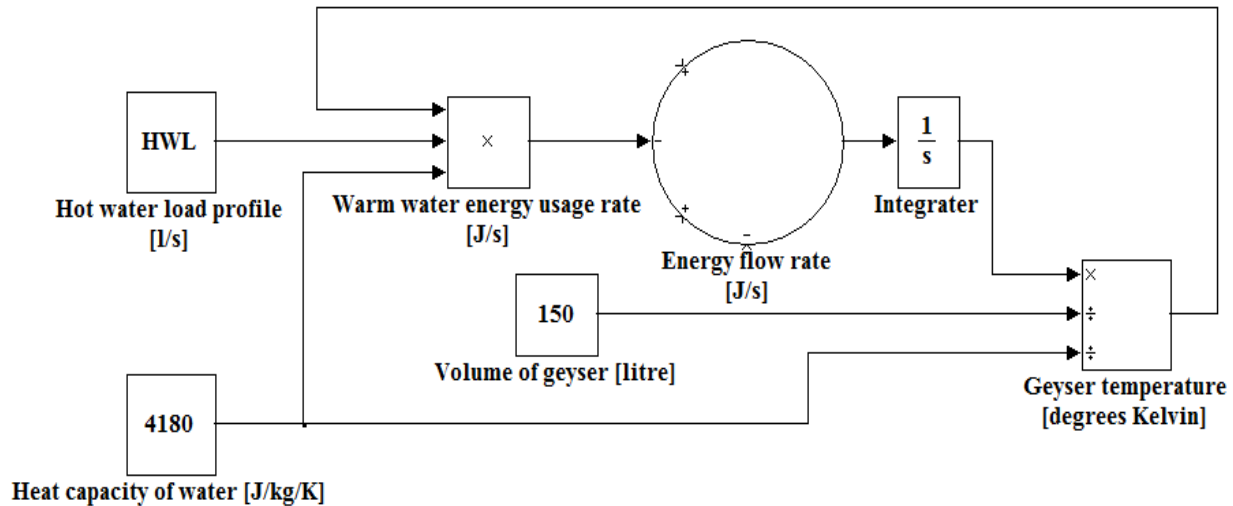
Figure 3.3.4: Model topology for calculating the energy flow rate of the hot water withdrawn from the geyser.

3.3.4   Standing losses

The standing losses, q_{loss}, are accounted for and calculated by the following equation:

$$q_{loss} = \left(\left(T_{hot} - (T_a + 273) \right) \times H \right) \times E_t$$
<div align="right">3.3.9</div>

<div align="right">71</div>

where

T_{hot} denotes the temperature of the water inside the geyser in degrees Kelvin;

T_a denotes the ambient temperature in degrees Celsius;

H denotes the heat loss factor and

E_t denotes the total energy in the geyser in joules.

The model topology for calculating the rate of energy lost from the geyser through standing losses is depicted in Figure 3.3.5.
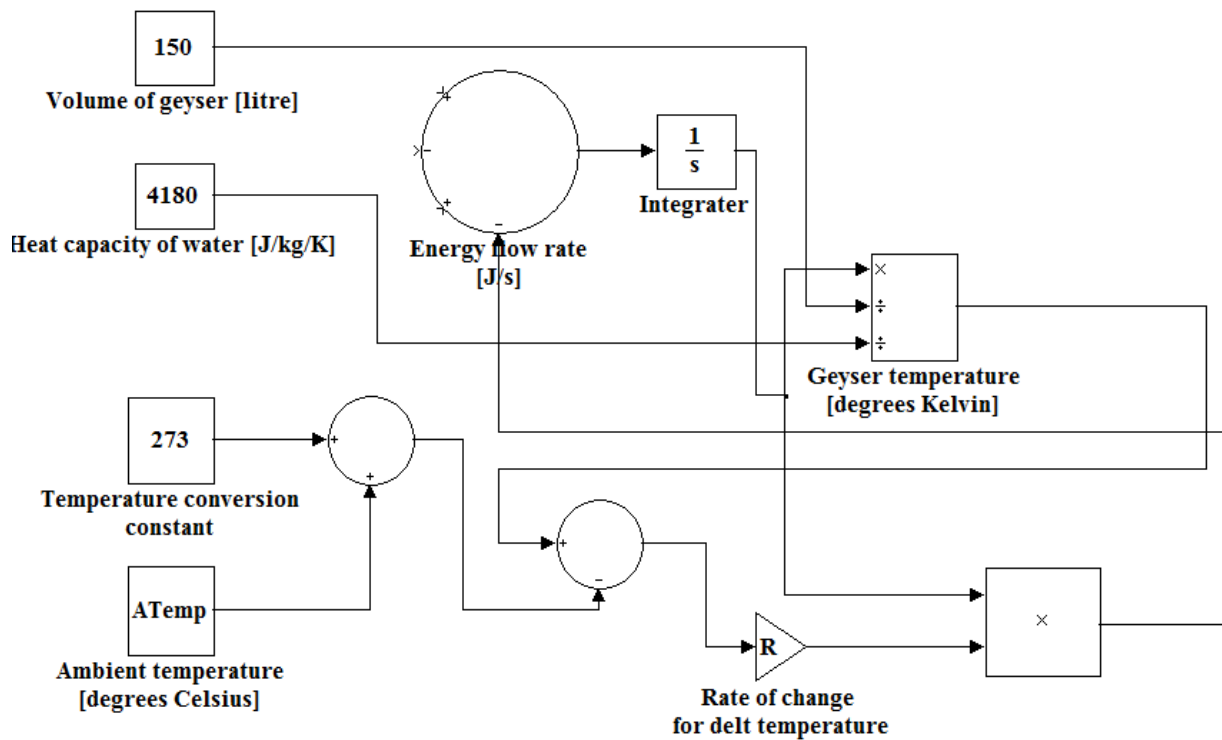
Figure 3.3.5: Model topology for calculating the rate of standing losses.

The values for the standing losses were obtained from the study conducted by Bosman et al. [36]. The losses were measured for geysers with and without geyser blankets. The model's standing losses were adjusted to reflect the findings of the Bosman et al. study.

The entire model topology is depicted in Figure 3.3.6. The *on* time of the relay represents the time the 3kW heater element is demanding electricity. This demand is integrated over each 30 minute period of the simulated days.
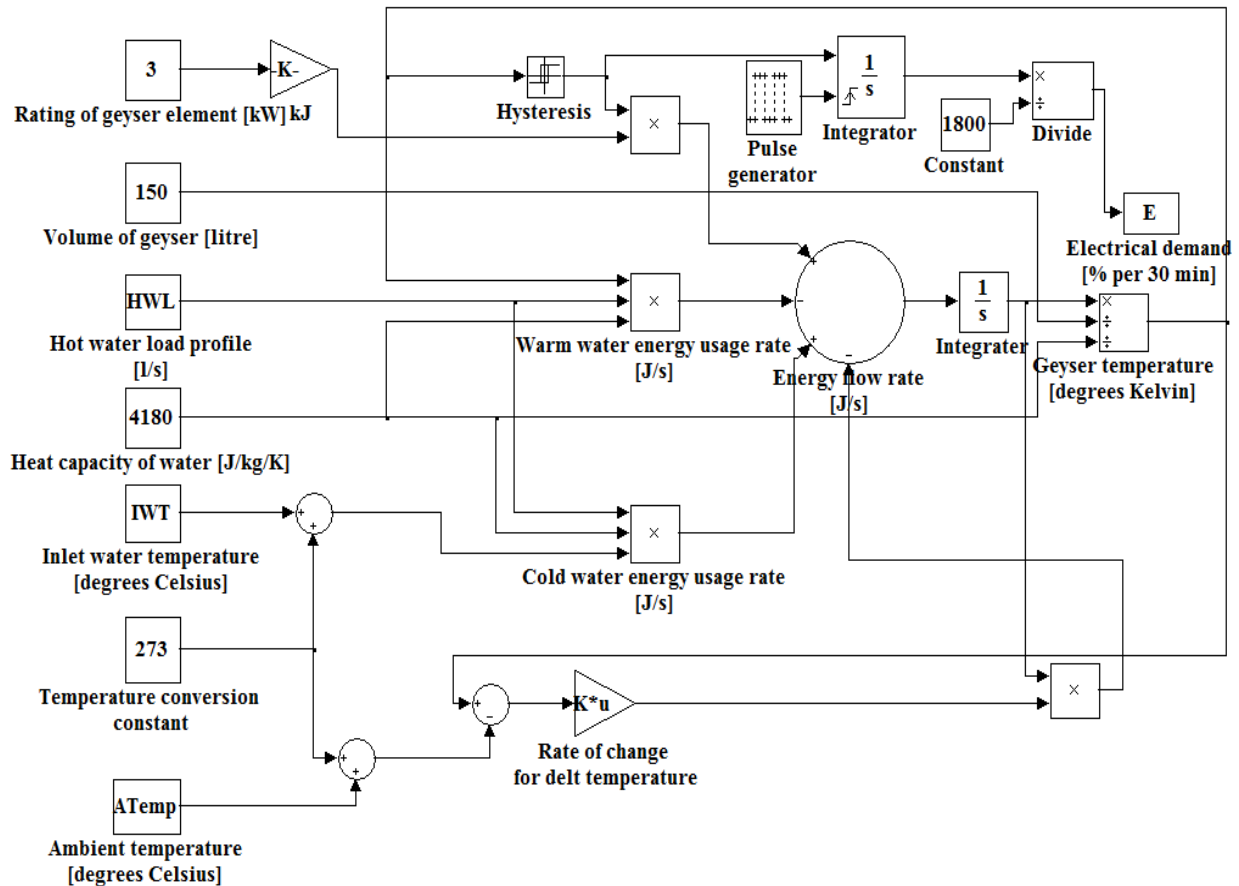
Figure 3.3.6: Model topology of a geyser.

3.4   Generated electrical demand

The output of the geyser model is the percentage of a 30 minute period the geyser was *on*. To best illustrate the energy demand data, a two-dimensional image plot is generated, shown in Figure 3.4.1. The coloured traces show the 30 minute period when the geyser was *on*. The colour bar on the right of the plot indicates the energy demand for the 30 minute period. It can be seen that during the summer months the colour traces are less dense. The two daily demand peak periods are more distinct during the winter months.

73
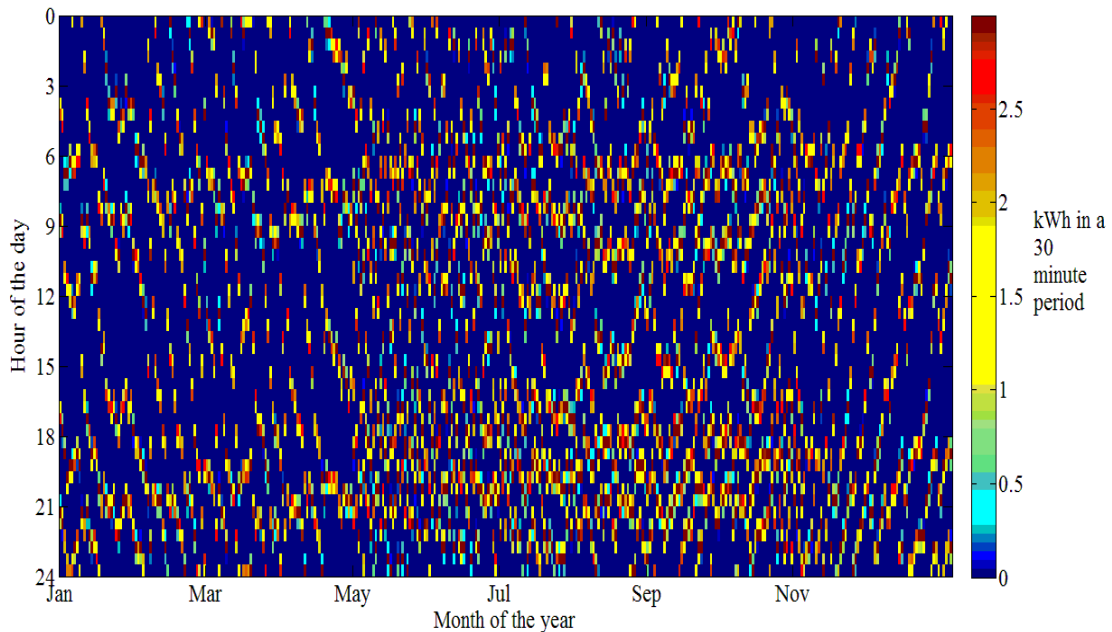
Figure 3.4.1: Image plot of the electrical energy demand simulated by the geyser model for a period of one year.

3.5 Data selection for input variables to kernel function

The demand data must be grouped into its characteristic season of the year, day of the week and 30 minute period of the day as was done in the short-term load forecasting of Charytoniuk et al. [27]. The function *unixSearchNumHH* was written to search for required data. This data included:

- ambient temperature;
- simulated electrical demand, and
- hot water consumption.

The function first generated all the date-time stamps of required category data by using MATLAB's *datenum* function. The *selectedData* function searched for the ambient temperature and electrical demand data that had equivalent date-time stamps. The located data was inserted into arrays, ready for insertion into the Gaussian kernel function. This combination of input variables, with the help of the bivariate kernel function, facilitated the evaluation of the relationship between the independent variable ambient temperature and dependent variable electrical demand by means of probability density functions. It was necessary to group the hot

74

water consumption data sets to facilitate its statistical analysis. The results of this statistical analysis are presented in Chapter four.

3.6   Bivariate Gaussian kernel function

The PDF for a particular 30 minute period, within a 24-hour day is generated with a PDF estimator with a kernel function. It was mentioned in Chapter two that the type of kernel function has much less of an influence on the outcome of the MISE compared to the selection of optimum bandwidth. As a result, the author chose to use a Gaussian kernel described by Charytoniuk et al., which allowed relatively simple bandwidth selection by making use of optimal bandwidth equations [27].

The bivariate Gaussian kernel function is described as follows:

$$f_t(p,T) = \frac{\sum_{i=1}^{n} \exp - \left[\frac{(p - p_i(t))^2}{2h_p^2(t)} + \frac{(T - T_i(t))^2}{2h_T^2(t)} \right]}{2\pi n h_p(t) h_T(t)}$$

3.6.1

where

p denotes the normalized demand;

T denotes the ambient temperature in degrees Celsius;

n denotes the number of samples;

h_p denotes the smoothing parameter for the demand data and

h_T denotes the smoothing parameter for ambient temperature data.

The smoothing parameters h_p and h_T are calculated respectively with the following equations and obtained from the plug-in method:

$$h_p(t) = \sigma_p(t)\left(1 - \rho^2(t)\right)^{\frac{5}{12}} \left(1 + \frac{\rho^2(t)}{2} \right)^{-\frac{1}{6}} n^{-\frac{1}{6}}$$

3.6.2

$$h_T(t) = \sigma_T(t)\left(1 - \rho^2(t)\right)^{\frac{5}{12}}\left(1 + \frac{\rho^2(t)}{2}\right)^{-\frac{1}{6}} n^{-\frac{1}{6}}$$

3.6.3

where

$\sigma_p(t)$ denotes the standard deviation of normalized demand data;

$\sigma_T(t)$ denotes the standard deviation of temperature data and

$\rho(t)$ denotes the correlation coefficient.

These equations minimise the MISE between the estimated density and true density but only if the true density function and the kernel used for estimation have Gaussian distributions. It was decided that this would be the simplest and most appropriate approach to initial density estimation of the geysers simulated demand. The function *KernelFunction* was written to calculate the bivariate probability density for the ambient temperature and demand data. The input parameters for this function were ambient temperature, electrical demand, ambient temperature bandwidth parameter and electrical demand bandwidth parameter. The function generated a matrix of probability density values. This matrix could then be plotted with MATLAB's surface function. Figure 3.6.1 shows a probability density function of the simulated normalized electrical demand of a geyser versus the recorded ambient temperature.

There are different methods of presenting the PFDs. In the Figure 3.6.1, the normalized demand versus ambient temperature is plotted. Demand, rather than normalized demand may also have been used to investigate its interdependency with ambient temperature. For the geyser simulations, it may be more beneficial to know the relationship between the time of day and the ambient temperature when the geyser is switched *on*.

Figures 3.6.2 and 3.6.3 respectively depict a surface and image plot of the probability density of ambient temperature and time of day. The two peak densities and their dependency on temperature can be seen in these figures.
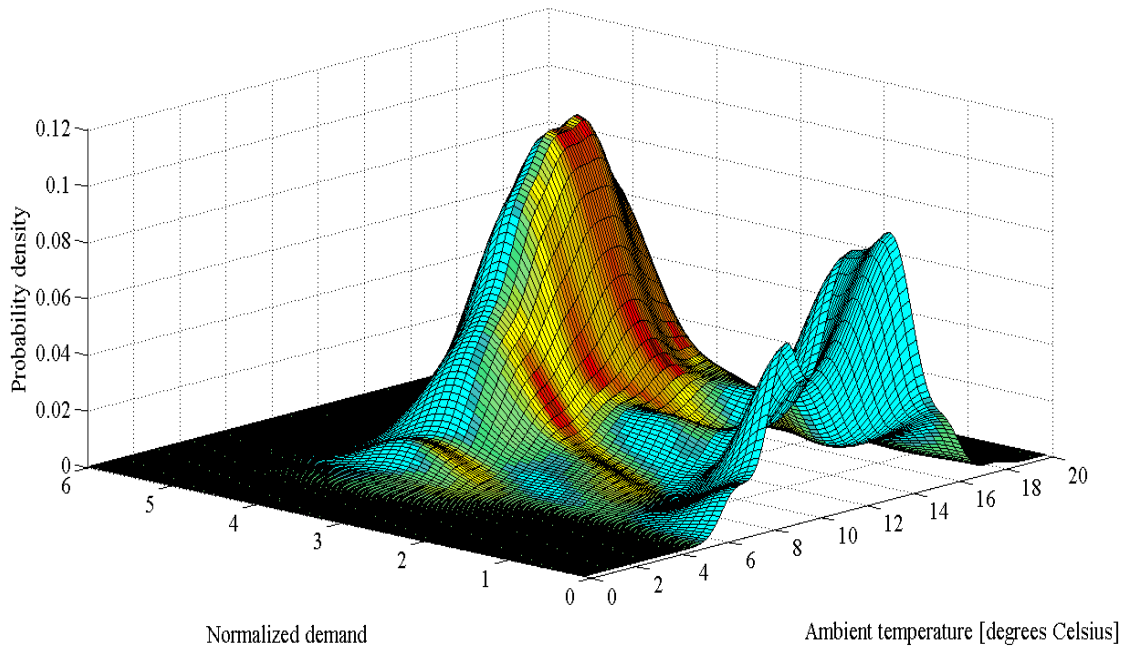
Figure 3.6.1: Three dimensional probability density plot of the simulated normalized electrical demand versus ambient temperature at 11h30 during a weekday in winter from 2001 to 2009.
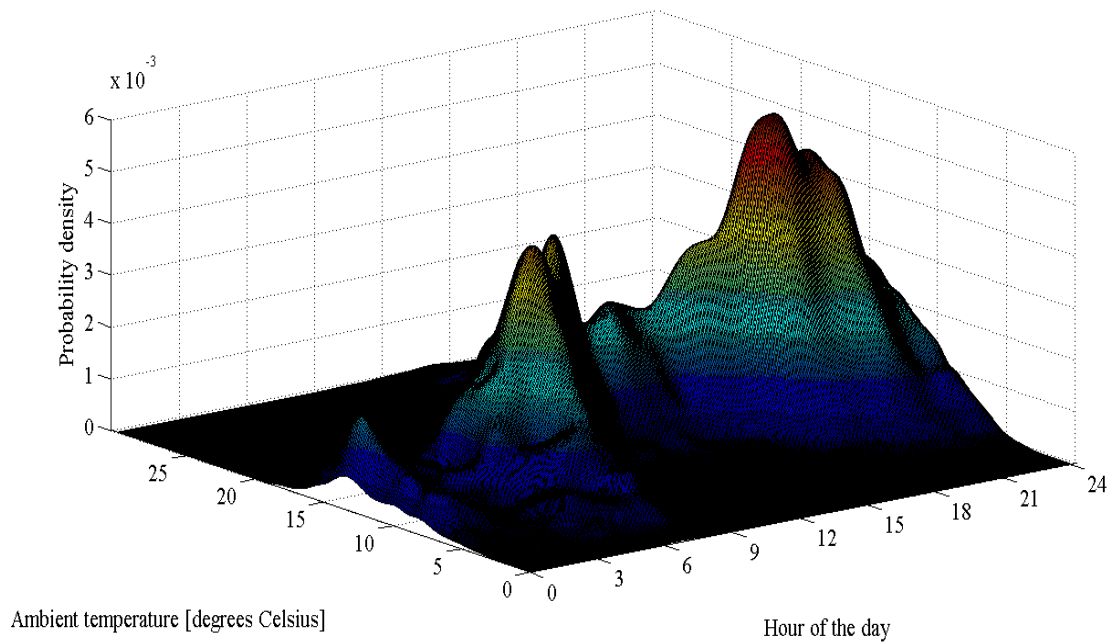
Figure 3.6.2: Three dimensional probability density plot of the ambient temperature versus time of day when the geyser is switched *on*, for weekdays during winter from 2001 to 2009.
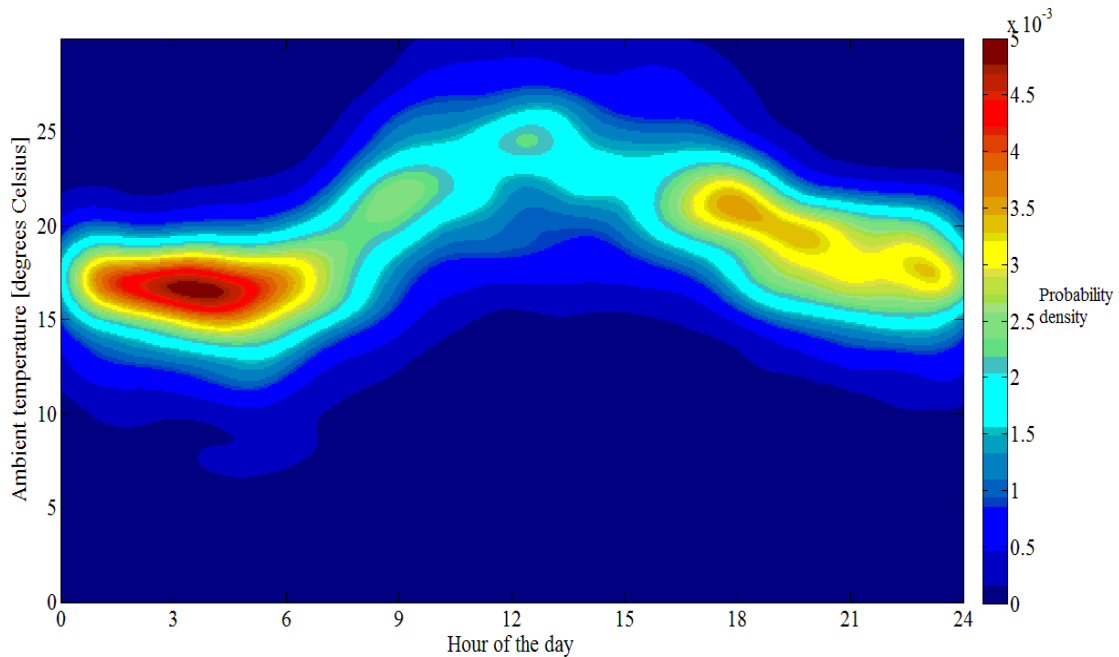
Figure 3.6.3: Two dimensional probability density plot of the ambient temperature versus time of day when the geyser is switched *on*, for weekdays during winter from 2001 to 2009.

3.7   Statistical analysis of a geyser's electrical demand

A description was given in Chapter two of the non-parametric approach to probability density estimation by Charytoniuk et al. [27]. This method was used to forecast short-term electrical demand. This same method was adopted to statistically analyse the demand impacts of the simulated ECMs of chapter five. The first decision to be made when following this approach is to identify which independent variables are statistically influential on electrical demand for the energy consumer. These variables could be any one of the following: ambient temperature, wind speed, light intensity levels and humidity. The chosen independent variables and electrical demand would then be inserted into the kernel function. The independent input variables must be analysed for characteristic patterns. These groupings must be represented in the statistical analysis. Patterns of time of day, day of the week and season of the year were evident in the hot water consumption data variable. These patterns necessitate a similar grouping of PDFs in the statistical analysis to determine the relationship between dependent and independent variables.

Ambient temperature was the independent variable chosen for representation in the statistical analysis because it was evident from the study conducted by Meyer et al. that it has a significant

78

influence on the other input variables to the geyser model which were hot water demand and inlet water temperature. Standing losses are also significantly affected by the difference between ambient temperature and the temperature inside the geyser. From the demand and temperature data, a PDF was generated for a single 30 minute period on a specific day of the week and season of the year. In the statistical analysis the PDFs must describe each one of the chosen categories of characterization.

The PDFs can be used to conduct short-term load forecasting. Alternatively, by changing the geyser model's parameters or adjusting a parameter of one of the generated input variables, it is possible to simulate the energy demand that would result from an implemented ECM.

Finally, a different approach may prove to be feasible in that two separate geyser models could be simulated. One model represents the pre-ECM demand profile and the other the post ECM demand profile. The kernel function of the post ECM could be subtracted from the kernel function of the pre-ECM which in theory results in a kernel function of the demand impact. The subtraction of the two bivariate kernel functions could be done by the multiplication of the Fourier transforms of the pre-ECM kernel function and the negative of the post ECM kernel function. An alternative method to this would be the multiplication of the characteristic functions of the kernel functions. The resultant function of either of these methods could then be inverse Fourier transformed to give the savings kernel function which could be used to generate the demand impact PDFs. This is suggested under future work in Chapter six of this thesis.

3.8   Conclusion

In summary this chapter on load modelling of a geyser has presented the methodology that was adopted to develop a model topology of a complex electrical load and the statistical analysis of its simulated demand. A brief description was given of the MATLAB functions that were developed to prepare the data for insertion into the geyser model and the kernel function. It was concluded that the model topology could be used to provide short-term load forecasting and the prediction of the impact of an ECM. A suggestion is finally given on how the pre- and post-ECM kernel functions could be subtracted from one another to produce an energy savings kernel density estimation function.

Chapter 4   **Statistical analysis of a case study of a geyser model's input data**

4.1   Overview

The aim of this chapter is to gain insight into the statistical nature of the input data of the geyser model. To design an effective and efficient DSM intervention, a comprehensive understanding is required of how the independent variables of the electrical consumer are related to its dependent variables. In a modelled version of an electrical consumer, a prescribed DSM intervention may effectively change the model of the electrical consumer itself or just change some of the parameters of the model. The output data of various simulations of the model is modified through the manipulation of the input data, the parameters of the model and/-or the model itself. A greater understanding of the relationship between input and output data is gained through their statistical analysis. The geyser model effectively manipulates the statistical nature of the data when processed by the model. By knowing the relationship and the statistical attributes of the input/output data it is possible to engineer the desired results through the optimal design of DSM interventions. As a result of the above information it was deemed necessary to statistically analyse the input data to the geyser model.

The input data sets for the geyser model were:

- ambient temperature;
- inlet water temperature, and
- hot water consumption rate.

All of these input data sets must have the same sampling period before they are entered into the geyser model. If the data is obtained at a different sampling rate it can be linearly interpolated or averaged respectively to a higher or lower frequency. The effects of this input data manipulation are investigated on:

- statistical parameters of the input data (in this chapter), and
- simulated output data and its statistical parameters (in Chapter five).

The statistical nature of all the input data sets to the geyser model will vary based on the geographical region of the country. The hot water consumption rate is also influenced by the housing density, which is the associated number of occupants in each household. It was seen in Chapter two that the hot water consumption per person was higher during the colder winter months compared to the warmer summer months. The DSM initiatives may need to be customised for each geographic region to optimise the desired impacts.

The hot water demand was generated from a combination of averages of recorded data and random number generation with statistical distributions. This data needed to be statistically analysed to see the total combined statistical nature of the hot water consumption. The analysis of the input data will commence with ambient temperature data sets, followed by the inlet water temperature data and concluded with the analysis of the hot water consumption data.

4.2   Analysis of input data through parametric and non-parametric methods

In this chapter parametric and non-parametric methods are used to analyse the input data to the geyser model. The methods chosen for the analysis are described in the next two subsections.

4.2.1   Parametric methods

A box plot is a statistical representation of some of the parameters of a data set [46]. This format conveys a large amount of statistical information in a compact easy to read form and allows for the easy comparison of other data set's box plots. With reference to Figure 4.2.1, the line within the box represents the median or the 50th percentile of the data group. The median differs from the mean of the data set in that it is more robust against the influence of outlying data samples. The top and bottom of the box represent the 25th and 75th percentiles respectively. The height of the box gives the inter-percentile range which is an indication of the variance of the data. The position of the median within the box indicates the degree of skewness of the probability density function of the data. The whiskers attached to the box indicate the distance of 1.5 times the inter-percentile range, either side of the 25th and 75th percentiles. The outlying data beyond the range of the whiskers is represented by red plus signs.
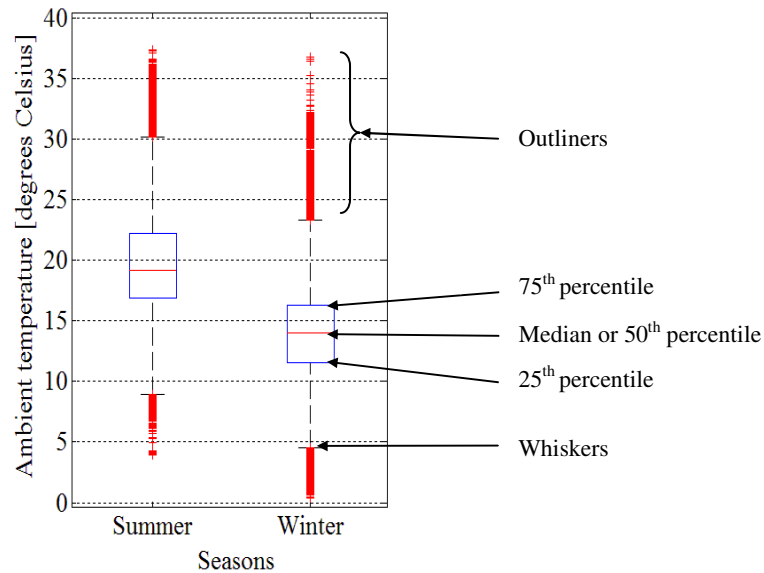
Figure 4.2.1: Summer and winter box plots of ambient temperature.

The reason for analysing the medians is to see if the grouping of multiple years of temperature data would have a significant effect in its original statistical properties. Normally, such an analysis can be done by means of an ANOVA; an acronym for the analysis of variance. However, the latter is misleading because it is in actual fact the analysis of the *means* of groups of data [46]. In an ANOVA three main assumptions are made, namely:

- sample groups of data are normally distributed;
- sample groups have equal variance, and
- all sample observations are mutually independent.

It will be seen during the statistical analysis of the temperature data that follows, that the second assumption is definitely violated. For this reason ANOVA tests were not performed on the temperature data.

4.2.2   Non-parametric methods

Histograms and kernel density estimation are two non-parametric methods used to present the statistical analysis of the input data. The MATLAB *ksdensity* function was used to estimate the densities of the data groupings [46]. This function made use of a normally distributed kernel and the bandwidth was optimised for normal distributions. The assumption of the recorded data and

82

generated data being normally distributed are reasonable, based on the central limit theorem and the fact that the statistical analysis is primarily searching for evidence of multi-modality and the characteristic shape of the density estimates. Non-optimal bandwidth selection will lead to this information being hidden by over- or under smoothing and this will be evident during the analysis of the density estimates.

4.3   Ambient temperature data

4.3.1   Overview

The two sets of ambient temperature data are used in the simulations presented in Chapter five and are described as follows:

- Stellenbosch data set (STL) was recorded from January 2009 to December 2009 and had an original sampling period of 10 minutes;
- Cape Town International Airport (CTIA) data set was recorded from January 2001 to December 2009 and had an original sampling period of 60 minutes.

These two sets of data were chosen because they had different original sampling periods. The STL data set was used to investigate the effects of different data sampling periods on simulation results as well as the effects that result from the processes of linear interpolation and averaging of data. The CTIA data set was used for all the remaining geyser simulations because the data set spanned nine years and thus increased the sample count for all characteristic groupings, especially the weekend grouping.

The first objective of the statistical analysis of the ambient temperature data was to determine what effect linear interpolation and averaging of the original data has on its statistical parameters. The second objective was to determine which of the years of the CTIA data set had the highest and lowest mean and variance in their data. These years were then used to test the sensitivity of the geyser model simulated data to ambient temperature. The results of this temperature sensitivity are presented in Chapter five. The third objective was to investigate the effect of grouping the original data set into subsets on the statistical parameters of the data. These subsets were formed by means of grouping the data according to years, seasons and 30

minute periods. This was necessary because the other input variables such as hot water consumption require characteristic groupings to be able to identify patterns in the data.

A balance needed to be found for the size of the grouped input data. The group size consideration factors were:

- optimal number of PDFs for effective statistical analysis of the demand data;
- optimal sample count within the group for statistical analysis, and
- grouping should show characteristic patterns within all the input data; e.g. hot water consumption patterns during summer weekends are different to winter weekends.

The increase in sample count results in an increase the robustness of the statistical interpretation, but this should not be done at the expense of significantly reducing the ability to show detail of characteristic patterns in the data. The aim of the multi-year box plots of the CTIA data is to identify the years in which the medians and variances were at the outer limits of their range. The simulated electrical demand data of these same years are then compared in Chapter five to ascertain the degree of impact the statistical parametric range of ambient temperature data had on the electrical demand.

4.3.2 Sampling rates of ambient temperature data

To achieve the first objective of the statistical analysis of the ambient temperature, a test environment was created to investigate the effects of linear interpolation and averaging on the input data. This process is depicted in Figure 4.3.1. This test environment was arranged by taking STL data set which had an original sampling period of 10 minutes and linearly interpolating it to 1 minute intervals and averaged to 30 minute and 60 minute sampling periods respectively. Subsequently, the 60 minute data set was then linearly interpolated back to 30 minute and 1 minute period data.

The purpose of generating the 1 minute and 30 minute data from the 10 minute sampling period data, was to investigate the statistical effects when a process of linear interpolation and averaging were applied to the data sampled at a higher frequency. The new data sets in turn generated from the data which had a 60 minute sampling period were used to investigate the statistical effects when a process of linear interpolation was applied to the data sampled at a lower frequency.
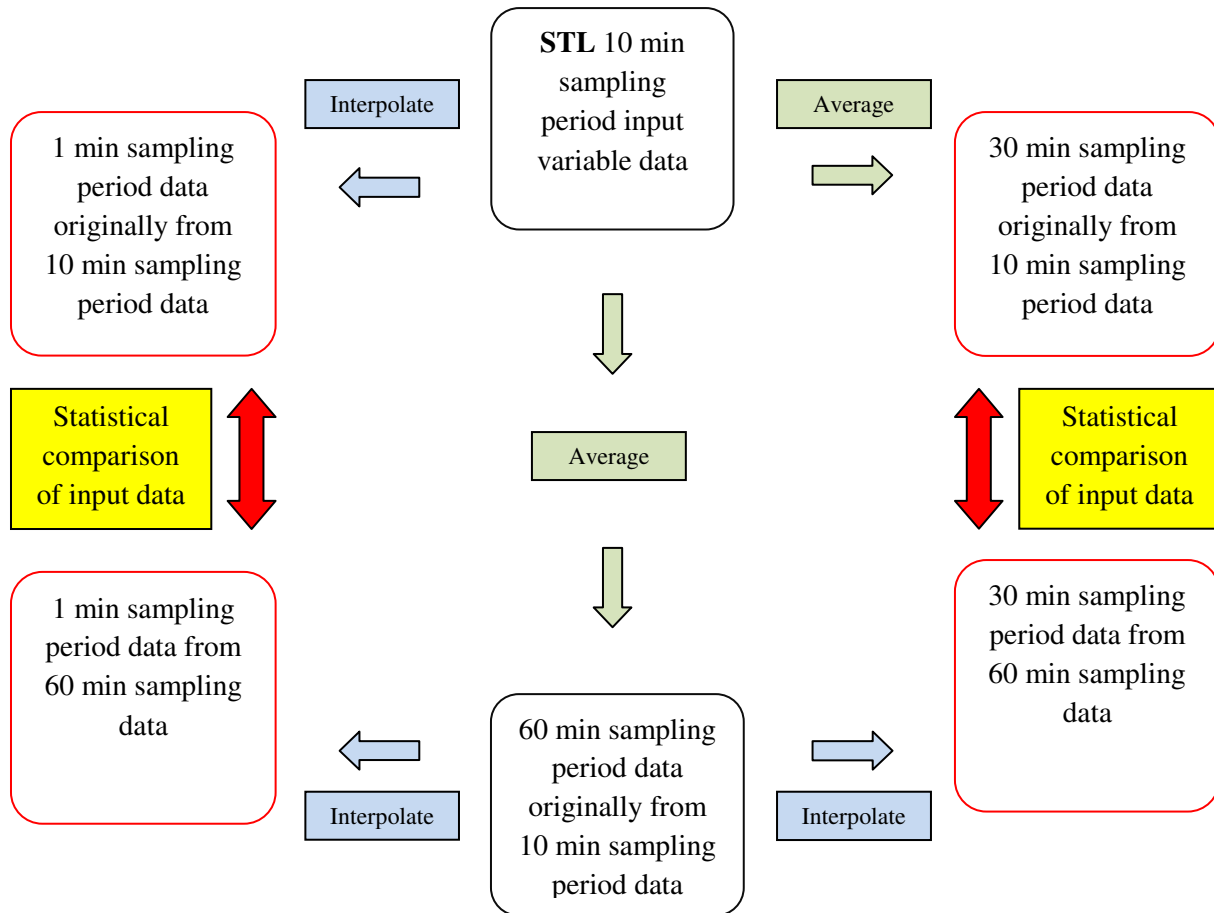
Figure 4.3.1: Data set preparation to investigate the effect of data set sampling period manipulation on statistical parameters of input variable and simulation results.

In summary, it was deemed necessary to investigate the sensitivity of the geyser model to the input data sampling period because this input data may be available in a variety of sampling periods. For instance the ambient temperature data from the South African Weather Bureau was supplied at a sampling period of 60 minutes. Alternatively, in the collection and processing of data, there is added complexity and cost associated with specifying a higher sampling frequency for the recorded data.

The graphical comparison indicated by the red arrows in Figure 4.3.1 of the two 1 minute and two 30 minute data sets for a period of 24-hours, are presented in Figures 4.3.2 and 4.3.3 respectively.
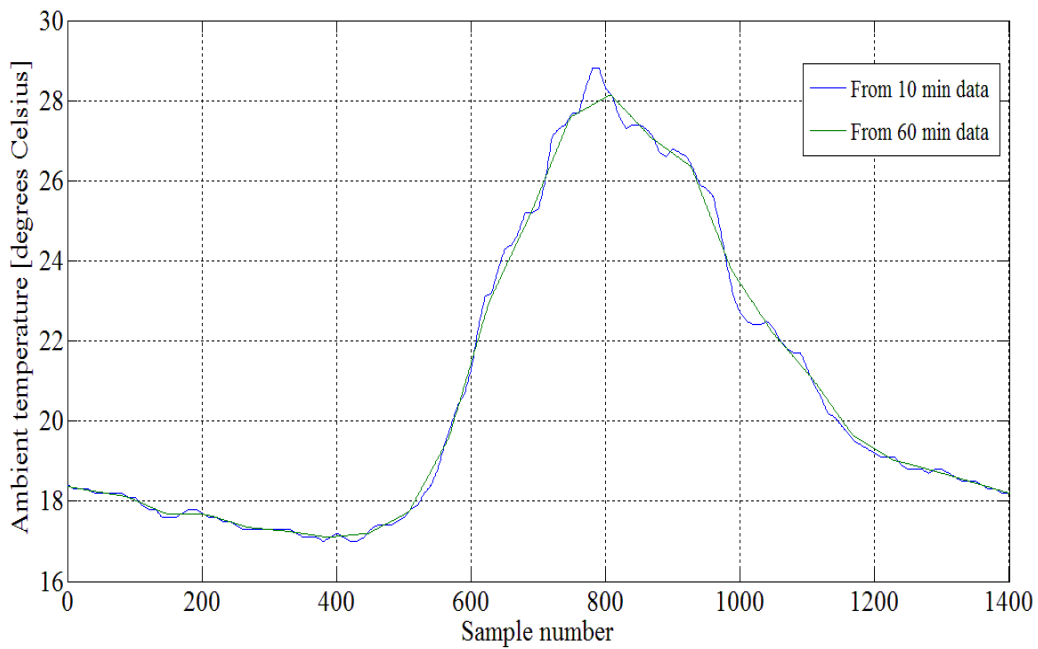
Figure 4.3.2: Ambient temperature data sets for a period of 24-hours of STL data at a 1 minute sampling period.
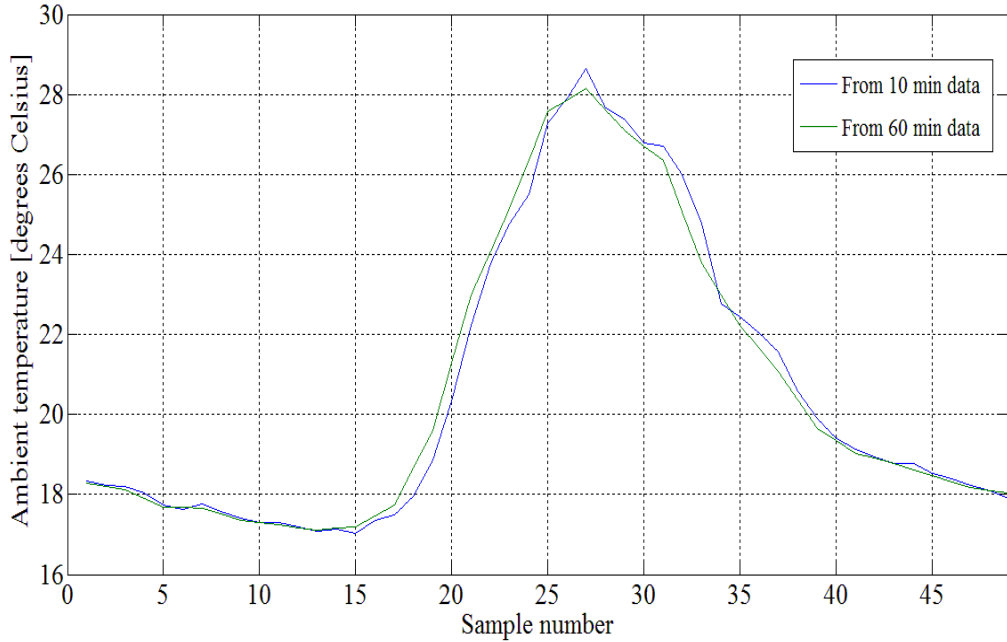
Figure 4.3.3: Ambient temperature data sets for a period of 24-hours of STL data at a 30 minute sampling period.

The averaging effect on the data is evident when comparing data sets. It can be seen that the data set produced from the 60 minute sampling period data has a lower variance in comparison to the data set with an original sampling frequency of 10 minutes. Table 4.3.1 shows the correlation coefficient and the covariance matrix of the two data sets. The 1 minute sampling period data is more closely correlated and had a greater covariance than the 30 minute sampling period data. The covariance between the two data sets is greater for the 1 minute period data because of the original sampling period from which the data sets were obtained. Information is lost during the averaging process that is not regained during the linear interpolation process.

Table 4.3.1: Correlation coefficient and covariance matrix of the two data sets for 1 minute and 30 minute sampling periods.

	Corr. coeff.	Cov. [1,1]	Cov. [1,2]	Cov. [2,1]	Cov. [2,2]
30 minute sampling period	0.9976	35.9922	35.7089	35.7089	35.5975
1 minute sampling period	0.9991	36.0370	35.7440	35.7440	35.5174

4.3.3   Statistical analysis of various time groupings of ambient temperature

The CTIA data set was used to investigate the various groupings of the ambient temperature data because of its sampling range that spanned nine years.

4.3.3.1   Statistical analysis of annual data

The purpose of the following investigation was to meet the second objective of the statistical analysis: to find the annual groupings of data with the greatest statistical extremes. When this data is used in the geyser simulation, it will represent the extreme range of the simulated data and its statistical nature which will give an indication of the sensitivity of simulated data of the geyser model to ambient temperature. The CTIA data set was grouped into years and the medians and variances of the years are plotted in Figures 4.3.4 and 4.3.5 respectively. This data set was linearly interpolated from 60 to 30 minute sampling periods. A comparison of the simulation results will be evaluated in Chapter five for the nine year period with maximum and minimum medians and variance.
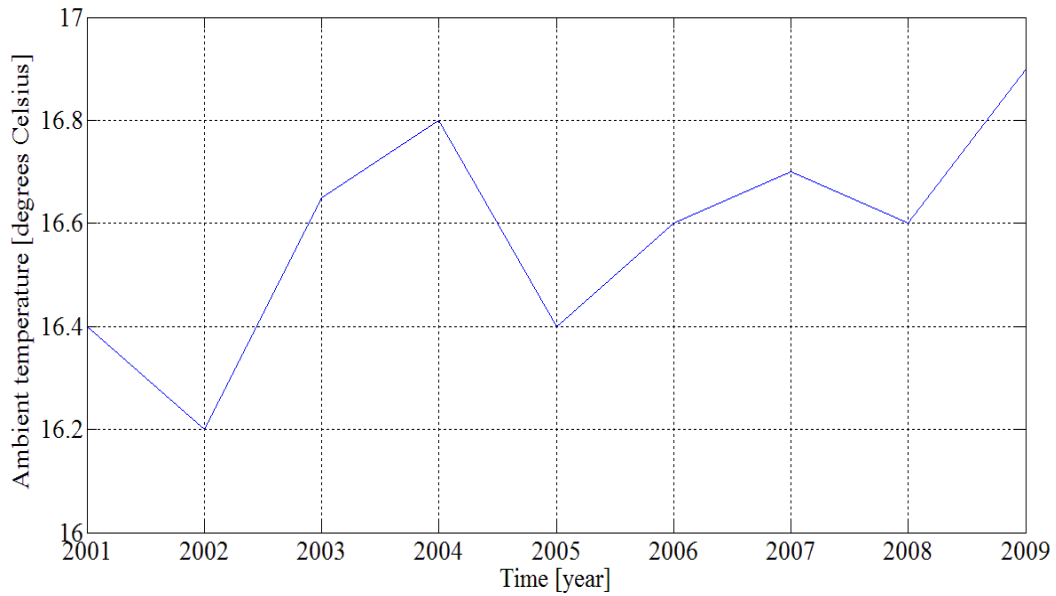
Figure 4.3.4: Annual median for ambient temperature of CTIA data from 2001 to 2009.
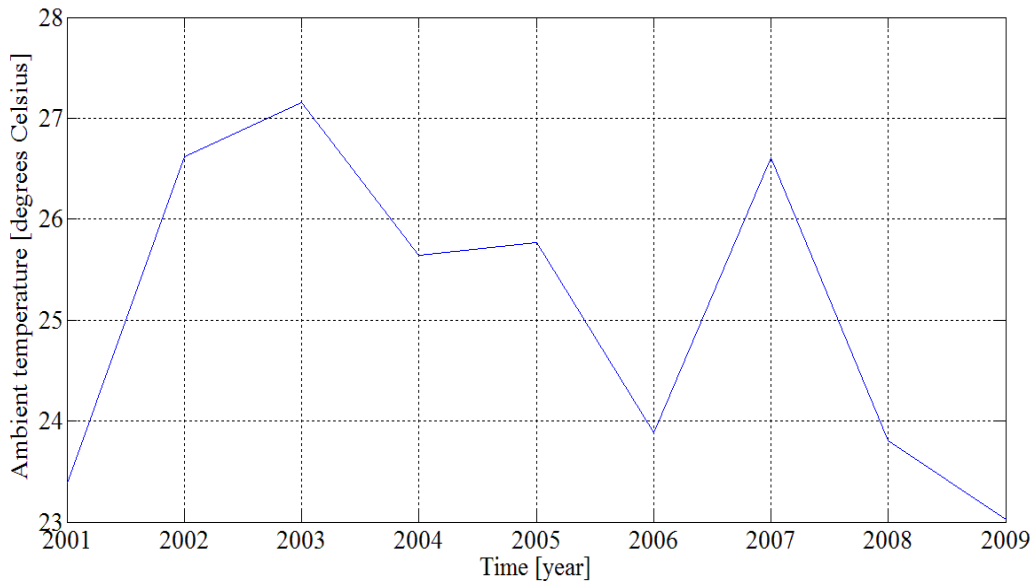
Figure 4.3.5: Annual variance for the ambient temperature of CTIA data from 2001 to 2009.

The temperature median and variance ranges for the data set were 0.7 and 4.2 degrees Celsius respectively.

Years 2002, 2007 and 2009 simulation results will be compared because of their relative range value combination of median and variance, described in Table 4.3.2.

88

Table 4.3.2: Selection of years of simulation due to their relative correlation coefficient from 2001 to 2009.

Year	Median	Variance	Corr. coeff. 2002	Corr. coeff. 2007	Corr. coeff. 2009
2002	Low	High	1	0.71	0.14
2007	High	High	0.71	1	0.18
2009	High	Low	0.14	0.18	1

It can be seen in the above table the relative correlation coefficients vary dramatically from 0.14 to 0.71 for the years 2002, 2007 and 2009.

The third objective of the statistical analysis of the ambient temperature was to investigate the effect which data grouping had on the statistical parameters. The CTIA data set of ambient temperature was divided up into years and this annual data was then box plotted. The resulting plot is depicted in Figure 4.3.6. It can be seen that the biggest difference between the box plots of the nine years is the inter-quartile range - the number and range of the outlying data. The statistical parameters of the entire data set are compared to each of the annual groupings.
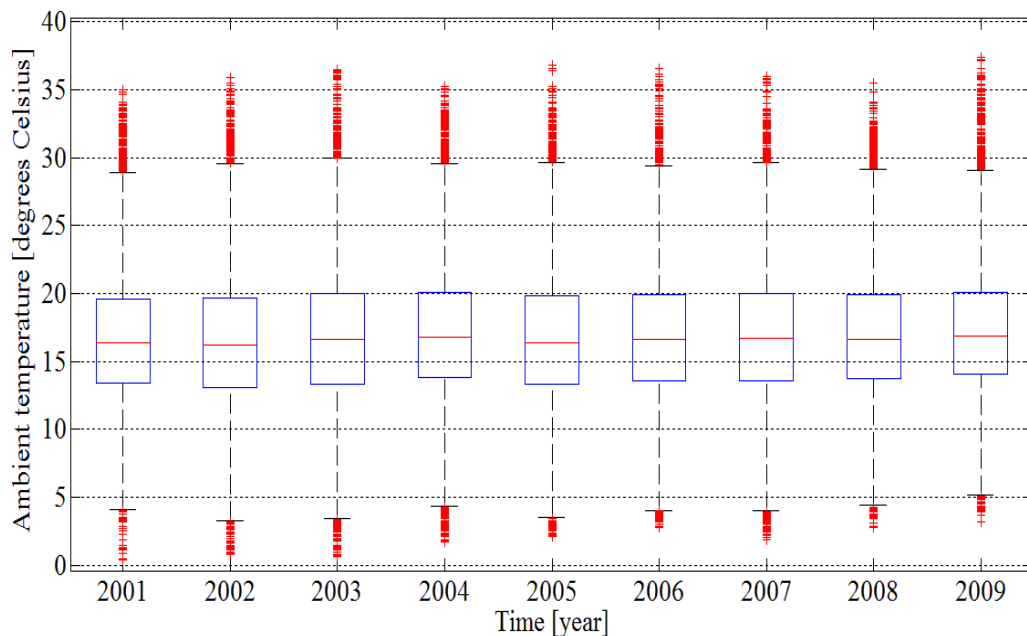
Figure 4.3.6: Box plots of CTIA data's annual recorded ambient temperature from 2001 to 2009.

The median for the nine years of CTIA data was 16.6 degrees Celsius and had a variance of 25.14 degrees Celsius. The percentage difference between the data set's median and variance and that for each year is tabulated in Table 4.3.3.

Table 4.3.3: Annual median and variance difference between the median and variance of the entire CTIA data set as a percentage.

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Median	-1.2	-2.4	0.3	1.2	-1.2	0	0.6	0	1.8
Variance	-7.3	6.1	8.4	2.1	2.6	-5.2	6.1	-5.5	-8.8

The greatest difference in median and variance was -2.4 percent and -8.8 percent for years 2002 and 2009 respectively. The negative sign indicates that the respective annual persentage was lower than the average of the entire data set. The histogram and kernel density estimates of the annual ambient temperature of CTIA data are depicted in Figure 4.3.7.
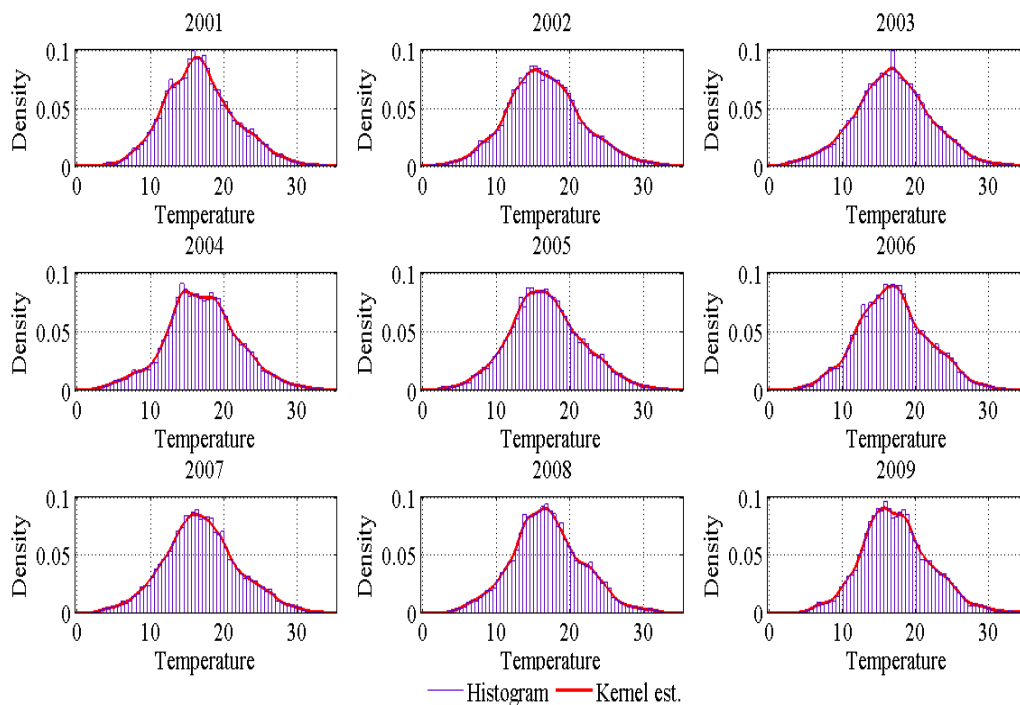
Figure 4.3.7: Probability density plots of the annual ambient temperature in degrees Celsius of the CTIA data from 2001 to 2009.

Years 2001, 2004, 2008 and 2009 suggest multi modal distributions which could be attributed to greater temperature ranges between summer and winter season groupings for those years. A pattern was seen in the correlation coefficient matrix. The coefficient was at a maximum when comparing the annual data with a five year period and at a minimum with a two to three year period, suggestive of a five year cycle for ambient temperature.

4.3.3.2   Statistical analysis of monthly data

In this section the ambient temperature of the CTIA data from January 2001 to December 2009 was grouped together into months of the year. These monthly data sets were then box plotted as depicted in Figure 4.3.8. It can be seen that the outlying data is higher than the median for the summer months, namely November to April. The inter-percentile range is smaller for the winter months, May to October, when compared to the summer months. The inter-percentile range is an indication of variance of the data. It can be seen that the median is not always positioned in the centre of the box; an indication of the varying range of skewness in the sampled data.
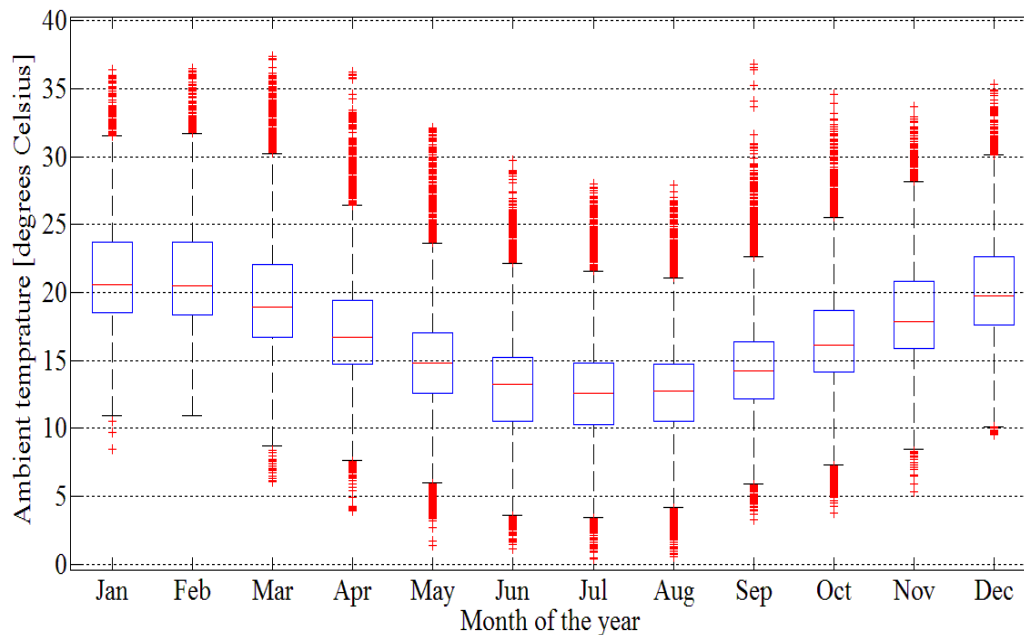
Figure 4.3.8: Monthly box plot of the CTIA data from 2001 to 2009.

The monthly histograms and kernel density estimates for summer and winter months are depicted in Figures 4.3.9 and 4.3.10 respectively.
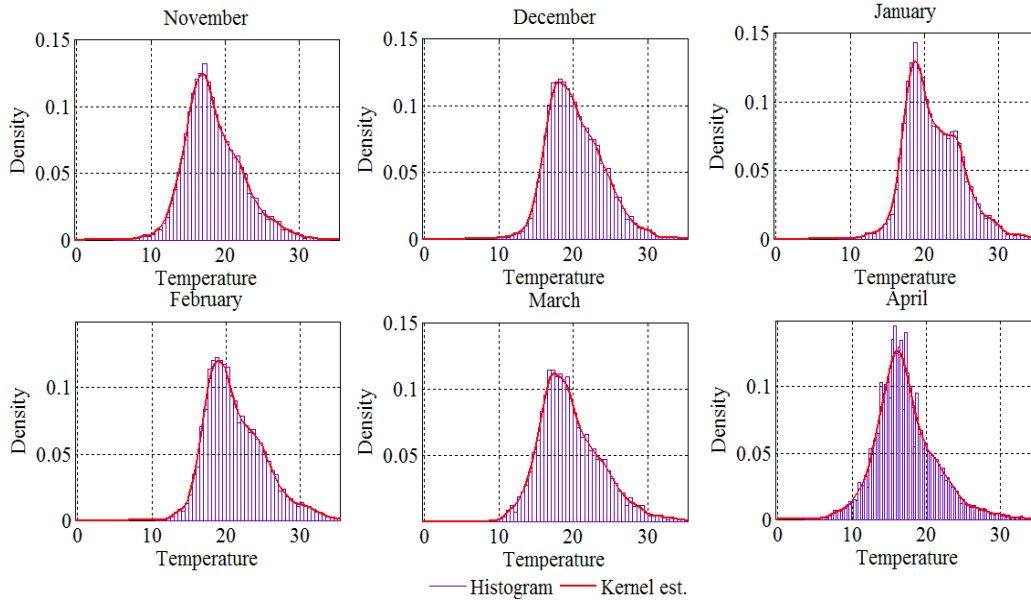
Figure 4.3.9: Monthly probability density plots of the CTIA ambient temperature data in degrees Celsius for the summer months from 2001 to 2009.
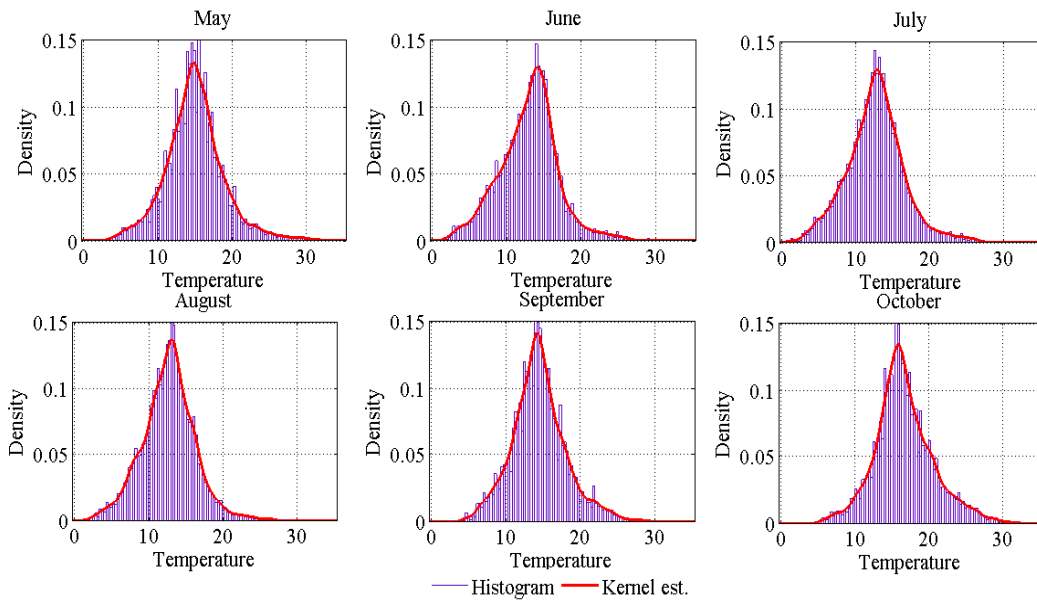
Figure 4.3.10: Monthly probability density plots of the CTIA ambient temperature data in degrees Celsius for the winter months from 2001 to 2009.

From Figures 4.3.9 and 4.3.10 it can be seen that January and June's density estimations are skewed above and below the median respectively. January's density estimation has a second less significant mode which is created by the higher day time average temperatures when compared

92

to the average night time temperatures. The transition months, representing autumn and spring, have little skewness in their density estimates.

4.3.3.3   Statistical analysis of seasonal grouping of the data

The data from the months are grouped into two data sets namely summer and winter. The box plots and density estimations are depicted in Figures 4.3.11 and 4.3.12 respectively.
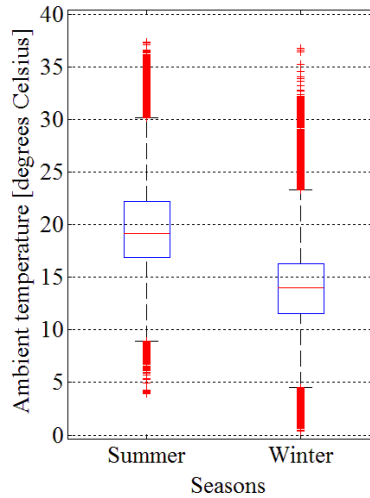
Figure 4.3.11: Box plots of seasonal grouping of the CTIA data from 2001 to 2009.
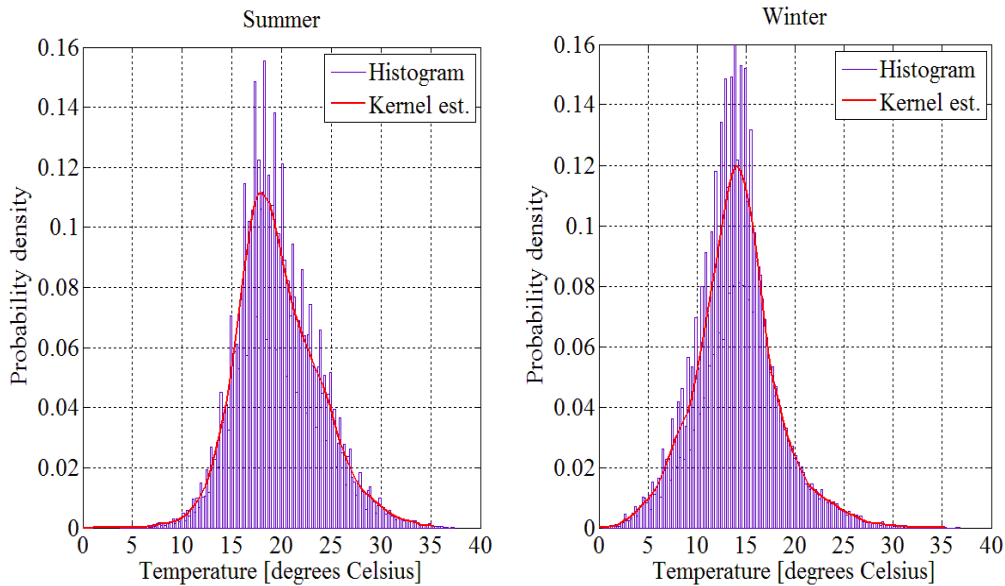
Figure 4.3.12: Probability density plots of summer and winter seasonal grouping of the CTIA ambient temperature data from 2001 to 2009.

The summer density estimate is skewed in the direction higher than the median and the opposite is true for the winter density estimate. The degree of skewness is less than that of summer and winter peak month namely January and June which can be seen in Figures 4.3.9 and 4.3.10 respectively. Temperature variance of summer is greater than that of winter. The secondary mode of January's density estimate is all but lost in the summer grouping of data.

4.3.3.4 Statistical analysis of 30 minute data

The temperature data for the months of January and June from January 2001 to December 2009 were grouped together into two sets of data. The 48 30 minute periods in these two data sets were box plotted. These two plots represent the peak summer and winter and are depicted in Figures 4.3.13 and 4.3.14 respectively.   The peak median temperature during January is experienced in the 26th 30 minute period which is equivalent to 13h00. The lowest median temperature experienced is in the 11th 30 minute which is equivalent to 05h30. The inter-percentile range is smallest at sunrise and greatest at midday. This can be attributed to hot and cool day-time temperatures which result in a greater inter-percentile range at approximately midday.
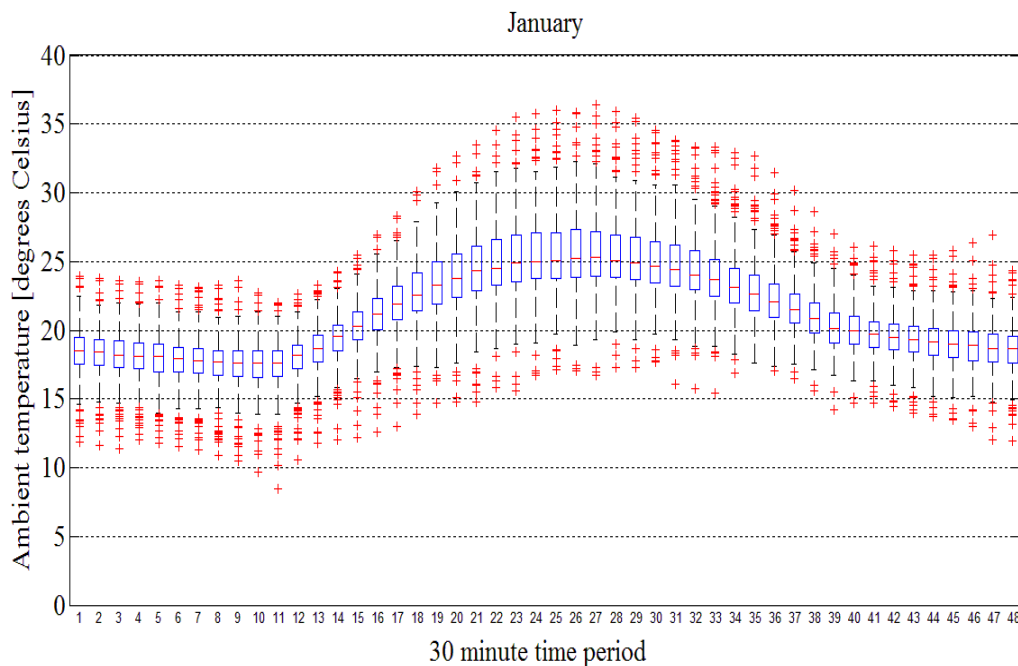
Figure 4.3.13: 30 minute box plots of January for the CTIA data from 2001 to 2009.

94
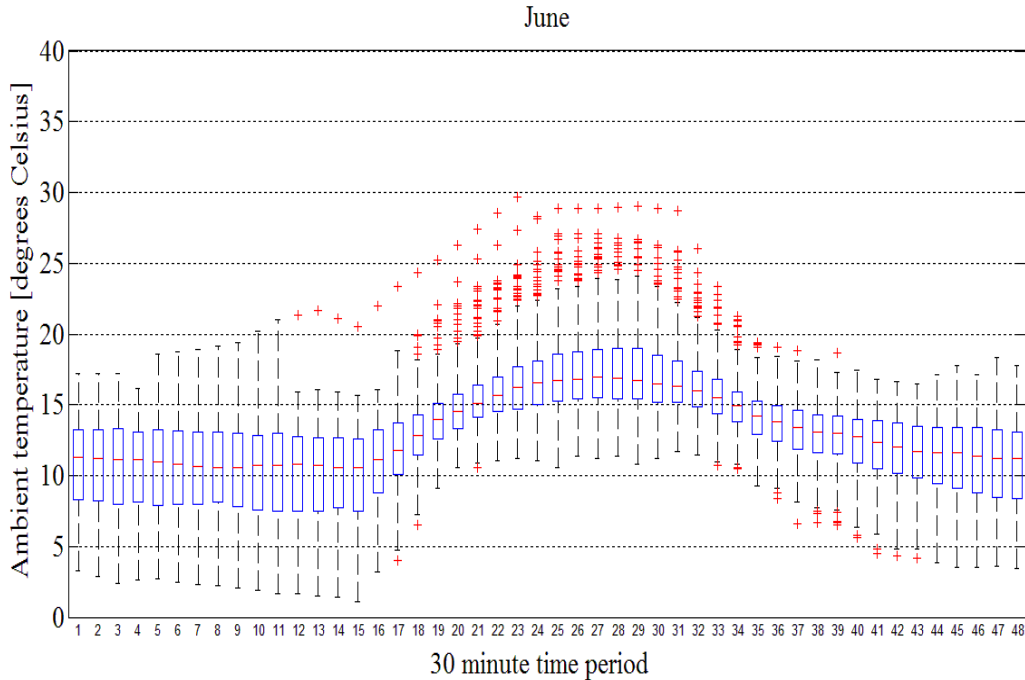
Figure 4.3.14: 30 minute box plots of June for the CTIA data from 2001 to 2009.

June has relatively few out-lying data below the medians, compared to January. June also has a greater inter-percentile range during the night time hours indicative of a greater variance in night time temperatures during the winter months. The shorter days are also evident in the 30 minute box plots of June, compared to January.

The seasonal data was now grouped into 48 30 minute periods and box plotted for the summer and winter months from January 2001 to December 2009. These are depicted in Figures 4.3.15 and 4.3.16 respectively. It can be seen in the summer data grouping that the diurnal (daily) median range is reduced when compared to January (refer to Figure 4.3.13). When comparing the box plots of January, June, summer and winter (refer to Figures 4.3.13, 4.3.14, 4.3.15 and 4.3.16) it can be seen that the seasonal grouping of the months has an averaging effect on the statistical parameters for each of the 48 30 minute periods.
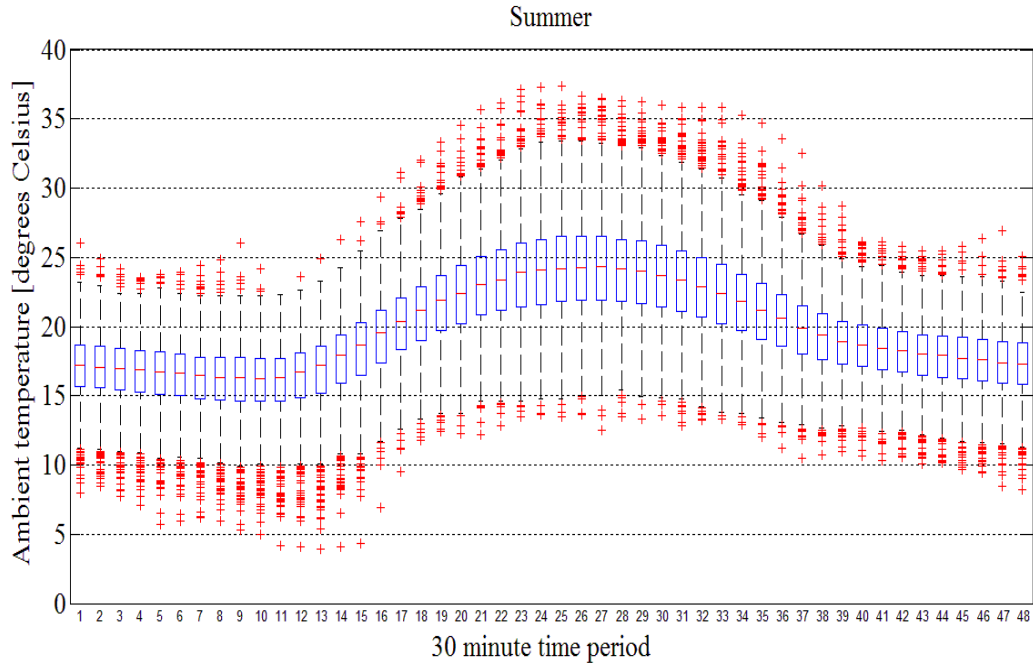
Figure 4.3.15: 30 minute box plots of the summer months for the CTIA data from 2001 to 2009.
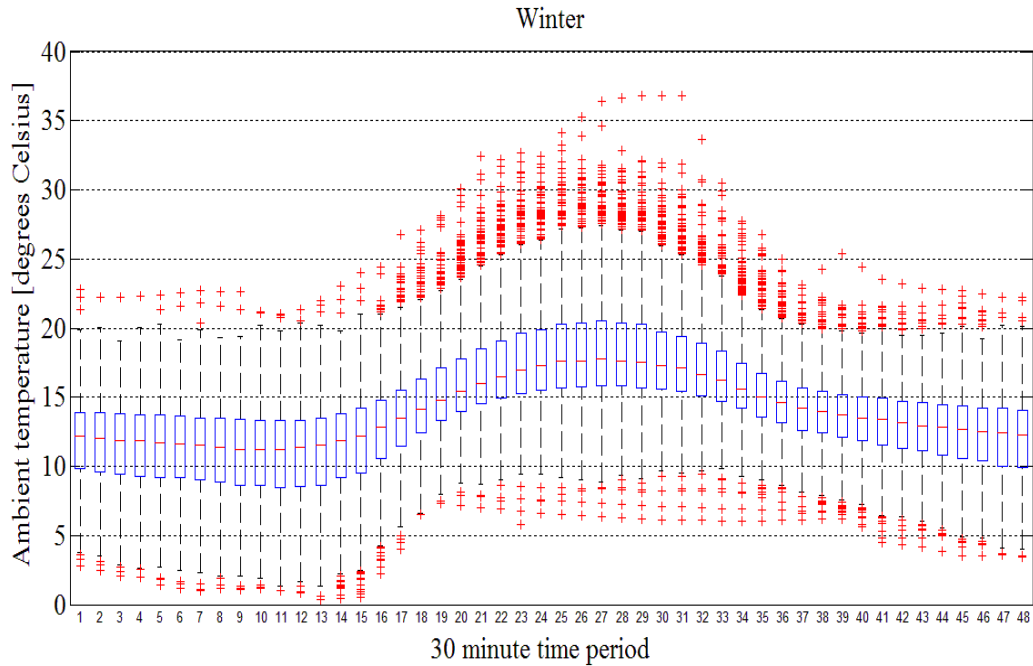
Figure 4.3.16: 30 minute box plots of the winter months for the CTIA data from 2001 to 2009.

In order to reduce the number of density plots, only four of the 48 30 minute time periods are presented for each season of the year. These four 30 minute periods 00h00, 06h00, 12h00 and

18h00 are the most significant for the geyser model. 30 minute periods 06h00 and 12h00 are respectively close to minimum and maximum ambient temperature for a 24-hour period, representing the respective times when the geyser potentially has its greatest and lowest standing losses. The 06h00 and 18h00 30 minute periods are normally the times at which there are peaks in hot water demand and subsequently the geyser thermostat would switch *on*. These are the times when two of the three input variables to the geyser model equal their extreme values. The distribution plots of summer and winter data sets are depicted in Figures 4.3.17 and 4.3.18 respectively. The median for all four summer 30 minute periods were higher than the corresponding winter periods. The direction of skewness of the night time periods is below the median and that of midday above the median. The 18h00 period has a more symmetrical density estimate.
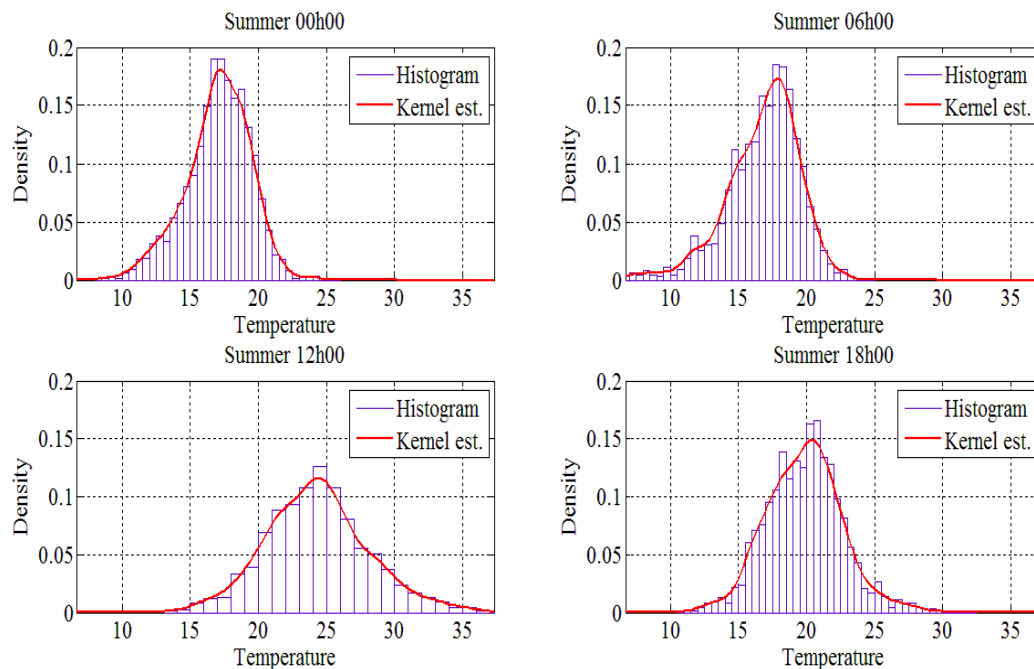
Figure 4.3.17: Probability density plots of the CTIA ambient temperature data in degrees Celsius of the summer months for time periods 00h00, 06h00, 12h00 and 18h00 from 2001 to 2009.
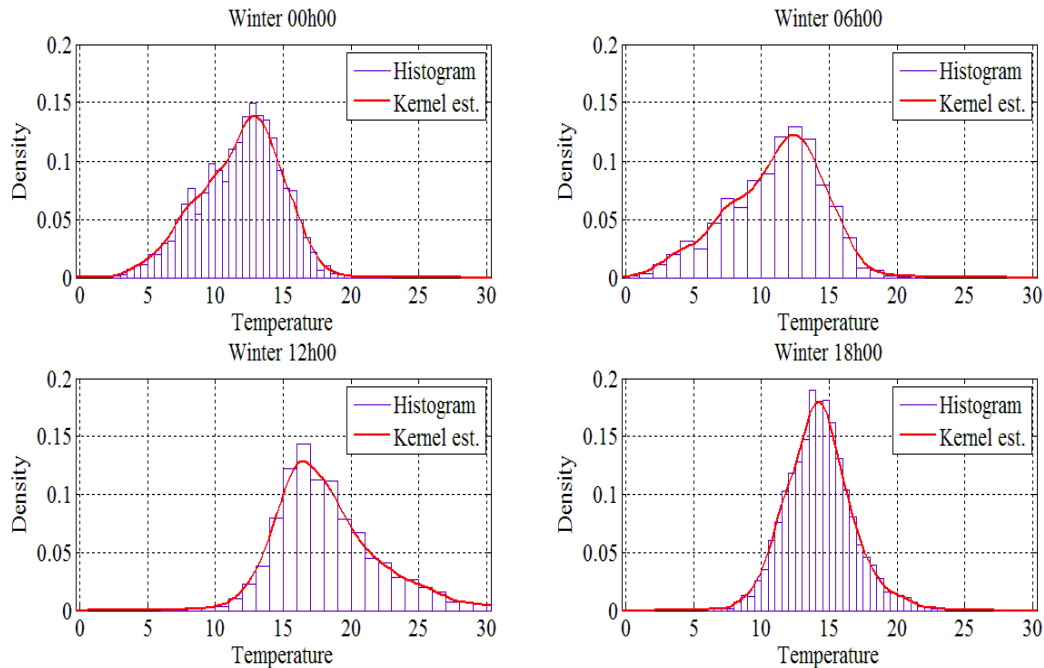
Figure 4.3.18: Probability density plots of the CTIA ambient temperature data in degrees Celsius of the winter months for time periods 00h00, 06h00, 12h00 and 18h00 from 2001 to 2009.

In Chapter five, the bivariate density estimates of ambient temperature and time of day the geyser was *on* will be presented. It is expected that the temperature density estimates should follow the same patterns for the same grouping of PDF plots because of the underlying dependency of electrical demand on ambient temperature.

4.4   Hot water consumption rate data

4.4.1   Overview

Two sets of hot water consumption data were generated. The first set was generated for use in the simulations using the STL ambient temperature data set. This hot water consumption rate data had data sampling periods of 1 minute and 30 minutes. The second set of consumption data was generated for use in the simulations that used the CTIA ambient data set. This second set of hot water consumption data had a sampling period of 30 minutes. The hot water consumption for both sets had a winter and summer profile with three different overall demand profiles. The three different overall demand profiles were representative of the different housing densities which had different rates of consumption because of the average number of occupants in each house and the level of affluence. A summary of the data sets is given in Table 4.4.1.

98

Table 4.4.1: Summary of the hot water consumption rate data sets.

Hot water demand	Simulation period	Sampling period	Housing density
Data set for STL simulations	1 year	1 & 30 min	High, medium, low
Data set for CTIA simulations	9 years	30 min	High, medium, low

The statistical parameters and distributions remain the same for both data sets. The data set for the CTIA data simulations has a greater sample count because it was evaluated over nine years. A set of random numbers were generated from a normal distribution with zero mean and 17 percent standard deviation in the summer, and 30 percent standard deviation in the winter, which is consistent with the Meyer et al. findings [31]. These random numbers were then added to the recorded average values from the Meyer et al. study [31]. The method used to generate the hot water consumption rate is described in detail in Chapter three. Nine years' worth of hot water consumption was generated for the CTIA simulations and the statistical analysis of this data is analysed in subsections 4.4.2 to 4.4.4. During the generation and analysis of the hot water consumption data, it was grouped together into four distinct profiles namely:

- summer weekday;
- summer weekend day;
- winter weekday, and
- winter weekend day.

4.4.2   Hot water consumption rate during summer

The box plots of the hot water consumption rates for the three types of housing densities during summer are depicted in Figure 4.4.1. It can be seen in these box plots that the consumption rate for low-density housing is more than double that of high-density housing. The consumption rate for low-density housing during weekdays is higher than on weekend days. The opposite is true for the hot water consumption rate in low-density housing. The box plots also suggest a skewness in the data, particularly in the high-density housing category. It is evident that there is a greater inter-percentile range or variance in the data for the low-density housing category compared to the other two densities. These observations are confirmed in the plots of the density estimates of the data, depicted in Figure 4.4.2.
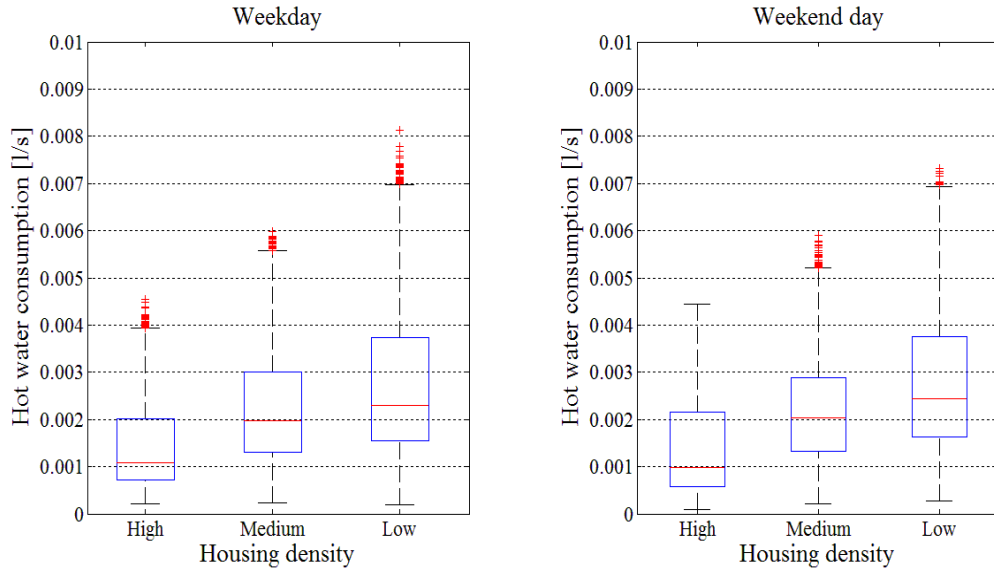
Figure 4.4.1: Box plots of the hot water consumption rates for the three types of housing densities during summer weekdays and weekend days.
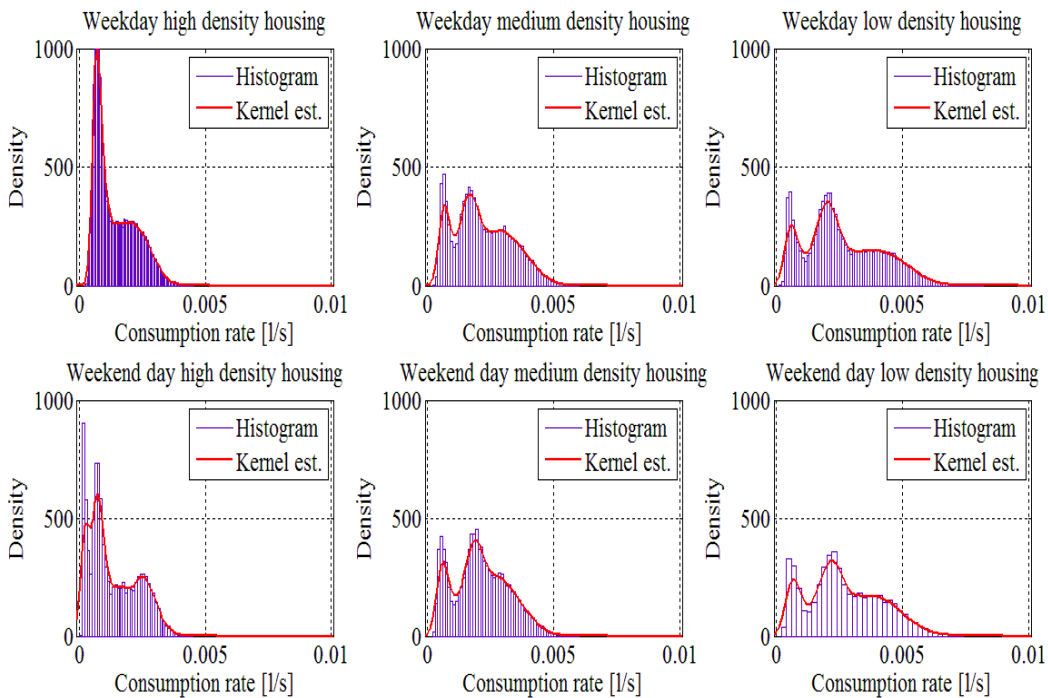
Figure 4.4.2: Probability density plots of the hot water consumption rate during summer for weekday and weekend day for high-, medium- and low-density housing.

It can be seen that all of the probability density estimates are bimodal which stems from the fact that the hot water consumption profiles had three distinct regions. There were two high

100

consumption rate periods: one in the morning and one in the evening. These two periods were associated with shower times. The third period, which occurred between the two periods of high consumption, was characterised by lower flow but over a longer period of time. The first mode in the probability density plots accounts for the period of low consumption rate. Its density is highest because it has more associated samples. The second lower mode is the period associated with higher consumption rate caused by demand from shower periods. Hot water demand from the high-density has a much smaller variance when compared to the low-density housing. There is no significant difference between the weekday and weekend day density plots. It can be concluded that the hot water demand volume remains similar for weekdays and weekend days but the time, at which the demand occurs, varies.

4.4.3   Hot water consumption rate during winter

The box plots of the hot water consumption rates for the three types of housing densities during winter are depicted in Figure 4.4.3.
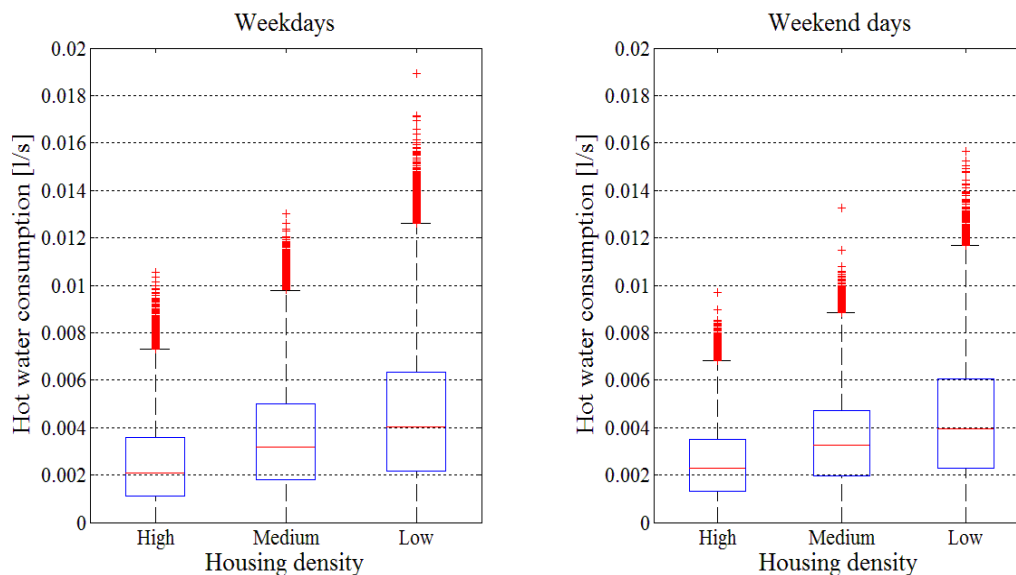
Figure 4.4.3: Box plots of the hot water consumption rates for the three types of housing densities during winter for week and weekend days.

The consumption rate is approximately double for the high-density housing grouping and approximately 1.5 times for the medium- and low-density groupings, when comparing the winter to the summer periods. The only noticeable difference between weekdays and weekend days is that the consumption rate is higher during weekend days for the high-density housing group,

101

which is evident from the value of the medians. There is a larger inter-percentile range or variance in the low- versus high-density housing hot water consumption rate data. These observations are confirmed in the plots of the density estimates of the data which are depicted in Figure 4.4.4. When the summer time probability density plots are compared to the winter plots it can be seen that the variance is greater during the winter period which is consistent with the higher demand for hot water for this period. The bimodal feature in the probability is less evident during the winter because of the increased variance in hot water demand. The only difference between week and weekend day, is in the high-density housing grouping where there are two modes for the weekend day which was previously attributed to the fact that the working class did not have to be at work as early on a weekend day as on a weekday.
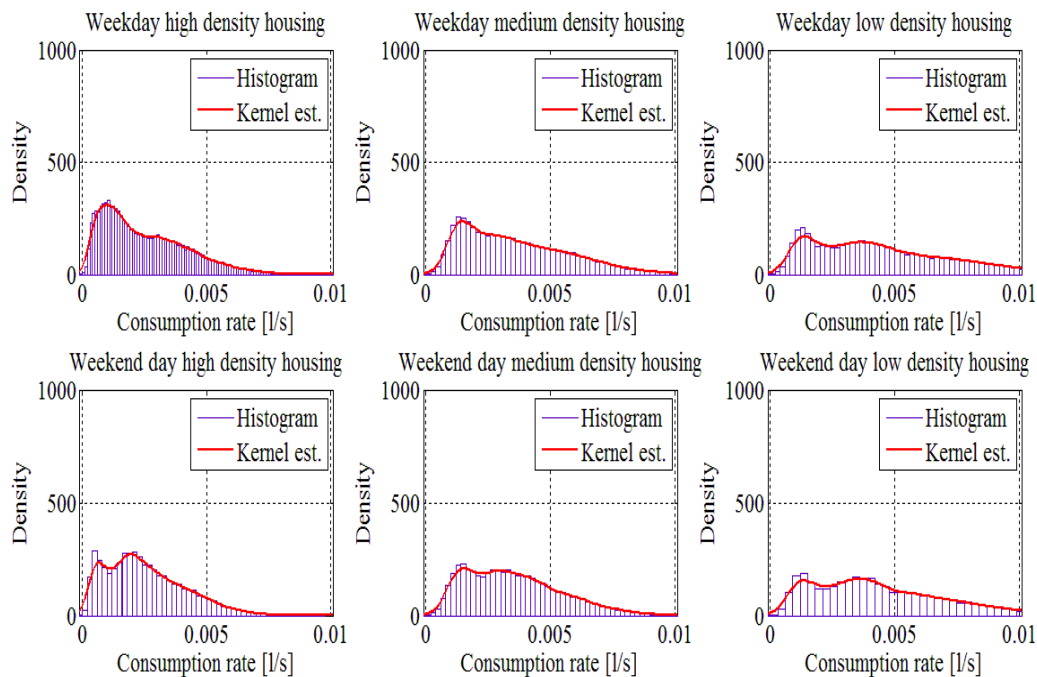
Figure 4.4.4: Probability density plots of hot water consumption rate during winter for week and weekend days for high-, medium- and low-density housing.

4.4.4 Hot water consumption rate for the evaluation period

In this section, the entire hot water consumption rate data is grouped together and statistically analysed. In Figure 4.4.5, the box plots of the three housing density groupings are depicted for an all year round hot water consumption rate. Once again, the high-density housing group's

consumption rate data has a smaller variance when compared to the other two groupings. Low-density housing has the largest consumption rate of the three groupings.
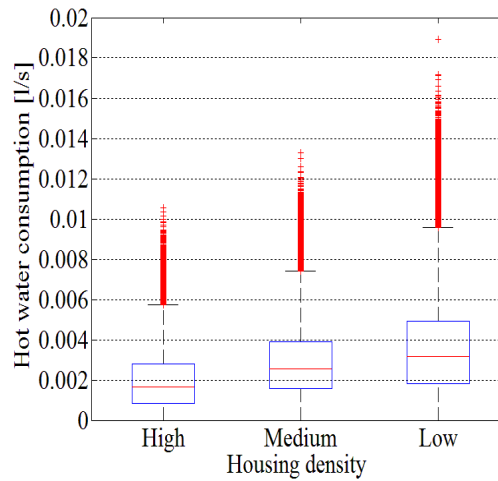
Figure 4.4.5: Box plots of the hot water consumption rate for high-, medium- and low-density housing groups.

The probability density estimates of the three housing density groupings are depicted in Figure 4.4.6.
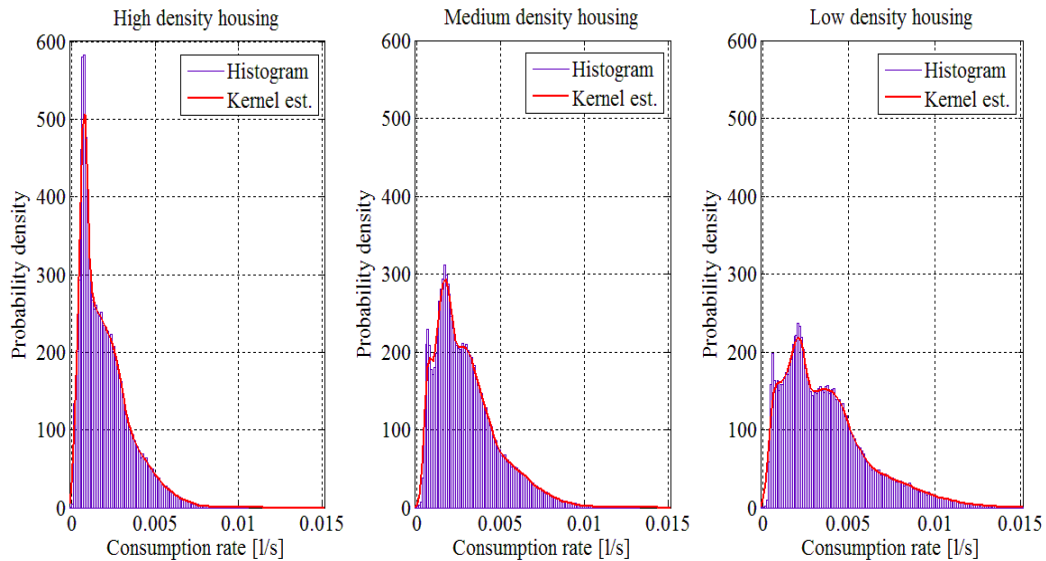
Figure 4.4.6: Probability density plots of hot water consumption for the evaluation period for high-, medium- and low-density housing groups.

103

The probability density estimate of the high-density housing grouping shows a high probability for a low consumption rate when compared to the other two groupings. The variance of this grouping is also the smallest. In the medium- and low-density housing groupings there is evidence of a tri-modal probability density estimate which is the result of the combined grouping of the seasons and characteristic days of the week.

4.5   Inlet water temperature data

4.5.1   Overview

The inlet water temperature is generated by a function which was described in detail in Chapter three. The advantage of this method is that multiple data sets can be generated by varying the input variables to the function. Utilising the function facilitates the ability to investigate the sensitivity of the geyser model simulation results to the inlet water temperature. The input variables to the inlet water temperature function are:

- depth at which the water supply pipe is buried in the soil;
- damping depth of the soil, and
- surrounding ambient temperature.

Different soil types will have different dampening depths. The inlet water temperature is represented by a sinusoidal function with a period of one year. The phase angle is dependent on the depth at which the water pipe is buried and the date at which the lowest ambient temperature was experienced in that year. The amplitude of the sinusoid is a function of the minimum and maximum temperatures recorded in that year. Its Y axis displacement is equivalent to the average ambient temperature. Geyser simulation sets require inlet water temperature data. The simulations require either a 1 or 30 minute sampling period. The theoretical depth at which the inlet water supply pipe was varies between one and seven meters. The range in depth should indicate the degree of influence the water supply pipe burial depth has on the simulated demand.

Seven data sets were required for the simulations in Chapter five and are described as follows in Table 4.5.1:

Table 4.5.1: Description of the data sets generated for inlet water temperature to the geyser model.

Data set code name	Depth of supply water pipe [m]	Original ambient temperature data set	Original sampling rate [minutes]	Generated sampling rate [minutes]
STL-01-10-01	1	STL	10	1
STL-01-10-30	1	STL	10	30
STL-01-60-01	1	STL	60	1
STL-01-60-30	1	STL	60	30
CTIA-01-60-30	1	CTIA	60	30
CTIA-07-60-30	7	CTIA	60	30
CTIA-1&7-60-30	1 (summer) 7 (winter)	CTIA	60	30

Each data set code name is formed by combining:

the original temperature data set (STL or CTIA) - the burial depth of water supply pipe - the original sampling period of the data - the new sampling period of the data.

4.5.2   Description of the inlet water temperature data sets generated by the STL data set

The first four inlet water data sets in Table 4.5.1 are used in the geyser simulation sets that are designed to investigate the effects of different input data sampling periods generated from different original sampling period data. Data set STL-01-10-01 is depicted in Figure 4.5.1. The average inlet water temperature was just less than 20 degrees Celsius. The maximum and minimum inlet water temperature was 23.8 and 16.5 degrees Celsius respectively. These maximum and minimum temperatures were generated in the months of April and October respectively. The fact that the maximum temperature only occurs in April is confirmed by the knowledge of the soil's ability to store energy resulting in a time delay between the maximum ambient temperature recorded and that of the inlet water temperature. This energy storing ability also acts as a high frequency filter which filters out the higher frequency diurnal ambient temperature fluctuation and leaves a lower frequency annual temperature fluctuation. This filtering effect becomes greater with an increasing depth below ground level. At one meter below ground level there is no evidence of the higher frequency daily ambient temperature fluctuation.
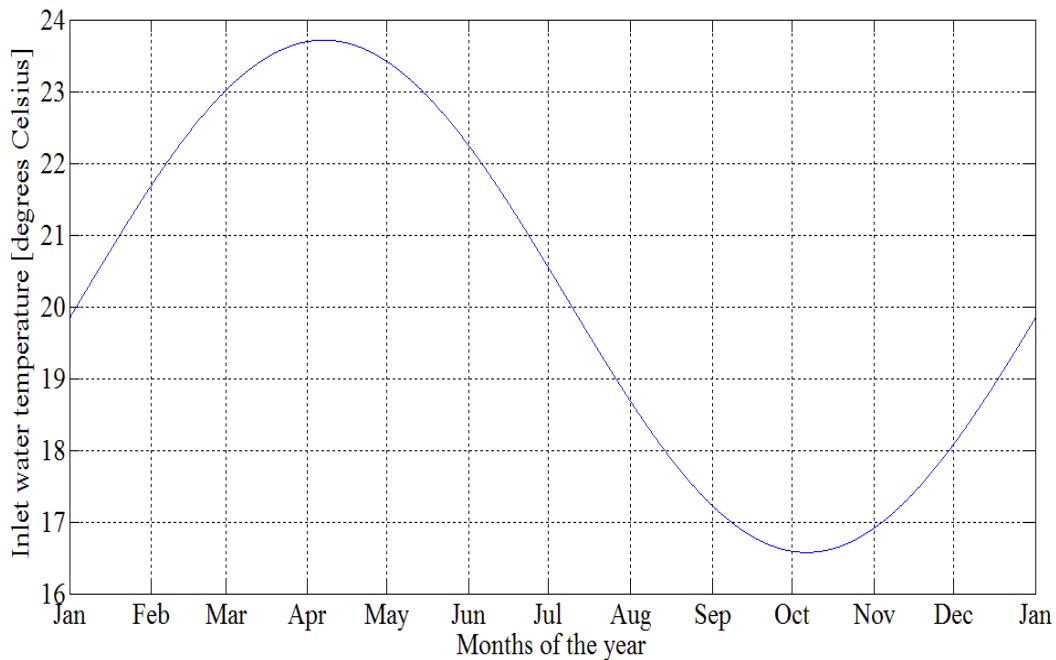
Figure 4.5.1: Inlet water temperature to the geyser generated from the STL ambient temperature data set at a supply water pipe depth of one meter.

4.5.3 Description of the data set generated by the CTIA temperature data and a water supply pipe burial depth of one meter

The data set, CTIA-01-60-30, was generated to investigate the effect of the range of inlet water temperature on the electrical demand. The data set was generated using the CTIA ambient temperature data set and a supply water pipe burial depth of one meter below ground level. This data set is depicted in Figure 4.5.2. Nine temperature cycles are plotted for the nine year simulation period. The minimum and maximum temperatures range from 16.8 to 21.8 degrees Celsius respectively. The mean temperature was 18.8 degrees Celsius. The variables of the function that generates the inlet water temperature were obtained by taking the averages of the minimum, maximum and mean of the annual temperature data over the nine year period in order to generate a continuous sinusoid of inlet water temperature. The density estimate of this data is depicted in Figure 4.5.3. It can be see that there is a greater density of data at the minimum and maximum temperature values.

106
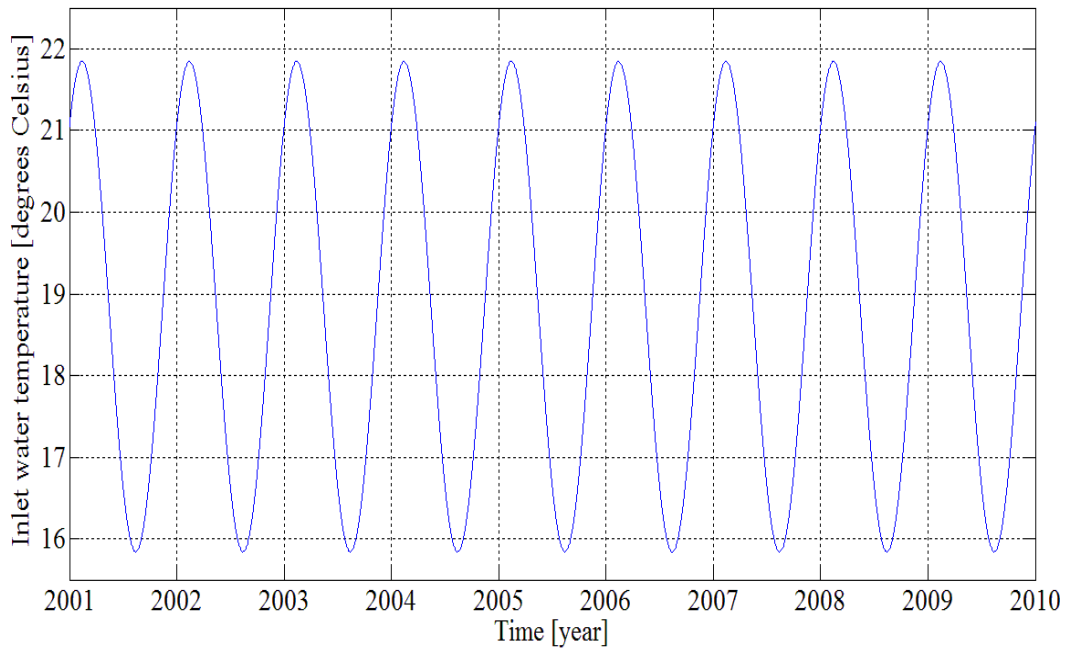
Figure 4.5.2: Inlet water temperature to the geyser with a supply water pipe burial depth of one meter using the CTIA temperature data from 2001 to 2009.
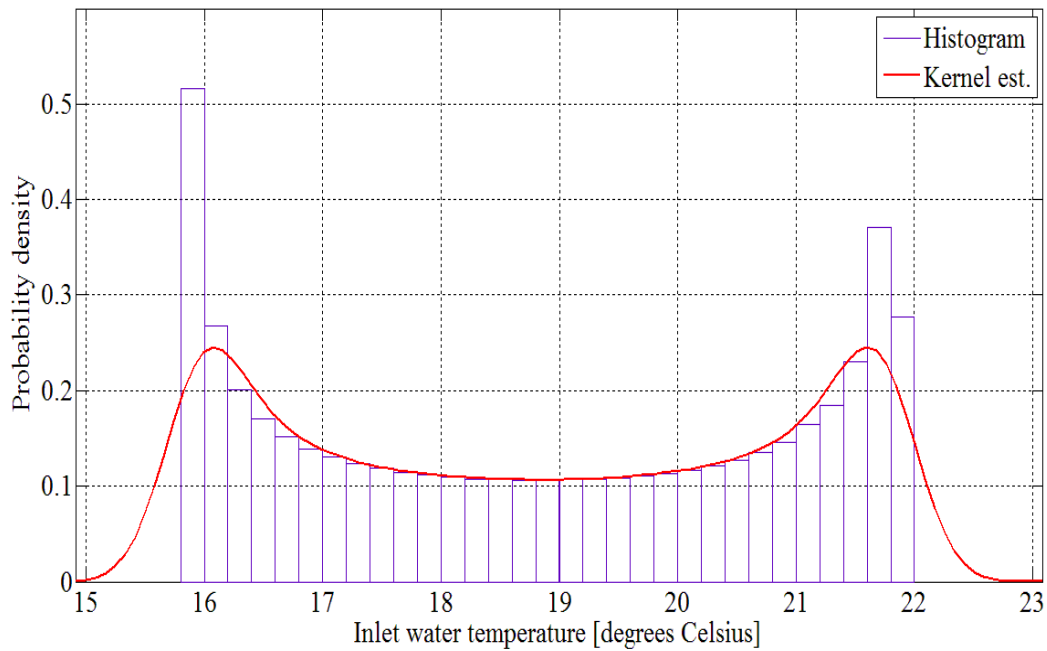
Figure 4.5.3: Probability density plot of the inlet water temperature of the geyser at a supply water pipe burial depth of one meter.

107

4.5.4 Description of the data set generated by the CTIA temperature data and a water supply pipe burial depth of seven meters

For this data set the burial depth of the water supply pipe was seven meters. It can be seen in Figure 4.5.4 that the amplitude of the temperature fluctuation has reduced from six to 0.35 degrees Celsius (refer to Figures 4.5.2 and 4.5.4). The time of year when the maximum temperature of the supply water is reached is also later than when the water pipe was buried at one meter below ground level. The average temperature has remained the same for both pipe burial depths, due to the fact that the inlet water temperature during summer has been reduced by approximately three degrees whereas the winter inlet temperature increased by the same amount. It is important to note that the hot water demand is substantially greater during winter and raising the inlet water temperature during that time of the year should reduce electrical demand. The extent of this reduction in demand will be documented in Chapter five. The variables of the function that generates the inlet water temperature were again obtained by taking the averages of the minimum, maximum and mean of the annual temperature data over the nine year period. This was done so that a continuous sinusoid of inlet water temperature could be generated.
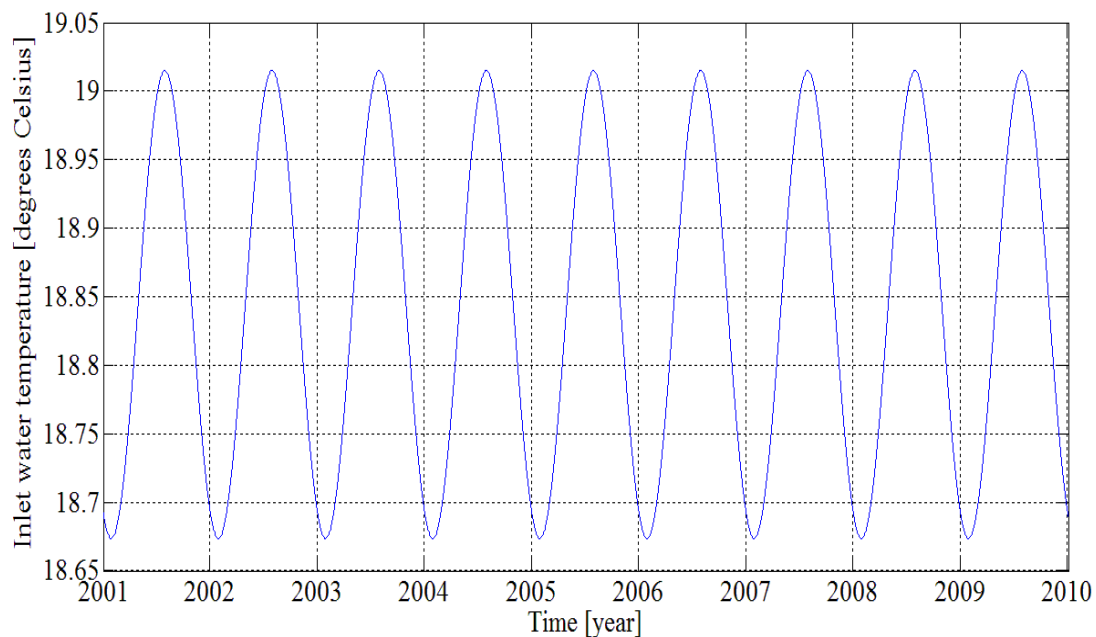
Figure 4.5.4: Inlet water temperature to the geyser with a supply water pipe burial depth of seven meters using the CTIA ambient temperature data from 2001 to 2009.

108

4.5.5   Description of the data set generated by the CTIA temperature data and a water supply pipe burial depth combination of one and seven meters

Data set CTIA-1&7-60-30 was generated by using a combination water supply burial depth of one and seven meters. This data set was generated based on the knowledge that higher inlet water temperature of the summer months are expected at a pipe burial depth of one meter below ground level and that higher inlet water temperature of the winter months are expected at a pipe burial depth of seven meters below the ground level. This hypothetical case was designed to investigate the outcome if a simple water storage system, which allows for the depth of water storage to vary during the course of a year, would make an economically significant impact on electrical demand. The cost of demand saving can then be used to amortise the cost of the mechanism installed to achieve the desired inlet water temperature profile. The generated inlet water temperature is depicted in Figure 4.5.5. The average temperature has been raised and the density estimate has been modified by the storage of water at a combination of burial depths.
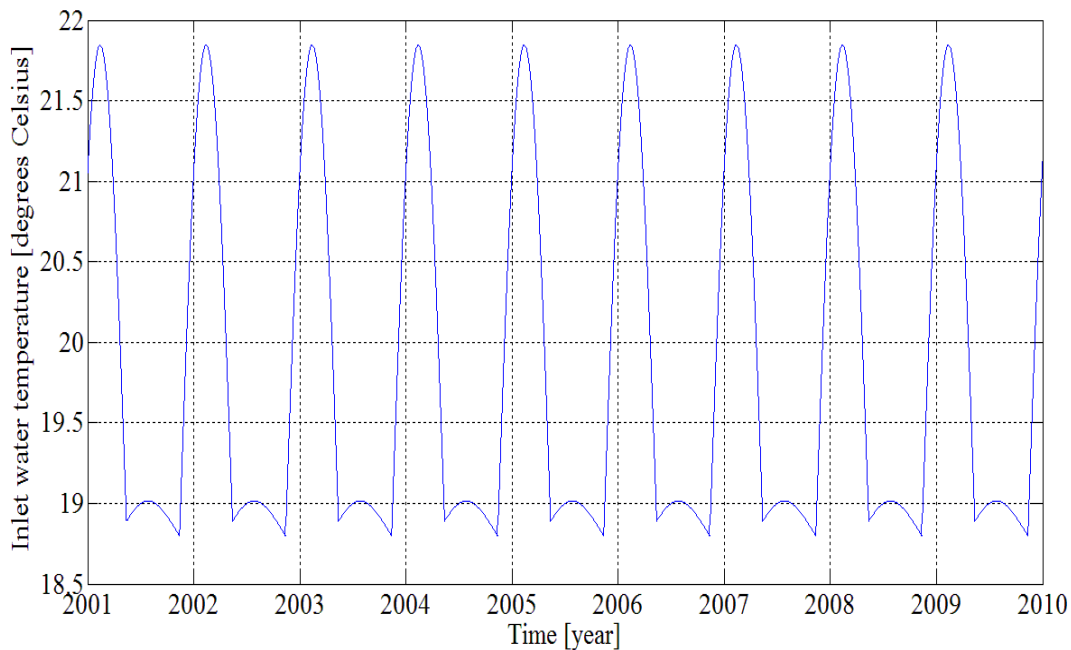
Figure 4.5.5: Inlet water temperature which was generated from combining a supply water pipe burial depth of one and seven meters.

The density estimate is depicted in Figure 4.5.6 and it can be seen clearly that the density at 19 degrees Celsius has dramatically increased. The results of the simulation using this data set for

109

inlet water temperature will give an indication whether it would be economically viable to develop a mechanism to achieve this inlet water temperature profile.
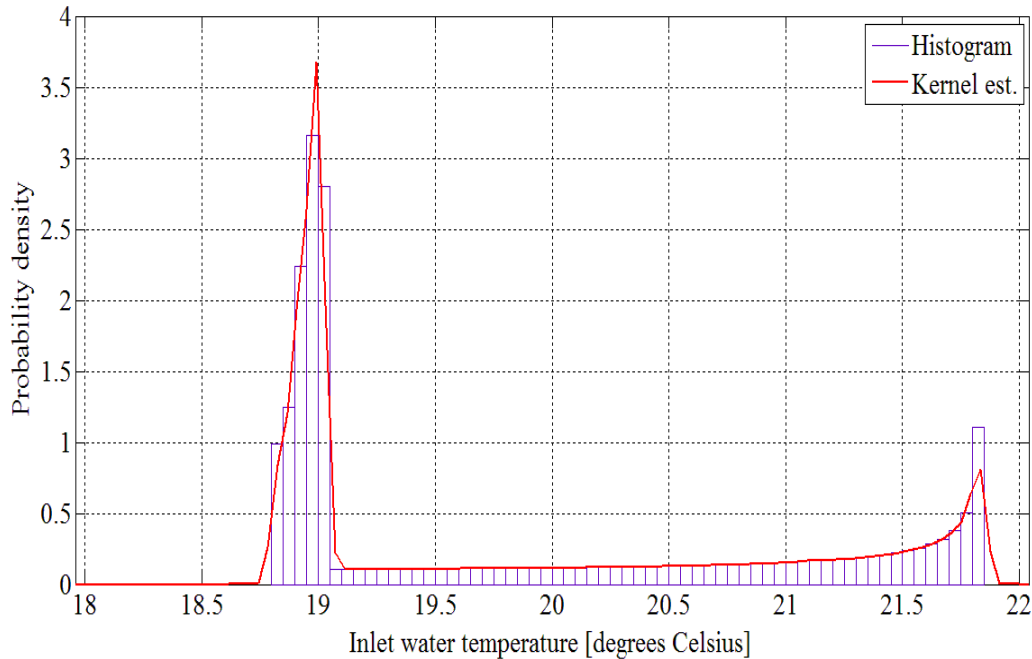
Figure 4.5.6: Probability density plot of the inlet water temperature to the geyser which makes use of a mechanism to store water at different depths below ground level.

4.6   Conclusions drawn from the statistical analysis of input data

The following outcomes with regards to the statistical properties for ambient temperature are noteworthy:

- median temperature was higher for the summer period than for the winter period;
- variance in ambient temperature was greater in the summer day time than in the winter day time, and
- variance in the ambient temperature for winter night time was greater than that of the summer night.

The following with regards to the statistical properties for the hot water consumption rate are noteworthy:

- average and variance in the consumption rate is greatest for low-density housing;

110

- variance in the consumption rate is greater in the winter for all three groupings of housing densities, and

- in the groupings of season and characteristic day type, the probability density is bimodal.

The following outcomes with regards to the statistical properties for inlet water temperature are noteworthy:

- average inlet water temperature was approximately equal to the average ambient temperature for an evaluation period of one or more years;

- amplitude of the seasonal temperature fluctuation decreased for an increase water supply depth below ground level, and

- time at which peak inlet temperature occurs during the year is a function of inlet water pipe burial depth below ground level.

4.7   Conclusion

In the first section of this chapter a detailed account was given of the statistical methods that were used to analyse the input data to the geyser model. This account was followed by the analysis of each of the input variables and the data sets that were produced for input into the simulations, which are further described in Chapter five. Finally a summary was given of the most noteworthy findings that resulted from the statistical analysis of the geyser input variables.

Chapter 5   **Statistical analysis of a case study of a geyser model's output data**

5.1   Overview

The output demand data of the geyser model simulations are statistically analysed in a similar way as was done in Chapter four. The difference in the approach of the statistical analysis of this chapter compared to Chapter four comes by way of the introduction of the bivariate density estimates which illustrate the dependency of the simulated demand versus ambient temperature. The following objectives are introduced in this chapter:

1. investigation of the sensitivity of the geyser model's simulated data to the sampling period of the input variable;

2. investigation of the geyser model's simulated demand sensitivity to the mean and variance of annual ambient temperature, and

3. evaluation of the geyser model's demand impact as a result of the following ECMs:

     o   the adjustment of the geyser's thermostat from 65 to 55 degrees Celsius;

     o   the retrofitting of low flow shower roses to reduce warm water demand, and

     o   modification of the geyser inlet water temperature profile through various methods.

These objectives were addressed as follows:

• The investigation of the sensitivity of the geyser model's simulated data to input variable sampling period was achieved by a four way comparison of simulated demand data. The details of this comparison are presented in the next subsection. This analysis technique was implemented to determine the most appropriate sampling period for input variables if data were to be collected for geyser simulations. Alternatively, these comparisons are also designed to determine the effect interpolation or averaging has on simulated demand of the recorded data.

• The simulations to generate the geyser model's simulated demand sensitivity to the mean and variance of annual ambient temperature were designed to determine if there would be

any significant implications on the statistical properties of the simulation data if the evaluation period's temperature data were grouped together into seasons. If the demand data were significantly dependent on the variation of annual temperature it may be deemed necessary to further group the ambient temperature into characteristic years and not just seasons.

- The evaluation of the geyser model's demand impact as a result of the ECMs. This was achieved by running two geyser models in parallel. The one model simulated the pre-ECM demand profile and the other the post ECM demand profile. One simulation was run with a zero hot water demand profile. This was done to investigate the dependency of the no load losses on ambient temperature.

All of the other simulations with regards to these three objectives were run with the three distinctive hot water demand profiles characteristic of the respective low, medium and high housing densities. These simulations gave an indication of the simulated demand dependency on hot water consumption profile. The results of objectives one and two are presented in the form of a comparison of the total kWh demand, correlation and covariance of the different demand simulations. The results of objective three are presented in the form of a kWh demand impact values and the PDFs of the pre- and post-ECM demand profiles versus temperature dependency.

5.2 Sensitivity analysis of the geyser model's simulated data with respect to the sampling period of the input variables

In order to test the sensitivity of the geyser model with regard to the sampling period of the input variables, model simulations were run with four sets of input variables with three different hot water consumption profiles.

5.2.1 Input variable data sets

A description of each of the input variable data sets is given in the next three subsections.

5.2.1.1 Ambient temperature

Four sets of ambient temperature data were generated from the STL data set as described in Chapter four. The methods used to generate the data were described in Chapter four and repeated here in Figure 5.2.1, as an adaptation of Figure 4.3.1.

1 min sampling period data originally from 10 min sampling period data	Interpolate ⬅	**STL** 10 min sampling period input variable data	Average ➡ 30 min sampling period data originally from 10 min sampling period data
Statistical comparison of input data ⬍		⬇ Average ⬇	⬍ Statistical comparison of input data
1 min sampling period data from 60 min sampling data	⬅ Interpolate	60 min sampling period data originally from 10 min sampling period data	Interpolate ➡ 30 min sampling period data from 60 min sampling data

Figure 5.2.1: Four sets of ambient temperature data used in the four sets of geyser simulations to test the sensitivity of the geyser model to the input variable sampling period.

5.2.1.2 Hot water demand

Three sets of hot water demand profiles were generated at 1 and 30 minute sampling periods namely:

- high-density housing hot water demand;
- medium-density housing hot water demand, and
- low-density housing hot water demand.

The hot water consumption profiles were generated at 30 minute sampling periods and then linearly interpolated to 1 minute sampling periods.

5.2.1.3 Inlet water temperature

A set of inlet water temperature data was generated at a 30 minute sampling period and from this a second set of inlet water temperature was generated with a sampling period of 1 minute by linearly interpolating the 30 minute data.

5.2.2 Description of the simulated demand's statistical analysis

Figure 5.2.2 portrays the sampling periods of the input variables that were used in the four sets of simulations for the three different hot water consumption profiles. Each alphabetically labelled statistical comparison (refer to Figure 5.2.2 and the list below) of the simulated output data investigates the following:

A. combined effect of using 1 or 30 minutes sampling period input data and the averaging process of the temperature data on the simulated demand;

B. effect of linear interpolation of the ambient temperature data from 60 minute original sampling period data on the simulated demand at a simulation sampling period of 30 minutes;

C. combined effect of using 1 or 30 minutes sampling period input data and the linear interpolating process of the temperature data on the simulated demand, and

D. effect of linear interpolation of the ambient temperature data from 60 minute original sampling period data on the simulated demand at a simulation sampling period of 1 minute.

Ambient temperature: 1 min sampling period; interpolated from 10 min data.
Hot water demand: 1 min sampling period data.
Inlet water temperature: 1 min sampling period data.

A

Ambient temperature: 30 min sampling period; averaged from 10 min data.
Hot water demand: 30 min sampling period data.
Inlet water temperature: 30 min sampling period data.

D

A four way statistical comparison of simulated demand data using the described input data sets

B

Ambient temperature: 1 min sampling period; interpolated from 60 min data.
Hot water demand: 1 min sampling period data.
Inlet water temperature: 1 min sampling period data.

C

Ambient temperature: 30 min sampling period; interpolated from 60 min data.
Hot water demand: 30 min sampling period data.
Inlet water temperature: 30 min sampling period data.

Figure 5.2.2: Four way statistical comparison of the geyser simulations using 1 and 30 minute sampling period data.

115

5.2.3 Comparison of the results of the geyser simulations at 1 and 30 minute input variable sampling periods

The first set of simulations made use of ambient temperature data that had an original sampling period of 10 minutes. The second set of simulations made use of ambient temperature data that was linearly interpolated from a 60 minute sampling period.

5.2.3.1 Simulations generated by ambient temperature data with an original sampling period of 10 minutes

This section provides a description of the statistical comparison **A** in Figure 5.2.2. Six simulations were run using three different hot water consumption profiles. Three of these simulations made use of 1 minute sampling period data and the remaining three made use of the data with a 30 minute sampling period. The output of the simulations was in the form of a vector of 30 minute averaged electrical demand in kWh. The output vectors of the 1 minute and 30 minute sampling period simulations are compared by correlation coefficient and covariance matrix. The results of the six simulations are recorded in Table 5.2.1

Table 5.2.1: Comparison of the correlation coefficient and covariance of simulated demand results using input data at a sampling period of 1 minute and 30 minutes for high-, medium- and low-density housing with an original ambient temperature sampling period of 10 minutes.

Housing density	kWh (1 min data)	kWh (30 min data)	Corr. Coeff.	Cov. [1,1]	Cov. [1,2]	Cov. [2,1]	Cov. [2,2]
High	3525.919	3525.269	0.620	0.190	0.118	0.118	0.191
Medium	4865.042	4864.517	0.628	0.248	0.156	0.156	0.248
Low	5901.915	5901.321	0.660	0.290	0.191	0.191	0.290

The greatest percentage difference in electrical demand between the 1 and 30 minute sampling period simulations was 0.018 percent and this was observed for the high-density housing group. The correlation coefficient of the electrical demand vectors varied from 0.628 to 0.660 for the different housing density groups. There was a greater variance in each of the demand vectors themselves than the covariance between the vectors.   This variance and covariance was determined from the covariance matrix. This set of simulations showed that the resulting

116

demand vectors were not closely correlated. The difference in total kWh demand between the different sampling period simulations was very small.

5.2.3.2   Simulations that made use of ambient temperature with an original sampling period of 60 minutes

This provides a description of the statistical comparison **C** in Figure 5.2.2. Six simulations were again run using three different hot water consumption profiles. Three of these simulations made use of 1 minute sampling period data and the other three made use of the data with a 30 minute sampling period. The results of the simulations are recorded in Table 5.2.2.

Table 5.2.2: Comparison of the correlation coefficient and covariance of simulated demand results using input data at a sampling period of 1 minute and 30 minutes for high-, medium- and low-density housing with an original ambient temperature sampling period of 60 minutes.

Housing density	kWh (1 min data)	kWh (30 min data)	Corr. Coeff.	Cov. [1,1]	Cov. [1,2]	Cov. [2,1]	Cov [2,2]
High	3525.872	3525.294	0.673	0.190	0.128	0.128	0.190
Medium	4864.396	4864.074	0.613	0.247	0.152	0.152	0.248
Low	5901.304	5900.552	0.655	0.290	0.190	0.190	0.290

The greatest percentage difference in electrical demand between the 1 and 30 minute sampling period simulations was 0.013 percent and this was observed for the low-density housing group. The correlation coefficient of the electrical demand vectors varied from 0.613 to 0.673 for the different housing density groups. There was a greater variance in each of the demand vectors themselves than the covariance between the vectors. This variance and covariance was determined from the covariance matrix. This set of simulations showed that the resulting demand vectors were not closely correlated. The difference in total kWh demand between the different sampling period simulations was again very small.

5.2.4   Comparison of the results of the geyser simulations that made use of ambient temperature data sets with 10 and 60 minute original sampling periods

The first set of simulations made use of ambient temperature data with an original sampling period of 10 minutes. The second set of simulations made use of ambient temperature data that

117

was linearly interpolated from a 60 minute sampling period. In the next subsection the simulations that used 30 minute sampling periods are compared, followed by a comparison of the simulations that made use of 1 minute sampling periods.

5.2.4.1 Simulations that made use of input variables with 30 minute sampling periods

This section provides a description of the statistical comparison **B** in Figure 5.2.2. The simulations that made use of input variables with 30 minute sampling periods were statistically compared and the results are presented in Table 5.2.3.

Table 5.2.3: Comparison of the correlation coefficient and covariance of simulated demand results using input data at a sampling period of 30 minutes for high-, medium- and low-density housing with an original ambient temperature sampling period of 10 and 60 minutes.

Housing density	kWh (10 min data)	kWh (60 min data)	Corr. Coeff.	Cov. [1,1]	Cov. [1,2]	Cov. [2,1]	Cov. [2,2]
High	3525.269	3525.294	0.705	0.191	0.134	0.134	0.190
Medium	4864.517	4864.074	0.790	0.248	0.196	0.196	0.248
Low	5901.321	5900.552	0.875	0.290	0.254	0.254	0.290

The greatest percentage difference in electrical demand between the simulations that made use of the 10 versus 60 minute original sampling period data was 0.013 percent and this was observed for the low-density housing group. The correlation coefficient of the electrical demand vectors varied from 0.705 to 0.875 for the different housing density groups. There was a greater variance in each of the demand vectors themselves than the covariance between the vectors. This set of simulations showed that the resulting demand vectors were more closely correlated than the demand vectors of the simulation sets in Tables 5.2.1 and 5.2.2. The difference in total kWh demand between the different sampling period simulations remained very small.

5.2.4.2 Simulations that made use of input variables with 1 minute sampling periods

This section describes the statistical comparison **D** in Figure 5.2.2. The simulations that made use of input variables with 1 minute sampling periods were statistically compared and the results are presented in Table 5.2.4.

Table 5.2.4: Comparison of the correlation coefficient and covariance of simulated demand results using input data at a sampling period of 1 minute for high-, medium- and low-density housing with an original ambient temperature sampling period of 10 and 60 minutes.

Housing density	kWh (10 min data)	kWh (60 min data)	Corr. Coeff.	Cov. [1,1]	Cov. [1,2]	Cov. [2,1]	Cov. [2,2]
High	3525.919	3525.872	0.599	0.190	0.114	0.114	0.190
Medium	4865.042	4864.396	0.796	0.248	0.197	0.197	0.247
Low	5901.915	5901.304	0.846	0.290	0.245	0.245	0.290

The greatest percentage difference in electrical demand between the simulations that made use of the 10 versus 60 minute original sampling period data was 0.013 percent and this was observed for the medium-density housing group. The correlation coefficient of the electrical demand vectors varied from 0.599 to 0.846 for the different housing density groups. There was a greater variance in each of the demand vectors themselves than the covariance between the vectors. This set of simulations showed that the resulting demand vectors were more closely correlated for the medium- and high-density groups compared to the demand vectors of the same housing groups in Tables 5.2.1 and 5.2.2. The difference in total kWh demand between the different sampling period simulations was again very small.

5.2.5   Conclusions drawn from the sensitivity analysis

Firstly, the kWh demand of the simulations is considered. The **A** comparison resulted in the greatest difference at 0.018 versus 0.013 percent for comparisons **B**, **C** and **D**. This difference was very small for a simulation period of one year. No trend could be seen when comparing the results of the different housing densities.

The correlation coefficient of the comparisons of the simulations is considered next. The correlation coefficients for comparisons **B** and **D** increased similarly with the increase in electrical demand, which in turn was dependent on hot water consumption rate. The correlation coefficients for comparisons **A** and **C** showed little variation in magnitude and no note worthy dependency on housing density. The correlation coefficients in the comparisons ranged from 0.599 to 0.875. From these observations it is concluded that the correlation coefficient was more

119

dependent on housing density (hot water demand) if the comparison of the 1 and 30 minute input sampling period were chosen. The correlation coefficients were similar in magnitude for the simulation that had the same input sampling period but had been manipulated by the processes of linier interpolation and averaging. It is deduced that the correlation coefficient is an indication of the time difference between the geyser's switching time intervals of the simulations. This deduction is based on the fact that the total kWh demands between the simulations varied by a small amount and that the correlation coefficient could then only be dependent on the switching time of the geyser.

Finally, the variance and covariance of the simulation comparisons are considered. For the comparisons of simulated demands of similar housing densities it can be seen that the variances and covariance remained similar. The variance and covariance increased as the hot water demand increased and this could be seen when the simulated demands of the three housing densities were compared. No further deduction could be made or patterns observed with regard to variance and covariance. The inclusion of the analysis of variance and covariance did not prove to be very informative for this set of simulations, but the simulation of a all-recorded input variable data set may lead to new conclusions been established with respect to input data sampling period and the resultant effect on simulated demand. It is therefore suggested that the analysis of the variance for different input variable sampling rates be included in the methodology to determine the effect it has on simulated demand results.

The total simulated demand was not adversely affected by input variable sampling period. The correlation coefficient of the simulation was dependent on hot water demand profile and the input variable sampling period. These results should not be considered conclusive in their own right because two of the three input variables to the geyser model were generated rather than recorded values. The conclusions that would be drawn from applying the same methodology on purely recorded input data may be different from these presented here. The value of this investigation lies in the methodology that was developed, to evaluate the effects of input variable sampling period and the data manipulation processes on the simulated demand. It is therefore recommended in future work, that the methodology presented in this chapter be followed for testing the simulated demand sensitivity to input sampling period and data manipulation for recorded data.

5.3   Sensitivity analysis of the geyser model's simulated demand with respect to the mean and variance of ambient temperature

An objective was set to determine the sensitivity of the simulated demand of the geyser model to ambient temperature. In Chapter four the annual mean and variance of the CTIA data was compared. The years with the extreme mean and variance ambient temperature data were selected to compare their simulated demand data. The 2002, 2007 and 2009 simulated demand were compared to determine if there would be any adverse effects on the statistical properties of the simulated demand if the annual demands were grouped together into seasons only, as opposed to seasons and years, for the statistical analysis. The mean and variance of 2002, 2007 and 2009 of the CTIA data are listed in Table 5.3.1.

Table 5.3.1: Mean and variance of the CTIA data for 2002, 2007 and 2009.

Parameter	2002	2007	2009
Mean [degrees Celsius]	16.2	16.6	16.9
Variance [degrees Celsius]	26.5	26.5	23

The simulated demand data of the three different housing densities for the chosen years are presented in Table 5.3.2.

Table 5.3.2: Simulated demand for the three housing densities for the years 2002, 2007 and 2009.

Housing density	kWh demand 2002	kWh demand 2007	kWh demand 2009
High	3689	3664	3641
Medium	5079	5044	5049
Low	6122	6136	6106

The percentage difference from the lowest simulated for each evaluation year demand in each housing density is given in Table 5.3.3.

Table 5.3.3: Percentage difference from the lowest simulated demand for each evaluation year for each housing density category.

Housing density	% demand diff. 2002	% demand diff. 2007	% demand diff. 2009
High	1.31	0.63	0
Medium	0.69	0	0.01
Low	0.26	0.49	0

Of the three years, 2002 had the lowest mean temperature. For high-density housing, which represent the lowest hot water demand profile of the three housing densities, the electrical demand increased with a decrease in mean temperature. A similar trend can be seen for the low-density housing hot water demand category but to a lesser degree compared to the high-density category.

The percentage difference in simulated demand calculated in Table 5.3.3 is small, and it would seem reasonable to group the years of simulated demand together without incurring a significant error in the statistical parameters. However it must be pointed out that the true relationship between simulated demand and ambient temperature is not represented in totality because the hot water demand and inlet water temperature were *generated* rather than *recorded* along with the ambient temperature. It will be seen in the next two subsections that of the three input variables to the geyser model, hot water demand has the greatest influence on electrical demand and it is therefore necessary to have recorded input variables where the underlying interdependency is present when attempting such a sensitivity analysis. However, significant for this research was the testing of the sensitivity of ambient temperature and the testing procedure that was followed.

5.4 Simulated electrical demand impacts that result from geyser ECMs

5.4.1 Overview

Three geyser ECMs were simulated as part of the case study of the geyser. In these ECM simulations two geyser models were run in parallel to calculate the resultant electrical demand of each model and the demand impact. The one model simulated the pre-ECM demand profile and

the other model the post ECM demand profile. The standing loss factors of both models were set in accordance with the Bosman et al. findings for geyser standing losses [36].

The first ECM that was simulated modelled the adjustment of the geyser thermostat's upper temperature limit down from the standard 65 to 55 degrees Celsius. The hysteresis relay of the post ECM's model was adjusted accordingly. The first simulation recorded the standing losses that resulted from zero hot water demand. Subsequent simulation made use of the three housing densities' characteristic hot water demand profiles.

The second geyser ECM that was simulated, modelled the reduction of hot water consumption that would result from the retrofitting of low flow shower roses in a household. The shower's mixed warm water demand was reduced in accordance with the claimed flow reduction figures.

The final geyser ECM simulation investigated the demand impact that would result from modified inlet water temperature profiles. The modified inlet water temperature profile could be realised by storing the inlet water to the geyser at various depths below the ground surface.

All three ECM simulations made use of the same CTIA ambient temperature data and the three characteristic hot water consumption profiles of the different housing densities. This allowed the evaluation of the ECMs' impacts for different characteristic hot water consumption profiles.

5.4.2 No load losses of a geyser with the thermostat adjusted from 65 to 55 degrees Celsius

Initially, a description is given of how the pre- and post-ECM models are adjusted and linked during the simulations of sections 5.4.2. and 5.4.3. The upper and lower limits of the hysteresis relay of the post ECM's geyser model were reduced by 10 degrees Celsius from standard factory geyser settings. The upper and lower limit had a factory setting of 65 and 55 degrees Celsius respectively. Both models made use of the same ambient and inlet water temperature data sets. In the simulations of section 5.4.2 the hot water demand of the pre-ECM model was set to zero (i.e. no load). This resulted in a zero warm and hot water demand for the post ECM model. In the simulations of section 5.4.3, the pre-ECM model made use of the three characteristic hot water consumption profiles of the different housing densities. This pre-ECM model also calculated the warm water demand for the post ECM geyser model at a mixing temperature of 39 degrees Celsius. It was necessary to calculate this warm water demand because it was the third

input variable to the post ECM model. Lowering the set point water temperature inside the post ECM's geyser would result in an increase in hot water demand thus maintaining the same warm water mixing temperature. The warm water temperature and volume is related to the hot water temperature and volume by equations 5.4.1 and 5.4.2.

$$T_{warm} \times f_{warm} \times C = \left(T_{cold} \times f_{cold} \times C\right) + \left(T_{hot} \times f_{hot} \times C\right) \qquad \textbf{5.4.1}$$

$$f_{warm} = f_{cold} + f_{hot} \qquad \textbf{5.4.2}$$

where

T_{warm} denotes warm water temperature demand in degrees Celsius.

f_{warm} denotes the flow rate of the warm water demand for the pre- and-post ECM model in litres per second;

f_{hot} denotes the hot water demand flow rate for the geyser pre-ECM model in litres per second;

T_{hot} denotes the temperature of the water inside the geyser pre-ECM model in degrees Celsius, and

T_{cold} denotes the inlet water temperature to the pre-ECM model in degrees Celsius.

C denotes the heat capacity of water in joules per kilogram per degrees Kelvin;

The solving of these two equations simultaneously results in equation 5.4.3 which was used to calculate the warm water demand of the pre- and post-ECM geyser model.

$$f_{warm} = f_{hot} \frac{\left(T_{hot} - T_{cold}\right)}{\left(T_{warm} - T_{cold}\right)} \qquad \textbf{5.4.3}$$

The post ECM model calculated the post ECMs hot water demand from the warm water demand of the pre-ECM model. This was done by using equation 5.4.4.

$$f_{hot} = f_{warm} \frac{\left(T_{warm} - T_{cold}\right)}{\left(T_{hot} - T_{cold}\right)} \qquad \textbf{5.4.4}$$

where

f_{hot} denotes the hot water demand flow rate for the post ECM model in litres per second;

f_{warm} denotes the flow rate of the warm water demand for the post ECM model in litres per second;

T_{hot} denotes the temperature of the water inside the post ECM model in degrees Celsius, and

T_{cold} denotes the inlet water temperature to the post ECM model in degrees Celsius.

5.4.2.1   Demand impact table

The first simulation with this parallel geyser model arrangement investigated the standing losses of the pre- and post-ECM periods. This ECM was simulated by allocating zero hot water demand to the pre-ECM model which resulted in zero warm water demand allocation for the post ECM's model. The geyser had a simulated volume of 150 litres and a heater element rating of 3 kW. The electrical demand that resulted from standing losses alone for the pre- and post-ECM periods are recorded in Table 5.4.1. This table also records the differences between the demands of the pre- and post-ECM and periods.

Table 5.4.1: Demand impact in kWh that resulted from the standing losses and the adjustment of the geyser's thermostat for the pre- and post-ECM simulation for the period 2001 to 2009.

Demand impact [kWh]									
Model	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Pre-ECM	729	732	727	724	729	727	724	726	720
Post ECM	557	561	559	553	559	555	555	557	550
Difference between	172	171	168	171	170	172	169	169	170

The annual demands that resulted from the standing losses compare favourably with the findings of SABS report [47]. The SABS report claimed an equivalent standing loss of 660.65 kWh per year for a 150 litre geyser that was subjected to a controlled environment with an ambient temperature and an internal geyser temperature of 20 and 69.5 degrees Celsius respectively. The annual demand difference between the pre- and post-ECM simulations also compared favourably with the findings of the Bosman et al. study. The study calculated that the annual difference in standing loss demand would be 168.28 kWh for a 100 litre geyser in the Cape Town region [36].

The next point of interest to investigate was the demand dependency on annual median and variance of ambient temperature. The annual medians and variances of the ambient temperature data set used in the demand simulations are recorded in Tables 5.4.2 and 5.4.3 respectively.

Table 5.4.2: Annual medians of ambient temperature of the CTIA data for the period 2001 to 2009.

Annual medians of ambient temperature [degrees Celsius]								
2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
16.4	16.2	16.7	16.8	16.4	16.6	16.7	16.6	16.9

Table 5.4.3: Annual variances of ambient temperature of the CTIA data for the period 2001 to 2009.

Annual variances of ambient temperature [degrees Celsius]								
2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
23.4	26.6	27.2	25.7	25.8	23.9	26.6	23.7	23.00

The resultant correlation between the demand data and the annual ambient temperature statistical parameters are summarised as follows:

- correlation coefficient of the annual demand versus ambient temperature medians for the pre- and post-ECM periods were -0.95 and -0.88 respectively;
- correlation coefficient of the annual demand versus ambient temperature variances for the pre- and post-ECM periods were 0.34 and 0.54 respectively, and

- correlation coefficient of the annual demand difference (between pre- and post-ECMs) versus ambient temperature median and variance were -0.29 and -0.46 respectively.

These results show that the median of the ambient temperature for an evaluation period has a strong influence on the overall electrical demand of the geyser that results from the no load losses. This result is consistent with the fact that the lower the ambient temperature experienced, the greater the resultant standing losses would be because of the greater temperature difference between ambient air and water temperature within the geyser. The post ECM's variance relative to annual demand was more positively correlated than that of the pre-ECM. This implies that for an increase in the variance of ambient temperature there will be a corresponding increase in the variance of electrical demand of the geyser. The stronger negative correlation between demand difference and the variance of the ambient temperature suggests the variance has greater negative influence on the demand difference of the simulated ECM compared to the median of the ambient temperature.

5.4.2.2   Image plots of demand data

The image plots of the electrical demand data of the pre- and post-ECM simulations of the year 2009 are depicted in Figures 5.4.1 and 5.4.2 respectively.
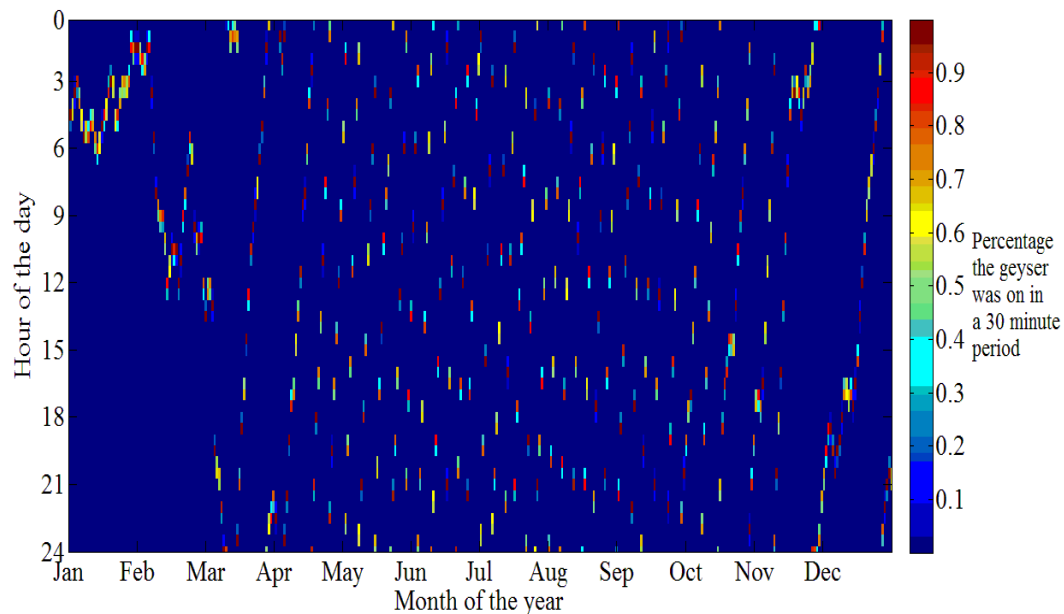
Figure 5.4.1: Image plot of the standing losses of the pre-ECM simulation for the year 2009.

127
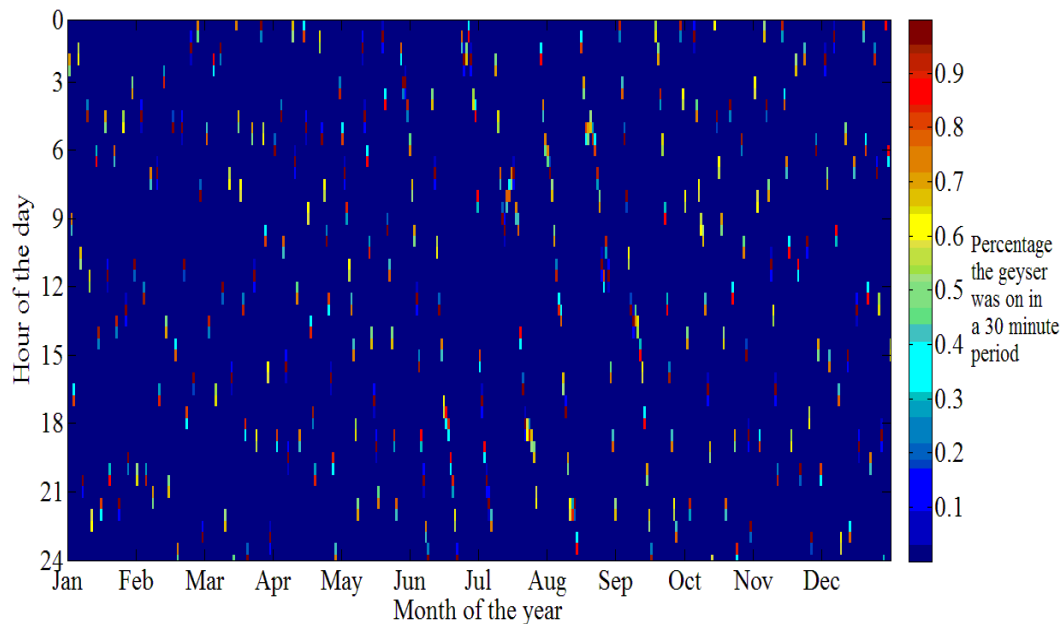
Figure 5.4.2: Image plot of the standing losses of the post ECM simulation for the year 2009.

It is evident from the summer months of the pre-ECM simulation in Figure 5.4.1, that the geyser was switching *on* approximately once a day and in the winter months this period was reduced to approximately every 21 hours. As shown in Figure 5.4.2, upon implementation of the ECM, the geyser was switching *on* approximately once a day in the winter months and during the summer months, this period was increased to approximately once every 35 hours. It is also evident from these two image plots that the geyser took about 35 to 40 minutes to recover from the low to the high temperature set points.

5.4.2.3   Probability density estimation of the recovery time of the geyser

Plots of the density estimates were generated for ambient temperature versus the time taken by the geyser to recover from the low to the high point temperature settings. The density estimation of the pre- and post-ECM simulations of the recovery times of the geyser are depicted in Figure 5.4.3.

The probability density estimates of the recovery time of the geyser shows its inversely proportional to ambient temperature. It was calculated that the difference between the medians' of the recovery times of the pre- and post-ECM simulations was 0.233 minutes.

The recovery time of the geyser is a directly proportional to its energy use in kWh. The recovery time of the geyser is dependent on the volume of water inside the geyser and the kilo watt rating of the geyser element. The high-density estimates of the geyser recovery times over a small temperature ranges are characteristic of a geyser with no hot water demand.
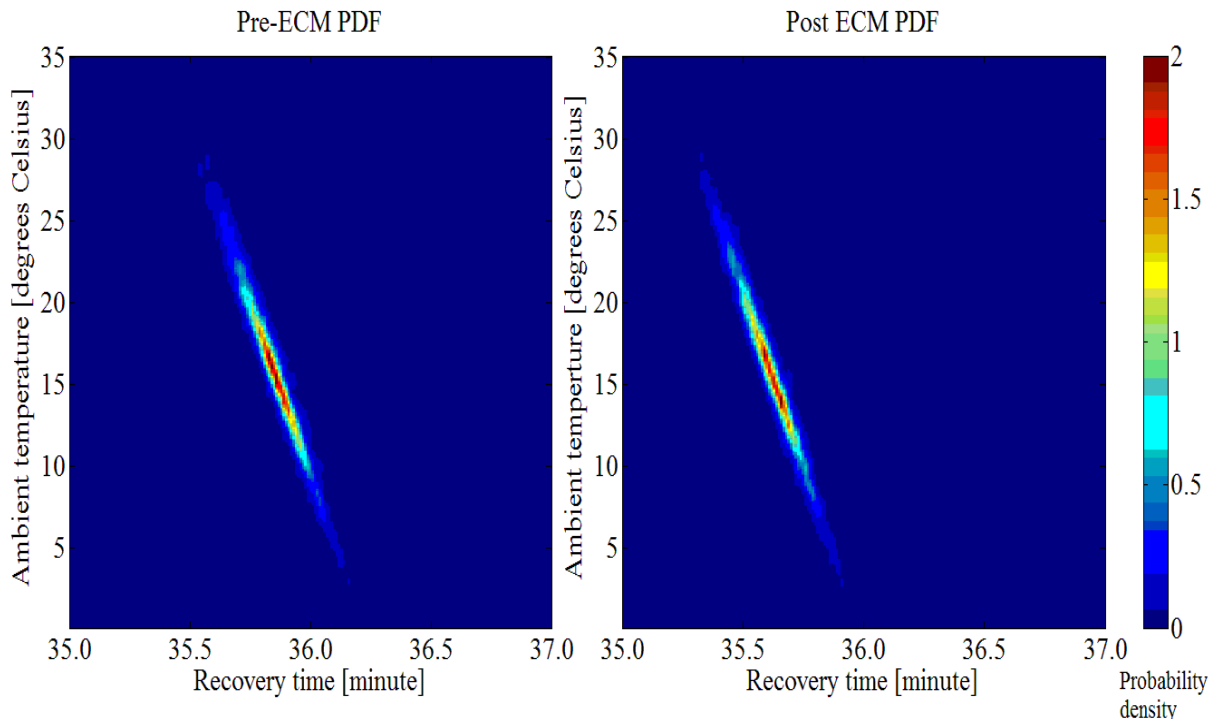
Figure 5.4.3: Probability density plots for the pre- and post-ECM recovery times of the geyser for the simulation period 2001 to 2009.

5.4.2.4   Probability density plots of ambient temperature versus time of day for geyser *on* times

The simulated electrical demand dependency of the geyser models on ambient temperature is subsequently investigated. The dependency of electrical demand versus temperature is depicted by means of a three dimensional mesh and a two dimensional image plot of the probability density of the ambient temperature versus time of day when the geyser was switched *on*. This investigation is conducted for both the pre- and post-ECM simulations. The plots of the PDFs of the pre- and post-ECM simulated data for the summer months are depicted in Figures 5.4.4 to 5.4.6.
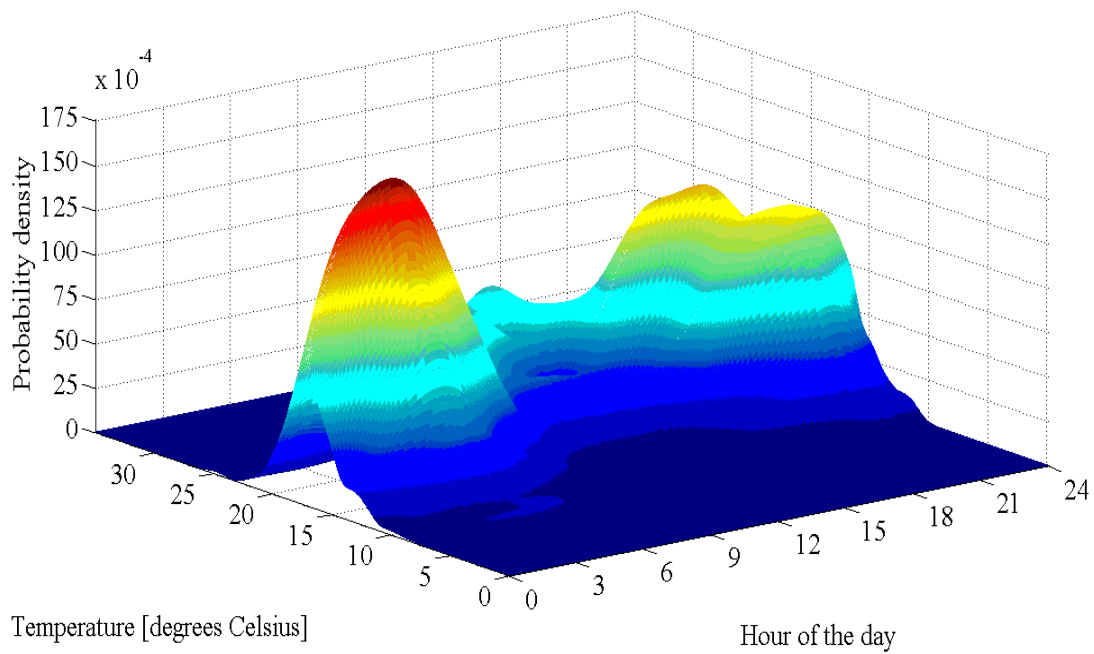
Figure 5.4.4: Probability density plot of the pre-ECM PDF of the ambient temperature versus time of day the geyser was *on* during the summer months from 2001 to 2009.
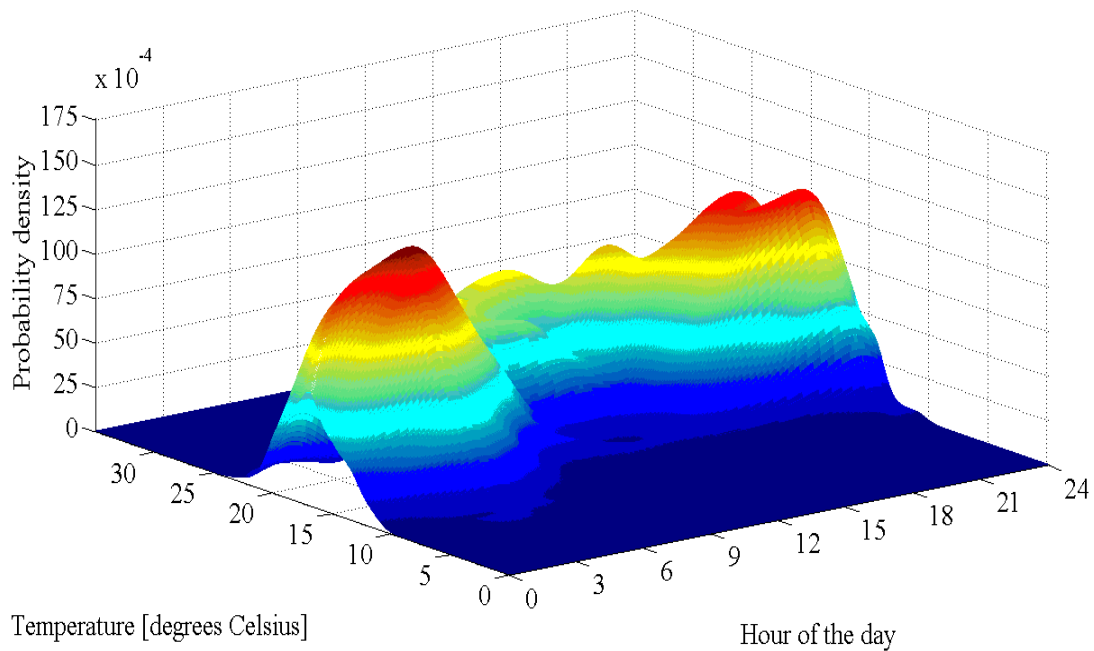
Figure 5.4.5: Probability density plot of the post ECM PDF of the ambient temperature versus time of day the geyser was *on* during the summer months from 2001 to 2009.
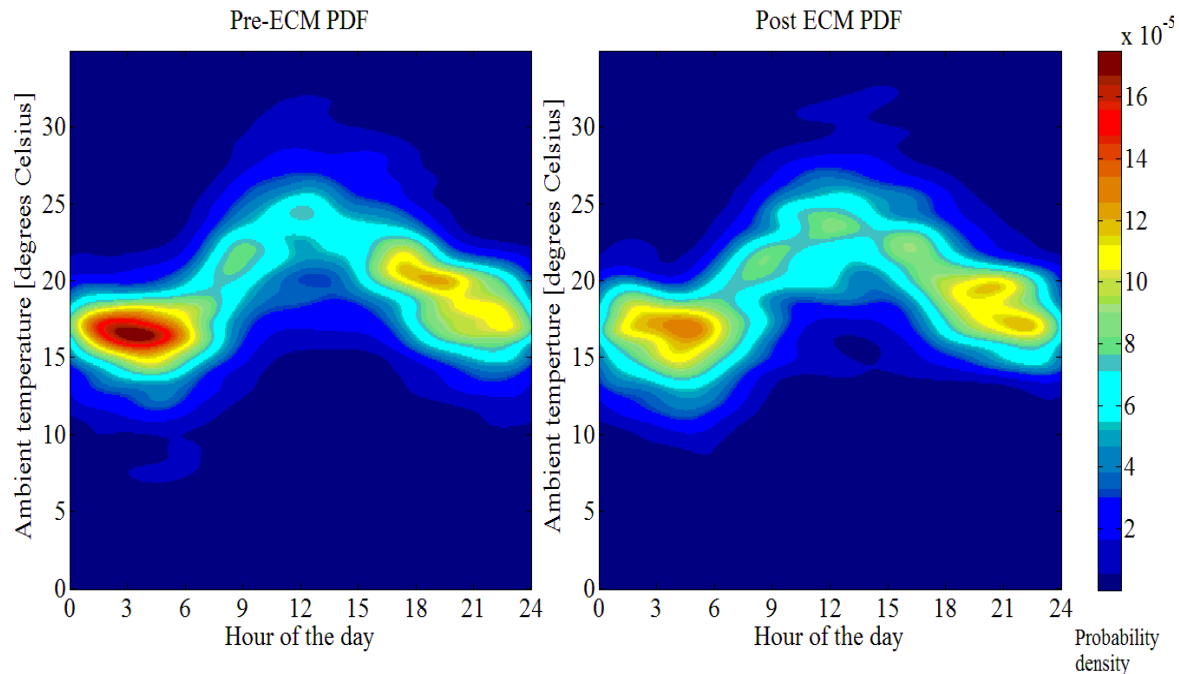
Figure 5.4.6: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the summer months from 2001 to 2009.

The following is evident when the probability density plots of the pre- and post-ECM are compared:

- These plots show that the electrical demand probability density is characteristically bimodal for the pre- and post-ECM simulations. This can be attributed to the diurnal ambient temperature fluctuations and higher summer average temperatures which resulted in lower standing losses, allowing the geyser to switch *on* less frequently over the midday period.

- In the pre-ECM simulation, the probability density is greater for the geyser switched-*on* in the early hours of the morning. The reason for this is because the pre-ECM simulation had a higher set point temperature than the post ECM simulation, resulting in a greater temperature difference between the ambient temperature and temperature within the geyser for the pre-ECM simulation. These standing losses were greatest during the low ambient temperatures of the early hours of the morning.

- The reduction in the set point temperatures of the post ECM simulation resulted in the electrical demand density shifting to a later part of the morning as a result of the reduced standing losses.

131

The plots of the probability density of the pre- and post-ECM simulated data for the winter months are depicted in Figures 5.4.7 to 5.4.9.

Figure 5.4.7: Probability density plot of the pre-ECM PDF of the ambient temperature versus time of day the geyser was *on* during the winter months from 2001 to 2009.

Figure 5.4.8: Probability density plot of the post ECM PDF of the ambient temperature versus time of day the geyser was *on* during the winter months from 2001 to 2009.

132

Figure 5.4.9: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the winter months from 2001 to 2009.

The following is evident when the probability density plots of the pre- and post-ECM are compared:

- The higher density bimodal features from the summer months have substantially reduced. There is a higher probability of the geyser being *on* over a wider temperature range from 09h00 to 18h00 during the winter months.

- The most notable feature of the winter month probability density plots is the 'saddle' that existed at approximately midnight.

- An equivalent demand density occurred at a lower average temperature during the winter months compared to the summer months.

- The demand density variance across the temperature range was greater during the winter months for all hours of the day.

In summary the patterns of the electrical demand density plots are dependent on ambient temperature, the size of the element, the volume of water, and the standing loss factor of the geyser.

5.4.3   Adjustment of a geyser's thermostat

The pre- and post-ECM geyser models were simulated in parallel. The pre-ECM simulation calculated the warm water demand for the post ECM geyser model. The post ECM geyser model's thermostat set points were set 10 degrees Celsius lower than the pre-ECM geyser model.

5.4.3.1   Demand impact tables

The three characteristic housing densities' hot water consumption profiles were used in the simulated demand impacts that resulted from the thermostat adjustment of the post ECM's geyser model and are recorded in Tables 5.4.4 to 5.4.6 for the nine year simulation period.

Table 5.4.4: Electrical demand impact that resulted from the adjustment of the geyser's thermostat from 65 to 55 degrees Celsius with high-density housing characteristic hot water consumption profile for the simulation period 2001 to 2009.

Demand impact [kWh]									
Model	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Pre-ECM	3679	3689	3682	3666	3663	3669	3664	3660	3641
Post ECM	3511	3521	3515	3499	3495	3504	3497	3494	3474
Difference between	168	168	167	167	168	165	167	166	167

Table 5.4.5: Electrical demand impact that resulted from the adjustment of the geyser's thermostat from 65 to 55 degrees Celsius with medium-density housing characteristic hot water consumption profile for the simulation period 2001 to 2009.

Demand impact [kWh]									
Model	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Pre-ECM	5051	5079	5047	5057	5063	5063	5044	5053	5049
Post ECM	4883	4911	4880	4889	4893	4895	4877	4887	4882
Difference between	168	168	167	168	170	168	167	166	167

Table 5.4.6: Electrical demand impact that resulted from the adjustment of the geyser's thermostat from 65 to 55 degrees Celsius with low-density housing characteristic hot water consumption profile for the simulation period 2001 to 2009.

Demand impact [kWh]									
Model	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Pre-ECM	6136	6122	6143	6125	6150	6140	6136	6134	6106
Post ECM	5967	5954	5978	5957	5983	5973	5970	5965	5939
Difference between	169	168	165	168	167	167	166	169	167

From the demand impact tables it is noted that not only did the difference between the calculated demand of the pre- and post-ECM simulations remain virtually constant across the years of the evaluation period, but also across the three hot water consumption profiles. These results can be expected because the standing loss factor remained the same between the sets of simulations.

The difference in demand impact between the simulations that did and did not have hot water demand was on average 2.9kWh per year. This result can be explained by the fact that there would be a small reduction in standing losses because of the marginally lower average temperature in the geyser as a result of the hot water demand.

5.4.3.2   Image plots of demand data

To provide a succinct graphical explanation, only the medium-density housing hot water consumption simulations were analysed and plotted. Medium-density housing characteristic hot water consumption profile was chosen because it represented the average of the three hot water consumption profiles. In a detailed analysis using the methodology presented in this thesis it is strongly recommended than the three hot water consumption profiles' simulations be analysed in detail because hot water consumption has a significant relationship with electrical demand of the geyser. The image plots of the electrical demand data of the pre- and post-ECM simulations for the year 2009 are depicted in Figures 5.4.10 and 5.4.11 respectively.
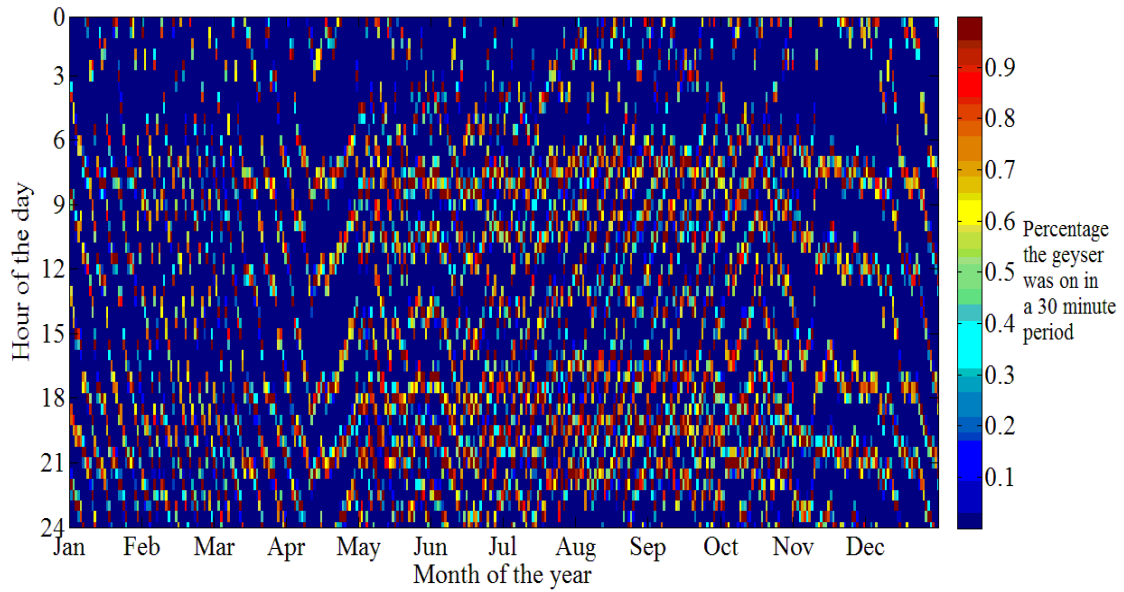
Figure 5.4.10: Image plot of the electrical demand data for the pre-ECM simulation for medium-density housing's characteristic hot water consumption for the year 2009.
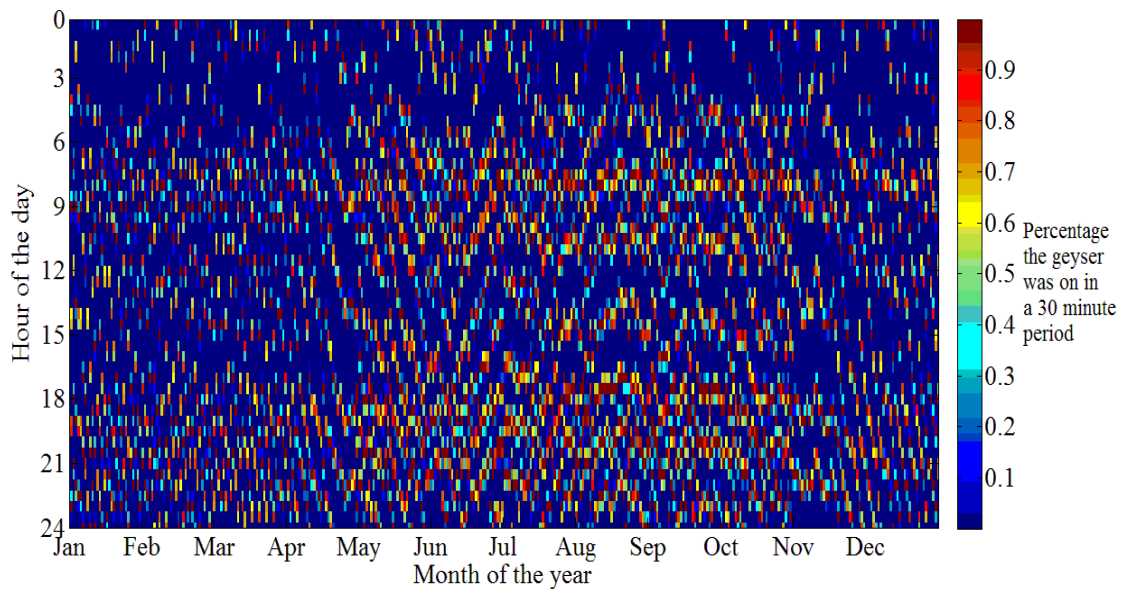
Figure 5.4.11: Image plot of the electrical demand data for the post ECM simulation for medium-density housing characteristic hot water consumption for the year 2009.

The higher density geyser *on* periods can be seen in the winter, compared to the summer months. There is also greater demand density surrounding the 09h00 and 19h00 time instances for both the pre- and post-ECM simulations. To draw any further conclusions from the image plots

136

becomes challenging. For this reason it becomes necessary to use statistical inference to analyse the demand data.

5.4.3.3   Probability density estimation of the recovery time of the geyser

Probability density plots were generated of the ambient temperature versus time taken by the geyser to recover from the low to the high point temperature settings. The density estimation of the pre- and post-ECM simulations of the recovery times of the geyser for the characteristic season and day type are depicted in Figures 5.4.12. to 5.4.15.
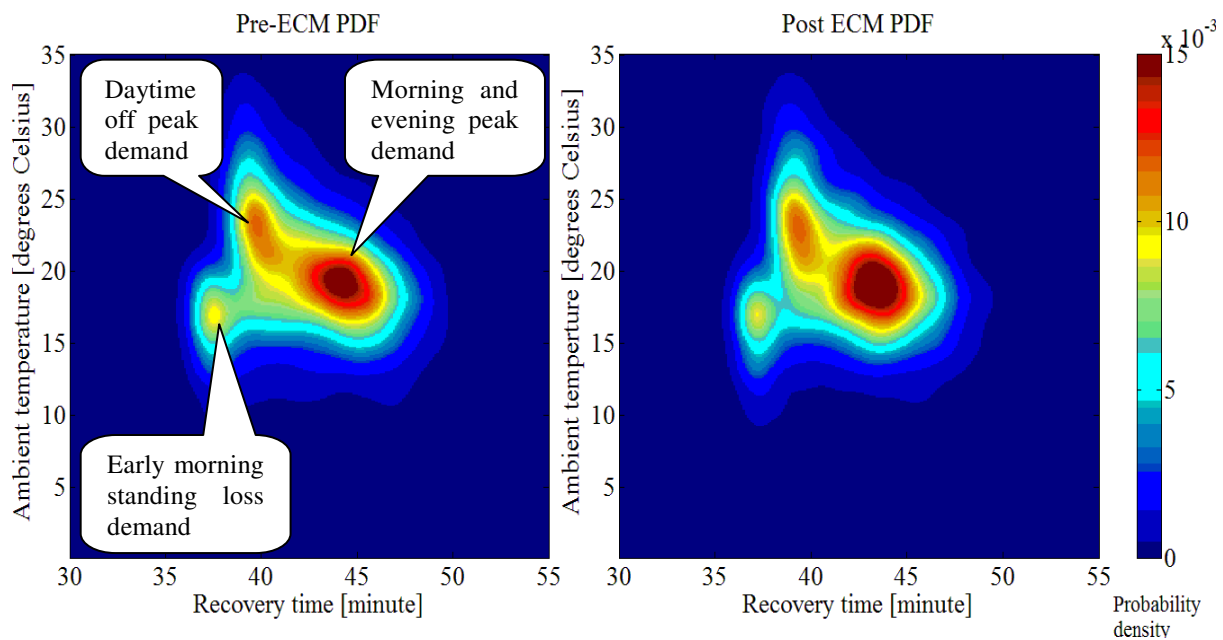
Figure 5.4.12: Probability density plots for the pre- and post-ECM recovery times of the geyser with a medium-density housing characteristic hot water demand profile for summer weekdays for the simulation period 2001 to 2009.

The density estimates of the geyser recovery times of the summer months have three distinct areas of interest. These areas are indicated in Figure 5.4.12. They are recognised by the temperature at which they occur. The density estimate of the morning and evening peak load during the weekend days is lower than that of the weekday. This observation is consistent with the observations made in the analysis of the hot water consumption. The weekday hot water consumption peak in is higher when compared to a similar time period for the weekend day. This pattern in the data is attributed to the fact that people don't have to keep the weekly schedule of getting to work at certain period of time and as a result of this the hot water demand

137

is spread over a larger time period in the weekend day demand profile. This results in similar patterns in the density estimates of recovery times and electrical demand of the geyser.
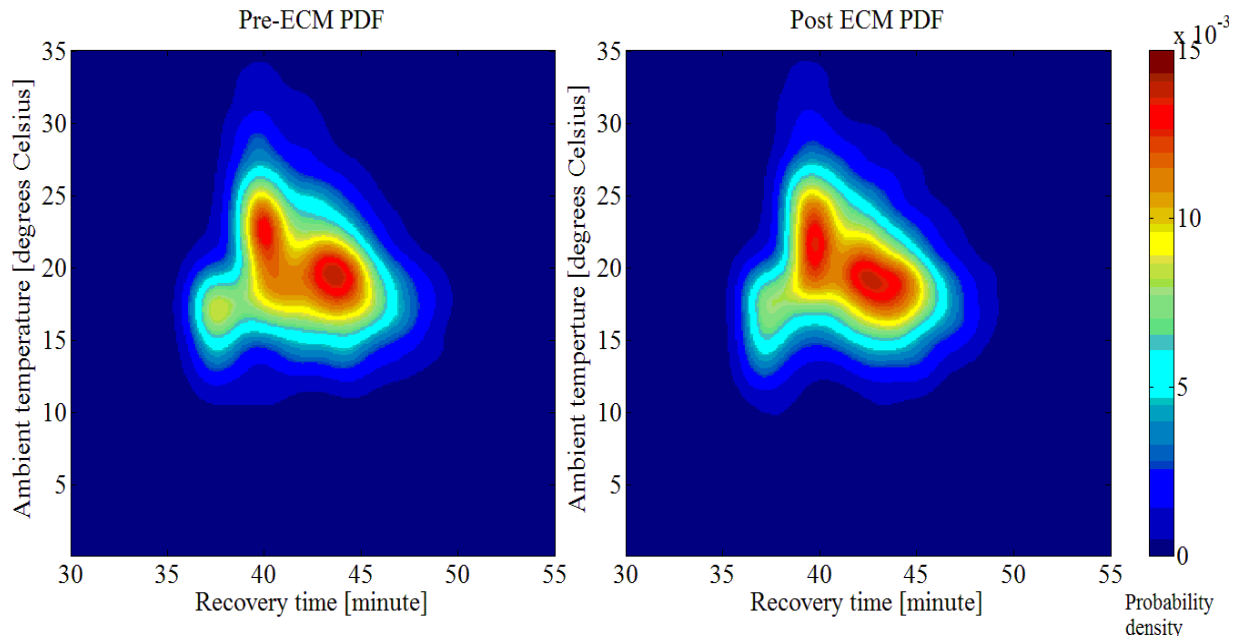
Figure 5.4.13: Probability density plots for the pre- and post-ECM recovery times of the geyser with a medium-density housing characteristic hot water demand profile for summer weekend days for the simulation period 2001 to 2009.
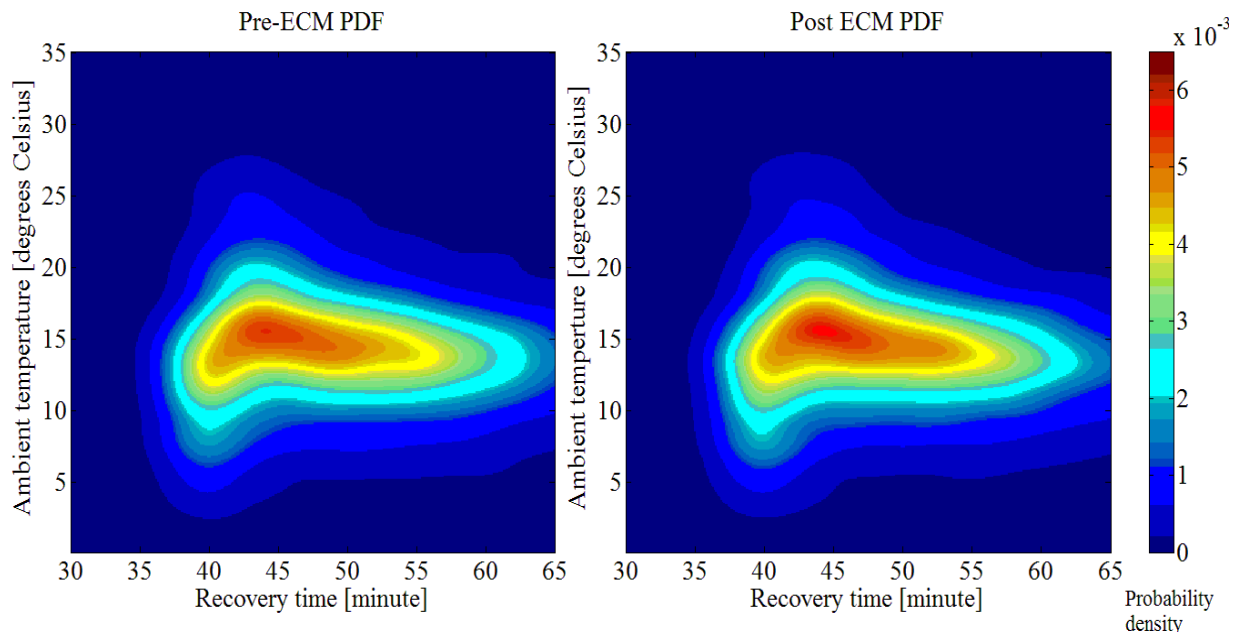
Figure 5.4.14: Probability density plots for the pre- and post-ECM recovery times of the geyser with a medium-density housing characteristic hot water demand profile for winter weekdays for the simulation period 2001 to 2009.
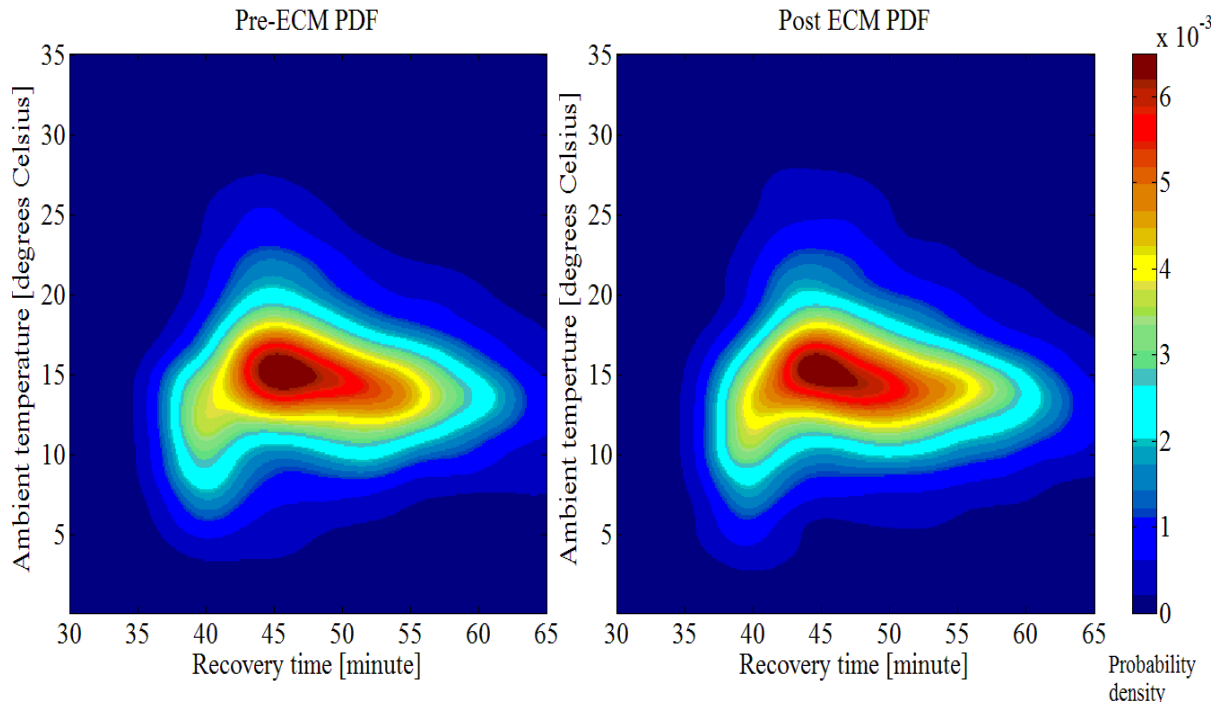
Figure 5.4.15: Probability density plots for the pre- and post-ECM recovery times of the geyser with a medium-density housing characteristic hot water demand profile for winter weekend days for the simulation period 2001 to 2009.

During the winter months the three peaks of density estimates of the summer simulation periods have reduced to a single mode. The ambient temperature at which the winter peak density occurs is less than that of the summer peak. In the summer and winter periods, the peak density increased by a small margin when comparing the post ECM to the pre-ECM. It is interesting to note that the peak density increased from the weekday to the weekend day density estimates during the winter months. This trend is directly opposite to the trend observed in the summer months. This could be explained by the reduce peak hot water consumption rates during the weekend days which results in shorter geyser recovery times.

5.4.3.4 Probability density plots of ambient temperature versus time of day for geyser *on* times

The probability density plots of the PDFs of the pre- and post-ECM simulations are depicted in Figures 5.4.16 and 5.4.17, respectively for the week and weekend days during the summer months.
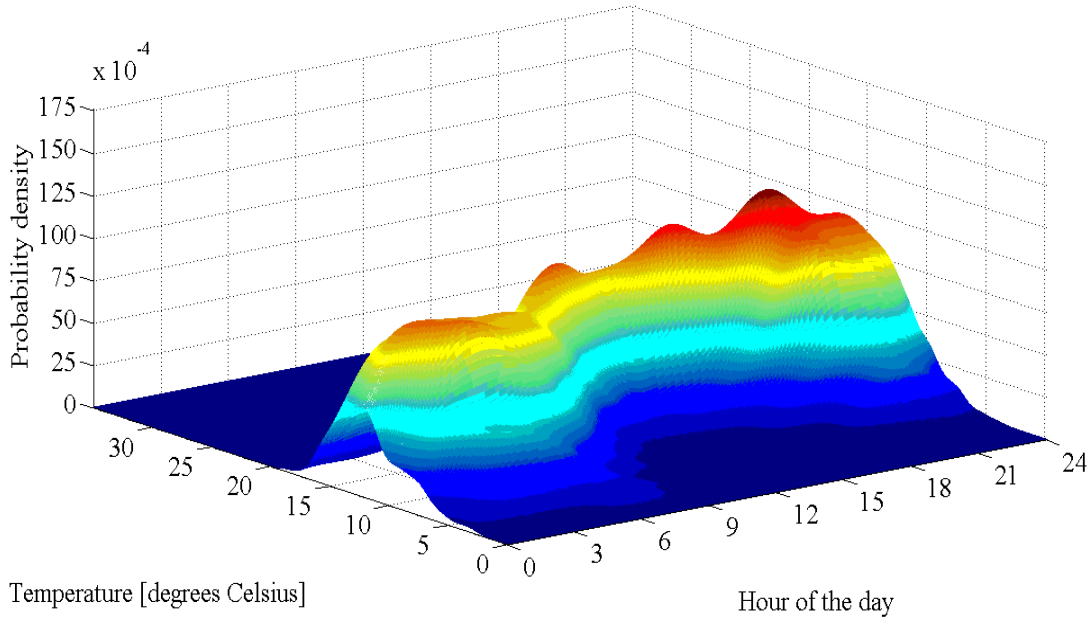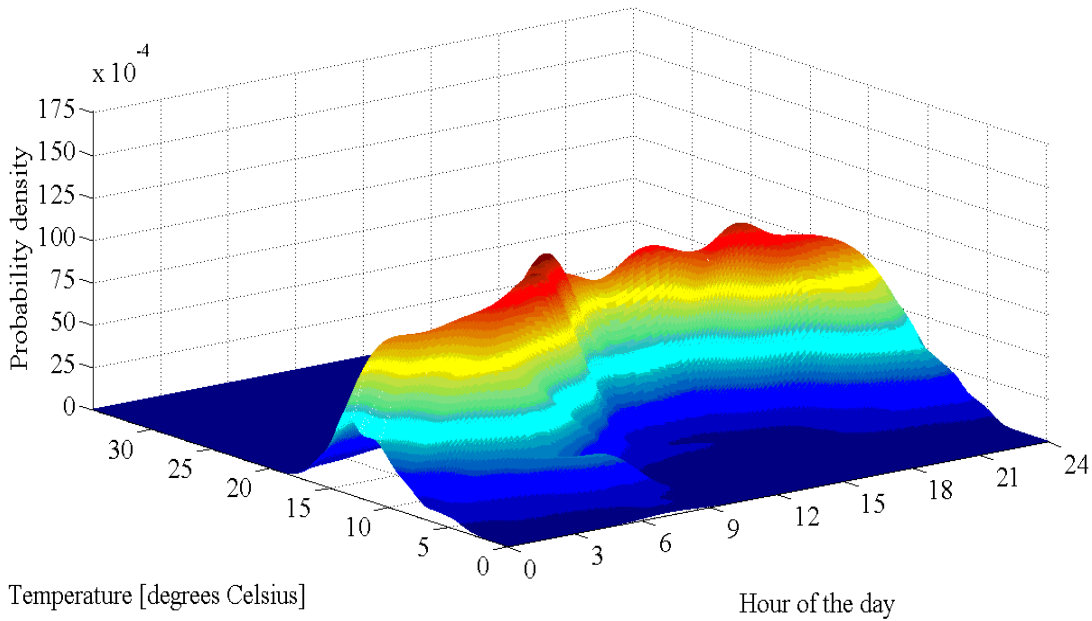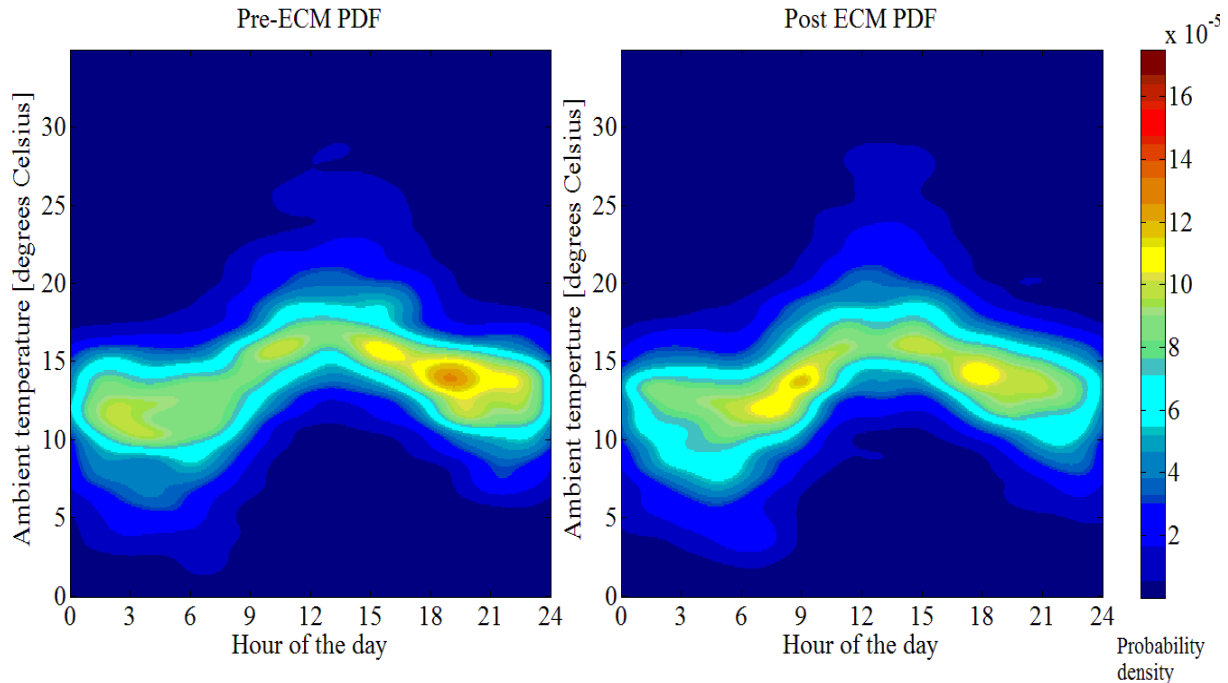
Figure 5.4.16: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the summer weekdays with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

Figure 5.4.17: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the summer weekend days with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

The following is evident when Figures 5.4.16 and 5.4.17 are compared:

- The maximum of the weekend modes are less than that of the weekday modes. The reason for this is the hot water demand during the weekend day was distributed more evenly over the given time period.
- From the weekday PDF plots it is evident that the ECM had reduced the overall demand density, especially in the first half of the day.
- The peak densities of the weekend PDFs occurred approximately an hour later when compared to the weekdays. This pattern in the demand results was influenced by the hot water consumption patterns.

The image plots of the PDFs for the pre- and post-ECMs simulated data are depicted in Figures 5.4.18 and 5.4.19, respectively for the week and weekend days during the winter months.

Figure 5.4.18: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the winter weekdays with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

Figure 5.4.19: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the winter weekend days with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

It was evident in Figures 5.4.18 and 5.4.19 that the demand peak densities are lower during the weekend days and occurred approximately an hour later compared to the weekdays. For both the week and weekend days of the pre- and post-ECM plots, the demand density was shifted later in time as a result of the ECM.

When the summer and winter PDF plots are compared it can seen that the evening density mode occurred approximately an hour earlier for the winter months. It was also interesting to note that there was a small mode at approximately 01h00 to 02h00 in the summer months that is dramatically reduced in the winter month plots.

5.4.4   Installation of low flow shower roses

In the study conducted by DeOreo et al, it was found that the showers were responsible for 25 percent of the domestic hot water demand [33]. Three independent sources claim that low flow shower roses reduce warm water demand on average by between 30 and 50 percent [48, 49 and 50]. The low flow simulation assumed that there would be a 40 percent reduction in warm water demand by the shower. The shower represents 25 percent of the total hot water demand in a household, resulting in a 10 percent overall reduction in hot water demand. The parallel models

142

that simulated the ECM where the geyser thermostat was adjusted from 65 to 55 degrees Celsius were modified by inserting a gain block in the warm water demand signal line to the post ECM geyser model. This gain block was set to reduce the warm water demand by 10 percent. This reduction represents the positive impact which low flow shower roses are claimed to have on reducing warm water demand.

5.4.4.1 Demand impact tables

The three characteristic housing densities' hot water consumption profiles were used in the simulated demand impacts that resulted from the retrofitting of low flow shower roses. The pre- and post-ECM simulated demands are recorded in Tables 5.4.7 to 5.4.9.

Table 5.4.7: Electrical demand impact that resulted from the installation of low flow shower roses with a high-density housing characteristic hot water consumption profile for the simulation period 2001 to 2009.

Demand impact [kWh]									
Model	2001	2002	2003	2004	2005	2006	2007	2008	2009
Pre-ECM	3679	3689	3682	3666	3663	3669	3664	3660	3641
Post ECM	3386	3393	3391	3373	3372	3377	3372	3370	3350
Difference between	293	296	291	293	291	292	292	290	291

Table 5.4.8: Electrical demand impact that resulted from the installation of low flow shower roses with a medium-density housing characteristic hot water consumption profile for the simulation period 2001 to 2009.

Demand impact [kWh]									
Model	2001	2002	2003	2004	2005	2006	2007	2008	2009
Pre-ECM	5051	5079	5047	5057	5063	5063	5044	5053	5049
Post ECM	4621	4647	4617	4624	4631	4630	4614	4622	4617
Difference between	430	432	430	433	432	433	430	431	432

143

Table 5.4.9: Electrical demand impact that resulted from the installation of low flow shower roses with a low-density housing characteristic hot water consumption profile for 2001 to 2009.

Demand impact [kWh]									
Model	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Pre-ECM	6136	6122	6143	6125	6150	6140	6136	6134	6106
Post ECM	5598	5585	5604	5587	5609	5601	5598	5593	5570
Difference between	538	537	539	538	541	539	538	541	536

From the demand impact tables it is noted that the difference between the calculated demand of the pre- and post-ECM simulations remained virtually constant across the years of evaluation for each characteristic hot water consumption profile. The percentage demand impact for each characteristic hot water consumption profile was similar, at an average of 8.5 percent.

5.4.4.2   Probability density plots of ambient temperature versus time of day for *on* times

The image plots of the PDFs for the pre- and post-ECM simulated data are depicted in Figures 5.4.20 and 5.4.21, respectively for the summer week and weekend days.

Figure 5.4.20: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the summer weekdays with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

144
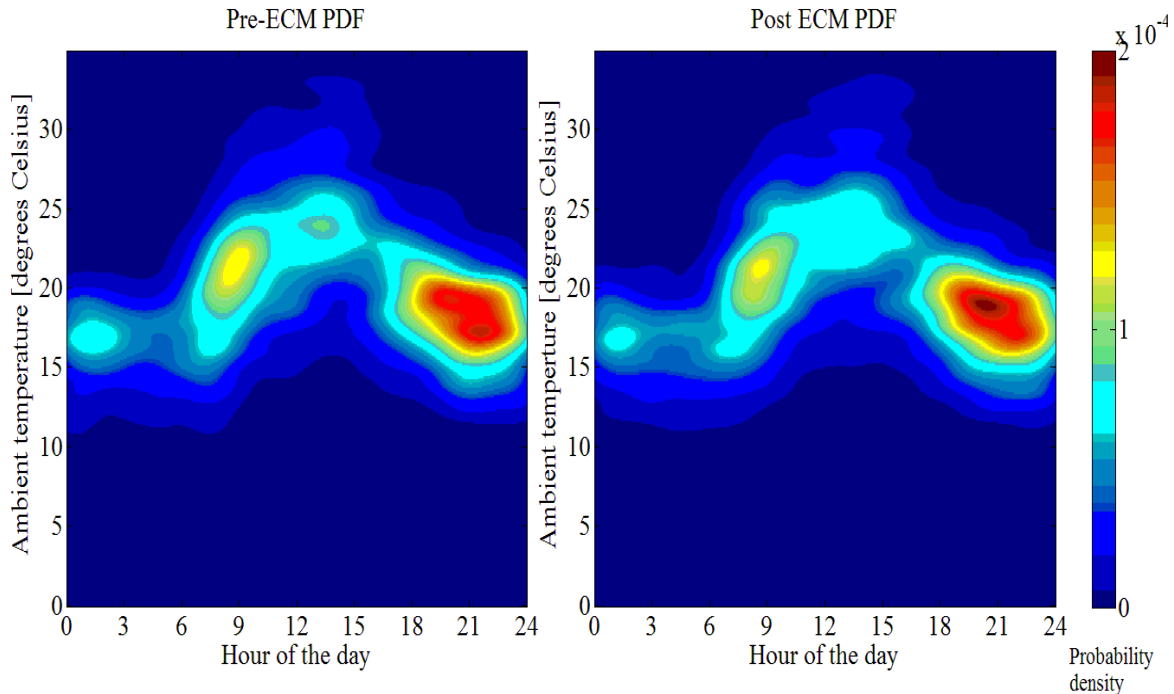
Figure 5.4.21: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the summer weekend days with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

The following is evident when Figures 5.4.20 and 5.4.21 are compared:

- The maximum of the weekend modes are less than that of the weekday modes, which can be ascribe to the fact that the hot water demand during the weekend day was distributed more evenly over the given time period.

- From the weekday and weekend day PDF plots it is evident that the ECM had reduced the overall demand density.

- The peak densities of the weekend PDFs occurred approximately an hour later when compared to the weekdays. This pattern in the demand results was influenced by the hot water consumption patterns.

The image plots of the PDFs of the pre- and post-ECM simulated data are depicted in Figures 5.4.18 and 5.4.19, respectively for the week and weekend days during the winter months.
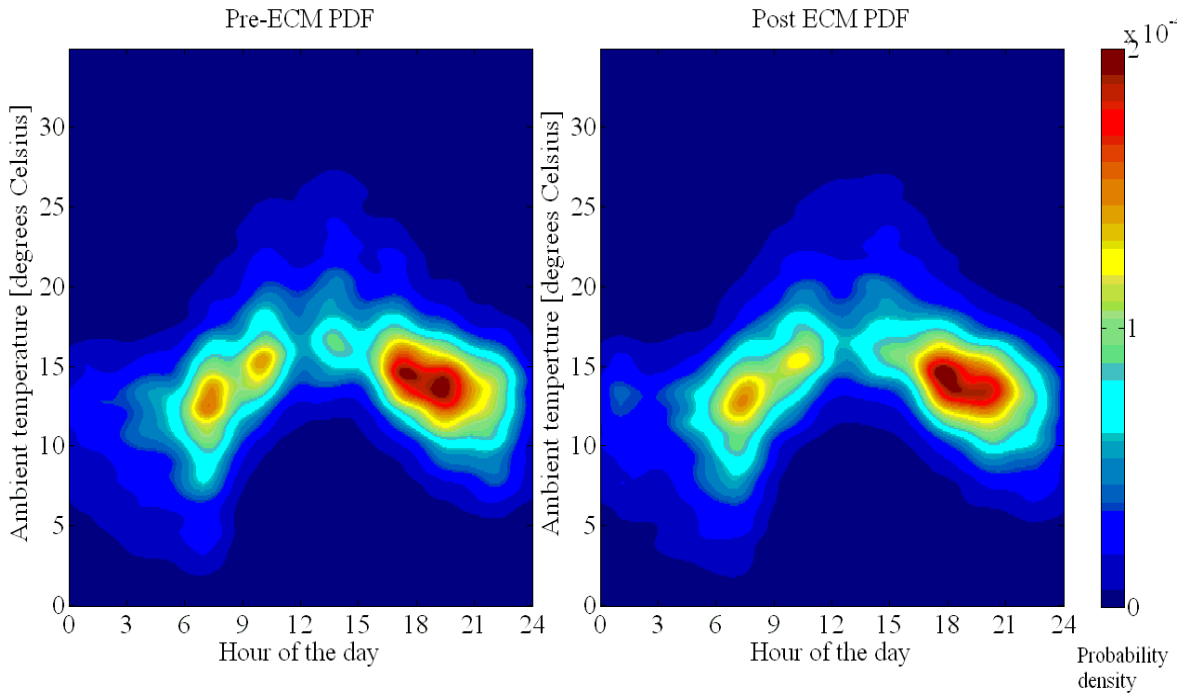
Figure 5.4.22: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the winter weekdays with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.
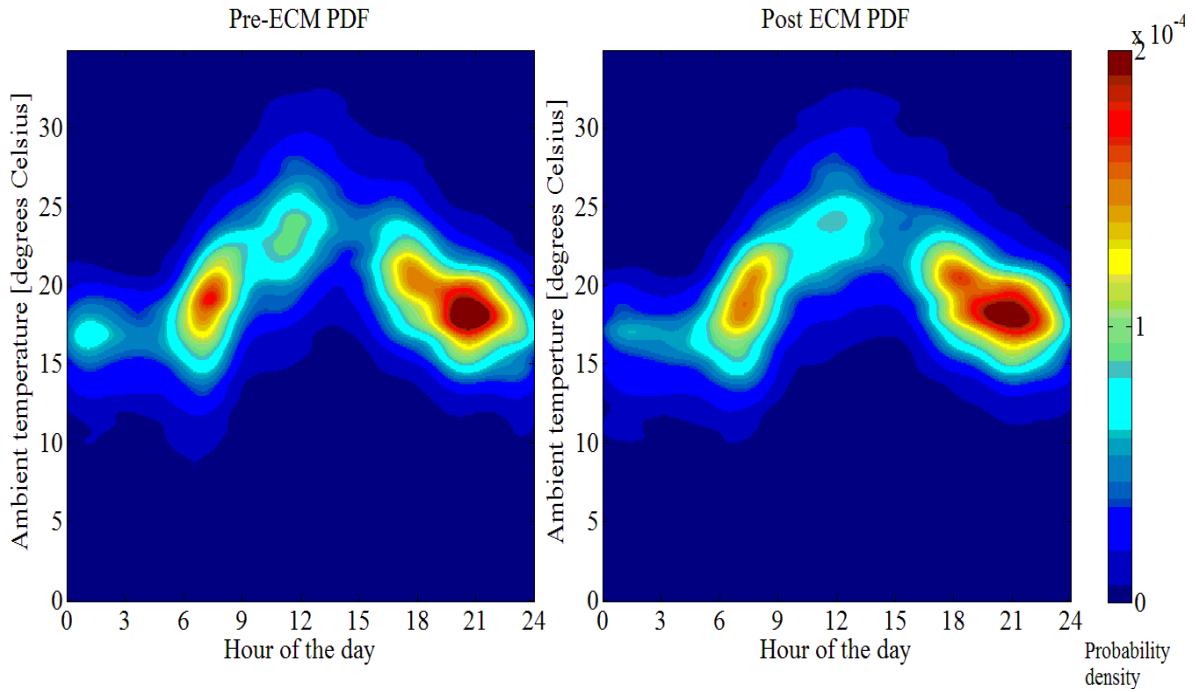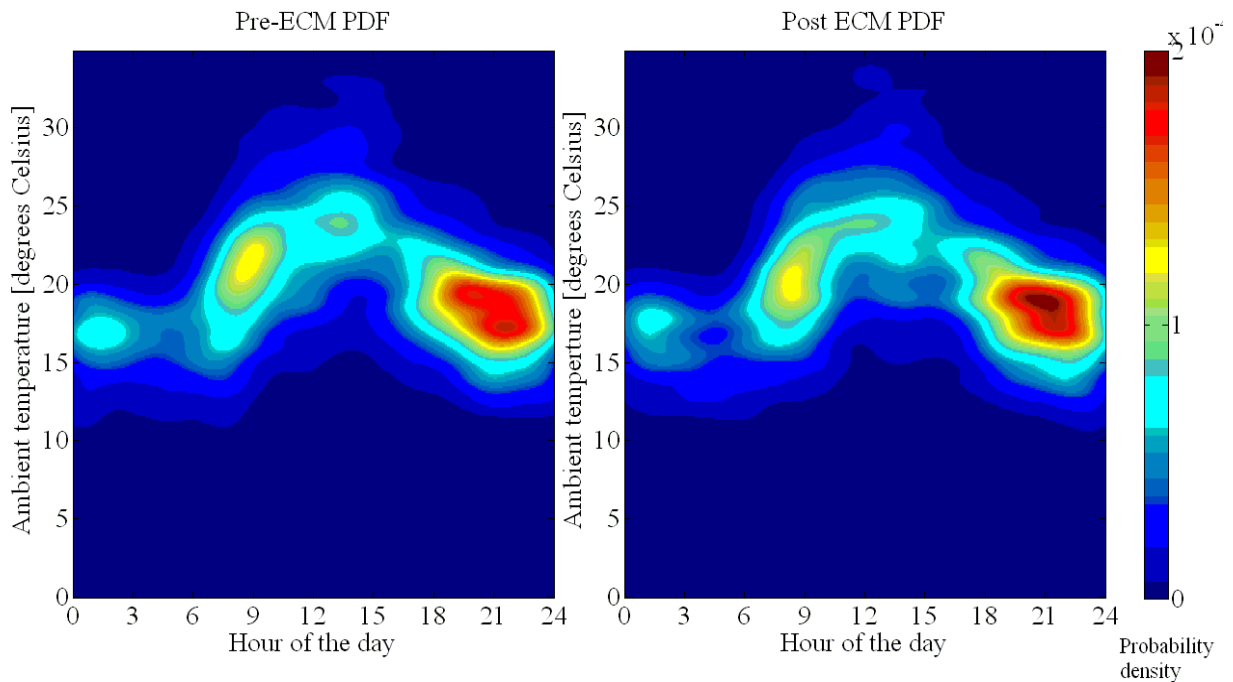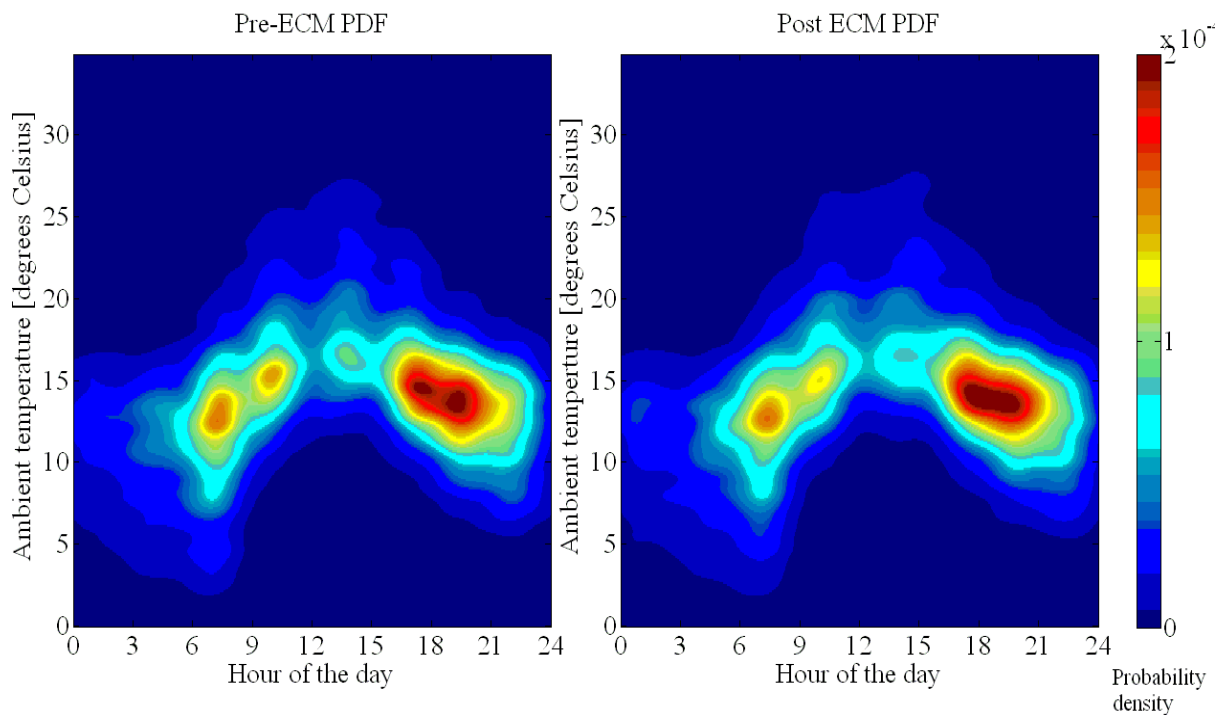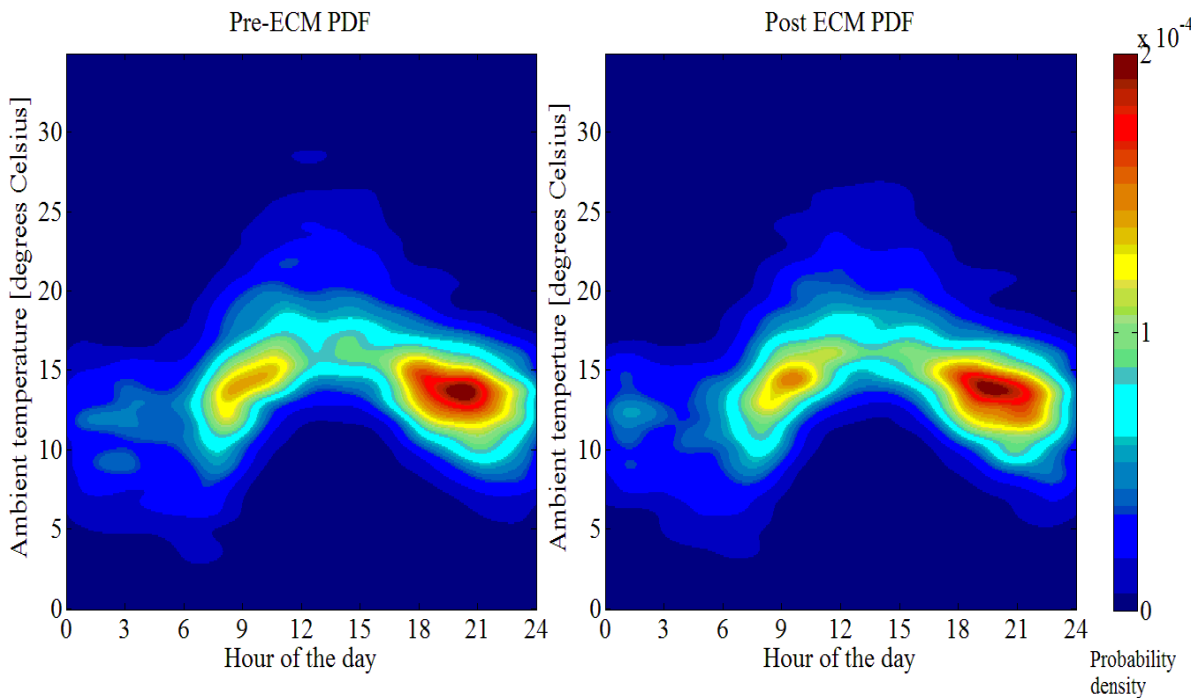

Figure 5.4.23: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the winter weekend days with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

When Figures 5.4.18 and 5.4.19 are compared it was noticed that demand peak densities are lower during the weekend days and occurred approximately an hour later compared to the weekdays. For both the week and weekend days of the pre- and post-ECM simulations, the demand density was shifted later in time as a result of the ECM.

When the summer and winter PDF plots are compared it can seen that the evening density mode occurred approximately an hour earlier for the winter months.

5.4.5 Modification of the geyser inlet water temperature profile

In the final ECM that was simulated, the geyser inlet water temperature profile was modified. This was a hypothetical scenario which endeavoured to investigate the demand impact a modified geyser inlet water temperature would have. The parallel model used in the first ECM was modified to take a fourth input variable which was the post ECM geyser model's inlet water temperature. The upper set points of the thermostats of both models were set at 65 degrees Celsius. A description of how the inlet water profile was generated is given in section 4.5.5.

5.4.5.1 Demand impact tables

The three characteristic housing densities' hot water consumption profiles were used in the simulated demand impacts as a result of the modified geyser inlet water temperature profile. The simulated demands of the pre- and post-ECM geyser models are recorded in Tables 5.4.10 to 5.4.12 for the nine year simulation period.

Table 5.4.10: Electrical demand impact that resulted from a modified geyser inlet water temperature profile with a high-density housing characteristic hot water consumption for the simulation period 2001 to 2009.

Demand impact [kWh]									
Model	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Pre-ECM	3679	3689	3682	3666	3663	3669	3664	3660	3641
Post ECM	3505	3513	3508	3495	3491	3496	3492	3489	3470
Difference between	174	176	174	171	172	173	172	171	171

Table 5.4.11: Electrical demand impact that resulted from a modified geyser inlet water temperature profile with a medium-density housing characteristic hot water consumption for the simulation period 2001 to 2009.

Demand impact [kWh]									
Model	2001	2002	2003	2004	2005	2006	2007	2008	2009
Pre-ECM	5051	5079	5047	5057	5063	5063	5044	5053	5049
Post ECM	4803	4829	4800	4806	4814	4812	4795	4806	4800
Difference between	248	250	247	251	249	251	249	247	249

Table 5.4.12: Electrical demand impact that resulted from a modified geyser inlet water temperature profile with a low-density housing characteristic hot water consumption for the simulation period 2001 to 2009.

Demand impact [kWh]									
Model	2001	2002	2003	2004	2005	2006	2007	2008	2009
Pre-ECM	6136	6122	6143	6125	6150	6140	6136	6134	6106
Post ECM	5821	5807	5829	5809	5832	5826	5821	5818	5792
Difference between	315	315	314	316	318	314	315	316	314

From the demand impact tables it is noted that the difference between the calculated demand of the pre- and post-ECM simulations remained virtually constant across the years of evaluation for each characteristic hot water consumption profile. The percentage demand impact for each characteristic hot water consumption profile was similar, at an average of 5 percent.

5.4.5.2   Probability density plots of ambient temperature versus time of day for geyser *on* times

The image plots of the PDFs for the pre- and post-ECMs' simulated data are depicted in Figures 5.4.24 and 5.4.25, respectively for the week and weekend days during the summer months.
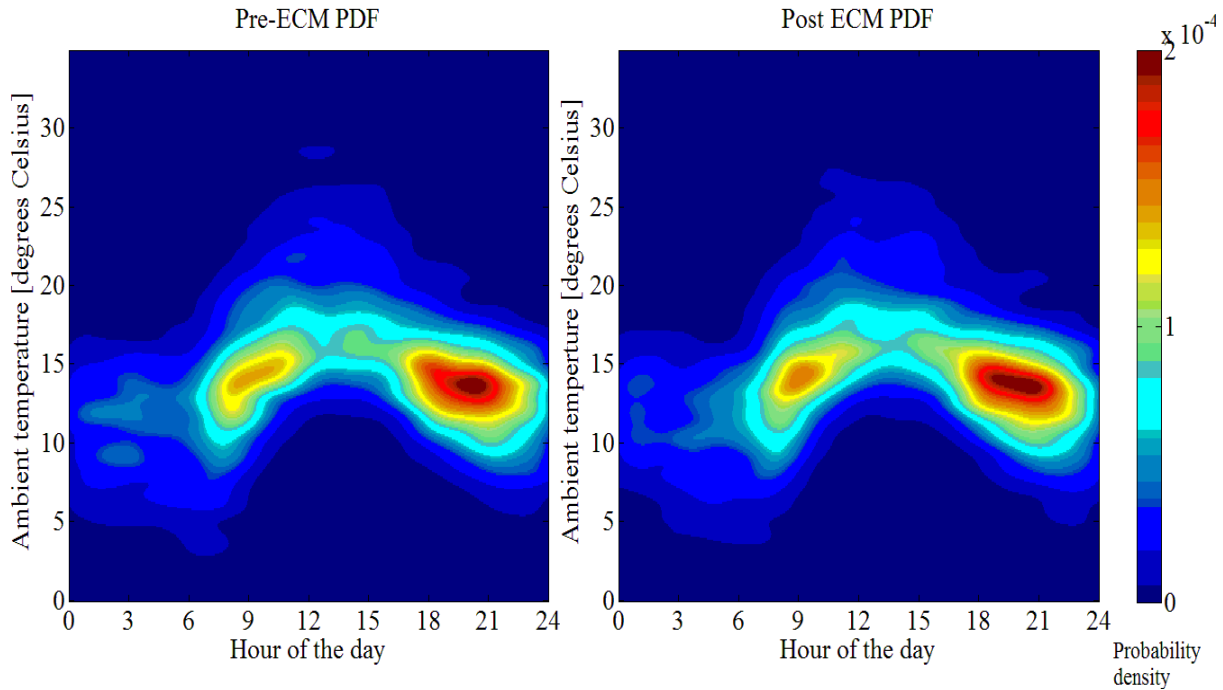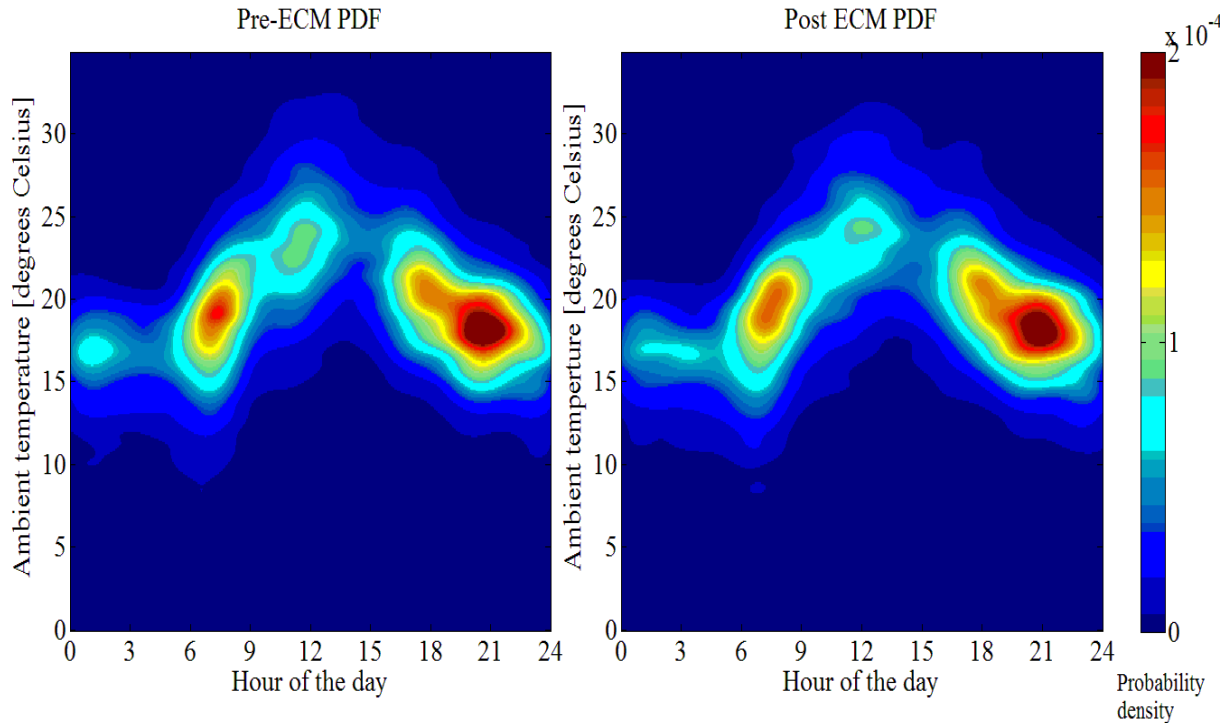
Figure 5.4.24: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the summer weekdays with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.
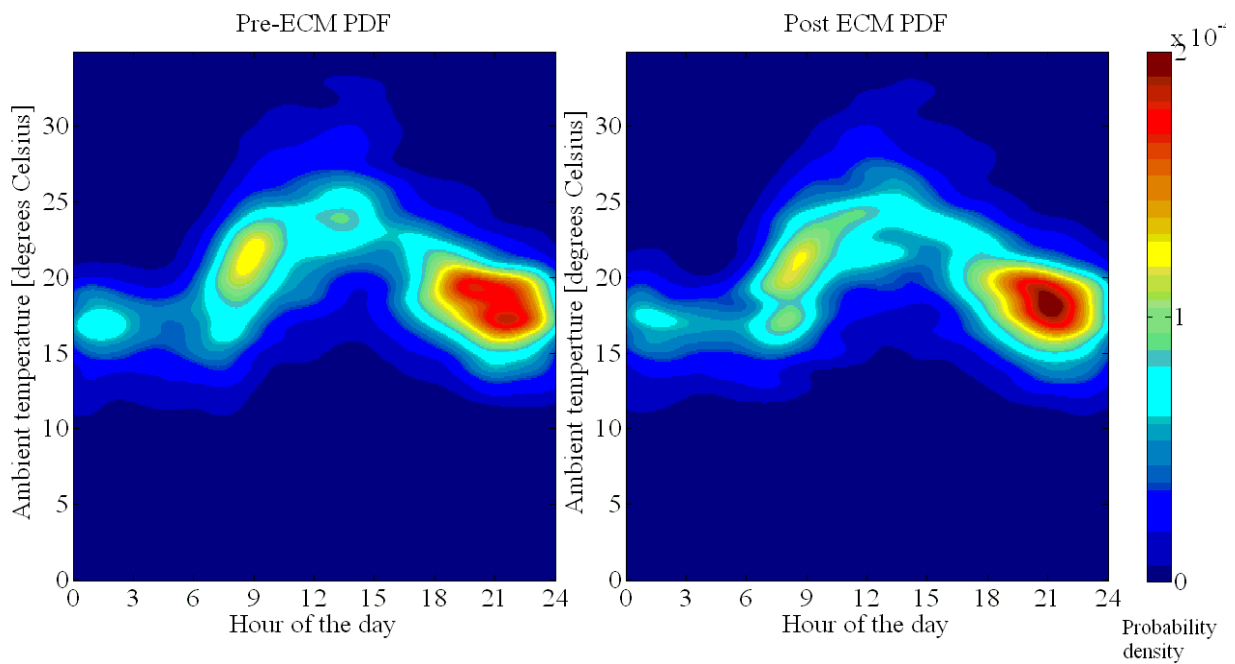
Figure 5.4.25: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the summer weekend days with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

149

The following is evident when Figures 5.4.24 and 5.4.25 are compared:

- The maximum of the weekend modes is less than that of the weekday modes, which can be ascribed to the fact that the hot water demand during the weekend day was distributed more evenly over the given time period.

- From the weekday PDF plots it is evident that the ECM had reduced the overall demand density.

- The peak densities of the weekend PDFs occurred approximately an hour later compared to the weekdays. This pattern in the demand results was influenced by the hot water consumption patterns.

The image plots of the PDFs for the pre- and post-ECMs' simulated data are depicted in Figures 5.4.26 and 5.4.27, respectively for the week and weekend days during the winter months.
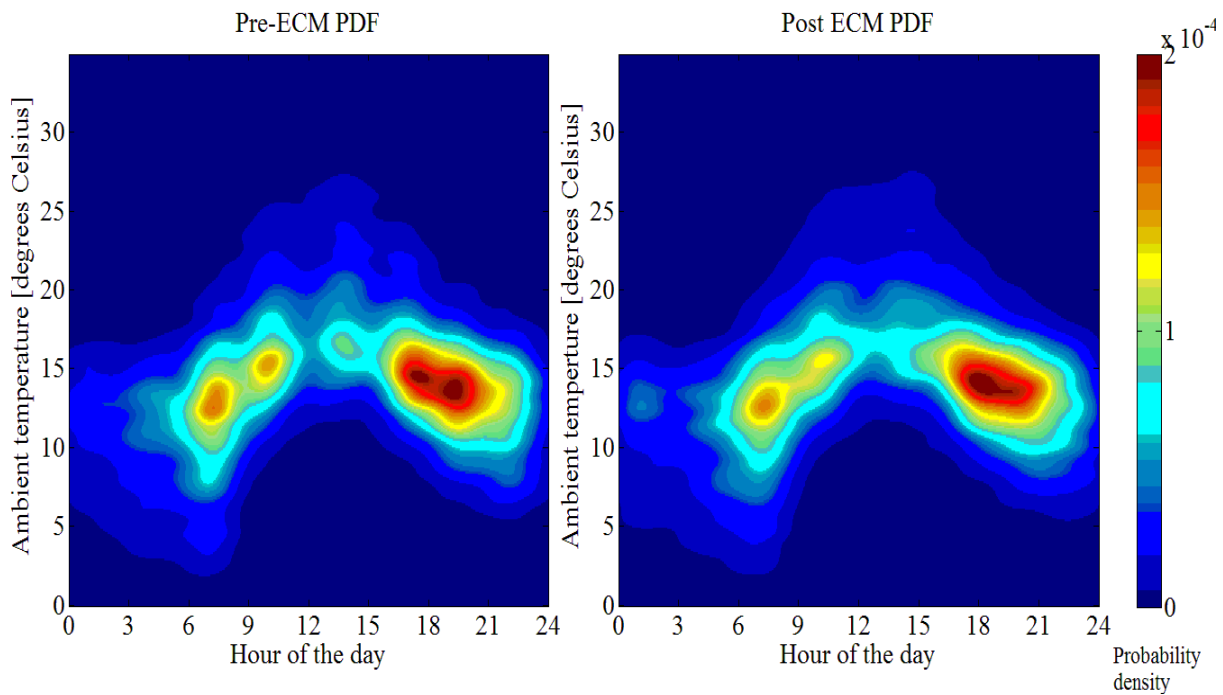
Figure 5.4.26: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the winter weekdays with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.
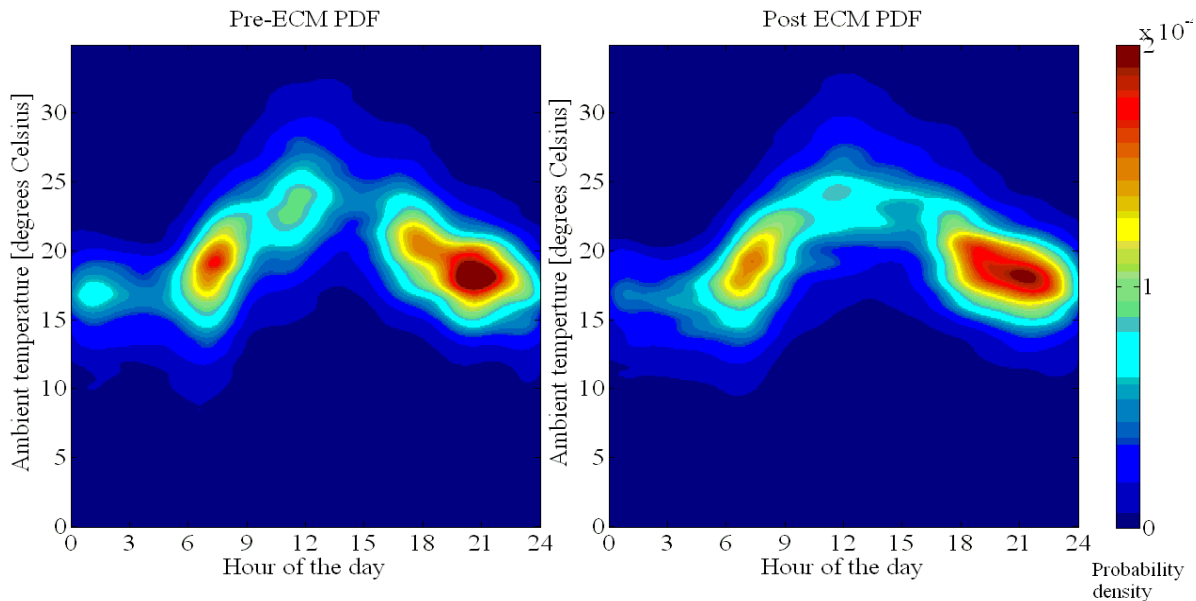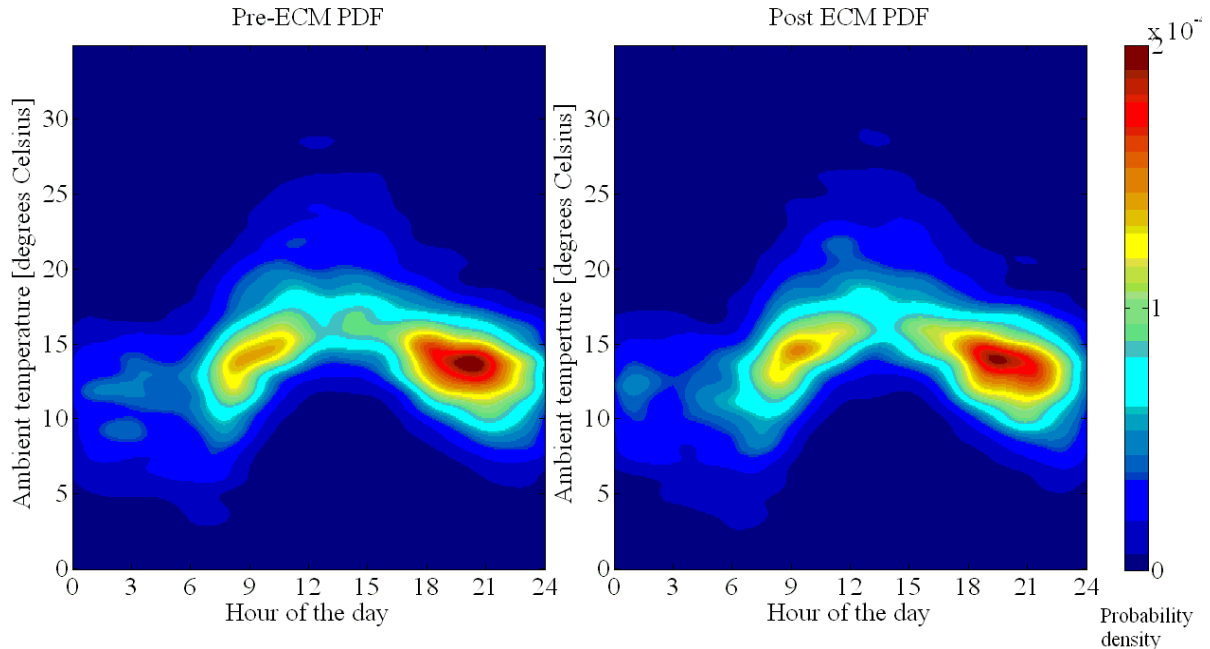
Figure 5.4.27: Probability density plots of the pre- and post-ECM PDFs of the ambient temperature versus time of day the geyser was *on* during the winter weekend days with a medium-density housing characteristic hot water consumption profile from 2001 to 2009.

The following is evident when Figures 5.4.26 and 5.4.27 are compared:

- The demand peak densities are lower during the weekend days and occurred approximately an hour later compared to the weekdays.
- For both the week and weekend days of the pre- and post-ECM simulations, the demand density was shifted later in time as a result of the ECM.

When the summer and winter PDF plots are compared it can seen that the evening density mode occurred approximately an hour earlier for the winter months. For both seasonal density plots the probability density was larger between the morning and evening mode during the weekend days.

5.4.6   Summary of conclusions drawn from simulation results

A brief summary of the conclusions attained from the three ECM simulations are presented in the next subsections.

5.4.6.1   Conclusions drawn from the demand impact tables

The conclusions drawn from the demand impact tables are summarised as follows:

151

- These results show that the median of the ambient temperature for an evaluation period has a strong correlation with the overall electrical demand of the geyser that results from the no load losses. This result is consistent with the fact that the lower the ambient temperature experienced, the greater the resultant standing losses would be because of the greater temperature difference between the ambient air and the water within the geyser.

- From the demand impact tables it was noted that the difference between the calculated demand of the pre- and post-ECM remained virtually constant across the years of evaluation for each characteristic hot water consumption profile for each of the ECMs.

- In the first simulated ECM where the geyser thermostat was adjusted from 65 to 55 degrees Celsius and a zero hot water demand profile was set, an average demand saving of 170 kWh was realised annually. When the three characteristic housing densities' hot water consumption profiles were applied to the models, the annual average demand saving was 167 kWh.

- In the second ECM, where the low flow shower roses were retrofitted, an average demand impact of 8.5 percent of pre-ECM demand profile was realised for each of the three characteristic housing densities' hot water consumption profiles.

- In the third ECM, where the inlet water temperature was modified, an average demand impact of 5 percent of pre-ECM demand profile was realised for each of the three characteristic housing densities' hot water consumption profiles.

5.4.6.2 Conclusions drawn from the probability density estimation of the recovery times of the geyser

The conclusions drawn from the probability density estimation of the recovery times of the geyser are summarised as follows:

- The recovery time of the geyser is directly related to its electrical demand.

- The probability density estimates of the recovery time as a result of the standing losses of the geyser shows it is inversely proportional to ambient temperature.

- The recovery time of the geyser is dependent on the volume of water inside the geyser and the kilo watt rating of the geyser element. The high-density estimates of the geyser recovery times over a small temperature ranges are characteristic of a geyser with no hot water demand

152

- Lowering the set point of the geyser thermostat shorten the recovery time of the geyser.

- The density estimates of the geyser recovery times during the summer months have three distinct modes.

- During the winter months the modes of the summer month simulation period have reduced to a single mode

- It is interesting to note that the peak density increased from the weekday to the weekend day density estimates during the winter months. This trend is directly opposite to the trend observed in the summer months. This could be explained by the reduce peak hot water consumption rates during the weekend days which results in shorter geyser recovery times.

5.4.6.3   Conclusions drawn from the image plots of the PDFs of the demand data

The conclusions drawn from the image plots of the PDFs of the demand data are summarised as follows:

- The maximum of the weekend modes was less than that of the weekday modes, which can be ascribed to the fact that the hot water demand during the weekend day was distributed more evenly over the given time period.

- From the weekday PDF plots it was evident that the ECMs had reduced the overall demand density.

- The peak densities of the weekend PDFs occurred approximately an hour later when compared to the weekdays. This pattern in the demand results was influenced by the hot water consumption patterns.

- The demand peak densities were lower during the weekend days and occurred approximately an hour later compared to the weekdays. For both the week and weekend days of the pre and post periods, the demand density was postponed a result of the ECMs.

5.5   Demand probability

The probability that the geyser will be *on* (demand at 3kW) can be calculated using equation 5.5.1 and the PDFs of the geyser *on* times, for a particular ambient temperature, time of day, day type and season of the year. The probability of the recovery time interval (indication of energy used in kWh) can also be calculated by using equation 5.5.1 and the PDFs of the geyser recovery

times for the same specified temperature and time interval. These calculations can be done for a pre- and post-ECM simulation. These calculated results presents the probability of the demand impact at a specified period of time and at a particular ambient temperature.

$$P(a < X < b) = \int_{a}^{b} f(x \mid T_{t}) dx \qquad \text{for all } a < b \qquad \qquad \textbf{5.5.1}$$

where

$f(x)$ denotes the probability density function at a specified ambient temperature and time

$P(a < X < b)$ denotes the probability of the variable X being in the range from a to b.

Chapter 6   **Conclusion**

6.1   Achievements of set objectives

The purpose of energy management is to ensure the sustainability of energy resources and assisting in ensuring the sustainability of the natural environment. The evaluation of the efficiency of energy consumers is at the heart of energy management. The DSM program drives the efficient use of electrical energy through ECMs. M&V plays an important role in the planning and evaluation of ECMs. The IPMVP was introduced as the foremost international protocol for M&V. In the author's opinion, the IPMVP fall short in its recommended approach to the statistical analysis of demand and demand impact that results from the implementation of an ECM. Further development is needed in this section of M&V and the IPMVP. Different approaches to statistical inference were investigated so that the most appropriate approach could be applied to the M&V process. The purpose of the statistical inference was to substantiate, with confidence levels, the energy impact claims provided by the M&V process as a result of an ECM.

The literature review revealed the importance of long and short-term load forecasting, in that load forecasting was not only applicable to savings but also to the forecasting of aggregated electrical load from a group to a network level of electrical consumers. Reference was made to a particularly relevant piece of work that was done on bivariate kernel density estimation to conduct short-term load forecasting. Whilst studying the methods advocated by this short-term load forecasting work, it became evident that, in the case of complex load systems, the relationship between load demand and the independent controlling variables needs to be established. The author deduced that the statistical methods used in load forecasting could be used in the evaluation of demand impacts that result from ECMs. To this end, the primary objective was set to develop a methodology in which a complex electrical load could be modelled both mathematically and statistically in order to determine the demand impact of an implemented ECM and the relationship of the demand impact to its controlling variables. A conventional geyser was chosen as the modelling environment to assist in the development of the methodology. The choice was based on the knowledge that hot water heating contributes significantly to the total domestic electrical demand. Many of the DSM initiatives have targeted

hot water heating and hot water heating technologies in an effort to manage and reduce domestic electrical load.

Through the literature review documented in Chapter two, comprehensive background knowledge was gained on hot water heating technologies, domestic hot water consumption patterns and the prediction of geyser inlet water temperature. This knowledge was used to develop the geyser model and the MATLAB code for the simulation environment.

To design an effective and efficient DSM intervention, a comprehensive understanding is required of how the independent variables of the electrical consumer are related to its dependent variables. This realisation resulted in the statistical analysis of the input and output variables to the geyser model in Chapters four and five. Parametric box plots and non-parametric univariate density estimation methods were used to analyse input variables; ambient temperature, hot water demand and inlet water temperature profiles to the geyser model. This analysis gave a better understanding of the statistical nature of the input variables. The bivariate kernel density estimation technique was used to produce the demand PDFs. Kernel density estimation is a modern technique used in probability density estimation and its strengths and weaknesses were discussed. The importance of, and methods used to determine bandwidth selection for kernel density estimation, were highlighted. The PDFs were used to determine the confidence factors associated with the predicted impacts of an ECM. The PDFs were also used to evaluate the relationship between the simulated demand and its controlling variable, ambient temperature.

The effects of input data manipulation to the geyser model were also investigated. Data manipulation in this respect was limited to the altering of the input variable's sampling period through processes of averaging and linear interpolation. Simulations in which input variables' sampling periods had been manipulated, revealed very little difference in the simulated energy demand.

The sensitivity of the geyser model towards the mean and variance of the ambient temperature was investigated. It was concluded that the method followed to determine the sensitivity was appropriate; however, the results would require verification with simulations that made use of recorded data sets for all the input variables and not just ambient temperature. The hot water demand and geyser inlet water temperature were generated for the simulations of the geyser

model since appropriate input simulation data was not available for the desired simulation period. The author believes that the true extent of the dependency of electrical demand on ambient temperature will only be realised when a data set becomes available where all three input variables to the geyser model have been recorded in a predetermined environment.

Finally, two geyser models were run in parallel: the one model simulated the pre-ECM demand and the other model simulated the demand of the post ECM period. This approach effectively simulated the demand impact that would result after the implementation of a geyser ECM. Three geyser ECMs were simulated:

- adjustment of a geyser thermostat to 10 degrees below the factory setting;
- retrofitting of low flow shower roses in order to reduce the hot water demand profile, and
- adjustment of the inlet water temperature profile to a geyser.

The ECM where low flow shower roses were installed proved to be the most effective. From this it was concluded that the hot water demand profile has the greatest influence on the electrical demand of the geyser. However, the Meyer et al. study concluded through evidence given in the summer and winter hot water demand profile that ambient temperature had a significant influence on hot water demand. The true dependency of electrical demand on ambient temperature will be realised when a data set of recorded input variables are used in this methodology that has been developed. The automatic bandwidth selection method that was used during the kernel density estimation of the geyser demand provided realistic results. A second source of bandwidth selection was investigated and a bivariate automatic bandwidth selection function was sourced from Mathworks [45]. Very little information was available regarding the principle that was used to develop the bandwidth selection for this function. The author of the function claims that it is most accurate when Gaussian distributions are estimated. The function uses recursion to estimate the diagonals of the bandwidth matrix. The density estimations in Chapter five made use of large data sample counts which caused a programmatic error in the form of a stack overflow when running the code in MATLAB to estimate the bandwidths. This function did therefore not reliably compare with the bandwidth selection function utilised in this thesis.

6.2   Recommendations and future work

Further research is required to find the most appropriate approach to establishing automatic bandwidth selection for the methodology developed in this thesis. The technique developed by Doung to estimate the off-diagonal elements of the bandwidth matrix should be further investigated [25]. This could prove to be particularly important in determining the dependency between the dependent and independent variables. An investigation regarding the influence which bandwidth selection has on the demand impact results would also be prudent.

The importance of acquiring recorded data sets cannot be stressed enough. Data is an essential component for the effective evaluation of ECMs and investment in data collection is much needed.

Future development based on this work could include the design of a graphical user interface for the software that was developed, to enhance the ease with which an ECM can be evaluated.

With the present level of development of the methodology, the pre- and post-ECM demand profiles can be compared with their respective confidence levels. A different approach may prove to be feasible in that the kernel function of the post ECM could be subtracted from the kernel function of the pre-ECM which in theory would result in a kernel function of the demand impact. The subtraction of the two bivariate kernel functions could be done by the multiplication of the Fourier transforms of the pre-ECM kernel function and the negative of the post ECM kernel function. An alternative method to this would be the multiplication of the characteristic functions of the kernel functions. The resultant function of either of these methods could then be inverse Fourier transformed to give the savings kernel function which could be used to generate the demand impact PDFs.

By employing the methodology presented in this thesis on the load modelling of a complex load system and the statistical analysis of its input/output variables, one would gain a far greater insight into the load system's relationship between the dependent and independent variables. It is recommended that this methodology be used to evaluated and develop future ECMs for reducing and moving load demand in load profiles. The methodology developed in this thesis is of particular importance for assessing the savings impacts of heat pump and solar water heater technologies.

Bibliography

[1] Ministry of Economic Development, "New Zealand Energy Strategy to 2050," New Zealand Government, Wellington, 2007.

[2] Ministry of Economic Development, "New Zealand Energy Efficiency and Conservation Strategy," New Zealand Goverment, Wellington, 2007.

[3] H. Willis, *Spatial electric load forecasting*, 1st ed. New York: CRC Press, 1998.

[4] Grunwald M., "Wasting Our Watts," *Time Magazine*, 12th January 2009.

[5] Efficiency Valuation Organization, "International Performance Measurement and Verification Protocol," Toronto Canada, September 2010.

[6] (2011, October) Eskom. [Online]. http:/www.Eskomidm.co.za/residential

[7] Anderson C. BassonJ. Eberhart.A Pickering M., "South African Energy Policy Discussion Document," South African Government, Pretoria, 1995.

[8] Eskom. (2011, Oct) Eskom. [Online]. http://www.eskomdsm.co.za

[9] SABS Standards Division, Measurement and Verification of Energy Savings, 2010.

[10] The Australian Energy Performance Contracting Association, "A Best Practices Guide to Measurement and Verification of Energy Savings," Sydney, 2004.

[11] The Association of Edison Illuminating Companies Load Research Committee, "Demand Response Measurement and Verification," March 2009.

[12] Nexant Inc, "M&V Guidelines: Measurment and Verification for Federal Energy Projects version 3.0," U.S. Department of Energy Federal Energy Management Program, Boulder CO., April 2008.

[13] New Brunswick Industrial Program, "Project-Level Measurement and Verification Manual," January 2009.

[14] Jeff S. Haberl, Charles Culp, and David E. Caridge, "ASHRAE's Guideline 14-2002 for Measurement of Energy and Demand Savings: How to Determine what was really saved by the retrofit.," , Pittsburgh Pennsylvania, October 2005.

[15] Prof LJ Grobler. (2010, March) International M&V Benchmarks and Eskom Practices. M&V Workshop.

[16] American College & University. (2009, April) Measurement and Verification and the IPMVP. Report.

[17] Simonoff J.S., *Smoothing methods in statistics*. New York: Springer-Verlag, 1996.

[18] Wasserman, *All of statistics: A Concise Course in Statistical Inference*.: Springer Texts in Statistics, 1998.

[19] Scott, *Multivariate density estimation: Theory, Practice, and Visualization*. New York: John Wiley & Sons Inc., 1992.

[20] "Ozean," *Ozean journal of applied sciences*, 1998.

[21] Silverman B.W., *Density Estimation for Statistical and Data Analysis*. London: Chapman and Hall, 1986.

[22] Scheild S. (2004, January) Introduction to Kernel Smoothing. Presentation.

[23] Brian Ambberg. (2008, December) A range of different kernels.

[24] Wand M.P. and Jones M.C., *Kernel smoothing*. London: Chapman and Hall Ltd, 1995.

[25] Tarn Duong, Bandwidth selectors for multivariate kernel density estimation, 2004, Doctoral thesis.

[26] Doung, T. & Hazelton M.L., "Plug-in bandwidth matrices for bivariate kernel density estimation," *Journal of Nonparemetric Statistics*, vol. 15, no. 17-30, p. 14, 2003.

[27] W. Charytoniuk M.S. Chen P. Kotas P. Van Olinda, "Demand Forecasting in Power Distribution Systems Using Nonparametric," *IEEE Transactions on Power Systems, Vol. 14, No. 4*, pp. 1200-1207, 1998.

[28] Mehra M., Demand Forecasting for Electricity, 2004.

[29] WTRG Economics. (2011, October) ioga.com. [Online]. http://www.ioga.com/Special/crudeoil_Hist.htm

[30] South African Press Association. (2011, October) Fin 24.com. [Online]. http://www.fin24.com/Economy/IFP-Is-Eskom-bankrupt-20080320

[31] Meyer J. and Tshimankinda M., "Domestic Hot Water Consumption in South African Houses for Developed and Developing Communities," *International Journal of Energy Research*, vol. 21, pp. 667 - 673, 1997.

[32] Central Statistical Services, October Household Survey, 1994.

[33] DeOreo W. MayerP., "The End Uses of Hot Water in Single Family Homes from Flow Trace Analysis," Seattle Public Utilities and US EPA, Seattle, 2000.

[34] Wilkes C. Mason A. Niang L. Jensen K. Hern S., "Quantification of Exposure-Related Water Uses for Various U.S. Subpopulations," United States Environmental Protection Agency, Las Vegas, 2005.

[35] Aguilar C. White D. Ryan D., "Domestic Hot Water Heating and Heater Energy Consumption in Canada," Canadian Building Energy End-Use Data and Analysis Center, CBEEDAC 2005-RP-02, 2005.

[36] Bosman I. Grobler L. Dalhleish A., "Determination of the Impact on Standing Losses of Installing Blankets to Electric Hot Heaters in South Africa," *Journal of South Africa*, vol.

17, no. 2, pp. 57 - 64, May 2006.

[37] (2011, July) Volker- Quaschning. [Online]. http://www.volker-quaschning.de/articles/fundamentals4/index.php,

[38] (2007) Greenterafirma. [Online]. http://greenterrafirma.com

[39] Department of buildings and housing. (2009) Department of buildings and housing. [Online]. http://www.dbh.govt.nz/codewords-38-2

[40] Hillel D., *Applications of soil physics*. New York: Academic Press, 1980.

[41] Houston A. Tranter G. Miller I. (2004) The Application of Field Soil Physical Methods on a Epicalcareous epipedal Black Vertosol, under Constraining Soil Managment.

[42] Fredlund D., *Advances in Measurement of Soil Physical Properties: Bringing Theory into Practice*.: Soil Science Society of America, 1992.

[43] Nofziger D. Wu J., "Soil Temperature Changes with Time and Depth," Department of Soil Sciences, Oklahoma State University, Stillwater OK, Stillwater, 2003.

[44] South African Weather Bureau, Weather data for Cape Town International Airport, 2010.

[45] (2009, November) Mathworks. [Online]. http://www.mathworks.com/index.html

[46] Mathworks. (2009) Matlab help file. Matlab 7b software.

[47] South African Bureau of Standards, "Fixed electric water storage heaters," South African Bureau of standards, Pretoria, Certificate SANS 151/SABS 151:2002, 2006.

[48] (2011, October) Hamilton city council. [Online]. http://hamilton.co.nz/page/pageid/2145847371

[49] (2011, October) Consumer. [Online]. http://www.consumer.org.nz/reports/shower-heads

[50] (2001, November) Satinjet. [Online]. http://www.satinjet.com/

[51] PJM Forward Market Operations, "Energy Efficiency Measurement &Verification," Manual 18B, March 2010.

Appendix A – M&V tables

Table A1: Summary of IPMVP option selection table part I. [5].

IPMVP Option	How Savings Are Calculated	Typical Applications
A. Retrofit Isolation: Key Parameter Measurement *Savings* are determined by field measurement of the key performance parameter(s) which define the *energy* use of the *ECM*'s affected system(s) and/or the success of the project. Measurement frequency ranges from short-term to continuous, depending on the expected variations in the measured parameter, and the length of the *reporting period*. Parameters not selected for field measurement are *estimated*. *Estimates* can based on historical data, manufacturer's specifications, or engineering judgment. Documentation of the source or justification of the *estimated* parameter is required. The plausible *savings* error arising from *estimation* rather than measurement is evaluated.	Engineering calculation of *baseline* and *reporting period energy* from: ○ short-term or continuous measurements of key operating parameter(s); and ○ *estimated* values. *Routine* and *non-routine adjustments* as required.	A lighting retrofit where power draw is the key performance parameter that is measured periodically. Estimate operating hours of the lights based on building schedules and occupant behavior.
B. Retrofit Isolation: All Parameter Measurement *Savings* are determined by field measurement of the *energy* use of the *ECM*-affected system. Measurement frequency ranges from short-term to continuous, depending on the expected variations in the *savings* and the length of the *reporting period*.	Short-term or continuous measurements of *baseline* and *reporting-period energy*, and/or engineering computations using measurements of proxies of *energy* use. Routine and non-routine adjustments as required.	Application of a variable-speed drive and controls to a motor to adjust pump flow. Measure electric power with a kW meter installed on the electrical supply to the motor, which reads the power every minute. In the *baseline period* this meter is in place for a week to verify *constant* loading. The meter is in place throughout the *reporting period* to track variations in power use.

Table A2: Summary of IPMVP option selection table part II. [5]

IPMVP Option	How Savings Are Calculated	Typical Applications
C. Whole Facility *Savings* are determined by measuring energy use at the whole *facility* or sub-*facility* level. Continuous measurements of the entire *facility's energy* use are taken throughout the *reporting period*.	Analysis of whole *facility baseline* and *reporting period* (utility) meter data. *Routine adjustments* as required, using techniques such as simple comparison or regression analysis. *Non-routine adjustments* as required.	Multifaceted energy management program affecting many systems in a *facility*. Measure energy use with the gas and electric utility meters for a twelve month *baseline period* and throughout the *reporting period*.
D. Calibrated Simulation *Savings* are determined through simulation of the *energy* use of the whole *facility,* or of a sub-*facility*. Simulation routines are demonstrated to adequately model actual *energy* performance measured in the *facility*. This Option usually requires considerable skill in calibrated simulation.	Energy use simulation, calibrated with hourly or monthly utility billing data. (Energy end use metering may be used to help refine input data.)	Multifaceted energy management program affecting many systems in a facility but where no meter existed in the *baseline* period. Energy use measurements, after installation of gas and electric meters, are used to calibrate a simulation. *Baseline* energy use, determined using the calibrated simulation, is compared to a simulation of *reporting period* energy use.

165

Table A3: Overview of M&V options A, B, C and D. [12]

M&V Option	Performance[1] and Usage[2] Factors	Savings Calculation
Option A—Retrofit Isolation with Key Parameter Measurement	This option is based on a combination of measured and estimated factors when variations in factors are not expected. Measurements are spot or short-term and are taken at the component or system level, both in the baseline and post-installation cases. Measurements should include the key performance parameter(s) which define the energy use of the ECM. Estimated factors are supported by historical or manufacturer's data. Savings are determined by means of engineering calculations of baseline and post-installation energy use based on measured and estimated values.	Direct measurements and estimated values, engineering calculations and/or component or system models often developed through regression analysis Adjustments to models are not typically required.
Option B—Retrofit Isolation with All Parameter Measurement	This option is based on periodic or continuous measurements of energy use taken at the component or system level when variations in factors are expected. Energy or proxies of energy use are measured continuously. Periodic spot or short-term measurements may suffice when variations in factors are not expected. Savings are determined from analysis of baseline and reporting period energy use or proxies of energy use.	Direct measurements, engineering calculations, and/or component or system models often developed through regression analysis Adjustments to models may be required.
Option C – Utility Data Analysis	This option is based on long-term, continuous, whole-building utility meter, facility level, or sub-meter energy (or water) data. Savings are determined from analysis of baseline and reporting period energy data. Typically, regression analysis is conducted to correlate with and adjust energy use to independent variables such as weather, but simple comparisons may also be used.	Based on regression analysis of utility meter data to account for factors that drive energy use Adjustments to models are typically required.
Option D—Calibrated Computer Simulation	Computer simulation software is used to model energy performance of a whole-facility (or sub-facility). Models must be calibrated with actual hourly or monthly billing data from the facility. Implementation of simulation modeling requires engineering expertise. Inputs to the model include facility characteristics; performance specifications of new and existing equipment or systems; engineering estimates, spot-, short-term, or long-term measurements of system components; and long-term whole-building utility meter data. After the model has been calibrated, savings are determined by comparing a simulation of the baseline with either a simulation of the performance period or actual utility data.	Based on computer simulation model (such as eQUEST) calibrated with whole-building or end-use metered data or both. Adjustments to models are required.

[1] Performance factors indicate equipment or system performance characteristics, such as kW/ton for a chiller or watts/fixture for lighting.

[2] Operating factors indicate equipment or system operating characteristics such as annual cooling ton-hours for chillers or operating hours for lighting.

Table A4: M&V system approach summary table I. [9]

1	2	3	4
Retrofit isolation		**Whole facility**	**Calibrated simulation**
Only the performance of the systems affected by the ESMs is of concern, either owing to the responsibilities assigned to the parties in an energy performance contract, or owing to the savings of the ESMs being too small to be detected in the time available using the whole facility option		The energy performance of the whole facility will be assessed, not just the energy savings measures	Either the baseline energy data or the reporting-period energy data are unavailable or unreliable
Interactive effects of the ESMs on the energy use of other facility equipment can be reasonably estimated, or assumed to be insignificant		There are many types of ESMs in one facility	There are too many ESMs to assess using other options
Possible changes to the facility, beyond the measurement boundary, would be difficult to identify or to assess		The ESMs involve activities whose individual energy use is difficult to separately measure, for example, operator training, and wall or window upgrades	The ESMs involve diffuse activities that cannot easily be isolated from the rest of the facility, such as operator training or wall and window upgrades
The independent variables, which affect energy use, are not excessively difficult or expensive to monitor		The savings are large compared to the variance in the baseline data during the reporting period	The performance of each ESM will be estimated individually within a multiple ESMs project, but the costs of other options are excessive
Sub-meters already exist to isolate energy use of systems		When retrofit-isolation techniques (option A or B) are excessively complex. For example, when interactive effects or interactions between ESMs are substantial	Interactions between ESMs or ESMs interactive effects are complex, making the isolation techniques of the retrofit options impractical
Meters added at the measurement boundary can be used for other purposes such as operational feedback or tenant billing		Major future changes to the facility are not expected during the reporting period	Major future changes to the facility are expected during the reporting period, and there is no way to track the changes or account for their impact on energy use (or both)
Measurement of parameters is less costly than option D simulations or option C non-routine adjustments		A system of tracking static factors can be established to enable possible future non-routine adjustments	An experienced energy-simulation professional is able to gather appropriate input data to calibrate the simulation model

167

Table A5: M&V system approach summary table II. [9]

1	2	3	4
Retrofit isolation		**Whole facility**	**Calibrated simulation**
Long-term testing is not warranted		Reasonable correlations can be found between energy use and other independent variables	The facility and the ESMs can be modelled by well-documented simulation software
There is no need to directly reconcile savings reports with changes in payments to energy suppliers			Simulation software predicts metered calibration data with acceptable accuracy
Key-parameter measurement	**All-parameter measurement**		
Estimation of key parameters may avoid possibly difficult non-routine adjustments when future changes happen within the measurement boundary	Meters added for isolation purposes will be used for other purposes such as operational feedback or tenant billing		Only one year's performance is measured in detail – typically immediately following installation and commissioning of the energy-management programme. After the first year only the parameters that serve as inputs to the simulation need to be measured
The uncertainty created by estimations is acceptable	Measurement of all parameters is less costly than calibrated simulation		
The continued effectiveness of the ESM can be assessed by simple routine inspection of estimated parameters	Savings or operations within the measurement boundary are variable		
Estimation of some parameters is less costly than measurement of them in using other options			
A key parameter used in computing savings is well known. Key parameters are parameters used to judge the performance of an ESM			

168

Appendix B – MATLAB code

Main file

```matlab
%////////////////////////////////////////////////////////////////////////
%////////////////////////////////////////////////////////////////////////
%Main script file

%////////////////////////////////////////////////////////////////////////
clear all
close all
clc
format long
memory
%////////////////////////////////////////////////////////////////////////
%Import of Ambient Temperature data
tic

inMat=txt2mat('Capetown4.txt');      %Import of temperature data

memory
toc
%%
%Formation of Time Temperature matrix
%Time stamp is attached to each temperature value
tic

Years=2001:2009;          %range of years to simulate
numberOfY=length(Years);
firstDataYear=2001;       %year of first data sample
firstDataMonth=1;         %month of first data sample
firstDataDay=1;           %day of first data sample
unixStartDate=datenum(firstDataYear,firstDataMonth,firstDataDay);
unixEndDate=datenum(firstDataYear+numberOfY,firstDataMonth,firstDataDay);
dayOfMonthColom=1;
firstDataHourColom=2;
%Function call to generate temperature matrix

[minTemp, hHourTemp,]=dataMatrix(inMat, unixStartDate, unixEndDate,...
    firstDataYear, firstDataMonth, firstDataHourColom, dayOfMonthColom);

memory
toc
%%
%Geyser inlet water temperature data generation
tic
%Function call to generate inlet water temperature

[ Twmin Twhh ] = InWaterTemp( hHourTemp );

memory
toc
```

169

```matlab
%%
%Hot water consumption generator from shower periods
tic
%Function call to generate hot water consumption

[ minConMatWd, minConMatWe, allMinConMat, hHourConMatWd, hHourConMatWe, ...
    allhHourConMat] = HotWgen(Years, firstDataMonth, numberOfY);


clear Years dayOfMonthColom firstDataHourColom firstDataMonth...
    firstDataYear numberOfY;


memory
toc
%%
%Hot water consumption generator from collected average data
tic
%Function call to generate hot water consumption

[ hHourConMatWd, hHourConMatWe, allhHourConMat] = ...
    HotWgenData2(Years, firstDataMonth, numberOfY);


clear Years dayOfMonthColom firstDataHourColom firstDataMonth...
    firstDataYear numberOfY;


memory
toc
%%
%Simulation of half hour hot water consumption data for half hour
%electricity demand of baseline geyser model
tic
%Function call for time series generation of input variables
[ var, var2, var3 ] = hhGyserMod( allhHourConMat, hHourTemp, Twhh,...
    unixStartDate );


simTime=60*30*size(allhHourConMat,1);     %simulation time
EI=208000000;                              %initial energy in geyser

%Call for simulation
sim('Geiser_model_65C');                   %per half hour simulation


%Demand matrix formation
Demand=Elec.signals.values;
elecDemandHHourMat=cat(2,allhHourConMat(1:end,1),Demand(1:end-1));


clear  var var2 var3 simTime EI;


memory
toc
%%
```

170

```matlab
%Simulation of per minute hot water consumption data
%for half hour electricity demand of baseline geyser model
tic

%Data sampling period adjustment
allhHourConMat=cat(1,allhHourConMat(:,:),...
    [unixStartDate+(datenum(0,0,0,0,30,0)*size(allhHourConMat,1))...
    allhHourConMat(1,2)]);
allMinConMat=cat(2,(allhHourConMat(1,1):datenum(0,0,0,0,1,0):...
    allhHourConMat(size(allhHourConMat,1),1))',...
    (jcbresample(1:1/30:size(allhHourConMat,1),...
    allhHourConMat(:,2)',1:1:size(allhHourConMat,1))));%
allMinConMat=allMinConMat(1:end-1,1:2);

%Function call for time series generation of input variables
[ var, var2,var3 ] = minGyserMod( allMinConMat, minTemp, Twmin,...
    unixStartDate);



simTime=60*size(allMinConMat,1);          %simulation time
EI=208000000;                             %initial energy in geyser

%Call for simulation
sim('Geiser_model_65C');                  %per min simulation

%Demand matrix formation
Demand=Elec.signals.values;
elecDemandMinMat=cat(2,allhHourConMat(1:end,1),Demand(:));


clear Elec var var2 var3 simTime EI;


memory
toc
%%
%Parallel model run for first ECM evaluation
%Thermostat adjustment
tic

%Function call for time series generation of input variables
[ var, var2, var3 ] = hhGyserMod( allhHourConMat, hHourTemp, Twhh,...
    unixStartDate );

simTime=60*30*size(allhHourConMat,1);     %simulation time
EI=208000000;                             %initial energy in geyser

%Call for simulation
sim('Geiser_model_55Cvs65C_NoBla');       %per half hour simulation

%Demand matrix formation
DemandBas=ElecDemBase.signals.values;
DemandRp=ElecDemRp.signals.values;
elecDemandHHourMat=cat(2,allhHourConMat(1:end,1),DemandBas(1:end-1),...
    DemandRp(1:end-1));
```

171

```matlab
clear  var var2 var3 simTime EI;

memory
toc
%%
%Parallel model run for second ECM evaluation
%Low flow shower rose
tic

%Function call for time series generation of input variables
[ var, var2, var3 ] = hhGyserMod( allhHourConMat, hHourTemp, Twhh,...
    unixStartDate );

simTime=60*30*size(allhHourConMat,1);    %simulation time
EI=208000000;                            %initial energy in geyser

%Call for simulation
sim('Geiser_model_65C_lowFlow');         %per half hour simulation

%Demand matrix formation
DemandBas=ElecDemBase.signals.values;
DemandRp=ElecDemRp.signals.values;
elecDemandHHourMat=cat(2,allhHourConMat(1:end,1),DemandBas(1:end-1),...
    DemandRp(1:end-1));

clear  var var2 var3 simTime EI;

memory
toc
%%
%Parallel model run for second ECM evaluation
%Modified geyser inlet water temperature
tic

%Function call for time series generation of input variables
[ var, var2, var3, var4 ] = hhGyserModInW( allhHourConMat, hHourTemp,...
    TwhhBas, TwhhRp, unixStartDate );

simTime=60*30*size(allhHourConMat,1);    %simulation time
EI=208000000;                            %initial energy in geyser

%Call for simulation
sim('Geiser_model_varInletTemp_65C_NoBla');      %per half hour simulation

%Demand matrix formation
DemandBas=ElecDemBase.signals.values;
DemandRp=ElecDemRp.signals.values;
elecDemandHHourMat=cat(2,allhHourConMat(1:end,1),DemandBas(1:end-1),...
    DemandRp(1:end-1));

clear  var var2 var3 var4 simTime EI;
memory
toc
```

```matlab
%%
%Sampling period statistics calculations for Chapter 5 from simulated
%demand data

DemandComp=[((Demand1)*1.5) ((Demand2)*1.5)];      %for a 3kW geyser element
AnnualDemand=sum(DemandComp)
correlationCoefficient=corrcoef(DemandComp(:,1),DemandComp(:,2))
Covariance=cov(DemandComp(:,1),DemandComp(:,2))
Table=[AnnualDemand(1,1) AnnualDemand(1,2) correlationCoefficient(1,2)...
    Covariance(1,1) Covariance(1,2) Covariance(2,1) Covariance(2,2)]

clear AnnualDemand Covariance Demand1 Demand2 DemandComp...
    correlationCoefficient
%%
%data search engine for half hour data
tic

%Data window selection
yArr=2001:2009;              %selected years

mnArr=[11 12 1 2 3 4];       %selected months

dWeekArr=[1 2 4 5 6 7 8];    %selected day of week

hPerArr=1:48;                %selected half hour periods (1st period 00h00)

%Function call for selected data
[ unixNumHH ] = unixSearchNumHH( yArr, mnArr, dWeekArr, hPerArr );

clear yArr mnArr dWeekArr hPerArr;

memory
toc
%%
%Search for selected temperature data
tic

%Function call for selected temperature data
[ selTempData ] = selectedData( hHourTemp, unixNumHH );

%Analysis and boxplotting of temperature data
temp1=reshape(selTempData(:,2),31,[]);
boxplot(temp1);

memory
toc

%%
```

173

```matlab
%Search for selected electrical demand data
tic

%Function call for selected demand data
[ selDemData ] = selectedData( elecDemandHHourMat, unixNumHH);

memory
toc
%%
%Search for hot water consumption data
tic

%Function call for selected hot water consumption data
[ selConData ] = selectedData( allhHourConMat, unixNumHH );

% Analysis of hot water consumption data
seasonalCon=((sum((reshape(selConData(:,2)',48,[])),2))/...
    (size(selConData,1)/48)).*(60*60);

%Function call for figure generation
configure1(seasonalCon);

%distribution fitting
[~,~,d]=find(selConData(:,2));

memory
toc

%%
%Function call for plotting of the selected consumption data
plotConData=reshape(selConData(:,2)',48,[]);

%%
%Function call for plotting of the selected temperature data
plotTempData=reshape(selTempData(:,2)',48,[]);

%%
%Function call for plotting of the selected normalized demand data
plotDemandData=reshape(normDemData(1:end)',48,[]);

%%
```

174

```matlab
%Chapter 5 information generating cells
tic

%Table generating data of the base and reporting periods of
%yearly electrical demand in kWh
Data=elecDemandHHourMat;

%Data window selection
yArr=2001:2009;             %selected years

mnArr=[11 12 1 2 3 4];      %selected months

dWeekArr=[1 2 4 5 6 7 8];   %selected day of week

hPerArr=1:48;               %selected half hour periods (1st period 00h00)

%Function call
[selDataBas,selDataRp] = YearlyDataCollection(Data,yearArr,monthArr,...
    dayArr,hhArr);

%Table population Chapter 5
sumBas=round((sum(selDataBas)).*(1.5));
sumRp=round((sum(selDataRp)).*(1.5));
sav=[sumBas;sumRp;(sumBas-sumRp)];

memory
toc
%%
%plotting of the base and reporting periods of yearly temperature
tic
Data=hHourTemp;

%Data window selection
yArr=2001:2009;             %selected years

mnArr=[11 12 1 2 3 4];      %selected months

dWeekArr=[1 2 4 5 6 7 8];   %selected day of week

hPerArr=1:48;               %selected half hour periods (1st period 00h00)

%Function call
[selDataTemp] = YearlyDataCollectionTemp(Data,yearArr,monthArr,dayArr...
    ,hhArr);

%Statistics
medianTemp=median(selDataTemp);
varTemp=var(selDataTemp);

clear yArr mnArr dWeekArr hPerArr;

memory
toc
%%
```

175

```matlab
% Image plotting of the selected demand data for the base and reporting
% periods (a selected year)
tic

%Data window selection
yArr=2001:2009;              %selected years

mnArr=[11 12 1 2 3 4];       %selected months

dWeekArr=[1 2 4 5 6 7 8];    %selected day of week

hPerArr=1:48;                %selected half hour periods (1st period 00h00)

%Function call for date time stamps
[ unixNumHH ] = unixSearchNumHH( yArr, mnArr, dWeekArr, hPerArr );

%Function call for search for electrical demand data
[ selDemData ] = selectedData( elecDemandHHourMat, unixNumHH);

plotDemandDataBas=reshape(selDemData(1:end,2)',48,[]);
plotDemandDataRp=reshape(selDemData(1:end,3)',48,[]);

%Function call for plotting
demandDataImage(plotDemandDataBas);
demandDataImage(plotDemandDataRp);

clear yArr mnArr dWeekArr hPerArr;

memory
toc
%%
%Histograms plots for on times of the geyser base and reporting periods
distDemDataBas=elecDemandHHourMat(find((elecDemandHHourMat(:,2)) > 0),...
    [1 2]);
distDemDataRp=elecDemandHHourMat(find((elecDemandHHourMat(:,3)) > 0),...
    [1 3]);
%Function call for ploting
histogramOnper(distDemDataBas(:,2));
histogramOnper(distDemDataRp(:,2));

%%
```

176

```matlab
%%
%//////////////////////////////////////////////////////////////////////
%Geyser recovery periods
%//////////////////////////////////////////////////////////////////////
%Window selection
tic

[ elecDemandHHourPer ]=hhPeriod ( elecDemandHHourMat );
[ PreECMrecTime ] = geyserRecover( elecDemandHHourPer(:,[1 2 4]));
[ PostECMrecTime ] = geyserRecover( elecDemandHHourPer(:,[1 3 4]));

toc
%%
%Geyser recovery periods
tic

%applicable year array
yArr=2001:2009;

%start month in year of array
mnArr=[1 2 3 4 5 6 7 8 9 10 11 12];

%day of week selection
dWeekArr=[1 2 3 4 5 6 7 8 9];

%half hour period of the day
hPerArr=1:48;                    %1st period is at 0h00

[ unixNumHH ] = unixSearchNumHH( yArr, mnArr, dWeekArr, hPerArr );


clear yArr mnArr dWeekArr hPerArr;
%Search for temperature data
[ selTempData ] = selectedData( hHourTemp, unixNumHH );

%Search for recovery data
[ PreRecData ] = selectedData( PreECMrecTime, unixNumHH );
[ PostRecData ] = selectedData( PostECMrecTime, unixNumHH );
selRecData=[PreRecData(:,1) PreRecData(:,2) PostRecData(:,2)
PreRecData(:,4)];

toc

%%
%geyser recovery periods
%Elimination of zero recovery times
tic

%selection Criteria from data
%geyser on selection function call
clear bandwidth PDFTempData PDFDemData selPDFbas selPDFrp

model=2;     %Model =2: base line, Model =3: reporting period
```

```matlab
[ PDFTempData, PDFRecData ] = recoverData( selTempData, selRecData ,model );

%Smoothing parameters

%Function call for bandwidth selection pre ECM
% [bandwidth]=kde2d([PDFTempData(:,2) PDFRecData(:,2)]);
% 'preECM bandwidth'
% ht=bandwidth(1,1)
% hd=bandwidth(1,2)

c=corrcoef(PDFRecData(:,2),PDFTempData(:,2))
rho=c(2);
hd = std(PDFRecData(:,2))*(1-(rho)^2)^(5/12)*((1+((rho)^2)/2)^(-
1/6))*(size(PDFRecData,1))^(-1/6)
ht = std(PDFTempData(:,2))*(1-(rho)^2)^(5/12)*((1+((rho)^2)/2)^(-
1/6))*(size(PDFRecData,1))^(-1/6)

[ selPDFbas ] = KernelFunction2( PDFTempData(:,2),PDFRecData(:,2), hd, ht );

clear bandwidth PDFTempData PDFDemData

PDF_figure4(selPDFbas);          %3D plot

model=3;     %Model =2: base line, Model =3: reporting period

[ PDFTempData, PDFRecData ] = recoverData (selTempData, selRecData, model );

%Function call for bandwidth selection post ECM
% [bandwidth]=kde2d([PDFTempData(:,2) PDFRecData(:,2)]);
% 'postECM bandwidth'
% ht=bandwidth(1,1)
% hd=bandwidth(1,2)

c=corrcoef(PDFRecData(:,2),PDFTempData(:,2))
rho=c(2);
hd = std(PDFRecData(:,2))*(1-(rho)^2)^(5/12)*((1+((rho)^2)/2)^(-
1/6))*(size(PDFRecData,1))^(-1/6)
ht = std(PDFTempData(:,2))*(1-(rho)^2)^(5/12)*((1+((rho)^2)/2)^(-
1/6))*(size(PDFRecData,1))^(-1/6)

[ selPDFrp ] = KernelFunction2( PDFTempData(:,2),PDFRecData(:,2), hd, ht );

PDF_figure4(selPDFrp);          %3D plot

PDF_figure5(selPDFbas, selPDFrp);          %image plots

clear bandwidth PDFTempData PDFDemData
memory
toc
%%
```

```
%////////////////////////////////////////////////////////////////////
%Geyser on time PDF generation
%////////////////////////////////////////////////////////////////////
tic

[ preMinOnTime ] = geyserOnTime ( hHourTemp, elecDemandHHourMat(:,[1 2]) );
[ postMinOnTime ] = geyserOnTime ( hHourTemp, elecDemandHHourMat(:,[1 3]) );

memory
toc

%%
%Geyser on time PDF generation
tic

%applicable year array
yArr=2001:2009;

%start month in year of array
mnArr=[5 6 7 8 9 10];

%day of week selection
dWeekArr=[1 2 4 5 6 7 8];

%half hour period of the day
hPerArr=1:48; %This must be retained as default values

[ unixNumHH ] = unixSearchNumHH( yArr, mnArr, dWeekArr, hPerArr );

clear yArr mnArr dWeekArr hPerArr;

[ selPreMinOnMat ] = selectedData( preMinOnTime, unixNumHH );
[ selPostMinOnMat ] = selectedData( postMinOnTime, unixNumHH );

[ selPreMinDataArr ] = minArr ( selPreMinOnMat );
[ selPostMinDataArr ] = minArr ( selPostMinOnMat );

selTempData=selPreMinDataArr(:,1);
selDemData=[selPreMinDataArr(:,1) selPreMinDataArr(:,2)
selPostMinDataArr(:,2) selPreMinDataArr(:,3)];

clear elecDemandHHourMat hHourTemp postMinOnTime preMinOnTime
selPostMinDataArr selPostMinOnMat...
    selPreMinDataArr selPreMinOnMat unixNumHH

memory
toc
%%
```

179

```matlab
%PDF generation for temperature verses minute periods for geyser on
tic
%selection Criteria from data
clear bandwidth PDFTempData PDFDemData
model=2;      %Model =2: base line, Model =3: reporting period

[ PDFTempData, PDFDemData ] = geyserOn (selTempData, selDemData, model );

%Smoothing parameters
%Function call for bandwidth selection pre ECM
% [bandwidth]=kde2d([PDFTempData(:,1) PDFDemData(:,2)]);
% 'preECM bandwidth'
% ht=bandwidth(1,1)
% hd=bandwidth(1,2)

c=corrcoef(PDFDemData(:,2),PDFTempData(:,1))
rho=c(2);
hd = std(PDFDemData(:,2))*(1-(rho)^2)^(5/12)*((1+((rho)^2)/2)^(-
1/6))*(size(PDFDemData,1))^(-1/6)
ht = std(PDFTempData(:,1))*(1-(rho)^2)^(5/12)*((1+((rho)^2)/2)^(-
1/6))*(size(PDFDemData,1))^(-1/6)

[ selPDFbas ] = KernelFunction1( PDFTempData(:,1),PDFDemData(:,2), hd, ht );

clear bandwidth PDFTempData PDFDemData %hd ht

PDF_figure2(selPDFbas);          %3D plot

model=3;      %Model =2: base line, Model =3: reporting period

[ PDFTempData, PDFDemData ] = geyserOn (selTempData, selDemData, model );

%Function call for bandwidth selection post ECM
% [bandwidth]=kde2d([PDFTempData(:,2) PDFDemData(:,2)]);
% 'postECM bandwidth'
% ht=bandwidth(1,1)
% hd=bandwidth(1,2)

c=corrcoef(PDFDemData(:,2),PDFTempData(:,1))
rho=c(2);
hd = std(PDFDemData(:,2))*(1-(rho)^2)^(5/12)*((1+((rho)^2)/2)^(-
1/6))*(size(PDFDemData,1))^(-1/6)
ht = std(PDFTempData(:,1))*(1-(rho)^2)^(5/12)*((1+((rho)^2)/2)^(-
1/6))*(size(PDFDemData,1))^(-1/6)

[ selPDFrp ] = KernelFunction1( PDFTempData(:,1),PDFDemData(:,2), hd, ht );

PDF_figure2(selPDFrp);         %3D plot

PDF_figure3(selPDFbas, selPDFrp);        %image plots

%clear bandwidth PDFTempData PDFDemData
memory
toc
```

```
%//////////////////////////////////////////////////////////////////////////
%//////////////////////////////////////////////////////////////////////////
Functions

function y = inpaintn(x,n,y0)
%
test4DCTNandIDCTN

if nargin==0&&nargout==0, RunTheExample, return, end

x = double(x);
if nargin==1, n = 100; end

sizx = size(x);
d = ndims(x);
Lambda = zeros(sizx);
for i = 1:d
    siz0 = ones(1,d);
    siz0(i) = sizx(i);
    Lambda = bsxfun(@plus,Lambda,...
        cos(pi*(reshape(1:sizx(i),siz0)-1)/sizx(i)));
end
Lambda = -2*(d-Lambda);

% Initial condition
W = isfinite(x);
if nargin==3
    y = y0;
    s0 = 0;
else
    if any(~W(:))
        [y,s0] = InitialGuess(x,isfinite(x));
    else
        y = x;
        return
    end
end
x(~W) = 0;

% Smoothness parameters: from high to negligible values
s = logspace(s0,-3,n);

RF = 2; % relaxation factor
Lambda = Lambda.^2;

h = waitbar(0,'Inpainting...');
for i = 1:n
        Gamma = 1./(1+s(i)*Lambda);
        y = RF*idctn(Gamma.*dctn(W.*(x-y)+y)) + (1-RF)*y;
        waitbar(i/n,h)
end
close(h)
```

```matlab
y(W) = x(W);

end

%% Test for DCTN and IDCTN
function test4DCTNandIDCTN
    if ~exist('dctn','file')
        error('MATLAB:smoothn:MissingFunction',...
            ['DCTN    and    IDCTN    are    required.    Download    DCTN    <a
href="matlab:web(''',...
            'http://www.biomecardio.com/matlab/dctn.html'')">here</a>.'])
    elseif ~exist('idctn','file')
        error('MATLAB:smoothn:MissingFunction',...
            ['DCTN    and    IDCTN    are    required.    Download    IDCTN    <a
href="matlab:web(''',...
            'http://www.biomecardio.com/matlab/idctn.html'')">here</a>.'])
    end
end


%% Initial Guess
function [z,s0] = InitialGuess(y,I)

if license('test','image_toolbox')
    %-- nearest neighbor interpolation
    [z,L] = bwdist(I);
    z = y;
    z(~I) = y(L(~I));
    s0 = 3;
else
    warning('MATLAB:inpaintn:InitialGuess',...
        ['BWDIST (Image Processing Toolbox) does not exist. ',...
        'The initial guess may not be optimal; additional',...
        ' iterations can thus be required to ensure complete',...
        ' convergence. Increase N value if necessary.'])
    z = y;
    z(~I) = mean(y(I));
    s0 = 6;
end

end

function y = jcbresample(tnew,yori,tori)
% jcbresample Changes the sampling rate of a signal.
% y = jcbresamle(tin, yori,tori) resamples the sequence in vector yori at
% time tori to samples at tnew
u = 1;

l = length(tnew);
r  = size(yori,1);
y = zeros(l,r(1));
yori = yori';

for i=1:l
    while tnew(i)>tori(u)
```

```matlab
        u = u+1;
    end
     if tnew(i) == tori(1)
        y(1,:) = yori(1,:);
     else
        y(i,:)    =    ((yori(u,:)-yori(u-1,:))/(tori(u)-tori(u-1)))*(tnew(i)-
tori(u-1))+yori(u-1,:);
     end
end

function [ var,var2,var3 ] = hhGyserMod(allhHourConMat,hHourTemp,Twhh,...
    unixStartDate)
%Variables preparation for gyser model simulation
format long

% Time series transformation
hHourTempTs=timeseries(hHourTemp(1:end,2),1:1800:1800*(size(hHourTemp,1)));


hHourTempTs.TimeInfo.StartDate=datestr(unixStartDate);
hHourTempTs.TimeInfo.Units='seconds';
hHourTempTs.DataInfo.Unit='degrees celcius';

TwhhTs=timeseries(Twhh(1:end),1:1800:1800*(size(Twhh,1)));

TwhhTs.TimeInfo.StartDate=datestr(unixStartDate);
TwhhTs.TimeInfo.Units='seconds';
TwhhTs.DataInfo.Unit='degrees celcius';

%time series generation
allhHourConMatTs=timeseries(allhHourConMat(1:end,2),1:1800:1800*(size(allhHou
rConMat,1)));

allhHourConMatTs.TimeInfo.StartDate=datestr(unixStartDate);
allhHourConMatTs.TimeInfo.Units='seconds';
allhHourConMatTs.DataInfo.Unit='liters per minute';

%Import structures to simulink for half hour input data simulation
t1=allhHourConMatTs.Time(1:end);
t2=hHourTempTs.Time(1:size(t1,1));
t3=TwhhTs.Time(1:size(t1,1));

var.time = t1;
var2.time = t2;
var3.time = t3;

var.signals.values=allhHourConMatTs.Data(1:size(t1,1));
var2.signals.values=hHourTempTs.Data(1:size(t1,1));
var3.signals.values=TwhhTs.Data(1:size(t1,1));
clear allhHourConMatTs hHourTempTs TwhhTs t1 t2 t3;

end
```

```matlab
function [ hHourConMatWd,hHourConMatWe,allhHourConMat] = ...
    HotWgenData2( evalY,firstDm,NOY)
%Hot water consumption generator
%Generate weekday and weekend consumption pattern
%Variables
Years=evalY;
firstDataMonth=firstDm;
numberOfY=NOY;

%Year month calendar generator
ca=cal(Years,firstDataMonth,12);

%Week day profile
selWd=cat(1,ca(:,[1 2 4]),ca(:,[1 2 5]),ca(:,[1 2 6]),ca(:,[1 2 7]),...
    ca(:,[1 2 8]));
wdDt=zeros(1,3);
for i=1:size(selWd,1)
    if selWd(i,3)==0
        continue;                       %removing all zero values
    end
        wdDt=cat(1,wdDt,selWd(i,:));
end

wdDt=wdDt(2:end,:);
dn=datenum(wdDt);
wdDt=cat(2,dn,wdDt);
wdDt=sortrows(wdDt,1);

%Weekend profile
selWe=cat(1,ca(:,[1 2 3]),ca(:,[1 2 9]));
weDt=zeros(1,3);
for i=1:size(selWe,1)
    if selWe(i,3)==0
        continue;                        %removing all zero values
    end
        weDt=cat(1,weDt,selWe(i,:));
end

weDt=weDt(2:end,:);
dn=datenum(weDt);
weDt=cat(2,dn,weDt);
weDt=sortrows(weDt,1);

%random number distribution variables
%Hot water consumption generator using recorded average data
%Hot water consumption for low, medium, high density houses
%Average recorded data input
%Litres per person per hour

np=6.2;                          %number of people in house hold

%Import of hot water demand data
inHwMat=txt2mat('hotWaterData.txt');
```

184

```matlab
%Setting up profile arrays
wdSum=inHwMat(3,:);
wdSum=wdSum*np./3600;


weSum=inHwMat(6,:);
weSum=weSum*np./3600;


wdWin=inHwMat(9,:);
wdWin=wdWin*np./3600;


weWin=inHwMat(12,:);
weWin=weWin*np./3600;


SumSD=17;                                   %percent standard deviation summer
WinSD=30;                                   %percent standard deviation winter

%Summer and winter week day count
s=0;
w=0;

for i=1:size(wdDt,1)

    switch wdDt(i,3)
        case {11 12 1 2 3 4}      %summer
            s=s+1;

        case{5 6 7 8 9 10}        %winter
            w=w+1;

        otherwise
            disp('Error:invalid month allocation');
    end
end


%Weekdays summer
%Data expation using statistical distribution
%Resampled to half hour distributions
%y = jcbresample(tnew,yori,tori)
%Summer
[ wdSum48 ] = (jcbresample(1:0.5:24, wdSum, 1:24))'; % Resampled to half
%hour periods
wdSum48(1,size(wdSum48,2)+1) = (wdSum(1,1)+wdSum(1,end))/2; % appending
%the final sample as a result of time rapping
wdSumMat=zeros(s,size(wdSum48,2));
for i=1:size(wdSum48,2)
    wdSumMat(:,i)=wdSum48(1,i)+((wdSum48(1,i)*(SumSD/100)).*randn(s,1));
end

%Weekdays winter
[ wdWin48 ] = (jcbresample(1:0.5:24, wdWin, 1:24))'; % Resampled to half
%hour periods

wdWin48(1,size(wdWin48,2)+1) = (wdWin(1,1)+wdWin(1,end))/2; % appending
%the final sample as a result of time rapping
```

185

```matlab
wdWinMat=zeros(w,size(wdWin48,2));

for i=1:size(wdWin48,2)
    wdWinMat(:,i)=wdWin48(1,i)+((wdWin48(1,i)*(WinSD/100)).*randn(w,1));
end

%per weekday hot water consumption profile
conProfileWd=zeros(1,4+24*2);
s=0;
w=0;

for i=1:size(wdDt,1)
    profile=zeros(1,24*2);
    switch wdDt(i,3)
        case {11 12 1 2 3 4}               %summer
            s=s+1;
        profile=wdSumMat(s,:);
        case {5 6 7 8 9 10}                %winter
            w=w+1;
        profile=wdWinMat(w,:);
        otherwise
            disp('Error:invalid month allocation');
    end
    dayProfileWd=cat(2,wdDt(i,:),profile);
    conProfileWd=cat(1,conProfileWd,dayProfileWd);
end
conProfileWd=conProfileWd(2:end,:);

%per weekend day hot water consumption profile
%Summer and winter weekend day count
s=0;
w=0;

for i=1:size(weDt,1)
    switch weDt(i,3)
        case {11 12 1 2 3 4}         %summer
            s=s+1;
        case {5 6 7 8 9 10}          %winter
            w=w+1;
        otherwise
            disp('Error:invalid month allocation');
    end
end

%Weekend day summer
%Data exspation using statistical distribution
%Resampled to half hour distributions
%y = jcbresample(tnew,yori,tori)

[ weSum48 ] = (jcbresample(1:0.5:24, weSum, 1:24))'; % Resampled to half
%hour periods
weSum48(1,size(weSum48,2)+1) = (weSum(1,1)+weSum(1,end))/2;

weSumMat=zeros(s,size(weSum48,2));
```

186

```matlab
for i=1:size(weSum48,2)
    weSumMat(:,i)=weSum48(1,i)+((weSum48(1,i)*(SumSD/100)).*randn(s,1));
end

%Weekend days winter
[ weWin48 ] = (jcbresample(1:0.5:24, weWin, 1:24))'; % Resampled to half
%hour periods
weWin48(1,size(weWin48,2)+1) = (weWin(1,1)+weWin(1,end))/2;

weWinMat=zeros(w,size(weWin48,2));

for i=1:size(weWin48,2)
    weWinMat(:,i)=weWin48(1,i)+((weWin48(1,i)*(WinSD/100)).*randn(w,1));
end

conProfileWe=zeros(1,4+24*2);

s=0;
w=0;

for i=1:length(weDt)
    profile=zeros(1,24*60);
    switch weDt(i,3)
        case {11 12 1 2 3 4}
            s=s+1;
            profile=weSumMat(s,:);
        case {5 6 7 8 9 10}
            w=w+1;
            profile=weWinMat(w,:);
        otherwise
            disp('Error:invalid month allocation');
    end
    dayProfileWe=cat(2,weDt(i,:), profile);
    conProfileWe=cat(1,conProfileWe,dayProfileWe);
end
conProfileWe=conProfileWe(2:end,:);

clear  SumSD WinSD Years ca dayProfileWd dayProfileWe dn firstDataDay ...
    firstDataMonth firstDataYear i numberOfY profile s selWd selWe w wdDt...
    wdSum wdSum48 wdSumMat wdWin wdWin48 wdWinMat weDt weSum weSum48...
    weSumMat weWin weWin48 weWinMat;
hour=reshape(cat(1,0:23,0:23),[],1);
min=reshape(cat(1,zeros(1,24),ones(1,24)*30),[],1);
second=zeros(48,1);

%Half hour hot water consumption data for week day
hHourConMatWd=zeros(1,2);
l=size(conProfileWd);

for i=1:l(1)
    hHourCon=conProfileWd(i,5:52);

date=cat(2,ones(48,1)*conProfileWd(i,2),ones(48,1)*conProfileWd(i,3),ones(48,
1)*conProfileWd(i,4));
```

187

```matlab
    unixNum=datenum(cat(2,date,hour,min,second));
    hHourConMat=cat(2,unixNum,hHourCon');
    hHourConMatWd=cat(1,hHourConMatWd,hHourConMat);
end


hHourConMatWd=hHourConMatWd(2:end,:);


clear l hHourCon date unixNum hHourConMat;


%Half hour hot water consumtion data for weekend
hHourConMatWe=zeros(1,2);
l=size(conProfileWe);


for i=1:l(1)
    hHourCon=conProfileWe(i,5:52);

date=cat(2,ones(48,1)*conProfileWe(i,2),ones(48,1)*conProfileWe(i,3),ones(48,
1)*conProfileWe(i,4));
    unixNum=datenum(cat(2,date,hour,min,second));
    hHourConMat=cat(2,unixNum,hHourCon');
    hHourConMatWe=cat(1,hHourConMatWe,hHourConMat);
end


hHourConMatWe=hHourConMatWe(2:end,:);


allhHourConMat=cat(1,hHourConMatWd,hHourConMatWe);
allhHourConMat=sortrows(allhHourConMat,1);
allhHourConMat=(allhHourConMat);%


clear conProfileWd conProfileWe date hHourCon hHourConMat hour i l min second
unixNum;


end

function [ PDFTempData, PDFDemData ] = recoverData( selTempData, selDemData
,model )
%Criteria selection function for kernel function input

PDFDemData=selDemData(find((selDemData(:,model)) > 0),[1 model]); %

PDFTempData=selTempData(find((selDemData(:,model)) > 0),:);
end
```

```matlab
function [ selPDF ] = KernelFunction1( selVar1, selVar2, h2, h1 )
%Kernel Density function
minVar1=0;
maxVar1=35;
stepVar1=0.1;
minVar2=0;
maxVar2=1440;
stepVar2=1;
selPDF=zeros(((maxVar2-minVar2)/stepVar2),((maxVar1-minVar1)/stepVar1));

r=1;
for var2=minVar2:stepVar2:maxVar2                    %for loop for half
hourly demand
    c=1;
    for var1=minVar1:stepVar1:maxVar1               %for loop for
temperature
        pdfValue=0;
        for hh=1:size(selVar2,1)
            pdfValue = pdfValue + (exp(-(((var2-
selVar2(hh))^2)/(2*((h2)^2))+(((var1-selVar1(hh))^2)/(2*((h1)^2))))))...
                /(2*pi*size(selVar2,1)*h2*h1);
        end
        selPDF(r,c)=pdfValue;
        c=c+1;
    end
    floor(r/((maxVar2-minVar2)/stepVar2)*100)
    r=r+1;
end

function [ selPDF ] = KernelFunction2( selVar1, selVar2, h2, h1 )
%Kernel function
minVar1=0;
maxVar1=35;
stepVar1=0.1;
minVar2=30;
maxVar2=65;
stepVar2=0.1;
selPDF=zeros(((maxVar2-minVar2)/stepVar2),((maxVar1-minVar1)/stepVar1));

r=1;
for var2=minVar2:stepVar2:maxVar2                    %for loop for half
hourly demand
    c=1;
    for var1=minVar1:stepVar1:maxVar1               %for loop for
temperature
        pdfValue=0;
        for hh=1:size(selVar2,1)
            pdfValue = pdfValue + (exp(-(((var2-
selVar2(hh))^2)/(2*((h2)^2))+(((var1-selVar1(hh))^2)/(2*((h1)^2))))))...
                /(2*pi*size(selVar2,1)*h2*h1);
        end
        selPDF(r,c)=pdfValue;
        c=c+1;
    end
    floor(r/((maxVar2-minVar2)/stepVar2)*100)
```

189

```matlab
    r=r+1;
end


selPDF=selPDF';


end


function [ timeData ] = selectedData( dataMat, unixNumHH )
%Unix numbers from search matrix is used to obtain demand,
%temperature and consumption data
format long g

%dataMat=hHourTemp;
timeData=zeros(length(unixNumHH),size(dataMat,2));

for i=1:size(unixNumHH,1)
        timeData(i,:)=dataMat((find((dataMat(:,1))==unixNumHH(i,1))),:);

end


end



function [Twmin Twhh ] = InWaterTemp( AmbientTemp )
%The calculation of the average soil temperature/water temperature at a
%Depth of water pipe in soil
z=7;

%Dampining depth of soil type
d=2.093;

%Number of days in the year from Jan 1st till annual min temp is recorded
y=[2001 2002 2003 2004 2005 2006 2007 2008 2009];
for a=1:size(y,2)
    YearNum(a)=datenum(y(a),0,1);
end

row=0;

for a=1:size(YearNum,2)
    [row,~,~]=find(AmbientTemp(:,1)>=YearNum(a)&AmbientTemp(:,1)<...
        (YearNum(a)+365));
    YearTemp(:,a)=AmbientTemp(row,2);
    [c,i]=min(YearTemp(:,a));
    DayOfYear(1,a)=floor(i/48);
end

t0=floor(sum(DayOfYear,2)/size(DayOfYear,2));

%Average ambient temperature
Ta=sum(AmbientTemp(:,2))/size(AmbientTemp,1)+2;     %add 2 degrees C

%Annual amplitude
```

190

```matlab
[row,~,~]=find(AmbientTemp(:,1)>=YearNum(1));
AllTemp(:,1)=AmbientTemp(row,2);
DayTemp=reshape(AllTemp,48,[]);
MinDayTemp=min(DayTemp);
MaxDayTemp=max(DayTemp);
A0=((sum(MaxDayTemp,2)/size(MaxDayTemp,2))-(sum(MinDayTemp,2)/...
    size(MinDayTemp,2)))/2;

%Water temperature function
for t=1:size(AllTemp,1)/48+1
Tw(1,t)= Ta+A0*exp(-z/d)*sin((2*pi*(t-t0))/365-(z/d)-(pi/2));
end

%resample from average daily temperature to perhalf hour temperatures
% y = jcbresamle(tin, yori,tori) resamples the sequence in vector yori at
% time tori to samples at tnew
[Twhh]=jcbresample(1:1/48:size(Tw,2), Tw, 1:size(Tw,2));
Twhh=Twhh(1:end-1);
[Twmin]=jcbresample(1:1/(48*30):size(Tw,2), Tw, 1:size(Tw,2));
Twmin=Twmin(1:end-1);


end

function [unixNumHH]=unixSearchNumHH(yArr,mnArr,dWeekArr,hPerArr)
%search for the required unix numbers
calMat=zeros(1,9);
for y=yArr
    for m=mnArr
        calMat=cat(1,calMat,cat(2,ones(6,1)*y,ones(6,1)*m,calendar(y,m)));
    end
end

calMat=calMat(2:end,:);
selMat=calMat(:,dWeekArr);
[r,~,v]=find(selMat(:,(3:(size(dWeekArr,2)))));
yr=selMat(r(:),1);
mn=selMat(r(:),2);
dy=v(:);

selDateMat=cat(2,yr,mn,dy);
unixDate=datenum(selDateMat);

hr=reshape(cat(1,0:23,0:23),[],1);
mn=reshape(cat(1,zeros(1,24),ones(1,24)*30),[],1);
sc=zeros(48,1);

timeMat=cat(2,hr,mn,sc);
selTimeMat=timeMat(hPerArr,:);
unixTime=datenum(cat(2,zeros(size(selTimeMat,1),3),selTimeMat));

unixNumHH=zeros(length(unixDate),length(unixTime));

for i=1:length(unixDate)
   unixNumHH(i,:)=unixDate(i)+unixTime;
end
```

191

```matlab
unixNumHH=reshape(unixNumHH,[],1);
unixNumHH=sort(unixNumHH);

clear calMat dWeekArr dy hPerArr hr i m mn mnArr r sc selDateMat selMat...
    selTimeMat timeMat unixDate unixTime v y yArr yr;
end

function            [minTemp,hHourTemp]            =            dataMatrix(
inMat,unixStartDate,unixEndDate,firstDataYear,firstDataMonth,firstDataHourCol
om,dayOfMonthColom)
%Time stamp allocation to data
format long;

countMonth=firstDataMonth-1;
countYear=firstDataYear;
dateMat=cat(2,[firstDataYear                              firstDataMonth
inMat(1,dayOfMonthColom)],inMat(1,firstDataHourColom:firstDataHourColom+23));

for i=1:length(inMat(:,1))
    if isnan(inMat(i,1))==1
        continue
    end
    if inMat(i,1)==1
        countMonth=countMonth+1;
        if countMonth==13
            countYear=countYear+1;
            countMonth=1;
        end
    end
    loopDate=[countYear                    countMonth                    inMat(i,1)
inMat(i,firstDataHourColom:firstDataHourColom+23)];
    dateMat=cat(1,dateMat,loopDate);
end

dateTimeMat=dateMat(2:length(dateMat),:);

%Unix number temperature array
tempMat=(dateTimeMat(1:end,4:end))';
outMat=zeros(1,8);

for i=1:length(dateTimeMat)
    date=ones(24,1)*(dateTimeMat(i,1:3));
    hour=(0:23)';
    min=zeros(24,1);
    sec=zeros(24,1);
    unixNum=datenum(cat(2,date,hour,min,sec));
    monthTemp=cat(2,date,hour,min,sec,unixNum,tempMat(1:24,i));
    outMat=cat(1,outMat,monthTemp);
end

outMat=outMat(2:(unixEndDate-unixStartDate)*24+2,:);

% Interpolating the missing temperature data
outMat(:,8)=inpaintn(outMat(:,8));
```

192

```matlab
% Resampling
hHourTemp=cat(2,(outMat(1,7):datenum(0,0,0,0,30,0):outMat(size(outMat,1),7))'
,...
    (jcbresample(1:0.5:size(outMat,1),outMat(1:end,8)',1:1:size(outMat,1))));
hHourTemp=hHourTemp(1:end-1,1:2);

minTemp=cat(2,(outMat(1,7):datenum(0,0,0,0,1,0):outMat(size(outMat,1),7))',..
.

(jcbresample(1:1/60:size(outMat,1),outMat(1:end,8)',1:1:size(outMat,1))));
minTemp=minTemp(1:end-1,1:2);

clear date dateMat dateTimeMat hour i loopDate min monthTemp sec tempMat...
    unixNum outMat inMat;

end



function [ calMat ] = cal( ya,sm,nm)
%Callender search
calMat=[];

for c=ya
    smMat=calendar(c,sm);
    yMat=ones(6,1)*c;
    mMat=ones(6,1)*sm;
    smMat=cat(2,yMat,mMat,smMat);
    calMat=cat(1,calMat,smMat);
    if nm==1
        continue;
    else
        for i=1:nm-1
            m=sm+i;
            y=c;
            if m>12
                y=c+1;
                yMat=ones(6,1)*y;
                m=m-12;
            end
            mMat=ones(6,1)*m;
            nmMat=calendar(y,m);
            nmMat=cat(2,yMat,mMat,nmMat);
            calMat=cat(1,calMat,nmMat);
        end
    end
end

end

function [ PDFTempData, PDFDemData ] = geyserOn( selTempData, selDemData
,model )
%Criteria selection function for kernel function input
```

193

```matlab
PDFDemData=selDemData(find((selDemData(:,model)) > 0),[1 4]); %

PDFTempData=selTempData(find((selDemData(:,model)) > 0),:);

end

function                       [selDataBas,selDataRp]                    =
YearlyDataCollection(Data,yearArr,monthArr,dayArr,hhArr)

selDataBas = zeros(17520,size(yearArr,2));
selDataRp = zeros(17520,size(yearArr,2));

for a=1:size(yearArr,2)

    %applicable year array
    yArr=yearArr(a);

    %start month in year of array
    mnArr=monthArr;

    %day of week selection
    dWeekArr=dayArr;

    %half hour period of the day
    hPerArr=hhArr;                  %1st perod is at 0h00

    [ unixNumHH ] = unixSearchNumHH( yArr, mnArr, dWeekArr, hPerArr );

    selDemData=selectedData( Data, unixNumHH);
    selDataBas(:,a) = selDemData(1:17520,2);
    selDataRp(:,a) = selDemData(1:17520,3);

end

function                        [selDataTemp]                            =
YearlyDataCollectionTemp(Data,yearArr,monthArr,dayArr,hhArr)

selDataTemp = zeros(17520,size(yearArr,2));


for a=1:size(yearArr,2)

    %applicable year array
    yArr=yearArr(a);

    %start month in year of array
    mnArr=monthArr;

    %day of week selection
    dWeekArr=dayArr;

    %half hour period of the day
    hPerArr=hhArr;                  %1st perod is at 0h00
```

194

```matlab
    [ unixNumHH ] = unixSearchNumHH( yArr, mnArr, dWeekArr, hPerArr );

    selDemData=selectedData( Data, unixNumHH);
    selDataTemp(:,a) = selDemData(1:17520,2);


end

function [ selMinData ] = minArr ( selMinOnMat )
%Geyser on function, per minute PDF simulations

temp=zeros(size(selMinOnMat,1),30);

for col=1:30
    temp(:,col)=(selMinOnMat(:,2));
end

temp=reshape(temp',[],1);

onPos=reshape((selMinOnMat(:,3:32))',[],1);

minPer=(1:1440);
period=zeros(size(minPer',1),(size(temp,1)/1440));
for col=1:(size(temp,1)/1440)
    period(:,col)=minPer';
end

period=reshape(period,[],1);

selMinData=cat(2,temp,onPos,period);

end

function [ minOnTime ] = geyserOnTime( temp, demand )
%generation of per minute on values
minOnTime=zeros(size(temp,1),32);

for pos=1:size(demand,1);
    minOnTime(pos,[1 2])=temp(pos,[1 2]);
    if pos==size(demand,1)
        break;
    end

    val=(demand(pos,2));
    next_val=(demand(pos+1,2));

    if (val==0)&&(next_val==0)
        continue
    elseif (val==0)&&(next_val>0)
        minOnTime(pos+1,(round(33-(next_val*30))):32))=1;
        continue
    elseif (val>0)&&(next_val>0)
```

195

```matlab
        minOnTime(pos+1,3:(round(next_val*30)+2))=1;
        continue
    else                    %(val>0)&&(next_val==0)
        continue
    end
end
end


function [ recoveryTime ] = geyserRecover( selDemDataECM )
%Density estimation of recovery time of geyser

recoveryTime=zeros(size(selDemDataECM,1),4);
onTime=0;
recTime=0;
for pos=1:size(selDemDataECM,1);
    recoveryTime(pos,[1 4])=selDemDataECM(pos,[1 3]);
    if pos==size(selDemDataECM,1)

        break;
    end

    recTime=0;
    val=(selDemDataECM(pos,2));
    next_val=(selDemDataECM(pos+1,2));

    if (val==0)&&(next_val==0)
        continue
    elseif (val==0)&&(next_val>0)
        onTime=(onTime+next_val);
        continue
    elseif (val>0)&&(next_val>0)
        onTime=(onTime+next_val);
        continue
    else                    %(val>0)&&(next_val==0)
        recTime=onTime;
        onTime=0;
    end
    recoveryTime(pos,[2 3])=[(recTime*30) (val*30)];
end

end

%%
%Create a monthly box plot of temperature data from 2001 to 2009
tic

ti={'January 00h00','January 06h00','January 12h00','January 18h00';...
    'February 00h00','February 06h00','February 12h00','February 18h00';...
    'March 00h00','March 06h00','March 12h00','March 18h00';...
    'April 00h00','April 06h00','April 12h00','April 18h00';...
    'May 00h00','May 06h00','May 12h00','May 18h00';...
    'June 00h00','June 06h00','June 12h00','June 18h00';...
    'July 00h00','July 06h00','July 12h00','July 18h00';...
    'August 00h00','August 06h00','August 12h00','August 18h00';...
```

```matlab
    'September    00h00','September    06h00','September    12h00','September
18h00';...
    'October 00h00','October 06h00','October 12h00','October 18h00';...
    'November 00h00','November 06h00','November 12h00','November 18h00';...
    'December 00h00','December 06h00','December 12h00','December 18h00'};

for i=1:size(tempData,2)

temp=reshape(tempData(:,i),numMn,[]);

subplot(1, 2, i);

boxplot(temp,'notch','on');

set(gca,'YGrid','on',...
    'XTickLabel',{'2001','2002','2003','2004','2005','2006','2007','2008',...
    '2009'}, 'XTick',[1 2 3 4 5 6 7 8 9],...
    'YTick',[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 ...
    23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40],...
    'FontSize',11,...
    'FontName','Times New Roman');

xlim([0.5 9.5]);

ylim([4 29]);

box('on');

% Create xlabel
xlabel('Years','Units','points','FontSize',12,'FontName','Times New Roman');

% Create ylabel
ylabel('Temperature','FontSize',12,'FontName','Times New Roman');

% Create title
title(ti(mnArr,i+2),'FontSize',12,'FontName','Times New Roman');

end

memory
toc
%%
%Creating monthly box plots of temperature over 48 half hour time periods
tic

%ti={'January','February','March','April','May','June','July',...
  %  'August','September','October','November','December'};

boxplot(tempData);

set(gca,'YGrid','on',...
    'XGrid','on',...
    'FontSize',20,...
    'FontName','Times New Roman');
```

197

```matlab
ylim([0 40]);

box('on');

% Create xlabel
xlabel('Half hour time periods','FontSize',20,'FontName','Times New Roman');

% Create ylabel
ylabel('Ambient temperature [degrees Celsius]','FontSize',20,'FontName',...
    'Times New Roman');

% Create title
title('June','FontSize',20,'FontName','Times New Roman');

memory
toc
%%
%Creat nonparametric distribution plot and histogram
%For four half hour time periods in a month
tic

%Lable matrix
ti={'January 00h00','January 06h00','January 12h00','January 18h00';...
    'February 00h00','February 06h00','February 12h00','February 18h00';...
    'March 00h00','March 06h00','March 12h00','March 18h00';...
    'April 00h00','April 06h00','April 12h00','April 18h00';...
    'May 00h00','May 06h00','May 12h00','May 18h00';...
    'June 00h00','June 06h00','June 12h00','June 18h00';...
    'July 00h00','July 06h00','July 12h00','July 18h00';...
    'August 00h00','August 06h00','August 12h00','August 18h00';...
    'September    00h00','September    06h00','September    12h00','September
18h00';...
    'October 00h00','October 06h00','October 12h00','October 18h00';...
    'November 00h00','November 06h00','November 12h00','November 18h00';...
    'December 00h00','December 06h00','December 12h00','December 18h00'};

tk={'Summer 00h00','Summer 06h00','Summer 12h00','Summer 18h00';...
    'Autumn 00h00','Autumn 06h00','Autumn 12h00','Autumn 18h00';...
    'Winter 00h00','Winter 06h00','Winter 12h00','Winter 18h00';...
    'Spring 00h00','Spring 06h00','Spring 12h00','Spring 18h00'};

% Loop to plot the time periods
for i=1:size(wdSumCon,2)

% Force all inputs to be column vectors
tempD = weSumCon(:,i);

% Set up figure to receive datasets and fits
f_=subplot(2,3,i+3);

legh_ = []; legt_ = {};   % handles and text for legend
ax_ = newplot;
set(ax_,'Box','on','FontSize',16,'FontName','Times New Roman');
```

198

```matlab
grid(ax_,'on');
%xlim([0 30]);
%ylim([0 0.2]);
hold on;

% --- Plot data originally in dataset "January 00h00"
t_ = ~isnan(tempD);
Data_ = tempD(t_);
[F_,X_] = ecdf(Data_,'Function','cdf'...
              );  % compute empirical cdf
Bin_.rule = 1;
[C_,E_] = dfswitchyard('dfhistbins',Data_,[],[],Bin_,F_,X_);
[N_,C_] = ecdfhist(F_,X_,'edges',E_); % empirical pdf from cdf
h_ = bar(C_,N_,'hist');
set(h_,'FaceColor','none','EdgeColor',[0.333333 0 0.666667],...
       'LineStyle','-', 'LineWidth',1);

xlabel('Temperature','FontSize',16,'FontName','Times New Roman');
ylabel('Density','FontSize',16,'FontName','Times New Roman')
legh_(end+1) = h_;
legt_{end+1} = 'Histogram';

% Nudge axis limits beyond data limits
xlim_ = get(ax_,'XLim');
if all(isfinite(xlim_))
   xlim_ = xlim_ + [-1 1] * 0.01 * diff(xlim_);
   set(ax_,'XLim',xlim_)
end

x_ = linspace(xlim_(1),xlim_(2),100);

% --- Create fit "Non-parametric"
t_ = ~isnan(tempD);
Data_ = tempD(t_);
y_ = ksdensity(Data_,x_,'kernel','normal',...
               'support','unbounded',...
               'function','pdf');
h_ = plot(x_,y_,'Color',[1 0 0],...
          'LineStyle','-', 'LineWidth',2,...
          'Marker','none', 'MarkerSize',6);
legh_(end+1) = h_;
legt_{end+1} = 'Kernel est.';
hold off;
leginfo_ = {'Orientation', 'vertical', 'Location', 'NorthEast'};
h_ = legend(ax_,legh_,legt_,leginfo_{:});  % create legend
set(h_,'Interpreter','none');
title(tk(3,i),'FontSize',16,'FontName','Times New Roman');
end

clear tempD Data_ y_ xlim_ x_ t_ legt_ leginfo_ legh_ h_ f_ ax_ X_ N_ F_ ...
    E_  C_ Bin_;
memory
toc
```

199