Evaluating the Uncertainty in the Performance of Small Scale Renewables

A Doctoral Thesis

by
Adam Thirkill

Submitted in partial fulfilment of the requirements for the award Doctor of Philosophy of Loughborough University

March 2015

“Writing is good, thinking is better. Cleverness is good, patience is better.”

Siddhartha**, Siddhartha,** Hermann Hesse

Abstract

The successful delivery of low-carbon housing (both new and retrofitted) is a key aspect of the UK's commitment to an 80% reduction in carbon emissions by 2050. In this context, the inclusion of small-scale building-integrated renewable energy technologies is an important component of low carbon design strategies, and is subject to numerous regulation and incentive schemes (including the Renewable Heat Incentive (RHI)) set up by government to encourage uptake and set minimum performance benchmarks.

Unfortunately there are numerous examples of in-use energy and carbon performance shortfalls for new and retrofitted buildings – this is termed the "performance gap". Technical and human factors associated with building subsystem performance, which are often not considered in design tools used to predict performance, are the root cause of performance uncertainty. The research presented in this doctoral thesis aims to develop and apply a novel probabilistic method of evaluating the performance uncertainty of solar thermal systems installed in the UK.

Analysis of measured data from a group of low carbon retrofitted dwellings revealed that the majority of buildings failed to meet the designed-for carbon emissions target with an average percentage difference of 60%. An in-depth case study technical evaluation of one of these dwellings showed significant dysfunction associated with the combined ASHP/solar thermal heating system, resulting in a performance gap of 94%, illustrating that the performance gap can be regarded as a "whole-system" problem, comprising a number of subsystem causal factors.

Using a detailed dataset obtained from the UK's largest field trial of domestic solar thermal systems, a cross-cutting evaluation of predicted vs. measured performance similarly revealed a discrepancy with a mean percentage difference in predicted and measured annual yield of -24%.

Having defined the nature and extent of underperformance for solar thermal technology in the UK, causal factors influencing performance were mapped and the associated uncertainty quantified using a novel knowledge-based Bayesian network (BN). In addition, the BN approach along with Monte Carlo sampling was applied to the well-established BREDEM model in order to quantify performance uncertainty of solar thermal systems by producing distributions of annual yield. As such, the modified BN-based BREDEM model represents a significant improvement in the prediction of performance of small-scale renewable energy technologies. Finally, financial analysis applied to the probabilistic predictions of annual yield revealed that the current UK RHI scheme is unlikely to result in positive returns on investment for solar thermal systems unless the duration of the payments is extended or electricity is the primary source of heating.

Key words

Domestic Buildings; Solar Thermal Systems; Renewable Energy; Performance Gap; Performance Evaluation; Modelling; Monitoring; Uncertainty; Bayesian Networks; Renewable Heat Incentive

Acknowledgements

I would like to take this opportunity to express my appreciation to those that have not only made this research possible but have also made my time working on one of the most demanding challenges in my life to date as pain free as is reasonable to ask.

Firstly I would like to thank Paul Rowley whose guidance and approach to supervision provided a comfortable, flexible and productive working environment. I would also like to acknowledge the Engineering and Physical Sciences Research Council (EPSRC).

Thanks to my parents for supporting me financially and lending a sympathetic ear during this challenging time. I'd also like to mention my brother, Luke.

In addition to my family I would like to thank my friends and colleagues at CREST for their lively debates and in-depth discussions which provided a valuable distraction from my research. To my friend Ming I'd like to say: I'm finally finished.

Table of Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

ANOVA	Analysis of Variance
ASHP	Air-Source Heat Pump
BEM	Building Energy Model
BN	Bayesian network
BRE	Building Research Establishment
BREDEM	Building Research Establishment Domestic Energy Model
BREHOMES	BRE Housing Model for Energy Studies
CDEM	Community Domestic Energy Model
CM	Causal Map
COP	Coefficient of Performance
CPF	Collector Performance Factor
CPT	Conditional Probability Table
DCF	Discounted Cash Flow
DECarb	Domestic Energy and Carbon Model
DECC	Department of Energy and Climate Change
DHW	Domestic Hot Water
DSA	Differential Sensitivity Analysis
ECO	Energy Company Obligation
EM	Expectation Maximisation
EPC	Energy Performance Certificate
EST	Energy Saving Trust
FiT	Feed in Tariff
FMEA	Failure Mode and Effects Analysis
FMECA	Failure Mode, Effects and Criticality Analysis
FTA	Fault Tree Analysis
GIS	Geographical Information System
GSHP	Ground-Source Heat Pump
MAP	Maximum a-posterior Probability
MC	Monte Carlo
MCS	Micro-generation Certification Scheme
MLE	Maximum Likelihood Expectation
NCM	National Calculation Method
NHBC	National House Building Council

NPV	Net Present Value
OOBN	Object Oriented Bayesian Network
PCA	Principal Component Analysis
PV	Photovoltaic
R4F	Retrofit for the Future
RES	Renewable Energy Strategy
RET	Renewable Energy Technology
RHI	Renewable Heat Incentive
RPI	Retail Price Index
SA	Sensitivity Analysis
SAP	Standard Assessment Procedure
SD	Standard Deviation
SF	Solar Fraction
SoS	System-of-Systems
SPF	Seasonal Performance Factor
SPT	Set Point Temperature
SSA	Stochastic Sensitivity Analysis
STS	Solar Thermal System
TFA	Total Floor Area
TSB	Technology Strategy Board
UKDCM	UK Domestic Carbon Model
UML	Unified Modelling Language
WBCSD	World Business Council for Sustainable Development

Mathematical notations

A_{ap}	Aperture area [m^2]
A_c	Collector area [m^2]
C_0	Capital cost [£]
C_m	Maintenance cost [£]
C_p	Specific heat capacity [J/kg.K]
f_1	Collector performance factor
f_2	Solar storage volume factor
G	Solar Irradiance [kW/m^2]
Gr	Grashof number
h	Enthalpy [J]
I	Net income [£]
k_1	First order heat loss coefficient [$W/m^2.K$]
k_2	Second order heat loss coefficient [$W/m^2.K^2$]
LR	Load ratio
$\dot{m}$	Mass flow rate [kg/s]
Med	Median
MS	Mean square error
N	Number of occupants
n_m	Number of days in month m
P	Probability

Par	Parent node
Q_{ASHP}	Thermal energy delivered by ASHP
Q_{aux}	Auxiliary input [kWh]
Q_{fs}	Fossil fuel energy saved
Q_{HW}	Hot water energy [kWh]
Q_L	Thermal losses [kWh]
Q_{SH}	Energy required for space heating [kWh]
Q_{sol}	Solar yield [kWh]
$\dot{Q}_{sol}$	Rate of heat transfer from the solar coil [W]
r	Discount rate [%]
Re	Reynolds number
R_f	Unit price of fossil fuel [p/kWh]
R_{RHI}	RHI tariff rate [p/kWh]
S	Incident Solar radiation [kWh/m^2]
SS	Sum of the squared error
T_{coil}	Temperature of working fluid in the solar coil [°C]
T_{cold}	Cold water temperature [°C]
T_{ext}	External ambient temperature [°C]
T_{load}	Hot water temperature [°C]
T_{sb}	Temperature of water in the solar buffer [°C]
T_{set}	Set point temperature [°C]
UF	Utilisation factor
VA	Volume-to-collector area ratio [L/m^2]
$V_{d,ave}$	Daily average volumetric consumption of hot water in a year [L/day]
$V_{d,m}$	Daily average volumetric consumption of hot water in month m [L/day]
V_{eff}	Effective solar volume [L]
V_{tank}	Storage tank volume [L]
W_{ASHP}	Electrical energy consumed by ASHP
Z_{panel}	Over-shading factor
β	Thermal expansion coefficient [1/°C]
ΔT	Temperature rise of hot water [°C]
η_0	Zero loss efficiency
η_{sys}	System efficiency
ν_{cold}	Viscosity of cold water [m^2/s]
ν_{hot}	Viscosity of hot water [m^2/s]

Subscripts

m	Month, m/measured value
res, r	Residuals of the restricted regression model
res, u	Residuals of the unrestricted regression model
p	Predicted value
t	Year, t

Chapter 1

Introduction

"There is always an easy solution to every human problem - neat, plausible, and wrong."

The Divine Afflatus**, New York Times (1917),** H.L. Mencken

1.1 Research Context

In response to the Climate Change Act of 2008 the UK is committed to reducing its greenhouse emissions by 80% by 2050 based on 1990 levels. Carbon dioxide (CO_2) is considered to be the most important greenhouse gas, comprising 82% of the UK greenhouse gases emissions in 2013 (DECC 2015), and its reduction is the focus of many strategies employed to meet UK & European emissions targets.

40% of the UK's carbon emissions come from the built environment with domestic buildings being responsible for two-thirds of this (Hall et al. 2013); Figure 1.1 shows housing to account for nearly 30% of the energy consumed in the UK with the majority of this used for space and water heating.

Carbon emission reduction is achieved by addressing the use of fossil fuels in the built environment. Reductions are achieved through reduced energy requirements by way of energy efficient technology or generation of energy using renewable energy technologies (Cherrington et al. 2013; Dowson et al. 2012).

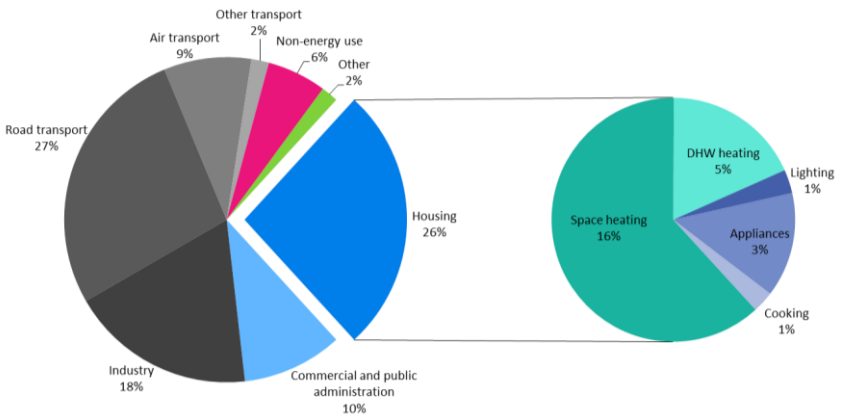

Figure 1.1: Breakdown of UK energy use 2011 (Palmer & Cooper 2013)

Many of the homes in Britain were built when trends in energy usage and expectations of comfort were very different from those of current times: 40 years ago houses had an average indoor temperature of 12°C (Palmer & Cooper 2013). In comparison, modern homeowners own and operate central heating, dishwashers, washing machines, refrigerators on top of which are added flat screen televisions, smart phones, laptops, tablets, and myriad other technologies; therefore reduction in domestic energy consumption and associated carbon is a continually changing challenge. Furthermore, with current housing stock being at around 27.4million homes and rising, the domestic building sector represents a key area in which to reduce carbon emissions.

The UK Government has rolled out numerous policies to address the issue of domestic energy consumption and its associated carbon emissions (Dowson et al. 2012); many of these advocate a "fabric first" approach to improving the thermal performance of a dwelling and thus reducing energy consumed for the purpose of space heating, which accounts for more than 60% of domestic energy consumption (Figure 1.1). New builds are covered by increasingly stringent building regulations which are expected to lead towards zero-carbon constructions, whilst existing housing stock (of which 70-85% is expected to be in use by 2050 (Power 2008; Ravetz 2008)) are targeted with retrofit strategies. The Green Deal is one such strategy and is the UK's main energy efficiency policy for existing homes (HM Government 2013) and will work alongside an Energy Company Obligation (ECO) to provide additional financial support for low income households (Ofgem 2015). Along with the Renewable Energy Strategy (RES) (HM Government 2009) and Micro-generation Certification Scheme (MCS) (MCS 2015), policies that encourage the generation of energy using renewable sources include the Feed in Tariff (FiT) scheme (HM Government 2015), which covers micro-generation of electricity, and the Renewable Heat Incentive (RHI), which is the world's first tariff scheme to pay homeowners for the thermal energy generated by an eligible renewable heating system (HM Government 2014) – fabric improvements identified by a Green Deal assessment are a prerequisite for the RHI.

These policies and incentive mechanisms represent the UK's commitment to reducing carbon and energy consumption associated with housing; however they are only effective if the expected improvements in building energy performance are realised in use. There is a great deal of evidence to indicate that the designed performance of a building (domestic or otherwise) is rarely met in actual operation (CarbonBuzz 2014; Bordass et al. 2004; Majcen et al. 2013; Kelly et al. 2012; Branco et al. 2004; Hens 2010; Cayre et al. 2011; Carbon Trust 2011; Fokaides et al. 2011; Turner & Frankel 2008). Designed performance predictions are obtained through the use of building energy models that project the energy consumption of buildings based on building physics, sub-system performance and energy demands of occupants; however it has been suggested that using designed performance as a predictor for actual performance is unreliable (Scofield 2009).

The discrepancy between the designed or predicted performance and the actual in-use performance is commonly referred to as the "performance gap". Numerous studies into the performance gap of domestic and non-domestic buildings have highlighted several reasons why designed performance is not met in reality: (Torcellini et al. 2004) indicate that estimates of occupancy levels made in the design stage were less than in actual operation and that building system dysfunction due to poorly designed control algorithms were to blame. Building system

dysfunction is a common cause of design estimates of energy consumption not being met (Zero Carbon Hub 2013; Mørck et al. 2012; Gupta & Dantsiou 2013). One cause for concern related to this is the appropriateness of test data related to building system performance used to predict as-built performance; these systems (including building materials) are often tested in isolation rather as integrated systems or assemblies in the field (Zero Carbon Hub 2013). Studies related to the performance of heat pumps systems suggest a general trend for underperformance compared to European counterparts (Boait et al. 2011) as well as wide variation in performance across a group of systems (EST 2010). System complexity was found to influence heat pump performance with simpler designs demonstrating higher performance (EST 2010); the issue of complexity of design has also been linked to whole-building performance discrepancies (Bannister 2009; Carbon Trust 2011). Ballarini & Corrado (2009) found occupant behaviour to be a key cause of discrepancies between estimated and actual energy consumption (Ballarini & Corrado 2009); this is especially the case when occupants are faced with complex control interfaces which may lead to ineffective operation of building sub-systems or a total bypass of the control systems altogether (Bordass et al. 2004).

(Hinge et al. 2008) summarises the causes of the performance gap as being: usage and occupancy patterns being different from design assumptions; sub-systems failing to meet assumed levels of performance; inadequate system commissioning and a lack of knowledge of how to maintain and operate the building properly - current modelling approaches used to make in-use energy predictions at the design stage are limited by these factors and depend on ideal world scenarios. To improve predictions of actual energy consumption it is recommended that in-use performance data is fed back into the models used at the design stage (Turner & Frankel 2008; Menezes et al. 2012; Diamond et al. 2006); however there is a current lack of energy performance data for new and existing housing (Lowe & Oreszczyn 2008). A common approach to using measured data to improve model estimates is by way of model calibration (Reddy 2006). Calibration is a retrospective exercise and therefore may not necessarily provide reliable predictions of performance for future buildings for which calibration data sets do not exist. Furthermore current modelling techniques do not allow for incorporation of uncertainty associated with the many input parameters required to describe the complex system interactions.

The performance gap represents a risk that the designed performance will not be met in actuality. This performance risk leads to financial risk to homeowners wishing to see a return on investment in renewable technologies and other retrofit strategies through tariff payments and reduced energy bills. This is an important consideration since the Green Deal's "golden rule" principle states that the *estimated* savings on energy bills must be equal to or greater than the costs attached to the energy bill thus the golden rule may be difficult to achieve due to the problems associated with accurately predicting the performance of retrofitted dwellings and the tendency of buildings and subsystems to underperform in comparison to design estimates (Dowson et al. 2012; Majcen et al. 2013). Performance risk also leads to issues associated with successful attainment of carbon targets (Majcen et al. 2013). The Retrofit for the Future (R4F) design competition was inspired by the UK's target to reduce carbon emissions by 80% by 2050; data from the competition presented in Figure 1.2 shows that many designs fall short of the 80% target and that measured reduction is often less than predicted.

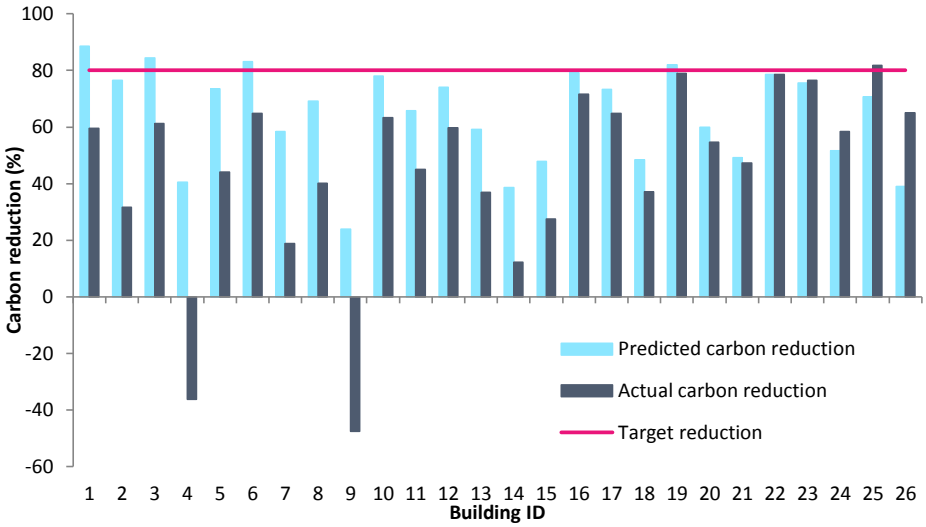

Figure 1.2: Predicted and measured carbon reduction for 26 R4F dwellings

There are several techniques used in engineering applications to manage and mitigate risk. When risk refers to the failure of component or system standard methods employed include fault tree analysis (FTA) and FMEA (or FMECA). In these methods risk is the product of the probability of the risk event occurring and the level of impact the risk event will have if it does occur. The drawback of this definition of risk is that the probability of certain risk events may be unknown due to a lack of data and the level of impact is a subjective quantity – Fenton et al. provide a good description of this in (Fenton & Neil 2013).

Risk is often quoted as a single value, in an ideal world risk would be described by a probability function with a distribution of the probability of failure and the consequences expressed in a meaningful way (INCOSE 2011). In terms of performance risk the probability of failure can be expressed by a probability distribution of energy performance in comparison with a benchmark, be this yield from a renewable energy system or consumption of energy by a building. The probability of failure is the sum of the probabilities for which the energy performance is less than the design estimate or target. Using energy performance as a metric is valuable since this is readily translated into monetary and carbon terms. Risks can be characterised by a set of uncertain events for which each event has a set of outcomes (Fenton & Neil 2006). In the context of performance risk, the model inputs required to make a design estimate are the uncertain events/variables and can take on a range of values on a continuous or discrete scale. Therefore evaluating the uncertainty associated with each of these system variables is the first step to characterising the performance risk of a system. This can be achieved using probabilistic modelling techniques that propagate the uncertainty of system variables through to the performance output.

The implications of the performance gap associated with domestic buildings and their subsystems are serious if CO_2 targets are to be met. The identification of the performance gap highlights the uncertainty related to in-use performance of strategies employed to improve the state of the housing stock. It cannot be assumed that because the models, predictions, test results and our own presumptions about low carbon designs suggest a certain level of performance that this will be met in reality. Understanding the nature of performance

uncertainty is the first step towards gaining insight into the true state of affairs and developing appropriate risk mitigation strategies.

1.2 Aims and Objectives

The focus of this research is to adopt a novel approach to describe, evaluate and quantify the uncertainty in performance of building energy subsystems as a result of uncertainties throughout the system domain. **The overarching goal of the study is to evaluate probabilistic approaches for quantifying performance uncertainty associated with renewable energy technologies (RET)**; a further aim is to suggest that this approach can be expanded to quantify performance uncertainty of a single building through to the entire housing stock by virtue of the systems engineering philosophy, and assess the likelihood that national carbon targets can be realised. In this context the research objectives are thus:

- Quantify the contribution made by building subsystems to the "performance gap" of a case study dwelling;
- Identify causal factors contributing to discrepancies between predicted and measured performance of a specific RET;
- Develop a flexible methodology for evaluating the effect of uncertainty related to system elements on the performance estimates of the case study RET;
- Apply this method to improve the predictions of solar thermal yields using a current UK compliance model.

1.3 Thesis Structure

This thesis is modular in structure in that each chapter is an independent piece of work which incorporates a review of relevant literature, technical evaluation and results, discussion, and references. The research presented in this thesis takes on three distinct phases: The first phase is to present a technical evaluation of a whole building system and its subsystems, breaking down the performance discrepancy into contributions from the subsystems in the process. This is contained within Chapter 3 which is centred around the concept of the performance gap, as a result literature reviewed to aid in the understanding of this concept and to direct the technical analysis performed is incorporated into this chapter. Subsequently a specific RET is selected for further investigations into the factors associated with performance discrepancies of building subsystems: Chapter 4 begins with a literature review of the classic modelling techniques used to make performance predictions of the RET of interest; these modelling techniques are then applied using the available data to present a comprehensive investigation into the use, advantages and limitations of each of the methods. Chapter 5 is concerned with how uncertainty is incorporated into performance modelling techniques starting with a literature review of commonly used methods and following with direct applications of the methods to the available data. Chapter 6 identifies the causal factors associated with the RET performance following the selection of a suitable modelling approach from Chapter 5. The final phase presented in Chapter 7 describes the final modelling approach taken. Concluding remarks are contained within Chapter 8. A flowchart of the thesis chapters is presented in Figure 1.3.

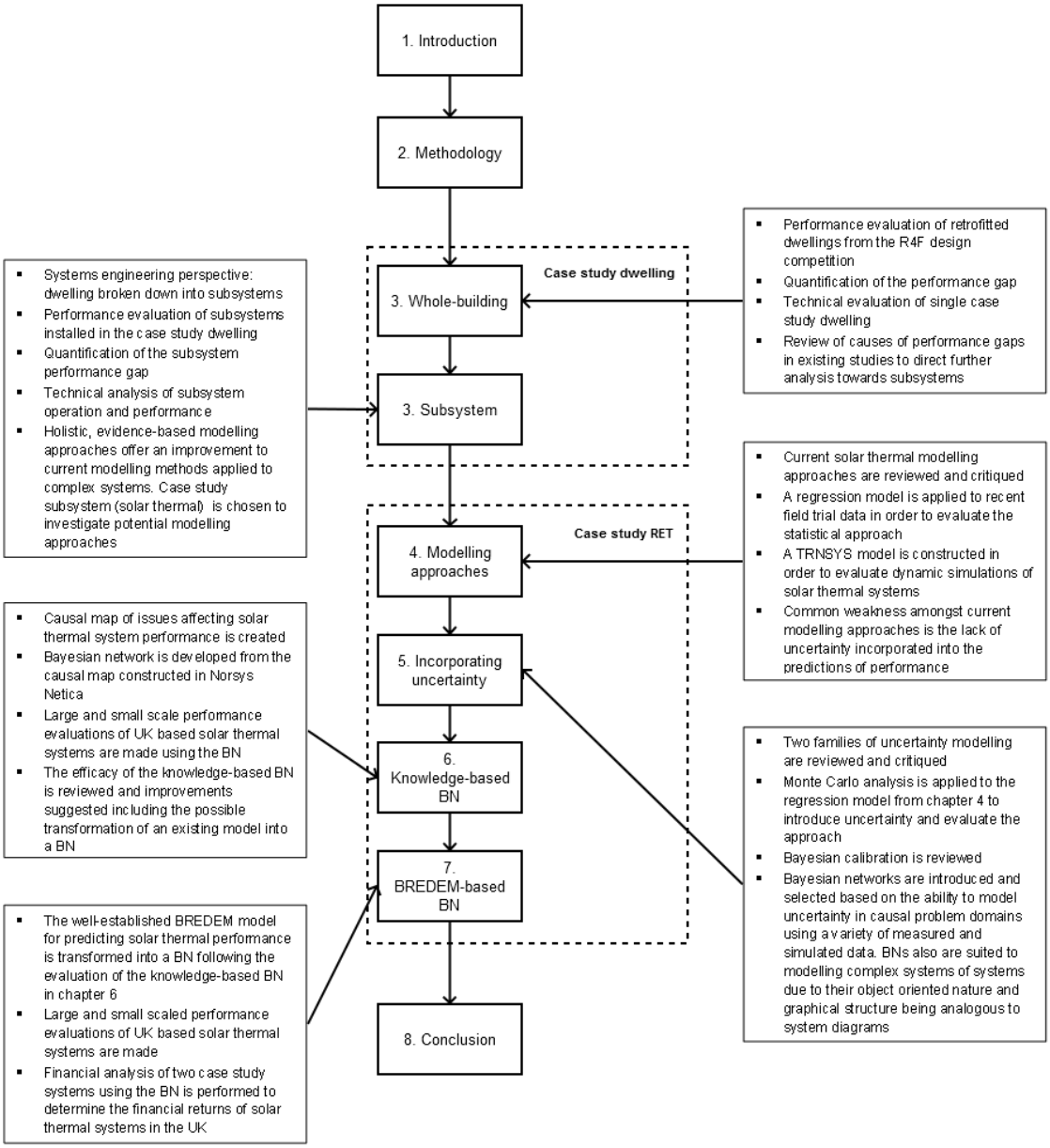

Figure 1.3: Thesis structure

1.4 Novelty

The research presented in this thesis makes the following novel contributions:

- A deeper understanding of the whole-system nature of performance risk associated with domestic buildings;
- Identification of the operational issues affecting building subsystem performance;
- Development of a novel, evidence-based probabilistic approach to evaluating the uncertainty throughout a system domain, which is applicable at all levels of the system hierarchy;
- A deeper understanding of the financial and carbon related implications of performance risk in the context of UK energy policy;

1.5 References

Ballarini, I. & Corrado, V., 2009. Application of energy rating methods to the existing building stock: Analysis of some residential buildings in Turin. *Energy and Buildings*, 41(7), pp.790–800.

Bannister, P., 2009. Why Good Buildings Go Bad, While Some Are Just Born That Way. *Ecolibrium*, pp.24–32.

Boait, P.J., Fan, D. & Stafford, a., 2011. Performance and control of domestic ground-source heat pumps in retrofit installations. *Energy and Buildings*, 43(8), pp.1968–1976.

Bordass, B., Cohen, R. & Field, J., 2004. Energy Performance of Non-Domestic Buildings: Closing the Credibility Gap. In *Building Performance Congress*.

Branco, G. et al., 2004. Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data. *Energy and Buildings*, 36(6), pp.543–555.

Carbon Trust, 2011. Closing the Gap: Lessons learned on realising the potential of low carbon building design.

CarbonBuzz, 2014. CarbonBuzz an RIBA CIBSE platform. Available at: http://www.carbonbuzz.org/index.jsp [Accessed February 9, 2015].

Cayre, E. et al., 2011. There are people in the house! how the results of purely technical analysis of residential energy consumption are misleading for energy policies. In *European Council for an Energy Efficient Economy (ECEEE) Summer School*. pp. 1675–1683.

Cherrington, R. et al., 2013. The feed-in tariff in the UK: A case study focus on domestic photovoltaic systems. *Renewable Energy*, 50, pp.421–426.

DECC, 2015. *2013 UK greenhouse gas emissions: final figures*, London.

Diamond, R. et al., 2006. Evaluating the energy\nperformance of the first generation of LEED-certified commercial buildings.

Dowson, M. et al., 2012. Domestic UK retrofit challenge: Barriers, incentives and current performance leading into the Green Deal. *Energy Policy*, 50, pp.294–305.

EST, 2010. *Getting warmer: A field trial of heat pumps*, London.

Fenton, N. & Neil, M., 2006. Measuring your Risks. *www.agenarisk.com*, pp.1–6.

Fenton, N. & Neil, M., 2013. *Risk Assessment and Decision Analysis with Bayesian Networks* First., CRC Press.

Fokaides, P. a. et al., 2011. Comparison between measured and calculated energy performance for dwellings in a summer dominant environment. *Energy and Buildings*, 43(11), pp.3099–3105.

Gupta, R. & Dantsiou, D., 2013. Understanding the Gap between "as Designed" and "as Built" Performance of a New Low Carbon Housing Development in UK. *Sustainability in Energy and Buildings*, 22, pp.567–580.

Hall, M.R. et al., 2013. Analysis of UK domestic building retrofit scenarios based on the E.ON Retrofit Research House using energetic hygrothermics simulation – Energy efficiency, indoor air quality, occupant comfort, and mould growth potential. *Building and Environment*, 70, pp.48–59.

Hens, H., 2010. Energy efficient retrofit of an end of the row house: Confronting predictions with long-term measurements. *Energy and Buildings*, 42(10), pp.1939–1947.

Hinge, A., Taneja, O. & Bobker, M., 2008. Sustainability in Commercial Buildings - Bridging the Gap from Design to Operations. *Proceedings of the 5th International Conference on Improving Energy Efficiency in Commercial Buildings: IEECB Focus 2008*, 1.

HM Government, 2014. Domestic Renewable Heat Incentive (RHI). Available at: https://www.gov.uk/domestic-renewable-heat-incentive [Accessed February 24, 2015].

HM Government, 2015. Feed-in Tariffs: get money for generating your own electricityour own electricity. Available at: https://www.gov.uk/feed-in-tariffs [Accessed February 24, 2015].

HM Government, 2013. Green Deal: energy saving for your home or business. Available at: https://www.gov.uk/green-deal-energy-saving-measures/how-the-green-deal-works [Accessed April 4, 2013].

HM Government, 2009. *The UK Renewable Energy Strategy*, London.

INCOSE, 2011. *Systems Engineering Handbook*, San Diego.

Kelly, S., Crawford-Brown, D. & Pollitt, M.G., 2012. Building performance evaluation and certification in the UK: Is SAP fit for purpose? *Renewable and Sustainable Energy Reviews*, 16(9), pp.6861–6878.

Lowe, R. & Oreszczyn, T., 2008. Regulatory standards and barriers to improved performance for housing. *Energy Policy*, 36(12), pp.4475–4481.

Majcen, D., Itard, L.C.M. & Visscher, H., 2013. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications. *Energy Policy*, 54, pp.125–136.

MCS, 2015. Microgeneration Certification Scheme. Available at: http://www.microgenerationcertification.org/ [Accessed April 15, 2015].

Menezes, A.C. et al., 2012. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97, pp.355–364.

Mørck, O., Thomsen, K.E. & Rose, J., 2012. The EU CONCERTO project Class 1 - Demonstrating cost-effective low-energy buildings - Recent results with special focus on comparison of

calculated and measured energy performance of Danish buildings. *Applied Energy*, 97, pp.319–326.

Ofgem, 2015. Energy Company Obligation (ECO). Available at: https://www.ofgem.gov.uk/environmental-programmes/energy-company-obligation-eco [Accessed April 15, 2015].

Palmer, J. & Cooper, I., 2013. *United Kingdom housing energy fact file*, London.

Power, A., 2008. Does demolition or refurbishment of old and inefficient homes help to increase our environmental, social and economic viability? *Energy Policy*, 36, pp.4487–4501.

Ravetz, J., 2008. State of the stock-What do we know about existing buildings and their future prospects? *Energy Policy*, 36, pp.4462–4470.

Reddy, T.A., 2006. Liteterature Review on Calibration of Building Energy Simulation Programs: Uses, Problems, Procedures, Uncertainty, and Tools. *ASHRAE Transactions*, 112(1), pp.226–240.

Scofield, J.H., 2009. Do LEED-certified buildings save energy? Not really... *Energy and Buildings*, 41, pp.1386–1390.

Torcellini, P.A. et al., 2004. Lessons Learned from Field Evaluation of Six High- Performance Buildings. In *ACEEE Summer Study on Energy Efficiency in Buildings*.

Turner, C. & Frankel, M., 2008. *Energy Performance of LEED for New Construction Buildings*, Washington.

Zero Carbon Hub, 2013. *Closing the Gap Between Design & As-Built Performance*, London.

Chapter 2

Research Framework and Methodology

2.1 Introduction

Chapter 1 introduced the concept of the building energy "performance gap" and, concluded that it arises from a series of technical and human factors associated with modelling assumptions made at the design stage and subsystem performance in the operational stage of a building. This chapter describes the overall research framework and specific methods that will be applied in order to achieve the research aim and objectives laid out in Chapter 1.

2.2 Systems Engineering and Uncertainty

As explored in Chapter 1, the performance gap arises from a range of technical and human issues related to ideal world assumptions made at the design stage about sub-system performance, construction and installation, and operation. It is clear that these issues are prevalent throughout the lifecycle of a building and therefore a method of addressing these issues from conception to operation would go some way to mitigating the risk of poor performance. Systems engineering is an effective way to manage complex and changeable systems and involves the completion of many processes at each of the stages in the lifecycle of a system (INCOSE 2011).

Systems engineering approaches provide a suitable framework for the analysis of a complex system such as a building by encouraging interactions between different interested parties and outlining numerous processes to be executed to achieve the goal of providing a quality product that meets the user's needs. Figure 2.1 shows the different systems engineering processes that are performed and the level of effort required for each process at different stages in the system lifecycle, a full description of the various different processes can be found in (INCOSE 2011). The figure also indicates where elements of uncertainty surrounding the actual in-use conditions are introduced throughout the lifecycle which combine to form the performance gap.

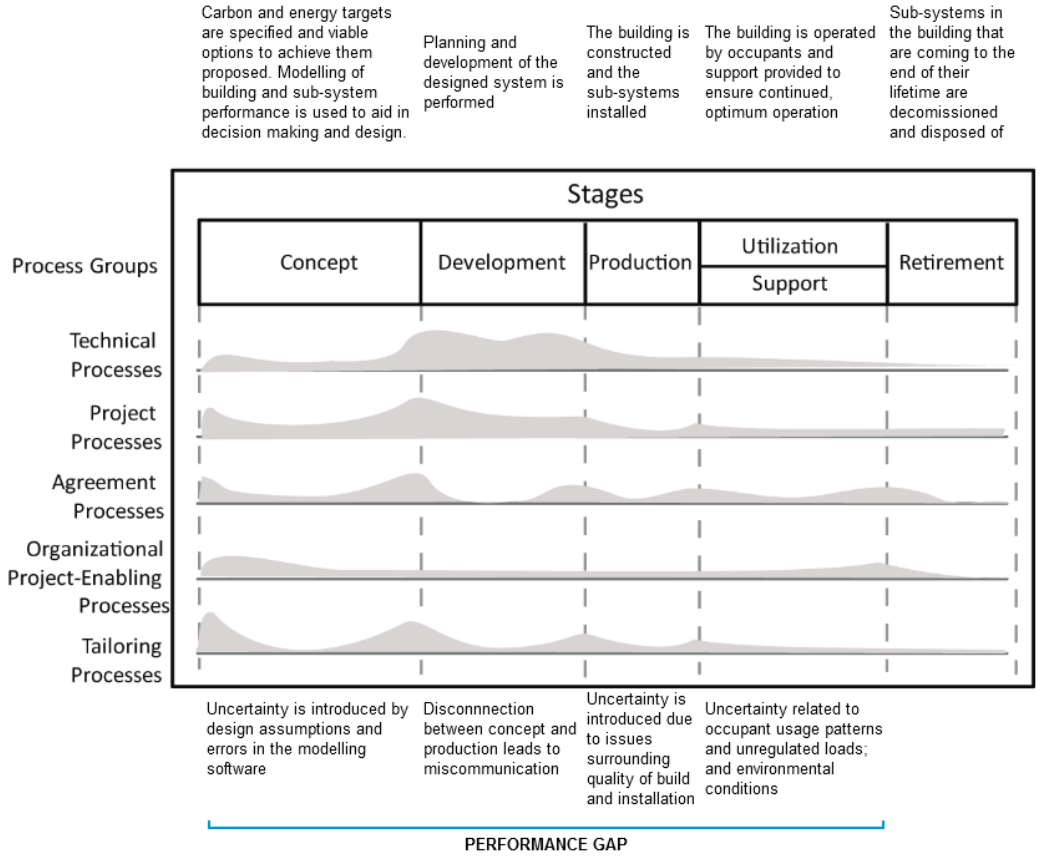

Figure 2.1: Systems engineering process effort level across lifecycle stages; adapted from (INCOSE 2011) **to include descriptions of the points at which uncertainty is introduced throughout the life of a building**

A building can be considered as a complex system-of-systems. Systems-of-systems (SoS) is a systems theory concept that states that the system elements within a system of interest are themselves systems; a system element is a major product, service or facility of a system, sometimes the term subsystem is used. The system-of-systems view means that the system of interest depends upon the perspective being taken; for example a system element (or subsystem) in one system of interest may itself be a system of interest depending on the perspective of the analysis. The hierarchy within a system of interest can be graphically represented as in Figure 2.2. The systems hierarchy diagram shows the interacting system elements for each of which responsibility can be delegated to different parties. In this manner the lifecycle processes outlined above can be executed by the respective parties to each of the system elements to resolve the requirements of the system of interest in a coordinated way.

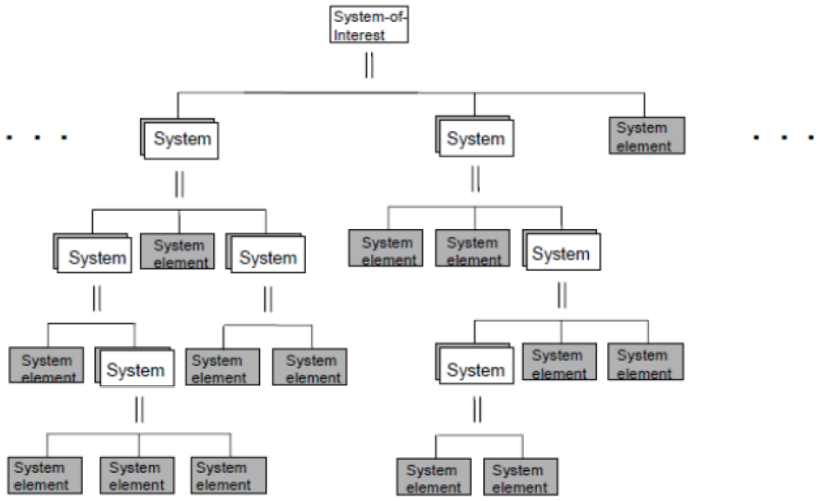

Figure 2.2: Hierarchy within a system (INCOSE 2011)

2.2.1 Probabilistic modelling

Bayesian networks represent a method by which uncertainty in a problem domain or system can be evaluated. The network aspect of a BN represents the causal links between different system elements identifiable from expert knowledge about the factors influencing the performance of the system. Uncertainty is propagated through each of these system elements through to the system output using Bayesian inference. Uncertainty regarding each of the system elements can be quantified in two ways: subjectively using distributions of values elicited from experts; or from data. The output of a BN is a probability distribution that quantifies the uncertainty related to system performance.

Bayesian networks are well supported by the systems-of-systems standpoint taken to mitigating risk in building performance due to their object oriented nature. Bayesian networks can be developed and applied to each of the system elements and the outputs of each network fed into an overall model of the system of interest. The overall model of the system of interest need not be a BN itself and may conceivably be a based on current modelling methods incorporating Monte Carlo techniques using the outputs from the system element BNs as input distributions which are sampled. Another advantage of the BN approach is that it can make use of real in-use performance data related to system elements and the system of interest; in this way the suggestion of feeding back measured data into the design stage of a building can be met.

2.3 Case Study Analysis

The above theoretical framework outlines the supporting structure within which the research will be conducted. This section describes the research methods used within the framework in order to address the aim and objectives of the study.

Since this research is interested in the discrepancy between predicted and actual in-use system performance then in-use performance data is required. Laboratory-style controls cannot be imposed on the operating conditions of a building system because this removes the uncertainties associated with in-use operation. Indeed controlled tests of building subsystems have been identified as contributing to the performance gap for the very reason that the test

conditions do not reflect in-use operation. It is in these circumstances that case study analysis proves to be a useful research method (Amaratunga & Baldry 2001).

A case study is a detailed examination of a system and is used to provide evidence of a theoretical principle or phenomenon. In the context of this research the system being studied is a domestic building, the purpose of the study being to evaluate the performance gap associated with its energy consumption. The case study dwelling will be evaluated within the framework of systems engineering to establish the system elements in the building and to quantify the discrepancy between modelled and measured performance of these elements arising from assumptions made at the design stage. The analysis will quantify the actual energy consumption associated with space and DHW heating and electrical equipment and compare this to the designed consumption. A detailed comparison between the assumptions made at the design stage and the actual in-use conditions will be made; these will relate to: the building fabric; climatic conditions; heating system performance; electrical loads/generation; and usage behaviour.

The case study analysis will delve deep into the performance aspects of the electricity and heating systems providing detailed discussion surrounding the actual energy consumption and generation. One limitation of case study analysis applied to a single building system is the difficulty in generalising the findings to apply them to other buildings – this is analogous to the limitation associated with calibrated models. Triangulation is a concept in case study analysis that describes the way in which this main limitation can be overcome. The use of multiple data sets and multiple case studies can be used to corroborate findings and identify general themes that may have a more widespread existence (Amaratunga & Baldry 2001). In the context of subsystem performance triangulation will be achieved by selection of a single case study subsystem – solar thermal. Field trial data that provides evidence of in-use performance of a number of solar thermal systems will be used to quantify the performance gap associated with delivered thermal energy termed the solar yield. Predicted performance will be obtained using the BREDEM model (Henderson & Hart 2013) and compared to measured annual yield. Interactions between the solar thermal system elements that influence performance will be elicited from expert knowledge contained in literature and graphically represented in a causal map.

The causal map provides an in depth understanding of the critical interactions between system elements that affect performance. It also provides the structure to the Bayesian network which will be used to evaluate the uncertainty related to each of these system elements and overall system performance. In-use data related to solar thermal system performance and each of the system elements will be obtained from several sources and used to quantify uncertainty. The uncertainty associated with each of the system elements can then be propagated through the network to provide a distribution of solar yield conditional on the knowledge held by the modeller about the system under evaluation. The use of multiple data sets and data for multiple systems allows the BN to quantify the uncertainty across a heterogeneous group of solar thermal systems as well as for a single system once evidence surrounding its parameters is known.

2.3.1 Case study selection

Retrofit for the Future was a competition held in 2009 funded by the Technology Strategy Board (TSB). The target was to reduce carbon emissions of a dwelling by 80% to reflect the UK's target of an 80% reduction in carbon emissions by 2050 based on 1999 levels. 194 projects were funded with £20,000 to develop strategies to achieve this aim and a further £150,000 was awarded to 86 projects to put these designs into practice. Some of these projects involved more than one home resulting in over 100 homes being retrofitted (TSB 2013b).

As part of the challenge the properties were monitored to enable evaluation of the whole-house performance in terms of energy consumption, primary energy consumption and carbon emissions. These data are contained in reports submitted by each of the project teams and can be found on the Low Energy Building Database (TSB 2013a).

The energy performance gap of 26 retrofitted dwellings was calculated and a single dwelling selected for case study analysis. The selection was based on the value of the performance gap, which approached the median value. Furthermore data for this dwelling was readily available, an important consideration when adopting the case study approach (Amaratunga & Baldry 2001). The case study dwelling is located in Wrestlingwoth, Bedford and for the remainder of this research will be referred to as the 'Bedford dwelling'.

The case study building system installed at Bedford and taken for further analysis is a solar thermal system. Up until 2012 solar thermal systems were the most widespread RET installed in the UK with almost 100,000 installations; however increases in electricity prices, high FiT rates and reduced installation costs led the number of PV installations to overtake the number of STS with 320,000 in 2012 (Balcombe et al. 2013). The number of PV installations installed each month has decreased since the FiT rate was halved in 2012 (Balcombe et al. 2013). More recently, the introduction of the domestic renewable heat incentive (HM Government 2014) may lead to an increase in the uptake of solar thermal systems (DECC 2013), payments of which are based on estimated performance (Ofgem 2014); this could incur a degree of competition over roof space between PV and solar thermal systems. Therefore analysis of the influential factors on solar thermal system performance and evaluation of the uncertainty related to performance is relevant when we consider the financial risks associated with poor performance of these systems and decisions to be made about what the alternatives may be. The RHI is also applicable to heat pump systems; however there have been several studies into the in-use performance of heat pumps including a nationwide field trial conducted in 2008 by the EST (Dunbabin & Wickins 2012; Kelly & Cockroft 2011; Boait et al. 2011). These revealed significant performance issues related to the quality of install and the clarity of the control systems which needed to be addressed and brought up to a reasonable standard before operational issues with performance could be reasonably evaluated. At the time of writing a second field trial of heat pumps installed throughout the UK suggested that these issues had been addressed (Dunbabin et al. 2013); however the data was not available for analysis to be included in this research. Data used to conduct the analysis into the performance of solar thermal systems came from expert knowledge elicited from literature and that collected as part of the domestic field trial conducted by the EST. This field trial provides a wealth of data associated with in-use performance of solar thermal systems. Several case study systems from this field trial were used to compare actual performance with that estimated using the BREDEM model. The BREDEM

model was used because its purpose is for prediction as opposed to the sister model SAP which is used for compliance (BRE 2013); furthermore it is presumed that a BREDEM based prediction model will be used to make performance estimates of systems being installed under the RHI (Crowther et al. 2010; MCS 2013).

2.4 Case Study I: Whole-Building Analysis

The Bedford dwelling is a 1950s mid-terrace bungalow, which has been comprehensively refurbished to improve energy consumption and CO_2 emissions of the property. Three renewable energy technologies were installed as part of this refurbishment including a 3.6kW$_p$ PV system; 4.5kW$_{th}$ ASHP; and three 2.1m^2 solar thermal collectors. Energy consumption and generation data has been collected since January 2011.

The Bedford dwelling is an "all-electric" dwelling meaning that prior to the retrofit all energy requirements including heating were met by mains electricity; it is occupied by two elderly people. The results from this project should provide valuable data for providers of small, compact dwellings in off-gas areas occupied by vulnerable members of society.

The project was led by SDC Builders Ltd and involved architects from Eco Design Consultants, mechanical and electrical consultants from Venables Associates; the client was Aragon Housing Association which has over 1700 post-war single story homes such as the Bedford dwelling. The objectives of the refurbishment were identified as (SDC Building Ltd 2010):

1. To improve the building envelope;
2. To validate a low carbon renewable energy proposal that would interlink fabric and the installations to provide controls suitable to the occupant behaviour;
3. To substantially reduce carbon emissions and to provide a cost effective whole house solution that would enhance the tenant wellbeing;
4. To monitor the installations and occupant behaviour;

2.4.1 Retrofit strategies

Electrical and heating loads of the case study dwelling were reduced through design improvements to the building fabric (including fenestration) and installed micro-generation technologies – these are summarised in Table 2.1.

Building fabric	External insulation – U-value 0.12W/m^{2}K Spacetherm floor insulation – U-value 0.15W/m^{2}K 400mm loft insulation – U-value 0.09W/m^{2}K Triple glazed windows – U-value 0.8W/m^{2}K Taped fenestration joints and wall/floor junctions for improved air tightness
Building services	**Electrical equipment**
	Mechanical ventilation and heat recovery (MVHR) 3.68kW$_p$ photovoltaic (PV) system
	Space and water heating
	4.5kW air-source heat pump (ASHP) 6.3m^2 flat plate collector

Table 2.1: Description of case study building retrofit installations

Passive solar and daylighting strategies were achieved through appropriate use of glazing to minimise the energy consumed for heating and lighting. The kitchen window was enlarged to fill the width of the room and maximise the amount of daylight entering; a sun pipe delivers the daylight factor target of 1.5% to the living room. The windows are triple glazed and have a U-value of 0.8W/m^2K ensuring that solar gains outweigh thermal losses through fenestration.

Space heating loads were minimised using external insulation applied to achieve a wall U-value of 0.12W/m^2K; the floor was insulated using Spacetherm insulation and chipboard to achieve a floor U-value of 0.15W/m^2K; mineral wool was added to the existing loft insulation to increase the thickness to 400mm and achieve a roof U-value of 0.09W/m^2K. The insulation strategy also contributes to space cooling during the summer by minimising heat transfer through the walls. The MVHR is used to ventilate the dwelling on occasions when metabolic and appliance heat gains cause excessive indoor temperatures. The MVHR system's primary function is to preheat incoming fresh air with exhaust air to reduce the thermal losses from the dwelling. The MVHR system adopts the Building Regulations Part F System 4 approach which allows air permeability of the dwelling to be reduced as much as possible. Air permeability (air tightness) was improved by installing high quality fenestration with taped joints; junctions between the walls and floor are also taped.

Three micro-generation systems were installed to reduce the amount of mains electricity used for electrical and heating loads: a 3.68kW_p PV system contributes to electrical loads of the dwelling with any surplus exported to the National Grid; space and water heating is met by a combined CO_2 ASHP (4.54kW) and 6.3m^2 flat plate collector system.

2.4.2 Occupancy

The dwelling is occupied by an elderly couple who both remain at home most of the time, with one member working from home and the other with reduced mobility due to a traffic accident and sight impairment.

2.5 Research Context

The systems engineering approach will be adopted in the analysis of the performance of the Bedford dwelling. The dwelling was identified as exhibiting discrepancy between predicted and actual in use energy consumption. As described above discrepancies between designed and in-use energy consumption arise from a complex blend of technical and non-technical factors. The technical and non-technical factors influencing system performance often overlap and so will be considered together in the analysis of the whole-building and subsystem performance.

2.6 Case Study Dwelling System Hierarchy

As per the system engineering approach the whole-building system at Bedford is broken down into its subsystems and system elements shown in Figure 2.3. The figure shows the four main building subsystems: electrical equipment; building fabric; heating system; and the occupants. The subsystems are broken down further into subsystems and system elements; however the fabric and occupant subsystems are not expanded into their subsystems because they are not the focus of this research.

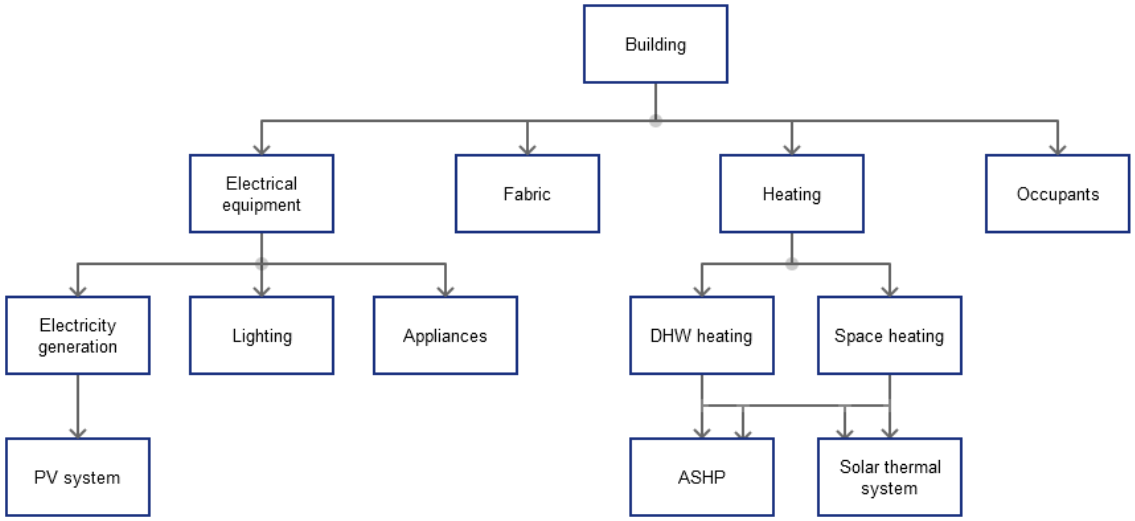

Figure 2.3: System hierarchy for the Bedford dwelling

The whole-building system performance will be evaluated in the first instance and the causes for performance discrepancy associated with design assumptions surrounding the four primary subsystems identified. A focus will be made on the energy consumed and delivered by the building systems. Following this, a technical evaluation of the three micro-generation subsystems is performed to discuss causes for subsystem performance discrepancy associated with design assumptions made when testing/modelling these systems as well as dysfunction arising during operation. A summary of how these issues may have been avoided at the different lifecycle stages of the building and its systems will be presented at the end.

2.7 Technical Evaluation of the Case Study Dwelling

2.7.1 Whole system

The measured annual energy consumption prior to the retrofit will be obtained from billing data. The designed energy consumption post-retrofit will be compared with the actual energy consumption measured in the second year of operation. The second year was chosen to allow a period of "settling down" of the building systems. The consumed energy is defined as that imported from the grid to allow comparison pre- and post-retrofit and should not be confused with the demand for energy. Building fabric improvements reduce the demand for energy required for heating (thus reducing the amount of grid electricity required), whereas the solar thermal and ASHP systems reduce the amount of energy consumed to meet the reduced demand. Therefore the energy consumption of the retrofitted dwelling should be less than the demand. Furthermore, the PV system should offset consumption of grid electricity thus reducing the amount imported from the grid.

2.7.2 Subsystems

Monitored data for the performance of three subsystems is used to make comparisons between designed and in-use performance. The amount of electricity generated by the PV system is measured and used to calculated system efficiency – these are compared to modelled results for the system. The amount of electricity consumed by the ASHP and the amount of heat delivered

are used to calculate the in-use COP compared to the manufacturer's expectations from field trial. Estimated performance of the combined solar thermal and ASHP system is compared to measured values for solar yield, system efficiency and COP. A detailed evaluation into the performance of the heating subsystem is made since space and water heating are a significant end-use in the UK – in particular the proportion of energy required for DHW heating is set to increase with the prevalence of the fabric first approach reducing the requirements for space heating and energy efficient electrical goods reducing electricity end-use loads (Steijger 2013). There is a high level of complexity and occupant-system interaction associated with the combined heating system which influences the performance; the monitored data will be used in conjunction with analysis of existing studies into these kinds of heating systems to propose reasons for system dysfunction. A summary of the causes for the performance gap associated with these building subsystems will be presented highlighting the roles of technical dysfunction (related to installation quality, system complexity and suitability for purpose) and design assumptions (related to ideal world beliefs about occupancy usage behaviour and subsystem performance) in performance discrepancies.

2.8 Case Study II: Solar Thermal

Solar thermal systems are a typically roof-mounted renewable heating system and as such are in direct competition with other roof-based systems such as PV and other renewable heating systems such as heat pumps. Householders choosing between solar thermal and PV systems must consider the financial return possible from the tariff mechanisms (RHI versus FiT) which are linked to predicted performance of these systems. Likewise householders who are looking to install a renewable heating technology must consider the financial returns of solar thermal systems versus other eligible technologies such as heat pumps. The relative immaturity of the RHI versus FiTs associated with PV systems means solar thermal systems may increase in popularity in the near future; therefore the uncertainty related to the in-use versus predicted performance is important to understand so that we may evaluate:

- the risk associated with actual carbon reductions;
- the actual fuel bill reductions seen by the occupant and subsequent savings made;
- the value of the RHI to the householder and policy maker (under versus over payment based on predicted performance);

The solar thermal system at Bedford is selected for further study to establish influencing factors on performance using the systems engineering framework.

The solar thermal system hierarchy represented in Figure 2.4 shows the different technical systems that constitute the physical aspect of the whole-system as well as the non-technical systems that are required for the whole-system to operate i.e. the occupants making thermal demands and the environment providing the solar radiation required to heat the working fluid.

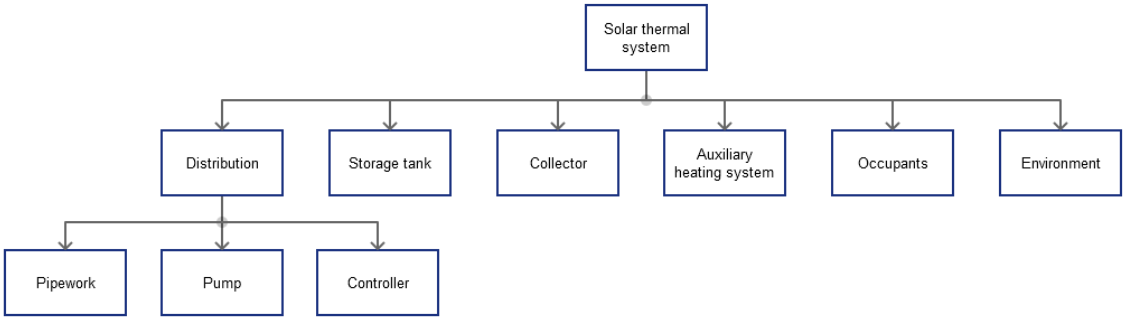

Figure 2.4: Solar thermal system hierarchy

As can be seen from the system hierarchies in Figure 2.3 & Figure 2.4 one common system element is the occupants; occupants interact with individual building subsystems that make up a dwelling and so are an important consideration when predicting in-use performance of buildings and systems. However occupant behaviour is often unpredictable and assumptions must be made at the design stage which as discussed above contributes to discrepancies in energy performance of systems. Evaluation of the uncertainty related to occupancy consumption behaviour and its effect on system performance is an important element of risk mitigation strategies employed under the systems engineering framework.

2.9 Technical Evaluation of Solar Thermal Systems

Following on from the technical evaluation of the Bedford dwelling, solar thermal system field trial data was used to investigate in-use performance of solar thermal systems installed throughout the UK. Designed performance of these systems was unavailable, however using the BREDEM model for predicting annual solar yield the performance gap could be identified for a subset of these systems. This model was used in two ways to achieve two different objectives: firstly to ascertain the discrepancies between actual and predicted values for solar radiation, DHW volumetric consumption and DHW energy demand and annual solar yield resulting from assumptions made about DHW usage patterns (including bathing habits and type of shower used), shading of the collector, and performance parameters of the collector (if unknown). The second objective was to ascertain the accuracy of the model itself by substituting actual data for the quantities usually calculated using regression equations and/or based on design assumptions – this would highlight any limitations in the calculation methods used in the model removing the effect of design assumptions. Causes for the performance gap associated with the BREDEM model are summarised in Chapter 7.

Following this analysis, various approaches to incorporating uncertainty into physics based models such as BREDEM and dynamic simulation models such as TRNSYS were evaluated. Arguments for the Bayesian network approach and details of how this was applied to the evaluation of solar thermal system uncertainty presented in Chapter 6. The Bayesian network approach necessarily required causal mapping of the influential factors on system performance which were elicited from expert literature and dynamic modelling.

2.10 Datasets and Monitoring

2.10.1 EST solar thermal field trial

The EST solar thermal field trial monitored the performance data for 86 systems installed throughout the UK over the period of one year from 1st April 2010 to 31st March 2011; all the systems were used to provide DHW only as per the renewable heat incentive. The objective of the field trial was to seek evidence of:

- In-use performance of solar thermal systems in UK homes to establish potential carbon savings from each installation;
- The factors that influence the performance of solar thermal systems;
- The customer experience and perceptions of the technology during acquisition, installation and operation as well as the benefits to the customer;

The EST appointed monitoring responsibilities to two contractors: EA Technology; and Gastec at CRE. Both companies have worked together on previous projects for the EST and the Carbon Trust monitoring condensing boilers, heat pumps and micro-CHP systems.

The monitored installations were selected from recipients of grant funding through the Low Carbon Buildings Programme as well as selections made by Sustainable Energy Ireland for the Irish installations. Selections were based on the following system characteristics: manufacturer; collector type; cylinder type; geographic location; household type.

Data for the following system parameters was collected:

- System design:
  - Collector type;
  - Collector area;
  - Tank volume;
  - Orientation;
  - Roof pitch;
  - Location;
  - Shading issues;
- Household:
  - Type of household;
  - Number of occupants;
- Usage behaviour:
  - DHW volume;
  - DHW energy content;
  - Auxiliary heating requirement;
  - Water use temperature;
  - Cold water temperature;
  - Pump and controller electricity consumption;
- Solar resource:
  - Solar irradiation;
  - Solar irradiance;

Figure 2.5 shows a diagram for the generic monitoring platform:

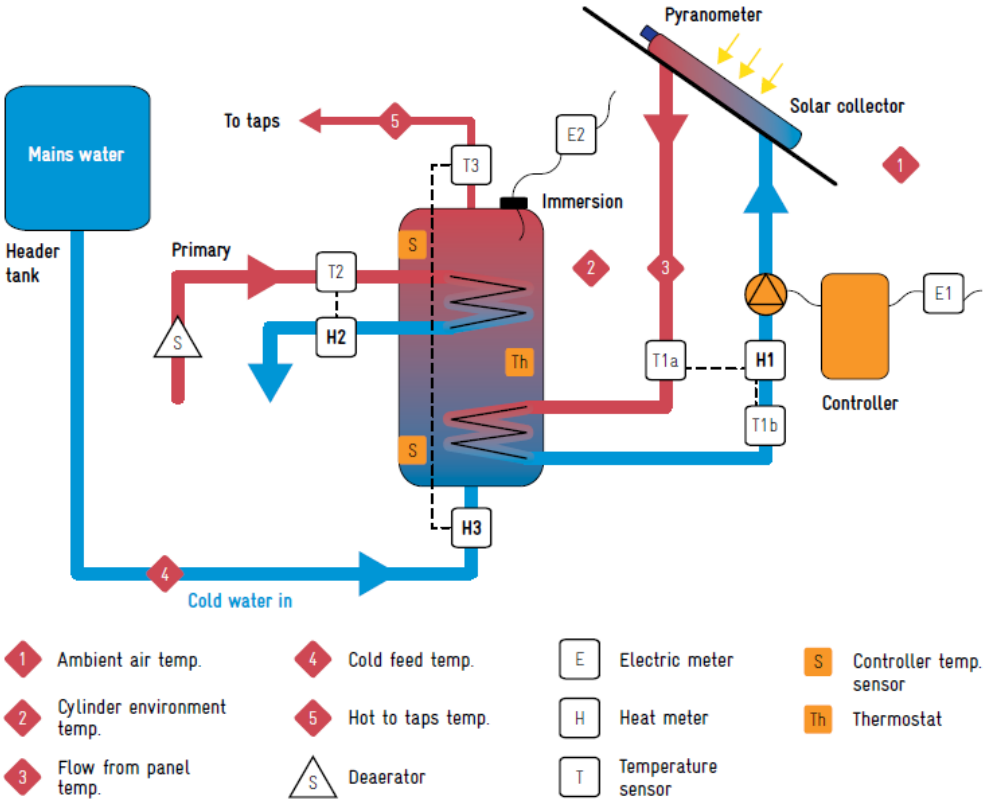

Figure 2.5: EST solar thermal field trial monitoring platform (EST 2011)

This data was used to investigate the performance gap associated with solar thermal systems in-use as well as to quantify the uncertainty related to solar thermal system parameters in the Bayesian network.

2.11 Additional Datasets

Additional data was used to quantify the uncertainty related to solar thermal system elements. These data were obtained from expert literature and small scale field trials used to identify causal relationships between influential factors on performance and provided additional data surrounding the different system elements; these studies are listed in the Chapter 6. In addition, an earlier EST field trial into the consumption of DHW throughout UK homes was used to provide additional data about occupant usage behaviour (EST 2008).

2.12 Summary

The systems engineering framework is adopted in order to identify the risk factors to the energy performance of a case study building and subsystem. Measured data associated with building and subsystem energy consumption was collected and compared to modelled predictions to identify the presence of the performance gap. An evaluation of the design assumptions made at the time of modelling was contrasted with actual in-use operating conditions to present potential causes for the performance discrepancy. Furthermore, detailed analysis of the

technical performance of the subsystems was made in order to highlight the contribution that subsystem dysfunction makes towards reduced performance compared to the designed value. The solar thermal system was chosen as a case study subsystem with which a deep dive into the root causes of performance uncertainty could be made using measured in-use data and published research. A Bayesian network approach to quantifying this performance uncertainty was developed and provides a novel approach to risk mitigation and design/investment based decision making in the presence of uncertainty. This method is supported by the systems engineering framework which prescribes improvement to the development of buildings through completion of coordinated processes at the different lifecycle stages of the building including risk management and decision making processes. In addition, the quantification of performance uncertainty is useful to those wishing to install solar thermal systems as part of the RHI – results from the BN can be used to assess potential financial returns based on energy bill reduction. The approach can be expanded to include alternative subsystems such as ASHPs and PV systems to establish the best investment opportunity. Furthermore, the object oriented nature of the approach means that performance distributions of different systems can be used in any modelling scenario that requires estimates of system performance such as building design and modelling of socio-economic impacts of renewable energy systems (Leicester et al. 2013).

2.13 References

Amaratunga, R. & Baldry, D., 2001. Case study methodology as a means of theory building: performance measurement in facilities management organisations. *Work Study*, 50(3), pp.95–105.

Balcombe, P., Rigby, D. & Azapagic, A., 2013. Motivations and barriers associated with adopting microgeneration energy technologies in the UK. *Renewable and Sustainable Energy Reviews*, 22, pp.655–666.

Boait, P.J., Fan, D. & Stafford, a., 2011. Performance and control of domestic ground-source heat pumps in retrofit installations. *Energy and Buildings*, 43(8), pp.1968–1976.

BRE, 2013. SAP 2012 The Government's Standard Assessment Procedure for Energy Rating of Dwellings.

Crowther, M. et al., 2010. *Report to DECC on Heat metering for the RHI*, London.

DECC, 2013. *Renewable Heat Incentive: Impact Assessment*, London.

Dunbabin, P., Charlick, H. & Green, R., 2013. *Detailed analysis from the second phase of the Energy Saving Trust's heat pump field trial*, London.

Dunbabin, P. & Wickins, C., 2012. *Detailed analysis from the first phase of the Energy Saving Trust's heat pump field trial*, London.

EST, 2011. *Here comes the sun: a field trial of solar water heating systems*, London.

EST, 2008. *Measurement of Domestic Hot Water Consumption in Dwellings*, London.

Henderson, J. & Hart, J., 2013. BREDEM 2012 – A technical description of the BRE Domestic Energy Model.

HM Government, 2014. Domestic Renewable Heat Incentive (RHI). Available at: https://www.gov.uk/domestic-renewable-heat-incentive [Accessed February 24, 2015].

INCOSE, 2011. *Systems Engineering Handbook*, San Diego.

Kelly, N.J. & Cockroft, J., 2011. Analysis of retrofit air source heat pump performance: Results from detailed simulations and comparison to field trial data. *Energy and Buildings*, 43(1), pp.239–245.

Leicester, P.A., Goodier, C.I. & Rowley, P., 2013. Using a bayesian network to evaluate the social, economic and environmental impacts of community deployed renewable energy. In *Proceedings of CISBAT, Clean Technology for Smart Cities and Buildings*.

MCS, 2013. Microgeneration Installation Standard MCS 024: Solar Domestic Hot Water Energy Calculation. , (1.1), pp.1–11.

Ofgem, 2014. Do I need metering for the Domestic RHI? , pp.1–3.

SDC Building Ltd, 2010. *Hard to Heat Homes*, Report available at: http://www.lowenergybuildings.org.uk/projectPDF.php?id=26.

Steijger, L.A., 2013. *Evaluating the Feasibility of ' Zero Carbon ' Compact Dwellings in Urban Areas*.

TSB, 2013a. Low Energy Building Database. Available at: http://retrofitforthefuture.org/ [Accessed January 21, 2014].

TSB, 2013b. *Retrofit Revealed: The Retrofit for the Future Projects - Data Analysis Report*, Report available at: https://retrofit.innovateuk.org/.

Chapter 3

The Energy Performance of Retrofitted Domestic Buildings

3.1 Introduction

In this chapter the annual energy performance of a series of retrofitted dwellings is presented; the aim is to evaluate the gap between the predicted annual energy performance and that measured referred to as the "performance gap". The data used to make this analysis are from dwellings in the TSB's Retrofit for the Future competition (R4F) (TSB 2013c). A single case study dwelling is selected for further analysis to identify reasons for the discrepancy between predicted and measured energy performance. Contributions of the different building subsystems to the whole-building performance gap are quantified and discussed. The aim of the chapter is to demonstrate that energy performance risk associated with a whole building is characterised by performance uncertainty in the building subsystems. This chapter lays the foundation for the proposal of a novel method for evaluating uncertainty related to subsystem performance in order to manage energy performance risk at different system levels.

3.2 The "Performance Gap"

There are numerous studies that show a difference between the actual and estimated energy consumption and performance of buildings (Bordass et al. 2004; Majcen et al. 2013; Kelly et al. 2012; Branco et al. 2004; Hens 2010; Cayre et al. 2011; Carbon Trust 2011; Fokaides et al. 2011) - this difference is often called the "performance gap" and is illustrated in Figure 3.1.

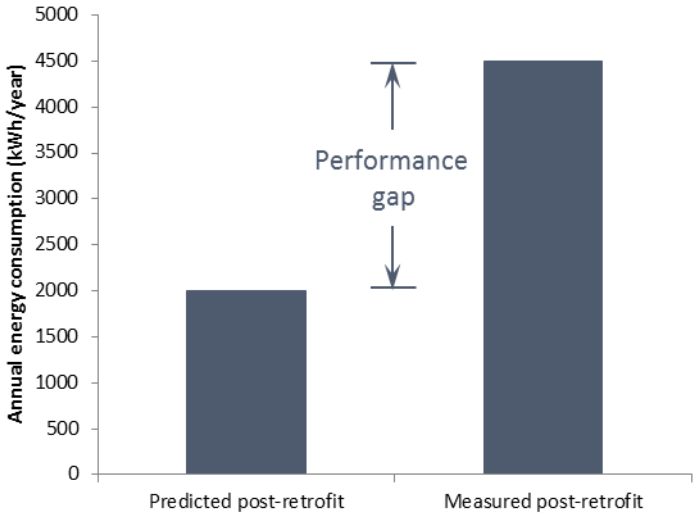

Figure 3.1: An illustration of the "performance gap"

Whole-building performance gaps have been attributed to several issues associated with the design, construction and operational stages of the project (Hinge et al. 2008; Zero Carbon Hub 2013; Mørck et al. 2012; Gupta & Dantsiou 2013; Bordass et al. 2004; Bannister 2009; Carbon Trust 2011; Torcellini et al. 2004; Menezes et al. 2012; Zero Carbon Hub 2010; NHBC Foundation 2012). Commonly reported issues can be summarised as:

- Design assumptions about occupancy levels and behaviour being different in reality;
- Systems failing to achieved the designed levels of performance;
- Inadequate system commissioning and build quality;
- Over complexity of building and/or system design;
- Lack of knowledge about how to maintain and operate the building and its subsystems;
- Lack of feedback of actual performance of buildings and subsystems;
- Limitations of current modelling techniques;

Majcen et al. (2013) concluded that in terms of average and low energy efficient homes predictive measures tend to result in an over-estimation of energy consumption. Conversely estimates of energy consumption for new and retrofitted homes tend to be lower than actual consumption (Majcen et al. 2013). Majcen et al. (2013) also discovered that the primary energy saved by improving a dwelling from a G to an A-rating is much less in reality than anticipated. This may lead to inaccurate estimations of payback times of the retrofit strategies employed, and doubt about whether carbon targets can be successfully attained (Majcen et al. 2013).

The next section evaluates predicted and actual performance of dwellings in the Retrofit for the Future competition.

3.3 Retrofit for the Future

Retrofit for the Future was a competition held in 2009 funded by the Technology Strategy Board (TSB). The target was to reduce carbon emissions of a dwelling by 80% to reflect the UK's target of an 80% reduction in carbon emissions by 2050 based on 1999 levels. 86 projects were funded

to put low carbon retrofit designs into practice. Some of these projects involved more than one home resulting in over 100 homes being retrofitted (TSB 2013c).

Energy consumption data related to building subsystems was monitored to provide information about the in-use energy and carbon performance of the whole-dwelling. Internal and external environmental conditions were also monitored to allow both the occupant comfort and individual subsystem performance to be evaluated. Section 3.6 describes in detail the monitoring platform installed and the parameters measured. These data are contained in reports submitted by each of the project teams and can be found on the Low Energy Building Database (TSB 2013a).

3.4 Performance Evaluation of the Retrofit for the Future Dwellings

Of the 86 projects only 26 provided both design and actual CO_2 emissions. Figure 3.2 shows the percentage difference between the actual carbon emissions and those predicted for the designed dwelling. The majority of dwellings emit more CO_2 than predicted with a mean percentage difference of 60% and a standard deviation of 70%.

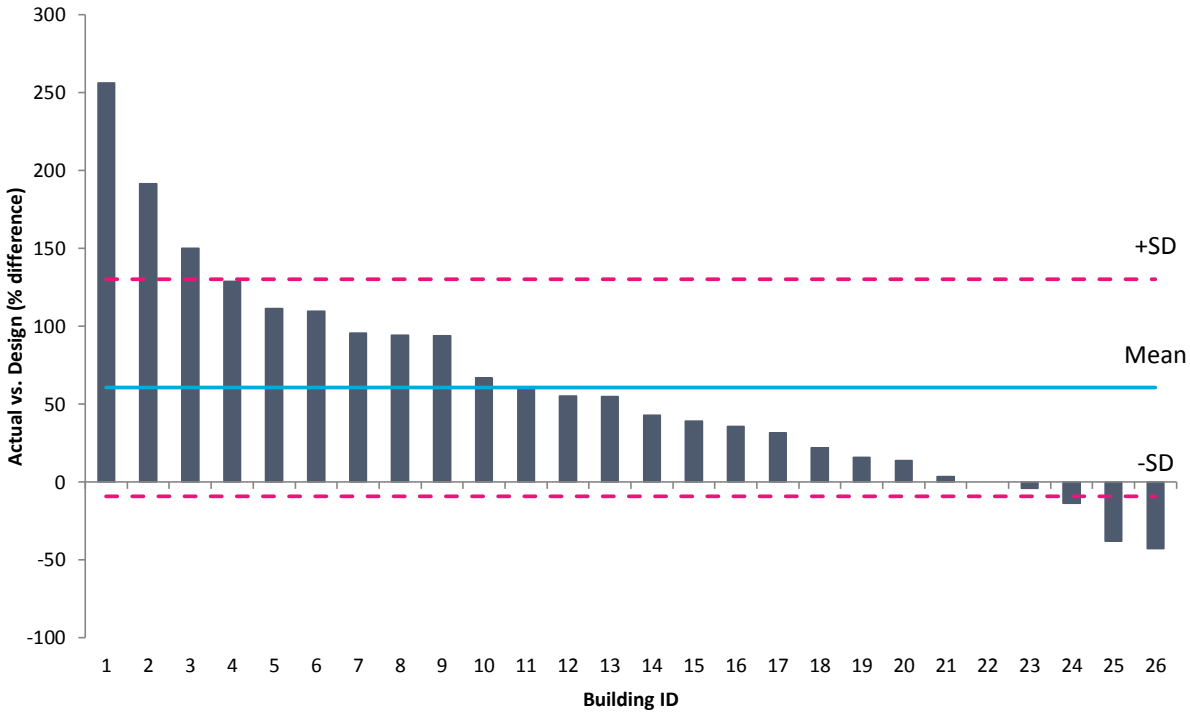

Figure 3.2: Actual versus design carbon emissions for 26 R4F dwellings

The case study dwellings are split between those that use both gas and electricity (one system uses wood and gas) for heating end uses and those that only use electricity, which are termed "all-electric" houses; 10 out of the 26 dwellings are "all-electric. Figure 3.3 shows the distribution of the actual vs. designed carbon performance for these two distinctions of dwelling. The mean percentage difference for gas & electric homes is 55% (SD = 71%) compared to 70% (SD = 70%) for "all-electric" homes.

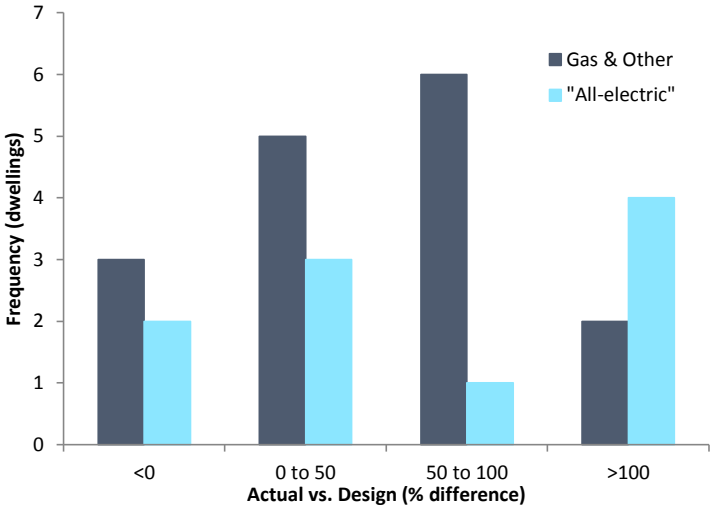

Figure 3.3: Frequency distribution of actual vs. designed carbon performance for gas & electric and "all-electric" dwellings

One reason for this difference in "performance gap" for these types of homes may be the technologies installed to obtain the designed performance: The gas & electric homes tend to make use of different micro-generation technologies to the "all-electric homes; for example although both types of home make use of two major roof based systems, PV and/or solar thermal, the majority of the gas heated homes have solar thermal installations (15/16 homes) as opposed to PV (9/16 homes). Conversely the "all-electric" homes make use of PV more than solar thermal (9/10 versus 6/10 homes respectively); in addition the "all-electric" homes have heat pumps systems installed in 9/10 homes.

Of the fifteen gas & electric homes with solar thermal systems, eight also have PV installed – the distribution of the difference between actual and designed performance for these homes is shown in Figure 3.4.

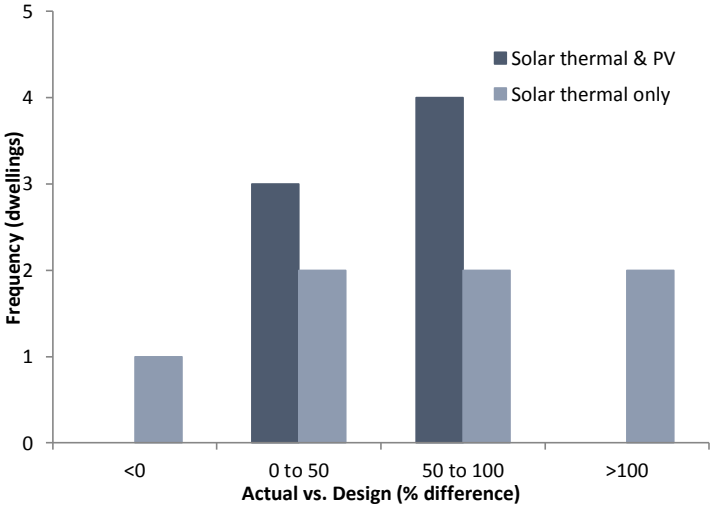

Figure 3.4: Frequency distribution of actual vs. designed carbon performance for gas & electric homes

In addition, one gas & electric home has only a photovoltaic system in addition to the gas boiler and shows a difference between measured and predicted performance of -38%; one system has

ASHP, solar thermal, PV and gas boiler with a difference between actual and designed performance of -4% - these are not included in Figure 3.4.

Figure 3.4 shows a broader range of performance discrepancy for those homes with only solar thermal systems installed (mean percentage difference = 82%, SD = 94%) compared to those with an additional PV installation (mean percentage difference = 49%, SD = 28%).

PV installations are more common for the "all-electric" homes with 9/10 homes having PV installations, eight of these have a heat pump (ASHP or GSHP) in addition with the remaining home using only solar thermal and PV (actual vs. design performance difference = -43%). Four homes have all three renewable technologies (heat pump, PV and solar thermal) installed. The distribution of the difference between actual and designed performance for these homes is shown in Figure 3.5.

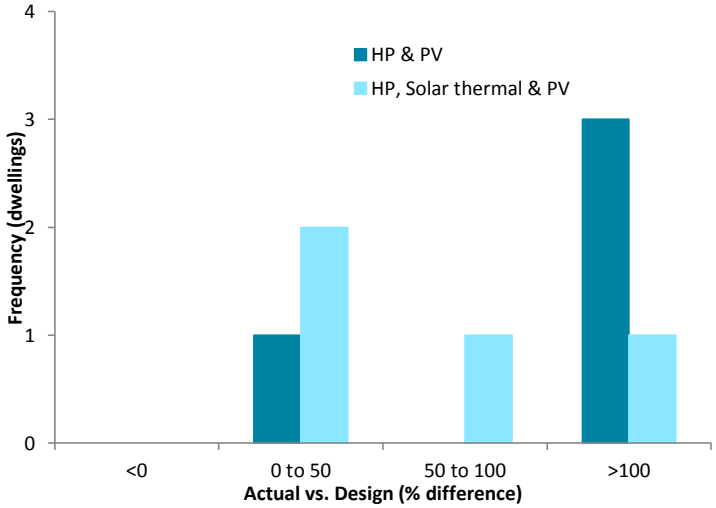

Figure 3.5: Frequency distribution of actual vs. designed carbon performance for "all-electric" dwellings

In addition, one system has only solar thermal and ASHP with actual versus designed performance difference of -14%, which is not included in Figure 3.5. The homes with HP and PV systems only show two distinct points in the distribution with three out of the four homes showing performance differences of >100% - these homes operate ASHP as opposed to a GSHP as is the case for the HP & PV home with 0-50% performance difference. The mean performance difference for the HP & PV homes is 117% (SD = 63%) compared to a mean of 71% (SD = 37%) for the homes operating all three micro-generation technologies. Again, of the homes with all three technologies installed, the one with the smallest performance difference (36%) uses a GSHP versus an ASHP.

The two figures above indicate that one potential reason for a higher performance difference for "all-electric" homes compared to gas & electric homes is the presence of a HP; in particular those systems using ASHP show higher differences between actual and designed performance. With respect to the gas & electric homes with no HP, those using only a solar thermal system in addition to the gas boiler have a broader range in "performance gap" with a higher mean percentage difference in actual and designed performance. It may be inferred from this evaluation that the performance discrepancies experienced by a dwelling in-use compared to the design may be due to building subsystem issues and that those subsystems that are more

intimately connected to occupant energy usage behaviour may present higher performance discrepancies. To investigate this hypothesis further, a systems engineering approach was applied to a case study dwelling from the R4F dataset to identify the subsystem contributions to the "performance gap" of low carbon retrofitted homes. (Hensen 2002) states that a building is to be treated as a whole system and not as a collection of separately designed and optimised sub-systems; dynamic interactions exist between the subsystems in a building and that all should be considered as a whole: The systems engineering approach allows a whole-building system to be viewed as a complex system-of-systems; therefore the performance of the building is dependent on the performance of the subsystems contained within.

3.5 Performance Evaluation of a Case Study Low Carbon Dwelling from a Systems Engineering Perspective

The case study dwelling is a 1950s mid-terrace bungalow, located in Bedfordshire, which has been comprehensively refurbished to improve energy consumption and CO_2 emissions of the property (Figure 3.6). Three renewable energy technologies were installed as part of this refurbishment including a 3.6kW_p PV system; 4.5kW_{th} ASHP; and three 2.1m^2 solar thermal collectors. Energy consumption and generation data has been collected since January 2011.

The collected data was used to evaluate the performance of the building's subsystems to investigate the hypothesis that the difference between actual and designed whole-house performance is a result of performance discrepancies between actual in-use and predicted subsystem performance.

Figure 3.6: The case study dwelling at Bedford

3.5.1 System-of-systems view of the case study dwelling

System-of-systems is a concept in the field of systems engineering that allows a complex system, in this case a low carbon dwelling, to be described by the dynamic interactions between

subsystems and system elements contained within the system of interest (INCOSE 2011). Figure 3.7 shows the system hierarchy of the case study dwelling at Bedford.

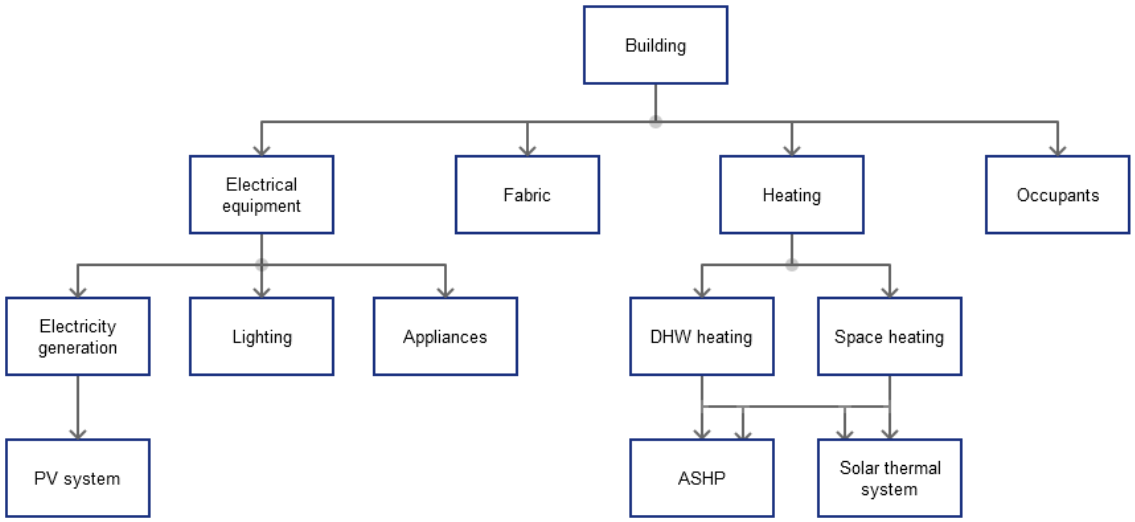

Figure 3.7: System hierarchy for the Bedford dwelling

The system hierarchy in Figure 3.7 is useful in showing which system elements belong to which system; this can be used in order to assign responsibility for design, installation/construction, and post-occupancy support of the different subsystems to different parties involved in the project (INCOSE 2011).

3.5.2 Origins of the "performance gap" associated with building subsystems

Performance evaluations of building subsystems such as the fabric and those installed to provide heating and electrical services are carried out by manufacturers using standard testing methods and conditions. These subsystems are typically tested in isolation from each other (NHBC Foundation 2012; Zero Carbon Hub 2013) and once installed in a building further tests are rarely carried out (Doylend 2014). Performance results obtained under standard test conditions may result in unrealistic estimates of actual performance (NHBC Foundation 2012). Discrepancies between designed and actual building subsystem performance including fabric heat loss parameters and heat pump efficiencies have been reported as contributing factors to overall "performance gaps" (Zero Carbon Hub 2010).

3.5.2.1 Building fabric

The "fabric first" approach to improving building performance includes the installation of insulating materials to the building envelope in order to reduce the thermal losses from the dwelling and thus reduce the space heating requirements and associated energy consumption and carbon emissions. Insulation methods often employed include loft insulation, cavity wall insulation, external and internal wall insulation and insulating the floor space as well as the use of glazing materials with low thermal conductivity. In addition, electrical loads associated with lighting and cooling can also be reduced with appropriate use of glazing and fenestration which can provide high levels of daylight and reflect solar radiation to prevent overheating of the dwelling due to solar gains in the summer months.

The "fabric first" approach is often believed to be a cost effective method of widespread energy and carbon reduction across UK housing stock (Shorrock et al. 2005) and fabric improvements identified by a Green Deal assessment are a prerequisite for the domestic Renewable Heat Incentive (RHI). However, the assumed performance of building fabrics used may differ in reality and thus affect the accuracy of predictions made by building energy models (de Wit 1995). Fabric heat loss was found to be a contributing factor to the "performance gap" of 16 domestic dwellings in a report by the Zero Carbon Hub (Zero Carbon Hub 2010). Unintended fabric losses are also reported by (Gupta & Dantsiou 2013). Assumed U-values for cavity masonry and other wall constructions were found to be higher in reality (in some instances as high as 50% greater than the design assumption (Siviour 1994)). Three causes for higher than predicted fabric heat losses are associated with construction quality: poorly fitting wall insulation; thermal bridging at junctions and air openings; and thermal bypassing through party wall cavities (Zero Carbon Hub 2010).

3.5.2.2 Building services

Building services include the heating and electrical systems including the ASHP, solar thermal and PV systems. Studies related to the performance of heat pumps systems in the UK suggest a general trend for underperformance compared to European counterparts (Boait et al. 2011) due to oversized heat pumps manufactured for the larger European homes; this leads to lighter loading of the heat pump and proportionately greater mechanical losses and parasitic loads associated with the circulation pump – they conclude that in-use performance of heat pumps could be improved with greater awareness from designers of the synergy between heat pumps and smaller, well insulated dwellings; improved controls to cope with the thermal properties of retrofitted UK homes; and smaller heat pump units.

The performance of heat pumps can be greatly affected by the quality of the installation (Miara et al. 2011; Loose et al. 2011) and installation practices (Boait et al. 2011; EST 2010). It is important that all hydraulic components are checked and installed correctly, for example Miara et al. (2011), found that in some systems the valves that switch the operation from DHW to space heating did not close completely which resulted in slow, but continuous draw of hot water; Kelly & Cockroft (2011) discovered that the temperature compensation strategy on ASHPs installed in Scotland had not been enabled due to installer error (Kelly & Cockroft 2011). There is a common practice amongst UK installers to program heat pumps to operate continuously throughout a 24 hour period, discouraging the option for a night-time setback. Setback is a period of time for which the heat pump is switched off in order to allow the dwelling to cool down and improve comfort in the night time. Currently installers often program heat pumps to operate continuously. It is suggested that dwellings with a high thermal time constant and conventional radiators maintain the continuous operation protocol; dwellings with low thermal time constants or radiators with high heat transfer coefficients may receive reduced energy consumption by the heat pump with a setback of 10 hours or more depending on external temperature (Boait et al. 2011). Therefore, control strategies are also a key contributor to heat pump performance. Heat pumps that are controlled based on a room set-point temperature T_s maintain a constant K which is described as:

$$K = \frac{(T_r - T_s)}{(T_s - T_e)} \tag{3.1}$$

Where T_r is the return temperature and T_e is the external ambient temperature. K is often set too high by installers keen to ensure warmth for the household resulting in high radiator temperatures and thus lower COP values; Boait et al. (2011) argue that heat pump controls should automatically set K in operation based on sensor measurements especially since the temperature of well insulated homes is affected by solar, appliance and metabolic heat gains. A common complaint of heat pump control by users is the complexity (Boait et al. 2011; EST 2010). Control issues reported in Miara et al. (2011) include permanent operation of pumps, incorrect charging of the thermal store, unnecessary space heating during the summer. Control plays a significant role in the performance of heating systems especially as the complexity of such systems increases (Loose et al. 2011); in fact the complexity of a system should be kept to a minimum according to results found by the EST (EST 2010).

The Energy Saving Trust carried out a field trial of 83 heat pump systems throughout the UK between April 2009 and April 2010 (Dunbabin & Wickins 2012). The mean system efficiency for 49 ground source heat pumps (GSHP) was 2.4 with a standard deviation of 0.45 (Figure 3.8a) whilst the mean efficiency of 22 air source heat pumps (ASHP) was lower at 1.8 with a standard deviation of 0.28 (Figure 3.8b).

The report details a number of technical factors related to design and installation responsible for lower than expected system performance. These are reproduced in Table 3.1 which provides the estimated reduction in system efficiency (COP) associated with each of the factors. Heat pump under-sizing was estimated to reduce the heat pump performance the most due to increased electrical loads from the auxiliary boost heater used to make up the shortfall in thermal output from the heat pump. Other significant factors include incorrect sizing of different subsystem components and insufficient levels of insulation on the pipework and storage cylinder. Refrigerant and brine leakage was experienced in two systems indicative of poor equipment reliability.

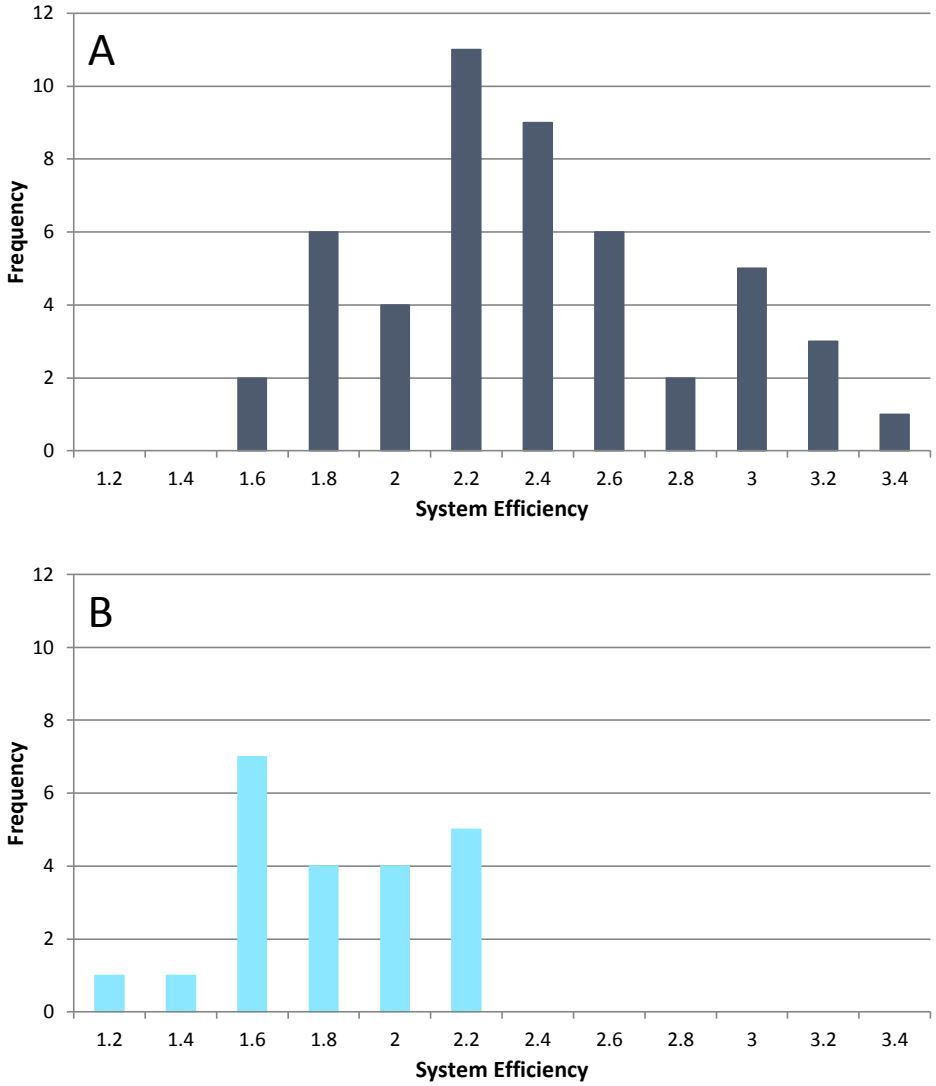

Figure 3.8: Distribution of measured heat pump efficiencies for a) GSHP; and b) ASHP (Dunbabin & Wickins 2012)

Category	**Factor**	**Estimated potential loss of performance as measured by system efficiency**
Design	Under-sizing of the heat pump	Up to 1.5
	Under sizing of the borehole/ground loop	Up to 0.7
	Insufficient insulation of pipework and hot water cylinders	0.3 - 0.6
	Under-sizing of hot water cylinder	Up to 0.4
	Too many circulation pumps	0.1 – 0.3
	Over-sizing/control strategy results in over-use of back-up heating	<0.1
Installation / commissioning	Central heating flow temperature too high: radiators	0.2 – 0.4
	Central heating flow temperature too high: underfloor heating	
	Circulation pump always on	0.1 – 0.3

Table 3.1: Factors influencing heat pump performance (Dunbabin & Wickins 2012)

The EST followed up the first heat pump field trial with a second in which 38 of the original systems were selected for interventions to be made in an effort to improve performance (Dunbabin et al. 2013). These included replacement of incorrectly sized heat pumps and other technical improvements such as software upgrades to reduce auxiliary input and installation of variable speed DC pumps. The mean system efficiencies for the ground source and air source units increased to 2.5 and 2.2 respectively with standard deviations of 0.47 and 0.44.

From 1998 to 2000, the EST evaluated the performance of eight solar thermal systems installed in new homes built in South Wales (EST 2001). Unfortunately, data losses and other monitoring issues meant that performance could only be quantified for three of these systems. The mean annual solar yield for these three systems was reported as 294kWh/m^2/yr with mean annual solar fraction and system efficiency of 55% and 29% respectively. The report compares the performance of these three systems with those installed in four different European countries – the results are reproduced in Table 3.2**.**

	Denmark	**Germany**	**Netherlands**	**Sweden**	**UK (South Wales)**
Annual yield (kWh/m^2/yr)	392	282	643	331	294
Solar fraction (%)	61	49	39	50	55
Annual irradiation in plane of collector (kWh/m^2/yr)	1223	1009	1083	1082	997

Table 3.2: Summary of annual European solar thermal system performance (EST 2001)

The UK systems have one of the lowest annual solar yields due to lower annual irradiation levels. Solar fractions for the Netherlands systems are noted to be low compared to the annual yield figure; although it is unclear why this is it was speculated that this may be due to the tendency of Dutch systems to have high collector area to storage tank volume ratios. It is also important to note that the total number of European systems (excluding the 3 from the UK) was 171 and so the validity of the UK performance comparison is questionable.

A later field trial conducted by Viridian Solar (Forward & Roberts 2008) in 2007 involved six systems, three in Suffolk and three in Sheffield. The mean annual yield was 304kWh/m^2/yr (range 260-364kWh/m^2/yr) with a mean annual solar fraction of 50% (range 26-70%). It was suggested that the performance of these systems was intimately linked to the behaviour of the user regarding the amount of hot water used and how the auxiliary boiler was used in conjunction with the solar thermal system. In well performing systems the solar performance of the system was limited by the amount of hot water used. Low performing systems tended to operate the boiler in competition with the collector limiting the amount of solar energy that can be stored due to maximum store temperatures being reached. The high cylinder temperatures of these systems lead to higher storage losses which were met by the boiler rather than the solar thermal system.

Arguably the most extensive solar thermal field trial was conducted by the EST in which 88 systems throughout the UK were monitored from April 2010 to April 2011 (EST 2011). Results were highly variable with some systems achieving solar fractions upwards of 60% and some as low as 9%; the median result for all systems was quoted as 39%. However, accompanying documents to the collected data reveal that the field trial was plagued with monitoring issues with many systems failing to provide an entire year's worth of data as well as technical issues

with the sensors used. 23 systems provided complete data for the annual solar yield. The mean annual yield for these systems was 275kWh/m^2/yr with a standard deviation of 98kWh/m^2/yr (Figure 3.9).

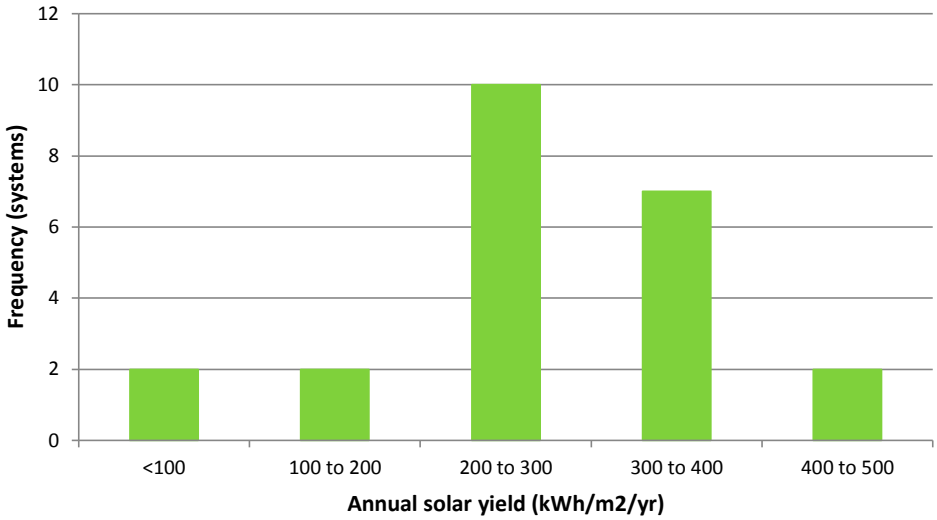

Figure 3.9: Distribution of measured solar thermal yield (EST 2011)

The performance of the solar thermal systems was linked to the way in which the householder used the system and the report advocates the provision of advice to users about how to control the system with respect to volumetric consumption of hot water, timing of auxiliary heaters and hot water draws and water temperatures to ensure optimum performance. Insufficient insulation installed on some storage tanks and pipe work was found to reduce the amount of hot water the solar collector could deliver. Technical faults with circulation pumps caused them to run when no solar energy was available increasing electrical parasitic loads to 23% of the heat delivered for one system. The report indicates that the performance of solar thermal systems is dependent on installation quality as well as occupant energy usage behaviour and interaction with the entire heating system including the auxiliary heater.

Two case studies were conducted by O'Flaherty & Pinder (O'Flaherty & Pinder 2011): the first evaluated the performance of 25 solar thermal systems in the West Midlands. Two systems appeared to show exceptionally high performance considered unlikely; however the remaining systems demonstrated a range of performance from 23-1541kWh/yr. The second case study presents the performance of 10 systems with high occupancy levels located in South Yorkshire; the annual solar yield ranged from 200-1990kWh over a two-year period with no system obtaining the design target of 2128kWh/2-years. Households with well-timed auxiliary firings had higher performing systems than those that did not alter their auxiliary firings or hot water usage patterns.

A field trial of 274 individual PV systems installed in 17 sites across the UK was conducted to evaluate the performance of these systems over the period 2000-2005 (Munzinger et al. 2006). The most common yield was found to be 700-800kWh/kW$_p$ (Figure 3.10).

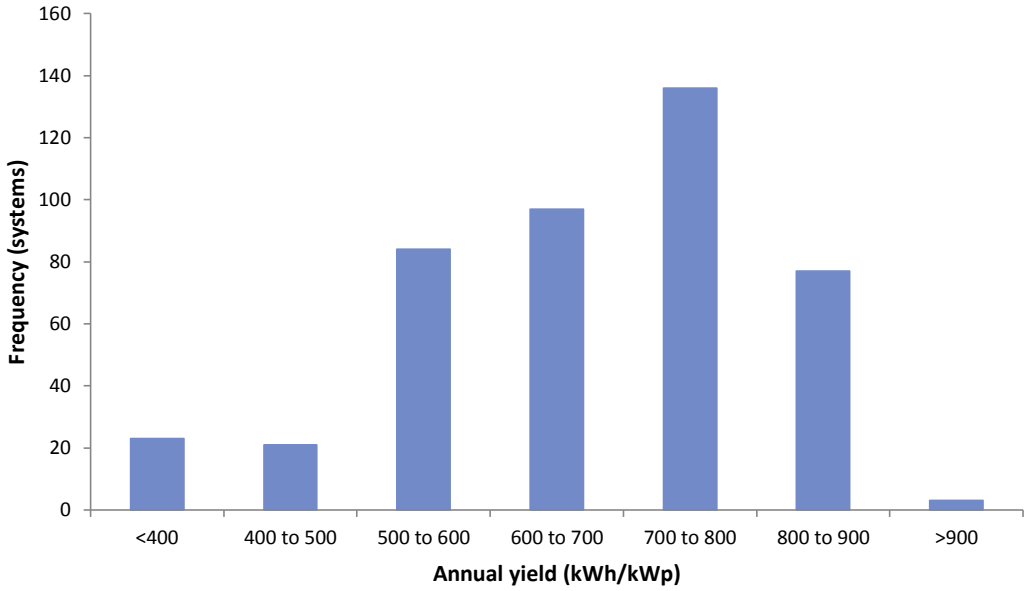

Figure 3.10: Distribution of measured PV system yield (Munzinger et al. 2006)

12 of the 17 sites provided two years' worth of performance data hence why Figure 3.10 contains 441 values. Two points to consider about the wide variety of performance values are that some values are not from the same annual period and different tilt and azimuth angles are represented in the dataset. Predicted yields for these sites suggested an annual yield of 740-840kWh/kW_p to be the most common with northern sites not seeing the predicted drop in performance associated with lower irradiation values. Technical issues associated with the inverter caused low performance in some systems as well as some modules having lower than the nominal rating quoted by the manufacturer. Design and installation issues were identified with shading of the array reducing the yield of some systems.

A case study by O'Flagherty & Pinder evaluated the performance of 23 PV systems in South Yorkshire (O'Flaherty & Pinder 2011). The mean yield over a two-year period was 999kWh/kW_p/2-years with a standard deviation of 352kWh/kW_p/2-years. The total amount of electricity generated over the two-year period ranged from 1512-5717kWh/2-years. None of the systems obtained the design target of 6000kWh generated over two years. Shading issues and unexplained system shutdowns were responsible for low performance in some systems with vandalism to the external power supply occurring at one location.

3.5.2.3 Occupant usage behaviour:

The occupants represent a living subsystem within the whole-house; their behaviour is stochastic by nature and therefore difficult to represent in building and subsystem energy models. Assumptions and simplifications about occupancy levels and usage patterns made at the design stage can therefore introduce differences between estimated and actual building energy performance (Menezes, 2012). In addition some occupant energy usage may be unaccounted for altogether; for example the absence of unregulated loads associated with non-heating and lighting end uses from models based on the National Calculation Method (NCM) (BRE 2009). The internal environment of a building can be influenced by occupants in ways not considered in the design stage for example the opening of windows which will undermine air-tightness strategies employed to improve the thermal performance of the building (Demanuele et al. 2010). The

performance of heating subsystems is also dependent on occupant interaction as discussed above and can lead to wide variation in system performance.

The "rebound effect" is an important issue related to occupant energy use behaviour. The rebound effect is the reduction of potential energy savings caused by the occupant through a change in energy use behaviour (NHBC Foundation 2011). A typical example of this is occupants leaving lights on for longer because they have been replaced with energy efficient bulbs or having the house at a higher internal temperature because it has been insulated. Rebound effects associated with space heating have been estimated at 20-30% (Haas & Biermayr 2000) whilst (WBCSD 2007) quantify the rebound associated with different end uses, reproduced in Table 3.3.

Energy end use	Magnitude of rebound
Space heating	10-30%
Space cooling	0-50%
Lighting	5-20%
Water heating	10-40%
Automobile	10-30%

Table 3.3: Magnitude of rebound effect associated with different energy end uses (WBCSD 2007)

The reliability of the rebound effect data is questionable due to methodological weaknesses with the associated studies as a result of low sample size, short monitoring periods, and the lack of a control group amongst other issues (Sorrell 2007).

3.6 Monitoring

An extensive monitoring platform was installed in the dwelling so that post-retrofit building performance could be evaluated. The monitoring stage of the project was outsourced to a third party (Automated Building Controls Limited) who installed the monitoring equipment and setup the wireless server. For full evaluation of the whole-house performance the following was measured:

- Mains electricity consumption (kWh)
- ASHP electricity consumption (kWh)
- Solar thermal pump and controller electricity consumption (kWh)
- PV electricity generation (but not the amount exported to the grid or used in-house) (kWh)
- ASHP heat delivery (kWh)
- Solar thermal heat yield (kWh)
- Three tank temperatures (top domestic buffer, bottom domestic buffer and solar portion) (°C)
- Internal temperatures (average, living space, kitchen and bedroom) (°C)
- Flow temperature to radiators (but not return) (°C)
- DHW volumetric consumption (L)
- Space heating requirement (kWh)
- CO_2 levels (living space, kitchen and bedroom) (ppm)
- Humidity (living space, kitchen and bedroom) (%)

- External temperature (°C)
- External humidity (%)
- Irradiance (W/m^2)

As part of the R4F competition requirements the measured data would be made available to other researchers and can be located here (TSB 2013b). A five-minutely time step for the measured data was the requirement laid out by the EST; however some of the quantities in the Bedford case study were measured every ten minutes for reasons unknown. This is indicative of a lack of communication between the data analysis team, the company responsible for installing the monitoring platform and the project coordinators. Furthermore, in order to improve the level of technical analysis that could have been performed, it would have been beneficial to measure the following quantities:

- Amount of PV generated electricity exported and used in-house (kWh)
- Solar thermal module temperature (°C)
- Solar thermal collector exit temperature (°C)
- Cold water temperature (°C)
- DHW delivery temperature (°C)
- Return temperature from radiators (°C)

Knowing the amount of PV-generated electricity used in-house would have been useful in determining the overall electricity consumption of the dwelling; this would help to establish whether the retrofit options installed reduced the total amount of electricity consumed. Solar thermal module temperature would have allowed the efficiency of the collector to be plotted against T*/G (where T* is the difference between the module and ambient external temperature, and G is the irradiance (W/m^2)); this would allow a comparison to the zero loss efficiency (plotted against T*/G) to determine a parameter analogous to the second law efficiency of heat pumps (i.e. to determine how close to the zero loss efficiency the collector is operating). The collector exit temperature could have been used to calculate the energy delivered by the system to validate the measurement from the heat meter. Cold water and DHW delivery temperatures would have been useful in evaluating the effects of these quantities on the combined heating system. Return temperatures from the radiators could have been used to confirm suspicions of high return temperatures reducing the performance of the heat pump and solar thermal system. By not knowing these quantities, analysis of the technical performance of the systems is limited and hypotheses about complex system interactions cannot be readily investigated to an adequate level of detail.

It is important for continual communication between data analysts and monitors (if they are different parties) to exist throughout the project so that the right level of performance analysis can be performed. This would have also helped to specify a more suitable resolution for ASHP and solar thermal performance measurements: The measurement of ASHP and solar thermal heat delivery was made using a Sontex Supercal 531 (Sontex 2003). The resolution of this device is 1kWh when measuring energy and registers a pulse when 1kWh of energy has been delivered by the ASHP and solar thermal systems. The disadvantage of this is that due to the low measurement resolution cycling of the ASHP and solar thermal systems cannot be observed. It is therefore unknown as to how these systems are interacting with each other and how often they

are operating throughout the day. A second limitation of such a low measurement resolution is that some of the energy delivered by the system on a particular day may not be included in the daily sum of energy. For example, if the ASHP delivers 4.6kWh of energy on a certain day then this will in fact be registered as 4kWh on that day; the 0.6kWh remaining will be included in the following day's energy total. This is particularly significant when analysing the solar thermal system performance as this may be linked to the temperatures in the tank which are affected by auxiliary heat input from the ASHP and DHW draws which recharge the tank with cold water. These effects are difficult to observe when the measured input from the solar thermal system may not be representative of the amount of energy actually delivered (a period of expected low solar input may in fact register more than expected due to the 1kWh being eventually reached), therefore, a Wh resolution would have been far more useful.

Another issue with the monitored data was the occurrence of downtime when for unknown reasons the server was unable to collect data. This occurred during the first year of monitoring. For analytical purposes the second year's data was used. This is for two main reasons: firstly the issues with monitoring were solved by this time and therefore the second year's data is complete; secondly, the renewable energy systems installed had one year to "settle down" meaning any peculiarities in system behaviour due to installation are minimised.

3.7 Whole-Building Performance

The case study dwelling is an "all-electric" house with pre-retrofit electrical and heating requirements met by mains electricity. Post-retrofit heating and electrical loads are supplied by three micro-generation technologies in an effort to reduce the consumption of mains electricity - the micro-generation systems are:

- 3.68kWp PV system;
- 6.3m^2 flat plate solar thermal system;
- 4.5kW CO_2 ASHP;

Improvements to the building fabric were also made to reduce space heating requirements.

The annual performance of the whole-house is predicted by the mains electrical energy consumed in a year and the associated carbon emissions based on SAP 2009 carbon data (BRE 2010). Pre-retrofit whole-house performance is based on energy consumption data obtained from March 2008 to June 2009. Post-retrofit modelled whole-house performance was obtained from several SAP analyses and mathematical modelling using TRNSYS (University of Wisconsin 2013). Post-retrofit measured whole-house performance was obtained from analysis conducted using measured data from January 2012 to December 2012. Results are summarised in Table 3.4:

	Pre-retrofit (Measured)	Predicted Post-retrofit	Actual Post-retrofit
Annual mains electricity consumption (kWh/year)	6593	1998	4503
Total annual carbon emissions (kgCO2/year)	3888	1200	2328

Table 3.4: Summary of measured and predicted annual whole-house energy consumption and carbon emissions

The enhanced building fabric and micro-generation technologies have improved the whole-house performance of the case study dwelling by reducing mains electricity consumption and

the associated carbon emissions by 32% and 40% respectively (Figure 3.11). However, despite a measurable reduction in carbon emissions, the dwelling fell short of the R4F and designed reduction targets (80% and 69% reduction in carbon) by 100% and 72% respectively.

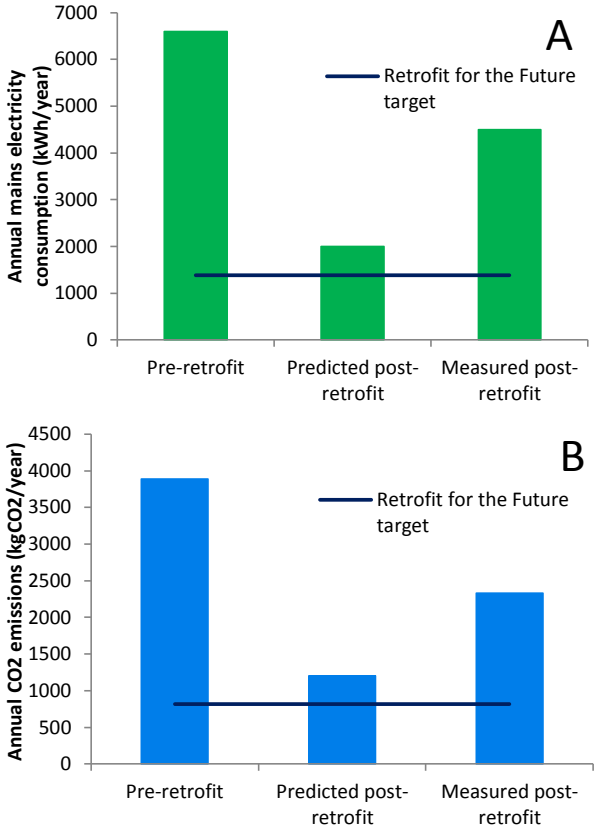

Figure 3.11: a) annual mains electricity consumption; b) corresponding annual carbon emissions

3.8 Subsystem Performance

The discrepancy between actual and designed performance of the case study dwelling is responsible for the failure to meet the designed and R4F carbon reduction targets. The mains electrical energy consumed by the dwelling is 94% higher than the design target. Significant differences in the expected and actual performance of the micro-generation systems is shown in Figure 3.12; in particular the space and water heating system is found to be underperforming compared to design estimates with the ASHP consuming almost five-times as much electricity as predicted and the solar thermal system providing 39% less thermal energy than modelled. The PV system on the other hand generates 11% more electrical energy than predicted. Although other electricity end-uses were not measured directly, the absence of unregulated loads from the mains consumption prediction is likely to introduce uncertainty about actual consumption figures. As discussed previously, differences between actual and assumed fabric performance and occupant usage patterns may introduce discrepancies.

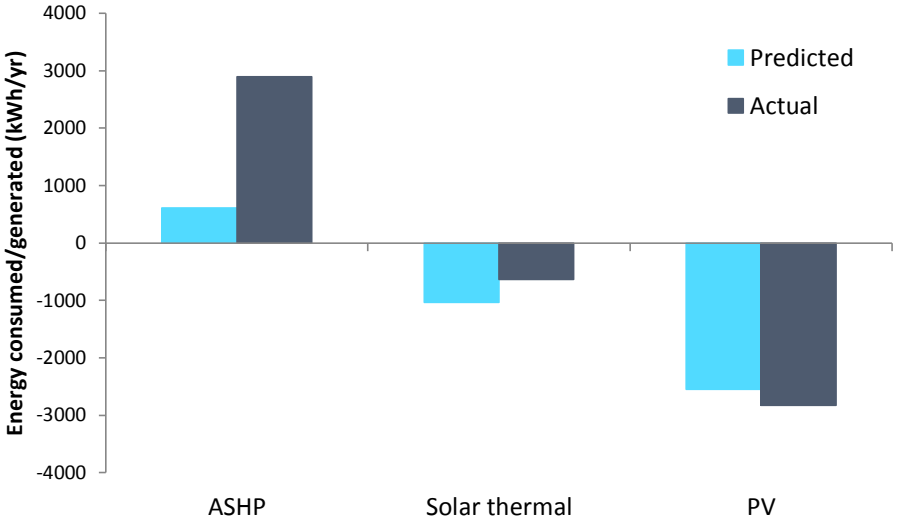

Figure 3.12: Actual vs. predicted micro-generation subsystem energy consumption

3.8.1 The PV system

The photovoltaic system is comprised of two different arrays mounted on east and west orientations with an inclination matching the roof pitch (30°):

- East facing array: 30*0.06kW$_p$ Abound AB1 thin film panels giving net rating of 1.8kW$_p$ (area = 21.6m^2)
- West facing array: 8*0.235kW$_p$ Yingli Solar YL235 crystalline panels giving net rating of 1.88kW$_p$ (area = 13.1m^2)

The overall rating of the system is 3.68kWp with a total area of 35.7m2.

The system inverter is a Power One PVI-3.6-OUTD and has an output of 3.6kW and an efficiency of 96%. Figure 3.13 shows the system diagram of the PV installation:

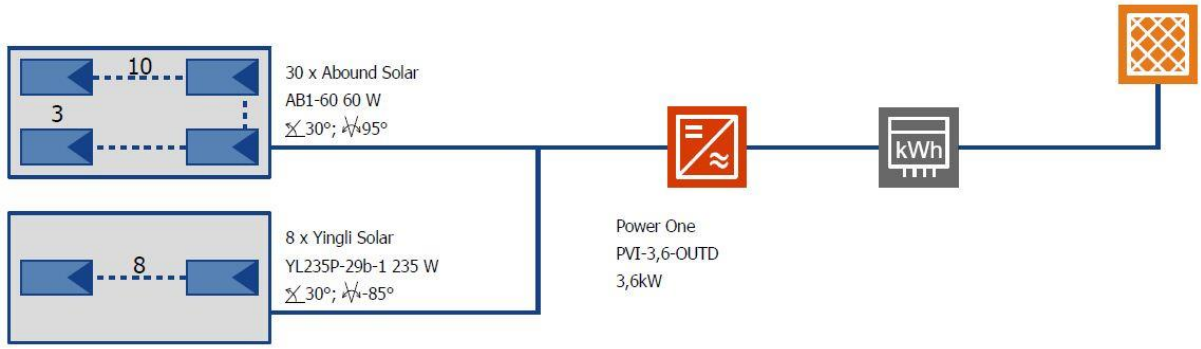

Figure 3.13: PV System diagram

3.8.2 PV system performance results

The predicted annual PV yield was obtained using a mathematical model by the manufacturer this is presented next to monitored data for 2012 in Table 3.5.

	Modelled	Measured
Annual electricity generated (kWh/year)	2552	2825
Annual specific yield (kWh/kW$_p$)	689.6	767.6
Annual system efficiency (%)	8.4	8.4

Table 3.5: Summary of the modelled and measured annual PV system performance

Figure 3.14 shows the predicted and measured monthly electricity yield generated by the PV system.

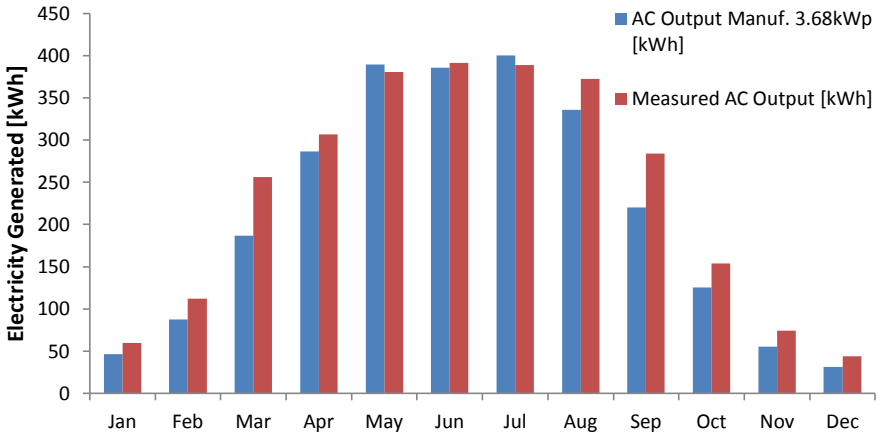

Figure 3.14: Modelled versus measured monthly PV yield

The actual annual PV yield is 10.7% greater than the predicted value; however there is no difference between predicted and measured system efficiency. Consider the equation for annual yield:

$$Annual\ PV\ Yield\ (kWh) = \eta_{sys} \times Annual\ incident\ irradiation\ (kWh) \qquad (3.2)$$

Since the predicted and measured system efficiency are equal then any difference between the predicted and measured annual PV yield may be a result of differences between modelled and measured irradiation. The measured irradiation amounted to 33648.55kWh/year which is 11.8% higher than the value used in the model (30089kWh/year); this indicates that one of the causes of the performance gap for the PV system is the inability of the model to represent actual weather conditions.

3.8.3 The combined solar thermal and air source heat pump system

In the Bedford dwelling hot water and space heating are provided by a combined solar thermal and ASHP heating system. According to the definitions laid out in the IEA SHC Task 44/HPP Annex 38 (Hadorn 2012) the Bedford system is a parallel system whereby the solar thermal and ASHP are connected in parallel to a combined storage tank and deliver heat directly; this is in contrast to a series system in which the solar thermal system feeds into a store which in turn acts as the heat source for the heat pump (the two systems are thereby connected in series). Two other combined system configurations exist: regenerative, where the solar thermal system warms the ground for GSHPs; and complex which is a combination of the three previous configurations (the solar thermal system then delivers heat directly to the combination store, delivers heat to a tank to act as a source for the heat pump, and regenerates the heat in the ground for GSHPs).

(Hadorn 2012) developed the square view energy flow diagram to represent and simplify the complex energy flows in combined systems; these are easier to comprehend than the classic system schematic. Figure 3.15 shows a square view diagram of the energy flows through the parallel combined system; the driving energy is transferred through the system using three different media: the ASHP refrigerant (CO_2), the solar thermal working fluid (water/glycol mix) and water delivered from the heat pump and storage tank to the end uses.

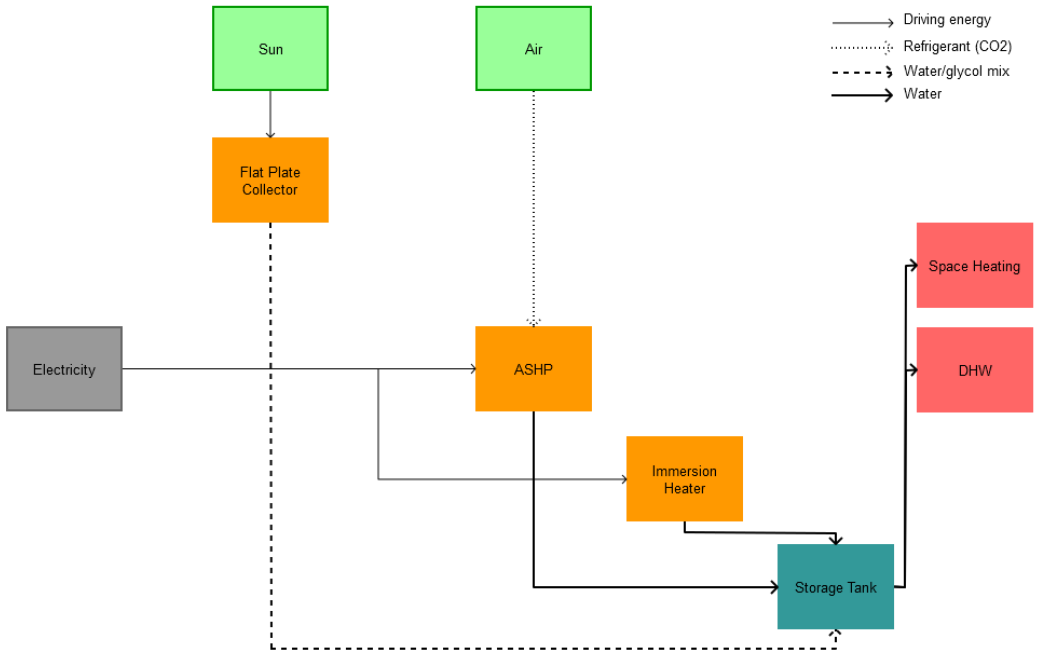

Figure 3.15: Square view energy flow diagram of the combined solar thermal and ASHP heating system (parallel system)

3.8.3.1 Modes of operation and control strategy

The possible modes of operation for the combined heating system are:

- Charging of the tank by the solar thermal system only;
- Charging of the tank by the ASHP only;
- Charging of the tank by the solar thermal system and ASHP;
- DHW provision from the tank;
- Space heating from the tank;
- Defrosting of the ASHP evaporator

All heating demands are satisfied by drawing from the combined storage tank – neither the heat pump nor solar thermal systems provide direct space heating. During the summer months, when the space heating demand is zero, the solar thermal system will only provide DHW; this may be with or without contributions from the ASHP depending on the amount of hot water required and the temperature in the tank. During the heating season the contribution from the solar thermal system is reduced due to lower irradiance, shorter days and colder external temperatures. At these periods of the year the ASHP delivers the majority of the heat used for DHW and to meet the increased space heating demand; however there may be periods during the heating season when solar contributions are made which will preheat the combined storage

tank and reduce the heat requirement from the ASHP; in this way the solar thermal system contributes to the DHW and space heating demands.

Figure 3.16 shows the layout of the combined store with inputs from the ASHP and solar thermal system.

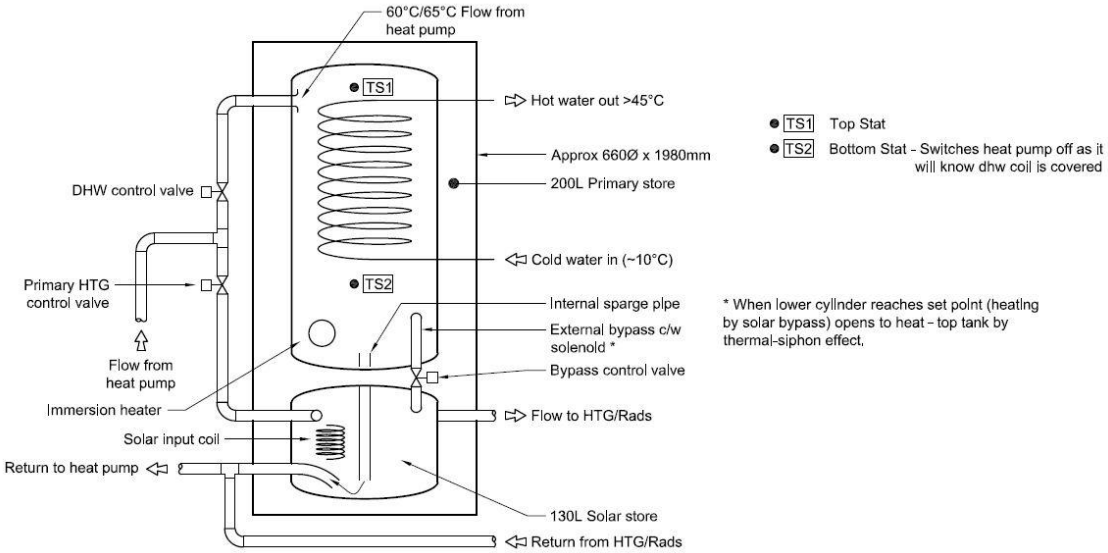

Figure 3.16: Schematic for the combined store

Water heated at 60-65°C by the ASHP is delivered to the top of the DHW buffer; when the temperature sensor below the DHW coil reads 65°C the ASHP switches off. When the ASHP is in space heating mode lower temperature water is delivered to the solar buffer, where it flows to the heating circuit. Cool water from the bottom of the solar buffer returns to the ASHP. When the solar buffer water temperature reaches 40°C the bypass valve opens to allow the solar heated water to enter the DHW buffer; however this valve may be activated during days of no solar contribution due to input from the ASHP. Table 3.6 describes the performance characteristics of the combined heating system.

Component	
Solar thermal collector	**Viridian Solar Clearline V20**
Type	Flat plate collector
Area	3x2.1m^2
Zero loss efficiency	0.81
Heat loss coefficients	a_1 = 3.9W/m^2.K a_2 = 0.0089W/m^2.K^2
Heat pump	**Sanyo Eco-Cute SHP-C45DEN**
Type	CO_2 ASHP
Heating capacity	4.5kW
COP (outdoor temperature 7°C, flow/return temperatures 65/35°C)	2.4
Storage tank	**Ecocat 300**
Type	Combined store
Volume	330L (200L domestic buffer + 130L solar buffer
Standing losses	2.2kWh/24h

Table 3.6: Performance specification of components in the combined heating system

3.8.4 CO_2 ASHPs

CO_2 ASHPs use CO_2 as the refrigerant flowing through the heat pump circuit. The thermodynamic properties of CO_2 as a heat pump refrigerant mean it is suited to raising the temperature of water from low temperatures to that suitable for DHW requirements whilst maintaining high coefficients of performance (COP) compared to conventional heat pumps (Stene 2007). This has made CO_2 ASHPs popular in Japan where living space is limited, making GSHP and large water tanks impractical, and a tradition of bathing in "onsen" exists which require large volumes of hot water (Chen et al. 2009).

In contrast to some Japanese homes, Northern European homes have lower DHW requirements with a greater demand for space heating due to cold climates (Chen et al. 2009). Space heating requires lower delivery temperatures than DHW applications especially with the presence of underfloor heating; therefore heat pumps that use conventional refrigerants may be preferable due to lower costs and high COPs associated with low sink temperatures. On the other hand the reduced space heating demand of well insulated and air tight homes means that DHW demand can make up to 50-85% of all heating requirements for which CO_2 heat pumps may prove beneficial (Stene 2007).

The case study property is a small, well insulated, air tight dwelling with a heat recovery system in place and so space heating demands are low (606kWh measured in 2012) in comparison to DHW demands (estimated at 1085kWh based on measured volumetric consumption and tank temperatures in 2012); for this reason a CO_2 ASHP is expected to be well suited to the Bedford dwelling although potential oversizing of the unit is an issue for small well insulated dwellings (Boait et al. 2011). All heating demands at Bedford are met by heat from the combined storage tank – the ASHP does not provide separate direct space heating.

3.8.5 Performance indicators

The parallel system configuration used in the Bedford property mean that the solar thermal and ASHP systems work independently from each other, therefore the performance of these individual subsystems will be evaluated using performance indicators relevant to each. Additionally since the performance of the combined heating system depends on the interaction of these subsystems and the control strategies employed (Loose et al. 2011), combined system performance will also be evaluated.

3.8.5.1 Solar thermal system performance indicators

Commonly used measures of performance for solar thermal systems are solar fraction, system efficiency and annual yield. The annual yield, symbolised by Q_{sol}, is simply the amount of energy delivered into the tank from the collector in a year measured in kWh. This energy may be used by the occupant to meet heating demands as well as be lost during distribution and storage. Since storage and distribution losses are a fundamental end-use of any heating system and must be covered by the system they will not be excluded from the annual yield. The system efficiency (η_{sys}) is the ratio of desired output over required input. For a solar thermal system the desired output is the annual yield and the required input is the irradiation incident (S) on the collector measured in kWh/m^2 multiplied by the collector area (A):

$$\eta_{sys} = \frac{Q_{sol}}{SA} \tag{3.3}$$

The annual solar fraction (SF) is the ratio of the heat delivered by the solar thermal system in a year to the total heating demand including losses:

$$SF = \frac{Q_{sol}}{Q_{DHW}+Q_{SH}+Q_L} \tag{3.4}$$

Where Q_{DHW} is the energy content of the hot water, Q_{SH} is the energy required for space heating and Q_L is the storage and distribution losses in the system. Balance of the energy flows into and out of the combined store gives:

$$Q_{sol} + Q_{aux} = Q_{DHW} + Q_{SH} + Q_L \tag{3.5}$$

$$SF = \frac{Q_{sol}}{Q_{sol}+Q_{aux}} \tag{3.6}$$

Where Q_{aux} is the energy required from the auxiliary system to meet remaining heating demands not covered by the solar thermal system.

3.8.5.2 ASHP performance indicators

The performance of the ASHP can be described using the amount of heat delivered (Q_{ASHP}), the amount of electrical consumption (W_{ASHP}) and the coefficient of performance (COP).

Similar to the definition for annual yield for a solar thermal system the annual heat delivered by the heat pump is that used to meet total heating demands including thermal losses. The electricity consumption provides information about the amount of mains (or renewable) electricity used to power the heat pump components including the compressor, control system and fans/pumps; the electricity consumed and heat delivered by the heat pump are used to calculate the coefficient of performance:

$$COP = \frac{Q_{ASHP}}{W_{ASHP}} \tag{3.7}$$

The COP of a heat pump is effectively the same quantity as efficiency: output over required input; however the function of the heat pump is to provide more thermal energy than electrical energy consumed leading to efficiencies of greater than 100%. For example a heat pump with a COP of 3 will provide 3kWh of heat for every 1kWh of electricity it consumes.

3.8.5.3 Combined system performance indicators

The combined system can be described using the seasonal performance factor (SPF). The seasonal performance factor again measures the desired output over required electrical input for the entire system. Depending on which system components are included in the definition of SPF different SPF values can be calculated:

$$SPF_0 = \frac{Q_{ASHP}}{W_{ASHP}} \tag{3.8}$$

$$SPF_1 = \frac{Q_{ASHP}+Q_{sol}}{W_{ASHP}+W_{sol}} \tag{3.9}$$

Where W_{sol} is the electrical energy required to operate the pumps and controllers of the solar thermal system. SPF_0 is simply the COP of the heat pump system and so from this point forward SPF will be used to refer to SPF_1.

3.8.6 ASHP and solar thermal performance results

Table 3.7 indicates that both the solar thermal system and ASHP perform worse than predicted. Estimated annual solar yield and system efficiency is 63% and 6.1% greater than measured respectively; this leads to an increased requirement from the ASHP to meet heating demands reflected in the comparatively low solar fraction for the measured system. Indeed the measured ASHP heat energy is higher than predicted as is the electricity consumption, due in part to the lower than predicted solar yield; however the amount of heat delivered by the ASHP is substantially higher than predicted at over three times the amount. Consequently the electricity consumption of the actual ASHP is also much higher than predicted; this is made worse by the lower measured COP value. Furthermore the measured and modelled annual COPs are lower than those obtained by the manufacturer during a field trial (Sanyo n.d.); however discrepancies are expected between results obtained from controlled experiments and from in-situ performance where variations in daily and seasonal outdoor temperatures affect the source temperature and may trigger defrost operations to occur (Loose et al. 2011).

Component	Modelled	Measured
Solar thermal collector		
Annual yield	1032.9kWh	634.0kWh
Annual system efficiency	16.8%	10.7%
Annual solar fraction	52.4%	16.6%
Heat pump		
Annual heat delivery	939.7kWh	3176.5kWh
Annual electricity consumption	612.3kWh	2892.2kWh
Annual COP	1.5	1.1
Combined system		
Annual SPF	2.76	1.28

Table 3.7: Summary of modelled and measured system performance

The high electrical consumption and low COP of the ASHP combined with the low solar yield of the solar thermal system leads to a SPF value of the actual combined system 46% lower than predicted.

Figure 3.17 shows solar thermal yield, ASHP heat delivered and seasonal performance factor for the combined heating system in the case study dwelling. The data shows peaking of SPF and Q_{sol} between May and August due to increased daily irradiation levels and external temperatures. Higher external temperatures and solar gains in the summer months lead to negligible space heating requirement which combined with increased solar thermal yields leads to an overall reduction in ASHP heat delivery and associated electricity consumption and subsequently higher solar fractions (Figure 3.18). The SPF of the combined heating system increases throughout the year until it peaks in August; this rise in SPF is due to increasing solar input for the aforementioned reasons and decreasing electricity consumption associated with reduced heat pump requirement. However this SPF could be improved if the solar yield and COP was higher leading to lower electricity consumptions for the same heat output.

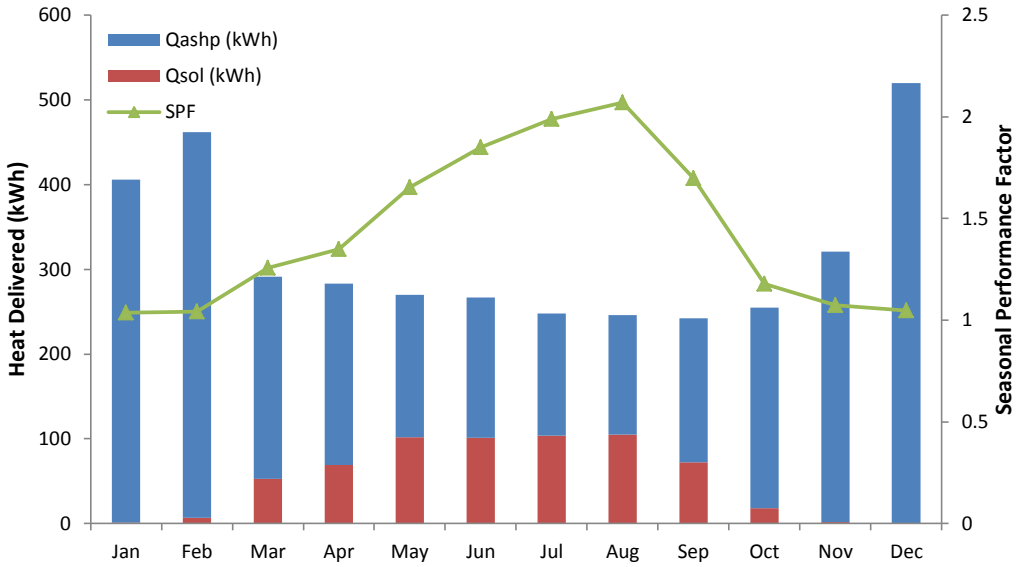

Figure 3.17: Monthly performance of the combined heating system

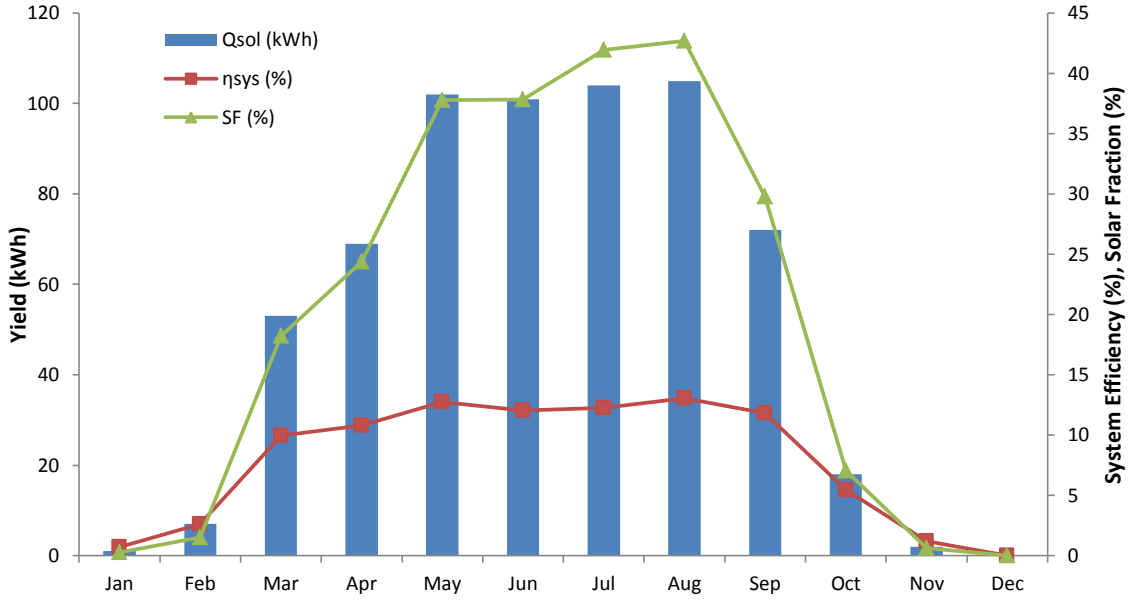

Figure 3.18: Monthly performance of the solar thermal system component

The efficiency of the solar thermal system increases from January to May after which the efficiency values plateau with slight variations between months; this is due to consistent irradiation levels and yields throughout these months. The efficiency drops off after September due to lower irradiation values. Although July sees the highest irradiation of any month (134.52kWh/m^2) with the second highest external temperature (17.27°C) the yield and efficiency are not the highest because the average daily DHW consumption of this month is the third lowest at 47L/day; there is evidence to suggest that solar yield increases with DHW consumption (Allen et al. 2010; Knudsen 2002). In contrast the peak efficiency, solar fraction and yield occur in August when the irradiation is 127.69kWh/m^2, the external temperature is 18.55°C (highest of any month) and the average DHW consumption is 65L/day (also highest of any month).

Figure 3.19 shows that the COP steadily increases throughout the year and peaks in August; the higher external temperatures mean that the compressor has to do less work to raise the temperature of the CO_2 to the desired temperature for heating demands, this leads to higher COPs. Furthermore, the higher external temperatures and lower humidity result in a reduced risk of frost formation on the evaporator; this maintains the heat transfer performance of the evaporator and minimises the number of defrost cycles being performed both of which limit ASHP performance.

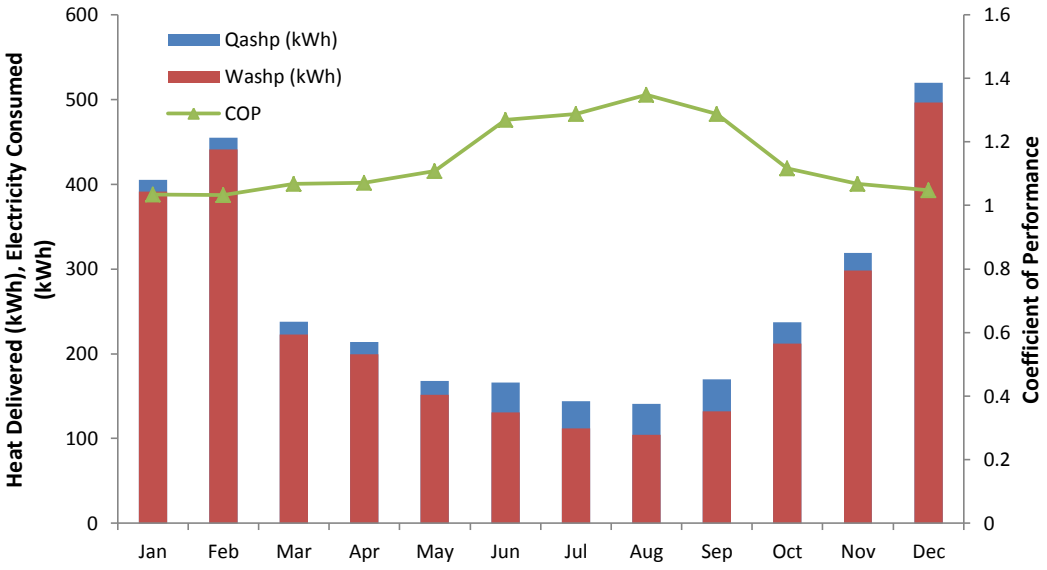

Figure 3.19: Monthly performance of the ASHP component

The case study combined heating system performs worse than expected with the heat pump consuming nearly five-times the amount of electricity predicted and the solar thermal system delivering 39% less thermal energy than modelled. The resulting measured annual solar fraction is 17% versus the predicted 52%. The heat pump is delivering 3.38 times the amount of heat as estimated resulting in a lower COP of 1.1 versus 1.5 obtained from the TRNSYS model and approximate COP of 3 from field trial results (Sanyo n.d.). The resulting measured SPF is 54% lower than the predicted value. Reasons for differences between measured and modelled results will be discussed in the next section.

3.9 Causes of the Combined Heating System Performance Gap

The performance gap is a result of the inability of the model to represent the real life condition (Menezes et al. 2012); this inability is due to assumptions made during the design stage, limitations of the model, the random nature of operating conditions (occupancy behaviour, weather patterns), installation quality, control strategies and unforeseen design changes made at the installation stage. This section aims to identify the differences between the model and actual system that contribute to the performance gap.

3.9.1 Design assumptions

Dynamic simulations such as the one used to model the combined heating system often require large amounts of detailed information about the input parameters of each of the components in the system (Zhao & Magoulès 2012); however this information may not always be known to the

modeller. Table 3.8-Table 3.10 show the required model parameters for the combined heating system and building envelope compared to those available to the modeller using manufacturer's technical data; the blank spaces show parameters for which no information is found. With regards to the solar thermal collector the performance can be described using the absoptance and emittance values or the heat loss coefficients and zero loss efficiency. Some manufacturers provide information about the aborptance and emittance, however this information is not provided by Viridian Solar. This means that the values used in the model are assumed values and may not correspond to the actual performance of the collector.

Apart from assumptions required due to a lack of data, further differences in the system parameters may be due to unknown information about the design as a result of a lack of communication between the design team and the modeller or due to changes in the design made after the model was completed (Bordass et al. 2004). Evidence of this is present in Table 3.8-Table 3.10 where differences exist between the modelled and installed ASHP make and model; combined store; glazing; and building fabric u-values. Differences in the u-values suggest that the actual building will have lower thermal losses than the modelled dwelling and that this will have an impact on the internal temperature and therefore heating demands provided by the ASHP and solar thermal system. Furthermore, differences in the thermal properties of the dwelling may also affect the control and performance of the combined heating system, especially considering that the modelled Mitsubishi Ecodan is not a CO_2 heat pump and therefore more appropriate for meeting the space heating demands of the modelled dwelling with greater thermal losses. Low thermal losses from the storage tank (much less than measured and average values which are typically more than 100 times than the assumed value) mean that less energy is required from the ASHP to ensure the set point temperature is met, in addition a 30° pitch on the solar thermal collector results in more solar energy being captured by the system compared to the actual pitch of 50° in the summer months; therefore, the solar thermal collector delivers more heat to the combined store and further reduces the input from the AHSP.

Component	Model	Actual
Solar Thermal System		
Type	Flat plate collector	Flat plate collector
Fin efficiency	0.7	
Edge loss coefficient	3.0kJ/hr.m^2.K	
Absorber plate emittance	0.15	
Absorptance	0.9	
Index of refraction of cover	1.526	
Extinction coefficient	0.0026	
Zero loss efficiency		0.81
Heat loss coefficients		k1 = 3.9W/m^2.K k2 = 0.0089W/m^2.K^2
Area	1x6m^2	3x2.1m^2
Inlet flow rate	360kg/hr	150L/hr
Pitch	30°	50°
Orientation	E/W	E/W
Ground reflectance	0.2	

ASHP	**Mitsubishi Ecodan**	**Sanyo Eco-Cute SHP-C45DEN**
Refrigerant	R410A	R744 (CO_2)
Max COP	4	3.75 (Ext Temp = 20°C)
Flow rate	360kg/hr	
Power input	1.75kW	1.2kW
Heat capacity	4.0kW	4.5kW
Combined Store		
Volume	250L	330L (200L + 130L)
Height	1.80m	1.98m
Diameter	0.42m	0.46m
Surface area	2.66m²	3.20m²
Storage loss coefficient	0.001W/m².K: calc. 0.0032kWh/24hr	calc. 0.57W/m².K: 2.2kWh/24hr
Cold inlet temp	20°C	Unknown, assumed 10°C

Table 3.8: Model and actual combined heating system parameters

Building envelope	**Model**	**Actual**
Floor area	48m²	48m²
Volume	116.5m³	112.5m³
Air tightness		2.43m³/m²
External wall		
Thickness	0.335m	
U-value	0.151W/m².K	0.12W/m²K
Solar absorptance	Front = 0.75, Back = 0.3	
Convective heat transfer	Front = 11kJ/h.m².K, Back = 64kJ/h.m².K	
Internal wall		
Thickness	0.024m	
U-value	2.296W/m²K	0W/m²K
Solar absorptance	Front = 0.6, Back = 0.6	
Convective heat transfer	Front = 11kJ/h.m².K, Back = 11kJ/h.m².K	
Ground		
Thickness	0.311m	
U-value	0.283W/m².K	
Solar absorptance	Front = 0.8, Back = 0.4	
Convective heat transfer	Front = 11kJ/h.m².K, Back = 999kJ/h.m².K	
Ceiling		
Thickness	0.462m	
U-value	0.087W/m²K	
Solar absorptance	Front = 0.6, Back = 0.6	
Convective heat transfer	Front = 11kJ/h.m².K, Back = 11kJ/h.m².K	
Roof		
Thickness	0.020m	
U-value	3.078W/m2K	0.09W/m2.K
Solar absorptance	Front = 0.35, Back = 0.75	
Convective heat transfer	Front = 11kJ/h.m2.K, Back = 64kJ/h.m2.K	
Floor		
Thickness	0.104m	

U-value	3.504W/m²K	0.15W/m²K
Solar absorptance	Front = 0.8, Back = 0.4	
Convective heat transfer	Front = 11kJ/h.m².K, Back = 11kJ/h.m².K	
Internal insulation		
Thickness	0.120m	
U-value	0.373W/m²K	
Solar absorptance	Front = 0.6, Back = 0.6	
Convective heat transfer	Front = 11kJ/h.m².K, Back = 11kJ/h.m².K	
Boundary wall		
Thickness	0.204m	
U-value	1.781W/m²K	
Solar absorptance	Front = 0.6, Back = 0.6	
Convective heat transfer	Front = 11kJ/h.m².K	

Table 3.9: Model and actual building envelope parameters

Fenestration	**Model**	**Actual**
Double glazed windows		Not installed
U-value	1.4W/m².K	
g-value	0.589	
Convective heat transfer	Front = 11kJ/h.m².K, Back = 64kJ/h.m².K	
INS3KR1 windows		Not installed
U-value	0.68W/m².K	
g-value	0.407	
Convective heat transfer	Front = 11kJ/h.m².K, Back = 64kJ/h.m².K	
Triple glazed windows	Not modelled	
U-value		0.80W/m².K
Triple glazed doors	Not modelled	
U-value		1.00W/m².K

Table 3.10: Model and actual fenestration parameters

In summary the modelled dwelling will have higher thermal losses than the actual dwelling, which is substantiated by the estimated requirement for space heating being twice that measured (1248kWh/year versus 606kWh measured in 2012). Higher thermal losses mean that control of the heating system is different from the real life system and that low storage losses result in a lower input from the ASHP. The difference between modelled and actual roof pitch may also lead to differences between the predicted and measured performance of the solar thermal system.

The differences in the parameters arise from the different ways the model calculates performance compared to how manufacturers describe performance which in turn leads to a lack of information about certain parameters required by the model. Other sources of discrepancy are assumptions made at the modelling stage due to a lack of communication between the modeller and design team, which is evident in the description of the combined store, building fabric U-values and ASHP; these discrepancies can also be caused by design changes occurring post modelling.

3.9.2 Operating conditions

Operating conditions represent another source of discrepancies that lead to a performance gap (Bordass et al. 2004; Carbon Trust 2011; Cayre et al. 2011). It is difficult for the model to accurately represent operating conditions due to the randomness often associated with them (de Wit 1995). Table 3.11 shows the primary differences between the operating conditions used in the model compared to the actual values.

Usage and control	Model	Actual
DHW		
Average daily volumetric consumption	60L/day	50L/day
Draw profile	2/3 07:00, 1/3 19:00	Random
Set point temperature	65°C	65°C
Electricity		
Tariff	Economy 7	Economy 10
Appliance usage	1386.6kWh	1610.77kWh
Weather		
Irradiation	1050.43kWh	940kWh
Average external temperature	10.78°C	10.60°C
Average humidity	77.28%	84.17%

Table 3.11: Model and actual usage behaviour

The most obvious difference between the model and actual system is the DHW usage behaviour of the occupants. On average the actual DHW consumption volume per day is less than that used in the model and the draw pattern of the measured system can be considered random compared to the consistent draw profile used in the model. Draw patterns and volumes are shown to have an effect on solar thermal system performance (Jordan & Vajen 2000; Knudsen 2002), and may have effects on the performance of the CO_2 heat pump which is suited to DHW provision compared to that modelled (Chen et al. 2009). The ratio of DHW energy to space heating is smaller in the modelled system; this difference in heating requirements between the modelled and measured systems can affect the performance of heat pumps with conventional ASHPs being more suited to low temperature applications such as space heating and CO_2 ASHPs performing better than conventional heat pumps in DHW heating applications (Stene 2007; Chen et al. 2009). Furthermore the higher modelled solar yield due to higher irradiation, higher external temperatures, more suitable pitch and higher, more consistent DHW draws will reduce the input required from the ASHP. In contrast the colder and more humid real life conditions will negatively affect the ASHP performance (Loose et al. 2011; Stark et al. 2014) and the higher storage losses combined with lower building thermal losses will affect the control of the heat pump.

The most obvious difference between the modelled and measured systems is the reduced input from the solar thermal system and high electricity draw combined with low COP from the ASHP – this constitutes an overall system underperformance. The next section compares the performance of the combined heating system with European systems (Loose et al. 2011; Miara et al. 2011) and aims to provide possible causes of the system's underperformance relating to the actual operating conditions that may not have been considered in the model.

3.10 Combined Heating System Underperformance

The installed combined heating system underperforms compared to the model and its European counterparts with a SPF of up to 5 demonstrated in (Loose et al. 2011). The COP of the heat pump component in this study was on average 3.7 which is comparable to other German (GSHP) and Swiss (100 brine/water heat pumps) heat pumps demonstrating average COPs of 3.4 and 3.7 respectively (Lokale Agenda-Gruppe 2006); (Erb et al. 2004) cited in (Loose et al. 2011). It is important to understand why the system is underperforming so that feedback can be given to future designers to improve their model assumptions and to perhaps improve in situ performance.

3.10.1 ASHP component

The ASHP installed in the case study dwelling underperforms compared to manufacturer's field trial data and the model in both electricity consumption and COP; the underperformance of ASHPs in the UK compared to other locations in Europe has also been highlighted elsewhere (Boait et al. 2011): The in-use COP over three years for 18 ASHP systems installed in Germany is on average 3.17 compared to results from a UK based field trial with average COP of 2.2 and this case study COP of 1.1 shows the difference in performance between UK and European systems.

There are fundamental reasons for this underperformance related to the actual operation of the ASHP in the case study dwelling. The first reason is related to the capacity of heat pump in comparison to the size of the house (Boait et al. 2011): In a Germany based field trial the average floor area was 199m^2 (with the smallest dwelling being 120m^2) compared to this case study dwelling's area of 48m^2 (Miara et al. 2011); however the capacity of the heat pump installed in the case study dwelling is 4.5kW, similar in size to those used in the German field trial, which ranged from 5-10kW. Furthermore, the case study dwelling is well insulated and air tight, which increases the thermal time constant of the dwelling (Boait et al. 2011). Boait et al. (2011) suggest that oversized heat pumps combined with small, well insulated dwellings means heat pumps will operate at lighter loads; this causes mechanical losses to be higher along with electricity consumption for the circulating pumps leading to lower COPs. The increased thermal time constant of well insulated, air tight homes means that appliance, solar and metabolic heat gains provide a variable, sometimes substantial, contribution to the overall heating requirement of the dwelling; with appropriate control this provides an opportunity to reduce electricity consumption and improve comfort levels by programming a setback period into the operation of the heat pump. Furthermore, the tendency of DHW loads overtaking the space heating requirement in small well insulated homes such as Bedford mean that CO_2 heat pumps may represent a better alternative to conventional heat pumps, but its heating capacity still needs to be smaller (Chen et al. 2009).

Heat pump performance is affected by the temperature difference between the source and sink; higher COPs are achieved when the difference between the source and sink is minimised. To achieve this, the space heating and DHW set-points should be as low as possible. Space heating set-points can be reduced with the use of underfloor heating which increases the surface area of the heat exchanger thus requiring a lower delivery temperature to provide the same amount of thermal energy as a traditional radiator. In the German field trial 72% of the all dwellings (GSHP, ASHP, WSHP) had underfloor heating which was shown to improve COP values (Miara et al. 2011). The heat source should be as high and consistent as possible; WSHP and GSHP tend to

have the more stable heat sources compared to ASHP, which have the least suitable heat source conditions for high COP values due to exposure to variable temperatures and humidity levels.

Miara et al. (2011) found that systems with direct heating rather than buffer storage had the highest efficiencies: Often the space heating demand was met by the buffer storage charged by the heat pump in DHW-mode; the higher inlet temperatures led to reduced COPs.

The set-point temperature for the heat pumps in the German field trial was 55°C compared to 65°C in the case study dwelling. Higher set-point temperatures, typically associated with DHW provision, reduce the COP of the heat pump, and it was noted that ASHPs in the German field trial would benefit with set-points of lower than 50°C (Miara et al. 2011). However the fact that the installed ASHP uses CO_2 as the refrigerant means that higher set-point temperature of 65°C should easily be met due to the high temperatures of the CO_2 exiting the compressor. Despite the improved performance of CO_2 ASHP over conventional heat pumps used for high temperature applications such as DHW provision, they are sensitive to the return temperature of the water entering the gas cooler from the storage tank. If the return temperatures are too high then the CO_2 exiting the gas cooler will not cool down sufficiently to reduce throttling losses during expansion and to reduce the enthalpy of the gas. The COP of a heat pump can be described by:

$$COP = \frac{h_2 - h_3}{h_2 - h_1} \tag{3.10}$$

Where h is the enthalpy of the CO_2 at three stages of the heat pump cycle:

1. At the compressor inlet, after the evaporator. Here is where the temperature and pressure of the CO_2 is lowest and therefore so is the enthalpy, ideally the temperature here will be as high as possible;
2. At the compressor outlet, before the gas cooler inlet. Here is where the temperature and pressure of the CO_2 is highest and so too is the enthalpy;
3. At the gas cooler exit. The pressure here is equal to the pressure at stage 2, but the temperature has dropped, ideally the temperature should be as low as possible to reduce the enthalpy and improve the COP;

The tank temperatures affect the return temperature of the water (Stene 2007) and so it is important to maintain stratification in the tank to minimise the temperature at the bottom of the tank and therefore the return temperature; this will minimise the temperature of the CO_2 at the gas cooler exit (stage 3). The external temperature (source temperature) affects the temperature of the CO_2 exiting the evaporator (stage 1) and so higher temperatures will reduce the enthalpy difference between the CO_2 here and at the compressor exit (stage 2); for this reason it is also beneficial to reduce the temperature of the CO_2 at the compressor exit by minimising the required set-point temperature (sink temperature) and reducing the amount of work needed from the compressor.

In the installed system the source temperature is weather dependent and cannot be changed by the user. Although in theory the sink temperature can be altered by the user, this remains at 65°C throughout the year. The return temperature for the Bedford system is quite likely to be very high judging from the high temperature at the bottom of the store and high levels of

insulation; Figure 3.20 shows a negative correlation between COP and temperature in the bottom of the tank.

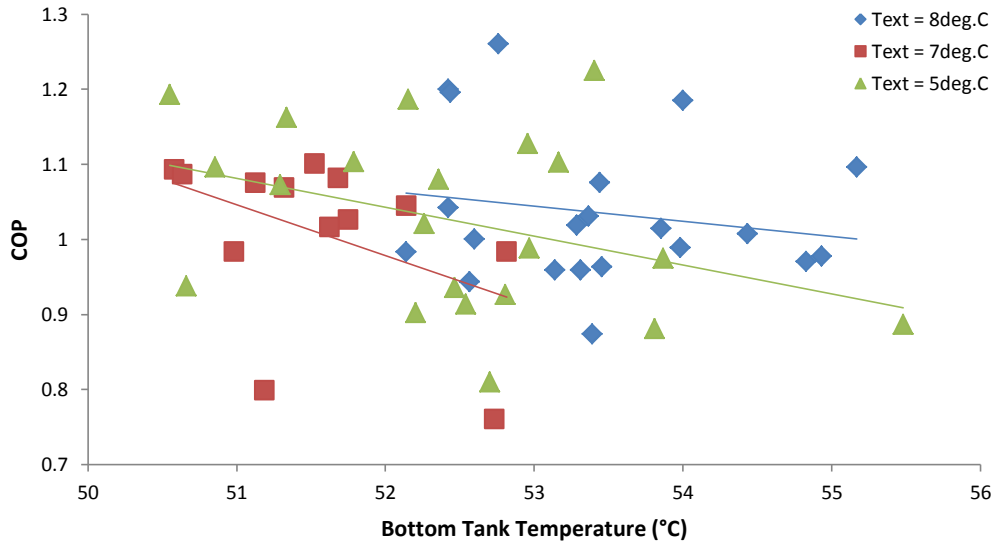

Figure 3.20: COP versus bottom tank temperature for three different external temperatures (5°C, 7°C and 8°C) and zero solar yield

From this relationship it is theorised that the solar thermal system may actually have a negative effect on the performance of the ASHP during periods when both are required.

Figure 3.21 shows that as the solar yield increases the required input from the ASHP decreases, this is expected behaviour of the system.

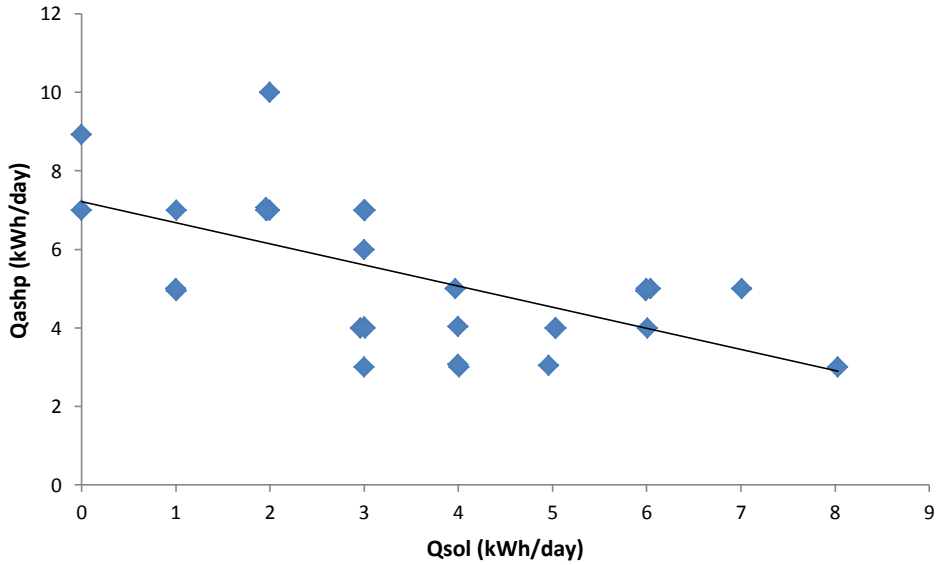

Figure 3.21: Q_{ASHP} versus Q_{sol} for days with external temperature 18°C; space heating demand is zero

However, Figure 3.22 shows that as the solar yield increases there is a tendency for the COP of the ASHP to decrease; this behaviour may be explained by the temperatures in the bottom of the store being higher during days with high solar yield (Figure 3.23). The high tank temperatures will increase the return temperature to the gas cooler and the outlet temperature

of the CO_2 from the gas cooler leading to higher enthalpies, throttling losses and ultimately lower COPs.

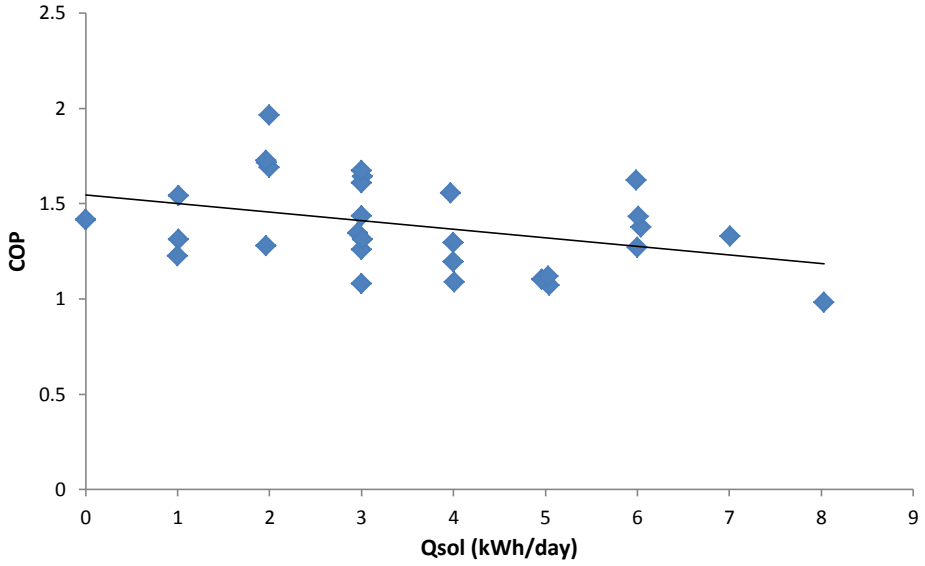

Figure 3.22: COP versus Q_{sol} for days with external temperature 18°C; space heating demand is zero

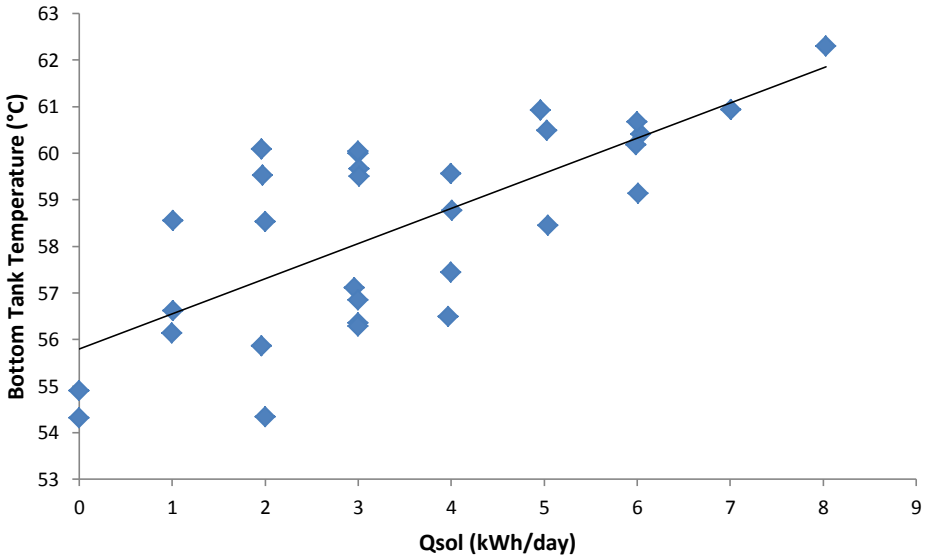

Figure 3.23: Temperature in the bottom of the tank versus Q_{sol} for days with external temperature 18°C; space heating demand is zero

In the colder, more humid times of the year ASHPs are susceptible to frost formation on the evaporator which reduces thermal efficiency in two ways: firstly a frost layer acts as an insulator and reduces heat transfer between the source and evaporator, this has the effect of increasing the work of the compressor in order to maintain a certain heating capacity additionally obstructed airflow due to ice formation will increase the work of the fan at the evaporator; secondly the frost formation requires additional energy to perform a defrost process, common defrost processes take the form of hot-gas bypass (HGBD) and reverse cycle defrosting (RCD) (Huang et al. 2009). The ASHP installed in the case study dwelling performs uses a HGBD process to divert warm CO_2 from the compressor outlet to the evaporator inlet and melt frost. The

consequence of this process is that work from the compressor is used to heat the CO_2 for the defrost process and not to heat water in the gas cooler. Since the heat output from the ASHP is measured after the gas cooler this results in low measured heat values for a high electrical input hence the low COP.

In the Sanyo Ecocute range a defrost process is initiated when the temperature at the evaporator air inlet approaches the temperature of the air outlet – this indicates reduced heat transfer and frost formation (Chen et al. 2009). However defrost process control strategies that do not measure the thickness of the frost formation can lead to a phenomenon known as mal-defrost (Wang et al. 2011). Mal-defrost manifests itself in two ways: the first is for defrost operations to occur long after a critical level of frost formations has occurred, this is common in temperate climates with moderately cold temperatures and high humidity; the second way occurs in cold dry climates and takes the form of unnecessary defrost operations occurring (Wang et al. 2011). Since the climate of Bedford lends itself to the first form of mal-defrost it is possible that there were periods during the winter when defrost operations occurred too late resulting in thick frost forming. This would lead to reduced heat transfer in the evaporator and increased work from the compressor to melt the ice. In addition the diversion of the hot CO_2 at the compressor outlet will lead to periods of little to no water heating and therefore no heat output. Tests performed by Wang et al. (2011) suggested that defrost operations performed in temperate conditions can be over 1hour overdue and lead to almost a 20% decrease in the COP at this temperature.

The number of on/off cycles during the year of the ASHP total 3885, unfortunately it is not possible to distinguish the number of these cycles that are used for defrost operations with the data that has been collected; for this the temperature at the evaporator would need to be measured. The number of cycles for a modelled ASHP based on an installed system in Scotland performed 3311 cycles in comparison and obtained an annual COP of 2.77 (Kelly & Cockroft 2011). Excessive on/off cycling can lead to increased electricity consumption due to the high current drawn by the compressor (Stafford & Lilley 2012). Cycling periods of less than 6min is known as short cycling in (Green 2012) and it is shown that this negatively affects the COP of the ASHP. The number of short cycles (<5min) in the Bedford installation totals 2005 cycles indicating that the ASHP is short cycling most of the time. The majority of the short cycles take place in the summer months, as seen in Figure 3.24.

Figure 3.24 shows that in the winter months the majority of the cycles are over 5min in length, this agrees with that found in (Green 2012) where cycle periods increased in length with decreasing external temperature. There are a fewer number of cycles over all during the spring and summer months as space heating demand decreases and solar yield increases reducing the requirement from the ASHP; however the number of short cycles increases in these months, which may limit the performance of the heat pump by increasing the power consumption. Short cycling can be a result of thermal losses from the building as a result of poor insulation or window opening by the occupant (Stafford & Lilley 2012). Since the Bedford property is well insulated indicated by low space heating demand, the warm temperatures in the house during the summer months may be encouraging the occupant to open windows in an effort to cool the dwelling down.

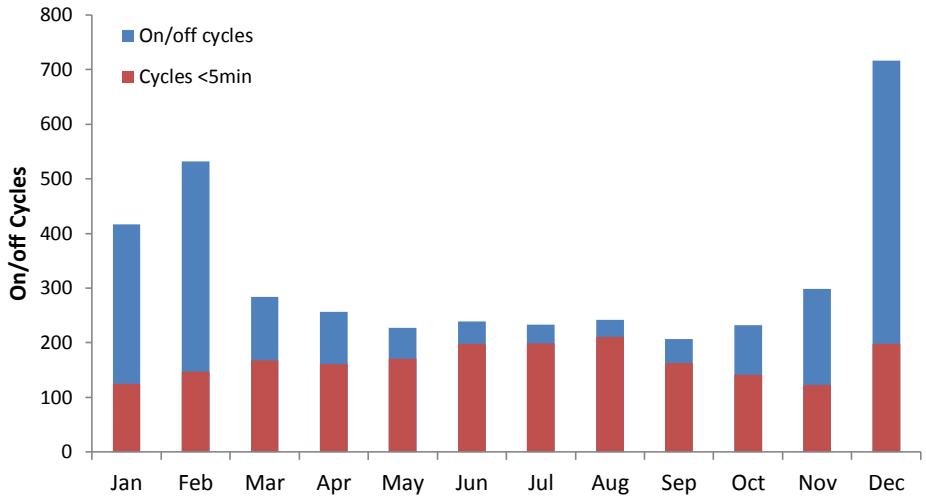

Figure 3.24: On/off cycles and short cycles throughout 2012

During a summer day the internal temperature does not appear to fall below 24°C; this is most likely due to appliance and solar gains as the space heating delivered by the ASHP is zero for this day. However, the ASHP appears to cycle throughout the day which is due to the tank temperature at the bottom of the tank (where the thermostat is) not reaching 65°C (when 65°C is reached here the heat pump switches off). When the solar thermal portion of the tank reaches more than 40°C the bypass valve opens to allow the solar heated water into the DHW buffer. Since the temperature in the solar buffer is approaching 70°C at its hottest the bypass valve must be open; however the ASHP is still cycling due to the low temperature at the thermostat level (<65°C) (Figure 3.25). As the water returning to the ASHP comes from the solar portion of the tank, high return temperatures are expected which will reduce the COP as explained previously.

On a winter's day (Figure 3.26) there is no solar thermal input and the demand for heating is both DHW and space heating. The ASHP switches on to heat the water in the tank, when the temperature near the thermostat reaches 65°C the heat pump stops producing DHW; however input is still required to meet space heating demand (3kWh for this day) indicated by the flow temperature in the radiators. In addition defrost cycles may be evident where the compressor is active but flow temperatures and temperatures in the tank do not alter or continue to fall.

Both days have a similar number of on/off cycles but cycles are shorter on the summer day due to higher external temperature limiting the ASHP requirement and temperatures in the tank being below the switch off limit causing the heat pump to operate. Although energy consumption per cycle is lower on this day due to the higher external temperature, the COP is reduced to 1.36 because of the higher return temperatures from the solar buffer and high start-up currents for the compressor versus the amount of heat actually delivered. Longer cycle periods are seen on the winter's day due to the need for space heating and DHW. Although longer cycle periods are better for performance, COPs remain low at 0.99 because of low external temperatures combined with high humidity (more potential for defrost operations to occur) and higher flow temperatures required for DHW and space heating.

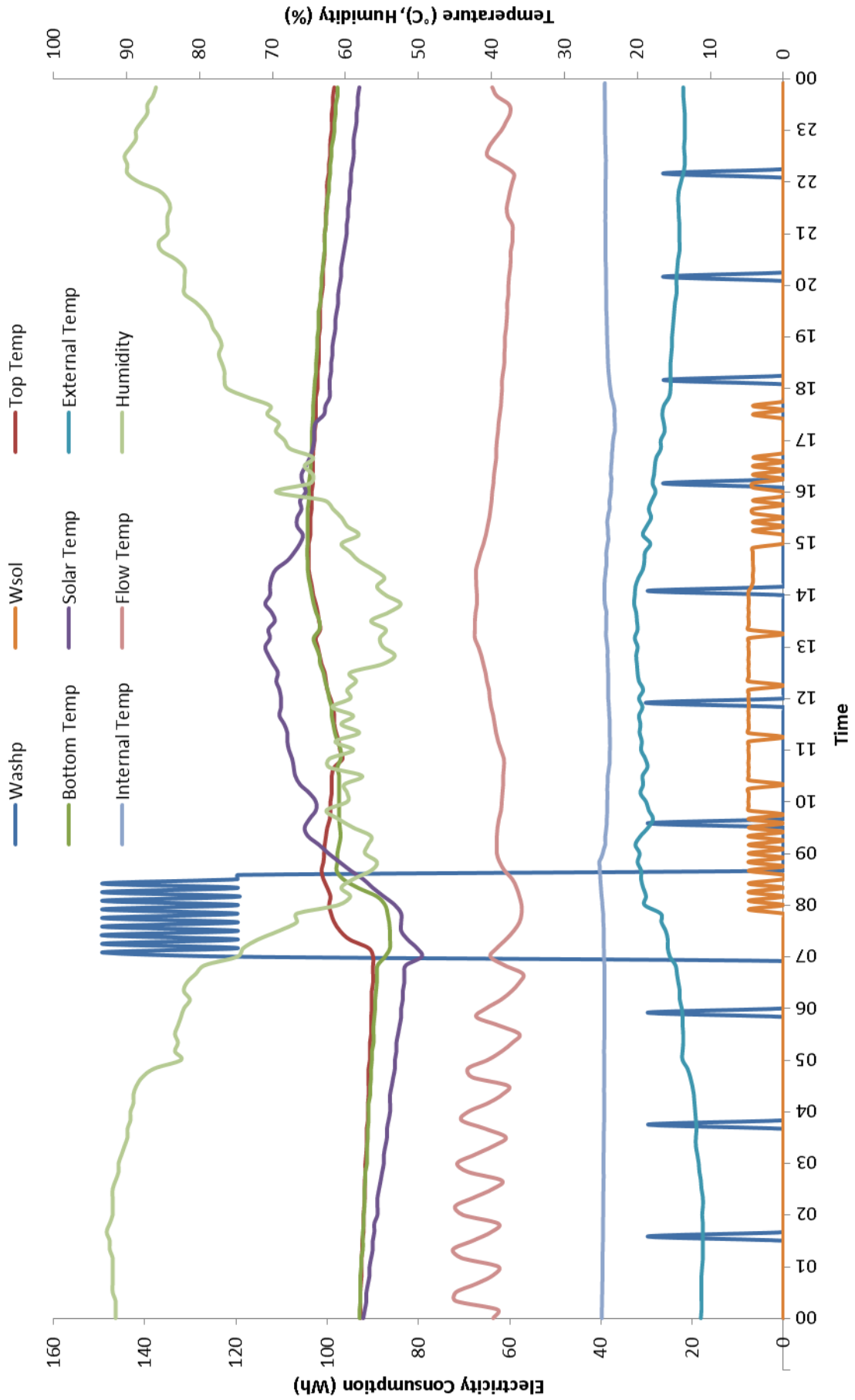

Figure 3.25: Heating and temperature profile for 16th June 2012

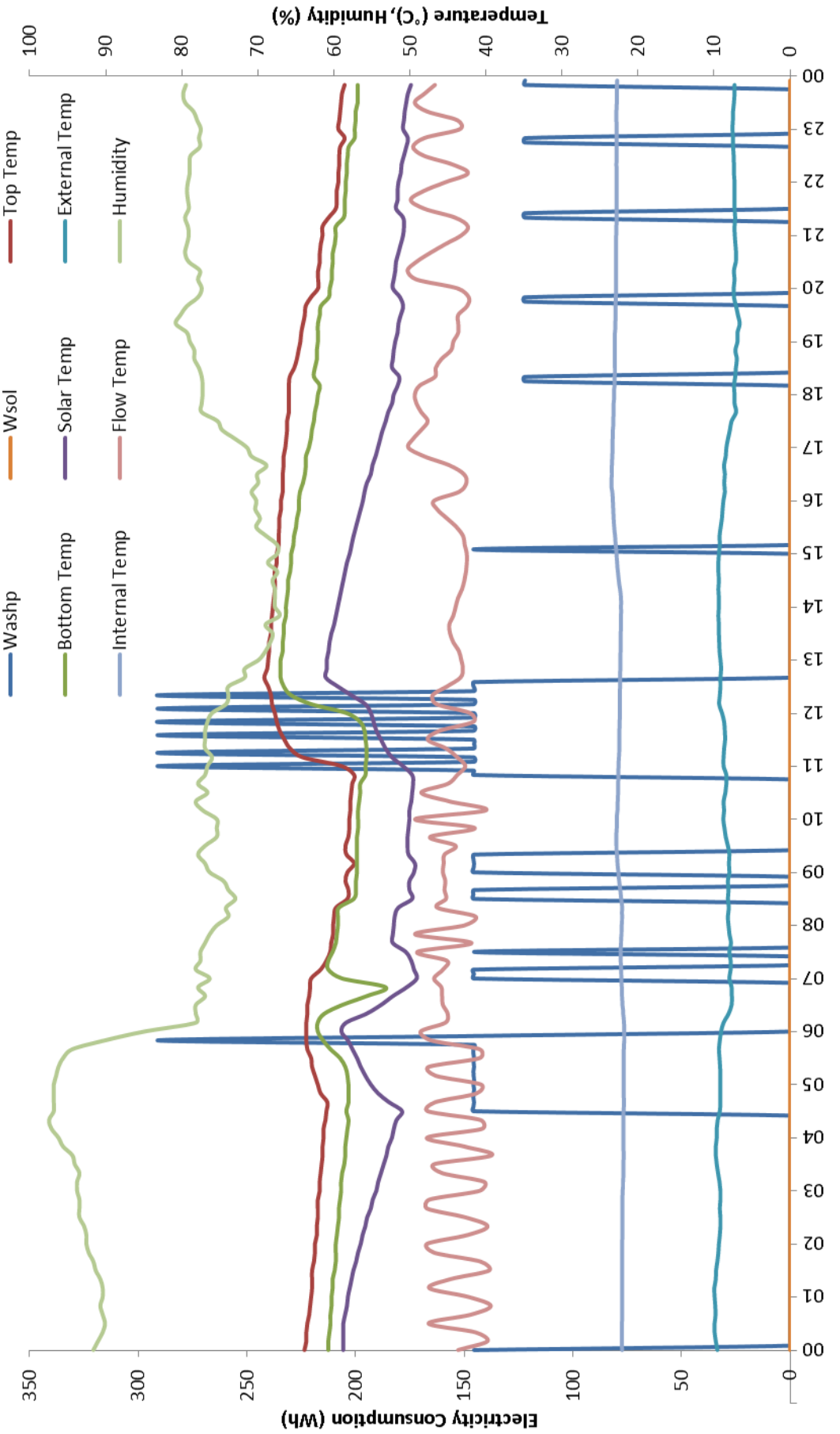

Figure 3.26: Heating and temperature profile for 7th January 2012

A further factor to be mentioned is that the solar fraction on the summer day is 60% (Q_{sol} = 5.99kWh, Q_{ASHP} = 4kWh). With a space heating demand of zero and DHW volumetric consumption of approximately 1.69kWh it is clear that the solar thermal system was capable of supplying 100% of the DHW requirement for this day. It appears that controlling the ASHP based on the bottom temperature of the DHW buffer is leading to unnecessary heat input from the heat pump on days of high solar yield; this will reduce both the solar fraction and the SPF. In contrast on the winter day the space heating demand was 3.03kWh and the DHW consumption only 11L (0.68kWh); the ASHP heat output was 10kWh resulting in over 6kWh of unused heat.

3.10.2 Solar thermal component

The solar thermal system provides the most heat between May and September; during these months the space heating requirement is low and the solar thermal system should act as the primary heating system with the ASHP providing back-up heating as required for those days with low irradiance, space heating demands and to boost the temperature of the DHW. During the remaining half of the year the irradiance and external temperatures are too low to expect any serious contribution from the solar thermal system and so any heat that it does provide is used to take the load off the ASHP.

Compared to the TRNSYS model outputs presented in section 3.8.6, the solar thermal system is underperforming; reasons for this discrepancy relating to the model assumptions have already been discussed; however reasons pertaining to the actual operating conditions will be presented here.

First and foremost the daily solar yield, solar fraction and system efficiency increase with irradiation and to an extent external temperature, which generally increases with irradiation as shown in Figure 3.27.

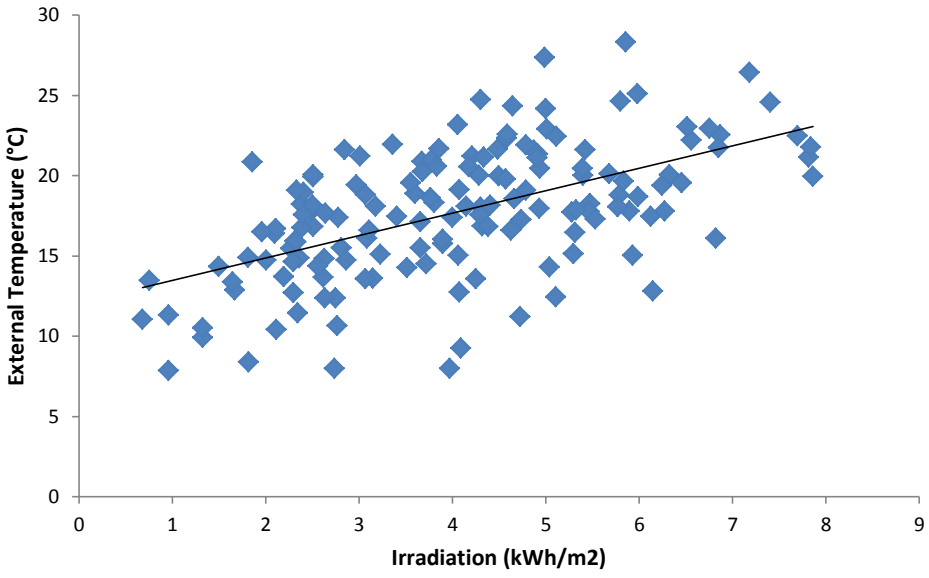

Figure 3.27: Average external day time temperature versus daily irradiation May – September

Figure 3.28-Figure 3.30 show the positive correlation between the daily values for Q_{sol}, solar fraction and system efficiency respectively for the period between May and September when the solar thermal system performs at its best. Daily irradiation (kWh/m^2) is the integral of

irradiance (W/m^2) with respect to time; therefore summer provides the highest irradiation days in general due to brighter and longer days meaning that the system can operate for longer periods hence higher solar yields overall at this time of year. The solar fraction increases with irradiation due to the higher solar yield but also because of the reduced ASHP requirement. The ASHP requirement is reduced because of the reduced need for space heating thanks to solar gains and warmer external temperatures as well as high solar yields from the solar thermal system.

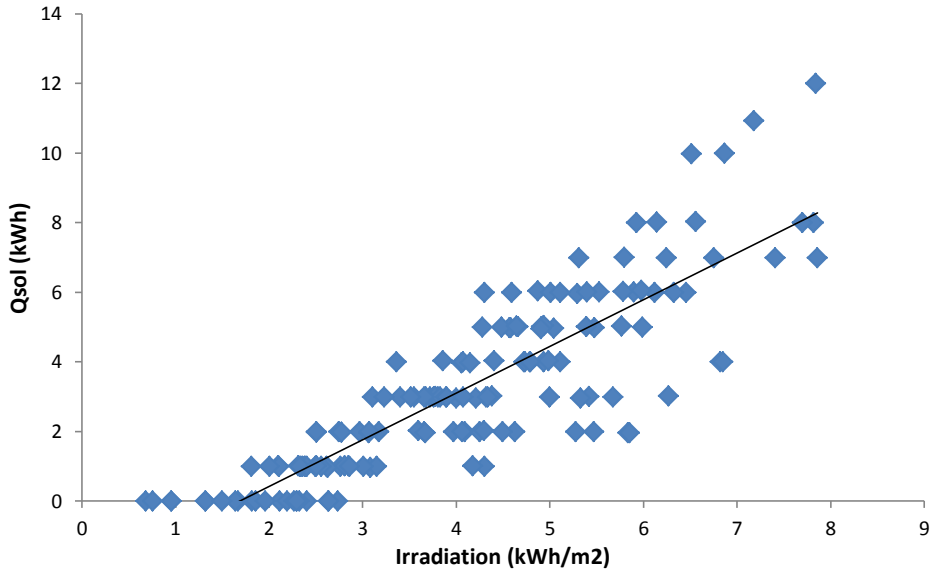

Figure 3.28: Daily Q_{sol} versus daily irradiation May – September

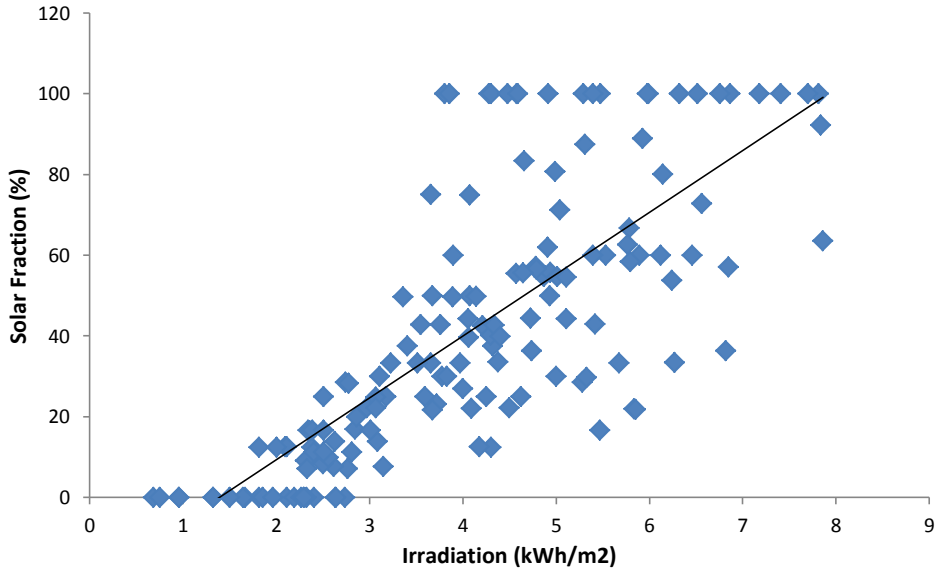

Figure 3.29: Daily solar fraction versus daily irradiation May - September

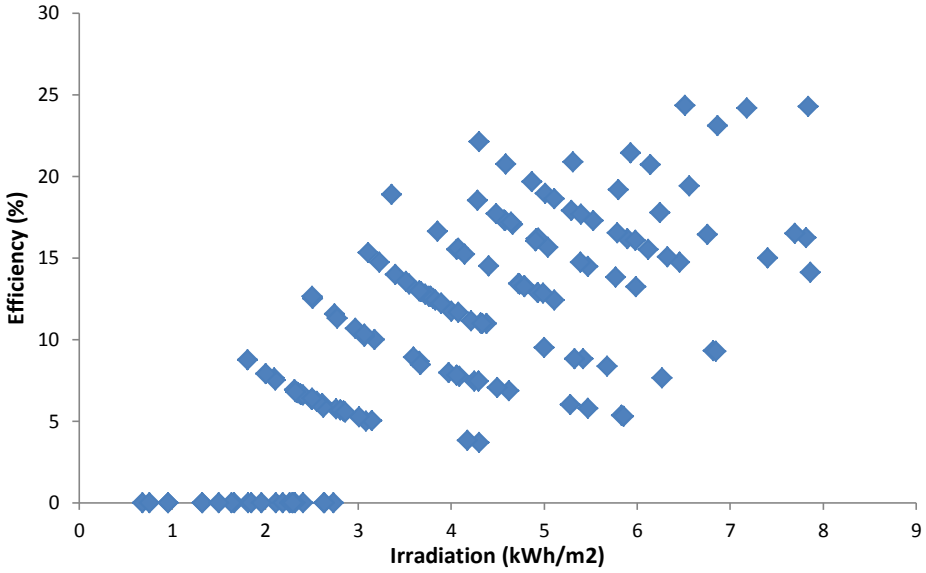

Figure 3.30: Daily system efficiency versus daily irradiation May - September

Figure 3.30 shows evidence of a plateau beginning to form at high irradiation levels; once the level of irradiation reaches a critical point the yield produced by the system is limited no longer by the available energy but by the temperature in the solar buffer. As the temperature of the water in the solar buffer increases the rate of heat transfer decreases because the temperature difference between the working fluid in the solar coil and the storage water decreases:

$$\dot{Q_{sol}} = \dot{m}C_p(T_{coil} - T_{sb}) \tag{3.11}$$

Where $\dot{Q}_{sol}$ is the rate of heat transfer from the solar coil; $\dot{m}$ is the flow rate of the working fluid through the solar coil; C_p is the specific heat capacity of the working fluid; T_{coil} is the temperature of the working fluid in the coil; and T_{sb} is the temperature of the water in the solar buffer. If the temperature difference between the coil and the solar buffer reaches a predetermined value (typically 5°C) then the solar pump is deactivated and heat transfer ceases ($\dot{m}$ = 0kg/s).

The temperature in the solar buffer is affected by the amount of DHW drawn throughout the day, with large volumes reducing the temperature in the area surrounding the solar coil and thus promoting heat transfer: As hot water is drawn out of the top of the DHW buffer, cold water is drawn into the bottom of the solar buffer; the graduation of temperature with hot water at the top of the tank and cooler water at the bottom is known as stratification. The degree of stratification is important for solar thermal system performance with higher levels of stratification leading to greater solar yields (Lavan & Thompson 1976; Phillips & Dave 1982; Hollands & Lightstone 1989; Furbo, Andersen, Thür, et al. 2005; Jordan & Furbo 2005). Good stratification means that cool water is contained in the solar buffer which means that the return temperature of the working fluid back into the collector is lower; this reduces collector losses because the temperature difference between the absorber and the ambient air is minimised (Martin & Watson 2002; Lavan & Thompson 1976). Lavan & Thompson (1976) suggest that stratified tanks also allow the operating time of the solar thermal system to be extended because the cooler tank temperature increases the temperature difference between the collector fluid and the storage water and means that the collector can deliver heat during the low irradiance conditions during the morning and late afternoon.

Stratification can be established during charging of the tank with auxiliary heat input into higher regions of the tank, which increases the tank temperature higher up the store; and during discharging, where DHW is drawn and cold water enters in at the bottom of the tank thus reducing the temperature here (Furbo, Andersen, Thür, et al. 2005). A fully mixed tank with a uniform temperature throughout and therefore zero stratification can be caused by mixing of cold and hot water stratum due to turbulence at cold water inlets; high water draws lowering the temperature of the entire tank more quickly than it can be heated leading to an overall cool tank; or by the heat being delivered into the tank more quickly than it is being removed – due to high heat inputs and/or low DHW consumption – leading to an overall warm tank. This suggests that the performance of the solar thermal system is not only governed by the weather conditions, but also by the user behaviour in terms of excess auxiliary heat and insufficient DHW consumption.

Figure 3.31 shows the system efficiency of the Bedford solar thermal system for days with no heat input from the ASHP, daily irradiation levels between 6-8kWh/m^2/day and average external day time temperatures greater than 19°C. It is evident that once the effects of temperature and auxiliary input have been removed, the daily efficiency of the system increases with increased DHW consumption.

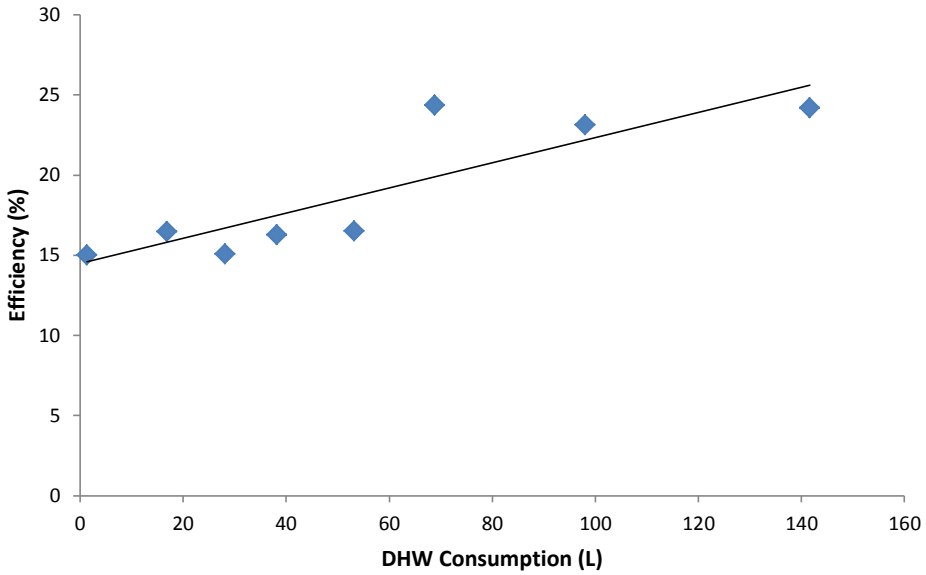

Figure 3.31: Daily system efficiency versus daily DHW consumption for Q_{ASHP} = 0, I = 6-8kWh/m^2/day and T_{ext} > 19°C

Figure 3.32 shows the daily system efficiency decreasing with increasing heat input from the ASHP.

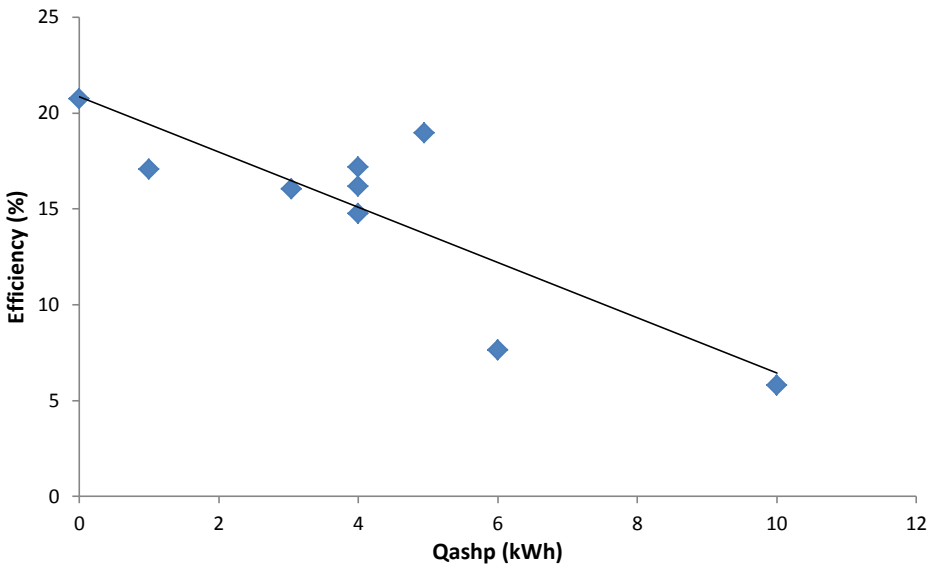

Figure 3.32: Daily system efficiency versus Q_{ASHP} for DHW consumption = 40-60L/day, I = 4-7kWh/m^2/day and T_{ext} > 17°C

It is important to understand that correlation does not necessarily indicate causality, especially between solar yield and ASHP input. During the winter the solar thermal system acts as the supporting heating system with most of the heat being delivered by the ASHP; therefore any heat delivered by the solar thermal system will reduce the amount required by the ASHP and a negative correlation will be seen between solar thermal system efficiency and ASHP input – this is not because the heat pump is negatively affecting the solar thermal system performance but rather because days of low solar yield in the winter require higher ASHP contributions. During the summer with higher external temperatures, solar gains and a reduced space heating demand the requirement of the ASHP depends on the amount of DHW drawn and whether the solar thermal system can provide for this. Days with high DHW demands that cannot be solely met by the solar thermal system will have higher ASHP inputs and again a negative correlation between system efficiency and ASHP input will be observed. However, when the DHW demand is kept constant and the weather conditions are suitable for the solar thermal system to be the primary heating system, as in Figure 3.32, then high ASHP heat input will negatively affect the solar thermal system performance by increasing the temperature of the tank and reducing the active period of the solar thermal system and therefore the system efficiency Figure 3.33:

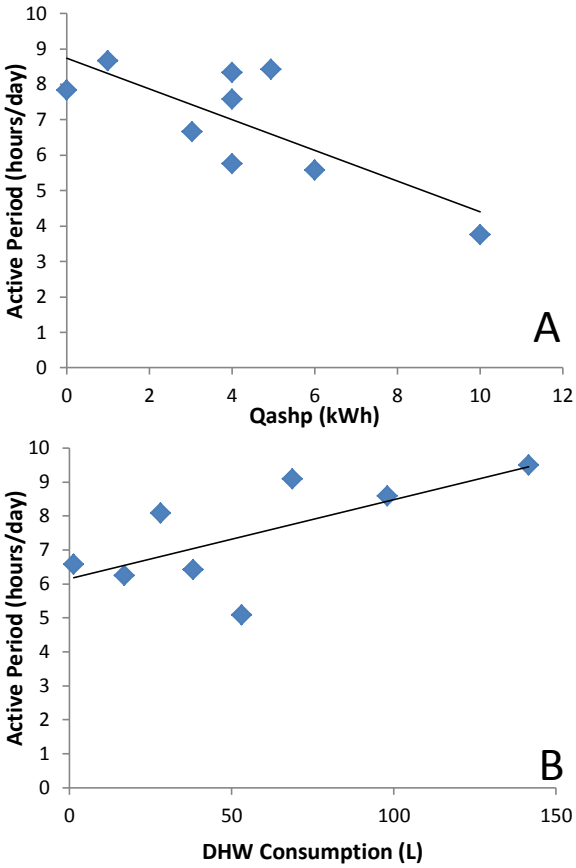

Figure 3.33: Total duration of solar pump activity (hours/day) versus a) Q_{ASHP}; b) DHW consumption

Zero flow conditions are a result of the solar controller deactivating the solar pump this is generally due to low irradiation, power failures in the pump and controller, the thermal capacity of the solar buffer being reached (high ASHP input and/or solar yield) and low DHW consumption (Harrison et al. 2004; Harrison & Cruickshank 2012). During zero flow periods when the irradiance is high collector losses increase until they equal the amount of energy incident on the collector, this is known as stagnation. During stagnation periods temperatures in the collector can become extremely high which can cause damage to the collector and circuit components as well as degrade the materials and working fluid resulting in a loss of performance over time (Harrison & Cruickshank 2012; Zhang et al. 2012). Therefore prolonged periods of high ASHP input and low DHW consumption during the key months for solar thermal activity not only results in reduced solar yields by destroying stratification in the tank, but also degrades the maximum achievable performance of the solar thermal system in the long term.

3.11 Discussion and Conclusion

The Bedford ASHP represents the single largest contributor to the overall whole-house performance gap with a five-fold increase in the actual electricity consumption versus the prediction. Possible reasons for were theorised as follows:

- Complex interactions between the solar thermal system and the ASHP;
- The heat pump's overcapacity in relation to the space heating demands of the dwelling;

- Possible control issues leading to excessive cycling especially during the summer months;
- High return temperatures to the gas cooler which increase the enthalpy of the CO_2.

Although these issues could be considered unique to the Bedford property, Figure 3.34 shows that in general the properties in the Retrofit for the Future competition with the highest carbon performance gap are those with ASHPs installed compared to those properties with only solar technologies and conventional heating systems:

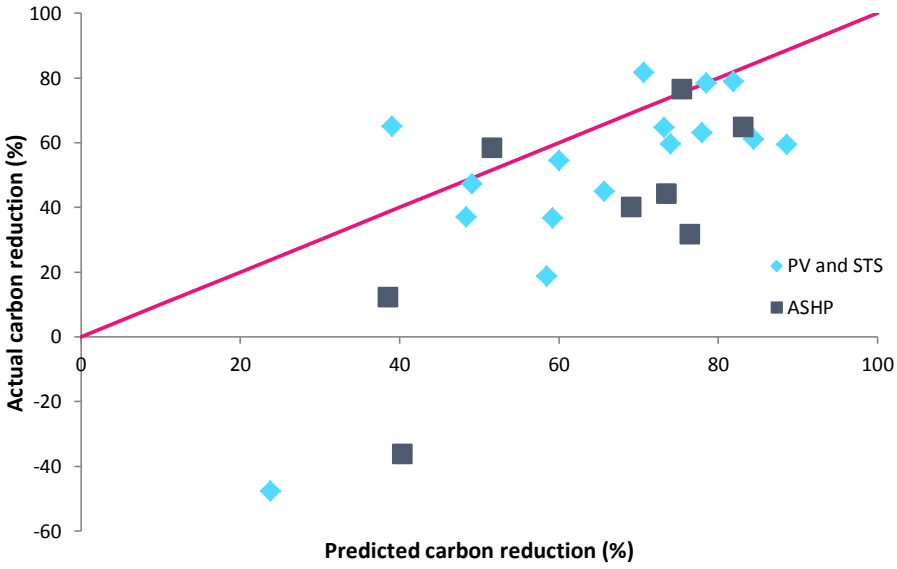

Figure 3.34: Measured carbon reduction versus predicted for properties with solar technology versus those with ASHPs (the ideal case plotted in red represents the scenario where predicted reductions are met in reality)

This suggests that for low and zero carbon homes at least, there is a real need for in-situ ASHP performance to be addressed before any real benefit can be gained from their use.

The conclusions to the chapter are:

- The performance gap exists in the majority of the Retrofit for the Future projects; those with the highest performance gaps were those properties with installed ASHPs. The performance gap represents an uncertainty about the true performance of the building and its subsystems. This uncertainty makes it unknown whether carbon and energy targets are being met in reality as well as making design and investment decisions difficult;
- The PV system at Bedford was the second smallest contributor to the whole-building performance gap after mains electricity consumption. The gap associated with PV was a positive gap meaning that the system generated more electricity than predicted. This was primarily due to actual weather conditions being better than those used in the model – the system efficiency prediction was accurate;
- The largest contributor to the performance gap at the Bedford case study property was the combined heating system with the ASHP component consuming four times more electricity than estimated; system dysfunction (including control issues and interaction with the solar thermal system) was a potential cause for this. Differences in actual and

modelled usage behaviour also contributed to the subsystem performance gap. The solar thermal system generated 39% less heat than predicted again due to interactions with the heat pump component and differences in modelled and actual operating conditions.

- Other causes of whole-building performance differences were due to discrepancies between modelled and actual building fabrics;
- During periods of joint solar thermal and ASHP heat input the performance of the combined heating system is affected by interactions between the solar thermal and ASHP components;
- Increased solar yields result in decreased ASHP requirement as expected however COP values decrease with increasing solar yield. This is due to increased return temperatures to the gas cooler causing higher enthalpy CO_2 at the gas cooler exit and throttling losses due to high temperature CO_2;
- Periods of high solar yield are accompanied by warm external temperatures which increase the COP values compared to the winter months; however the COP is limited further by the presence of short cycling (<5min) which although compressor input energy is lower per cycle at this time, the heat output is proportionately less since compressor start-up currents are higher;
- Solar yield increases with irradiation and external temperature; however system efficiency begins to plateau at irradiation levels above 7kWh/m^2/day; these limitations are caused by consumption behaviour – if DHW consumption does not increase with increasing irradiation then the rate of change of solar yield with respect to irradiation is zero;
- For constant weather conditions and ASHP input the system efficiency increases with DHW consumption; this is due to improved stratification in the tank which decreases return temperatures and collector losses as well as reduces the occurrence of zero flow conditions;
- For constant weather conditions and DHW consumption the system efficiency decreases with increasing ASHP input; this is due to reduced stratification and higher temperatures in the tank leading thermal capacity being reached and zero flow conditions to be imposed by the controller;
- Stagnation is a potential threat to the system's long term performance due to high temperatures resulting from high irradiance days coupled with low DHW consumption and excessive ASHP input;
- Low space heating requirements mean that the ASHP is potentially oversized resulting in uncomfortably high indoor temperatures and low COPs. Furthermore the solar thermal system contributes very little to the space heating demands which occur only in the winter and therefore is predominantly used to address DHW needs; these needs are low in comparison to the size of the collector and tank which results in lower system efficiencies due to reduced operating periods in the summer. This may also lead to a higher possibility of stagnation;
- Excessive ASHP input during the summer months reduces the solar fraction. Temperatures in the solar buffer are more than adequate to meet the DHW needs of the occupants; however the temperature in the bottom of the DHW buffer is low enough to maintain ASHP input for many of the days between May and September. Since DHW

requirements are generally low (average daily consumption of 50L/day) then the solar buffer storage volume (130L) is adequate, and may even be beneficial according to (Furbo, Andersen, Knudsen, et al. 2005) who show that the net utilised solar energy increases when the tank volume matches the DHW consumption. During summer days with zero space heating demands it would be better to draw DHW from just the solar buffer, which would improve stratification in this part of the combined store and therefore prolong the operating period of the solar thermal system. During the winter, with no solar yield, average temperature in the solar buffer is greater than 50°C, which is reducing the COP of the heat pump.

- There is a positive correlation between space heating demand and COP due to the effect that higher space heating demands reduce the temperature in the solar buffer and consequently reduce the return temperature to the gas cooler.
- A more appropriate strategy may be to deactivate the ASHP entirely when outdoor temperatures exceed 20°C, the high thermal time constant of the dwelling means that solar, appliance and metabolic gains may be enough to maintain a comfortable indoor temperature during the summer. Without interference from the ASHP the solar thermal system can deliver more heat for DHW purposes which can be topped up by the immersion heater powered by the PV system.
- To improve performance for both the ASHP and solar thermal components stratification in the tank has to be improved. The tank in tank system negatively affects the ASHP by maintaining high temperatures in the solar buffer leading to high return temperatures to the gas cooler and thus lower COPs – this is caused by no cold water entering in the solar buffer; the arrangement negatively affects the solar thermal system because temperatures in the solar buffer become very high, but are not drawn directly from the solar buffer which increases the level of mixing, the collector losses and reduces the operating time of the solar system. In addition the cold water does not enter into the solar buffer and so stratification is destroyed further. The fact that the temperature in the bottom of the DHW buffer are lower than those in the solar buffer means that as hot water from the solar buffer flows into the DHW buffer the cooler water from the buffer is displaced into the solar buffer; however this displaced water is often more than 60°C which maintains high temperatures in the solar buffer and possible de-stratification. In fact the tank in tank configuration means that the solar thermal system is heating cold water from below which is contrary to the usual method of cold water entering below the solar coil, which is beneficial for stratification and therefore solar thermal performance (Furbo, Andersen, Thür, et al. 2005).
- The poor performance of the ASHP during the summer may be due to high return temperatures to the gas cooler from the tank as a result of solar input; the poor winter performance may be due to mal-defrost operations as well as low source temperatures.

From the discussion above it is evident that interactions between building envelope/fabric, solar thermal systems, heat pumps and occupants result in uncertainty relating to the actual energy and carbon performance of a dwelling. These interactions are part of the actual operating conditions experienced by the installed systems and are more often than not impossible for current modelling strategies to accurately portray. As a result it is difficult to say with conviction which retrofit strategies will return the biggest benefit in terms of household bill reduction,

carbon reduction and reduced energy consumption. Furthermore with the introduction of the domestic RHI the uncertainty in performance of renewable heating systems such as heat pumps and solar thermal systems means that there is a considerable financial risk to both the householder and policy maker.

An evidence base combined with a holistic representation of the system in the field accounting for user behaviour, building envelope/fabric and interactions with other systems would provide the necessary elements for addressing the issues of performance uncertainty.

The following chapters will demonstrate the potential benefit of a holistic, evidence based approach to predicting the in-situ performance of solar thermal systems following a critique of current modelling approaches in Chapter 4.

3.12 References

Allen, S.R. et al., 2010. Integrated appraisal of a Solar Hot Water system. *Energy*, 35(3), pp.1351–1362.

Bannister, P., 2009. Why Good Buildings Go Bad, While Some Are Just Born That Way. *Ecolibrium*, pp.24–32.

Boait, P.J., Fan, D. & Stafford, a., 2011. Performance and control of domestic ground-source heat pumps in retrofit installations. *Energy and Buildings*, 43(8), pp.1968–1976.

Bordass, B., Cohen, R. & Field, J., 2004. Energy Performance of Non-Domestic Buildings: Closing the Credibility Gap. In *Building Performance Congress*.

Branco, G. et al., 2004. Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data. *Energy and Buildings*, 36(6), pp.543–555.

BRE, 2009. National Calculation Method. Available at: http://www.ncm.bre.co.uk/ [Accessed January 20, 2014].

BRE, 2010. SAP 2009 The Government's Standard Assessment Procedure for Energy Rating of Dwellings.

Carbon Trust, 2011. Closing the Gap: Lessons learned on realising the potential of low carbon building design.

Cayre, E. et al., 2011. There are people in the house! how the results of purely technical analysis of residential energy consumption are misleading for energy policies. In *European Council for an Energy Efficient Economy (ECEEE) Summer School*. pp. 1675–1683.

Chen, Y. et al., 2009. CO2 heat pumps for the Swedish market: Test and analysis of the SANYO ECO-CUTE heat pump modified for Swedish conditions.

Demanuele, C., Tweddell, T. & Davies, M., 2010. Bridging the gap between predicted and actual energy performance in schools. In *World Renewable Energy Congress*. pp. 1–6.

de Wit, M.S., 1995. Uncertainty analysis in building thermal modelling. In *Proceedings of international building performance simulation association*.

Doylend, N., 2014. *Evaluating Building Energy Performance: A Lifecycle Risk Management Methodology*. Loughborough University.

Dunbabin, P., Charlick, H. & Green, R., 2013. *Detailed analysis from the second phase of the Energy Saving Trust's heat pump field trial*, London.

Dunbabin, P. & Wickins, C., 2012. *Detailed analysis from the first phase of the Energy Saving Trust's heat pump field trial*, London.

Erb, M., Hubacker, P. & Ehrbar, M., 2004. Feldanalyse von Wärmepumpenanlagen FAWA 1996-2003.

EST, 2010. *Getting warmer: A field trial of heat pumps*, London.

EST, 2011. *Here comes the sun: a field trial of solar water heating systems*, London.

EST, 2001. Solar hot water systems in new housing. *GIR88*.

Fokaides, P. a. et al., 2011. Comparison between measured and calculated energy performance for dwellings in a summer dominant environment. *Energy and Buildings*, 43(11), pp.3099–3105.

Forward, D. & Roberts, C., 2008. *Viridian Solar - Clearline Solar Thermal Field Trial*, Watford.

Furbo, S., Andersen, E., Thür, A., et al., 2005. Performance improvement by discharge from different levels in solar storage tanks. *Solar Energy*, 79(5), pp.431–439.

Furbo, S., Andersen, E., Knudsen, S., et al., 2005. Smart solar tanks for small solar domestic hot water systems. *Solar Energy*, 78(2), pp.269–279.

Green, R., 2012. *The Effects of Cycling on Heat Pump Performance*, Capenhurst.

Gupta, R. & Dantsiou, D., 2013. Understanding the Gap between "as Designed" and "as Built" Performance of a New Low Carbon Housing Development in UK. *Sustainability in Energy and Buildings*, 22, pp.567–580.

Haas, R. & Biermayr, P., 2000. The rebound effect for space heating Empirical evidence from Austria. *Energy Policy*, 28, pp.403–410.

Hadorn, J., 2012. IEA Solar and Heat Pump Systems. Solar Heating and Cooling Task 44 & Heat Pump Programme Annex 38.

Harrison, S. & Cruickshank, C. a., 2012. A review of strategies for the control of high temperature stagnation in solar collectors and systems. *Energy Procedia*, 30, pp.793–804.

Harrison, S.J., Lin, Q. & Mesquita, L.C.S., 2004. Integral stagnation temperature control for solar collectors. In *SESCI 2004 Conference*. pp. 1–9.

Hens, H., 2010. Energy efficient retrofit of an end of the row house: Confronting predictions with long-term measurements. *Energy and Buildings*, 42(10), pp.1939–1947.

Hensen, J., 2002. Hensen, J.L.M. (2002). Simulation for performance based building and systems design: some issues and solution directions. Proceedings of the 6th International Conference on Design and Decisions Support Systems in Architecture and Urban Planning, 7-10 July. , pp.7–10.

Hinge, A., Taneja, O. & Bobker, M., 2008. Sustainability in Commercial Buildings - Bridging the Gap from Design to Operations. *Proceedings of the 5th International Conference on Improving Energy Efficiency in Commercial Buildings: IEECB Focus 2008*, 1.

Hollands, K.G.T. & Lightstone, M.F., 1989. A review of low flow, stratified tank solar water heating systems. *Solar Energy*, 43(2), pp.97–105.

Huang, D., Li, Q. & Yuan, X., 2009. Comparison between hot-gas bypass defrosting and reverse-cycle defrosting methods on an air-to-water heat pump. *Applied Energy*, 86(9), pp.1697–1703.

INCOSE, 2011. *Systems Engineering Handbook*, San Diego.

Jordan, U. & Furbo, S., 2005. Thermal stratification in small solar domestic storage tanks caused by draw-offs. *Solar Energy*, 78(2), pp.291–300.

Jordan, U. & Vajen, K., 2000. Influence of the DHW load profile on the fractional energy savings: A case study of a solar combi-system with TRNSYS simulations. *Solar Energy*, 69, pp.197–208.

Kelly, N.J. & Cockroft, J., 2011. Analysis of retrofit air source heat pump performance: Results from detailed simulations and comparison to field trial data. *Energy and Buildings*, 43(1), pp.239–245.

Kelly, S., Crawford-Brown, D. & Pollitt, M.G., 2012. Building performance evaluation and certification in the UK: Is SAP fit for purpose? *Renewable and Sustainable Energy Reviews*, 16(9), pp.6861–6878.

Knudsen, S., 2002. Consumers' influence on the thermal performance of small SDHW systems—Theoretical investigations. *Solar Energy*, 73(1), pp.33–42.

Lavan, Z. & Thompson, J., 1976. Experimental study of thermally stratified hot water storage tanks. *Solar Energy*, 19, pp.519–524.

Lokale Agenda-Gruppe, 2006. Heating with environmental heat and power. Available at: http://www.agenda-energie-lahr.de/leistungwaermepumpen.html [Accessed March 16, 2015].

Loose, A. et al., 2011. Field test and performance monitoring of combined solar thermal and heat pump systems. In *ISES Solar World Congress*. Kassel.

Majcen, D., Itard, L.C.M. & Visscher, H., 2013. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications. *Energy Policy*, 54, pp.125–136.

Martin, C. & Watson, M., 2002. *Further testing of solar water heating systems*, Department of Trade and Industry (DTI), London.

Menezes, A.C. et al., 2012. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97, pp.355–364.

Miara, M. et al., 2011. *Analysis and Evaluation of Heat Pump Efficiency in Real-life Conditions*, Freiburg: Fraunhofer Institute for Solar Energy Systems.

Mørck, O., Thomsen, K.E. & Rose, J., 2012. The EU CONCERTO project Class 1 - Demonstrating cost-effective low-energy buildings - Recent results with special focus on comparison of calculated and measured energy performance of Danish buildings. *Applied Energy*, 97, pp.319–326.

Munzinger, M. et al., 2006. *PV domestic field trial final technical report*, Department of Trade and Industry (DTI), London.

NHBC Foundation, 2011. *How occupants behave and interact with their homes*, Bracknell: IHS BRE Press.

NHBC Foundation, 2012. *Low and zero carbon homes: understanding the performance challenge*, Bracknell: IHS BRE Press.

O'Flaherty, F. & Pinder, J., 2011. *The role of micro-generation technologies in alleviating fuel poverty*, Sheffield: Sheffield Hallam University Press.

Phillips, W.F. & Dave, R.N., 1982. Effects of stratification on the performance of liquid-based solar heating systems. *Solar Energy*, 29(2).

Sanyo, *Sanyo CO2 Field Test in the UK*, Unpublished.

Shorrock, L.D., Henderson, J. & Utley, J.I., 2005. *Reducing carbon emissions from the UK housing stock*, Watford: BRE Bookshop.

Siviour, J.B., 1994. Experimental U-values of some house walls. *Building Services Engineering Research and Technology*, 15(1), pp.35–36.

Sontex, 2003. The multi-functional integrator Supercal 531. *Design*, i, pp.1–11.

Sorrell, S., 2007. *The Rebound Effect: An assessment of the evidence for economy-wide energy savings from improved energy efficiency*, London: UKERC.

Stafford, a. & Lilley, D., 2012. Predicting in situ heat pump performance: An investigation into a single ground-source heat pump system in the context of 10 similar systems. *Energy and Buildings*, 49, pp.536–541.

Stark, S., Loose, A. & Drück, H., 2014. Field Test Results from Combined Solar Thermal and Air-source Heat Pump Systems with a Special Focus on Defrosting. *Energy Procedia*, 48, pp.654–663.

Stene, J., 2007. Integrated CO2 heat pump systems for space heating and hot water heating in low-energy houses and passive houses. In *International Energy Agency Heat Pump Programme Annex 32*. pp. 1–14.

Torcellini, P.A. et al., 2004. Lessons Learned from Field Evaluation of Six High- Performance Buildings. In *ACEEE Summer Study on Energy Efficiency in Buildings*.

TSB, 2013a. Low Energy Building Database. Available at: http://retrofitforthefuture.org/ [Accessed January 21, 2014].

TSB, 2013b. Retrofit for the Future. Available at: https://retrofit.innovateuk.org/ [Accessed January 21, 2014].

TSB, 2013c. *Retrofit Revealed: The Retrofit for the Future Projects - Data Analysis Report*, Report available at: https://retrofit.innovateuk.org/.

University of Wisconsin, 2013. A Transient Systems Simulation Program. Available at: http://sel.me.wisc.edu/trnsys/.

Wang, W. et al., 2011. Field test investigation of the characteristics for the air source heat pump under two typical mal-defrost phenomena. *Applied Energy*, 88(12), pp.4470–4480.

WBCSD, 2007. *Energy Efficiency in Buildings: Business realities and opportunities*, Geneva.

Zero Carbon Hub, 2010. *Carbon Compliance for Tomorrow's New Homes: A review of the modelling tool and assumptions*, Milton Keynes.

Zero Carbon Hub, 2013. *Closing the Gap Between Design & As-Built Performance*, London.

Zhang, X. et al., 2012. Test result analysis of the stagnation effect on the thermal performance of solar collector. *Energy Procedia*, 30, pp.824–828.

Zhao, H. & Magoulès, F., 2012. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6), pp.3586–3592.

Chapter 4

Predicting Solar Thermal System Performance

4.1 Evaluation of a Solar Thermal System Model from a Systems Engineering Perspective

From the perspective of systems engineering it is possible to breakdown a solar thermal system (the system of interest) into a collection of subsystems and system elements – i.e. a system-of-systems, represented in a system hierarchy in Figure 4.1.

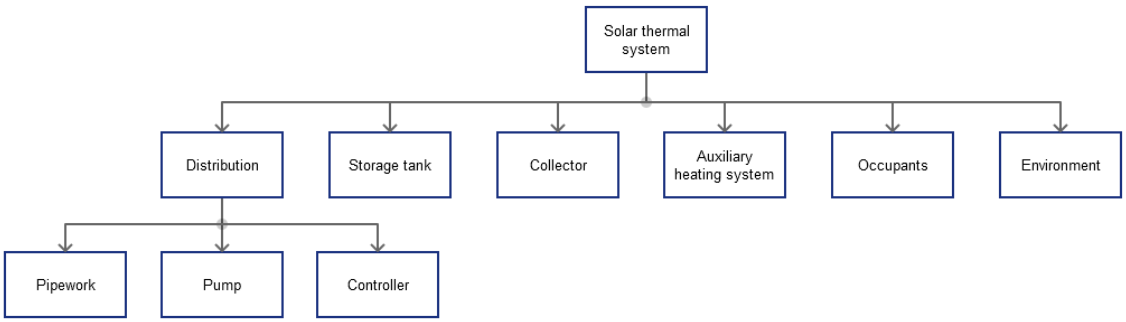

Figure 4.1: Solar thermal system hierarchy

Unified modelling language (UML) is an object oriented approach used in software engineering to model software projects prior to coding – it is analogous to the use of blueprints and engineering drawings prior to manufacture (Object Managment Group 2014). UML diagrams are able to show the static structure of a system model as well as the dynamic interactions between sub-models contained within. Figure 4.2 shows the UML diagram of a solar thermal system model with three sub-models.

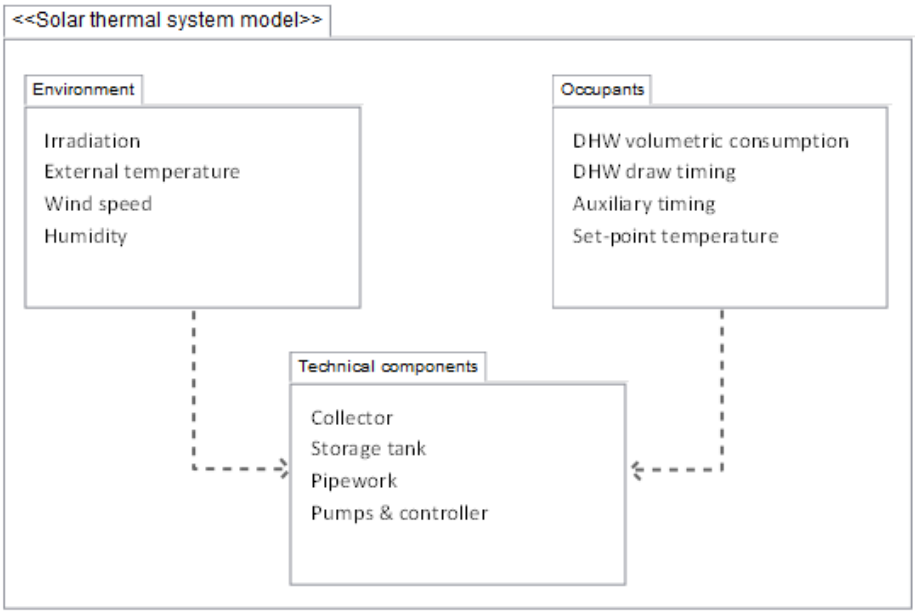

Figure 4.2: Solar thermal system model with sub-models

In each sub-model key system elements are listed for each of which the modeller will be required to set values for a number of input parameters. The level of detail and number of input parameters that need to be provided by the modeller depends on the complexity of model used. For example hot water usage profiles are typically not required for simplistic solar thermal models such as the F-chart and BREDEM-based models. This is in contrast to dynamic simulations which require time series inputs for weather patterns and occupant hot water usage.

4.2 Modelling occupant hot water usage behaviour

Occupant hot water usage behaviour varies between occupant, household and location. It is influenced by: personal needs/wants for hot water, which are a response to individual schedules or habits; climatic factors such as outside temperature; and socioeconomic factors such as the cost of water and energy, the changing size of families and households, and policies designed to encourage reduced consumption (Knudsen 2002; Michaelides 1993; Evarts & Swan 2013). Despite the uncertain nature of occupant hot water usage, it is a key component of any solar thermal system model and many hot water usage profiles have been developed and used in simulations and experiments (Mutch 1974; Furbo et al. 2005; Martin & Watson 2001; Jordan & Vajen 2000). Domestic hot water (DHW) profiles can be simple, such as a single daily draw, or more complex such as the realistic profiles developed by Jordan and Vajen (Jordan & Vajen 2000). Generally speaking the complexity of the DHW draw profile increases with its realism.

4.2.1 Single draw

Single daily draws can be taken during the morning, afternoon or evening and therefore allow the effect of drawing at different times of day to be starkly contrasted (Jordan & Vajen 2000; Martin & Watson 2001). The obvious drawback of this type of use profile is that it does not allow the effect of a more distributed usage pattern on solar thermal system performance to be investigated (Martin & Watson 2001).

4.2.2 Multiple draws

There are different standards for multiple draw patterns in existence and are used by different research bodies. For example the three a day draw profile laid out in the standard EN 12977 has been widely adopted (Figure 4.3) (Furbo et al. 2005; Martin & Watson 2001; Jordan & Vajen 2000).

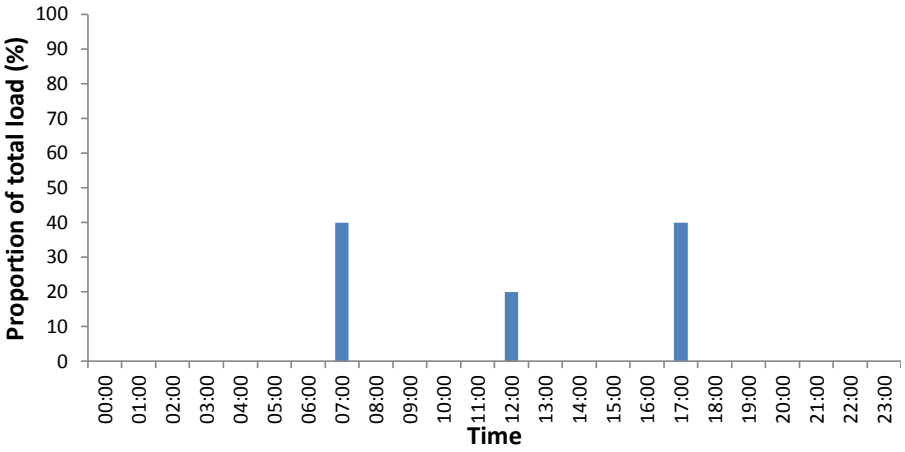

Figure 4.3: Draw profile according to EN 12977; flow rate is maintained at 10L/min

The Building Research Establishment (BRE) have developed a draw profile consisting of nine draw offs per day from 7:15 to 21:30. This profile is used for three draw volumes: 168L/day; 298 L/day; and 383L/day; the temperature difference between hot and cold water is assumed to be 55°C and the flow rate depends on the amount of hot water drawn during the day (Spur et al. 2006). The advantage of multiple draw profiles is that they allow the effect of a distributed draw profile on the solar thermal system performance to be evaluated they can be used to compare and contrast time of day effects; for example Knudsen performs early morning and evening draws consisting of five equal draws spread one hour apart (02:00, 03:00, 04:00, 05:00 and 06:00; and 14:00, 15:00, 16:00, 17:00 and 18:00) (Knudsen 2002).

4.2.3 Realistic draws

Realistic profiles are the most complex type of draw profile that can be used in a solar thermal system model. Raw data is used to construct a statistically representative profile: The RAND profile (Mutch 1974) is an early example of such a profile and has been used widely (Shariah & Lof 1996; Shariah & Löf 1997; Shariah & Ecevit 1995; Vine et al. 1987; Michaelides 1993). Jordan and Vajen use measured data to develop a DHW profile that changes from day to day based on the probabilities of four load types (small, medium, shower and bath loads) occurring throughout the day. The model also takes into consideration the probability of the daily profile shape based on the time of year (using a sine function to vary the daily DHW load), the day of the week (the probability of the load type is dependent on the day of the week) and the occurrence of holidays (Jordan & Vajen 2000). The advantage of this probabilistic DHW profile is that is doesn't remain fixed for each day of the year, but varies as would real usage patterns. However the disadvantage of this method is that when it comes to performing laboratory experiments it become impractical since it can only be used for an entire year (Spur et al. 2006). Spur et al overcome this limitation by reducing the complexity slightly; instead of a changing daily profile based on the probability of occurrence of load types, three fixed profiles were

developed based on average daily profiles from the same dataset used in the Jordan and Vajen profile. Three profiles were developed to address three types of household water use: light (100L/day), medium (180L/day) and heavy (320L/day) usage.

Researchers have used water use data for different nations to produce solar thermal system models in different parts of the world (Hasan 1997; Ijumba & Sebitosi 2010). Therefore realistic profiles can more accurately represent the water use behaviour of occupants in different countries.

Since water usage is affected by climate and socioeconomic factors (Michaelides 1993) realistic profiles will differ from country to country and cannot be applied across all locations. In addition water saving policies, more efficient technologies, changing family sizes and changing water prices will all affect usage behaviour of occupants (Knudsen 2002) and therefore realistic profiles will change accordingly.

4.2.4 Hot water usage in simple physics based models

The above describes the development and use of DHW profiles for dynamic simulations of solar thermal systems; however more simplistic models such as the F-chart and BREDEM based models do not require time-based usage profiles. In BREDEM a single figure for the average daily volumetric demand for hot water is obtained from a regression equation developed using measured data that correlates volumetric demand to the number of occupants. This makes the model quite simple to use with the ability to make predictions in common software packages such as Microsoft Excel. Occupant hot water usage behaviour is incorporated in the statistical relationships used in simplified models provided that the empirical data collected is representative of the population (Swan & Ugursal 2009); however since usage behaviour is hidden away in the statistical relationships they cannot be used to investigate the effects of different usage patterns on the performance of solar thermal systems. Further details regarding the prediction of DHW usage in BREDEM based models can be found in section 4.4.2.

It is important to have an accurate estimate of water usage behaviour to aid in system sizing and system performance estimation for returns on investment, carbon predictions and renewable energy policies.

4.3 Modelling the solar resource

Climatic data of some form is required for all solar thermal system models; the data requirements depend on the complexity of the model. For example simple modelling approaches make use of monthly average climatic data (Henderson & Hart 2013; Klein et al. 1976), whereas more complex simulations require hourly data for the period of one year (Solar Energy Laboratory 2006). Climatic input data is produced using measured data over a number of years for specific locations. BREDEM based models provide monthly average horizontal irradiance data for different regions of the UK, in contrast TRNSYS and other dynamic simulation software have pre-installed climatic data obtained from database packages such as Meteonorm. Alternatively data collected by the user can replace default options.

4.3.1 Meteonorm

Meteonorm (Remund et al. 2008) was developed in the early 80's and is a database package that contains climatic data from over 8000 weather stations around the world. Climatic data collected over a period of at least 10 years is used to give monthly average radiation values. A combination of ground stations and satellites are used to collect climate data; satellites are used to interpolate radiation values over remote areas where ground stations do not exist. Daily and hourly radiation data is obtained from the monthly data using stochastic models; this overcomes the difficulties associated with collecting and interpolating hourly radiation data (storage space and time limitations). There are several models incorporated in Meteonorm that calculate the radiation on tilted planes using the global radiation data (Perez model (1986 and 1991); Hay's model (1979); Skartveit and Olset model (1986); Gueymard's model (1987)).

Meteonorm weather files can be used in the vast majority of solar thermal modelling tools in existence (for example: TRNSYS, T*SOL, EnergyPlus, POLYSUN).

4.3.2 PVGIS

PVGIS is a web-based climate model (Šúri et al. 2005). It combines the r.sun solar radiation model (Suri & Hofierka 2004) with the geographical information system (GIS) GRASS ((Neteler & Mitasova 2002) cited in (Šúri et al. 2005)). Solar radiation models use empirical equations to estimate irradiation over large areas including the effects of surface inclination, orientation and shadowing; GIS provide information about the terrain. The aim is to provide annual and monthly predictions of solar radiation for areas with dynamic terrain, which Meteonorm is unable to do due to the coarse spatial resolution of the measurements (Šúri et al. 2005). Dynamic terrain describes areas with varying elevations which can lead to terrain shadowing and significant fluctuations in local irradiance levels. The temporal resolution of PVGIS is limited by the resolution of the data collected by ground stations and satellites used in the radiation model.

PVGIS is not used in dynamic simulation software, but is ideal for quick and easy estimations of solar radiation that may be useful for policy makers and owners of solar thermal and PV systems.

4.4 Simplified physics based models

Simplified physics models are methods that use physical principles and empirical data to determine the monthly energy performance of solar thermal systems. There are several models based on this approach that are used for estimating the energy performance of UK building stock (BREHOMES (Shorrock & Dunster 1997), the Johnston model (Johnston 2003), UKDCM (Boardman et al. 2005), DECarb (Natarajan & Levermore 2007), CDEM (Firth et al. 2010)). These are all based on the Building Research Establishment Domestic Energy Model (BREDEM) (Henderson & Hart 2013).

The application of a simplified physics approach exclusively to solar thermal systems is not quite so widespread. One of the earliest methods that uses this simplified approach is the F-chart (Klein et al. 1976). In the UK BREDEM can be used to calculate the energy performance of solar thermal systems. The UK national calculation procedure, SAP (Standard Assessment Procedure) (BRE 2013), is based on algorithms contained in BREDEM and therefore also has the ability to calculate the thermal energy output from a solar thermal system. Since 2005 SAP has been

instrumental in UK climate policy, acting as the calculation method for delivering estimates of the energy performance of buildings and energy performance certificates (EPC) in line with the Energy Performance of Buildings Directive (EPBD) (BRE 2006; Lee et al. 2013; Murphy et al. 2011). In addition, it is suggested that BREDEM/SAP will play a central role in future government policies such as the renewable heat incentive (RHI) (Kelly et al. 2012; Crowther et al. 2010), being used to estimate heat output from renewable heat technologies; therefore it is important to understand the way in which these models estimate the performance of solar thermal systems and what their strengths and limitations are, discussed in section 4.5.

4.4.1 F-chart

The *f*-chart method (Klein et al. 1976; Haberl 2004) uses the Hottel-Whillier equation (4.1) ((Hottel & Whillier 1955) cited in (Klein et al. 1976)) for describing the performance of a solar thermal collector:

$$Q_u = F_R A_c [H_T(\tau\alpha) - U_L(T_i - T_a)] \quad (4.1)$$

- A_c = Area of solar collector (m^2)
- F_R = Collector heat removal factor (%)
- U_L = Collector overall energy loss coefficient (kJ/hr.K)
- H_T = Monthly average daily radiation incident on the collector surface (kJ/hr.m^2)
- Q_u = Rate of energy collection by the flat plate solar collector (kJ/hr)
- $(\tau\alpha)$ = Product of cover transmittance and plate absorptance accounting for dirt and shading
- T_a = Ambient temperature (°C)
- T_i = Temperature of fluid at collector inlet (°C)

Assuming that the rate of energy collected by the collector is equal to the heat transferred to the storage tank ((de Winter 1975) cited in (Klein et al. 1976)) then:

$$Q_u = \frac{F'_R}{F_R} F_R A_c [H_T(\tau\alpha) - U_L(T_i - T_a)] \quad (4.2)$$

In the *f*-chart method f refers to the monthly solar fraction: i.e. the monthly fraction of the heating load (space and DHW) provided by the solar thermal system. The parameter f is correlated to two dimensionless parameters X and Y which are taken from the reduced Hottel-Whillier equation; the correlation was obtained after more than 300 simulations conducted in TRNSYS. *X*, *Y*, and *f* are given as follows:

$$X = F'_R U_L \left(T_{ref} - \bar{T}_a\right) \Delta t \frac{A_c}{L} \quad (4.3)$$

$$Y = F'_R (\tau\alpha) \bar{H}_T N \frac{A_c}{L} \quad (4.4)$$

Where:

- F'_R = Collector heat exchanger efficiency factor (%)
- U_L = Collector overall energy loss coefficient (W/m^2.K)
- Δt = Total number of seconds or hours per month
- T_a = Average monthly ambient temperature (°C)
- L = Monthly total heating load for space heating and DHW (kJ)

- H_T = Monthly average daily radiation incident on the collector surface (MJ/m^2)
- N = Number of days per month
- ($\tau\alpha$) = Monthly average transmittance-absorptance product (%)
- T_{ref} = An empirically derived reference temperature (100°C)

$$f = \begin{cases} 1.04Y - 0.065X - 0.159Y^2 + 0.00187X^2 - 0.0095Y^3, \ Air\ system \\ 1.029Y - 0.065X - 0.245Y^2 + 0.0018X^2 - 0.0215Y^3, \ Liquid\ system \end{cases} \quad (4.5)$$

The parameter X provides a ratio of collector losses to heating load and represents collector losses, Y gives a ratio of absorbed solar energy to heating load and represents collector gains. The annual solar fraction (F) of the system is given by:

$$F = \frac{\sum fL}{\sum L} \quad (4.6)$$

The *f*-chart method was developed for Windows platforms by Klein and Beckman in 1993 and is still commercially available (F-Chart Software 2014). The *f*-chart method is used to estimate the long term performance of different solar thermal systems and collectors.

4.4.2 BREDEM

The Building Research Establishment Domestic Energy Model (BREDEM) was first developed in the 1980's by the Building Research Establishment (BRE). It is a simplified, steady state calculation procedure that is used to estimate the energy consumption of a dwelling based on its physical characteristics.

BREDEM is modular in its structure and provides estimates of energy consumption for various end-uses. These energy consumption predictions can be easily converted into fuel cost and CO_2 emission values using the relevant conversion factors provided. Therefore BREDEM provides an estimate of a building's energy performance in terms of three key indicators: energy consumption, cost and CO_2. It can be used to model the energy consumption of building stock and to evaluate the potential benefits of energy efficiency improvements.

The BREDEM procedure is comprised of 10 modules:

1. Energy consumption for lights, appliances and cooking
2. Energy requirement for water heating
3. Dwelling specific heat loss
4. Dwelling thermal mass
5. Solar gain
6. Internal heat gain
7. Mean internal temperature
8. Energy consumption for space heating
9. Energy consumption for cooling
10. Electricity generated by photovoltaics and wind turbines

BREDEM performs calculations on a monthly basis and sums these to provide an annual prediction of energy performance. One advantage of BREDEM is its simplicity allowing calculations to be performed in a spreadsheet.

4.4.2.1 Solar thermal performance predictions in BREDEM

The second module of the BREDEM procedure calculates the energy requirement of the hot water. This module is split into five sections:

2. Energy required to heat water

- 2.1. The volume and energy content of heated water
- 2.2. Water heating system losses
- 2.3. Energy required for electric showers
- 2.4. Hot water from solar water heating systems
  - 2.4.1.Calculating the solar energy incident on a solar collector
  - 2.4.2.Calculating the heat output of a solar water heater
- 2.5. Net water heating energy requirement

Section 2.4 provides calculations to estimate the annual energy provided by the solar thermal system. This section is broken down further into two parts: 2.4.1 calculates the solar energy incident on the solar collector (also applicable to PV and glazing calculations); and 2.4.2 calculates the heat output from the solar thermal system. Before 2.4.2 can be completed an estimate of the annual energy content of the heated water is needed; this is obtained from section 2.1. Since BREDEM is a monthly calculation procedure it is important to understand which quantities change on a monthly basis and which remain constant throughout the year.

The predicted solar heat yield is included in the net water heating energy requirement calculation (section 2.5), which is used to calculate the monthly and annual whole-building energy consumption.

4.4.2.2 Calculating the solar energy incident on the solar collector

The BREDEM calculation procedure to estimate the annual solar radiation incident on each square metre of solar collector requires input data provided by the user or from tables contained within the BREDEM documents. These input data, related to pitch, orientation and latitude of the collector, are combined with monthly horizontal solar flux data to produce estimates of the monthly incident solar flux; this is converted into monthly incident solar energy, which is summed to provide the annual incident solar energy for the system.

A pitch factor (f_{pitch}) is used in the calculation of the orientation parameters (A, B, C) for the system:

$$f_{pitch} = \sin\left(\frac{\pi}{180} \times \frac{p}{2}\right) \tag{4.7}$$

$$A = k_1 f_{pitch}^3 + k_2 f_{pitch}^2 + k_3 f_{pitch} \tag{4.8}$$

$$B = k_4 f_{pitch}^3 + k_5 f_{pitch}^2 + k_6 f_{pitch} \tag{4.9}$$

$$C = k_7 f_{pitch}^3 + k_8 f_{pitch}^2 + k_9 f_{pitch} + 1 \tag{4.10}$$

Where p is the pitch of the surface in degrees and $k_{1\text{-}9}$ are orientation constants obtained from Table 15 of the BREDEM method, which provides values for $k_{1\text{-}9}$ based on the orientation of the collector (North, NE/NW, East/West, SE/SW, or South).

Next the solar height factor ($f_{\varphi\delta}$) is calculated:

$$f_{\phi\delta} = \cos\left(\frac{\pi}{180}(\phi - \delta_m)\right) \quad (4.11)$$

Where φ is the latitude of the system and δ_m is the solar declination for month m available in Table 16 of the BREDEM method, this value is the same for all latitudes.

Next the ratio to convert horizontal solar flux to solar flux incident on the collector for each month ($R_{h\text{-}p,m}$) is calculated:

$$R_{h-p,m} = Af_{\phi\delta}^2 + Bf_{\phi\delta} + C \quad (4.12)$$

$$Fx_m = Fx_{h,m}R_{h-p,m} \quad (4.13)$$

Where A, B, and C are orientation parameters calculated as above; $f_{\varphi\delta}$ is the solar height factor; Fx_m is the solar flux incident on the collector of the specified pitch, orientation and latitude for each month m; and $Fx_{h,m}$ is the horizontal solar flux for each month m obtained from Table A1 in Appendix A or site data if available.

Fx_m is converted into solar energy incident on the collector for each month (S_m) using:

$$S_m = 0.024Fx_m n_m \quad (4.14)$$

Where n_m is the number of days in each month (28 in February); these monthly values are summed to give the annual incident solar energy on the collector:

$$S = \sum S_m \quad (4.15)$$

The next step of the solar thermal performance prediction is to calculate the heat output of the solar thermal system for which an estimate of the annual energy content of the heated water is needed.

4.4.2.3 Calculating the volume and energy content of the heated water

This section of the BREDEM procedure is required for section 2.4.2 of the solar thermal energy performance prediction. It uses information about the occupants including the number of occupants and bathing habits. These are fed into formulae that determine the average daily hot water requirement and the average daily hot water requirement for each month. Input data about the water temperature rise is used to arrive at an estimate for annual energy content of the heated water.

The average daily hot water volumetric demand for each month ($V_{d,m}$) is calculated as follows:

$$V_{d,m} = V_{d,ave}f_{hw} \quad (4.16)$$

Where f_{hw} is the monthly hot water use factor obtained from Table 7 and:

$$V_{d,ave} = V_{d,shower} + V_{d,bath} + V_{d,other} \quad (4.17)$$

$$V_{d,shower} = n_{shower}V_{PS} \quad (4.18)$$

Where n_{shower} is the number of showers per day and V_{PS} is the amount of hot water used per shower. The amount of hot water used per shower can be obtained from Table 6 or from

information the user may have. The number of showers may also be known to the user; alternatively it can be approximated using the number of occupants N:

$$n_{shower} = 0.45N + 0.65 \tag{4.19}$$

Similarly of the number of baths taken is known then this value may be used; alternatively it can be approximated as follows:

$$n_{baths} = \begin{cases} 0.35N + 0.5, & if\ no\ shower\ present \\ 0.13N + 0.19, & is\ shower\ is\ also\ present \end{cases} \tag{4.20}$$

$$V_{d,bath} = 50.8 n_{bath} \tag{4.21}$$

$$V_{d,other} = 9.8N + 14 \tag{4.22}$$

The monthly energy content of the heated water ($Q_{HW,m}$) can then be found using:

$$Q_{HW,m} = \frac{4.19 V_{d,m} n_m \Delta T_m}{3600} \tag{4.23}$$

Where 4.19 is the specific heat capacity of water; n_m is the number of days in each month (28 in February); ΔT_m is the monthly rise in temperature required obtained from Table 8; and 3600 is the factor to convert the estimate from J to kWh.

The annual energy content of the heated water is simply the sum of the monthly values:

$$Q_{HW} = \sum Q_{HW,m} \tag{4.24}$$

This value is used in the calculation of the heat output of the solar thermal system.

4.4.2.4 Calculating the heat output of the solar thermal system

To calculate the heat output of the solar thermal system input data related to the collector performance, hot water requirement and available solar energy are used. Additional parameters that introduce losses in the system are also calculated.

The collector performance factor (f_1) is a parameter that is calculated using the zero loss efficiency (η_0) and the first and second order collector heat loss parameters (a_1 and a_2); it represents the heat losses from the collector:

$$a^* = 0.892(a_1 + 45a_2) \tag{4.25}$$

$$f_1 = \begin{cases} 0.97 - 0.0367\left(\frac{a^*}{\eta_0}\right) + 0.0006\left(\frac{a^*}{\eta_0}\right)^2, & \left(\frac{a^*}{\eta_0}\right) < 20 \\ 0.693 - 0.0108\left(\frac{a^*}{\eta_0}\right), & \left(\frac{a^*}{\eta_0}\right) \geq 20 \end{cases} \tag{4.26}$$

The utilisation factor (UF) allows for the fact that not all of the solar energy captured by the system directly translates to useful heating of DHW. It uses the solar energy available to the collector and the user's hot water demand:

$$LR = \frac{A_{ap}\eta_0 S Z_{panel}}{Q_{HW}} \tag{4.27}$$

$$UF = 1 - e^{(-1/LR)} \tag{4.28}$$

Where LR is the load ratio; A_{ap} is the aperture area; S is the annual incident solar radiation on the collector (calculated in the previous section); Z_{panel} is the over-shading factor obtained from Table 18; and Q_{HW} is the annual energy content of the heated water (obtained from section 2.1).

The solar storage volume factor (f_2) represents storage losses of the solar thermal system:

$$f_2 = 1 + 0.2 \ln \frac{V_{eff}}{V_{d,ave}} \tag{4.29}$$

Where $V_{d,ave}$ is obtained from section 2.1 and V_{eff} is the effective solar volume of the storage tank. V_{eff} is dependent on the configuration of the storage system:

- If a separate solar storage tank exists then V_{eff} is simply the volume of this tank;
- If a combined cylinder with twin coils (solar and non-solar) is installed then V_{eff} is the volume of the dedicated solar storage. This is the volume of the tank up to the height of the non-solar coil plus 0.3 times the remaining volume;
- If a thermal store with twin coils is installed, where the solar coil is within the store, V_{eff} is the volume of the dedicated solar storage (the volume up to the height of the non-solar coil);
- If a direct system is installed (with no dedicated solar volume) then V_{eff} is 0.3 times the volume of the cylinder

If f_2 is found to be greater than unity then it is reset to one.

The annual output of the solar thermal system (Q_{sol}) is then calculated using the loss factors and available solar radiation to the collector:

$$Q_{sol} = SZ_{panel}A_{ap}\eta_0 UF f_1 f_2 \tag{4.30}$$

Monthly solar thermal heat yield is calculated using the ratio of monthly incident solar energy (S_m) to annual incident solar energy (S):

$$Q_{sol,m} = \frac{S_m}{S} Q_{sol} \tag{4.31}$$

4.4.3 SAP

BREDEM forms the foundation of the UK's national calculation method, SAP. SAP is a compliance tool and does not strictly predict the energy performance of buildings. However the algorithms and calculations used in SAP are based on those contained within BREDEM and the modular structure is maintained. The output of the SAP method is a total annual energy consumption value for the dwelling or dwelling archetype in question. This can be converted into an annual cost to run and annual CO_2 emission value using the relevant conversion factors provided. These values are then used to give the dwelling a SAP and Environmental Impact (EI) rating: SAP ratings score the dwelling based on its annual running costs; EI ratings score the dwelling based on its annual CO_2 emissions.

The module that calculates annual solar thermal heat yield uses an identical calculation procedure to that in BREDEM with a few subtle differences that make it suitable for a compliance tool.

The main difference between SAP and BREDEM is the way in which the average daily hot water volume requirement is estimated. BREDEM uses more detailed information about the number of occupants and their bathing habits, relying on regression models only when this information is unknown to the user. However due to the usage of SAP as a compliance tool the number of occupants to inhabit a dwelling is unknown since this procedure is performed prior to occupation. Therefore the occupancy has to be estimated based on the size of the dwelling: a regression model for occupancy based on the total floor area (TFA) of the dwelling is implemented:

$$N = \begin{cases} 1 + 1.76\left[1 - e^{(-0.000349(TFA-13.9)^2)}\right] + 0.0013(TFA - 13.9), & TFA > 13.9 \\ 1, & TFA \leq 13.9 \end{cases} \quad (4.32)$$

Once the occupancy is determined the annual average daily hot water demand can be estimated; again since SAP is performed prior to occupation, the bathing habits of the users are unknown. Therefore in order to estimate hot water demand a single variable linear regression model is used:

$$V_{d,ave} = 25N + 36 \quad (4.33)$$

This regression model was obtained from a study conducted by the EST, which found the number of occupants to be the primary cause for variation in hot water demand (EST 2008b). This regression equation explains 12% (R^2 = 0.12) of the variation in the hot water consumption suggesting that occupancy is not a good predictor of DHW consumption.

After $V_{d,ave}$ is determined the SAP and BREDEM methods for calculating the annual energy content of the heated water is the same, making use of monthly use factors and temperature rises.

The next difference between SAP and BREDEM is the way in which the annual incident solar energy is calculated. The SAP methodology is far more simplistic than the BREDEM method and does not make use of pitch factors, orientation parameters or latitude. The annual incident solar radiation is obtained from a table which provides estimates of annual solar energy per square metre of collector based on the pitch and orientation alone.

After Q_{HW} and S are determined in SAP the method follows the BREDEM procedure to arrive at an estimate of annual solar thermal heat yield.

4.5 Strengths and weaknesses of simplified physics based models

The major strength of simplified physics models is their simplicity; in contrast to dynamic simulations, which require specialist knowledge to use correctly, simplified models can be used by a wide variety of non-experts including government departments, policy makers and local authorities along with more adept users such as energy auditors and companies (Kelly et al. 2012). However there is a danger that policy makers will overlook the limitations of such models in favour of their ease of use (Kavgic et al. 2010).

The simplicity lies with the implicit inclusion of complex thermodynamic relationships within the statistical relationships obtained from empirical data (Murphy et al. 2011). However because

complex physical relationships are incorporated in the empirical relationships there is a great demand for empirical data. There are two ways in which empirical data may be collected: laboratory experiments; and field trials. Laboratory experiments are less expensive and easier to perform compared to field trials, however assumptions about occupant hot water usage behaviour are needed, which are not likely to be representative of real life (Martin & Watson 2001). Therefore field trials and monitoring schemes such as those performed by the Energy Saving Trust (EST) (EST 2008b; EST 2010; EST 2011)and others (BRE 2005; Munzinger et al. 2006; Forward & Roberts 2008) are a valuable source of empirical data for which to extract statistical relationships between variables. However widespread monitoring schemes are fraught with difficulties and considerations:

It is not practical to monitor every household or system in the field and so a sample of the population must be chosen. It is important that this sample is representative of the population; for example when monitoring solar thermal systems in the field different collector types, system configurations, orientations, roof pitches and locations exist and should be accounted for in the sample (EST 2008a). Although the BREDEM based models are continually being reviewed as new data emerges (Henderson 2002; Anderson 2005), the data used to develop the models have been criticised for being unrepresentative due to small, homogenous samples (Kelly et al. 2012; Kavgic et al. 2010).

Before any field trial or experiment is performed there needs to be full understanding of what parameters will need to be measured, how long they will need to be measured for and at what accuracy and frequency (EST 2008a). These considerations are not required in dynamic simulations for which detailed output information for each parameter can be obtained (Murphy et al. 2011).

The field trial or experiment needs to capture variation in the measured parameters and performance over time; therefore if long term performance is to be evaluated then monitoring periods of at least one year are required. Most field trials collect data for the period of one year; however it is preferable to collect data for more than one year to allow systems to settle down and for any monitoring issues to be addressed in the first year.

The frequency of measurement also needs to capture variation; for example solar radiation can vary rapidly throughout the course of the day and for deeper understanding of solar system performance to be obtained measurements need to be made at appropriate time intervals. Often the 5 minutely data is collected (EST 2011).

When a parameter is measured there will be a difference between the measured value and the true value; this is determined by the accuracy of the measurement equipment and will lead to uncertainties in the analysis. If the inaccuracies in the measurement are too great then the uncertainty in the true value of a parameter can render the analysis useless. However, costs of measurement equipment increases as accuracy increases and so there needs to be a compromise between expenditure and accuracy.

The number of systems, monitoring period, number of parameters and measurement frequency mean that there will be masses of raw data, which needs storing, quality checking, processing and finally analysing.

The major disadvantage of monitoring schemes is the cost. There are costs associated with the measurement equipment; installation; data acquisition system; and server costs or personnel costs depending on how the data will be retrieved. A field trial of 100 systems/households can cost upwards of £250,000 (EST 2008a).

Often during field trial studies there will be periods of downtime. This may be due to sensor faults; installation errors; server issues or caused by the occupants. Downtime can be small and provided the majority of the data for each month is available (according to the requirements of the project) may not cause any issues. However serious levels of downtime can lead to systems being removed from the sample; for example Evarts & Swan collected hot water usage data for 1594 dwellings, a third of these properties had missing data (Evarts & Swan 2013). In an earlier study conducted by the EST twelve households had to be removed entirely from the database due to problems during data collection (EST 2008b). Other studies had similar issues (EST 2001; Kelly & Cockroft 2011; Munzinger et al. 2006; EST 2011);

These difficulties make frequent, long term monitoring schemes hard to implement and therefore the validity of models that depend upon this data for empirical relationships is in question as time goes by; for example BREDEM is validated against a small number of similar building types and may not be relevant to the next generation of low carbon homes (Kelly et al. 2012). Similarly historic data used to produce the empirical relationships associated with solar thermal system performance in BREDEM and SAP may not be applicable to new systems with improved efficiencies (Murphy et al. 2011), and data pertaining to hot water usage patterns is already 6 years old (EST 2008b) and may no longer reflect current trends especially in low energy homes that contain efficiency measures such as low flow appliances and grey water recycling.

The large amounts of empirical data required to produce the algorithms used in simplified models may be obtained from dynamic simulation (Murphy et al. 2011; Klein et al. 1976). The first principles approach taken by dynamic simulation techniques to solve thermodynamic equations means that although detailed input data are needed they do not rely on empirical relationships, therefore, overall data requirements are lower than those of simplified models (Murphy et al. 2011). In addition, provided that the expert knowledge needed to perform such simulations is available then they are vastly quicker than laboratory experiments and field trials. However dynamic simulations are faced with their own set of limitations as described in section 4.13. Although empirical data requirements are high, once the algorithm has been developed the input data required from the user is relatively low compared to dynamic simulations. This has several benefits: detailed information about the solar thermal systems and usage behaviour is often unknown, thus reducing the accuracy of dynamic simulations (Evarts & Swan 2013; Zhao & Magoulès 2012); with simpler and smaller data input requirements the chance for errors due to calculation and human input are minimised (Murphy et al. 2011).

BREDEM based models are customisable due to their modular structure, minimal data requirements and empirical relationships and can be easily imported in Excel to be used widely. They allow the incorporation of data from several different sources, and encourage the use of measured data where possible. In addition the modular structure allows complex dynamic simulations to be performed in conjunction with the BREDEM model; for example a designer may choose to use TRNSYS to model a solar thermal system and use this result in the BREDEM

calculation for whole-building energy consumption. The BREDEM model has been used to construct other energy performance models in the UK (CDEM, BREEHOMES, Johnston, etc.) owing to its customisability. This is advantageous since predictions made early in the design stage using such models can be refined as knowledge is gained about the system in question through more complex simulations techniques and measured data.

Simplified models are unable to disaggregate the occupant usage behaviour from the relationships used to make predictions; for example in the BREDEM based models the volume of the hot water is obtained from regression equations (unless it has been measured) but the time of use is not specifiable. This is a limitation to simplified models since it has been shown extensively that time of use can affect solar thermal system performance (Jordan & Vajen 2000; Knudsen 2002; Rodríguez-Hidalgo et al. 2012; Spur et al. 2006; Morrison et al. 1992). Although the temperature rise values of the hot water are based on measured data, they are average monthly values obtained from 112 households (EST 2008b). The dependency on simple, single variable regression equations and average values excludes uncertainty from the predictions of occupant usage behaviour. The uncertainty of occupant behaviour, as a direct result of its variable and unpredictable nature will impact on the performance of real solar thermal systems and should be included in predictions to enable informed investment decisions to be made. In addition single variable regression equations fail to describe all the variation in the hot water usage and do not offer explanations as to the causes behind the correlations between variables and so the different elements of occupant behaviour have unknown effects on the water usage and solar thermal system performance. The utilisation factor in SAP describes the losses in a solar thermal system relating to a mismatch between available solar energy and the load. However Murphy et al. show that the SAP predicted utilisation factor fails to take into consideration tank size, DHW volume consumed and weather (Murphy et al. 2011). The SAP utilisation factor also does not make allowance for DHW profile because the empirical data used to construction the relationship was obtained from laboratory experiments of eight systems (Martin & Watson 2001).

Any modelling approach that relies on measured data must face the fact that the data may become obsolete over time as new energy efficient measures are introduced and socioeconomic factors change (Kavgic et al. 2010).

4.6 Statistical modelling

Statistical modelling involves using empirical data to produce regression equations that allow an output (dependent variable) to be predicted based on inputs (independent variables), regression coefficients and intercepts (Miller et al. 1990). They differ from simplified physics models because the regression equations are not necessarily based on physical principles, but instead are statistically determined using curve fitting techniques such as the method of least squares estimation. However, regression equations are often used in simplified models for estimating parameters used in the model's governing equations; for example SAP uses a single variable regression equation to estimate the annual average daily volumetric DHW demand based on the number of occupants. The estimated volume is used to calculate the energy content of the hot water using the physical relationship:

$$Q_{DHW,m} = V_{d,m} c \Delta T_m \quad (4.34)$$

$$V_{d,m} = u_m(25N + 36) \quad (4.35)$$

Here, the subscript 'm' stands for the month 1-12; $V_{d,m}$ is the average daily volumetric demand for DHW for month m and is given by the annual average daily demand represented by the statistical relationship ($25N+36$); 'u_m' is the monthly factor by which the annual average volumetric demand is multiplied to give the monthly average demand; 'c' is the specific heat capacity of water; 'ΔT_m' is the temperature rise of the water in the tank for month m.

4.6.1 Types of regression

There are different types of regression as detailed in (Miller et al. 1990), these include:

- Linear regression;
- Curvilinear regression;
- Multiple regression;

4.6.1.1 Linear regression

This is the simplest form of regression and relates the value of a dependent variable (y) to a single independent variable (x). Typically the error in the value of x is negligible compared to the variation that can exist in y; for example assuming x to be controlled to a fixed value, the resulting value of y may differ over several tests. This variation in y can be due to error in the measurement of y, or because of other factors that exist during the experiment. Therefore a single value of x produces a distribution of possible y values. Regression modelling is concerned with the mean value of y for a given x; differences from the mean value of y as predicted by x are denoted by an error term ε. A linear regression model of y on x is given by the following general formula:

$$y = \alpha + \beta x + \varepsilon \quad (4.36)$$

The error term, ε is a random variable and a value for α can be chosen so that the mean of ε is zero. To estimate values for α and β a non-subjective method is required that can fit straight lines; this method is called the method of least squares.

Given n pairs of (x_i, y_i), for which the relationship is linear, a prediction for y can be given by:

$$\check{y} = a + bx \quad (4.37)$$

The error associated with predicting y given x_i is:

$$e_i = y_i - \check{y}_i \quad (4.38)$$

The line of best fit occurs when the error is minimised; therefore the values of a and b need to be determined so that the error is as close to zero as possible. This is achieved by minimising the sum of the squares of the error:

$$\sum_{i=1}^{n}(e_i)^2 = \sum_{i=1}^{n}[\check{y}_i - (a + bx_1)]^2 \quad (4.39)$$

This is equivalent to minimising the sum of the squares of the vertical distances from the data points to the line of best fit. The method of least squares gives values of a and b which are estimates for α and β.

4.6.1.2 Curvilinear regression

Curvilinear regression is concerned with non-linear relationships between y and x; the method of least squares can still be applied. In the past a non-linear regression equation could be found by plotting data points on different types of graph paper to observe for which transformed scale the data points produced a straight line. For example if the points (y_i, x_i) produce a straight line when $logy_i$ is plotted against x_i then this indicates that the regression curve of y on x is exponential and given by the general equation:

$$y = \alpha\beta^x \tag{4.40}$$

Taking the log of this equation linearises the relationship ($\log y = \log\alpha + x\log\beta$) allowing α and β to be found using the method of least squares on ($logy_i$, x_i).

Other non-linear forms that arise in engineering applications are the reciprocal and power functions:

$$Reciprocal: y = \frac{1}{\alpha+\beta x} \tag{4.41}$$

$$Power: y = \alpha x^{\beta} \tag{4.42}$$

The method of least squares can be applied to the linearised forms of these relationships:

$$\frac{1}{y} = \alpha + \beta x \; applied\; to\; the\; points\; (\frac{1}{y_i}, x_i) \tag{4.43}$$

$$\log y = \log\alpha + \beta\log x \; applied\; to\; the\; points\; (\log y_i\,, \log x_i) \tag{4.44}$$

If none of the above non-linear functions fit the data, that is the regression model of y on x has an unknown form, then it is assumed that the regression model is a polynomial and that the first few terms of a Taylor series expansion will fit the data, giving the general form:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p \tag{4.45}$$

The method of least squares is applied to minimise the distance between the data points and the curve; this is achieved by minimising:

$$\sum_{i=1}^{n}\left[y_i - (\beta_0 + \beta_1 x_i + \beta_2 {x_i}^2 + \cdots + \beta_p {x_i}^p)\right]^2 \tag{4.46}$$

Following differentiation and rearranging of the above equation as described in (Miller et al. 1990) the following equations are produced, which are solved simultaneously to find b_0 to b_p, estimates for β_0 to β_p:

$$\begin{aligned} \textstyle\sum y &= nb_0 + b_1\sum x + \cdots + b_p\sum x^p \\ \textstyle\sum xy &= b_0\sum x + b_1\sum x^2 + \cdots + b_p\sum x^{p+1} \\ &\vdots \\ \textstyle\sum x^p y &= b_0\sum x^p + b_1\sum x^{p+1} + \cdots + b_p\sum x^{2p} \end{aligned} \tag{4.47}$$

4.6.1.3 Multiple regression

Multiple-regression involves the prediction of y based on more than one independent variable, x_1-x_r.

The method of least squares is used in all of the regression models described above. Nowadays, commercial software such as Microsoft Excel allows the user to perform simple curve fitting techniques for single variable problems or more complex regression analysis for obtaining a regression model based on multiple independent variables.

4.7 Predicting long term solar thermal performance using the method of least squares

There are few regression models in existence for long term solar thermal system performance predictions. Martin and Watson regressed daily solar yield onto daily solar radiation for 8 solar thermal systems for different draw profiles (Martin & Watson 2001). This study found strong positive correlations between daily solar yield and daily solar radiation levels, it was also discovered that there was very little difference in solar yield between a three a day and single draw profile; however effects of auxiliary timing were not investigated in this study. In addition a regression equation for each system was generated in order to make long term predictions of performance; therefore these predictions can only be made with the appropriate equation limiting the usefulness of such estimations in general terms. The systems tested were subjected to consistent draw volumes each day so that the real life performance is unknown. A follow up study to investigate the effects of auxiliary inputs was performed, however the auxiliary schedule was optimised and so the real impacts of auxiliary input are unknown from these results (Martin & Watson 2002). Similar regression equations giving collector output as a function of daily solar radiation are provided for four in-situ solar thermal systems (The Energy Monitoring Company Ltd 2001). Again these regression equations are only applicable to the appropriate system. Data and from these studies were used in the development of the solar thermal prediction method in BREDEM (Henderson 2002).

In terms of predicting the long term performance of unspecified domestic solar thermal systems installed in the UK no regression analysis has been performed. A reason for this may be the dearth of field trial studies in this area and therefore the lack of empirical data, which is required to produce regression models with the least squares method.

Field trial studies provide valuable data about the in-situ performance of solar thermal systems and have the advantage that occupant hot water usage behaviour does not need to be simulated as is the case for dynamic models and laboratory experiments. In the UK field trial studies of solar thermal systems are limited to a few in number; an early field trial performed by The Energy Monitoring Company Ltd on behalf of the Department of Trade and Industry (DTI) consisted of two years monitoring of four solar thermal systems in England and Scotland (The Energy Monitoring Company Ltd 2001). Following this a field trial of six systems in Sheffield and Suffolk was performed by Viridian Solar, a solar thermal systems manufacturer (Forward & Roberts 2008). Ayompe et al (2010) monitored flat plate and evacuated tube system for a period of one year, however this was a long running experiment as opposed to an in-situ field trial (Ayompe, Duffy, Mc Keever, et al. 2011). In 2010 a large scale solar thermal system field trial

was performed by the Energy Saving Trust in which 88 systems throughout the UK were monitored for a period of one year.

Armed with this data it is possible to produce a regression model of the in-situ long term performance of solar thermal systems. The method of least squares as described above was used to estimate the linear relationship between the specific solar thermal heat yield (the dependent variable) and two independent variables describing the system size and DHW demand.

4.7.1 Organising the data

The first step to be performed before the least squares method could be applied to the data was to organise the data from different sources. The regression model is concerned with long term performance predictions of solar thermal systems, which is defined as the amount of thermal energy yielded from the collector in one year; therefore annual data was required for the model.

One of the disadvantages of using regression to model long term performance is that data for dependent and independent variables needs to be complete for the year. There were numerous issues with the monitored data from the EST solar thermal field trial which meant that of the 88 systems originally monitored only 24 could be used for the regression modelling. The issues with monitoring during the EST field trial generally result in missing data for large parts of the year meaning that the true annual values of the variables is unknown.

Data from three other monitoring studies (The Energy Monitoring Company Ltd 2001; Forward & Roberts 2008; Ayompe, Duffy, Mc Keever, et al. 2011) were combined with the EST data into a single spreadsheet containing the following data:

- Annual specific solar heat yield - Q_{sol} (kWh/m^2/yr) – this is the dependent variable;
- Annual solar incident irradiation – S (kWh/yr);
- Annual average daily DHW consumption – $V_{d,ave}$ (L/day);
- Annual average temperature rise of the hot water – ΔT (°C);
- Annual auxiliary input – Q_{aux} (kWh/yr);
- Tank volume-to-collector area ratio – $VA\ ratio$ (L/m^2);
- System type (evacuated tube or flat plate)
- Tank volume-to-annual average daily DHW consumption – $V_{tank}/V_{d,ave}$;
- Annual average DHW consumption-to-collector area – $V_{d,ave}$/A (L/m^2);

These data describe the system type and size, climate and energy usage of the systems and were used to produce predictions of annual specific solar heat yield.

The sample size used in the regression analysis was 37 systems: 10 evacuated tube and 27 flat plate collectors.

4.7.2 Variable selection

The initial selection of independent variables is typically based on theory and the modeller's prior knowledge of the system and the factors that influence its performance (Farrar & Glauber 1967). It has been shown that solar radiation affects the heat yield of solar thermal systems (Martin & Watson 2001; Martin & Watson 2002); the VA ratio is also shown to affect

performance (Shariah & Lof 1996) as does DHW demand (Allen et al. 2010; Knudsen 2002) and auxiliary input (Shariah & Löf 1997); the size of the storage tank in relation to DHW demand has also been shown to be an important consideration when designing solar thermal systems (Knudsen 2002).

After the initial set of variables has been established the aim becomes to find the "best" regression model; the best model is the simplest model that adequately fits the data (Crawley 2015). Models are simplified by removing variables that add very little to the explanation of the variance in the dependent variable. A measure of the explained variance is the coefficient of determination (R^2); a regression model that produces an R^2 value of 1 explains 100% of the variation in the dependent variable. However obtaining an R^2 value of 1 is extremely unlikely when fitting to data due to factors that affect the dependent variable that may not be captured in the independent variables, i.e. there are unknown variables needed to explain the variation fully. For example consider the prediction of DHW volume based on occupancy; There is a wide variation in the DHW demand for a given occupancy level which means that additional variables are required that explain this variation and thus give better predictions of DHW demand. Adding variables to a regression equation will either increase the R^2 value or do nothing – it will never reduce the explained variance.

Despite the seemingly obvious requirement of a regression model to explain as much variance as possible, those models with the highest R^2 values are not necessarily the best. The concept of parsimony is generally considered favourable in statistical modelling since an overly complex model is sensitive to noise (Crawley 2015; Kelly 2011). The aim is to produce a model with the fewest variables but still explains a portion of the variance comparable to that explained by the fully unrestricted model. The term unrestricted model refers to the model based on all independent variables; a restricted model is one with a number of those variables removed.

There are several methods to remove variables from an unrestricted regression model: comparisons of R^2 and correlation (ρ); forward elimination; backward elimination; and stepwise regression.

To use R^2 to reduce variables every conceivable regression equation using the variables is considered ranging from no variables to all the variables. The R^2 values for all of these models are considered and the highest ones selected. Following this correlations between the independent variables are investigated and where there is significant correlation (ρ<-0.7, ρ>0.7) a variable is eliminated. For a model with n-independent variables the number of possible regression models is 2^n. It is for this reason that this method was not employed; there are 7 independent variables giving a possible 128 regression models. Obtaining the R^2 values for this number of equations would be impractical.

Backward elimination was the method used to simplify the unrestricted regression model. It makes use of the ANOVA (**AN**alysis **O**f **VA**riance) table (Table 4.1).

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	150939.83	21562.83	4.57	0.0018
Residual	27	127480.62	4721.50		
Total	34	278420.45			

Table 4.1: ANOVA table for the unrestricted regression model

4.7.2.1 Backward elimination:

To use the backward elimination technique, start with the unrestricted model and systematically remove variables based on a partial F-test. This differs from forward elimination which begins with no variables and systematically adds variables based on a partial F-test.

A partial F-test involves calculating the F-value of the restricted model and comparing this to the F-value obtained from the F-distribution at a given significance level. The significance level is a probability given by the F-distribution and is denoted α; α often takes the values 0.05 or 0.01 which are 5% and 1% significance respectively. Significance represents the probability at which a type I error is made; rejecting a true null hypothesis is a type I error. The first pass of the backward elimination method involves removing each variable in turn from the unrestricted equation and calculating the F-value using:

$$F = \frac{SS_{res,u} - SS_{res,r}}{MS_{res,u}} \tag{4.48}$$

The subscript 'u' refers to the unrestricted model and 'r' to the restricted model. SS_{res} is the sum of the squared error for the residuals and MS_{res} is the mean square error of the residuals; these values are provided in the ANOVA table. The F-values obtained for each of the restricted models is compared to F_α where α represents the significance level of 0.05. F_α is determined from the F-distribution using the degrees of freedom (d_1 and d_2) of the model:

$$d_1 = p - q \tag{4.49}$$

Where p is the number of variables in the unrestricted model not including the intercept (7 in this case) and q is the number of variables of the restricted model not including the intercept (6 in this case).

$$d_2 = (N - 1) - q \tag{4.50}$$

Where N is the number of data points (in this case 35) and q is as before. The probability of a variable X with an F-distribution, $X \sim \mathrm{F}(d_1, d_2)$ is given by the probability density function, where B is the Beta function:

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{xB\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \tag{4.51}$$

The inverse of this function with probability 0.05 gives F_α. The F-values obtained from data in the ANOVA table are compared to F_α: If the lowest F-value is $< F_\alpha$ then the corresponding variable is removed from the equation. This is a form of null hypothesis testing: the null hypothesis in this case is that the restricted model is *not* worse than the unrestricted model. The null hypothesis is

rejected when the F-value is less than F at significance level 0.05. When the F-value is less than $F_{0.05}$ it means that the probability of the null hypothesis being rejected purely by chance is less than 5% or in other words that the probability of rejecting the null hypothesis given that it is true is 5%. Once a variable has been removed the resulting model is redefined as the unrestricted model and the process is repeated until the lowest F-value obtained is $>F_\alpha$ at which point the model is considered to be the optimum model providing simplicity and an adequate fit of the data. An example of how the backward elimination procedure is performed is shown in Table 4.2.

Variable removed	RSS unrestricted	MSE unrestricted	RSS restricted	F-value	d1	d2	$F_{0.05}$
$V_{d,ave}/A$	127480.621	4721.504	127761.909	0.060	1.000	27.000	4.210
Q_{aux}	127480.621	4721.504	133426.118	1.259	1.000	27.000	4.210
$V_{tank}/V_{d,ave}$	127480.621	4721.504	146133.391	3.951	1.000	27.000	4.210
VA ratio	127480.621	4721.504	189565.286	13.149	1.000	27.000	4.210
$V_{d,ave}$	127480.621	4721.504	129489.038	0.425	1.000	27.000	4.210
ΔT	127480.621	4721.504	130144.072	0.564	1.000	27.000	4.210
S	127480.621	4721.504	131831.463	0.921	1.000	27.000	4.210

Table 4.2: Partial F-test for the first iteration of the backward elimination procedure

The lowest F-value is associated with $V_{d,ave}/A$ and is 0.060; since $0.060<F_\alpha$ then this variable is removed and the new unrestricted model is a function of 6 variables (Q_{aux}, $V_{tank}/V_{d,ave}$, $VA\ ratio$, $V_{d,ave}$, ΔT *and* S).

The 7-variable unrestricted model explained 54.2% of the variation in Q_{sol}; the new simplified 2-variable model explains 47.6%.

4.7.3 The resulting regression equation for predicting long term solar thermal system performance

The backward elimination procedure as outlined above provided the following regression equation for predicting solar thermal system performance in terms of kWh/m^2/yr:

$$Q_{sol} = \beta_1 \left(\frac{V_{tank}}{A_c}\right) + \beta_2 \left(\frac{V_{tank}}{V_{d,ave}}\right) + \alpha \tag{4.52}$$

$$\alpha = 184.60;\ \beta_1 = 3.00;\ \beta_2 = -23.13 \tag{4.53}$$

Where V_{tank} is the volume of the tank; A_c is the collector area; $V_{d,ave}$ is the annual average daily DHW volumetric demand; and β_1, β_2, and α are the regression coefficients and intercept respectively.

The simplified regression model for predicting the specific solar heat yield suggests that the system size and DHW usage (or more specifically the usage with respect to the size of the storage tank) are the most important variables to consider and explain enough of the variance in Q_{sol}.

Figure 4.4 shows the annual specific solar thermal heat yield as a function of the two variables in the simplified regression equation.

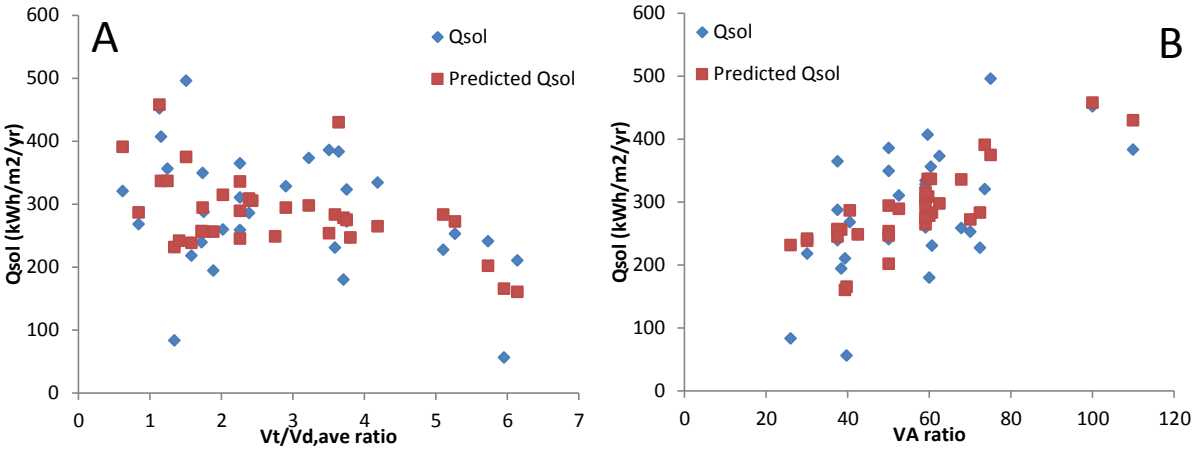

Figure 4.4: Predicted and measured Q_{sol} as a function of a) $V_{tank}/V_{d,ave}$; and b) *VA ratio*

A negative correlation between $V_{tank}/V_{d,ave}$ and Q_{sol} exists because a large storage tank combined with a small DHW draw leads to increased tank temperatures due to a lack of recharge of the storage tanks with cold water as suggested by (Wolf et al. 1984; Rodríguez-Hidalgo et al. 2012; Bates et al. 1999). The increased tank temperature means that the solar controller switches the pump off due to the temperature difference between the collector fluid and tank being too small; in addition the heat transfer between the solar coil and the storage tank is reduced as the difference in temperature between the two reduces. Conversely there is a positive correlation between $VA\ ratio$ and Q_{sol} suggesting that as the storage tank increases for a given collector then more energy is delivered by the system (Hobbi & Siddiqui 2009; Shariah & Ecevit 1995; Shariah & Lof 1996). This can be explained by larger tanks providing more capacity for solar heat since more thermal energy is required to heat a larger amount of water than a smaller amount. However if this water is not used and a high tank temperature is maintained then solar yield is limited as explained previously.

The correlation between Q_{sol} and annual irradiation, S is only 0.11 and explains only 1.2% of the variation in Q_{sol}. Although in general a higher irradiation value has been shown to increase the solar yield for a given system on a daily time scale (Figure 3.28 in Chapter 3), the amount of annual irradiation between systems in the sample for the regression analysis is relatively consistent ($S \sim \mathrm{N}(993.15,\ 139.60)$); therefore any variation in Q_{sol} is not likely to be due to S, but to other factors which are more variable between systems such as occupancy usage behaviour.

4.7.3.1 Residual plots

The residual plots (Figure 4.5) indicate whether the regression model is adequate or not. The residuals should be evenly spread, with no correlations or patterns being obvious.

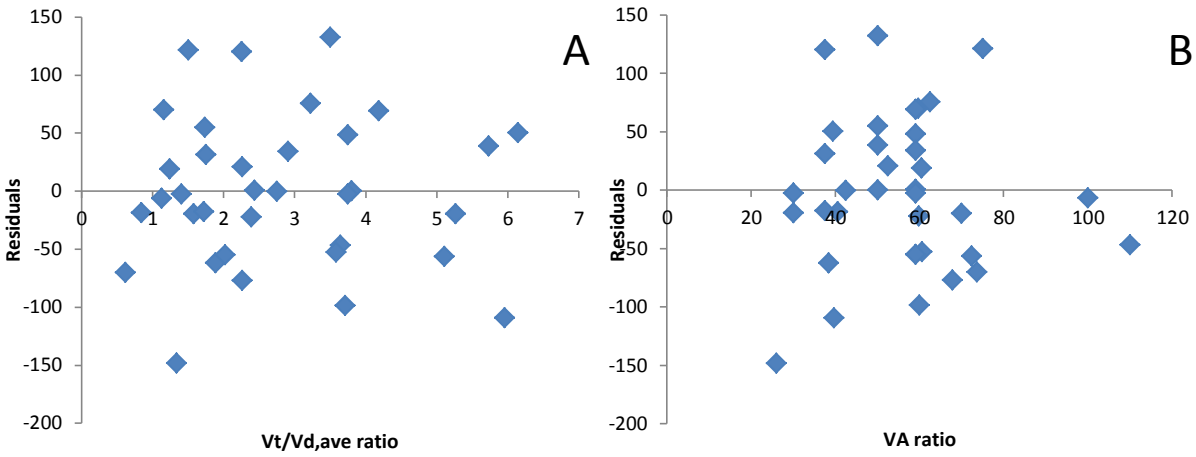

Figure 4.5: Residuals plotted as a function of a) $V_{tank}/V_{d,ave}$; and b) *VA ratio*

The even spread of residuals for each of the variables in the simplified regression model suggests that the model is adequate.

4.7.3.2 Multicollinearity

Multicollinearity is a potential issue for regression models. Multicollinearity arises when two or more of the independent variables are correlated and are therefore not truly independent. The Pearson correlation matrix (Table 4.3) shows that there are significant correlations between variables in the unrestricted regression model above, significant is taken to mean correlations of >0.7 and <-0.7 (Dancey & Reidy 2004) cited in (Sokunbi 2014).

The variables shown in bold type are those in the simplified regression model; the correlation between these two variables is negligible being only 0.063 suggesting that multicollinearity is not an issue in the simplified model.

	S	$V_{d,ave}$	ΔT	Q_{aux}	**VA ratio**	**$V_{tank}/V_{d,ave}$**	$V_{d,ave}$ /A
S	**1.000**	-0.277	0.070	-0.160	0.000	0.077	-0.355
$V_{d,ave}$	-0.277	**1.000**	0.202	0.751	0.279	-0.706	0.894
ΔT	0.070	0.202	**1.000**	0.340	0.047	-0.350	0.173
Q_{aux}	-0.160	0.751	0.340	**1.000**	0.451	-0.458	0.714
VA ratio	0.000	0.279	0.047	0.451	**1.000**	0.063	0.457
$V_{tank}/V_{d,ave}$	0.077	-0.706	-0.350	-0.458	0.063	**1.000**	-0.622
$V_{d,ave}$ /A	-0.355	0.894	0.173	0.714	0.457	-0.622	**1.000**

Table 4.3: Pearson correlation matrix for independent variables in the unrestricted regression model

A further check for multicollinearity involves comparing the multiple-correlation (R) of the regression model with the correlations between independent variables shown in Table 4.3. If the R-value is greater than the correlation between independent variables then multicollinearity is not an issue; the multiple correlation of a regression equation is given by the square root of the coefficient of determination of that model:

$$R = \sqrt{R^2} \tag{4.54}$$

The multiple-correlation in the simplified model is 0.69, which is greater than the correlation between $V_{tank}/V_{d,ave}$ and *VA ratio* (0.063); therefore multicollinearity is not an apparent issue.

4.7.4 Making predictions with the regression model

The regression model was constructed using annual data from 35 systems throughout the UK. It can be used to make predictions about the long term annual performance of a solar thermal system in terms of specific heat yield. The data input requirement for the model to make predictions is low requiring only three items: collector area; tank volume; and annual average DHW demand in litres.

The validity of this model is tested using the Retrofit for the Future case study dwelling in Bedford. The annual average DHW demand for this property of two occupants is 51.98L/day; the tank volume is 350L and the collector area is 6.3m^2:

$$Q_{sol,p} = \left[3 \times \left(\frac{350}{6.3}\right)\right] + \left[-23.13 \times \left(\frac{350}{51.98}\right)\right] + 184.6 \quad (4.55)$$

$Q_{sol,p}$ 195.52kWh/m^2/yr
$Q_{sol,m}$ 100.63kWh/m^2/yr

The regression model predicts the specific solar yield to be 195.52kWh/m^2/yr; however the measured specific yield was 100.63kWh/m^2/yr, which means that the model over-predicted the yield by 94.3%.

There are two reasons for this large discrepancy between the predicted and measured Q_{sol} for Bedford: Firstly, the sample size used to construct the model is small and is not likely to be representative of all the systems installed in the UK. Secondly, the regression model has only two variables which explain 47.62% of the variation in Q_{sol}; there is a further 52.38% of the variation that goes unexplained by the variables in the model. This suggests that there are other variables that have not been measured that may improve the estimate of system performance.

One factor that is not explicitly included in the model is the time of use of the water, which is suggested to affect system performance (Bates et al. 1999; Buckles & Klein 1980; Knudsen 2002; Lundh et al. n.d.; Morrison et al. 1992; Jordan & Vajen 2000; EST 2011). However this parameter is difficult to include in regression models as an explicit descriptor of performance due to the temporal nature of the variable. However an advantage of regression models in this respect is their ability to implicitly incorporate patterns of use provided that the sample is representative. For example, a larger sample size will incorporate a larger variety of usage patterns within the DHW demand variable and the model will explain more of the variation in Q_{sol} leading to better predictions. Different system configurations and types will also affect performance but these are qualitative variables and cannot be easily incorporated into a single regression model unless the sample is representative of the many types of system in existence.

For a deeper understanding of the discrepancy between the model prediction and measured performance of Bedford, a more detailed analysis is required. Further analysis shows that the household has poor control of the auxiliary heating system with an average daily auxiliary input of 8.68kWh.

Figure 4.6 shows that the ASHP is inputting heat throughout the entire day; this limits the amount of heat that can be delivered by the solar thermal system, especially when the peak auxiliary input coincides with the beginning of the peak solar irradiation. Although the backward

elimination procedure removed auxiliary energy input from the regression model because it did not explain a large proportion of the variation in Q_{sol}, in extreme cases auxiliary input explains why the measured performance is so different from predicted.

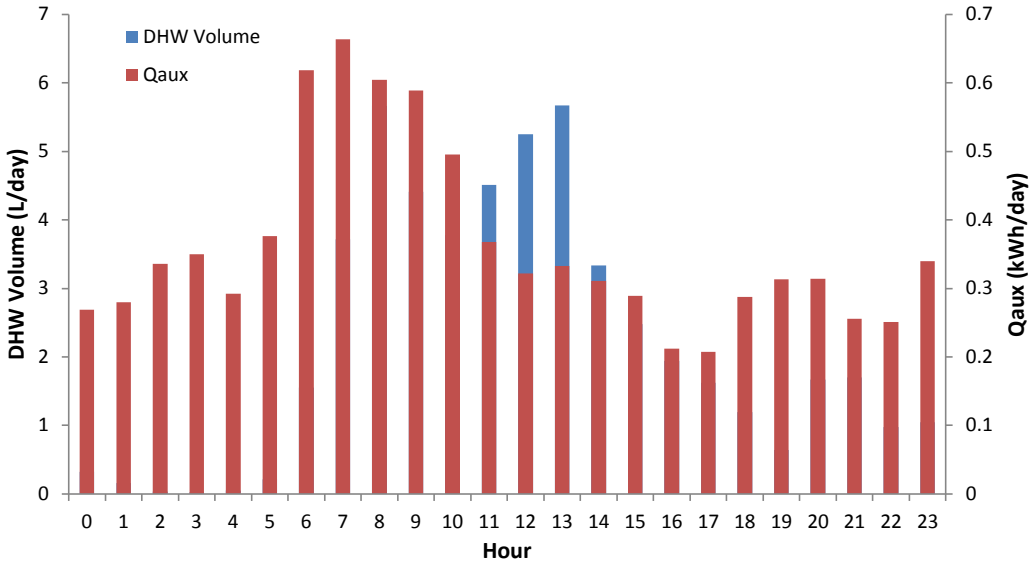

Figure 4.6: Average daily DHW draw profile and auxiliary timing for Bedford – Q_{sol} = 100.63kWh/m^2/yr

Nevertheless there is some information that can be taken from the prediction of Q_{sol} at Bedford: the system is clearly oversized for the amount of demand. The collector area and tank size seem to be well matched in terms of their sizing indicated by the VA ratio of 55.56L/m^2, which would suggest a nominal specific yield of ~250kWh/m^2; however the tank size is too large for the DHW demand, with a ratio of 6.73, which would indicate an annual yield of less than 200kWh/m^2/yr. Results from the EST field trial data suggest that lower tank size-to-demand ratios and larger tank-to-collector area ratios are best for long term performance; Knudsen states that for DHW volumes of less than 200L/day, tank volumes of 100-175L are better (Knudsen 2002). Larger tanks-to-collector ratios ensure that there is enough capacity for the collector to input heat, this requires suitably large DHW demands to maintain a cool tank for the solar thermal system to deliver heat into, which Bedford does not.

4.8 The effect of control and timing on solar thermal system performance

The regression equation highlights that there is a substantial amount of variation yet to be explained; however many of the original variables have been eliminated from the model because the restricted model explains a comparable amount of the variation. The result of the unexplained variation in Q_{sol} is that some systems of similar size and DHW demand have different annual yields. To offer an explanation for this it is worth considering that the control of the solar thermal system can affect performance (Forward & Roberts 2008; EST 2011).

To investigate this further, comparable systems from the EST field trial were selected for in depth comparison of their DHW profile and auxiliary timing. An average daily profile of DHW usage and auxiliary timing was constructed to highlight any differences between the control of similar systems and their subsequent performance.

Table 4.4 shows the systems under comparison; the systems are studied as pairs with similar size characteristics but different annual yields. From Table 4.4 alone it can be observed that in general the better performing systems of a given size use on average less auxiliary energy, more DHW, or both. This is the case for 4 of the 5 pairs. The pair that stands out is 560WWH/569GRI; the auxiliary usage for both of these properties is comparable as is the DHW volume. However 560WWH produces 142.16kWh/m^2/yr more solar yield than 569GRI; therefore there is another reason for the stark contrast in performance between the two systems.

System	Qsol	Vtank/Vd,ave	VA ratio	Qaux	Vd,ave
536MEA	385.74	3.50	50.00	1.00	85.68
535GAL	246.47	3.80	50.00	2.25	78.88
534CRO	210.84	6.14	39.37	1.02	48.84
565MAC	56.59	5.95	39.68	2.34	41.99
559RDE	406.68	1.15	59.66	6.44	182.04
570GWY	355.97	1.24	60.42	7.39	116.85
560WWH	372.82	3.22	62.50	1.73	68.35
569GRI	230.66	3.59	60.61	1.66	55.78
531KIL	240.64	5.73	50.00	1.23	52.37
Bedford	100.63	6.35	52.38	8.68	51.98

Table 4.4: Systems for comparison; there are 5 pairs of systems separated by a black line, the grey row shows the better performing system of the pair

560WWH (Figure 4.7) shows a predominant usage of DHW in the morning with a peak at around the 08:00, the main auxiliary input occurs during the evening at around 16:00 ready for the smaller, evening peak draw. This large DHW draw in the morning recharges the tank with cool water for the peak period of solar thermal generation. The peak auxiliary input in the evening is enough to top up the heat of the water for the peak evening usage and may allow the tank to store enough energy through the night for sufficiently hot water in the morning.

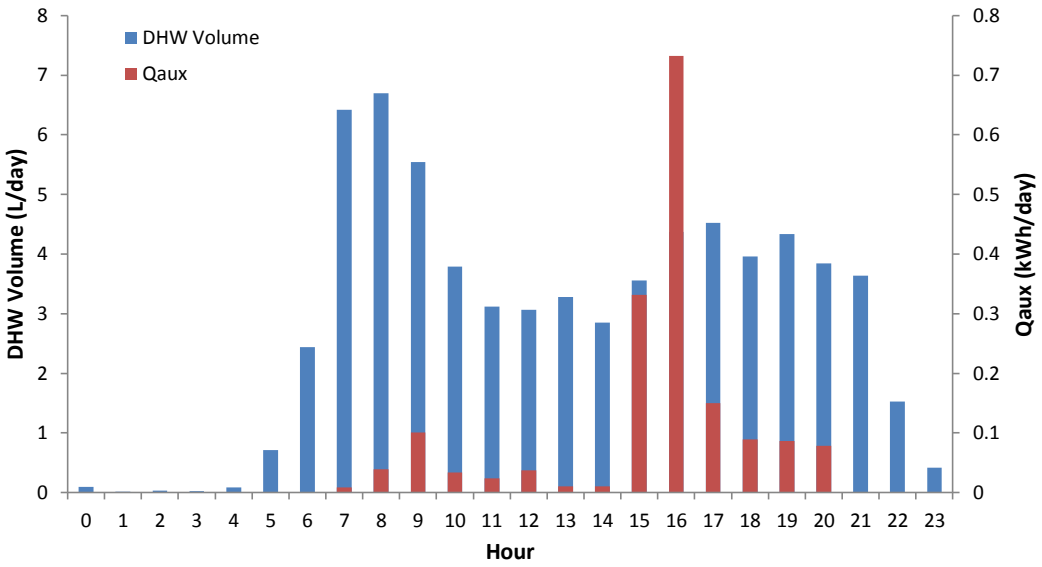

Figure 4.7: Average daily DHW draw profile and auxiliary timing for 560WWH – Q_{sol} = 372.82kWh/m^2/yr

569GRI (Figure 4.8) shows a different control strategy: DHW draw is more uniform throughout the day, although three peaks can be distinguished at 06:00, 12:00 and 17:00. The volume of

water used by 560WWH and 569GRI from 04:00-09:00 is 21.9L and 19.7L respectively; however 569GRI inputs more auxiliary heat during this time than 560WWH which increases the temperature of the tank and reduces the amount of heat that can be delivered by the solar thermal system. The uniform auxiliary input throughout the day, especially the afternoon, also reduces the input of the solar thermal system as the collector and the boiler are effectively competing (Forward & Roberts 2008). Therefore the difference in performance is due to the way in which the two systems are controlled

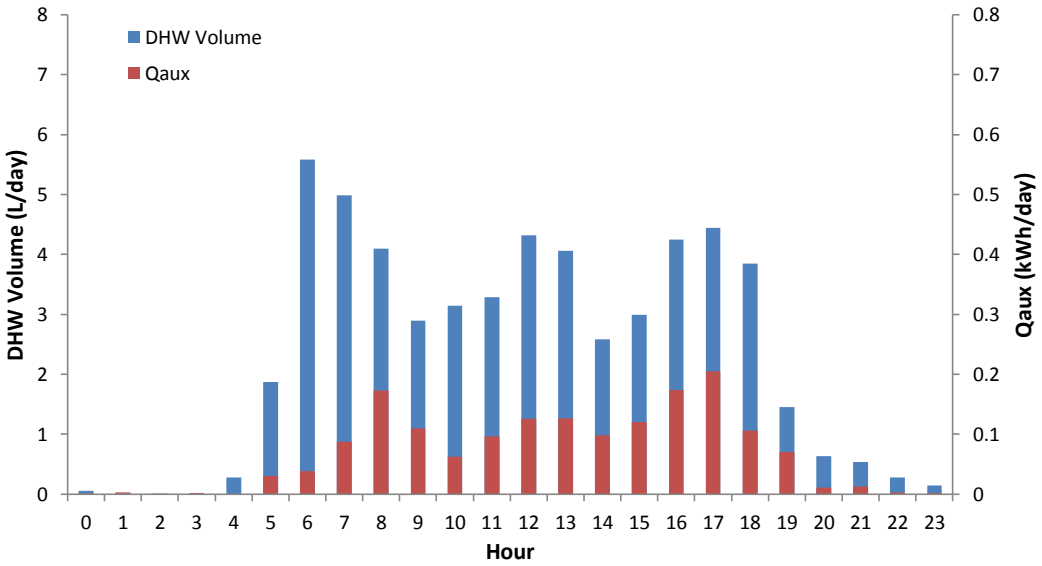

Figure 4.8: Average daily DHW draw profile and auxiliary timing for 569GRI – Q_{sol} = 230.66kWh/m^2/yr

The regression analysis suggests three ways to maximise solar yield:

- Maximise VA ratio to allow the collector to input maximum amounts of energy;
- Minimise $V_{tank}/V_{d,ave}$ to ensure the tank is always charged with cooler water;
- Minimise auxiliary input especially during the morning and afternoon possibly by employing weather predictive control such as that described in (Liao & Dexter 2010);

4.9 Strengths and weaknesses of regression models

Regression models provide one of the simplest methods by which to make predictions of a dependent variable; however the raw data requirements required to produce a model that adequately fits the data are large. A key drawback of regression is that for models to be suitable for making predictions about the performance of solar thermal systems the sample needs to be representative of the population. As shown above, a sample of 35 systems is not enough to represent the different system configurations and occupancy usage behaviour thus limiting the accuracy of the model when predicting annual performance of systems.

Despite this regression models do allow patterns and trends to be discovered such as the relationships between system size (in terms of tank volume-to-collector area and tank volume-to-DHW demand) and Q_{sol}. In addition regression analysis reveals which variables are the least and most influential on system performance; for example, it has been shown that the annual incident irradiation explains very little of the variation in Q_{sol}. A conclusion that can be drawn

from the previous two points is that system size seems to be the primary concern for determining the performance of solar thermal systems and the system should be designed in such a way that the VA ratio is as large as possible whilst minimising the tank volume-to-DHW demand ratio; this is an example of how regression analysis can be used to inform design decisions.

Although the data requirements to produce a statistically representative regression model are high, the data inputs required to make predictions using the model are low amounting to three items of data in the model above: collector area; tank size; and DHW demand. This is an advantage of regression over dynamic simulations and even simplified physics based models. However the data pertaining to DHW demand may not readily obtained and would require estimations based on regression relationships such as those used in BREDEM and SAP. As shown previously there is a discrepancy between estimated and actual DHW demand, which limits the confidence that can be had in predictions made using these estimates.

Assuming that a sufficiently large sample of data can be obtained for each of the variables used in the regression model, it can implicitly include the effects of DHW draw profiles in predictions. This overcomes the issue of using standard draw patterns in dynamic simulations, which vary widely between households and throughout the year (Swan & Ugursal 2009). However the downside to implicit inclusion of draw patterns is that their effects on performance cannot be extracted with ease. This becomes an issue when there is a want or need to determine the effects of draw patterns on performance.

A general limitation to regression models is that the correlation between variables does not necessarily explain the causes behind the relationship (Fenton & Neil 2011; Fenton & Neil 2007; Fenton et al. 2002; Armstrong 2012). For example a negative correlation between daily auxiliary input and daily solar heat yield does not give any indication as to whether the solar heat yield is low because there is a large auxiliary input heating the tank, or whether the auxiliary input is high because there is very little input from the solar system. This relationship is more complex and depends on other factors such as time of year (winter leading to greater auxiliary inputs and lower solar yield), occupancy usage behaviour and control, for example, whether the occupants understand that auxiliary input can limit the input from the solar thermal system if it is mistimed for instance.

In summary, regression models provide simple to use equations that are useful for making predictions and design decisions about solar thermal systems, whilst at the same time indicating which factors affect performance the most. However their predictive power is limited by their ability to represent the population of installed systems due to insufficient data, which also leads to a large amount of the variation in the dependent variable remaining unexplained – regression models do not accurately quantify the uncertainty in a system domain (Booth & Choudhary 2011).

4.10 Dynamic simulation

Dynamic simulations of solar thermal systems are performed in specialist software such as TRNSYS (University of Wisconsin 2013), WATSUN (National Resources Canada 2013) and T*SOL (Valentin Software 2015). Of these TRNSYS is commonly used (German Solar Energy Society 2010) and reference to dynamic simulations in this chapter will generally refer to those conducted in TRNSYS.

Dynamic simulations of solar thermal systems are performed in order to facilitate economic analyses of such systems based on expected thermal yield and reductions in auxiliary energy consumption (Duffie et al. 1988). They are also used in design applications (Hobbi & Siddiqui 2009), and as to produce datasets for simplified models (Klein et al. 1976; Murphy et al. 2011).

4.11 Background to TRNSYS

TRNSYS (TRaNsient SYStems Simulation Program) was developed in the early 1970's by the Solar Energy Laboratory at the University of Wisconsin (Klein et al. 2010); it became commercially available in 1975. TRNSYS allows the user to assemble subroutines that represent system components using a graphical interface (Figure 4.9).

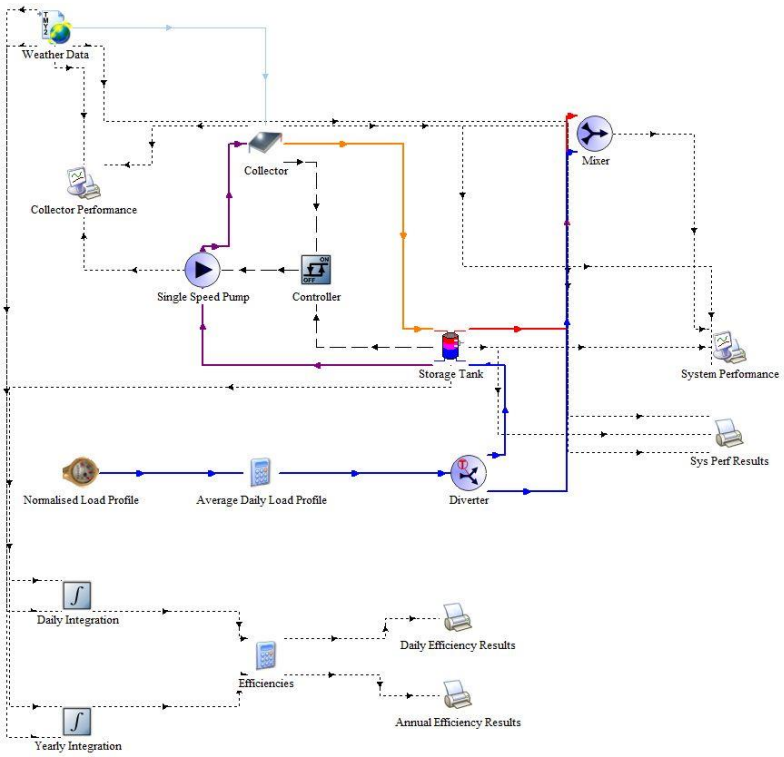

Figure 4.9: Screenshot of a solar thermal hot water system modelled in TRNSYS

The components of a TRNSYS model are called "types". TRNSYS solves the equations in the system model numerically at user specified time steps. The equations used in TRNSYS are based on first principles using climatic input data to drive the simulation. Outputs of the calculations give the temperatures, energy flows, control functions and anything else that occurs in the system during operation, and can be exported to spreadsheet format for easy analysis (Duffie et al. 1988).

One of the advantages of TRNSYS is the ability to customise existing components or construct new components that better describe the system in question; to aid in this there is a strong, global network of TRNSYS users who can communicate issues and solutions using the forum (Solar Energy Laboratory 2014).

4.12 Principles of dynamic simulations

Dynamic simulations solve differential equations based on principles of thermodynamics and heat transfer. Simulations can contain multi-zone building models along with heating system components and so are suitable for evaluating complex interactions between spaces and components including ventilation, and active and passive heating/cooling (German Solar Energy Society 2010).

4.12.1 Validation and calibration

One of the advantages of dynamic simulations over laboratory based experiments is their ability to quickly perform long term simulations of solar thermal systems; a year's worth of performance data can be obtained in a fraction of the time it would take a monitored system to produce (i.e. one year). It also avoids other complications associated with monitoring systems such as cost and reliable data quality over extended periods; however dynamic models used for this purpose need to reflect the system being evaluated as accurately as possible. In order for researchers to be confident that their simulations reflect the real situation, a validation procedure is performed. Early validation studies of TRNSYS models include those by Mitchel et al (1978) and Duong & Winn (1977) cited in (Haberl 2004) and show that the predicted collector output in was within 5% of measured data; therefore valid predictions of solar thermal system performance can be made using TRNSYS (Yohanis et al. 2006). Further validation tests using experimental data have been performed by subsequent researchers (Ayompe, Duffy, McCormack, et al. 2011; Jordan & Vajen 2000; Spur et al. 2006; Wolf et al. 1984; Taherian et al. 2011; Cadafalch 2009; Knudsen 2002). Similarly, simple physics based models such as the *f*-chart method may be validated against experimental data or against validated dynamic simulations (Klein 1976) cited in (Haberl 2004). Once a model has been validated it may be used to produce long term predictions of system performance in different locations and climatic conditions before real systems are installed. In addition they can be used as a design aid to allow the optimum size of system components to be determined for different load profiles (Ayompe, Duffy, McCormack, et al. 2011; Rodríguez-Hidalgo et al. 2012).

Models used to predict performance of systems may present results that are different from measured data; this difference is called here the "performance gap" and has been discussed in Chapter 3. Models struggle to represent the actual in-use installation/build quality and other peculiarities associated with the way in which systems are operated leading to a mismatch between modelled and measured performance. Model parameters can be modified so that the outputs match well with the monitored data; this process is called calibration. Calibration is performed in order to improve the predictive ability of the model. It is a common procedure in the building sector and considered an important step in building energy performance modelling (Ahmad & Culp 2006; Heo et al. 2012; Yoon et al. 2003). It is achieved by site visits; interviews; and taking field measurements of physical properties, and occupancy and usage patterns (Heo et al. 2012). Calibration of a solar thermal system has been performed by Degelman (2008) to

enable an evaluation of the degradation in performance of a system over 22 years; base measurements were collected in 1982 and used to validate a contemporary model. Measurements were repeated in 2004 and used to adjust the collector performance parameters in the model. The difference between 1982 and 2004 efficiency curves obtained using the calibrated performance parameters represented the degradation in performance over time (Degelman 2008).

4.13 Strengths and weaknesses of dynamic simulations

Swan and Ugursal (2009) provide a detailed description of the strengths and weaknesses of engineering models used to predict building energy performance; many of these strengths and weaknesses apply to dynamic simulation of solar thermal systems (Swan & Ugursal 2009).

Dynamic simulations do not require empirical data in order to operate giving them a distinct advantage over simplified models based on statistical relations obtained from laboratory experiments or field trials. The advantages related to low data requirements are:

- No costly, time consuming monitoring schemes or experiments;
- Do not require large samples of systems to be monitored to provide validation or show statistical significance – validation of the model against data from a relatively short monitoring period is sufficient;
- Simplified models should be statistically representative of all the system types in use in the population, this is not a requirement of dynamic simulations which can model any system configuration from first principles;
- Able to model new technologies and system designs that have no prior historical data (Swan & Ugursal 2009)

However dynamic simulations cannot reflect the peculiarities in operation of systems and components without calibration procedures. Such peculiarities can arise from manufacturing inconsistencies, installation quality and occupancy usage behaviour. Simplified models may implicitly include such phenomena in the empirical relationships provided the sample is statistically representative; however this may not explicitly explain the variation/uncertainty in performance estimates that arise from such phenomena unless regression equations contain associated variables.

Although dynamic simulations have very low data requirements in terms of their development and validation, the input data requirements needed for each simulation are higher than for simplified models (Zhao & Magoulès 2012). Users must input very detailed information about the system being modelled including collector performance parameters, thermal storage, auxiliary input, weather data, possibly building envelope and occupancy usage patterns (timing and energy demand). Since dynamic simulations are often used in the design stage, when much of this information is unknown or uncertain (Domínguez-Muñoz et al. 2012), the simulations must rely on assumptions made by the modeller. The accuracy of these assumptions is dependent on the experience of the designer, although standards exist to guide modellers when it comes to occupancy usage patterns; however as discussed in section 4.2 these have their own limitations.

Dynamic simulation software has steep learning curves requiring expertise to use them effectively. The users are required to make informed assumptions about the system parameters as well as the system configuration and how it will be modelled in the software. They must also be able to deal with complications in the simulation such as convergence errors that may develop. These issues do not arise in the use of simplified models because the assumptions about occupancy usage for example are contained within the regression equations developed from empirical data. The simplified model user is directed stepwise through the calculation procedure and encouraged to use real data where possible.

In the past the computational intensity of dynamic simulations was a concern; these days modern computing has made this less of a problem. However running many hundreds of simulations for the purpose of including uncertainty in the outputs (Monte Carlo methods) may extend the processing time considerably (Domínguez-Muñoz et al. 2012).

In general dynamic simulations are flexible and a cheaper and more convenient method to laboratory experiments or field trials; however they require skill and expertise to produce accurate results and validation/calibration processes to provide confidence in the results. In addition they are unable to accurately represent occupancy usage behaviour, which is highly uncertain and can reduce the accuracy of the simulation, though it is possible to introduce uncertainty into dynamic simulations using Monte Carlo methods.

4.14 Using TRNSYS to demonstrate the effects of DHW profile on the performance of solar thermal systems

Dynamic modelling approaches to predicting the performance of solar thermal systems can readily be criticised due to the simplified assumptions made about DHW usage which in reality varies greatly from household to household and is dependent on many complicating factors (Knudsen 2002; Michaelides 1993; Evarts & Swan 2013). However, it is ironic that the limitations of dynamic simulations in this respect may be one of its strengths when exploratory modelling exercises are needed. The necessity for the modeller to specify DHW usage profiles means that the effect of DHW volumetric consumption, flow rate, duration of draw and time of use can be easily investigated whilst keeping the remaining model inputs constant. These kinds of investigations are difficult to achieve with real data sets such as those provided in the EST solar thermal field trial due to the number of variables associated with real life solar thermal systems which firstly need to be measured and secondly vary greatly from day to day and system to system. Larger data sets with several systems of similar size and configuration and exposed to similar weather conditions would go some way to enabling investigative analyses to be performed using real data; however real life DHW usage still remains variable and random in nature. In order to have control over system parameters including weather conditions and DHW profile laboratory experiments can be performed. Unfortunately the length of the experiment is determined by the time period over which the analysis is required making long term performance analysis a time consuming endeavour. In addition in order to control the weather conditions access to an environmental chamber is necessary. Therefore modelling, with all its drawbacks, provides the most practical means by which to investigate the effects of DHW profiles on the performance of solar thermal systems.

The DHW profile is determined by the volumetric consumption, the duration of the draw and the time at which it is drawn. The timing of hot water consumption has been shown to affect the performance of solar thermal systems: Investigations made by (Buckles & Klein 1980) and (Knudsen 2002) indicate that the mid-afternoon (defined in these studies as 14:00-18:00) is the best time to draw DHW providing the highest solar fraction and solar yield; drawing DHW in the early morning (02:00-06:00) is shown to provide the worst solar fraction due longer storage periods leading to higher losses and increased auxiliary input. In contrast, results from (Jordan & Vajen 2000) indicate that the early afternoon (between 12:00 and 16:00) is the best time to draw hot water with fractional energy savings at this time being greater than both morning (07:00) and evening (19:00) draws. In the studies conducted by (Buckles & Klein 1980), (Knudsen 2002) and (Jordan & Vajen 2000) no solid reason as to why midday draws are superior in terms of solar yield is given however realistic draws developed by (Jordan & Vajen 2000; Jordan & Vajen 2001) are shown to have lower solar yields than recurring draws due to a 14 day period in August of no DHW use; (Buckles & Klein 1980) also found that a non-recurring draw profile led to a reduction in performance.

The DHW profile has been linked to the thermal performance of the storage tank, in particular the level of stratification the tank exhibits when different draw profiles exist (Jordan & Vajen 2000; Jordan & Furbo 2005). An increase in thermal stratification has been shown to increase the thermal performance of the solar thermal system (Jordan & Furbo 2005; Knudsen 2002; Andersen & Furbo 1999 cited in (Jordan & Vajen 2000)). These studies all demonstrate that DHW profile is an influential parameter in the performance of solar thermal systems. To corroborate the findings from these studies a comprehensive simulation strategy has been developed in which the effect of DHW volume, time of draw and initial thermal conditions of the storage tank on the solar thermal performance is presented. The objectives of the study are to offer reasons behind the differences in thermal performance due to DHW profile; quantify the level of influence DHW usage parameters have on short term and long term performance of solar thermal systems; establish causal relationships between system parameters. Information from the simulations can therefore be used to suggest usage patterns that would lead to maximum thermal performance and quantify the uncertainty in the solar thermal system performance owing to differences in DHW usage behaviour.

4.14.1 System description

A simple solar thermal system is modelled in TRNSYS and is based on a standard configuration as described in (German Solar Energy Society 2010) and shown in Figure 4.10. This type of system is the most common system monitored in the EST solar thermal field trial with 71 of the 88 systems being of the standard configuration.

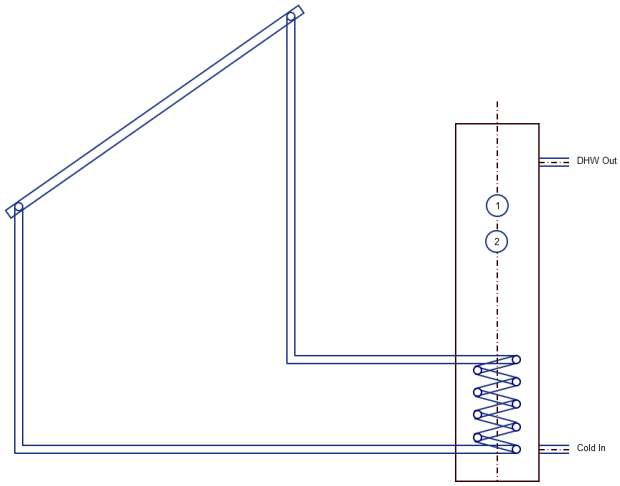

Figure 4.10: Diagram of a standard solar thermal system

The TRSNYS model schematic is shown in Figure 4.11; the flat plate collector is modelled as a Type73 with an area of 6m^2, south facing orientation and an inclination of 35°. Solar thermal heat is delivered into the storage tank by way of a solar coil with an inlet and outlet 700mm and 200mm from the bottom respectively. The tank is loosely based on the Ecocat 300 (Chelmer Heating Solutions 2013) modelled using a Type60; the tank has a volume of 350L and a height of 2000mm. Cold water enters the tank 200mm from the bottom at a temperature of 12°C; hot water exits the tank 1800mm from the bottom. The tank is split into 8 nodes of equal size to model the effects of stratification. There are two electric auxiliary heaters located 1400mm and 1600mm from the bottom of the tank to simulate a heating element of 200mm in length; these switch on simultaneously and are controlled using a timer and thermostat and will activate at a specified time of day if the temperature in the top of the tank is less than 60°C. If the temperature of the fluid exiting the solar coil is more than 5°C warmer than the temperature exiting the solar collector then the Type2 controller deactivates the Type114 pump. The weather data used is a Meteonorm file for London.

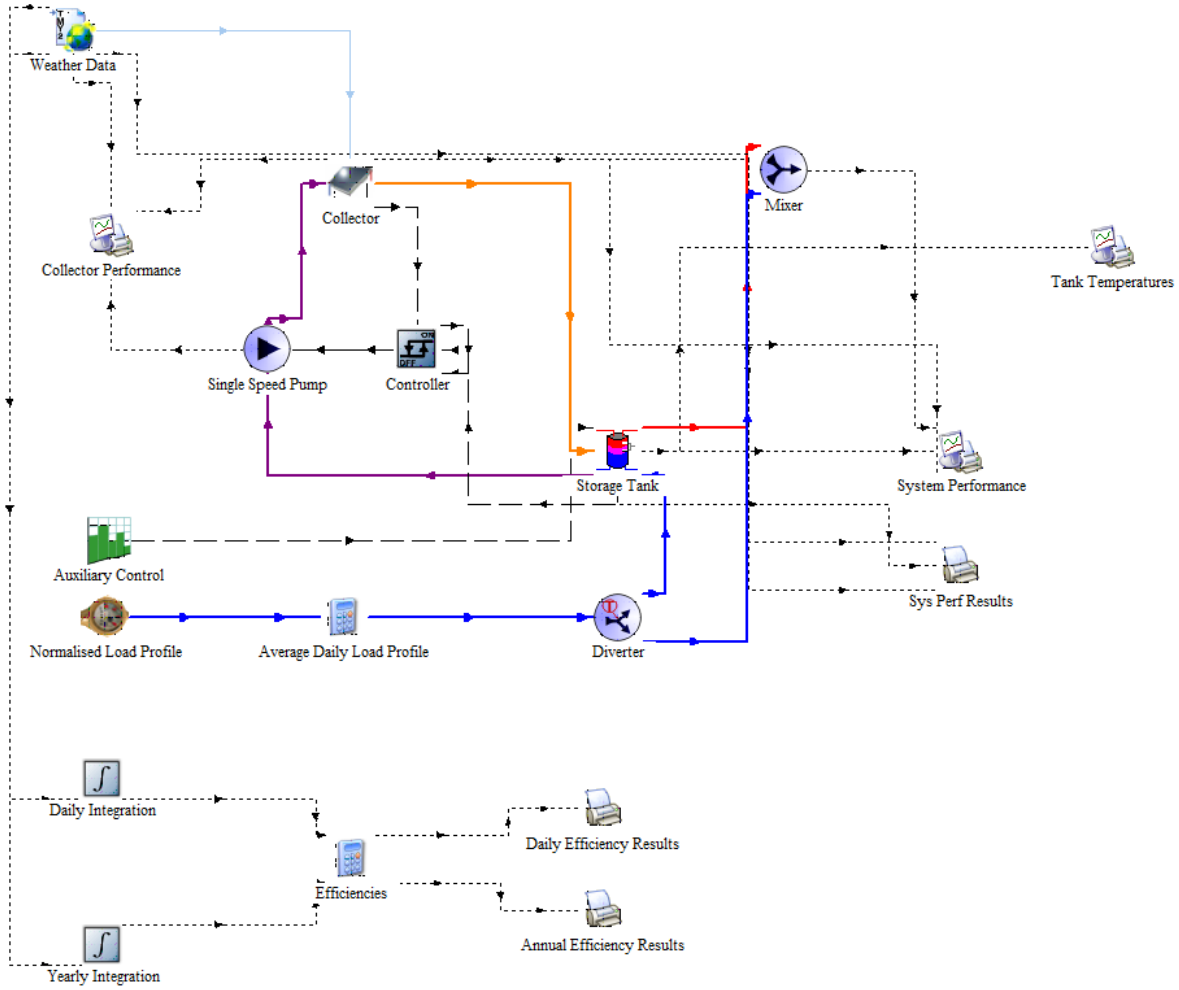

Figure 4.11: TRNSYS model for standard solar thermal system

A normalised DHW draw profile is specified using the Type14 forcing function which is multiplied by the DHW volume specified in the calculator labelled "Average Daily Load Profile". As hot water is drawn out of the top of the tank cold water feeds into the bottom of the tank at the same flow rate determined by the length of draw (specified in the Type14 forcing function) and the volume. Another Type14 forcing function is used to control the timing of the auxiliary heaters in the tank, these will switch on at the specified time if the temperature of the water is less than 60°C.

Simulations were performed for a single summer's day and a two-day period during the summer in which the second day had the same weather conditions as the single day simulations; these simulations were performed at a resolution of 1 minute. In addition annual simulations were performed at a resolution of 1 hour.

4.14.2 DHW profile

Six DHW consumption volumes were used in the simulations given as a proportion of the total tank volume: $0.25V_t$; $0.5V_t$; $0.75V_t$; $1V_t$; $1.25V_t$; $1.5V_t$. Each volume was drawn over a period of 1 hour at four different times of the day to give four different draw profiles shown in Figure 4.12.

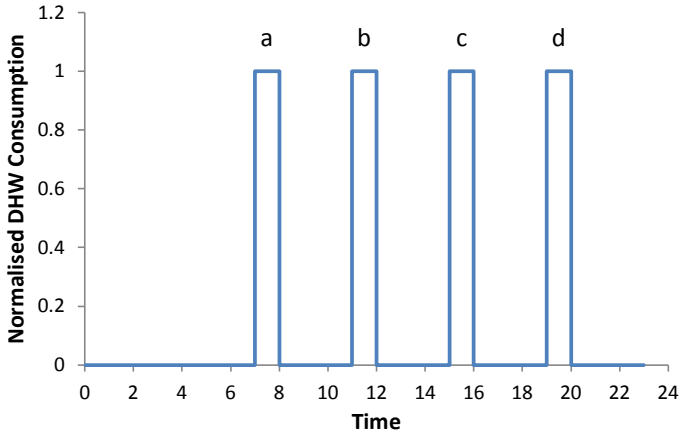

Figure 4.12: Four DHW profiles: a) morning; b) early afternoon; c) late afternoon; d) evening

4.14.3 Auxiliary control

Large draw volumes and wintery weather conditions occurring throughout the annual simulations make auxiliary heating a necessity to ensure that the DHW is drawn at a useable temperature. A 10kW auxiliary heater is situated in the second and third temperature nodes to ensure that water in the top portion of the tank is between 55-60°C.

The auxiliary heater is active between 06:00-20:00 and will activate if the temperature in the top part of the tank falls below 55°C; it will remain active until the temperature reaches 60°C. In all simulations the initial conditions of the tank are 12°C, fully mixed (i.e. a cold tank) – fully mixed means that the tank is a uniform temperature throughout.

Table 4.5 details the variables and constants used in the model.

Component	
Solar Thermal System	
Type	Type73 Flat plate collector
Fin efficiency	0.7
Edge loss coefficient	0.8333W/m^2.K
Absorber plate emittance	0.1
Absorptance	0.97
Index of refraction of cover	1.526
Extinction coefficient	0.0026
Area	1x6m^2
Inlet flow rate	240kg/hr
Inclination	35°
Orientation	South
Ground reflectance	0.2

Combined Store	
Type	Type60 Detailed fluid storage tank
Volume	350L
Height	2000mm
Diameter	472mm
Storage loss coefficient	0.57W/m²K
Cold inlet temp	12°C
Starting tank temperature	Fully mixed 65°C; fully mixed 55°C; fully mixed 45°C; fully mixed 12°C; stratified
Usage Patterns	
DHW draw volume	87.5L/day; 175L/day; 262.5L/day; 350L/day; 437.5L/day; 525L/day
DHW draw profile	Single draw 07:00-08:00; single draw 11:00-12:00; single draw 15:00-16:00; single draw 19:00-20:00
Auxiliary timing	No auxiliary; 10kW auxiliary 06:00-20:00

Table 4.5: Table of variables and constants used in the simulations

4.14.4 Single day simulation results

4.14.4.1 The effect of varying the initial temperature conditions of the storage tank

These simulations are performed for a single summer's day with no auxiliary heat delivered to the storage tank. The storage tank has four different initial temperature conditions: 65°C fully mixed; 55°C fully mixed; 45°C fully mixed; 12°C fully mixed. Fully mixed describes a tank in which all nodes are of the same temperature meaning that the tank is a uniform temperature from top to bottom.

Figure 4.13 shows that the daily solar yield (Q_{sol}) increases with increasing DHW consumption volume for all draw profiles except the evening draw. The evening draw takes place after the operating period of the solar thermal system has come to an end, thus the solar yield remains constant regardless of how much water is drawn after this period. For the remaining profiles increased draw volumes reduce the temperature of the bottom of the tank at periods prior to and during high solar radiation. This improves the rate of heat transfer between the solar coil and surrounding storage water due to a greater temperature difference between the fluid in the coil and surroundings. Furthermore, cooling the temperature in the bottom of the tank increases the temperature difference between the collector fluid and the fluid returning from the solar coil, which will extend the period of operation of the collector provided the temperature difference is greater than 5°C. Signs of a plateau beginning to form are visible at large draw volumes; this is expected since solar yield is dependent on tank temperature. Once a critical draw volume has been reached then increasing the amount of water drawn will not reduce the tank temperature any further and solar yield values will remain constant.

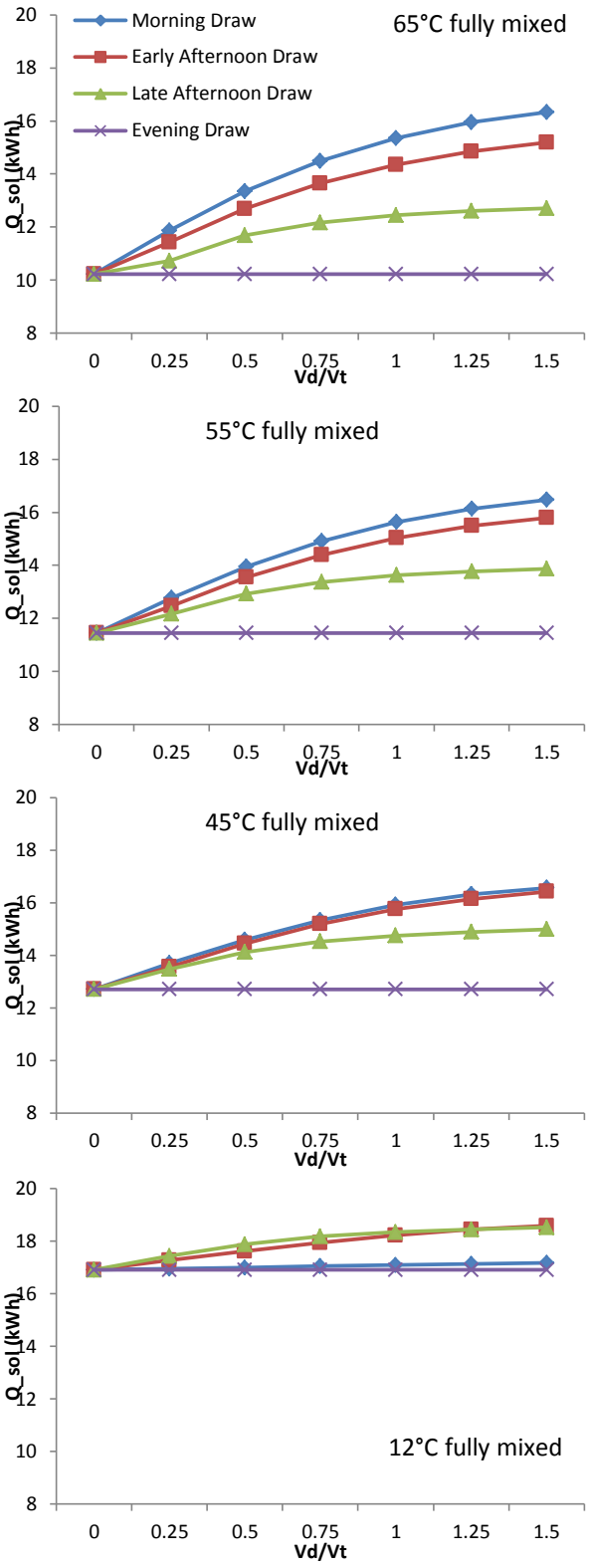

Figure 4.13: Daily solar yield versus DHW draw volume for four draw profiles and four initial tank conditions

The timing of the draw is shown to have an effect on the daily solar yield collected by the solar system with certain times of day providing more solar yield than others; however which profile provides maximum solar yields depends on the initial temperature conditions of the storage tank. Tanks that start the day hot favour the early draw profiles due to the influx of cold water into the bottom of the tank prior to peak operation of the solar collector; this promotes longer

operating periods due to a greater temperature difference between the collector fluid and the return temperature to the collector, and increased heat transfer rates due to greater temperature differences between the solar coil and the surrounding storage water. The hotter tank conditions show a greater difference between the best and worst draw profiles (morning and evening respectively) in terms of daily solar yield: daily solar yield is a maximum of 59.7% greater for the morning profile versus the evening profile when the tank is 65°C compared to 30.5% when the tank starts out at 45°C. The cold tank condition favours midday draw profiles: late afternoon draws extend the period of operation of the system at a time when solar radiation is dwindling; reducing the tank temperature at this time of day increases the amount of time at which the temperature difference between the collector and return fluid is greater than 5°C. Morning and evening profiles have high tank temperatures at this time and so the solar collector is inactive. An early afternoon draw reduces the tank temperature prior to peak solar radiation therefore maximising the heat transfer rate between the solar coil and surrounding storage water. Cold starting conditions give morning and evening a similar performance in terms of solar yield; when the tank is cold there is little difference between the temperature of the tank before and after a morning draw and so a similar tank condition is experienced for both morning and evening draws. The two profiles involve a draw either side of the most active period for the solar thermal system therefore the Q_{sol} values are very similar; however the evening draw makes use of the solar thermal heat and water draw temperatures are of a useable value, in contrast the morning profile results in cold water being drawn and so is not useful to the occupant. These observations are only relevant for a single day simulation where the thermal heat stored at the end of the day does not impact the performance of the system on the next day.

Figure 4.14 shows the solar yield obtained for the four draw profiles starting with three different tank temperature conditions for a constant DHW draw volume of 175L/day.

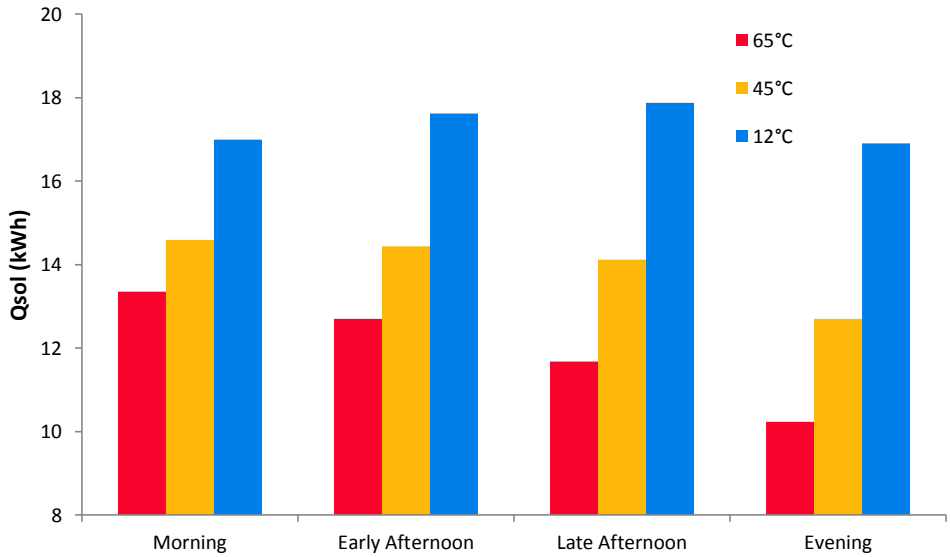

Figure 4.14: Daily solar yield versus DHW draw profile for initial tank temperature 65°C, 45°C and 12°C and draw volume 0.5V_t (175L/day)

It is seen that for all draw profiles the solar yield increases with decreasing initial tank temperature; the highest increase is seen for an evening draw when an improvement of 65.2%

can be obtained by starting with a cold tank. It is also possible to see that drawing water in the middle of the day increases the yield by a maximum of 5.8% when the tank is initially cold. This is in contrast to beginning the day with a hot tank when morning draws provide the highest solar yield, 30.5% higher than an evening draw. It is clear that hot tanks at the beginning of the day limit the performance of solar thermal systems, especially those with late afternoon and evening draws.

Table 4.6 shows the average temperature of the water throughout the draw for each draw profile, volume and initial tank temperature condition. The reduction in temperature of the DHW as it is drawn is known as loss of load.

Initial Tank Temperature	Draw Volume	Average Draw Temperature (°C)			
		Morning	Early Afternoon	Late Afternoon	Evening
65°C	0.25Vt	63.6	74.0	89.1	89.3
	0.5Vt	63.4	73.5	88.8	89.0
	0.75Vt	62.3	72.3	87.0	87.4
	1Vt	59.7	69.4	83.3	83.6
	1.25Vt	55.9	65.2	77.8	77.8
	1.5Vt	51.6	60.3	71.6	71.2
45°C	0.25Vt	44.5	58.3	76.3	77.2
	0.5Vt	44.3	57.7	75.9	77.0
	0.75Vt	43.6	56.9	74.5	75.6
	1Vt	42.1	54.9	71.4	72.4
	1.25Vt	39.7	51.8	67.0	67.5
	1.5Vt	37.0	48.3	61.8	62.0
12°C	0.25Vt	14.2	33.3	54.5	57.5
	0.5Vt	14.1	32.4	54.1	57.4
	0.75Vt	13.9	31.8	53.2	56.4
	1Vt	13.8	31.1	51.3	54.2
	1.25Vt	13.7	30.0	48.5	50.8
	1.5Vt	13.6	28.6	45.2	46.9

Table 4.6: Matrix indicating loss of load

High initial tank temperatures ensure that the average temperature of the draw is above the usable limit (taken as 35°C) for all draw volumes and timings. Late afternoon and evening draw profiles are able to provide solar fractions of 100% on this summer's day; however it is clear from this table that cold initial tank temperatures are unable to provide water at a usable temperature before midday indicating a need for auxiliary heating.

- Solar yield increases with increasing draw volume for all draw profiles except the evening draw with maximum yields being obtained at the highest draw volume (525L/day);
- Solar yield increases for all draw volumes as initial tank temperatures are reduced, maximum increases occur for evening draws when yields can be increased by 65.2% when temperature is reduced from 65°C to 12°C;
- Evening draw profiles lead to the lowest solar yield values for all draw volumes and initial tank temperature conditions;

- Morning draws are preferable when initial tank temperatures are high providing a maximum of 59.7% increase in yield compared to an evening draw when initial tank temperature is 65°C;
- Midday profiles are preferable for cold starting tank temperatures providing a maximum of 9.9% improvement in yield compared to an evening draw when initial tank temperature is 12°C;
- Although morning draws provide higher solar yield values than evening draws when tank conditions are initially cold (maximum 1.5% higher) this water is unusable in comparison due to low temperatures;

4.14.4.2 The effects of auxiliary heating

The auxiliary heating is activated between 06:00-20:00 for each draw profile and is thermostatically controlled in this period so that the heater is active if the temperature of the water in the top of the tank falls below 55°C and deactivates when the water reaches 60°C; the initial condition of the tank is fully mixed at 12°C.

By introducing auxiliary heat the solar yield has been decreased by a maximum of 12.1% occurring for all evening draws and draws of 0L/day (Figure 4.15).

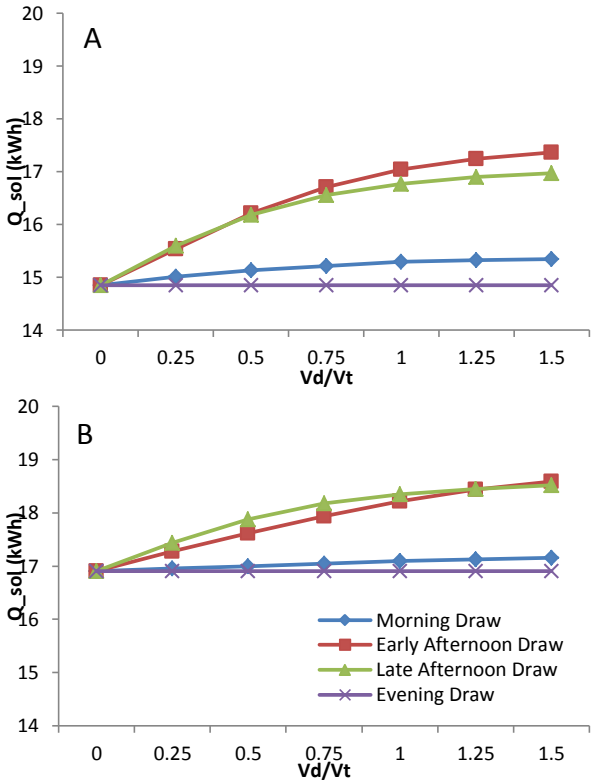

Figure 4.15: Daily solar yield versus draw volume for four DHW draw profiles on a single summer's day A) 10kW auxiliary 06:00-20:00; B) No auxiliary

For a single day starting with a fully mixed tank at 12°C later draw profiles (late afternoon and evening) provide the highest solar fractions (Figure 4.16). Water drawn at these times can make full use of the solar heat collected by the system with auxiliary heat added to boost the temperature during the draw. The morning profile has the lowest solar fraction over a single day due to the low temperature of the water at this time of day which has received the least amount

of solar energy at the time of draw than any other profile being a maximum of 18% lower than the late afternoon profile. The solar yield that is collected after the draw is stored in the tank, but if it is not used is wasted.

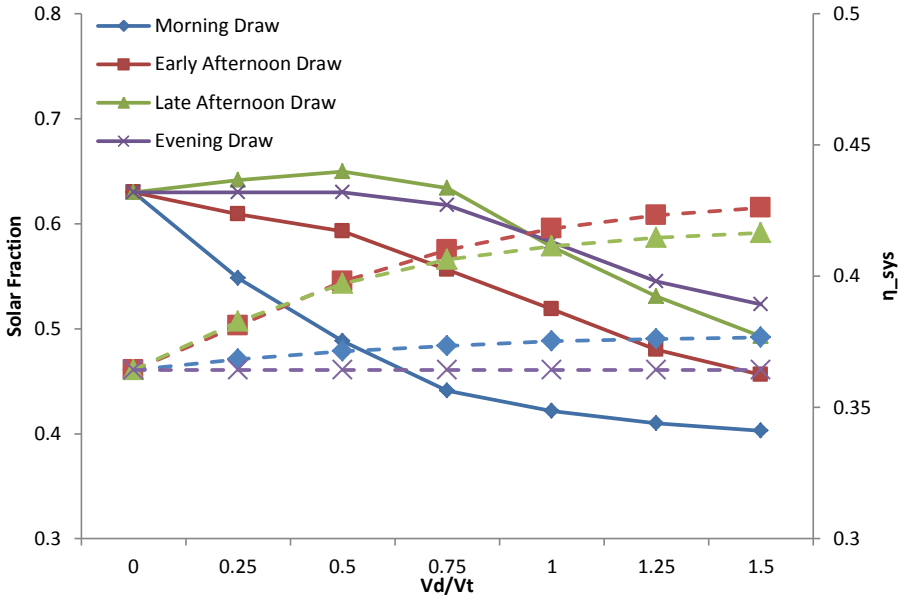

Figure 4.16: Solar fraction (solid line) and system efficiency (dashed line) versus draw volume for four DHW draw profiles on a single summer's day, 10kW auxiliary 06:00-20:00

The system efficiency (η_{sys}) is simply a ratio of the solar yield to the amount of irradiation incident on the solar collector; therefore this parameter follows the same pattern as the solar yield in Figure 4.15 and increases with increasing DHW consumption volume. The solar fraction decreases with increasing consumption volume despite increases solar yield sue to an increased requirement for auxiliary heating to ensure the desired set-point temperature.

Ranking the four profiles in terms of solar fraction and system efficiency reveals that the best time to draw DHW on a single day is in the late afternoon followed by the early afternoon; morning is the worst time to draw water providing low system efficiencies and solar fractions on a single day (Table 4.7).

	Solar Fraction Rank	System Efficiency Rank	Overall Score
Morning	1	2	2
Early Afternoon	2	4	8
Late Afternoon	4	3	12
Evening	3	1	3

Table 4.7: Profiles ranked based on solar fraction and system efficiency after a single day simulation

- The introduction of auxiliary heating has the effect of reducing the solar yield by a maximum of 12.1% for all DHW draw profiles when comparing results for a single summer's day simulation with initial tank conditions 12°C fully mixed – from this it can be concluded that the auxiliary input affects the solar yield thus clarifying the direction of the causal relationship between these two parameters;
- Late afternoon is the best time to draw DHW providing high system efficiencies and solar fractions

4.14.5 Sensitivity study

A sensitivity analysis was performed using a single day simulation on a summer's day to discover which of the three main variables studied in the above simulations had the greatest influence on the performance of the solar thermal system in terms of yield. The three variables in question are the time of use, initial tank temperature and DHW consumption volume.

Each of the variables had a base value established that would provide a medial solar yield; the base values for each of the variables were a 6am draw; 150L consumption volume; and 40°C initial tank temperature. The values were then changed one at a time by +/-50%; the resulting tornado plot is shown in Figure 4.17.

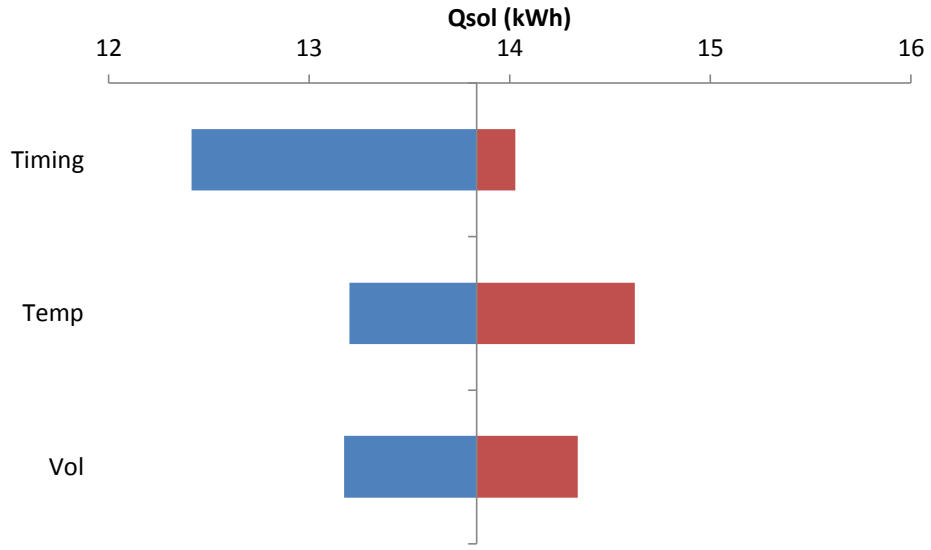

Figure 4.17: Tornado plot for three variables for a single summer's day simulation

The daily solar yield for the base case was 13.8kWh; the red bars indicate the case with a 50% increase in each value and the blue bar a 50% decrease. It can be seen that in terms of a single day, the time of DHW draw has the largest influence on the solar thermal performance; when the hot water is drawn at 6pm the yield is reduced by 10% compared to the 6am base case, when drawn at 12pm 1.4% increase in yield. A 50% increase in the initial tank temperature provides a 4.3% reduction in the yield, a 50% reduction in the tank temperature gives a 5.7% increase in yield. A 50% increase in the DHW draw volume results in a 3.6% increase in yield compared to a 4.3% decrease when the volume is decreased by 50%.

4.14.6 Thermal performance over a two-day period

The previous simulation scenarios were conducted over a single day and so the effect of thermal storage on the performance of subsequent days was not considered. In order to observe the thermal performance of the solar thermal system over a continuous period, simulations were conducted over a two-day period (Figure 4.18). In these simulations the second day is identical in terms of weather conditions as the single day in the previous set of simulations. The auxiliary heater was maintained and was in operation between 06:00-20:00, thermostatically controlled to maintain temperatures between 55-60°C. The tank was initially in the fully mixed condition at 12°C and a constant draw volume of 350L/day was used for all simulations.

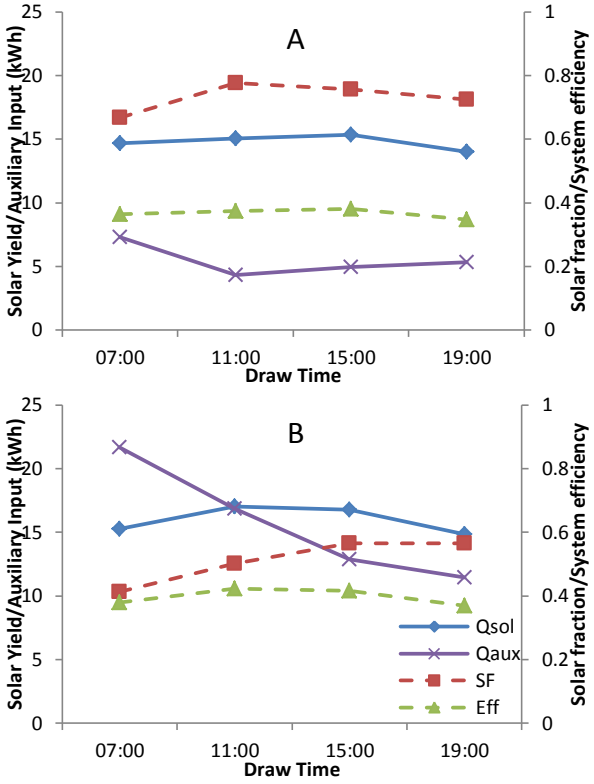

Figure 4.18: Thermal performance versus time of draw for a 350L/day draw volume a) second day in two-day simulation; b) single day in one-day simulation note that these two days are the same in terms of weather conditions

The thermal storage enables heat collected by the solar thermal system to be used on the following day if it is not made use of on the day it is collected. The effect of this is to reduce the amount of auxiliary heat required on the following day for all draw profiles, notably the morning draw which has a reduction in Q_{aux} of 66.3% (Figure 4.18). Ranking the draw profiles as before reveals that for the second-day in a two-day simulation the best time to draw water is the early afternoon followed by the late afternoon; these profiles have the highest solar yields and lowest auxiliary inputs. The morning and evening draw are equal in ranking with solar yield being higher for morning usage but auxiliary requirement being less for evening draws. Solar yield is reduced on the second day compared to the single day performance due to an increase in initial tank temperature as a result of stored thermal energy.

The difference in solar yield between the best and worst cases (late afternoon and evening respectively) for the single day simulation is 11.9%; for the second day in the two-day simulations the difference between best and worst (late afternoon and evening) is 8.8% suggesting that over a continuous period the difference in performance due to draw profile is reduced compared to a single day.

Figure 4.19 shows that a smaller morning auxiliary boost is required for all profiles on the second day in the two day simulation but the late afternoon and evening draws reduce the tank temperature for the next day leading to higher morning auxiliary inputs for these profiles. The early afternoon profile maintains high tank temperatures on the following morning reducing the morning auxiliary requirement, the solar collector then increases the temperature to above 60°C

prior to the draw thus reducing the need for auxiliary heat during the draw compared to the morning profile.

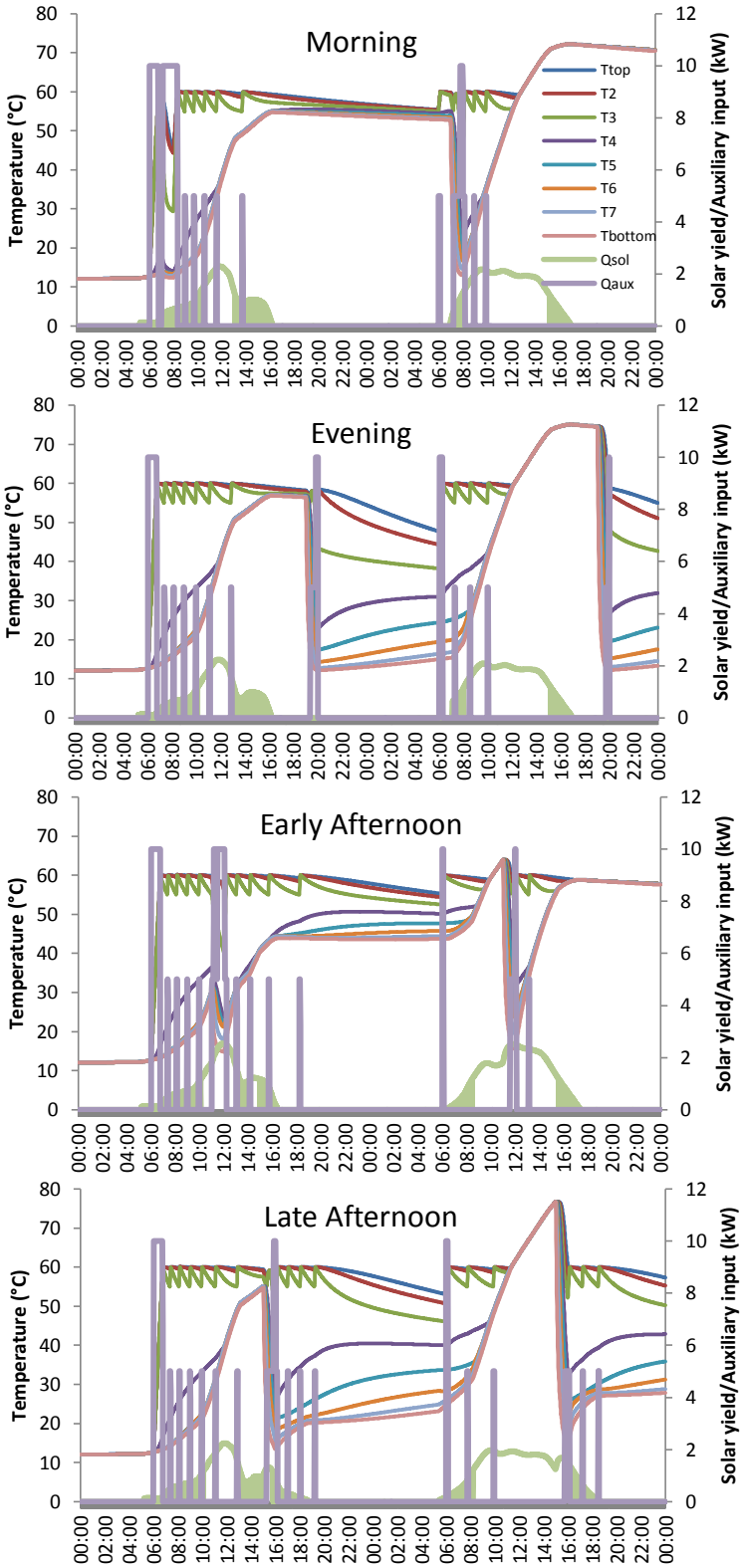

Figure 4.19: Tank temperatures and auxiliary firing pattern for two-day simulation 350L/day, 10kW auxiliary 06:00-20:00

The above results suggest that the solar yield affects the amount of auxiliary heat required when considering performance over a continuous period. This is a reversal of the causal relationship linking the parameters Q_{sol} and Q_{aux} obtained from the single day simulations. It can therefore be said that over a period of more than one day, owing to the thermal storage allowing unused solar heat from the previous day to be carried forward to subsequent days, the solar yield determines the amount of auxiliary heat required.

- Early afternoon profile is best providing high system efficiencies and solar fractions;
- Over a continuous period the amount of solar yield collected during the day affects the amount of auxiliary heat required rather than the other way around; this is due to thermal storage allowing the previous day's heat to be used in the early draws of the second day, and the current day's solar heat to be used when the draws are later in the same day. This has reversed the direction of the causal relationship from Q_{aux} - Q_{sol} to Q_{sol} - Q_{aux}.

4.14.6.1 The effect of resetting the tank temperature

It has been shown that the thermal performance of the solar thermal system and the thermal energy subsequently stored in the tank affects the performance of the system for the following day. To highlight this, further simulations were performed over a two-day period during the summer and over the course of a year for the four DHW profiles with a 06:00-20:00 auxiliary heating control strategy. In these simulations the temperature in the storage tank was 12°C in all temperature nodes at the beginning of each day; this was achieved with a large water draw of twice the volume of the tank from 02:00-03:00.

By resetting the tank temperature at the start of the second day the performance of the solar thermal system with respect to time of draw follows the same pattern as the single day simulation (Figure 4.20). After ranking the draw profiles in terms of solar fraction and system efficiency the results indicate that the late afternoon profile gives the best performance on the second day when the temperature is reset compared to the scenario when the temperature was not reset which suggests the early afternoon is the best time to draw water. There exists a small difference in the solar yields obtained for each draw profile in Figure 4.20a compared to Figure 4.20b: Morning draw -2.1%; early afternoon draw -0.76%; late afternoon draw +0.06%; evening draw -1.1%. This confirms that the initial temperature condition of the tank as a result of thermal storage affects the thermal performance of the solar thermal system.

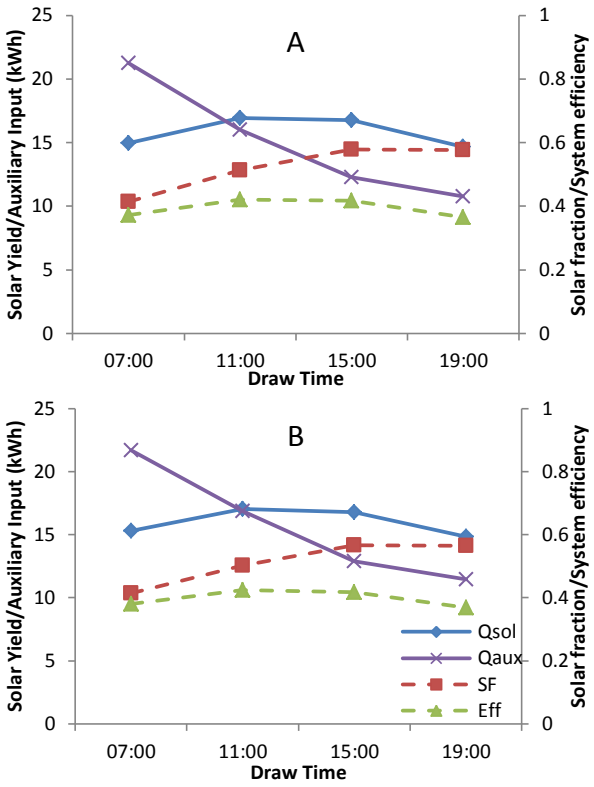

Figure 4.20: Thermal performance versus time of draw for a 350L/day draw volume a) second day in two-day simulation with tank temperature reset to fully mixed condition at 12°C; b) single day in one-day simulation note that these two days are the same in terms of weather conditions

4.14.7 Annual results

For a comprehensive study in which these simulations can be used to corroborate those of previous studies and to provide useful insights into long term performance, annual simulations are performed.

Throughout an entire year there are periods of low solar irradiance and external temperature, which will require the use of auxiliary heat to ensure suitable temperatures for hot water use. The auxiliary control strategy employed is a 10kW heater activated from 06:00-20:00 maintaining water between 55-60°C. Hot water is drawn in accordance with the four profiles used above at the predetermined volumes. All simulations begin on 1st January with a fully mixed tank at 12°C and continue for a period of one year at a time step of one hour.

The familiar relationship between solar yield and DHW consumption volume can be seen in Figure 4.21. Over the course of a year the early afternoon draw profile provides the highest annual solar yield for all consumption volumes whilst the evening draw gives the lowest - the maximum difference between the two occurs at a draw volume equal to the volume of the tank (350L/day) when drawing in the early afternoon leads to a 21% improvement in solar yield compared to the evening draw. Up to draw volumes of $0.75V_t$ a late afternoon draw provides higher solar yields than a morning draw profile; beyond volume of $0.75V_t$ drawing in the morning provides greater solar yield than a late afternoon profile. After $0.75V_t$ the solar yield values for a late afternoon draw show signs of a plateau whereas the morning draw profile continues to rise with increasing draw volume.

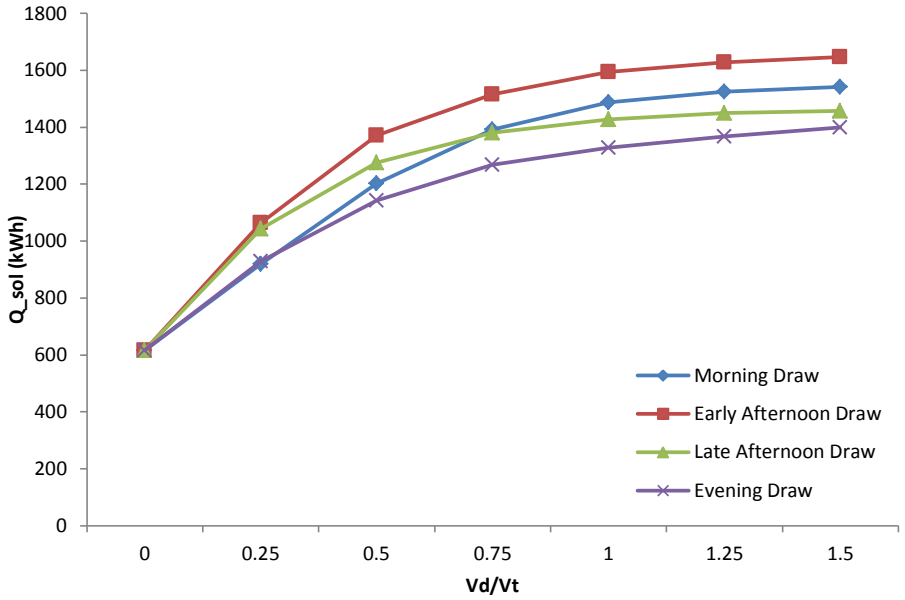

Figure 4.21: Annual solar yield versus draw volume for four DHW draw profiles with 10kW auxiliary input 06:00-20:00

The annual solar fraction and system efficiency is shown in Figure 4.22; the profile which provides the best solar fraction and system efficiency over the year is the early afternoon, the evening profile provides the worst annual solar fraction and system efficiency. The biggest difference between the performance of the best and worst profiles occurs at a draw volume of 350L/day: solar fraction is 3.3% higher for the evening draw and system efficiency is 4.3% higher.

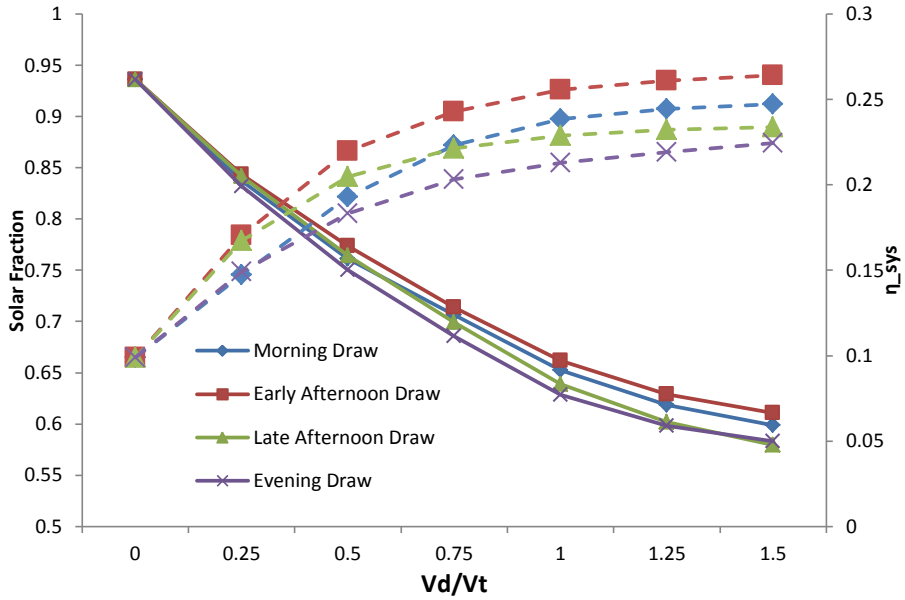

Figure 4.22: Annual solar fraction (solid line) and system efficiency (dashed line) versus draw volume for four DHW draw profiles, 10kW auxiliary 06:00-20:00

Over the course of the year the early afternoon profile is the best time to draw hot water providing the highest system efficiency, solar yield and solar fraction. In contrast the evening draw profile is worst for solar fraction and system efficiency being a maximum of 3.3% and 4.3% lower than the early afternoon draw profile respectively; this difference in best and worst

performance over a continuous period of a year is much smaller than that over a single day (3.3% difference versus 18% difference in solar fraction for annual simulations versus single day simulations).

4.14.8 Sensitivity study

A sensitivity analysis was conducted for annual simulations; the base values for each of the variables were a 6am draw; and 150L/day consumption volume. The values were then changed one at a time by +/-50%; the resulting tornado plot is shown in Figure 4.23.

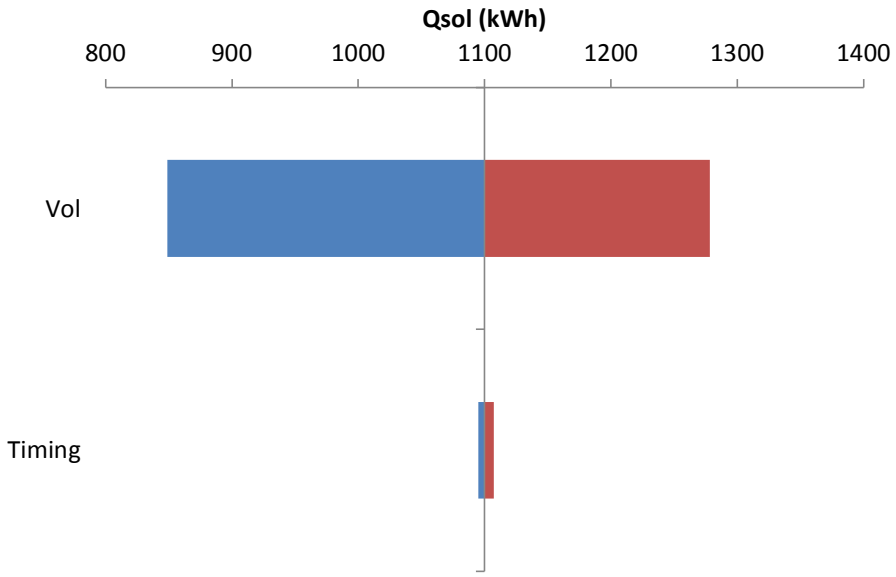

Figure 4.23: Tornado plot for two variables for a yearly simulation

The annual solar yield for the base case was 1100kWh/year; a 50% increase in DHW consumption volume increases the annual solar yield by 16.2%, decreasing the consumption by 50% results in a decrease in annual yield of 22.8%. Drawing water at 6am represents the base case producing 1100kWh/year; drawing at 12pm increases the annual yield by 0.7% compared to a decrease of 0.5% when water is drawn at 6pm each day. For long term performance the daily consumption volume of DHW has the greatest impact compared to the timing of the draw.

4.14.9 Conclusions to the TRNSYS simulations

A comprehensive series of simulations was performed in TRNSYS for the most common configuration of a solar thermal system in the UK based on data collected from the EST solar thermal field trial. The simulations investigated the effects of DHW profiles, specifically the timing and volume of the draw, as well as initial tank temperature conditions and auxiliary heating on the thermal performance of the solar system. The thermal performance was described by the solar yield, solar fraction and system efficiency. The thermal performance of the system was modelled over a single day and two-day period at a resolution of 1 minute; this allowed complex interaction within the tank to be explored and the effects of thermal storage to be determined. Following this the long term performance was simulated which provided a more useful insight into the effects of usage behaviour that may be experienced by installed systems, the performance of which is typically quoted in annual terms.

The simulations allowed control over all system parameters and therefore enabled the effects of a single parameter to be quantified. This kind of analysis was impossible to perform using data collected in the EST field trial where many system parameters vary from day to day and system to system including weather conditions, usage patterns and system configuration. Although non-recurrent draw profiles are often seen in reality with changing draw timings and volumes from day to day, by fixing the volume and timing in the simulations it was possible to quantify the effect on the system performance over a day and a year as well as provide advice to potential users on the best times to draw DHW to maximise performance. Furthermore the investigations have highlighted which usage parameters have the greatest effect on short term and long term performance as well as aiding in the determination of the direction of the causal relationships between solar yield and auxiliary input. This information can be used to develop the causal map of factors that affect solar thermal system performance. The annual simulations can also be used to produce CPTs to be used in the Bayesian network which will be discussed in Chapter 6.

Initial temperature conditions of the tank affect the performance of the solar thermal system with hotter tanks reducing the amount of solar yield that can be collected due to reduced heat transfer rates and operating periods resulting from a smaller temperature difference between the solar heated fluid and the storage water. It was found that when the initial condition of the tank is hot then earlier draw profiles lead to improved solar yields compared to later draws; this is because temperatures in the bottom of the tank are reduced by the influx of cold water, which occurs early on in the operating period of the solar thermal system leading to improved heat transfer rates and extended operating periods. Colder starting temperatures in the tank provide the highest solar yield overall; however the earlier draws and large draw volumes struggle to provide water at a usable temperature without the use of auxiliary heating. When the tank is cold solar yield obtained by a morning versus evening profile are similar in magnitude owing to the initial conditions of the tank being identical if no auxiliary is present; in addition water draw occurs before or after the operating period of the system and so the solar thermal system delivers a similar yield over the course of a day. Midday draw profiles extend the operating period of the system and reduce the temperature in the bottom of the tank during the main operating period of the collector thus improving heat transfer rates at this time; this all leads to improved solar yields over the course of a day. Loss of load is an issue when auxiliary heating is not provided.

Auxiliary heating has the effect of reducing the solar yield when the short term (daily) performance of the solar thermal system is concerned. This is due to an increase in the temperature in the storage tank. Midday draw profiles provide the highest solar yield values; however lower amounts of auxiliary heating is required for later draw profiles which can make use of the thermal energy delivered by the collector throughout the day. For this reason the late afternoon and evening draws provide the best solar fraction values after a single day. When the performance of the system is explored over a continuous period greater than a single day then the thermal storage allows for heat energy to be used on subsequent days. This has the effect of reducing the auxiliary requirement for all draw profiles – this means that the causal relationship between solar yield and auxiliary flows from solar yield to auxiliary input. The annual solar yield, solar fraction and system efficiency are affected more greatly by the volume of the draw rather than the timing but it can be said that the early afternoon is the best time to draw hot water provided that 100% of the required volume is drawn at this time. The effect of a non-recurrent

draw profile has been presented by (Buckles & Klein 1980) who show that it reduces the performance of the solar thermal system.

The fact that these simulations do not quantify the effects of a non-recurrent draw is irrelevant when considering the incorporation of these results into a Bayesian network for probabilistic modelling of performance. The data from the EST field trial provides a distribution of the daily solar yields which can be explained by the presence of known variables (similar to the inclusion of more variables in a regression equation increasing the R^2 value). The data as it stands with no consideration of draw profile means that a proportion of the spread of solar yield represented by the distribution arises from differences and variation in the draw profile of the DHW – this is a known unknown and is represented by a uniform probability distribution across all possible draw timings. Despite not knowing the effect of draw profile on the distribution of the solar yield the uncertainty represented by the distribution can still be quantified and decisions made based on this. However, by including the information discovered in the simulations it is possible to calibrate or modify the shape of the performance distribution given that information about the draw timings has become known. Although the data is simulated it provides the best evidence of performance variability due to DHW draw timing.

4.15 Discussion and conclusions

The three methods of modelling described above are deterministic in nature; input variables are entered into a series of equations, either physics based or statistically determined, and a result is calculated.

These different deterministic models have a common disadvantage: uncertainty in the performance of solar thermal systems is excluded from prediction. The performance of a solar thermal system will differ from year to year due to variation in occupancy behaviour, annual irradiation, and degradation of the system. Performance between different systems will vary for the same reasons as well as installation quality, system configuration and size. Therefore the single figure for annual yield predicted by most deterministic models does not account for the variation in the input parameters and therefore does not provide sufficient confidence for design and investment decisions to be made.

Uncertainty can be categorised into two types: aleatoric uncertainty, which represents the uncertainty in the solar thermal system performance due to uncertainties in the input parameters; and epistemic uncertainty, which represents the uncertainty in the solar thermal system performance that is due to the effect of some unknown or unmeasured parameter, for example ignoring the effects of auxiliary timing on solar yield because it has not been measured or considered.

Uncertainty in performance predictions can be incorporated to produce a probability distribution of solar thermal yield, which will indicate the possible range and most likely yield obtainable for a given system. By producing a distribution of annual performance an investor can make an informed decision about the risks involved in installing a solar thermal system: broad distributions represent risky investments whereas narrow distributions are less risky. This will not only help homeowners to decide on whether a solar thermal system is worth installing based

on their personal circumstances relating to system configuration, location and usage behaviour, but also allow policy makers to target solar thermal systems to those families that may have lifestyles that are compatible with the technology.

The following chapter discusses methods of incorporating uncertainty into predictions of solar thermal system performance.

4.16 References

Ahmad, M. & Culp, C.H., 2006. Uncalibrated Building Energy Simulation Modeling Results. *HVAC&R Research*, 12(4), pp.1141–1155.

Allen, S.R. et al., 2010. Integrated appraisal of a Solar Hot Water system. *Energy*, 35(3), pp.1351–1362.

Andersen, E. & Furbo, S., 1999. Thermal destratification in small standard solar tanks due to mixing during tapping. In *Proceedings of ISES Solar World Congress*.

Anderson, B., 2005. *Treatment of solar water heating in SAP 2005*, Unpublished.

Armstrong, J.S., 2012. Illusions in regression analysis. *International Journal of Forecasting*.

Ayompe, L.M., Duffy, A., Mc Keever, M., et al., 2011. Comparative field performance study of flat plate and heat pipe evacuated tube collectors (ETCs) for domestic water heating systems in a temperate climate. *Energy*, 36(5), pp.3370–3378.

Ayompe, L.M., Duffy, A., McCormack, S.J., et al., 2011. Validated TRNSYS model for forced circulation solar water heating systems with flat plate and heat pipe evacuated tube collectors. *Applied Thermal Engineering*, 31(8-9), pp.1536–1542.

Bates, J., Bertarelli, L. & Schmidt, G., 1999. Active solar heating system performance and data review. *ETSU S/P3/00270/REP*.

Boardman, B. et al., 2005. *40% House*, University of Oxford: Environmental Change Institute.

Booth, A. & Choudhary, R., 2011. Calibrating micro-level models with macro-level data using Bayesian regression analysis. In *Building Simulation*. pp. 641–648.

BRE, 2006. *Energy Performance of Buildings Directive*, Available at: http://www.bre.co.uk/filelibrary/Scotland/Energy_Performance_of_Buildings_Directive_%28EPBD%29.pdf.

BRE, 2005. *Energy Use in Homes: Fuel Consumption*, Report available at: https://www.bre.co.uk/filelibrary/pdf/rpts/FuelConsumption.pdf.

BRE, 2013. SAP 2012 The Government's Standard Assessment Procedure for Energy Rating of Dwellings.

Buckles, W.E. & Klein, S.A., 1980. Analysis of solar domestic hot water heaters. *Solar Energy*, 25, pp.417–424.

Cadafalch, J., 2009. A detailed numerical model for flat-plate solar thermal devices. *Solar Energy*, 83(12), pp.2157–2164.

Chelmer Heating Solutions, 2013. Ecocat Thermal Store Technical Information. Available at: http://www.chelmerheating.co.uk/ecocat_thermal_store_technical_info.html.

Crawley, M.J., 2015. *Statistics An Introduction Using R* Second Edi., John Wiley & Sons Ltd.

Crowther, M. et al., 2010. *Report to DECC on Heat metering for the RHI*, London.

Dancey, C. & Reidy, J., 2004. *Statistics Without Maths for Psychology: Using SPSS forWindows* Third Edit., London: Prentice Hall.

Degelman, L.O., 2008. Calibrated simulation of a solar hot water system to match degraded performance over a 22-year period using two models. *Building and Environment*, 43(4), pp.628–637.

de Winter, F., 1975. Heat exchanger penalties in double-loop solar water heating systems. *Solar Energy*, 17(6), pp.335–337.

Domínguez-Muñoz, F. et al., 2012. Design of solar thermal systems under uncertainty. *Energy and Buildings*, 47, pp.474–484.

Duffie, J.A., Klein, S.A. & Beckman, W.A., 1988. *TRNSYS Evolution*, Solar Energy Laboratory, University of Wisconsin.

EST, 2008a. *Appendix 1 In-situ Measurement of Solar Thermal Performance in Dwellings: A technical specification for field monitoring*, London.

EST, 2010. *Getting warmer: A field trial of heat pumps*, London.

EST, 2011. *Here comes the sun: a field trial of solar water heating systems*, London.

EST, 2008b. *Measurement of Domestic Hot Water Consumption in Dwellings*, London.

EST, 2001. Solar hot water systems in new housing. *GIR88*.

Evarts, J.C. & Swan, L.G., 2013. Domestic hot water consumption estimates for solar thermal system sizing. *Energy and Buildings*, 58, pp.58–65.

Farrar, D.E. & Glauber, R.R., 1967. Multicollinearity in regression analysis: the problem revisited. *The Review of Economics and Statistics*, 49(1), pp.92–107.

F-Chart Software, 2014. F-Chart Software. Available at: www.fchart.com [Accessed April 14, 2014].

Fenton, N., Krause, P. & Neil, M., 2002. Software Measurement: Uncertainty and Causal Modeling. *IEEE Software*, (July/August).

Fenton, N. & Neil, M., 2007. Managing Risk in the Modern World. *Knowledge Transfer Report for the London Mathematical Society and the KTN for Industrial Mathematics*.

Fenton, N. & Neil, M., 2011. The use of Bayes and causal modelling in decision making, uncertainty and risk. *CEPIS Upgrade*, 12(5), pp.10 – 22.

Firth, S.K., Lomas, K.J. & Wright, A.J., 2010. Targeting household energy-efficiency measures using sensitivity analysis. *Building Research & Information*, 38(1), pp.25–41.

Forward, D. & Roberts, C., 2008. *Viridian Solar - Clearline Solar Thermal Field Trial*, Watford.

Furbo, S. et al., 2005. Smart solar tanks for small solar domestic hot water systems. *Solar Energy*, 78(2), pp.269–279.

German Solar Energy Society, 2010. *Planning and Installing Solar Thermal Systems* Second., London: Earthscan.

Haberl, J.S., 2004. *Literature review of uncertainty of analysis methods*, ESL-TR-04/08-04.

Hasan, A., 1997. Thermosyphon solar water heaters: Effect of storage tank volume and configuration on efficiency. *Energy Conversion and Management*, 38(9), pp.847–854.

Henderson, J., 2002. *Updating the solar water heating procedure in BREDEM*, Building Research Establishment.

Henderson, J. & Hart, J., 2013. BREDEM 2012 – A technical description of the BRE Domestic Energy Model.

Heo, Y., Choudhary, R. & Augenbroe, G. a., 2012. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47, pp.550–560.

Hobbi, A. & Siddiqui, K., 2009. Optimal design of a forced circulation solar water heating system for a residential unit in cold climate using TRNSYS. *Solar Energy*, 83(5), pp.700–714.

Hottel, H.C. & Whillier, A., 1955. Evaluation of flat-plate solar collector-performance. In *Transactions of the Conference on the Use of Solar Energy*. pp. 74–104.

Ijumba, P. & Sebitosi, A.B., 2010. Evaluating the impact of consumer behaviour on the performance of domestic solar water heating systems in South Africa. *Journal Of Energy In Southern Africa*, 21(1), pp.25–34.

Johnston, D., 2003. *A Physically-Based Energy and Carbon Dioxide Emission Model of the UK Housing Stock*. Leeds Metropolitan University.

Jordan, U. & Furbo, S., 2005. Thermal stratification in small solar domestic storage tanks caused by draw-offs. *Solar Energy*, 78(2), pp.291–300.

Jordan, U. & Vajen, K., 2000. Influence of the DHW load profile on the fractional energy savings: A case study of a solar combi-system with TRNSYS simulations. *Solar Energy*, 69, pp.197–208.

Jordan, U. & Vajen, K., 2001. Realistic Domestic Hot-Water Profiles in Different Time Scales Realistic Domestic Hot-Water Profiles in Different Time Scales. *Solar Heating and Cooling Program of the International Energy Agency, Task 26: Solar Combisystems*, pp.1–18.

Kavgic, M. et al., 2010. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, 45(7), pp.1683–1697.

Kelly, N.J. & Cockroft, J., 2011. Analysis of retrofit air source heat pump performance: Results from detailed simulations and comparison to field trial data. *Energy and Buildings*, 43(1), pp.239–245.

Kelly, S., 2011. Do homes that are more energy efficient consume less energy?: A structural equation model of the English residential sector. *Energy*, 36(9), pp.5610–5620.

Kelly, S., Crawford-Brown, D. & Pollitt, M.G., 2012. Building performance evaluation and certification in the UK: Is SAP fit for purpose? *Renewable and Sustainable Energy Reviews*, 16(9), pp.6861–6878.

Klein, S.A., 1976. *A Design Procedure for Solar Heating Systems*. University of Wisconsin-Madison.

Klein, S.A., Beckman, W.A. & Duffle, J.A., 1976. A design procedure for solar heating systems. *Solar Energy*, 18, pp.113–127.

Klein, S.A., Duffie, J.A. & Beckman, W.A., 2010. Trnsys 17: A Transient System Simulation Program. , pp.1–29.

Knudsen, S., 2002. Consumers' influence on the thermal performance of small SDHW systems—Theoretical investigations. *Solar Energy*, 73(1), pp.33–42.

Lee, T., Yao, R. & Coker, P., 2013. An analysis of UK policies for domestic energy reduction using an agent based tool. *Energy Policy*, 66, pp.267–279.

Liao, Z. & Dexter, A.L., 2010. An inferential model-based predictive control scheme for optimizing the operation of boilers in building space-heating systems. *IEEE Transactions on Control Systems Technology*, 18(5), pp.1092–1102.

Lundh, M., Wäckelgård, E. & Ellegård, K., *Design of hot water user profiles for Swedish households based on time diaries*, Report available at: http://www.researchgate.net/publication/228472499_Design_of_hot_water_user_profiles_for_swedish_households_based_on_time_diaries.

Martin, C. & Watson, M., 2002. *Further testing of solar water heating systems*, Department of Trade and Industry (DTI), London.

Martin, C. & Watson, M., 2001. *Side by side testing of eight solar water heating systems*, Department of Trade and Industry (DTI), London.

Michaelides, I.M., 1993. *Computer simulation and optimisation of solar heating systems for Cyprus*. University of Westminster.

Miller, I., Freund, J.E. & Johnson, R.A., 1990. *Probability and Statistics for Engineers* Fourth., Prentice Hall.

Morrison, G.L., Gilliaert, D. & Tebaldi, P., 1992. Outdoor testing of solar water heaters - effects of load pattern and auxiliary boosting. *Solar Energy*, 49(4), pp.299–308.

Munzinger, M. et al., 2006. *PV domestic field trial final technical report*, Department of Trade and Industry (DTI), London.

Murphy, G.B., Samuel, A. & Counsell, J., 2011. Enhancement of the UK standard assessment procedure (SAP) solar water heating prediction algorithm using parametric dynamic thermal simulations. In *Proceeedings of Building Simulation 2011*. pp. 1195–1202.

Mutch, J.J., 1974. *Residential water heating: Fuel conservation, economics and public policy*,

Natarajan, S. & Levermore, G.J., 2007. Predicting future UK housing stock and carbon emissions. *Energy Policy*, 35(11), pp.5719–5727.

National Resources Canada, 2013. WATSUN 2009. Available at: https://www.nrcan.gc.ca/energy/software-tools/7435 [Accessed April 27, 2015].

Neteler, M. & Mitasova, H., 2002. *Open Source GIS: A GRASS GIS Approach* First., Boston: Kluwer Academic Publishers.

Object Managment Group, 2014. Unified Modeling Language™ (UML®) Resource Page. Available at: http://www.uml.org/ [Accessed March 30, 2015].

Remund, J., Kunz, S. & Schilter, C., 2008. *Meteonorm Version 6.0*, Bern: Meteotest.

Rodríguez-Hidalgo, M.C. et al., 2012. Domestic hot water consumption vs. solar thermal energy storage: The optimum size of the storage tank. *Applied Energy*.

Shariah, A.M. & Ecevit, A., 1995. Effect of hot water load temperature on the performance of a thermosyphon solar water heater with auxiliary electric heater. *Energy Conversion and Management*, 36(5), pp.289–296.

Shariah, A.M. & Lof, G.O.G., 1996. The optimisation of tank volume - to - collector area ratio for a thermosyphon solar water heater. *Renewable Energy*, 7(3), pp.289–300.

Shariah, A.M. & Löf, G.O.G., 1997. Effects of auxiliary heater on annual performance of thermosyphon solar water heater simulated under variable operating conditions. *Solar Energy*, 60(2), pp.119–126.

Shorrock, L.D. & Dunster, J.E., 1997. The physically-based model BREHOMES and its use in deriving scenarios for the energy use and carbon dioxide emissions of the UK housing stock. *Energy Policy*, 25(12), pp.1027–1037.

Sokunbi, M.O., 2014. Sample entropy reveals high discriminative power between young and elderly adults in short fMRI data sets. *Frontiers in Neuroinformatics*, 8(July), pp.1–12.

Solar Energy Laboratory, 2006. *TRNSYS 16 Volume: Weather Data*, University of Wisconsin.

Solar Energy Laboratory, 2014. TRNSYS users mailing list. Available at: https://mailman.cae.wisc.edu/listinfo/trnsys-users.

Spur, R. et al., 2006. Influence of the domestic hot-water daily draw-off profile on the performance of a hot-water store. *Applied Energy*, 83(7), pp.749–773.

Suri, M. & Hofierka, J., 2004. A New GIS-based Solar Radiation Model and Its Application to Photovoltaic Assessments. *Transactions in GIS*, 8(2), pp.175–190.

Šúri, M., Huld, T. & Dunlop, E., 2005. PV-GIS: a web-based solar radiation database for the calculation of PV potential in Europe. *International Journal of Solar Energy*, 24(2), pp.55–67.

Swan, L.G. & Ugursal, V.I., 2009. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8), pp.1819–1835.

Taherian, H. et al., 2011. Experimental validation of dynamic simulation of the flat plate collector in a closed thermosyphon solar water heater. *Energy Conversion and Management*, 52(1), pp.301–307.

The Energy Monitoring Company Ltd, 2001. *Analysis of performance data from four active solar water heating installations*, Department of Trade and Industry (DTI), London.

University of Wisconsin, 2013. A Transient Systems Simulation Program. Available at: http://sel.me.wisc.edu/trnsys/.

Valentin Software, 2015. T*SOL. Available at: http://www.valentin-software.com/en/products/solar-thermal/14/tsol.

Vine, E., Diamond, R. & Szydlowski, R., 1987. DHW consumption in four low income apartment buildings. *Energy*, 12(6), pp.459–467.

Wolf, D., Sembira, A.N. & Kudish, A.I., 1984. Dynamic simulation and parametric sensitivity studies on a central solar domestic hot water system. *Energy*, 9(2), pp.169 – 181.

Yohanis, Y.G. et al., 2006. The annual number of days that solar heated water satisfies a specified demand temperature. *Solar Energy*, 80(8), pp.1021–1030.

Yoon, J., Lee, E.J. & Claridge, D.E., 2003. Calibration procedure of energy performance simulation of a commercial building. *Journal of Solar Energy Engineering*, 125(3), pp.251–257.

Zhao, H. & Magoulès, F., 2012. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6), pp.3586–3592.

Chapter 5

Incorporating Uncertainty into the Prediction of STS Performance

5.1 Introduction

Three different modelling strategies used to make predictions about the performance of buildings and installed renewable energy technologies such as PV, solar thermal and heat pump systems have been comprehensively described in Chapter 4. These predictions are usually point-estimates of the energy consumed by a dwelling or the energy consumed/delivered by one of its subsystems; however it has been shown in Chapter 3 that these point-estimates are often different from the measured value and therefore decisions made based on these predictions may result in under or over-prediction regarding financial returns, energy consumed or generated and carbon emitted. It is therefore necessary to quantify the uncertainty of important model parameters in order to provide a distribution of estimated performance as opposed to point values; this view is shared by many researchers in the field of building energy performance (Lomas & Eppel 1992; de Wit 1995; Burhenne et al. 2013; Booth et al. 2012; Heo et al. 2012; Booth & Choudhary 2013; Macdonald & Strachan 2001; Hopfe & Hensen 2011; de Wit & Augenbroe 2002; Lu et al. 2013; Wang et al. 2012).

Uncertainty in model input parameters causes uncertainty in the outputs generated by a model. De Wit (1995) suggests the sources of uncertainty in input parameters may be categorised as follows (de Wit 1995):

1. A lack of knowledge about the system details (building or some subsystem such as a solar thermal installation);
2. Deviations during manufacture of system components and installation quality;
3. Unpredictable behaviour of the users of the system;
4. A lack of knowledge about the underlying physical processes;

A lack of knowledge of the system details results in assumptions being made by modellers which may be oversimplified or unrealistic (Menezes et al. 2012; Carbon Trust 2011; de Wit & Augenbroe 2002). The quality of the installation of systems or build quality is a common influence on the energy performance of a building or the yield obtained from a renewable energy system (Carbon Trust 2011; Menezes et al. 2012; de Wilde 2014; Morant 2012; Bordass

et al. 2004; Miara et al. 2011; Loose et al. 2011). Occupants may operate appliances that account for unregulated loads and influence the internal environment by opening windows etc. – this is what is meant by user behaviour in terms of building energy performance and refers to the unpredictable energy consumption patterns of users and the way in which they influence the operation of the system (building or otherwise) by modifying the boundary conditions (usually described by temperature) (Carbon Trust 2011; Menezes et al. 2012; de Wilde 2014; Morant 2012; Cayre et al. 2011). With respect to domestic solar thermal systems used to provide hot water user behaviour includes the patterns of DHW usage described by: the volumetric consumption; set point temperature; time of use; duration of draw; and interactions with auxiliary heating systems – these quantities affect such factors as the return temperatures and system losses of the system and therefore the amount of heat that can be delivered (Jordan & Vajen 2000; Jordan & Furbo 2005). The consumption patterns of DHW are unpredictable and affected by several complicating factors (Knudsen 2002; Michaelides 1993; Evarts & Swan 2013).

There are few examples of research that incorporate uncertainty into the design/performance predictions of solar thermal systems (Mathioulakis et al. 2012; Domínguez-Muñoz et al. 2012); however there is far more research in this area conducted in the field of building performance estimation. Methods for quantifying uncertainty include sensitivity analysis techniques such as differential sensitivity analysis (DSA) (Macdonald & Strachan 2001; Lomas & Eppel 1992); stochastic sensitivity analysis (SSA) (Lomas & Eppel 1992); Monte Carlo analysis (Macdonald & Strachan 2001; Lomas & Eppel 1992; Hopfe & Hensen 2011; Burhenne et al. 2013; de Wit & Augenbroe 2002; Lu et al. 2013; Wang et al. 2012; Lee et al. 2013). Bayesian methods also exist including Bayesian calibration of models (Booth et al. 2012; Heo et al. 2012; Booth & Choudhary 2013; Kennedy & O'Hagan 2001) and Bayesian networks (Weber et al. 2012; Pearl 1988; Fenton & Neil 2013).

5.2 Monte Carlo methods

Monte Carlo analysis is an unstructured method of uncertainty analysis; unstructured methods differ from structured methods which rely on design of experiments to analyse the outcome of a model based on changing the input parameters – Differential Sensitivity Analysis (DSA) is an example of a structured method in which the input parameters to a model are altered individually to assess the influence on the model output (Macdonald & Strachan 2001).

Monte Carlo analysis as used in building energy modelling can be generalised by the graphical representation in Figure 5.1.

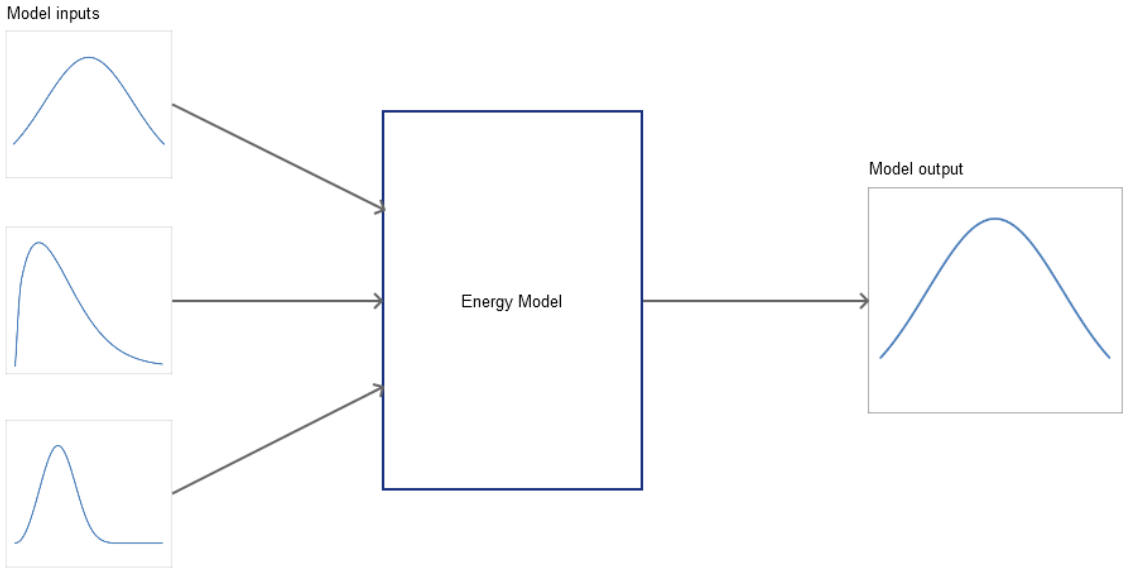

Figure 5.1: Graphical representation of the Monte Carlo method applied to energy performance modelling of systems

Uncertain model inputs are represented by probability distributions which are sampled a given number of times, *N*, and used in the energy model; the energy model may be a simplified physics based model, a dynamic simulation or a regression equation used to simulate any system or subsystem such as a building or renewable energy technology and is run *N* times to give *N* outputs presented as a distribution. Further explanations of the MC method can be found in (Lomas & Eppel 1992; Burhenne et al. 2013; Domínguez-Muñoz et al. 2012; Mathioulakis et al. 2012).

The distributions of the model inputs may be established by fitting a distribution to empirical data (Lu et al. 2013; Lee et al. 2013); alternatively they may take on an assumed form such as a Gaussian distribution which may or may not be determined using procedures for eliciting expert judgement (de Wit & Augenbroe 2002). For each run of the model a single value is randomly sampled from each of the input distributions – values with a higher probability of occurrence are more likely to be sampled. For *N* simulations each input variable will be sampled *N* times and there will be *N* output values used to make a distribution; the accuracy of this method is improved with higher values of *N*.

Monte Carlo simulations performed in this way have the advantage that interactions between the input variables are automatically taken into consideration because all variables are altered simultaneously. This is in contrast to DSA where only one of the variables is changed at each run of the simulation in order to ascertain the sensitivity of the model output to each of the inputs. A key disadvantage of Monte Carlo simulations is the increased computation time required to run a simulation many times over; for example Lee et al. (2013) performed a total of 20,000 simulations using EnergyPlus to provide a probability distribution of the energy savings of two retrofit options – these simulations were performed using 10 computers with two quad-core processors taking 8 hours to complete thus demonstrating the computing intensity of MC methods applied to complex system-of-systems (Lee et al. 2013).

5.3 Monte Carlo methods applied to solar thermal modelling

There are very few examples for which Monte Carlo methods have been applied to modelling solar thermal systems; however Mathioulakis et al (2012) introduced uncertainty to the performance prediction of solar thermal systems by applying Monte Carlo methods to a model that combines a regression equation with equations based on physical principles (Mathioulakis et al. 2012). This method calculates the daily yield of the system based on the daily collector output (obtained by regression); daily usage; heat losses over the night; and solar yield from the previous day. Annual yield is obtained by producing 365 iterations linking the previous day's "leftover" solar energy with that of the current day. This allows the effects of the previous day to be incorporated into the performance calculation. The limitation of this method lies in the specificity of the regression parameters with a given solar thermal system meaning new parameters would have to be found for each individual system and therefore require testing to be performed thus negating one of the benefits of modelling. Nevertheless it shows an example of how Monte Carlo methods can be used to propagate uncertainties in input variables through to the solar thermal yield prediction.

An alternative approach was made by Dominguez-Munoz et al. (2012) whereby Monte Carlo methods were applied to TRNSYS simulations of solar thermal systems. Uncertainties here were categorised into noise factors and design factors. Noise factors are uncontrollable by the designer and relate to climate and user behaviour; design factors are controllable and include system sizes and configurations. Both noise and design factors can be certain or uncertain and it is the designer's task to decide which variables will be considered uncertain. Uncertainty related to the modelling environment was included using a trigger factor which randomly selected different models used to simulate the system. Differences between models include such things as the different algorithms used to calculate incident irradiance on an inclined plane for example. Different modelling environments give different results for the same system and therefore introduce uncertainty in the design process. Dominguez-Munoz et al. (2012) showed how the Monte Carlo methods for incorporating uncertainty are an improvement on deterministic (single value answers) simulations for design purposes: Using the deterministic method design requirements for solar fraction and utilisation factor could be met with a collector area of 47m^2; however once uncertainty had been incorporated the probability of achieving the design requirements with 47m^2 was only 44.53%. Therefore, Monte Carlo simulations can be run with different design values until a minimum probability of success is achieved (90% in Dominguez-Munoz (2012)) (Domínguez-Muñoz et al. 2012).

An important discovery made by Dominguez-Munoz et al. (2012) was that solar fraction and utilisation factor were most sensitive to the set-point temperature of the auxiliary heater. From their Monte Carlo simulations they could make the suggestion that auxiliary SPT should remain fixed and uncontrollable by the user; this reduces the variation in the performance of the solar thermal system. It can therefore be concluded that it is variation in parameters that affects performance and not necessarily the actual value of the parameter; for example although high auxiliary SPTs were shown to reduce solar fraction and utilisation factor, they can be designed around provided that they remain constant.

5.4 Applying Monte Carlo methods to a simplified regression equation for long term solar thermal system performance

Monte Carlo methods have been applied to the simplified regression equation described in Chapter 4. The advantage of applying Monte Carlo methods to this long term performance equation is that the regression model is fitted to data from 35 different systems and, therefore, can be used to make predictions for any system provided the size and DHW demand are known; this is in contrast to the model in Mathioulakis (2012) which can only be applied to specific systems for which the regression coefficients are known.

Monte Carlo methods are applied by sampling from distributions of V_{tank}, $V_{d,ave}$ and A to give a distribution of $V_{tank}/V_{d,ave}$ and $VA\ ratio$. Values of $V_{tank}/V_{d,ave}$ and $VA\ ratio$ are then sampled from the distributions and inputted into the regression equation to provide a distribution of Q_{sol}; this distribution will give information of the likely solar heat yield for an unspecified system (a system for which $V_{tank}/V_{d,ave}$ and $VA\ ratio$ are unknown) and so will represent the distribution of performance of all systems in the UK provided the sample used to form the regression equation is representative of the total population.

V_{tank}, $V_{d,ave}$ and A were sampled in two ways:

1. The first way was to take 1000 random samples from the 35 data points of V_{tank}, $V_{d,ave}$ and A to calculate 1000 values of $V_{tank}/V_{d,ave}$ and $VA\ ratio$; these values were used in the regression equation to give a distribution for Q_{sol}.
2. The second way was to fit a distribution to the V_{tank}, $V_{d,ave}$ and A frequency histograms and to sample 1000 times from these fitted distributions. Fitting a continuous distribution has the benefit of reducing the granularity of the continuous variables caused by a limited number of data points. The distributions were fitted using the maximum likelihood method (MLE) and provides a continuous distribution of V_{tank}, $V_{d,ave}$ and A rather than being limited to 35 measured values. All of the distributions were found to be Log-Normal with parameters given in Table 5.1.

	Mean	**Standard deviation**
$V_{d,\,ave}$ (L/day)	80.851	0.561
V_{tank} (L)	191.995	0.275
A (m^2)	3.624	0.299

Table 5.1: Log-Normal parameters

The regression coefficients used in the simulations were the average values found by the least squares method detailed in Chapter 4.

Figure 5.2 shows the measured and predicted distributions using Monte Carlo simulations. The mean annual yield and standard deviations for the distributions are summarised in Table 5.2.

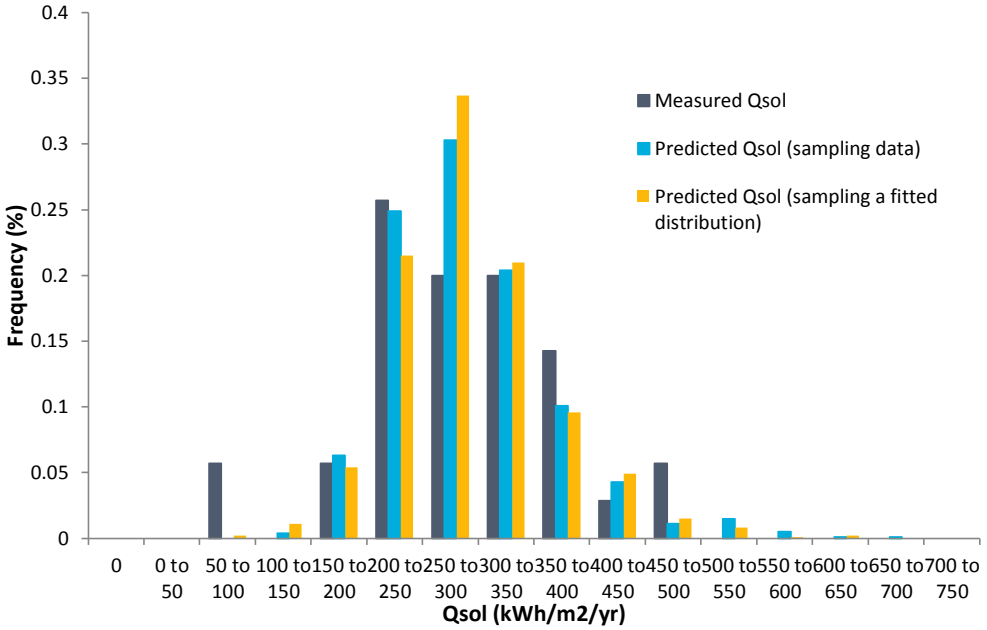

Figure 5.2: Measured and predicted distributions of Q_{sol} using Monte Carlo (1000 simulations); average values for the regression coefficients used

	Mean (kWh/m^2/yr)	Standard deviation (kWh/m^2/yr)
Measured	285.36	90.49
Sampled from data	290.59	73.78
Sampled from a fitted distribution	290.41	70.62

Table 5.2: Measured and predicted mean and standard deviation annual Q_{sol}

The limitation of this Monte Carlo simulation is that for a specified system (with known V_{tank}, $V_{d,ave}$ and A) the prediction from the regression model will be a point estimate and no uncertainty in year-to-year performance will be introduced. This issue can be overcome by sampling distributions for both the independent variables and the regression coefficients.

If the residuals are normally distributed then the regression coefficients can also be assumed to be normal since there is a linear relationship between the residuals and the coefficients (Miller et al. 1990):

$$e_n = y_n - \widehat{y_n} = y_n - (a + b_1 x_{1,n} + b_2 x_{2,n}) \tag{5.1}$$

Here e is the residual value; a, b_1 and b_2 are the least squares estimates of α, β_1, and β_2. The normal scores plot in Figure 5.3 suggests that the residuals are normally distributed therefore the regression coefficients can also be considered normal.

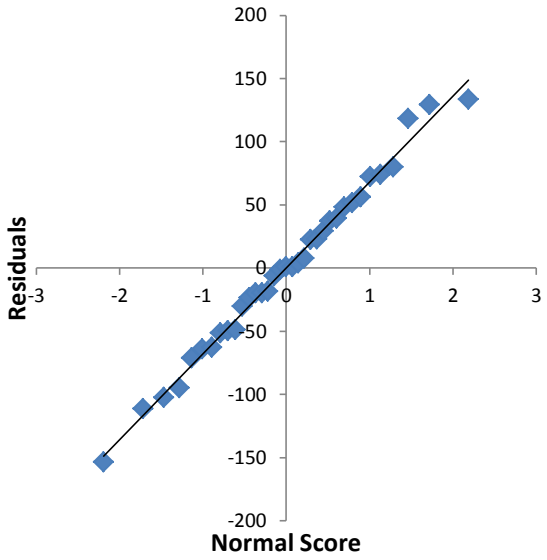

Figure 5.3: Normal scores plot of the residuals

By sampling the distributions of the regression coefficients, uncertainty for the annual solar thermal yield for both a single system and a heterogeneous group of systems can be described. Figure 5.4 shows the measured and predicted distributions for Q_{sol} for heterogeneous group of systems used to develop the regression model in Chapter 4.

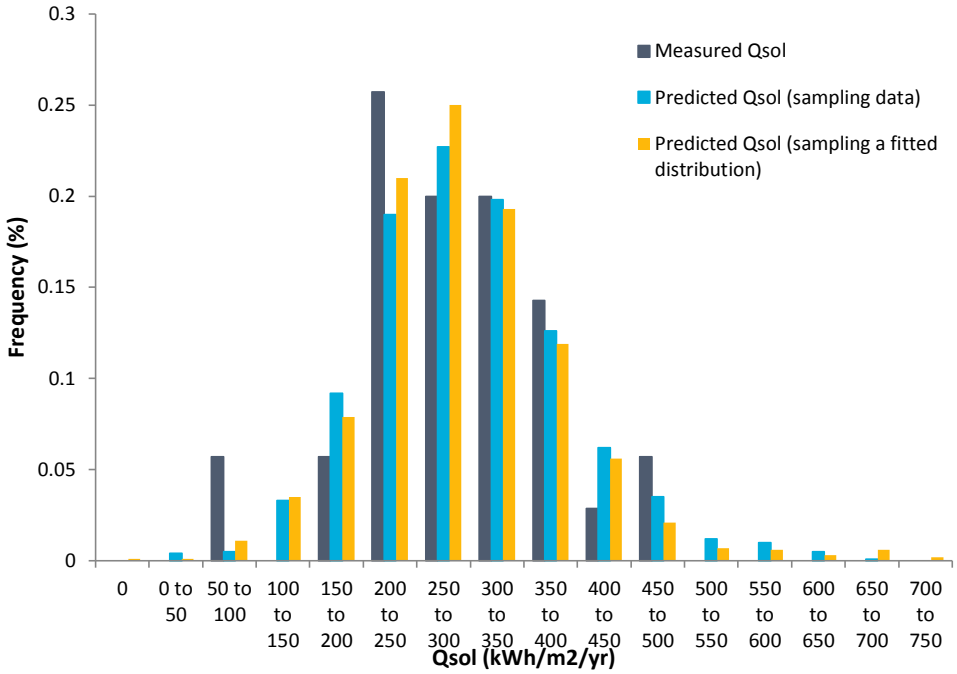

Figure 5.4: Measured and predicted distributions of Q_{sol} using Monte Carlo (1000 simulations); normal distributions for regression coefficients sampled 1000 times

The mean and standard deviations of the distributions in Figure 5.4 are summarised in Table 5.3.

	Mean ($kWh/m^2/yr$)	Standard deviation ($kWh/m^2/yr$)
Measured	285.36	90.49
Sampled from data	295.32	94.13
Sampled from a fitted distribution	290.00	93.31

Table 5.3: Measured and predicted mean and standard deviation for annual Q_{sol}

The R^2 values of the four predicted distributions are summarised in Table 5.4. The higher the R^2 value the closer the predicted distribution is to that produced from measured data.

	R^2
Average regression coefficient Sampled data	0.84
Average regression coefficient Sampled fitted distribution	0.75
Normal regression coefficient Sampled data	0.89
Normal regression coefficient Sampled fitted distribution	0.90

Table 5.4: Goodness of fit of predicted distributions to measured data

The above table suggests that sampling the regression coefficients from a normal distribution and sampling a fitted distribution of $V_{d,ave}$, V_{tank} and A provides the closest fit to the measured Q_{sol} distribution.

Sampling a distribution of the regression coefficients enables the annual performance distribution of the solar thermal system installed at the Bedford case study dwelling to be produced (Figure 5.5).

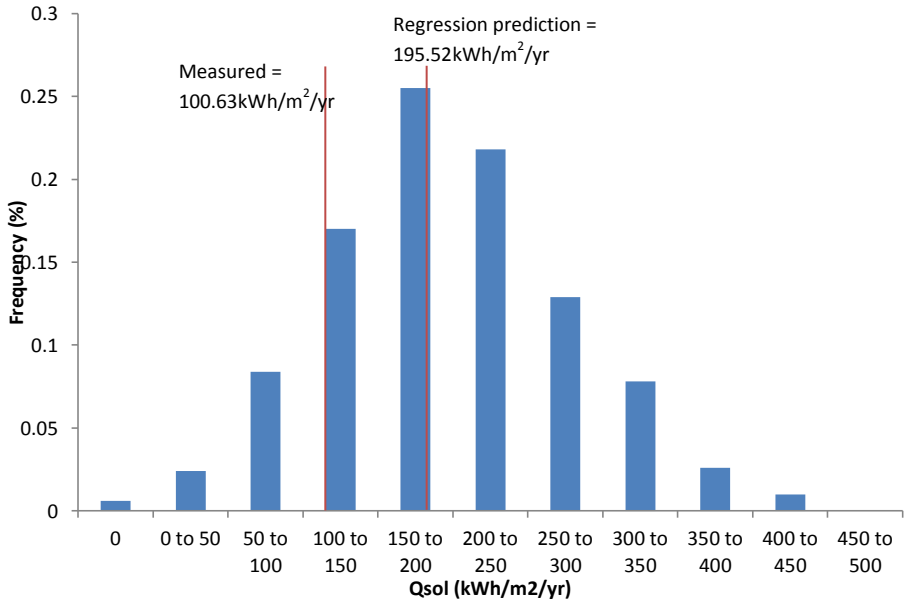

Figure 5.5: Predicted annual performance distribution for Bedford (measured Q_{sol} = 100.63kWh/m^2/yr)

The regression model gave a point estimate of annual yield of 195.52kWh/m^2/yr for the Bedford dwelling which was shown to be 94% higher than the measured value is 100.63kWh/m^2/yr; the Monte Carlo method suggests that there is approximately a 17% probability of obtaining an

annual yield between 100 and 150kWh/m^2/yr and 0.9% chance of achieving the deterministic value predicted by the regression model. Therefore the value obtained by the deterministic method should be taken with caution since it does not represent the likely performance of the system. Thus design and investment decisions made by deterministic models carry considerable risk. This is an improvement on the regression method since it incorporates uncertainty in the predictions as a result of unexplained variance in the model due to both aleatory (unknown values of parameters) and epistemic uncertainties (excluded parameters such as auxiliary timing).

5.5 Strengths and weaknesses of Monte Carlo methods

Monte Carlo methods are shown to propagate uncertainty in the value of model parameters through to the prediction of performance. In this way a deterministic model can be used to provide a probabilistic result that can be used to facilitate decision making in uncertain circumstances. Monte Carlo methods are also shown to improve the simple regression model by incorporating uncertainty in the regression coefficients into the prediction. The uncertainty in the regression coefficients is represented by the normal distribution of said coefficients, which is a result of deviations of data points from the regression line. The spread of data about the regression line indicates unexplained variation in the performance of solar thermal systems by the independent variables; in this way Monte Carlo methods can make up for unexplained variation in Q_{sol} giving predictions of the distribution of performance that are comparable to that measured.

An advantage of Monte Carlo methods is the relative simplicity by which they can be employed: simple Monte Carlo methods can be readily incorporated into a spreadsheet model using random sampling techniques.

In effect Monte Carlo methods are not a separate modelling technique, but are simply a method by which to incorporate an element of uncertainty to a deterministic model. Therefore the limitations of the core modelling method are still of concern when Monte Carlo methods are applied and in some cases may be made worse, for example the increased processing times of dynamic simulations that use Monte Carlo methods; consider a complex model that may take up to 20 seconds to run; if distributions of the uncertain parameters, for which there are many in dynamic models, are sampled 1000 times and therefore 1000 runs of the simulation performed this can drastically increase the runtime of the model.

5.6 Bayesian methods

Bayesian methods for incorporating uncertainty in model inputs make use of Bayes' theorem which allows the posterior distribution of a parameter given evidence, $P(\theta|Z)$, to be determined using prior distributions $P(\theta)$ and $P(Z)$:

$$P(\theta|Z) = \frac{P(Z|\theta)P(\theta)}{P(Z)} \tag{5.2}$$

5.6.1 Bayesian calibration

Bayesian calibration is one such Bayesian method used in building performance modelling (Kennedy & O'Hagan 2001; Booth et al. 2012; Heo et al. 2012). Calibration of a model involves

using measured performance data to adjust the value of model parameters (calibrate) to produce a closer approximation of the model estimates to the observed data – in this way a model can be used to provide closer estimates of building performance to evaluate retrofit strategies that may be employed (Heo et al. 2012); Bayesian calibration employs Bayesian methods to make adjustments to model parameters.

Bayesian methods use the concept of prior and posterior distributions; prior distributions are the probability distributions assigned to an uncertain variable in the presence of no evidence – they are elicited by assumption or expert judgement, or from observed data. These prior distributions, or simply priors, are updated in the presence of evidence to produce posterior distributions or posteriors; the posterior distribution is the probability of a variable θ given evidence e and may also be elicited from data or by expert judgement. A model, such as a building energy model, will have model parameters which are uncertain due to a lack of knowledge about the true values – this is called epistemic uncertainty. In Bayesian calibration prior distributions are assigned to these uncertain parameters, θ. The model is run using the uncertain inputs to obtain n outputs, $y = [y_1, y_2, y_3, \ldots, y_n]$ which are compared to n observations $z = [z_1, z_2, z_3, \ldots, z_n]$; the probability of Z given θ is determined from this comparison. The priors for the uncertain variable, $P(\theta)$, the prior for the observations $P(Z)$, and the distribution $P(Z|\ \theta)$ are substituted into Bayes' theorem above to provide the posterior distribution of the uncertain variables given the observed data, $P(\theta|Z)$. This method is used by (Booth et al. 2012; Heo et al. 2012) and a similar method by (Tian & Choudhary 2012); these posterior distributions can then be used in Monte Carlo analysis of energy models (Booth et al. 2012; Booth & Choudhary 2013). Although the data requirements of Bayesian calibration is less than that of other statistical methods, it requires the number of model runs to be equal to the amount of data collected so that a comparison can be made between the two to find $P(Z|\ \theta)$; even with simplified models, the Bayesian calibration method can be time consuming if the number of parameters to be calibrated is high; therefore a method of finding the most influential parameters is necessary, such as the Morris method (Factorial Sampling Analysis) (Heo et al. 2012; Booth et al. 2012). Despite steps to minimise the processing time, Heo et al. (2012) suggest that the Bayesian calibration method is only applicable to the case of a single building and not suitable for large scale analysis.

5.6.2 Bayesian networks

An alternative Bayesian approach to incorporating uncertainty into model predictions is a probabilistic graphical model called a Bayesian network (BN). Bayesian networks contain both probabilistic and graphical components which enable problems of uncertainty to be represented in an intuitive manner. Bayesian networks have been used in medical diagnostics (Spiegelhalter 1987), fault diagnostics (Lampis 2010; Cai et al. 2014); decision making for safety applications (Washington & Oh 2006), weather prediction (Yacef et al. 2012), energy consumption predictions (Shipworth 2006; Shipworth 2005b; Shipworth 2010) and evaluating building design under uncertainty (Naticchia & Carbonari n.d.; Jensen et al. 2009).

As with Bayesian calibration, Bayesian networks work on the principle of Bayesian inference, which is the process by which the belief in a hypothesis, H, can be updated given evidence, E. Bayesian inference is made possible by Bayes' theorem rewritten here to give the probability of a hypothesis, H, given evidence, E:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \tag{5.3}$$

Bayes' theorem is a result of the concept of conditional probability. Consider two independent events A and B with the joint distribution $P(A, B)$; the probability of any joint event is given by the probability of A, $P(A)$, and the probability of B, $P(B)$:

$$P(A \cap B) = P(A) \times P(B) \tag{5.4}$$

Now consider the scenario where the event B is impacted by the event A, i.e. the events A and B are no longer independent; the probability of a joint event happening in this case is given by:

$$P(A \cap B) = P(A) \times P(B|A) \tag{5.5}$$

It follows from the above that if $P(B|A) = P(B)$, then A and B are independent events. Rewriting this equation gives the fundamental rule of conditional probability:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A,B)}{P(A)} \tag{5.6}$$

This is used to give Bayes' theorem:

$$P(A,B) = P(B|A)P(A) = P(A|B)P(B) \tag{5.7}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{5.8}$$

Consider two uncertain variables, X and Y where the value of Y is dependent on the value of X. The probability that $x = [x_1, x_2, x_3, \ldots, x_n]$ is denoted $P(X)$, likewise the probability that Y takes on the range of values $y = [y_1, y_2, y_3, \ldots, y_n]$ is denoted $P(Y)$. $P(X)$ and $P(Y)$ are prior distributions and may be obtained from data or subjectively assigned a probability by an expert in the field. Since the value of Y is dependent on the value of X the prior, $P(Y)$, may be updated given some evidence about the value of X to give the posterior distribution $P(Y|X)$. Bayes' theorem may be applied to find this posterior distribution using the prior distributions along with the likelihood of $P(X|Y)$ which gives the probability that evidence about the value of X is observed given that the value of Y is observed – in other words it is the probability that the evidence is observed given that the hypothesis about the value of Y is true. Bayes' theorem facilitates the inference of probabilities to occur in both directions i.e. from X to Y and from Y to X; this is a key advantage of Bayesian methods in that they allow predictions as well as diagnoses to be made. The above relationship between X and Y can be represented graphically as a simple, two-node Bayesian network (Figure 5.6).

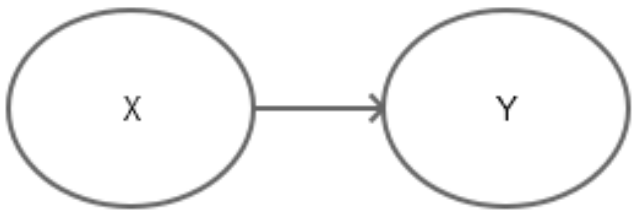

Figure 5.6: Simple two-node BN representing the probabilistic relationship between *X* and *Y*

Bayesian networks are an example of a probabilistic graphical model (PGM); another example of a PGM is a Markov chain (Pearl 1988). A PGM is comprised of nodes and arcs (or edges); nodes represent factors or variables in a problem domain and arcs link those factors that are considered connected.

A Bayesian network is a directed acyclic graph (DAG) meaning that the arcs cannot link two nodes in both directions to form a cycle. DAGs are useful in representing complex problems where causality flows from one variable or factor to another. This concept can be combined with Bayesian inference to provide a graphical representation of the flow of conditional probabilities through connected variables. Therefore Bayesian networks have the ability to describe the causal links between variables in a problem domain and provide the conditional probabilities associated with those variables (Neapolitan 2003).

Each variable, X_i, in a Bayesian network has a conditional probability distribution $P(X_i|Par(X_i))$ where $Par(X_i)$ refers to the parent node of variable X_i. The joint probability distribution of a Bayesian network is determined by the structure of the graph (Shipworth 2010) and calculated by the chain rule which is an extension of the rule of conditional probability (Fenton & Neil 2013):

$$P(X) = P(X_1, \dots, X_n) = P\big(X_n\big|Par(X_n)\big) \times \dots \times P(X_2|X_1)P(X_1) \tag{5.9}$$

$$P(X) = \prod_{i=1}^{n} P\big(X_i|Par(X_i)\big) \tag{5.10}$$

The chain rule reduces the computational demands of the joint probability because it assumes independence between unconnected nodes i.e. it only considers the probability of X_i given $Par(X_i)$ rather than considering all variables in the graph to be dependent on each other (Charniak 1991; Naticchia & Carbonari n.d.; Shipworth 2010). All of the probabilities in the joint distribution must be non-negative and sum to one.

The conditional independence assumptions present in the joint probability distribution are identifiable from three main node structures present in a BN: linear structures, converging structures; and diverging structures (Figure 5.7).

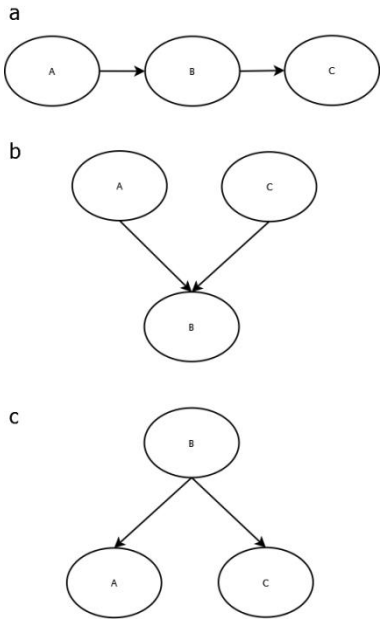

Figure 5.7: a) linear structure; b) converging structure; c) diverging structure

The linear structure is used when a single variable influences another single variable; converging structures denotes a common variable that two other variables may influence; and diverging nodes denote a common influence on two variables by a single variable.

In the linear structure A can influence C when information about B is unknown; if any information about the true value of B is known then this blocks the flow of influence from A to C rendering A conditionally independent of C given B. It is also a key attribute of Bayesian networks that C can affect A given that B is unknown; this is a result of Bayes' theorem and allows a Bayesian network to both predict and diagnose. In the converging structure A and C both influence B. If the value of B is unknown then A does not influence C (and vice versa); however if the value of B is known then any evidence about the A will affect the probabilities contained within C (and vice versa). For example imagine that B is the probability a car fails to start, A is the probability there is no fuel in the tank and C is the probability the battery is dead. Given no information we have a probability that the car won't start conditional on the probabilities of there being no fuel in the tank and the battery being dead. If there is no fuel in the tank this does not affect the probability of the battery being dead and vice versa. However if it is known that the car does not start then the probability of there being no fuel in the tank is dependent on the battery being dead or not; if there is fuel in the tank and the car won't start then the battery must be dead. In this way knowing information about the common effect makes the causes conditionally dependent. In the diverging structure B influences both A and C. If no information about B is known then A can influence C; however if the state of B is known then evidence about A does not affect the probabilities in C. For example if B represents the probability that car won't start, A is the probability that the parent is late for work and C is the probability that the children are late for school. Both A and C are influenced by B; if there is evidence about the car not starting then the probability of being late for work and school are updated. However if it is unknown whether the car started but it is known that the parent is late for work then the probability of the child also being late for school increases and is therefore influenced by A given B is unknown. The flow of influence in a diverging structure is similar to

the linear structure because Bayes' theorem allows the influence to flow in both directions. Table 5.5 summarises the flow of influence in the three node structures found in Bayesian networks.

Structure	Evidence about B	No evidence about B
Linear	A cannot influence C	A can influence C
Converging	A can influence C	A cannot influence C
Diverging	A cannot influence C	A can influence C

Table 5.5: When can A influence C?

The influence that variables have on each other can be interpreted as the flow of causality from cause to effect and the probability distribution associated with the effect variable is conditional on the probability distribution of the cause (parent) variable.

5.7 Approaches to constructing Bayesian networks

Nadkarni & Shenoy (2004) identify two different approaches used to construct Bayesian networks: The first is a data-based approach in which conditional independencies identified in the raw data are used to produce graphical models (Lauría & Duchessi 2007).

5.7.1 Data-based approach

There are six steps in the data-based approach outlined by Lauria and Duchessi (2007):

1. Reduce dimensionality: the amount of data required to estimate variables in a problem domain grows exponentially with the number of variables in the system. The number of variables that can be accurately estimated is, therefore, limited by the sample size. In order to improve the accuracy of estimates it is necessary to reduce the number of variables in the problem – this is what is meant by reducing dimensionality. Several methods can be used including principal component analysis (PCA) and factor analysis;
2. Rescale and discretise variables: Continuous variables will need to be discretised into intervals known as bins. Bins may be a regular width or of different widths, with areas of the distribution with many data being broken down into narrower bins. The amount of data limits the number of bins (states) a variable may have because it is necessary for each bin to be represented in the data for example with a 10-node network of binary nodes where each node has two parents the number of states to estimate is 40, if each node has 4 bins then the total number of states in the network is 160 thus requiring more data to ensure representativeness. Reducing the number of bins in each variable limits the accuracy of the model;
3. Define semantic constraints: If all relationships between variables are to be ascertained from data then the number of variables in the problem directly affects the number of DAGs that can be returned. A problem with 3 variables has 25 possible DAGs where as a problem with 10 variables has $3x10^{17}$ DAGs. If some relationships can be predetermined by the modeller based on expert theory/opinion then the number of DAGs returned by search algorithms is reduced. This is particularly useful in large-scale problems where domain knowledge can be used to reduce the number of node combinations (Neil et al. 2000);

4. Search for candidate models: search and score based, heuristic algorithms search the data for potential networks. This is a problem of optimisation where the optimum DAG has the highest score and represents the best given the data. The score is based on the network's posterior probability distribution. K2 (Cooper & Herskovits 1992) and simulated annealing (Kirkpatrick et al. 1983) are two examples of search algorithms, Bayesian Information Criteria and Info-Geo (Lauría & Duchessi 2007) are two scoring functions;
5. Benchmark candidate models: Candidate models produced by step 4 are evaluated to identify the network which best describes the domain or process. A commonly used criteria for evaluation is the predictive accuracy of the network;
6. Apply the selected model: the best performing network model is used to make probabilistic inferences about the domain;

Data-based approaches require large amounts of data to overcome the issue of granularity which is a function of the number of available cases identifiable in the data - more cases in the data improve the granularity. A case represents a combination of different variable states that can occur in the domain.

5.7.2 Knowledge-based approach

The second approach is a knowledge-based approach and makes use of the knowledge held by domain experts to identify causal relationships in the problem domain. The knowledge based approach is particularly useful when there is a dearth of data related to a problem domain and causal knowledge is critical. This approach first constructs the graphical structure then applies data; this allows expert knowledge to direct the construction of the network and therefore the data requirements are far less than that of the data-based approach. In addition the application of causal reasoning means that parsimony in the network can be achieved without the use of techniques such as PCA and factor analysis. A similar approach that makes use of the cause-effect relationships elicited from literature and experts is provided by Shipworth in which a method known as realist synthesis is used to construct the graphical structure (or architecture) of the BN (Shipworth 2005a). Briefly, realist synthesis is an alternative method of evaluation in which the data collection phase is focused on finding evidence to support theories that have been put forward by the researcher (Pawson et al. 2004); its primary focus has been in decision support for policy guidance rather than decision making, which Shipworth argues makes it a complementary process to the development of BNs used in decision support situations.

Due to the lack of recent, in-situ performance data for solar thermal systems in the UK and the dependency of performance on causal factors such as weather, installation and usage characteristics the knowledge-based approach to developing the BN will be implemented. The knowledge-based approach that will be used in this research follows the systematic procedure outlined by Nadkarni & Shenoy in which causal knowledge elicited from experts will be combined with data to produce a Bayesian network for the probabilistic modelling of solar thermal system performance. This procedure was also followed by (Sedki & Bonneau de Beaufort 2012) to analyse the fishing activities in the West of France. A full description of this process is provided in Chapter 6.

5.8 Discussion

Bayesian networks provide an intuitive means of representing a causal problem domain open to uncertainty and provide both a qualitative and quantitative description of the system. Relationships are obtained from data rather than deterministic equations (Shipworth 2010) making them ideal for modelling problem domains such as the performance of solar thermal systems which is influenced by uncertain variables such as user behaviour, weather patterns and system configuration; not only are these variables open to uncertainty but are not linked by a single equation to determine the performance of solar thermal systems (Mathioulakis et al. 2012). As with Monte Carlo simulations the interactions between variables are maintained, but in addition the individual effects of variables may also be observed due to the marginalisation of nodes; this therefore allows sensitivity analysis to be performed to determine the most influential factors after the model has been constructed. With Bayesian calibration and Monte Carlo methods sensitivity analysis is performed prior to uncertainty quantification in order to reduce the number of uncertain variables to those that are the most influential on the system in an effort to reduce the computational demands of the analysis. Once a Bayesian network has been constructed the conditional independence assumptions resulting from the graphical structure allows the joint probability of the graph to be determined with relative speed thanks to the chain rule. Bayesian inference and marginalisation of nodes in the BN allows a homogenous group of systems or single system to be modelled and therefore enables small scale analysis to be performed whilst graphically displaying the aleatory uncertainty i.e. the random variation in variables of a homogenous group or single system. By not marginalising nodes, i.e. modelling a group of heterogeneous systems, large scale analysis can be performed which graphically represents the uncertainty in the problem domain due to a lack of knowledge about the true values of variables in the system (epistemic uncertainty).

5.9 References

Booth, a. T. & Choudhary, R., 2013. Decision making under uncertainty in the retrofit analysis of the UK housing stock: Implications for the Green Deal. *Energy and Buildings*, 64, pp.292–308.

Booth, A.T., Choudhary, R. & Spiegelhalter, D.J., 2012. Handling uncertainty in housing stock models. *Building and Environment*, 48, pp.35–47.

Bordass, B., Cohen, R. & Field, J., 2004. Energy Performance of Non-Domestic Buildings: Closing the Credibility Gap. In *Building Performance Congress*.

Burhenne, S. et al., 2013. Uncertainty quantification for combined building performance and cost-benefit analyses. *Building and Environment*, 62, pp.143–154.

Cai, B. et al., 2014. Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network. *Applied Energy*, 114, pp.1–9.

Carbon Trust, 2011. Closing the Gap: Lessons learned on realising the potential of low carbon building design.

Cayre, E. et al., 2011. There are people in the house ! how the results of purely technical analysis of residential energy consumption are misleading for energy policies. In *European Council for an Energy Efficient Economy (ECEEE) Summer School*. pp. 1675–1683.

Charniak, E., 1991. Bayesian Networks without Tears. *American Association for Artificial Intelligence*.

Cooper, G.F. & Herskovits, E., 1992. A bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9, pp.309–347.

de Wilde, P., 2014. The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction*, 41, pp.40–49.

de Wit, M.S., 1995. Uncertainty analysis in building thermal modelling. In *Proceedings of international building performance simulation association*.

de Wit, M.S. & Augenbroe, G., 2002. Analysis of uncertainty in building design evaluations and its implications. *Energy and Buildings*, 34(9), pp.951–958.

Domínguez-Muñoz, F. et al., 2012. Design of solar thermal systems under uncertainty. *Energy and Buildings*, 47, pp.474–484.

Evarts, J.C. & Swan, L.G., 2013. Domestic hot water consumption estimates for solar thermal system sizing. *Energy and Buildings*, 58, pp.58–65.

Fenton, N. & Neil, M., 2013. *Risk Assessment and Decision Analysis with Bayesian Networks* First., CRC Press.

Heo, Y., Choudhary, R. & Augenbroe, G. a., 2012. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47, pp.550–560.

Hopfe, C.J. & Hensen, J.L.M., 2011. Uncertainty analysis in building performance simulation for design support. *Energy and Buildings*, 43(10), pp.2798–2805.

Jensen, K., Toftum, J. & Friishansen, P., 2009. A Bayesian Network approach to the evaluation of building design and its consequences for employee performance and operational costs. *Building and Environment*, 44(3), pp.456–462.

Jordan, U. & Furbo, S., 2005. Thermal stratification in small solar domestic storage tanks caused by draw-offs. *Solar Energy*, 78(2), pp.291–300.

Jordan, U. & Vajen, K., 2000. Influence of the DHW load profile on the fractional energy savings: A case study of a solar combi-system with TRNSYS simulations. *Solar Energy*, 69, pp.197–208.

Kennedy, M.C. & O'Hagan, A., 2001. Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society*, 63(3), pp.425–464.

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P., 1983. Optimization by Simulated Annealing. *Science*, 220(4598), pp.671–680.

Knudsen, S., 2002. Consumers' influence on the thermal performance of small SDHW systems—Theoretical investigations. *Solar Energy*, 73(1), pp.33–42.

Lampis, M., 2010. *Application of Bayesian Belief Networks to System Fault Diagnostics*. Loughborough University.

Lauría, E.J.M. & Duchessi, P.J., 2007. A methodology for developing Bayesian networks: An application to information technology (IT) implementation. *European Journal of Operational Research*, 179(1), pp.234–252.

Lee, P. et al., 2013. Probabilistic risk assessment of the energy saving shortfall in energy performance contracting projects–A case study. *Energy and Buildings*, 66, pp.353–363.

Lomas, K.J. & Eppel, H., 1992. Sensitivity analysis techniques for building thermal simulation programs. *Energy and Buildings*, 19(1), pp.21–44.

Loose, A. et al., 2011. Field test and performance monitoring of combined solar thermal and heat pump systems. In *ISES Solar World Congress*. Kassel.

Lu, Y., Huang, Z. & Zhang, T., 2013. Method and case study of quantitative uncertainty analysis in building energy consumption inventories. *Energy and Buildings*, 57, pp.193–198.

Macdonald, I. & Strachan, P., 2001. Practical application of uncertainty analysis. *Energy and Buildings*, 33(3), pp.219–227.

Mathioulakis, E., Panaras, G. & Belessiotis, V., 2012. Uncertainty in estimating the performance of solar thermal systems. *Solar Energy*, 86(11), pp.3450–3459.

Menezes, A.C. et al., 2012. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97, pp.355–364.

Miara, M. et al., 2011. *Analysis and Evaluation of Heat Pump Efficiency in Real-life Conditions*, Freiburg: Fraunhofer Institute for Solar Energy Systems.

Michaelides, I.M., 1993. *Computer simulation and optimisation of solar heating systems for Cyprus*. University of Westminster.

Miller, I., Freund, J.E. & Johnson, R.A., 1990. *Probability and Statistics for Engineers* Fourth., Prentice Hall.

Morant, M., 2012. *CEW1005 The Performance Gap - Non Domestic Building: Final Report*,

Naticchia, B. & Carbonari, A., Causal modelling based on bayesian networks for preliminary design of buildings. , pp.293–316.

Neapolitan, R.E., 2003. *Learning Bayesian Networks* First., Prentice Hall.

Neil, M., Fenton, N. & Nielsen, L., 2000. Building Large-Scale Bayesian Networks. *The Knowledge Engineering Review*, 15, pp.257–284.

Pawson, R. et al., 2004. Realist synthesis: an introduction. *ESRC Research Methods Programme*.

Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems - Networks of Plausible Inference* First., San Francisco: Morgan Kaufmann Publishers, Inc.

Sedki, K. & Bonneau de Beaufort, L., 2012. Cognitive Maps and Bayesian Networks for Knowledge Representation and Reasoning. In *24th International Conference on Tools with Artificial Intelligence*. pp. 1035–1040.

Shipworth, D., 2006. *Bayesian Network modelling of home energy use*, University of Reading, Reading.

Shipworth, D., 2005a. Synergies and conflicts on the landscape of domestic energy consumption : beyond metaphor. In *ECEEE*. pp. 1381–1391.

Shipworth, D., 2005b. Synergies and conflicts on the landscape of domestic energy consumption: beyond metaphor. In *ECEEE 2005 Summer Study*. pp. 1381–1391.

Shipworth, D., 2010. The vernacular architecture of household energy models. In *Grune-Yanoff, T. and Morgan, M. Proceedings of Modelling in the Social Sciences: Interdisciplinary Comprison*. pp. 1–14.

Spiegelhalter, D.J., 1987. Probabilistic Expert Systems in Medicine: Practical Issues in Handling Uncertainty. *Statistical Science*, 2(1).

Tian, W. & Choudhary, R., 2012. A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater London. *Energy and Buildings*, 54, pp.1–11.

Wang, L., Mathew, P. & Pang, X., 2012. Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building. *Energy and Buildings*, 53, pp.152–158.

Washington, S. & Oh, J., 2006. Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: example applied to at grade railroad crossings in Korea. *Accident Analysis and Prevention*, 38(2), pp.234–47.

Weber, P. et al., 2012. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, 25(4), pp.671–682.

Yacef, R., Benghanem, M. & Mellit, a., 2012. Prediction of daily global solar irradiation data using Bayesian neural network: A comparative study. *Renewable Energy*, 48, pp.146–154.

Chapter 6

Knowledge-Based Bayesian Network to Model In-Use Performance of STS

6.1 Introduction

A knowledge-based Bayesian network (BN) to quantify the uncertainty in the performance of solar thermal systems resulting from uncertainty in system variables related to user behaviour, ambient conditions and system configuration has been constructed. The BN approach was taken for the following reasons:

- Provides a graphical representation of the causal factors affecting solar thermal system performance thus enabling a deeper level of understanding about the system domain;
- Incorporates qualitative and quantitative data obtained from field trials of in-situ performance, expert opinion elicited from peer reviewed literature and dynamic simulations;
- Resilient to missing data values and limited datasets of varying quality (Shipworth 2005);
- Exhibits the uncertainty inherent in the system domain related to both the system parameters and performance metrics;
- Propagates uncertainty from the system parameters to system performance nodes given evidence about the system being analysed;
- Facilitates prognostic and diagnostic inference via Bayes' theorem;

The method of construction follows that of (Nadkarni & Shenoy 1999; Nadkarni & Shenoy 2004), in which a causal map was first produced using textual analysis of literature to identify causal factors affecting solar thermal system performance. The conditional probabilities were obtained from data making use of the expectation maximisation (EM) algorithm in Norsys Netica. In addition, because the relationship between system performance and DHW draw timing was unobtainable from field trial data collected by the EST (EST 2011), it was elicited from a series of simulations conducted in TRNSYS. This chapter shows the resulting network that was produced and demonstrates how the network can be used to:

- Predict the performance of solar thermal systems on the small (individual system) scale and large (heterogeneous group of systems) scale;
- Diagnose system performance using backward propagation;

The aim is to demonstrate a novel probabilistic approach to modelling solar thermal system performance which considers and quantifies the uncertainty associated with the technical and non-technical factors shown to influence system performance.

6.2 Causal maps

Causal knowledge elicited from experts will be represented in a graphical model called a causal map. Scavarda et al. (2004) describe the different types of causal maps used in operations management each with a different graphical structure and function (Scavarda et al. 2004). Some of these include Ishikawa diagrams, issues trees, fault trees, FMEA, and cause and effect diagrams. In this research the term causal map will be used to specifically refer to the cause and effect diagrams of Scavarda et al. (2004).

Causal maps have been used in health sciences (Greenland & Brumback 2002), operations management (especially in identifying root causes for quality management processes such as Six Sigma) (Scavarda et al. 2004), control systems (Francis & Leitch 1989), management science and policy analysis (Roos & Hall 2014).

A causal map has three constituent parts (Nadkarni & Shenoy 2004):

- Causal concept: A single term used to represent an attribute, issue, factor or variable of a system – concepts reduce the amount of cognition required to store and organise knowledge (cognitive economy) and allow conclusions to be drawn;
- Causal connection: A link between two causal concepts represented in a causal map by a directed arrow from cause to effect;
- Causal value: the strength of the causal connection linking two concepts and can be used to indicate whether a connection exists and/or whether it is positive or negative – how the causal value is determined depends on the method of eliciting expert knowledge;

Causal maps are a useful means to construct BNs because both are a graphical structure with arcs and nodes depicting an expert's estimation of the occurrence of certain event or action leading to a particular result.

6.3 Developing a Bayesian network from a causal map

The systematic procedure for creating a BN from a causal map according to Nadkarni & Shenoy is as follows:

1. Data elicitation;
2. Derivation of causal map;
3. Modification of causal map to construct Bayesian graphical structure;
4. Assignment of conditional probabilities;

6.3.1 Data elicitation

Techniques for eliciting causal knowledge from experts can be categorised as structured or unstructured. Structured methods involve providing the experts with a list of predetermined causal concepts with which they are asked to specify the direction, sign and value of the

relationship; structured methods are therefore used to confirm and validate causal knowledge. Unstructured methods are exploratory in nature and can provide a deeper understanding of the system or problem domain; typically in-depth qualitative and open ended questions are posed to the expert in unstructured elicitation methods from which a narrative is created which can be textually analysed to extract causal knowledge. Brainstorming provides an alternative method for elicitation of expert knowledge but requires experts to be grouped together to be most effective (Scavarda et al. 2004). Interviews are a time consuming approach that require a certain level of skill on the part of the modeller to perform well. However there is another source of expert knowledge in the form of a narrative that can be made use of; namely journal papers. In this research a comprehensive survey of literature pertaining to solar thermal system performance will be made in order to identify causal concepts and value the causal connections using textual analysis. This is a similar approach to that used by Washington and Oh to determine the most effective method for improving safety at railroad crossings in South Korea (Washington & Oh 2006).

6.3.2 Derivation of the causal map

Nadkarni & Shenoy (2004) describe four steps in the procedure for deriving a causal map from expert data in the form of a narrative:

1. Identify causal statements in the narrative;
2. Construct a raw causal map;
3. Design a coding scheme;
4. Convert the raw causal map into a coded causal map;

6.3.2.1 Identify causal statements in the narrative

Causal statements are statements that include a cause phrase and an effect phrase linked by a causal connector. Identifying the causal connector is an important part of identifying a causal statement. Examples of some causal connectors are: if-then; because; depends on; so; as; therefore etc. In addition some words not only indicate the causal statements but also the sign of the connection; for example the word “increases” when used to connect a cause and effect phrase indicates that the causal connection is positive.

Causal statements can be identified manually or automatically. The manual procedure has the advantage that humans are able to identify new and unusual causal connectors whilst reading the narrative. Unfortunately the manual procedure is a long and laborious process. In the literature surveyed for causally mapping solar thermal system performance, many different and unexpected causal connectors were identified, and human judgement was required to determine whether a particular sentence was indeed a causal statement. In addition the use of graphs in literature provides a visual representation of causal concepts (although care must be taken not to confuse correlation with causality) and so the manual procedure was decided to be the most effective approach to analysing the text.

6.3.2.2 Construct a raw causal map

Once the causal statement has been identified and broken down into its three constituent parts: cause phrase, effect phrase and causal connector these parts can be arranged into a raw causal map.

Figure 6.1 shows a selection of causal statements identified from literature and the resulting raw causal map showing **Causal Phrase-Causal Connector-Effect Phrase** relationships in textual form.

Causal Statement	Causal Phrase	Causal Connector	Effect Phrase
Low storage tank volume means the temperature difference between collector and storage tank is maximised	Low tank volume	means	maximised temp difference
Lower collector area would provide a lower proportion of a household's hot water requirements	Lower collector area	would provide	lower proportion of hot water
Systems where most hot water energy is used in the afternoon perform better than those using in the morning	hot water usage patterns	affects	performance
Daily solar irradiance affects daily system performance	Daily solar irradiance	affects	performance
Mains cold water temperature affects daily system performance	Mains cold water temp.	affects	performance
Solar fraction depends critically on the hot water demand place on the system	Hot water demand	depends on	solar fraction
…run off pattern will influence the performance of the system	run off pattern	influence	performance
Wide variations in the daily draw pattern can significantly reduce the sytem thermal performance	daily draw pattern	reduce	performance
Increasing the size of the gas boiler or thermal store reduces the operational temperature (of the store)	increasing size of thermal store	reduces	operational temperature
thermal stratification in the solar tank has a great impact on the thermal performance of SDHW systems	thermal stratification	impact	thermal performance
Climate and season affect event schedule through mains temperature	Climate	affect	mains temperature
Climate and season affect event schedule through mains temperature	Season	affect	mains temperature
Refer to Fig. 4 in Furbo et al. (2005)	increase in DHW volume	increases	solar yield
The average tank temperature is a function of load temperature, collector area and tank volume	tank volume	function of	average tank temperature

Figure 6.1: Causal statements and resulting raw causal map identified from a selected sample of literature

As is clear in some causal statements in Figure 6.1 the cause and effect phrases are reversed in the causal statement due to the grammatical rules of the English language; for example **hot water demand-depends on-solar fraction**: clearly the hot water demand is not dependent on the solar fraction but rather the solar fraction is dependent on the hot water demand - in order to maintain the correct flow from cause to effect it is sometimes necessary to ignore grammatical rules.

6.3.2.3 Design a coding scheme

Language used in the narrative is not suitable for use in a causal map due to the complexity of written word; therefore it is necessary to simplify the cause and effect phrases into coded concepts, thus providing a key component of a causal map. A coded concept is a term that describes the cause phrase or effect phrase for example the cause phrase '**hot water demand'** can be coded into '*DHW volume'*.

The causal statements in Figure 6.1 have been broken down into the cause and effect phrases with a causal connector; Figure 6.2 shows the corresponding coded concepts.

Coded Concepts	Coded Concepts
Cause	**Effect**
Tank volume	Temp. difference
Collector area	Solar fraction
DHW usage	Solar yield
Irradiance	Solar yield
Cold water temp.	Solar yield
DHW volume	Solar fraction
DHW usage	Solar yield
DHW usage	Solar yield
Tank volume	Tank temp.
Stratification	Solar yield
Climate	Cold water temp.
Time of year	Cold water temp.
DHW volume	Solar yield
Tank volume	Tank temp.

Figure 6.2: Coded concepts corresponding to the causal statements in Figure 6.1

When using an unstructured approach to data elicitation the number of coded concepts is unknown to the modeller as is the case in this research; a list of the all the coded concepts identified are given below:

- Auxiliary control
- Auxiliary input
- Auxiliary rating
- Boiler efficiency
- Climate
- Cold water temp.
- Collector area
- Collector losses
- Collector performance
- Collector properties
- Collector temp.
- Collector Type
- Control
- Design
- DHW energy
- DHW usage
- DHW volume
- Distribution losses
- Duration of operation
- Efficiency
- Electric shower
- External temp.
- Flow rate
- Heat transfer
- House archetype
- Inlet design
- Installation quality
- Insulation
- Irradiance
- Irradiation
- Location
- Mixing (Stratification)
- Occupants
- Orientation
- Performance (Solar yield)
- Performance characteristics
- Pipe length
- Pitch
- Return temp.
- Set point temp.
- Solar fraction
- Solar yield
- Storage design
- Storage losses
- Stratification
- System characteristics
- System interaction
- Tank aspect ratio
- Tank geometry
- Tank height
- Tank temp.
- Tank volume
- T_a-T_{main}
- Temp. difference
- Temp. rise
- Time of year
- VA ratio
- Weather conditions

Performance is taken to mean solar yield since this is the root of all performance metrics associated with solar thermal systems and the primary output of such systems. In addition some concepts are antonyms of each other and so can be combined into a single term following the requirement of a causal map to be parsimonious (Scavarda et al. 2004). For example mixing in the tank is the opposite of stratification in the tank, in other words mixing is a special case of stratification in which the level of stratification approaches zero. For this reason the coded concept *'Mixing'* was incorporated into the coded concept *'Stratification'*.

By developing coded concepts it is possible to determine the adequacy of the sample size i.e. how many occurrences of the same connections between coded concepts have been identified. It also allows the point of redundancy to be established, the point at which further elicitation provide no additional concepts (Scavarda et al. 2004).

6.3.2.4 Convert the raw causal map into a coded causal map

Using the coded concepts a coded causal map can be constructed with nodes given by the coded concepts and arcs replacing the causal connectors in the raw causal map. This gives the first graphical structure which must be modified in the next step to be transformed into a Bayesian graphical model.

Figure 6.3 shows an example of a coded causal map that was created using the coded concepts presented in Figure 6.2. The map does not show all causal concepts or connections for the entire solar thermal system domain; the completed coded causal map can be found in Appendix A.

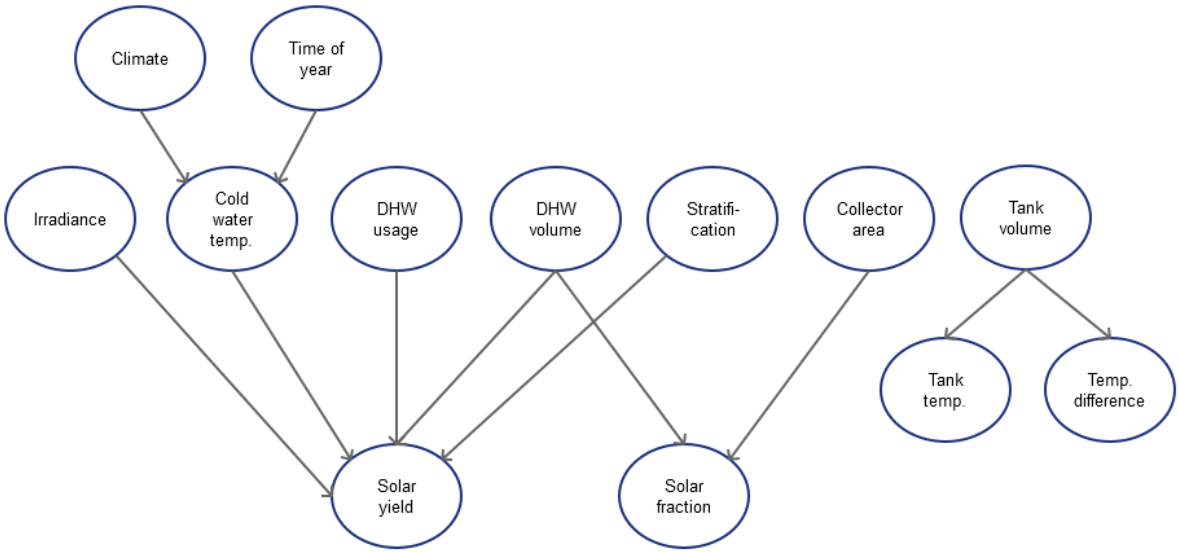

Figure 6.3: Example of a coded causal map developed from Figure 6.2

6.3.3 Modification of causal map to construct Bayesian graphical structure

There are four issues to consider when transforming a causal map into a Bayesian network. These issues are related to:

- Conditional independencies that must be represented in the Bayesian network but may not necessarily be depicted in the causal map;

- Ensuring that the direction of causality is correct and well-reasoned;
- Distinguishing between direct and indirect relationships;
- Eliminating circular relationships in order to adhere to the acyclic requirement of a BN.

Following these considerations it may be possible for the coded concepts in the causal map to be simplified, combined or eliminated provided sound reasoning is used and the conditional independencies are maintained.

There are two widely used structured methods for addressing these issues: structured interviews and adjacency matrices. In structured interviews the causal concepts are presented with which the interviewee is asked to provide the direction of the relationship (assuming there is one) and the sign (positive or negative). Adjacency matrices are another structured method for making the causal map compatible with a Bayesian network. This method is very similar to the structured interview except that the concepts are presented in a matrix with rows representing causes and columns effects. The interviewee must then assign a '0' (for no relation), '+' (positive relationship) and '-' (negative relationship) in each of the matrix cells.

In this research the structured interview approach provides a simple and clear method for summarising the causal relationships identified in the development of the causal map. Using this method, direct and indirect relationships were also identified and considerations about the availability of data for certain parameters were also taken into account. Before the structured interview is presented, background theory into the four modelling issues mentioned above is necessary.

6.3.3.1 Conditional independency, D-maps and I-maps

Graphically, causal maps look very similar to Bayesian networks – both are directed graphs with arcs and nodes that display relationships between system variables. However graphical models may be categorised as dependence-maps (D-maps) or independence-maps (I-maps) and this distinction is important when developing a Bayesian network from a causal map. In a D-map the arcs connecting nodes indicate that these nodes are dependent; however the lack of an arc between nodes does not necessarily mean that they are not dependent. Conversely the lack of an arc between nodes in an I-map indicates that these nodes are indeed conditionally independent given other variables, but the presence of an arc between nodes does not necessarily mean that they are related. It may be that further exploration of the system's behaviour reveals that these concepts are dependent in some way, or that the expert from whom the relationships are elicited may not have explicitly stated that the concepts are dependent, but does not necessarily believe them to be independent. Therefore a causal map is a D-map but this does not automatically imply it is an I-map.

On the other hand, in a Bayesian network conditional independence is an important concept when making inferences because it indicates which variables have information that can update the probabilities of connected variables. Conditional independence is represented by an absence of an arc from one node to successive nodes and therefore Bayesian networks must be I-maps. The issue when transforming causal maps into Bayesian networks is to ensure that the causal map is not only a D-map, but also an I-map.

6.3.3.2 Direction of causality

There are two types of reasoning used by humans to perceive cause-effect relationships: deductive and abductive reasoning. Deductive reasoning takes the direction of reason from cause to effect whereas abductive reasoning is when individuals reason from effect to cause. For example a faulty coil pack in a car engine causes the engine to misfire – this is deductive reasoning; conversely knowing the effect of a misfiring engine leads us to reason that a faulty coil pack is probably the cause – this is abductive reasoning. Of course there are many other causes associated with this issue such as damaged spark plugs, HT leads or engine management faults. Our knowledge and experience allows us to update our beliefs as to what the cause is and to find the most likely one; for example if we know that the coil pack is new then our belief that this is the cause of a misfiring engine is reduced and the probability of it being any of the other causes increases; a more experienced individual or an expert mechanic has collected many data over the years and their beliefs will take this into account (the prior distributions of likely causes will be based on many more data points compared to a layman who may be basing their reasoning on one prior scenario). This is why Bayesian networks are such powerful tools when used in decision making under uncertainty because they mimic the reasoning process of humans which is very good at handling uncertain data.

Although Bayesian networks allow the flow of inference to move from effect to cause (courtesy of Bayes' theorem) it is important to ensure that when the network is being created, abductive reasoning is eliminated and that arcs move from cause to effect. This helps to minimise the chance of including circular relationships in the graph and provides a more accurate portrayal of the problem domain.

6.3.3.3 Direct and indirect relationships

The coded concepts identified in the literature may be linked directly or indirectly. The causal map does not necessarily make a distinction between these two types of connection; two concepts identified in the literature as being connected may not be a direct connection but linked through some other variable. For example the link between *DHW volume* and *Solar yield* is not necessarily a direct link but rather linked by the effect of DHW draw volume on the temperature in the bottom of the storage tank. Conditional independencies are reliant on the distinction between direct and indirect relationships. For example considering that the solar yield is dependent on the DHW volume consumed if the BN can be modified to include the variable *Tank temperature* then the direct connection between *DHW volume* and *Solar yield* can be removed. This would mean that if we know the tank temperature then any information about the amount of hot water consumed would do nothing to update the distribution of *Solar yield* – in this way the conditional independency in the network is affected by the indirect connection between *DHW volume* and *Solar yield*.

Fully identifying the direct and indirect causal relationships between system variables gives a better understanding of the behaviour of the system – it shows whether two system parameters are actually related by a third intermediate variable. If two nodes are connected indirectly through an intermediary, then the direct connection between the two nodes is redundant and can therefore be removed in order to reduce the complexity of the network (Figure 6.4).

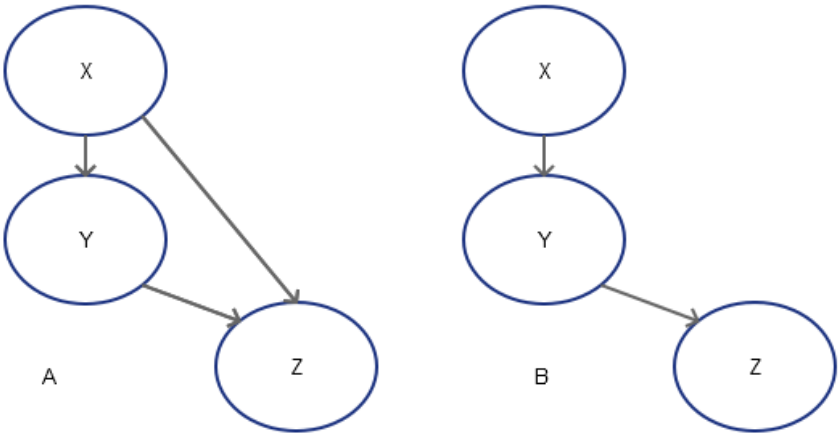

Figure 6.4: a) direct and indirect relationships between X and Z; and b) indirect relationship between X and Z through Y

In the solar thermal system domain the temperature of the tank is an influential variable on the performance of the system. The temperature of the tank can be affected by the amount of hot water drawn, the cold water temperature, the set point temperature, the level of stratification, and the amount of insulation. It is also affected by the amount of heat delivered by the solar collector and so it is possible to incorrectly reason that high tank temperatures in the bottom of the tank lead to higher solar yields. It is therefore important to consider the temperature in the bottom of the tank prior to the operating period of the solar thermal system. However reductions in the temperature of the tank during the middle of the day due to hot water consumption mean that the initial tank temperature is not the only consideration. Tank temperature is therefore a very complex and dynamic system variable and so is not suitable for inclusion in the Bayesian network. Instead the direct connection between the aforementioned system parameters and solar yield will be maintained at the expense of not fully describing the relationships in the system. Furthermore data related to the temperature in the storage tank was not monitored for a large proportion of the systems included in the EST field trial, which will provide a large part of the data to construct the BN; only 6 out of 88 systems had tank temperatures measured.

Any direct relationships that were broken down into indirect relationships between nodes in the literature surveyed were identified at the time of creating the raw causal map.

6.3.3.4 Circular relationships

Bayesian networks are directed acyclic graphs and therefore may not contain circular relationships between nodes. On the other hand, causal maps may identify system variables as being dependent on each other. When transforming a causal map into a Bayesian network it is important to remove any circular relationships firstly by considering whether the relationship is a result of deductive and abductive reasoning. There are some instances where a circular relationship occurs due to temporal effects in which the current state of a variable affects another which in turn affects the future state of the previous variable. For example, assuming that the amount of caffeine consumed in a day will affect the number of hours of sleep an individual receives that night, then the number of hours sleep received that night may affect the next day's caffeine intake. This suggests a circular relationship between caffeine intake and

hours of sleep. In order to account for this in a Bayesian network the relationship can be disaggregated into two time frames t_1 and t_2. This breaks the cyclic relationship between two nodes into an acyclic relationship between three nodes (Figure 6.5).

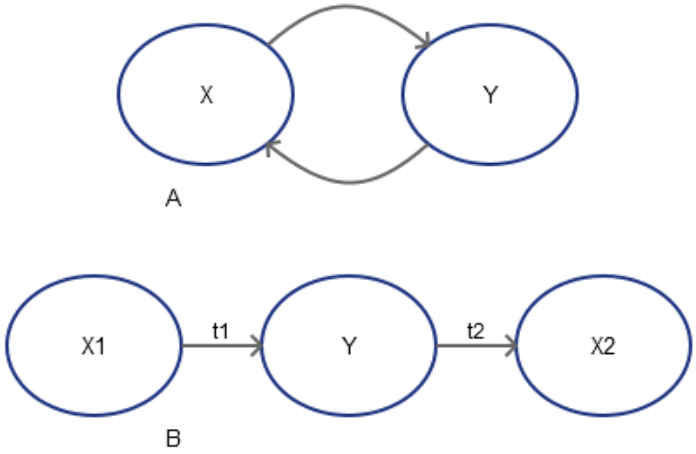

Figure 6.5: a) circular relationship between X and Y and; b) disaggregated relationship into two time frames t_1 and t_2

Alternatively, rather than representing two time frames it is possible to retain one of the time frames and exclude the other from the graph. In the modelling of solar thermal system performance, a circular relationship appears to exist between *Solar yield* and *Auxiliary input*. As the solar yield increases the amount of auxiliary input required decreases; as the auxiliary input increases the amount of solar yield decreases due to an increase in tank temperature. It is also worth considering that during the winter months the auxiliary input is high and the solar yield is low; these two facts are not related i.e. the high auxiliary in the winter does not cause low solar yield at this time, rather the solar yield is low due to the time of year and this increases the demand for the auxiliary heating. From the TRNSYS simulations (Chapter 4) the direction of the causal relationship between *Solar yield* and *Auxiliary input* was determined to be unidirectional and travelled from *Solar yield* to *Auxiliary input*. This is because the effects of having thermal storage allows the heat collected by the solar thermal system to be carried forward to the next day and thus reduces the following day's auxiliary requirement. This may not be the case if the auxiliary heater is poorly controlled and in competition with the solar collector (Forward & Roberts 2008; EST 2011). In the models, the auxiliary heating was thermostatically controlled and free to turn on any time between 06:00 and 20:00. This control strategy meant that the auxiliary heater would react to the consumption timing of the draw by activating when necessary. In order to model a poorly controlled system it would be necessary to activate the auxiliary at time when it is unnecessary (i.e. when the tank is up to the set-point temperature); this is in contrast to the thermostatic control used in domestic heating systems. Therefore the causal relationship goes from *Solar yield* to *Auxiliary input*.

6.3.3.5 Structured approach to developing a Bayesian graphical model

The structured interview approach was used to identify the four modelling considerations detailed above. It also allowed certain variables to be combined and to on the whole simplify the original coded causal map.

6.4 Assignment of conditional probabilities

The causal map provides the graphical structure of the Bayesian network and represents the qualitative knowledge of the problem domain elicited from experts. Causal maps make no allowance for the incorporation of uncertainty associated with each of the variables in the graph because the primary focus of these maps is to analyse the causal relationships in the system or problem domain. A function of Bayesian networks is to facilitate inferences about the different variables in the graph to be made. For this function to be possible, the uncertainties associated with each of the variables need to be quantified; these uncertainties are conditional on the states and values the parent nodes take.

Nadkarni & Shenoy (2004) describe the method of probability assessment, or probability encoding, which enables experts to provide their degree of belief of a variable conditional on the parents. In this method the state space of the variables needs to be determined followed by the conditional probabilities. This method is useful when relying solely on expert opinion and input; however in this research the conditional probabilities will be determined from the raw data collated from field trials and literature. This will make use of the algorithms provided in Norsys Netica, the commercial Bayesian network software used to construct the BN, and will be described more fully in the following section.

6.5 Norsys Netica

Netica is commercially available Bayesian network development software created by Norsys (Norsys 1995). The Bayesian networks presented in this research were developed in Netica because of the following software features (Norsys 1995):

- Intuitive GUI for creating networks with nodes and arcs;
- Tabular CPTs allowing easy manual entry of probabilities;
- Ability to incorporate deterministic functions;
- Repertoire of statistical functions available;
- Ability to import data from Excel spreadsheets;
- Ability to learn CPTs using a choice of three different algorithms with imported data;
- Inexpensive $285 for single user price;
- Completes Bayesian inferences very quickly compared to other software on the market;

6.6 Nodes, states and case files

The graphical structure obtained from eliciting expert knowledge from literature can be directly replicated in Netica quickly and easily. Nodes represent system variables and therefore take on a value or state. The conditional probabilities of a node show the probability of the variable taking on each state – the total probability of the variable must equal unity i.e. the variable must have some state/value. How the states are represented by the conditional probability distribution depends on whether the node is continuous or discrete, qualitative or quantitative.

The conditional probability tables that describe the probability of nodes having some state given the probability of other nodes can be obtained from data in the form of case files or by an equation. The equation may be a probabilistic equation such as that describing e.g. normal and

Weibull distributions – these are probabilistic relationships. Alternatively the equation may be one that describes the physical relationship between two variables – this is a deterministic relationship.

6.6.1 Continuous/discrete, qualitative/quantitative nodes

Nodes are categorised as continuous or discrete depending on the type of information the system variable represents. Continuous variables have states/values that are measured on a continuous scale and therefore are always quantitative. For example, DHW consumption volume is a continuous variable as the measured consumption can take any possible value between the minimum and maximum. For continuous nodes the states are presented as discrete intervals; this process is called discretisation. Each discrete interval (also called a bin) can be given a constant width or the width may vary to allow different parts of the distribution to have different resolutions depending on the available data. Areas of the distribution with large amounts of data (around the mean) can have finer bin widths compared to the extreme areas of the distribution which may require wider bin widths to ensure it is represented in the data. Discrete nodes operate on a discrete scale such as household occupancy in which the occupancy can only be described by an integer number (not fractions of people). Each integer value in the data becomes a state in the node. Discrete nodes may also represent qualitative variables such as collector type which may be flat plate or evacuated tube. Alternatively, some variables may be qualitative or quantitative such as quality of installation which can be described as excellent, good, average, or poor or quantified as 4,3,2,1 using a scoring scheme.

6.6.2 Probabilistic versus deterministic nodes

Nodes in Netica can have their conditional probability tables determined by equations or data. Parentless probabilistic nodes may have conditional probabilities described by probability distribution functions for example Normal or Weibull equations if it is a continuous node, or Poisson or Binomial functions if it is a discrete node. Nodes with parents may have the probability distributions obtained from data using the algorithms present in Netica (detailed later) which give the probability of the node having a certain state given the probabilities of the parents having certain states. Probabilistic relationships can be used to replace regression based equations and therefore capture the entire variation in the dependent variable.

Some relationships can be described by deterministic relationships and these can be captured in Bayesian networks. In Netica such a relationship is given by the equation that links the variables. For example there is a deterministic relationship between voltage, current and resistance given by Ohm's law. In Netica the probability distributions for current and resistance can be obtained from measured data or by a probability distribution function such as a Normal distribution used to introduce the variation in current that occurs in a system (uncertainty due to measurement accuracy can also be included in the BN using the measurement idiom described by Fenton and Neil (Fenton & Neil 2013)); the voltage node is determined by the equation V=IR. In Netica, nodes described by an equation will need their equations converting to a CPT; this is done by sampling. For example the current and resistance conditional probability distributions will be sampled N times and the product calculated to give N voltage values which are placed in the appropriate state intervals in the voltage node to provide a distribution of voltage. This is effectively a simple Monte Carlo simulation. The same principle is used when the nodes are discrete. This approach was applied to the regression equation developed in Chapter 4 where

uncertainty was introduced by sampling from the distributions of the variables as well as from normal distributions associated with the regression coefficients (Chapter 5). In this way even deterministic relationships can have uncertainty due to variation in the parent nodes quantified.

6.7 Learning algorithms used in Netica

As mentioned briefly above, Netica uses case files to learn the conditional probability tables (CPTs) from data held by the modeller. Case files typically take the form of an Excel spreadsheet with a column representing each of the nodes/variables in the network and a row for each data entry. The data across all variables in a row must come from the same source/system. If gaps in the data for a system do exist then these must be represented by an empty cell in the case file. Bayesian networks are resilient to missing data values; this is advantageous for long term solar thermal system performance prediction based on evidence because missing data often occurs (EST 2011; O'Flaherty et al. 2012; The Energy Monitoring Company Ltd 2001). Case files can be constructed using data from different sources such as field trials performed by separate bodies at different times as well as those that provide data for different system variables; for example data from hot water field trials can be combined with solar thermal field trials conducted by different researchers.

Once the case file has been constructed it can be incorporated into the Bayesian network in Netica using three available learning algorithms: the count algorithm; expectation maximisation (EM) and gradient learning.

6.7.1 Count algorithm

The count algorithm is the preferred method of developing CPTs automatically as it is the simplest and quickest algorithm. The method is simple: for each combination of parent node states the number of data points that satisfy each of the child node states is counted and a frequency table is generated (example given by Table 6.1) with the number of cells being given by the product of the number of parent node combinations and the number of child node states. The number of parent node combinations is given by:

$$(Number\ of\ states\ in\ X_1) \times (Number\ of\ states\ in\ X_2) \times \ldots \times (Number\ of\ states\ in\ X_n) \quad (6.1)$$

Household Type	**Number of Occupants**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Single	100	0	0	0	0	0	0
Working family	0	0	20	60	15	4	1
Working couple	0	100	0	0	0	0	0
Retired couple	0	100	0	0	0	0	0
Retired single	100	0	0	0	0	0	0

Table 6.1: CPT of the node *Number of Occupants* given the parent node *Household Type* there is 1 parent node with 5 states giving 5 parent node combinations, multiply this by the 7 states of *Number of Occupants* and the number of cells in the CPT is 35.

This method is more effective with increasing amounts of available data, decreasing numbers of parent nodes and decreasing numbers of node states. Consider a case file that contains 10000 data points for a child node Z and corresponding data for four parent nodes A, B, C, D. Now

consider that each parent node has 5 states/bins and node Z has 6 states/bins, this gives 5x5x5x5 = 625 parent node combinations and 3750 cells in CPT which means 3750 individual probabilities/frequencies required from the data. Assuming that the data collected represents all parent node combinations uniformly then each node combination will have just 16 data points which will then be divided amongst the child node bins and each bin will only be represented by 2-3 data points assuming the distribution is uniform. This problem is made worse if the distributions in the parent and child nodes are not uniform since some of the extreme regions of the distribution may be represented by very few data points and may possibly have no data at all. Imagine that the number of parent nodes in the above example is reduced to 3 and the states reduced to 4, this gives 64 node combinations giving 156 data points per node combination (assuming the distribution of the data is uniform). This demonstrates the importance of minimising the number of parent nodes and states that a child node has. This can be achieved with latent variables, which describe and combine several parent nodes into a single node which can then be connected to a child node thus reducing the number of parent nodes; this can be achieved using statistical approaches such as PCA used by Lam et al. to reduce three climatic variables (dry bulb temperature, wet bulb temperature and global solar radiation) into a single latent variable Z for use in a regression model to predict building energy consumption (Lam et al. 2010). However one of the stipulations of the count algorithm is that no latent variables are present since there would be no data in the case file for such a variable. Furthermore the count algorithm is not reliable when there is a significant amount of missing data.

There are two alternative algorithms that can be used to construct CPTs from data with missing values and/or latent variables: expectation maximisation (EM); and gradient learning – Norsys suggest that Netica's EM algorithm provides more robust results over a wide variety of situations; it is for this reason that this algorithm was used and so only the EM algorithm will be discussed.

6.7.2 Expectation maximisation (EM) learning

The EM algorithm is so called because it is an iterative process with an expectation step followed by a maximisation step (Dempster et al. 1977). To illustrate the process, an example is presented of how conditional probabilities in a simple Bayesian network with missing data entries can be learned using the EM algorithm (Neapolitan 2003).

Let X be a random variable that can take on two values 1 and 2; let F be a random variable which represents the belief concerning the relative frequency of random variable X being equal to 1; F can take on the states true or false which are represented by 1 and 0 respectively. The expected value of F is defined as the estimate of the relative frequency; this is noted as $E(F)$. Now let f be the relative frequency of occurrence of $X = 1$, then given this information the belief that $X = 1$ in the first execution of an experiment will be f:

$$P(X = 1|f) = f \tag{6.2}$$

In other words the subjective probability of $X = 1$ in the first experiment is equal to the estimate of the relative frequency:

$$P(X = 1) = E(F) \tag{6.3}$$

Consider the scenario where three urns exist each of which contains a coin. The first urn is labelled X_1; the second urn is labelled $X_2|X_1 = 1$ and will have its coin tossed in the event that the

coin in $X_1 = 1$ (1 here represents heads); the third urn is labelled $X_2|X_1 = 2$ and will have its coin tossed in the event that the coin $X_1 = 2$ (2 here represents tails). This can be represented by the following Bayesian network (Figure 6.6):

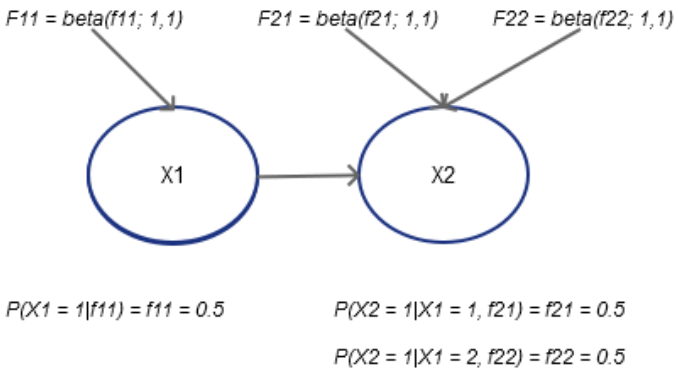

Figure 6.6: Bayesian network with prior beliefs about f_{11}, f_{21} and f_{22}

The outcomes of X_1 and X_2 can be represented by a beta distribution with parameters a and b – written as $beta(f; a, b)$; Neapolitan shows that the subjective probability of $X_1 = 1$ can be given by:

$$P(X_1 = 1) = \frac{a}{a+b} \tag{6.4}$$

Assuming that prior experience suggests that out of 100 tosses $X_1 = 1$ fifty times then $a = 50$ and $b = 50$ giving $P(X_1 = 1) = 0.5$ – this is the estimate of the relative frequency $E(F)$. It can be seen that if the relative frequency of $X_1 = 1$ is 0.5 then a will be equal to b. In this example three random variables F_{11}, F_{21}, and F_{22} are used to represent the belief concerning the relative frequencies (f_{11}, f_{21} and f_{22}) of the events $X_1 = 1$, $X_2 = 1|X_1 = 1$ and $X_2 = 2|X_1 = 1$ respectively; the probability distributions of F_{11}, F_{21}, and F_{22} are given by the beta distribution:

$$F_{11} = beta(f_{11}; a_{11}, b_{11});\ E(F_{11}) = P(X_1 = 1) = \frac{a_{11}}{a_{11}+b_{11}} \tag{6.5}$$

$$F_{21} = beta(f_{21}; a_{21}, b_{21});\ E(F_{21}) = P(X_2 = 1|X_1 = 1) = \frac{a_{21}}{a_{21}+b_{21}} \tag{6.6}$$

$$F_{22} = beta(f_{22}; a_{22}, b_{22});\ E(F_{22}) = P(X_2 = 2|X_1 = 1) = \frac{a_{22}}{a_{22}+b_{22}} \tag{6.7}$$

In four tosses of X_1 it is expected that $X_1 = 1$ two times; $X_2 = 1|X_1 = 1$ once; and $X_2 = 2|X_1 = 1$ once therefore $a_{11} = 2$, $b_{11} = 2$; $a_{21} = 1$, $b_{21} = 1$; $a_{22} = 1$, $b_{22} = 1$ therefore $P(X_1 = 1) = 0.5$, $P(X_2 = 1|X_1 = 1) = 0.5$ and $P(X_2 = 2|X_1 = 1) = 0.5$.

The above lays out the expected frequencies of occurrence of different events in the experiment based on the prior belief that each event has a 50% chance of occurring. It is useful to now demonstrate how these beliefs can be updated when data about the frequency of occurrence of these events is available.

- s_{11} = the number of times $X_1 = 1$;
- t_{11} = the number of times $X_1 = 2$;
- s_{21} = the number of times $X_2 = 1|\ X_1 = 1$;
- t_{21} = the number of times $X_2 = 2|\ X_1 = 1$;

- s_{22} = the number of times $X_2 = 1 | X_1 = 2$;
- t_{22} = the number of times $X_2 = 2 | X_1 = 2$;

Case	X_1	X_2
1	1	1
2	1	1
3	1	1
4	1	2
5	2	2

Table 6.2: Data for 5 tosses of the coins

From Table 6.2 it can be seen that $s_{21} = 3$ and $t_{21} = 1$; Neapolitan shows that:

$$\rho(f_{21}|d) = beta(f_{21};\ a_{21} + s_{21}, b_{21} + t_{21}) \tag{6.8}$$
$$= beta(f_{21}; 1 + 3, 1 + 1)$$
$$= beta(f_{21}; 4, 2)$$

$E(F_{21})$ has now been updated from $beta(f_{21};\ 1,\ 1)$ to $beta(f_{21};\ 4,\ 2)$ owing to the presence of data, d.

Now consider updating the prior beliefs with data, d that has missing entries shown in Table 6.3.

Case	X_1	X_2
1	1	1
2	1	?
3	1	1
4	1	2
5	2	?

Table 6.3: Data for 5 tosses of the coins with missing data for cases 2 and 5

A reasonable estimate for the value of X_2 in cases 2 and 5 would be to use $P(X_2 = 1|X_1 = 1)$ which is equal to 0.5; therefore in cases 2 and 5 X_2 has a 50% chance of being equal to 1 – this is used to produce a new table of data, d' (Table 6.4):

Case	X_1	X_2	Number of occurrences
1	1	1	1
2	1	1	0.5
2	1	2	0.5
3	1	1	1
4	1	2	1
5	2	1	0.5
5	2	2	0.5

Table 6.4: Estimated missing values for *d*

The beta functions can now be updated using the data d' in Table 6.4: s'_{21} = (1+0.5+1) = 2.5 and $t'_{21} = (0.5+1) = 1.5$; $s'_{22} = 0.5$ and $t'_{22} = 0.5$.

$$\rho(f_{21}|d') = beta(f_{21};\ a_{21} + s'_{21}, b_{21} + t'_{21}) \tag{6.9}$$
$$= beta(f_{21}; 1 + 2.5, 1 + 1.5)$$

$$= beta(f_{21}; 3.5, 2.5)$$

And

$$\rho(f_{22}|d') = beta(f_{22};\ 1.5, 1.5) \tag{6.10}$$

This calculation process is the first pass of the EM algorithm and updates the network in Figure 6.6 to that in Figure 6.7:

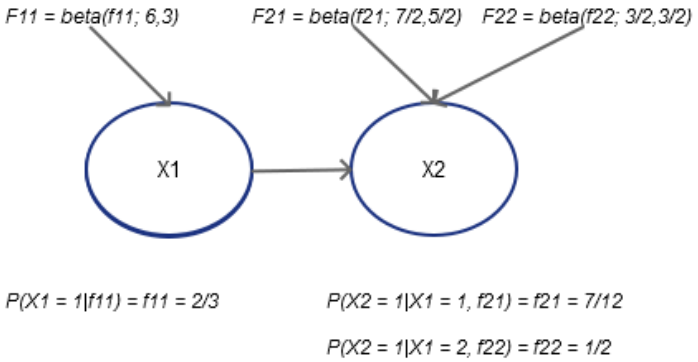

Figure 6.7: Bayesian network with beliefs about f_{11}, f_{21} and f_{22} updated after the first pass of the EM algorithm

By considering s_{ij} and t_{ij} as random variables, the above method updates the value of these variables based on their expected values given data and the prior value of the relative frequency, f':

$$\mathrm{f}' = \{f'_{11}, f'_{21}, f'_{22}\} = \{f_{11}, f_{21}, f_{22}\} = \{0.5, 0.5, 0.5\} \tag{6.11}$$

$$s'_{21} = E(s_{21}|d, \mathrm{f}') = \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 1|d, \mathrm{f}') \tag{6.12}$$

$$= \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 1|\mathrm{x}^{(h)}, \mathrm{f}') \tag{6.13}$$

$$= \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 1|x_1^{(h)}, x_2^{(h)}, \mathrm{f}') \tag{6.14}$$

$$= 1 + 0.5 + 1 + 0 + 0 = 2.5$$

$$t'_{21} = E(t_{21}|d, \mathrm{f}') = \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 2|d, \mathrm{f}') \tag{6.15}$$

$$= \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 2|\mathrm{x}^{(h)}, \mathrm{f}') \tag{6.16}$$

$$= \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 2|x_1^{(h)}, x_2^{(h)}, \mathrm{f}') \tag{6.17}$$

$$= 0 + 0.5 + 0 + 1 + 0 = 1.5$$

These are the same values for s'_{21} found above from Table 6.4. These estimates of the beta density functions are obtained from prior data about the chances of $X_2 = 1|X_1 = 1$ and not based on the data, d in Table 6.3. It is seen in Table 6.3 that $X_2 = 1|X_1 = 1$ occurs twice and $X_2 = 2|X_1 = 1$ occurs only once; therefore to incorporate this data in the estimates of the beta functions the computation to find s_{ij} is repeated using the probabilities in the updated network:

$$\rho(f'_{21}|d') = beta(f'_{21};\ a_{21} + s'_{21}, b_{21} + t'_{21}) \tag{6.18}$$

$$= beta(f'_{21}; 1 + 2.5, 1 + 1.5)$$

$$= beta(f'_{21}; 3.5, 2.5)$$

$$P(X_2 = 1|X_1 = 1, f'_{21}) = f'_{21} = \frac{a_{21}}{a_{21}+b_{21}} \quad (6.19)$$

$$f'_{21} = \frac{3.5}{3.5 + 2.5} = \frac{7}{12}$$

Repeating for f_{11} and f_{22} gives:

$$\mathrm{f}' = \{f'_{11}, f'_{21}, f'_{22}\} = \left\{\frac{2}{3}, \frac{7}{12}, \frac{1}{2}\right\} \quad (6.20)$$

$$s'_{21} = E(s_{21}|d, \mathrm{f}') = \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 1|d, \mathrm{f}') \quad (6.21)$$

$$= \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 1|\mathrm{x}^{(h)}, \mathrm{f}') \quad (6.22)$$

$$= \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 1|x_1^{(h)}, x_2^{(h)}, \mathrm{f}') \quad (6.23)$$

$$= 1 + 7/12 + 1 + 0 + 0 = 2\frac{7}{12}$$

Note for t'_{21} $P(X_2 = 2|X_1 = 1) = 2.5/(2.5+3.5) = 5/12$:

$$t'_{21} = E(t_{21}|d, \mathrm{f}') = \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 2|d, \mathrm{f}') \quad (6.24)$$

$$= \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 2|\mathrm{x}^{(h)}, \mathrm{f}') \quad (6.25)$$

$$= \sum_{h=1}^{5} 1 \times P(X_1^{(h)} = 1, X_2^{(h)} = 2|x_1^{(h)}, x_2^{(h)}, \mathrm{f}') \quad (6.26)$$

$$= 0 + 5/12 + 0 + 1 + 0 = 1\frac{5}{12}$$

This process of calculating the variables s_{ij} and t_{ij} is known as the expectation step and the re-computation of f' using these updated quantities is the maximisation step; these steps can be repeated as shown above until $\rho(\mathrm{f}|d)$ is maximised. The value of f that maximises $\rho(\mathrm{f}|d)$ is denoted $\tilde{\mathrm{f}}$ and is termed the maximum a-posterior probability (MAP) – this is the value of f required rather than one that leads to a local maximum of $\rho(\mathrm{f}|d)$; Neapolitan provides a method for avoiding local maxima (Neapolitan 2003).

6.8 Using TRNSYS simulations to determine a DHW draw timing CPT

The results from the TRNSYS simulations presented in Chapter 4 were used to develop the probabilistic relationship between system performance and DHW draw timing, which were unobtainable from the EST field trial data. The following procedure was used to obtain the CPT:

1. Construct a BN from the TRNSYS simulation;
2. Obtain the prior probability distribution for *Daily* Q_{sol} using the TRNSYS BN;

3. Change the time of hot water draw using the *DHW usage* node in the TRNSYS BN to obtain the updated distributions of *Daily* Q_{sol};
4. Calculate the scale factor between the prior *Daily* Q_{sol} and each of the updated distributions in the TRNSYS BN;
5. Multiply the CPT values of *Daily* Q_{sol} in the knowledge-based BN by the scale factors obtained for the four different DHW usage timings (step 4) – this increases the number of conditional probabilities in the table by four times;
6. Connect the *DHW usage* node to the *Daily* Q_{sol} node in the final BN and paste in the modified CPT;

6.8.1 Using simulated data to construct a Bayesian network

The TRNSYS model used to investigate the annual performance of a solar thermal system described in Chapter 4 was transformed into a Bayesian network. The model parameters were used to provide the structure of the network (Figure 6.8) whilst the CPTs were developed using the simulated results and the count algorithm (section 6.7.1).

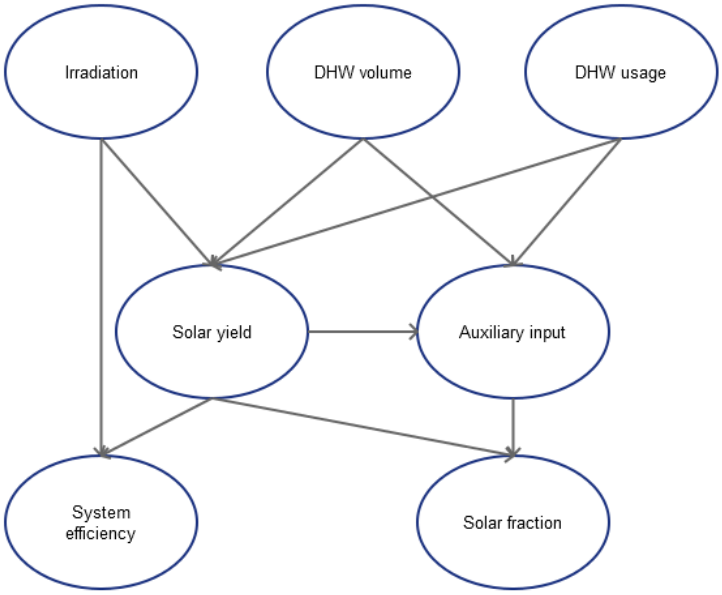

Figure 6.8: Graphical network structure of annual TRNSYS model

In a real system, fluctuations in the cold water temperature over the year would have an effect on the amount of energy required to heat the mains water (Thur et al. 2006). However in these simulations the cold water temperature is fixed throughout the year at 12°C.

The annual simulation had to be run 28 times in order to populate the CPTs (Table 6.5), providing a total of 10220 days of data. The states in each node are given as daily values; for example, the distribution of Q_{sol} is a distribution of the daily solar yields that are provided over the course of a year.

Variable	States	Number of states
DHW timing	Morning; early afternoon; late afternoon; evening	4
DHW volume	0L; 87.5L; 175L; 262.5L; 350L; 437.5L; 525L	7

Table 6.5: Variables and states used in the daily TRNSYS simulations

The Bayesian network constructed from annual data is shown in Figure 6.9.

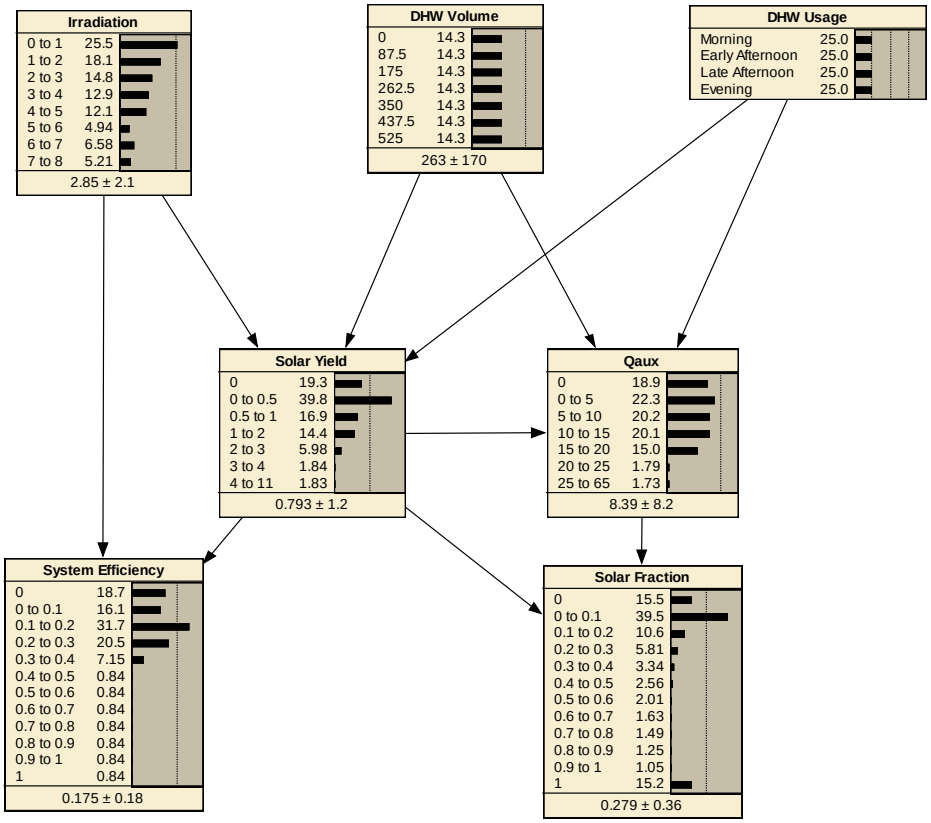

Figure 6.9: Netica BN for annual TRNSYS simulations

6.8.2 Prior distribution of solar yield

The prior distribution of solar yield refers to the distribution of Q_{sol} when no evidence about the time of hot water draw is given; it is presented in Figure 6.10.

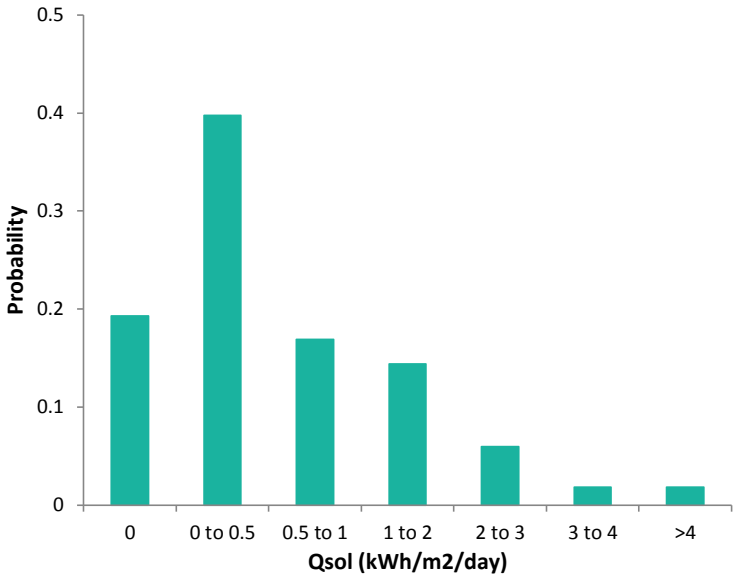

Figure 6.10: Prior distribution obtained from TRNSYS simulations

The average daily solar yield is 0.793kWh/m^2/day with standard deviation 1.2kWh/m^2/day.

6.8.3 The effect of DHW consumption timing

The posterior distributions of Q_{sol} given evidence about the time of DHW use are shown in Figure 6.11, updated mean solar yield values are presented in Table 6.6.

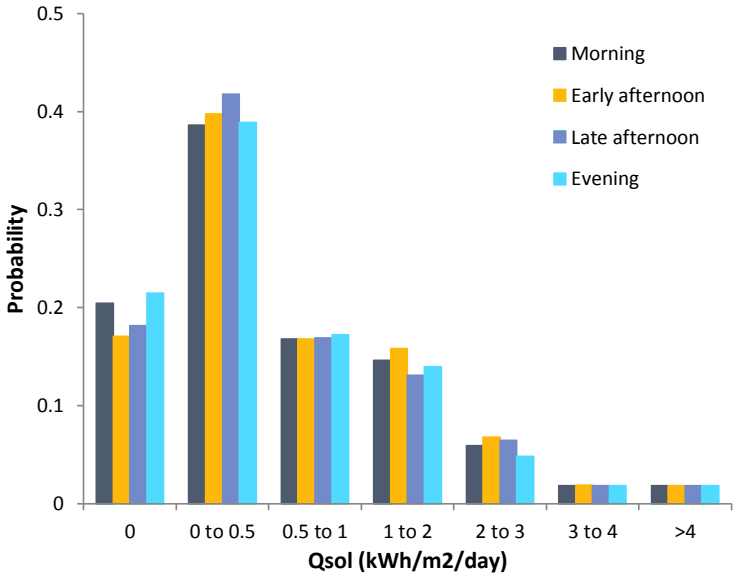

Figure 6.11: P(Q_{sol} | Morning, Early Afternoon, Late Afternoon, Evening)

	Qsol (kWh/m^2/day)	
	Mean	**SD**
Morning	0.79	1.2
Early afternoon	0.835	1.3
Late afternoon	0.79	1.2
Evening	0.757	1.2

Table 6.6: Mean and standard deviation daily solar yield versus hot water draw time

Results of the simulations are discussed in detail in Chapter 4. In summary Table 6.6 indicates that an early afternoon draw provides highest average solar yields and an evening draw provides the lowest.

6.8.4 Obtain scale factors

The prior probabilities in the distribution for Q_{sol} using the TRNSYS BN are shown in column 2 of Table 6.7; the probabilities for the updated Q_{sol} distributions given the time of hot water drawn are also shown (columns 3-6). The scale factors are the ratio of the posterior Q_{sol} distribution given DHW usage to the prior Q_{sol} distribution:

$$Scale\ factor_i = \frac{P(Q_{sol,i}|DHW\ usage_i)}{P(Q_{sol,i})} \tag{6.27}$$

Where subscript i denotes the i^{th} scale factor/probability for which there are 7 in the distribution for *Daily* Q_{sol}.

Qsol	TRNSYS prior	Morn	Early aft.	Late aft.	Eve.	Morn. scale factor	Early aft. scale factor	Late aft. scale factor	Eve. scale factor
0	0.1928	0.2044	0.1710	0.1813	0.2146	1.0601	0.8866	0.9404	1.1129
0 to 0.5	0.3976	0.3860	0.3979	0.4175	0.3890	0.9708	1.0007	1.0501	0.9785
0.5 to 1	0.1693	0.1678	0.1682	0.1691	0.1722	0.9910	0.9936	0.9985	1.0170
1 to 2	0.1438	0.1462	0.1584	0.1310	0.1394	1.0170	1.1020	0.9111	0.9698
2 to 3	0.0598	0.0590	0.0676	0.0645	0.0482	0.9862	1.1304	1.0783	0.8051
3 to 4	0.0184	0.0183	0.0186	0.0183	0.0183	0.9961	1.0117	0.9961	0.9961
4 to 10	0.0183	0.0183	0.0183	0.0183	0.0183	1.0000	1.0000	1.0000	1.0000

Table 6.7: TRNSYS prior distribution with distribution of Q_{sol} given time of draw; scale factors given by P(Q_{sol} | DHW usage)/P(TRNSYS prior)

6.8.5 Apply the scale factor to the knowledge-based BN Q_{sol} prior

Taking the first row of the original CPT for *Daily* Q_{sol} in the knowledge-based BN gives Table 6.8:

Qsol	Original prior
0	0.3333
0 to 0.5	0.3333
0.5 to 1	0.3333
1 to 2	0
2 to 3	0
3 to 4	0
4 to 10	0

Table 6.8: Probabilities from the first row of the CPT for *Daily* Q_{sol} from the final BN prior to the inclusion of *DHW usage*

By multiplying these probabilities by the scale factor the first row of the modified CPT gives Table 6.9:

Qsol	Original prior	Morn	Early aft.	Late aft.	Eve.
0	0.3333	0.3534	0.2955	0.3135	0.3710
0 to 0.5	0.3333	0.3236	0.3336	0.3500	0.3262
0.5 to 1	0.3333	0.3303	0.3312	0.3328	0.3390
1 to 2	0.0000	0.0000	0.0000	0.0000	0.0000
2 to 3	0.0000	0.0000	0.0000	0.0000	0.0000
3 to 4	0.0000	0.0000	0.0000	0.0000	0.0000
4 to 10	0.0000	0.0000	0.0000	0.0000	0.0000

Table 6.9: Scaled probabilities from the first row of the CPT for *Daily Q_{sol}* | *DHW usage*

These probabilities then need to be normalised to ensure that they sum to 1 giving Table 6.10:

Qsol	Original prior	Morn	Early aft.	Late aft.	Eve.
0	0.3333	0.3508	0.3077	0.3146	0.3580
0 to 0.5	0.3333	0.3213	0.3474	0.3513	0.3148
0.5 to 1	0.3333	0.3279	0.3449	0.3341	0.3272
1 to 2	0.0000	0.0000	0.0000	0.0000	0.0000
2 to 3	0.0000	0.0000	0.0000	0.0000	0.0000
3 to 4	0.0000	0.0000	0.0000	0.0000	0.0000
4 to 10	0.0000	0.0000	0.0000	0.0000	0.0000

Table 6.10: Scaled and normalised probabilities from the first row of the CPT for *Daily Q_{sol}* | *DHW usage*

For the complete scaled distributions each value in the original CPT is multiplied by the appropriate scale factor; this will increase the number of probabilities in the modified CPT fourfold because there are four states in the additional node (*DHW usage)*. The same procedure is required for the *Daily Q_{aux}* node.

6.8.6 Limitations with using simulated data

- The number of simulations needed to be performed is equal to the total number of node probabilities required to populate the CPT;
- Distributions look rather granular and coarse due to small sample size: an increased sample size allows finer bin widths to be used but this will increase the size of the CPTs
- Usage patterns are often non-recurring and change each day (Buckles & Klein 1980)

6.9 Description of the Bayesian network model

The knowledge-based Bayesian network is an object oriented Bayesian network (OOBN) made up of three groups of nodes referred to as objects: user behaviour object; system configuration object; and solar resource object. The BN is shown in Figure 6.12 where the three objects are designated by the dashed boxes.

6.9.1 Object oriented Bayesian networks

Solar thermal performance is dependent on user behaviour, the operating environment and system configuration. As has been previously stated these system elements cannot be combined into a single predictive formula for system performance and are themselves dependent on complex interactions between random variables (Mathioulakis et al. 2012). As such a Bayesian network topology can be applied to each of these three subsystems and combined to produce

the overall BN for solar thermal performance modelling. These types of network designs are known as object oriented Bayesian networks (OOBN) where each of the sub-networks are known as objects (Naticchia & Carbonari n.d.).

Nodes in a BN can be categorised as: input nodes; intermediate nodes; and output nodes. Input nodes exist for the user to insert information about the random variables in the problem domain; intermediate nodes are used to perform calculations using data from input nodes; output nodes display information that can be used by decision makers (Naticchia & Carbonari n.d.). In OOBNs the output node of one object will form the input node for another object and may also feed into an intermediate node. For example the three objects in the solar thermal BN (user behaviour, system configuration, and environment) feed information into the system performance node from their output nodes (Figure 6.13).

In Figure 6.13 a combination of UML and Bayesian network graphology has been used to remind the reader that the objects in the OOBN refer to the different sub-system models described in Chapter 4 and represented by the UML diagram. The figure is only an example but shows how the input nodes (dashed bordered nodes) are used to generate output nodes (solid bordered nodes) within an object and that these outputs can be used to provide information for input nodes in another object as well as feed into the overall output node (system performance). The outputs may be generated probabilistically using CPTs or deterministically by computing equations in which case the output node feeds into an intermediate node such as Node2 in the System configuration object.

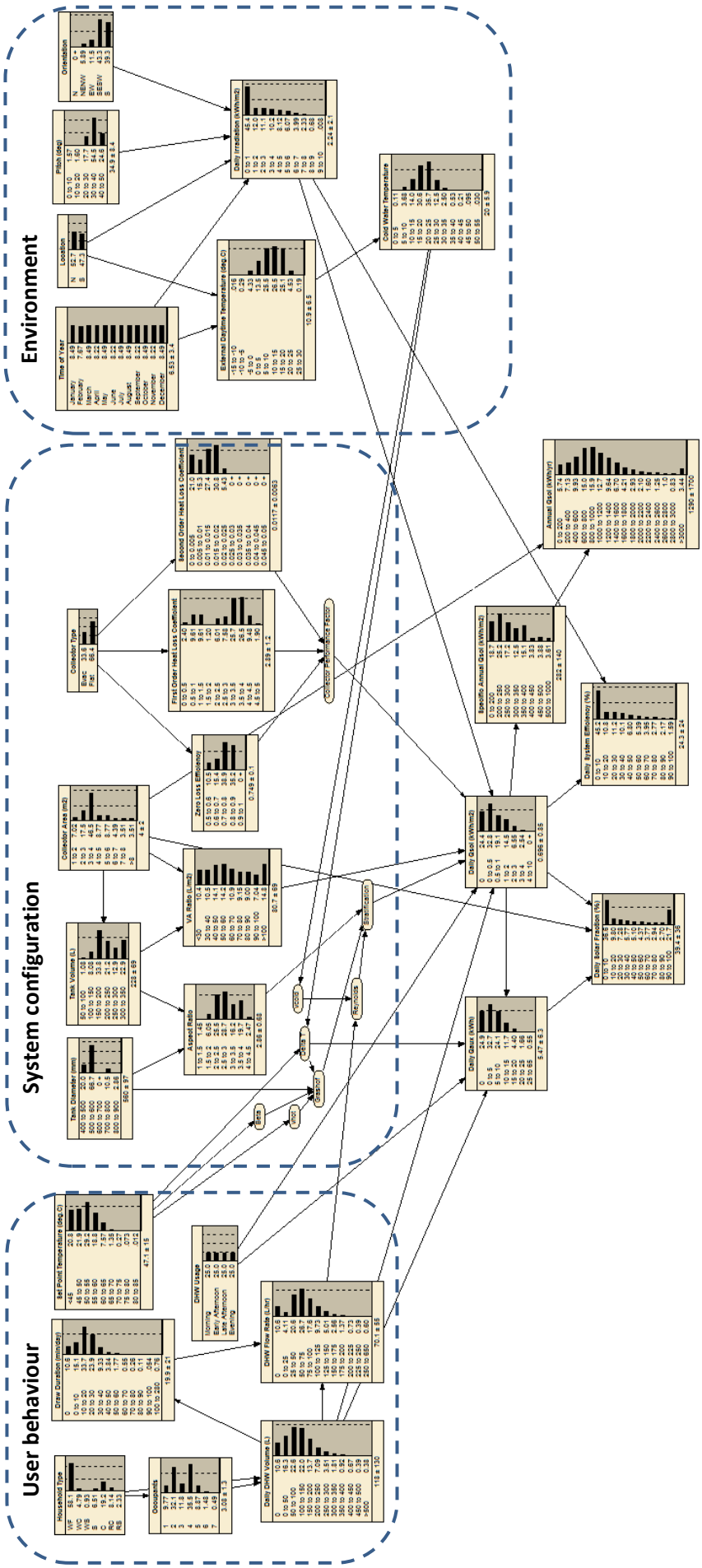

Figure 6.12: Completed Bayesian network for the prediction of solar thermal system performance

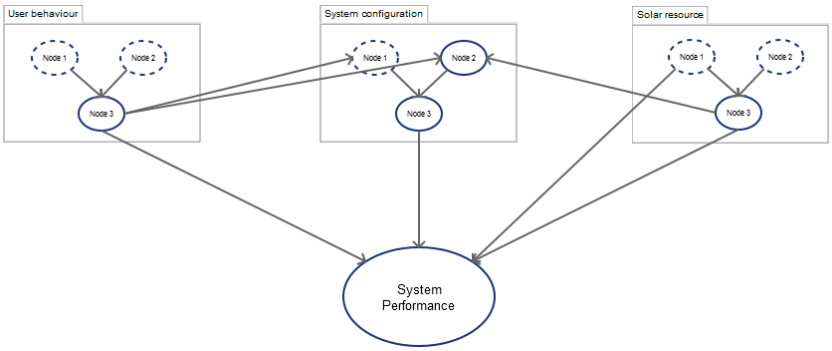

Figure 6.13: Example of a simplistic OOBN for solar thermal system performance

Using the object oriented approach made BN construction simpler and faster. These advantages stem from the reduced number of nodes that are processed using the EM algorithm. Furthermore the outputs from this BN which focuses on the technical aspects of a solar thermal system may be fed into another Bayesian object that deals with the socio-economic impacts of different renewable technologies. In this way the outputs from BNs that model technologies in detail can be used in higher level BNs that model the effects of different policies on society and design decisions on the energy and carbon performance of buildings, both of which require a distribution of performance of the technical systems concerned.

6.9.2 User behaviour object

User behaviour is complex and random; it may be described by many random variables that interact with each other but are open to a great deal of variation themselves due to personal needs/wants, which are a response to individual schedules or habits; climatic factors; and socioeconomic factors such as the cost of water and energy, the changing size of families and households, and policies designed to encourage certain behaviours (Knudsen 2002; Michaelides 1993; Evarts & Swan 2013). User behaviour is modelled using regression equations in simplified physics based models (SAP, BREDEM) and based on assumptions on the part of the modeller in dynamic based models (TRNSYS). Although these relationships may stem from collected data, both of these deterministic methods fail to capture the inherent uncertainty in the variables that describe the system. In addition there are certain qualitative variables that influence user behaviour which are not able to be incorporated into the traditional forms of modelling. The 'user behaviour' Bayesian object aims to take random variables associated with user behaviour and establish causal links between them whilst displaying the uncertainty in each variable given the value of the parents using the concept of conditional probability. The conditional probabilities between the variables were established using real data collected from the EST hot water survey (EST 2008) and EST solar thermal field trial (EST 2011).

Since user behaviour is difficult to describe deterministically, all of the nodes within the user behaviour object are probabilistic; there are seven nodes in this object (Appendix B).

6.9.3 Solar resource object

The solar resource object contains nodes related to the solar irradiation as well as other climatic factors such as the external ambient daytime temperature. There are several methods employed to model the solar resource and overview of which can be found in Chapter 4; however the Bayesian method allows distributions for daily irradiation to be determined from real data and

indicates which variables are influential on this parameter. The outputs of the solar resource object are daily irradiation and cold water temperature which are influenced by the time of year, location, pitch and orientation of the collector, and external ambient temperature; there are seven nodes in this object (Appendix B).

6.9.4 System configuration object

The system configuration object contains information about the size of the tank and collector as well as the performance parameters of the collector. This object contains both probabilistic and deterministic nodes. Some input and intermediate nodes in this object are output nodes of the previous two objects. The conditional probabilities of this object have been determined using the EST solar thermal field trial data; in addition a subset of nodes has had additional data sets incorporated into separate case files; there are 17 nodes contained within this object (Appendix B).

6.9.5 Performance nodes

The performance of the solar thermal system is described by daily solar yield, annual solar yield, daily solar fraction and daily system efficiency – these constitute the performance nodes. Daily auxiliary is also represented within this group of nodes due to its dependence on the daily solar yield. The key node in this group is the daily solar yield node because from this all other performance parameters can be inferred. For inference of *Daily* Q_{sol} to be made possible a case file is required that contains the output variables from the previously described Bayesian objects. In the performance case file the measured values for each of these output variables from the EST solar thermal field trial systems are used. This case file is then incorporated into the group of performance nodes, which in actual fact can be considered a fourth object.

The three Bayesian objects provide distributions of the parent variables to *Daily* Q_{sol} these can be denoted P(*Par(Daily* Q_{sol}*)*) where *Par* stands for parent(s). The case file uses measured data (evidence) from the field trial to elicit the conditional probabilities of *Daily* Q_{sol} given the parent nodes – this makes use of Bayes' theorem:

$$P(Daily\ Q_{sol}\ |Par(Daily\ Q_{sol})) = \frac{P(evidence\ Par(Daily\ Q_{sol})|Daily\ Q_{sol})P(Daily\ Q_{sol})}{P(Par(Daily\ Q_{sol}))} \quad (6.28)$$

P(*evidence Par(Daily* Q_{sol}*)* | *Daily* Q_{sol}) is the likelihood function and provides the probability that the value of the parent variables is observed in the data given the value of *Daily* Q_{sol}; P(*Daily* Q_{sol}) is the prior distribution of the daily yield measured in the field trial and used in the performance case file; P(*Par(Daily* Q_{sol}*)*) is the distribution of the parent variables provided by each of the aforementioned Bayesian objects. The result is the posterior distribution of *Daily* Q_{sol} updated given information about the parent nodes. There are 6 nodes in this group (Appendix B).

6.10 Information provided by the network

The knowledge-based BN displays the causal connections between variables related to solar thermal system performance through its graphical structure; this provides a deeper understanding of the factors that influence the annual yield. The causal structure of the system domain is provided by data elicited from expert opinion and can be updated or modified as new research arises.

The BN gives an overview of the uncertainties that exist in the domain variables and shows how this uncertainty is propagated through the network to allow a distribution of annual yield to be determined and thus quantifying the risk associated with variable performance. The uncertainty in each parameter is the epistemic uncertainty and describes the random variation in the values of those parameters. This epistemic uncertainty causes variation in the performance of a group of heterogeneous systems. Large scale analysis of a group of heterogeneous systems can be modelled in the BN because the prior distribution of the annual solar yield shows the variation in performance across the whole collection of systems in the dataset used to populate the CPTs. By incorporating more systems' data in the dataset a more representative portrayal of the large-scale performance of systems across the UK can be shown. Marginalising the input nodes (i.e. inserting information known about a system or group of homogenous systems) will update the *a-priori* knowledge regarding the annual solar yield. The variation that then exists in the annual solar yield is known as the aleatory uncertainty and represents the random variation that occurs in the output within a group of homogenous systems or from year to year in a single system; therefore BNs can be used to perform large scale analysis across a range of different systems as well as small scale analysis which is confined to a homogenous group of systems or a single system.

One advantage of Monte Carlo methods over methods such as differential sensitivity analysis is the ability to maintain the interactions between system variables; on the other hand DSA allows for the influence of individual variables on the output to be determined. Bayesian networks allow both of these analyses to be performed by choosing to hold variables at a constant value or allowing uncertainty to propagate through them given information about the parent nodes. Therefore BNs can provide information about the most influential variables as well as keeping variable interaction intact.

6.11 Large-scale analysis of solar thermal system performance

When no evidence about system parameters is known, the knowledge-based BN displays the epistemic uncertainty in system performance for a heterogeneous group of systems in the UK based on in-situ performance data collected by the EST (EST 2011). The BN therefore allows large-scale prediction of solar thermal system performance. Figure 6.14 and Figure 6.15 show the distributions of specific and total annual solar yield respectively.

As one would expect, the measured and predicted specific annual solar yield match closely (Figure 6.14). This is because the BN predicted specific Q_{sol} is derived from the measured data, small discrepancies are likely due to the EM algorithm probability approximations where missing data values exist.

Mean measured annual solar yield is 1124.2kWh/yr with a standard deviation of 450.1kWh/yr, this is lower and less variable than the BN predicted mean of 1290kWh/yr (SD = 1700kWh/yr). The discrepancy between the prediction and measured distributions is due to a combination of factors. The most influential of these is that BN-predicted annual solar yield is a deterministic product of specific annual yield and collector area. By sampling the collector area and specific annual yield distributions 1000 times, many more possible values of annual solar yield are

produced; this is in contrast to the measured distribution for annual yield, which describes a single year's performance for 24 systems in the EST field trial.

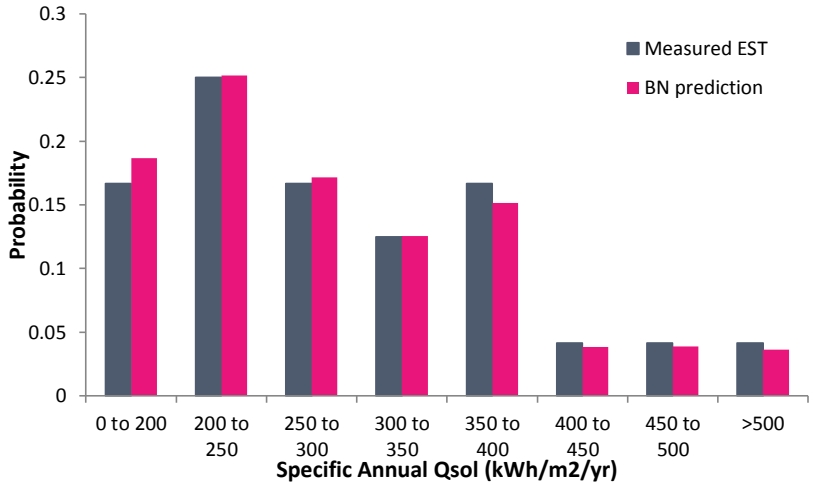

Figure 6.14: Measured and BN predicted specific annual solar yield

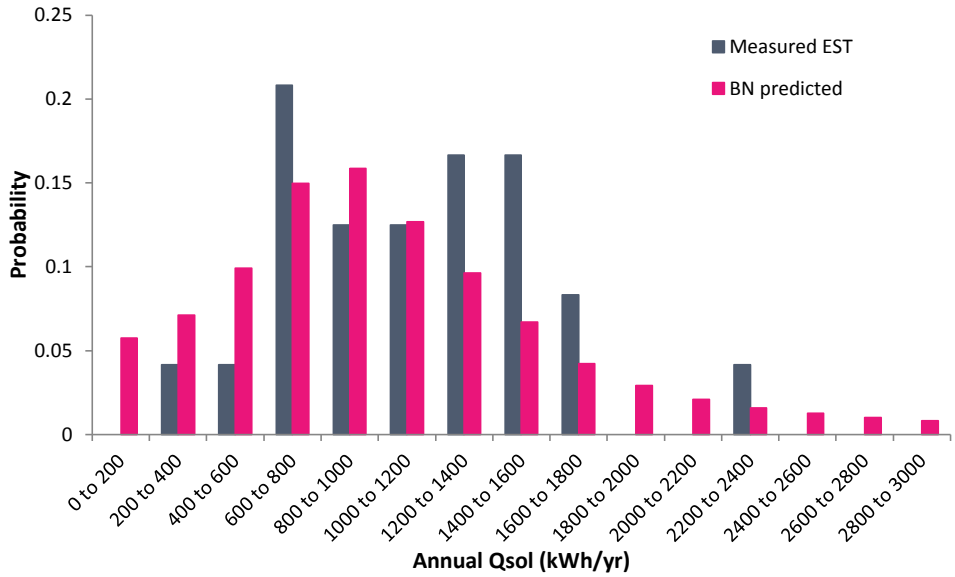

Figure 6.15: Measured and BN predicted total annual solar yield

Adding more data points to the measured distribution for annual solar yield increases its representativeness of different kinds of systems provided there is a mix of system types contained in the dataset. Figure 6.16 shows the BN predicted annual yield distribution versus a distribution of measured annual yield for 86 systems based in the UK and Europe.

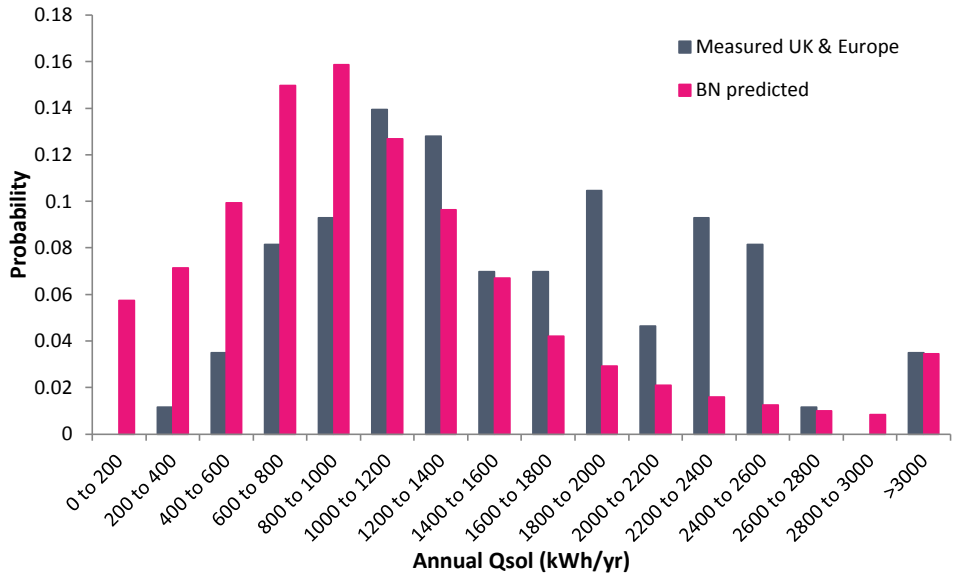

Figure 6.16: Measured (UK & European systems) and BN predicted total annual solar yield

The figure shows that more data records improves the representativeness of the distribution, with unpopulated regions of the distribution in Figure 6.15 now being represented by measured data from European systems.

6.12 Small scale analysis of solar thermal system performance

One major advantage of the BN method of prediction is the ease at which different levels of analysis can be performed. Evidence about a solar thermal system's household, operating environment or system configuration can be readily input into the model and uncertainty in performance updated. Evidence about system parameters enables two small scale analysis perspectives to be taken:

1. Analysis of a single system with performance distributions representing possible variations in annual solar yield;
2. Analysis of a homogenous group of systems in which the performance distribution can also be considered as system-to-system variation in a single year in addition to year on year variation.

Two systems have been analysed using the knowledge-based BN to give annual solar yield predictions. These systems were shown to exhibit large discrepancies between predicted performance using the BREDEM model and measured annual yield (Chapter 7). A summary of the system description is provided in Table 6.11.

	System 501	System 518
Household	Working couple	Retired couple
Occupancy	2	2
Location	Cumbria	Aberdeenshire
Orientation	South	South-east
Pitch	35.5°	23°
Collector area	2.58m^2	5m^2
Collector type	Evacuated tube	Flat plate
Collector performance	η_0 = 0.775 k_1 = 1.476W/m^{2}K k_2 = 0.0075W/m^{2}K^2	η_0 = 0.791 k_1 = ? k_2 = ?
Tank volume	175L	300L

Table 6.11: System description for small scale analysis

6.12.1 Performance results

Annual solar yield distributions are for systems 501 and 518 are presented in Figure 6.17 and Figure 6.18 respectively.

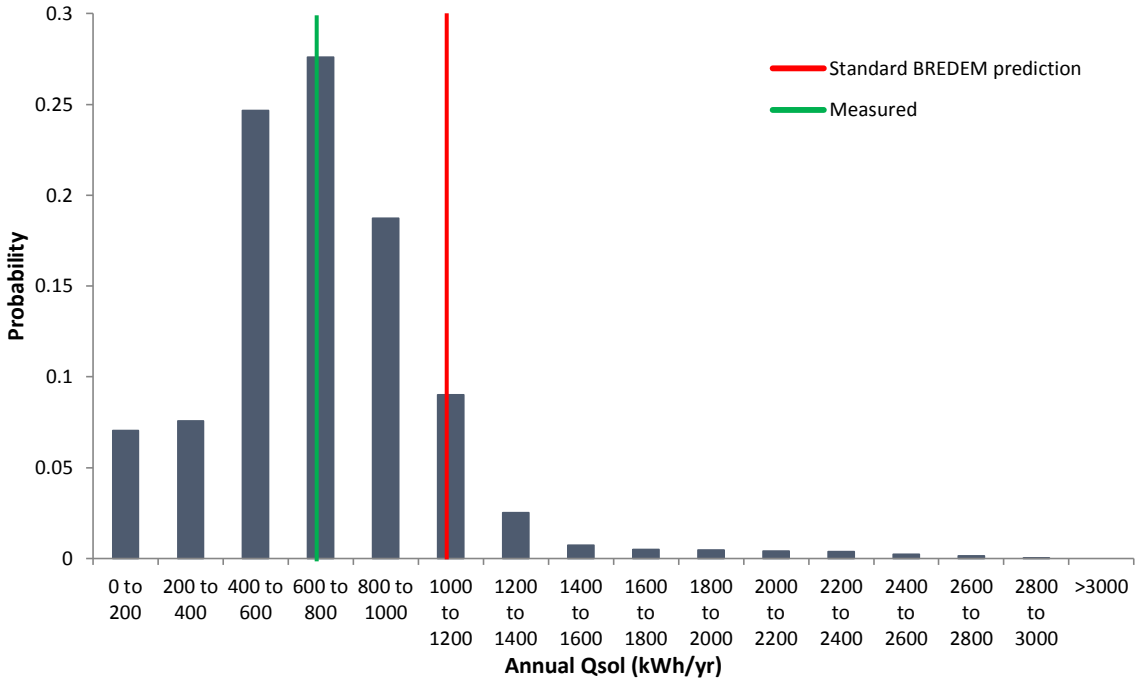

Figure 6.17: Actual vs. predicted annual solar yield for system 501

The knowledge-based BN applied to system 501 shows an improvement on the standard BREDEM prediction in two ways.

1. It appears to show an improvement in the accuracy of predicted annual performance: Based on the mean annual Q_{sol}, the performance discrepancy between the BN and measured yield is -5% versus -43% when the standard BREDEM model is used.
2. It demonstrates the variation in annual performance for the system and quantifies the uncertainty; an alternative perspective would be to view the distribution as a variation in performance on a system-by-system basis.

The distribution also allows the probability of the measured performance and BREDEM prediction to be established: The probability of the measured performance falling within the same range the following year (or of multiple identical systems obtaining performance in the range 600-800kWh/yr) is 28%. The probability of the performance falling within the same range as the BREDEM prediction is 9%. The probability distribution allows us to quantify the uncertainty related to obtaining the same measured performance year on year/system to system as well as the uncertainty related to obtaining the BREDEM estimated performance.

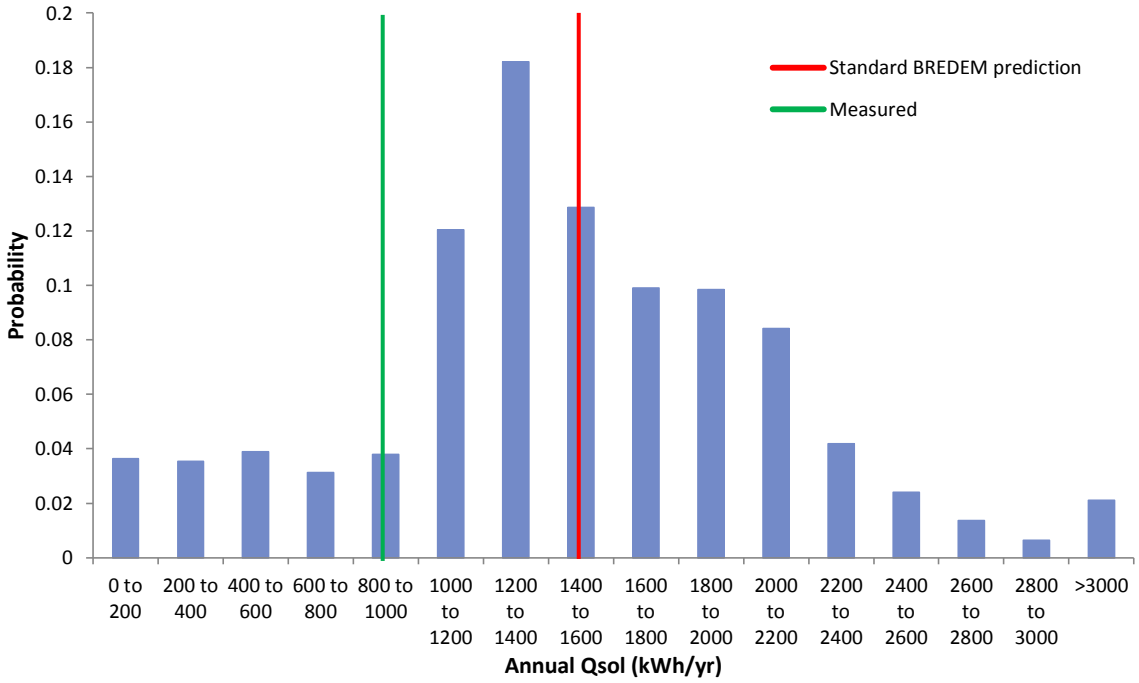

Figure 6.18: Actual vs. predicted annual solar yield for system 518

Figure 6.18 shows a different performance characterisation for system 518 compared to 501. The mean predicted performance is much higher than for system 501 (1580 vs. 703kWh/yr), but the variation in performance is expected to be much greater with standard deviation of 1300kWh/yr vs. 370kWh/yr. This is because annual yield is determined by the product of collector area and specific annual yield. The higher collector area increases the mean annual yield, but also results in the possibility of very high and low yield.

For system 518, it is difficult to make a comment on the accuracy of the BN prediction because there are not enough measured data with which to compare results. There would have to be either more data from identical systems or multiple years' monitored data for this system. This does not mean that the BN results do not provide useful information. The distribution suggests that performance of system 518 is low compared to the mean value, which is corroborated by the BREDEM estimation. The BN distribution allows us to say that this level of performance is lower than expected because of its low probability of occurrence (3.8%), but that it is not impossible. Furthermore, the BN allows us to investigate why this may be by virtue of the two-way application of Bayes' theorem.

By selecting the node state 800-1000kWh/yr for annual yield the parent distributions in the network are updated to reflect the likely values leading to this level of performance. Figure 6.19

and Figure 6.20 show how the distributions for DHW consumption and daily irradiation are updated given the evidence that measured performance is 800-1000kWh/yr compared to the performance shown in Figure 6.18. The figures show that DHW consumption and irradiation are skewed towards the lower ends of the range for annual yield 800-1000kWh/yr. All parent nodes respond to the evidence about the annual performance; however irradiation and DHW consumption exhibit the biggest changes. An in depth analysis of system performance using the backward propagation of uncertainty is made later in the chapter.

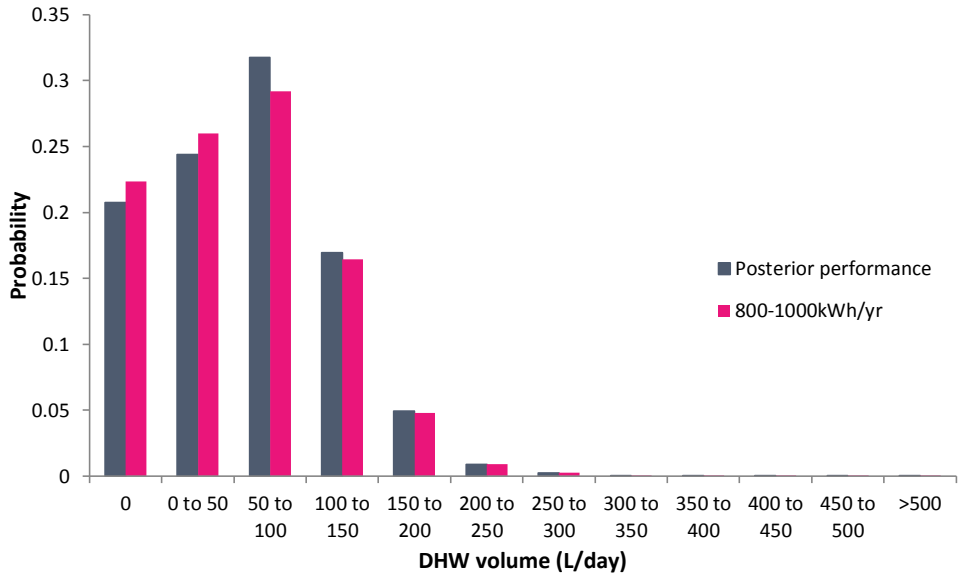

Figure 6.19: Distribution of daily DHW volumetric consumption for posterior distribution and 800-1000kWh/yr annual yield

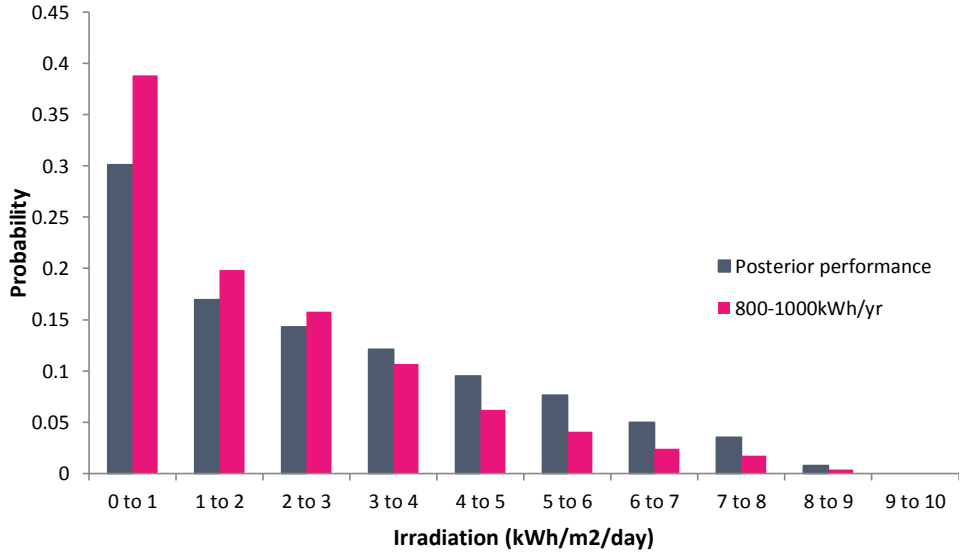

Figure 6.20: Distribution of daily irradiation for posterior distribution and 800-1000kWh/yr annual yield

6.13 Sensitivity analysis

Netica allows the sensitivity of an output node to the input nodes to be determined using various sensitivity measures described in the handbook (Norsys n.d.) referencing work by (Pearl 1988).

The summary report produced by Netica's in-built sensitivity analysis provides information about the reduction in variance of the daily solar yield node by each of the parent nodes; this is presented in Table 6.12.

Node	Variance Reduction (%)
Daily Irradiation (kWh/m^2)	40.8
VA Ratio (L/m^2)	4.42
Daily DHW Volume (L)	0.745
Stratification	0.385
Collector Performance Factor	0.153
DHW Usage	0.0123

Table 6.12: Sensitivity analysis results obtained from Netica showing the most influential variables for solar thermal performance

The reduction in variance shows how sensitive the target variable is to changing the states of a control variable. If the variance in the target reduces when the state of the control variable is known, then it is sensitive to the control variable. If the variance does not change then it is insensitive to the control variable and can be considered independent.

Sensitivity analysis is a useful method in which a BN can be simplified because insensitive nodes can be removed (Shipworth 2006). Table 6.12 shows that daily solar yield is most sensitive to daily irradiation, VA ratio and DHW consumption. If the network were to be simplified for reasons discussed later in the chapter then these nodes would be maintained, whilst the other three would be removed.

6.14 Evaluating STS Performance Using Backward Propagation

This section demonstrates the use of the BN in evaluating in-situ solar thermal system performance under specific ambient conditions. The network effectively and clearly shows the relationships between system parameters and the associated uncertainty, whilst Bayes' theorem facilitates diagnosis of system performance given evidence related to parents of the solar yield node.

6.15 Selecting Ambient Conditions

The number of data points associated with each combination of daily irradiation and external temperature is indicative of the most common weather conditions during the monitoring period 2010-2011. Figure 6.21 shows that throughout the UK between April 2010 to April 2011 the

most common combination of irradiation and external temperature were 0-1kWh/m^2/day and 5-10°C representing 11% of the data[1].

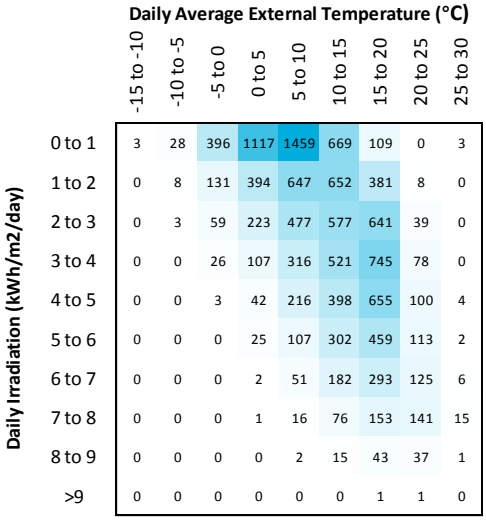

Figure 6.21: Number of data points associated with irradiation and external temperature node state combinations

Figure 6.22 shows the average daily solar yield for each of the weather node state combinations for which more than 20 data records exist. It shows higher solar yields on days with high irradiance and ambient temperature. In addition for each temperature bin, average Q_{sol} increases with daily irradiation. For each irradiation bin an increase in average daily solar yield with external ambient temperature is not apparent except for irradiation level 0-1kWh/m^2/day and 8-9kWh/m^2/day (although this node combination only has two average Q_{sol} values). Figure 6.22 also indicates that above irradiation of 6kWh/m^2/day the variation between average solar yields is relatively small with a standard deviation across these values of 0.1kWh/m^2/day (5% of the mean of these average yields).

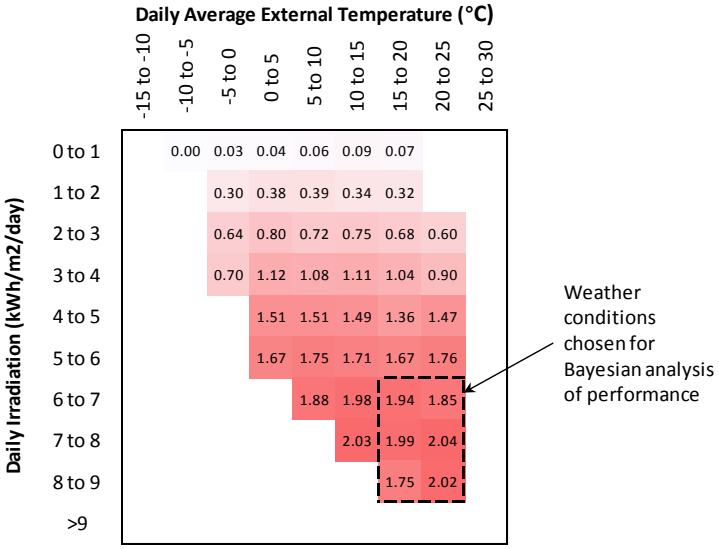

Figure 6.22: Average daily Q_{sol} for different irradiation and external temperature node state combinations

[1] Out of 25915 total data records used in the BN from the EST field trial 13404 provided complete records for daily irradiation, solar yield and external temperature used in Figure 6.21.

In order to use the BN to make inferences about solar thermal system performance it is important to consider two competing aspects:

1. Useful inferences made about system performance should be made under conditions which suit the solar thermal system, i.e. high daily irradiation;
2. A suitably high number of data records associated with the conditions under which the solar thermal system is to be evaluated is required to ensure results are represented by data;

Figure 6.23 shows the average daily solar yield mapped against irradiation and external temperature for which more than 500 data records are present. It shows that the data requirements for valid inferences to be made give data for low to moderate performance. However, owing to the low level of variation between average solar yield for irradiation above 6kWh/m^2/day and different external temperature bins, a subset of these weather conditions were used. For irradiation level 6-9kWh/m^2/day and average external temperature 15-25°C, 792 data records exist and average solar yield is 1.95kWh/m^2/day; these are chosen for performance evaluation using the BN (dashed box in Figure 6.22).

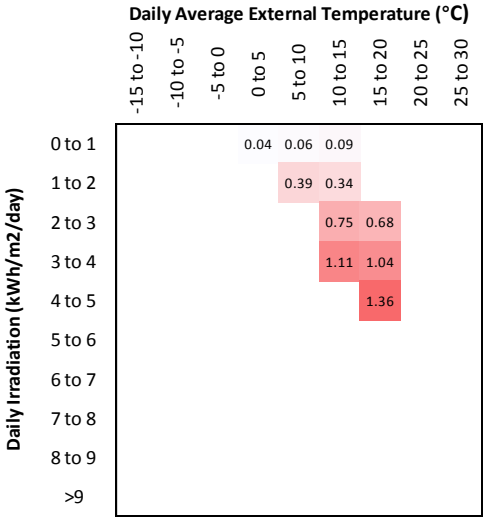

Figure 6.23: Average daily Q_{sol} for different irradiation and external temperature node state combinations with >500 data records

6.16 The Effect of Daily DHW Volumetric Consumption

Volumetric consumption of hot water is a stochastic variable unpredictable by designers and policy makers. It is influenced by climatic and socioeconomic factors including cost of water and energy, the changing size of families and households, and policies designed to encourage reduced consumption (Knudsen 2002; Michaelides 1993; Evarts & Swan 2013). DHW consumption has been shown to influence the performance of solar thermal systems (Knudsen 2002; Furbo et al. 2005; O'Flaherty et al. 2012; EST 2011) and is a key variable in predictive models.

It has been suggested that increased volumetric consumption increases solar yield (Knudsen 2002; Furbo et al. 2005; O'Flaherty et al. 2012; EST 2011); reasons for this include improved

stratification from cold water influx (Furbo et al. 2005) and increased heat transfer from the solar coil (O'Flaherty et al. 2012). Figure 6.24 shows the interquartile range (IQR) and median values of daily Q_{sol} for each bin of DHW volumetric consumption for the selected ambient conditions, the mean is also plotted.

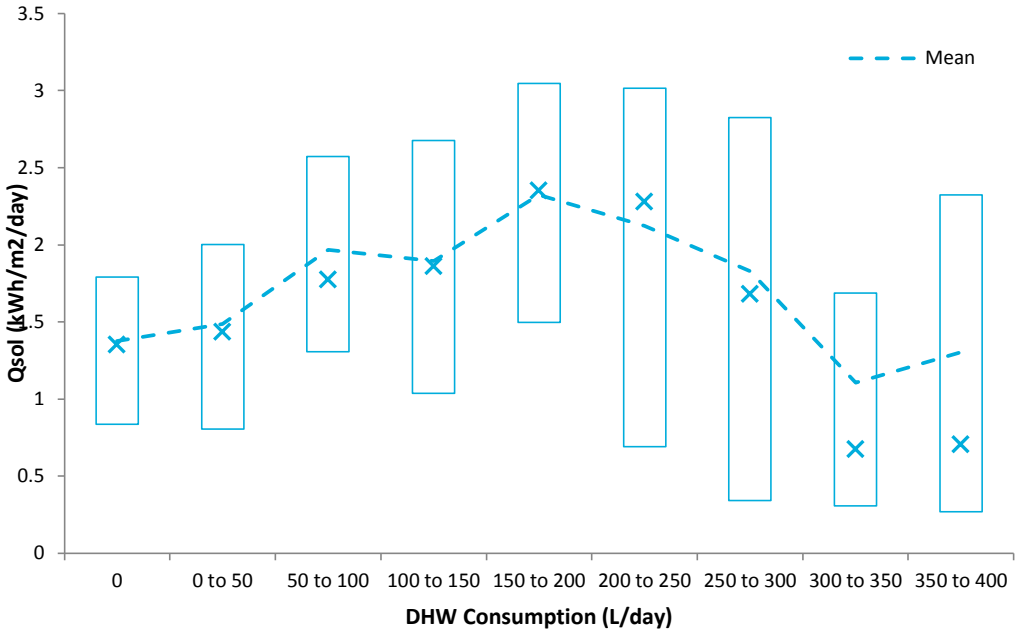

Figure 6.24: Box plot of daily Q_{sol} versus DHW consumption showing 1st and 3rd quartiles and median (crosses) (I = 6-9kWh/m^2/day; T_{ext} = 15-25°C) obtained from the BN

There is a distinct relationship shown suggesting solar yield increases with increasing hot water consumption up to 200L/day. However, consumption rates greater than 200L/day result in a reduction of solar yield. This relationship has not been found in existing literature for two reasons:

1. Hot water consumption is less than 200L/day in most studies (Knudsen 2002; Furbo et al. 2005; Jordan & Vajen 2001; Ayompe et al. 2010; Martin & Watson 2001; Forward & Roberts 2008);
2. This relationship is not modelled by simulation tools, which show an increase in solar yield as DHW consumption increases (Figure 6.25);

Figure 6.25 shows the relationship between mean solar yield and DHW consumption obtained from the BN versus that predicted by the BREDEM and TRNSYS models. The predictive models indicate a steady increase in solar yield with increased hot water consumption which begins to reach a maximum at 375L/day. This is in contrast to the relationship obtained from the BN which approximates probability of Q_{sol} using the EM algorithm and measured data. However there are two potential causes for the unusual relationship demonstrated by the BN:

- The EM algorithm approximates the probabilities for all nodes for which data records are absent – large amounts of missing data are problematic for the EM algorithm which may get stuck in local maxima leading to extreme estimates of probabilities (Lauritzen 1995; Liao & Ji 2009);

- Erroneous measurements or unique systems may introduce bias in the EM algorithm probability approximations for Q_{sol}.

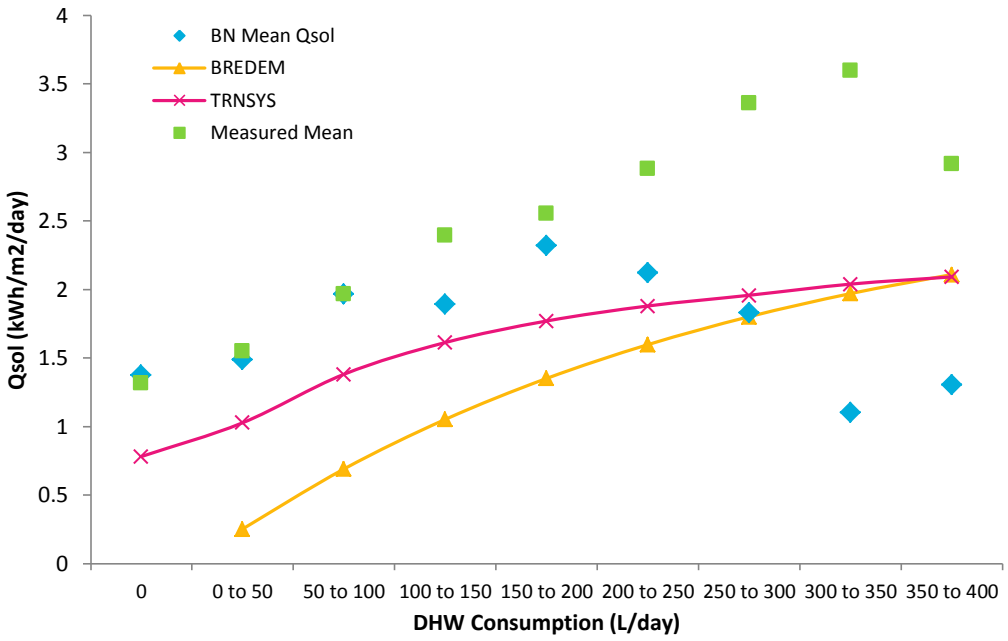

Figure 6.25: Modelled solar yield versus DHW consumption for BN, BREDEM and TRNSYS

It was found that one system (538HOU) in particular appeared to exhibit a large consumption of hot water, which may have been a result of technical issues with the hot water meter. For the ambient conditions specified above, this system showed an average DHW consumption of 389L/day and average solar yield of 1.20kWh/m^2/day. The high water consumption and low performance of this system introduces bias in the EM approximations: data collected from this system dominate the cases where >300L/day hot water is consumed accounting for 50% of the data records. The performance of this system under these conditions is low but comparable to that of the same configuration (513TOM) (Table 6.13):

	System 538HOU	**System 513TOM**
Irradiation	6-9kWh/m^2/day	6-9kWh/m^2/day
V:A ratio	38.46L/m^2	38.46L/m^2
Collector performance	η_0 = 0.586	η_0 = 0.586
	k_1 = 3.9W/m^2.K	k_1 = 3.9W/m^2.K
	k_2 = 0.011W/m^2.K^2	k_2 = 0.011W/m^2.K^2
Average solar yield	**1.20kWh/m^2/day**	**1.80kWh/m^2/day**

Table 6.13: Performance comparison between system 5378HOU and 513TOM

When system 538HOU is removed from the analysis, the measured data gives a more typical response (green squares in Figure 6.25). Note that as the DHW consumption volume increases the number of data records associated decreases, limiting the representativeness of the mean; for example only 6 data records exist for the bin 350-400L/day.

Although the EM algorithm is affected by large amounts of missing data it may provide useful estimates of probabilities where data is missing. For example consider the auxiliary requirement versus DHW consumption for the specified ambient conditions shown in Figure 6.26.

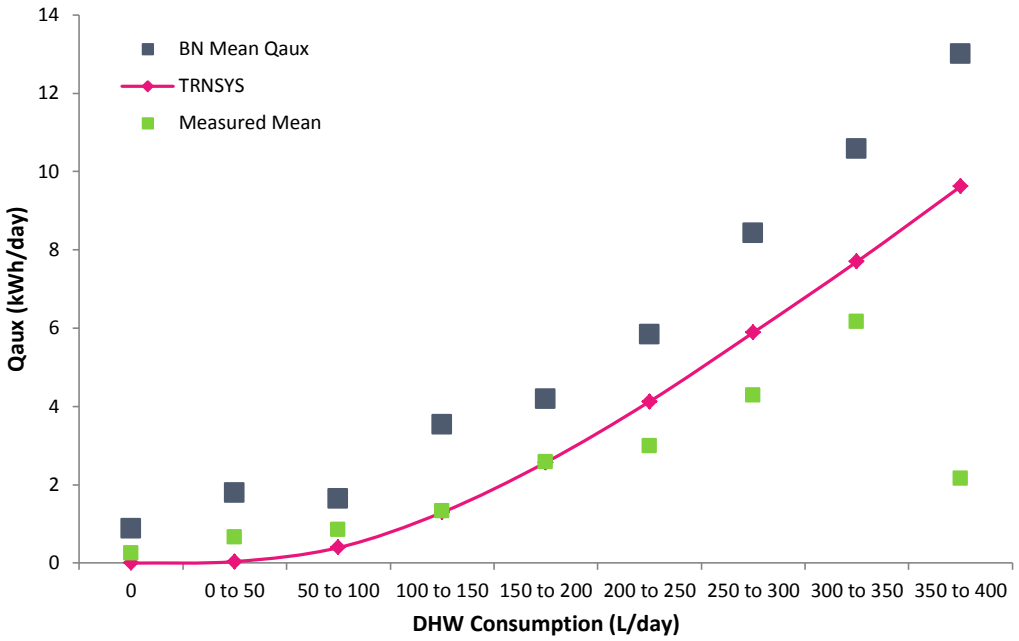

Figure 6.26: Modelled auxiliary requirement versus DHW consumption for BN, TRNSYS and measured data

The TRNSYS prediction shows a steady increase in auxiliary requirement as DHW consumption increases; this relationship is mirrored by the mean values obtained from the BN. However the measured mean data shows a lower level of auxiliary requirement for each DHW bin than the BN and shows the auxiliary input to decrease at 350-400L/day consumption. This decrease is a result of a small number of data records for this DHW condition (6 data points) which may be extreme values. The EM approximation is an improvement on relying on a small number of data points because it approximates the probability of each missing auxiliary input value belonging to each bin in the node given evidence about observed auxiliary input values.

In summary the use of the EM algorithm in the BN can produce unusual predictions of solar yield due to biases in the data and large quantities of missing data records leading to extreme probability approximations. On the other hand when more data is available the EM algorithm can improve predictions of auxiliary requirement compared to measured mean values when missing data is a concern.

6.17 Diagnosis of System Performance

Bayes' theorem allows inference to flow from child to parent as well as parent to child. This is useful from the perspective of designers because it can be used for the following:

- Inferring system design parameters likely to provide a target performance;
- Diagnosis of poor performance under optimum operating conditions

Two cases of solar thermal system performance for the specified ambient conditions were diagnosed: the first and third quartiles of solar yield were compared for the zero-auxiliary case and median auxiliary (0-5kWh/day); Figure 6.27 shows the resulting solar yield distributions for these two auxiliary cases.

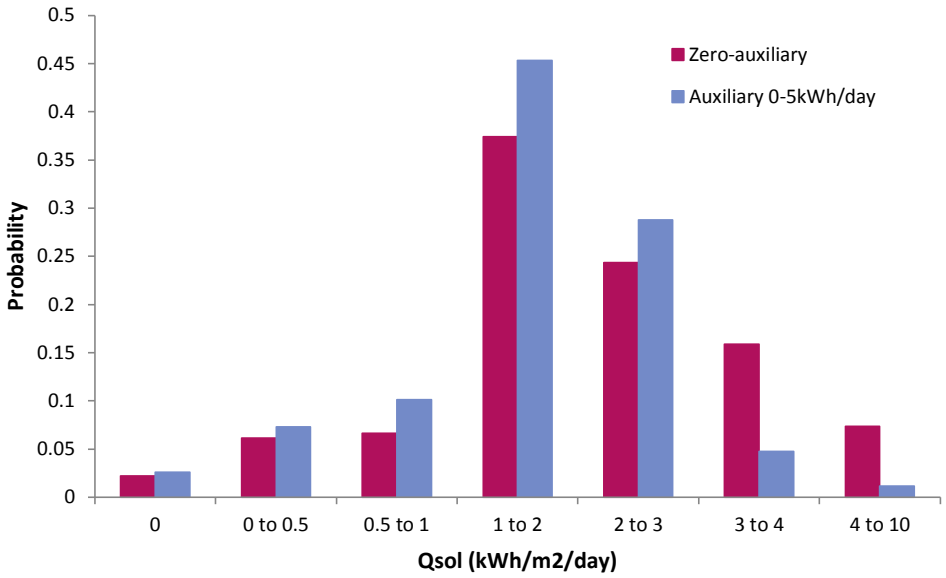

Figure 6.27: Distribution of daily solar yield for auxiliary = 0kWh/day and 0-5kWh/day

Daily solar yield is higher for the zero-auxiliary case with a mean of 2.31kWh/m^2/day (SD = 1.7 kWh/m^2/day) versus 1.74 kWh/m^2/day (SD = 1.1 kWh/m^2/day) for the 0-5kWh/day auxiliary case; however it is not apparent whether the low auxiliary input improves solar thermal performance or whether increased solar thermal performance reduces the requirement for auxiliary heating.

Figure 6.28 and Figure 6.29 show a series of distributions for the parent nodes of solar yield for the zero-auxiliary and 0-5kWh/day auxiliary cases respectively, Table 6.14 summarises the findings.

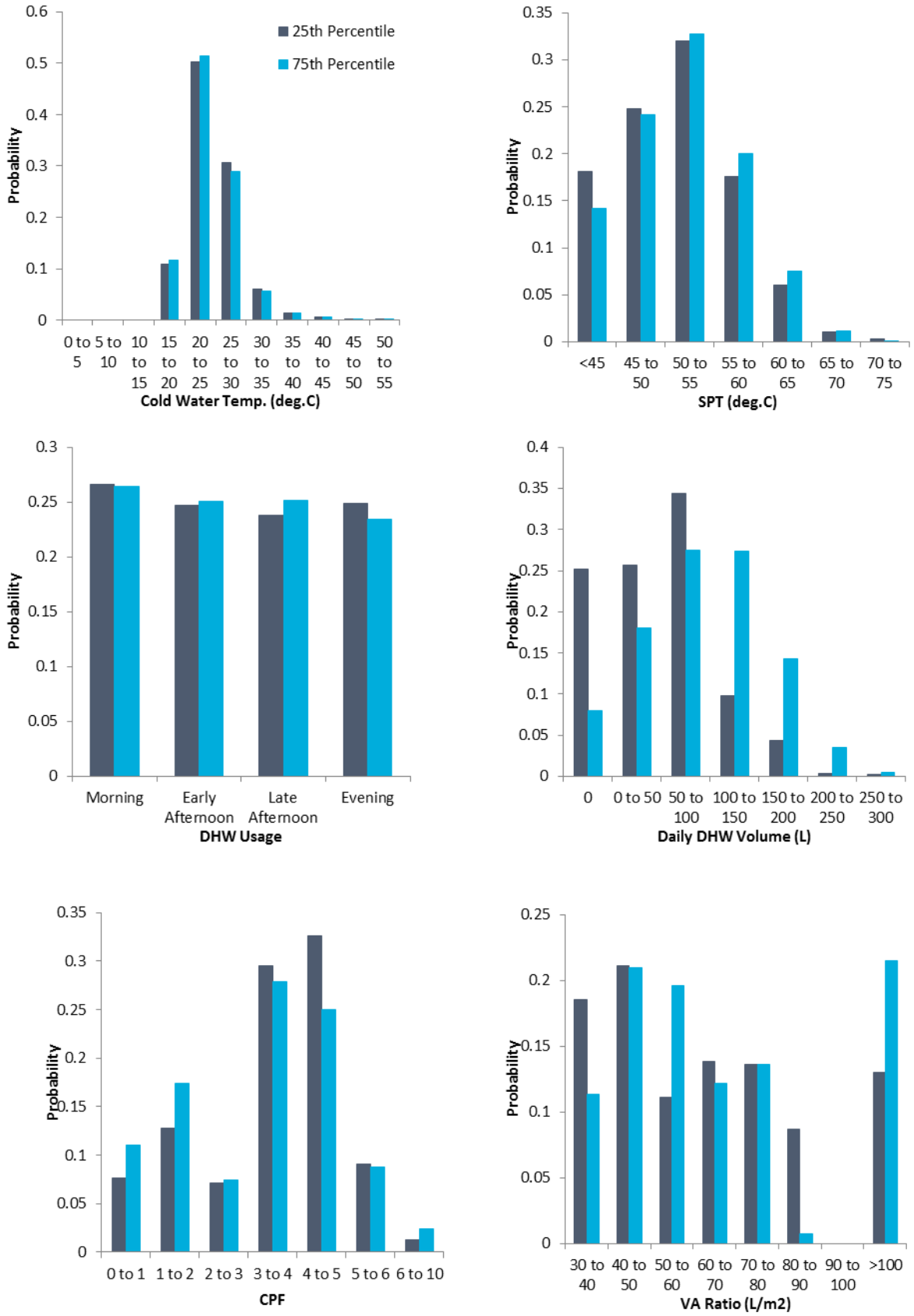

Figure 6.28: Parent node distributions for first and third quartiles of solar yield under the zero-auxiliary condition

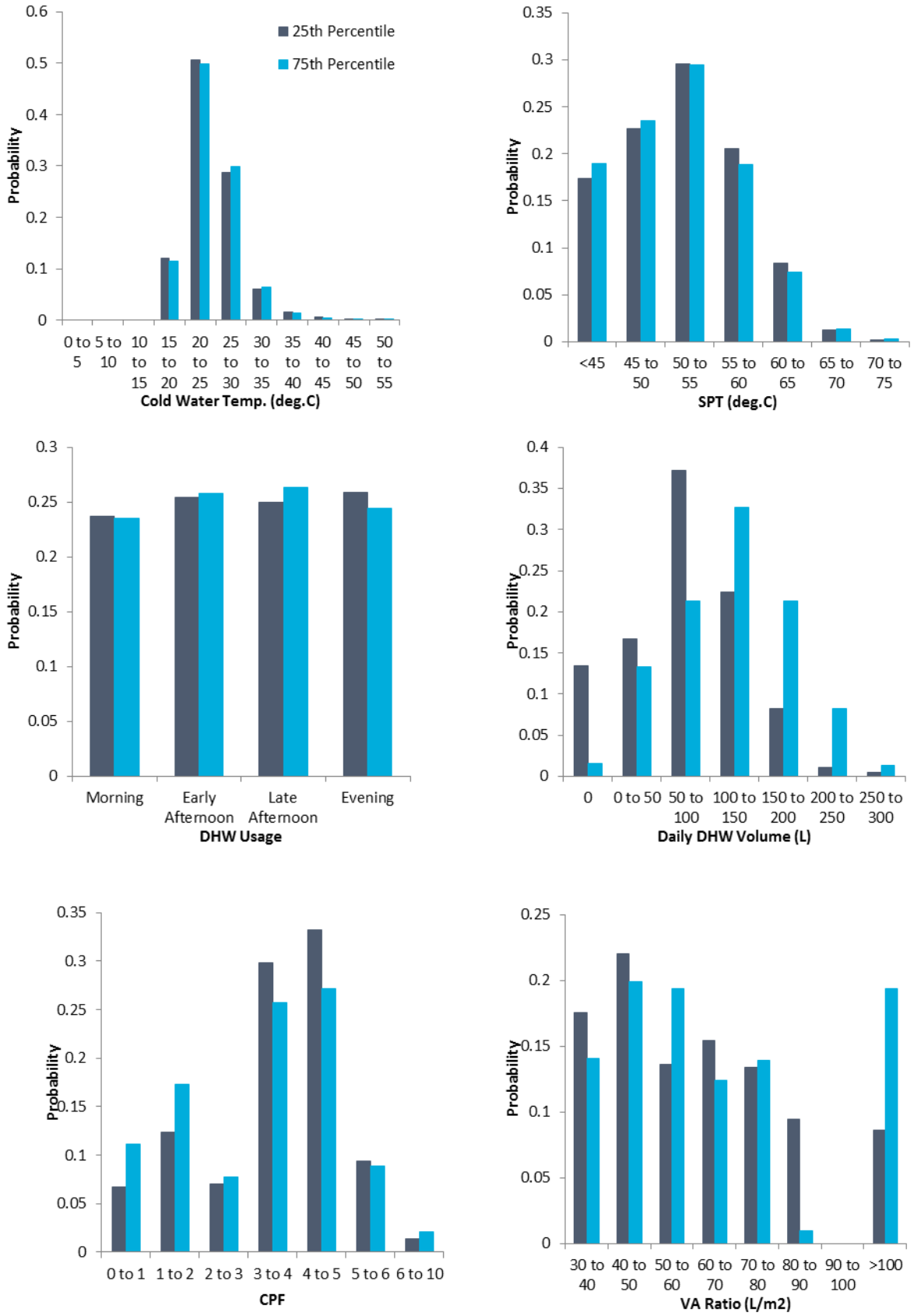

Figure 6.29: Parent node distributions for first and third quartiles of solar yield under the 0-5kWh/day auxiliary condition

	Zero-auxiliary		0-5kWh/day Auxiliary	
	1st Quartile	**3rd Quartile**	**1st Quartile**	**3rd Quartile**
DHW volume (L/day)	53.6 SD = 52	104 SD = 140	81.6 SD = 86	121 SD = 64
DHW usage	Morning	Morning	Evening	Late Afternoon
SPT (°C)	47.6 SD = 14	49 SD = 13	48.3 SD = 14	47.6 SD = 14
Cold water temp. (°C)	24.5 SD = 4.7	24.3 SD = 4.7	24.4 SD = 4.8	24.5 SD = 4.8
V:A ratio	77.8 SD = 64 Med = 59.27	91 SD = 78 Med = 59.02	70.9 SD = 54 Med = 57.69	87.1 SD = 75 Med = 58.25
Collector performance	3.51 SD = 1.5	3.28 SD = 1.7	3.56 SD = 1.5	3.28 SD = 1.7

Table 6.14: Summary of mean performance results

The results suggest that low performance is a result of reduced consumption of hot water with higher mean consumption volumes being found for performance in the 3rd quartile for both auxiliary cases. In addition it can be seen that for the same performance to be achieved when auxiliary is used versus the case of zero auxiliary, DHW consumption must increase; however it may be argued that higher auxiliary requirement is a result of increased DHW consumption. Figure 6.30 shows the distribution of solar yield for the two auxiliary cases when the analysis is fixed to DHW consumption 150-200L/day and V:A ratio of >100L/m^2.

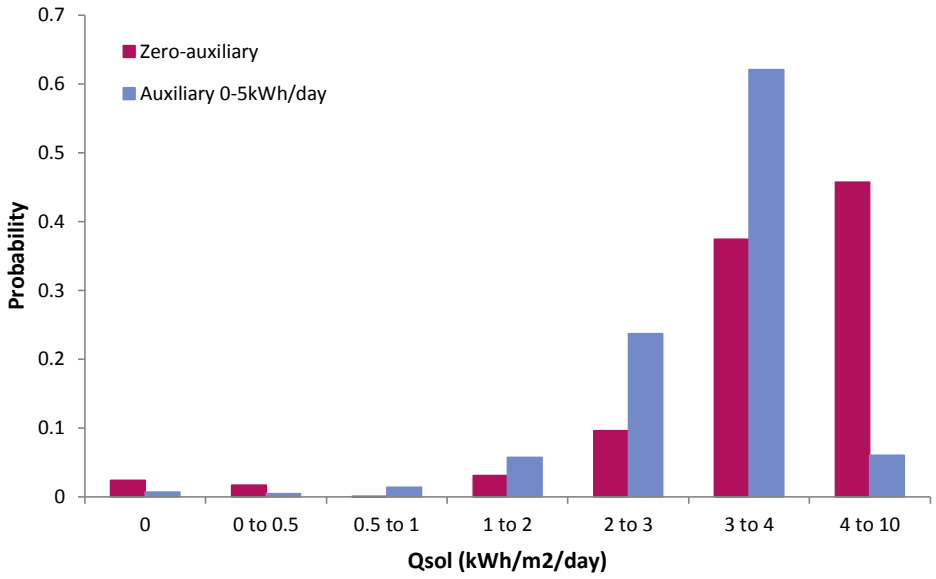

Figure 6.30: Distribution of daily solar yield for auxiliary = 0kWh/day and 0-5kWh/day; DHW consumption = 150-200L/day; V:A ratio >100L/m^2

Mean yield for the zero-auxiliary case is 4.8kWh/m^2/day (SD = 2.5kWh/m^2/day) versus 3.29kWh/m^2/day (SD = 1.7kWh/m^2/day) when auxiliary is 0-5kWh/day. It may be inferred by considering for similar ambient conditions, system sizing and DHW consumption for both auxiliary cases that auxiliary input is causing a reduction in solar yield. This demonstrates that BNs can show at least as much information as a correlation between variables, but that

additional consideration of other system parameters may allow speculation about causality to be made by the modeller.

Higher mean V:A ratios are found for performance in the 3rd quartile versus the lower performance band; However median values vary very little with standard deviation across the medians being 0.72L/m^2. The mean is affected by the probability of V:A ratio being >100L/m^2; for higher performance the probability of V:A ratio >100L/m^2 increases to 22% and 19% for zero-auxiliary and 0-5kWh/day auxiliary input respectively versus 8.6% and 13% for lower performance. Lower performance is characterised by an increased probability of V:A ratio being 80-90L/m^2 being approximately 9% for both auxiliary conditions versus 1% for the higher performance under both auxiliary conditions. An optimum V:A ratio has been stated in other studies (Hobbi & Siddiqui 2009; Shariah & Ecevit 1995; Shariah & Lof 1996); however when inspecting the modelled results of Shariah & Lof (1996) it can be seen that over the range of collector areas and upper tank temperatures experienced in the EST data, solar fraction increases with increasing V:A ratio (Shariah & Lof 1996). (Hobbi & Siddiqui 2009) also show a similar pattern for a 6m^2 system with a set point temperature of 60°C although a decrease in solar fraction is shown for V:A ratios of 200 and 300L/m^2 is shown; nevertheless solar fraction is still within 5% of the stated optimum VA ratio (50-75L/m^2). These high VA ratios are non-existent in the dataset for the BN and contradict current design guidelines (Viridian Solar n.d.).

6.18 Discussion and Conclusions

A BN has been constructed using expert knowledge elicited from literature to develop the causal links and graphical structure, along with field trial and simulated data to develop the node probability tables following the method outlined in Nadkarni & Shenoy 2004; Nadkarni & Shenoy 1999). Missing data values in the case files meant that the count algorithm in Netica could not be used and so the EM algorithm was applied to approximate probabilities of missing values.

The network shows the ability to present two types of uncertainty:

- Uncertainty in performance of a group of heterogeneous systems, known as epistemic uncertainty;
- Uncertainty in the performance of a single system or group of homogenous systems, known as aleatory uncertainty.

This allows both large and small scale analysis to be performed. The network also captures the uncertainty related to the system variables and through Bayes' theorem propagates it through the network to achieve a distribution of performance characterised by uncertain causal factors. Bayes' theorem can be applied in the reverse direction (child to parent) to help explain system performance; this may be used in a diagnostic sense to identify technical and non-technical issues causing lower than expected performance. Alternatively it may be used in a design framework to aid in the design of a system to achieve a target performance. Using the target performance as a measure of project success, risk of failure can be quantified using the performance distributions.

A summary of the findings from system performance evaluation using the knowledge-based BN is given below:

- Measured data, TRNSYS and BREDEM models show similar relationships between Q_{sol} and DHW consumption for the specified ambient conditions i.e. increases solar yield with increased consumption. The BN suggests solar yield decreases for consumption volume >200L/day. This is due to biases in the data as a result of unique or erroneous measurements influencing the probability approximations of the EM algorithm.
- Measured data for auxiliary consumption versus DHW consumption indicates a decrease in additional heating for DHW consumption of 350-400L/day; this is contradicted by TRNSYS simulation results which show an increase in auxiliary input as DHW consumption increases. The reason is a lack of representativeness in the data due to a small number of data points at this consumption level. The EM algorithm approximations indicate much higher auxiliary requirement than measured and TRNSYS simulated results, but follow a typical trend exhibited by the TRNSYS simulation.
- Auxiliary input for the specified ambient conditions is accompanied by a lower solar yield compared to the zero-auxiliary condition. However it is not clear whether this is due to negative feedback from the auxiliary system or as a result of low output from the solar thermal system. Despite this inferences about the direction of causality between these variables might be possible when other system parameters are considered – this is not possible with analysis of correlations as in regression.
- Improved system performance is accompanied by higher DHW consumption and higher V:A ratios.

Limitations to the knowledge based BN exist as a direct result of limited data quality and quantity:

- Large quantities of missing data affect the results of the EM algorithm probability approximations for Q_{sol};
- Possible erroneous measurements limit the confidence in the dataset as well as introducing biases to the probability approximations;
- Data from a relatively low number of systems (71) leads to certain node state combinations being represented by very few or no data records making CPT learning difficult. The result is uniform distributions which offer no useful information (Onisko et al. 2001);
- The direction of causality between solar yield and auxiliary input is not explicitly clear from the data without intensive data manipulation. Relationships between these two factors has been speculated in the EST field trial report but are not easy to obtain from the data (EST 2011).

Potential improvements to the knowledge-based BN model that could be made are:

- Improved data quality and quantity including continual measurement of systems and measurement of systems installed under the RHI, which is not currently offered for solar thermal systems (Ofgem 2014). Feedback of performance data into predictive models is a common suggestion amongst researchers (Turner & Frankel 2008; Menezes et al. 2012; Diamond et al. 2006) and would be particularly useful for BN models that require a great deal of data as the complexity of the network increases (Fenton & Neil 2013);

- Simplification of the BN to remove unnecessary nodes would prevent over-fitting of the data to the model thus increasing the number of data records associated with each node state. Although this would reduce the level of understanding about the system domain by removal of causal factors as well as reduce the accuracy of the model in describing dependences between variables, it would improve the computational efficiency of the model (Bensi et al. 2011; Zhang & Poole 1994; Baker & Boult 1990). This can be achieved by removing insensitive variables (Shipworth 2006);
- CPT development with small datasets can also be improved using the Noisy-OR gate approach described in (Onisko et al. 2001);
- A dynamic BN could be developed to introduce the potentially circular relationship between solar yield and auxiliary input by splitting the analysis into two time frames: pre-solar input and post-solar input.

The concept of parsimony in a BN model is an important consideration when the amount of data is limited (Fenton & Neil 2013): increasing the number of parameters in a BN (or any model) increases the complexity of that model and requires more data to support it. A classic example of this is fitting an n^{th} degree polynomial to explain all data points; when another data point is added the model fails because it has been over fitted to the data. In terms of causal models such as BNs each parameter added is an explanatory cause of some outcome (e.g. the amount of solar yield provided by a solar thermal system). Adding more causes increases the risk of discovering complex relationships when simple ones might be enough for prediction, these complex models will need more data to support CPT development due to increased number of node states.

This chapter showed how a dynamic model constructed in TRNSYS can be transformed into a BN. In this way variation in the system parameters and performance could be displayed clearly and uncertainty updated rapidly. The ability of combining Bayesian network topology with existing modelling techniques presents an alternative improvement to solar thermal system prediction. The next chapter describes the process of transforming the BREDEM solar thermal prediction model into a BN, which enables regression equations and assumptions about parameter values to be replaced with probabilistic relationships and distributions obtained from real data thus incorporating uncertainty. Uncertainty is further incorporated into the BREDEM deterministic relationships by way of sampling prior distributions associated with the input parameters.

6.19 References

Ayompe, L.M. et al., 2010. Validated real-time energy models for small-scale grid-connected PV-systems. *Energy*, 35(10), pp.4086–4091.

Baker, M. & Boult, T.E., 1990. Pruning Bayesian networks for efficient computation. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*.

Bensi, M., Der Kiureghian, A. & Straub, D., 2011. Bayesian network modeling of correlated random variables drawn from a Gaussian random field. *Structural Safety*, 33(6), pp.317–332.

Buckles, W.E. & Klein, S.A., 1980. Analysis of solar domestic hot water heaters. *Solar Energy*, 25, pp.417–424.

Dempster, A.P., Laird, N.M. & Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(1), pp.1–38.

Diamond, R. et al., 2006. Evaluating the energy\nperformance of the first generation of LEED-certified commercial buildings.

EST, 2011. *Here comes the sun: a field trial of solar water heating systems*, London.

EST, 2008. *Measurement of Domestic Hot Water Consumption in Dwellings*, London.

Evarts, J.C. & Swan, L.G., 2013. Domestic hot water consumption estimates for solar thermal system sizing. *Energy and Buildings*, 58, pp.58–65.

Fenton, N. & Neil, M., 2013. *Risk Assessment and Decision Analysis with Bayesian Networks* First., CRC Press.

Forward, D. & Roberts, C., 2008. *Viridian Solar - Clearline Solar Thermal Field Trial*, Watford.

Francis, J.C. & Leitch, R.R., 1989. A feedback control scheme based on causal modelling. *Engineering Applications of Artificial Intelligence*, 2(3), pp.182–189.

Furbo, S. et al., 2005. Performance improvement by discharge from different levels in solar storage tanks. *Solar Energy*, 79(5), pp.431–439.

Greenland, S. & Brumback, B., 2002. An overview of relations among causal modelling methods. *International Journal of Epidemiology*, 31, pp.1030–1037.

Hobbi, A. & Siddiqui, K., 2009. Optimal design of a forced circulation solar water heating system for a residential unit in cold climate using TRNSYS. *Solar Energy*, 83(5), pp.700–714.

Jordan, U. & Vajen, K., 2001. Influence of the DHW load profile on the fractional energy savings: A case study of a solar combi-system with TRNSYS simulations. *Solar Energy*, 69, pp.197–208.

Knudsen, S., 2002. Consumers' influence on the thermal performance of small SDHW systems—Theoretical investigations. *Solar Energy*, 73(1), pp.33–42.

Lam, J.C. et al., 2010. Principal component analysis and long-term building energy simulation correlation. *Energy Conversion and Management*, 51(1), pp.135–139.

Lauritzen, S.L., 1995. The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, 19(2), pp.191–201.

Liao, W. & Ji, Q., 2009. Learning Bayesian network parameters under incomplete data with domain knowledge. *Pattern Recognition*, 42(11), pp.3046–3056.

Martin, C. & Watson, M., 2001. *Side by side testing of eight solar water heating systems*, Department of Trade and Industry (DTI), London.

Mathioulakis, E., Panaras, G. & Belessiotis, V., 2012. Uncertainty in estimating the performance of solar thermal systems. *Solar Energy*, 86(11), pp.3450–3459.

Menezes, A.C. et al., 2012. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97, pp.355–364.

Michaelides, I.M., 1993. *Computer simulation and optimisation of solar heating systems for Cyprus*. University of Westminster.

Nadkarni, S. & Shenoy, P.P., 2001. A Bayesian network approach to making inferences in causal maps. *European Journal of Operational Research*, 128(3), pp.479–498.

Nadkarni, S. & Shenoy, P.P., 2004. A causal mapping approach to constructing Bayesian networks. *Decision Support Systems*, 38(2), pp.259–281.

Naticchia, B. & Carbonari, A., Causal modelling based on bayesian networks for preliminary design of buildings. , pp.293–316.

Neapolitan, R.E., 2003. *Learning Bayesian Networks* First., Prentice Hall.

Norsys, 1995. Netica Application. Available at: https://www.norsys.com/netica.html.

Norsys, Sensitivity Equations. Available at: http://www.norsys.com/WebHelp/NETICA/X_Sensitivity_Equations.htm [Accessed February 12, 2015].

O'Flaherty, F., Pinder, J. & Jackson, C., 2012. Evaluating the performance of domestic solar thermal systems. In *Retrofit 2012*. University of Salford.

Ofgem, 2014. *Do I need metering for the Domestic RHI?*, Available at: https://www.ofgem.gov.uk/publications-and-updates/do-i-need-metering-domestic-renewable-heat-incentive-rhi.

Onisko, A., Druzdzel, M.J. & Wasyluk, H., 2001. Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. *International Journal of Approximate Reasoning*, 27, pp.165–182.

Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems - Networks of Plausible Inference* First., San Francisco: Morgan Kaufmann Publishers, Inc.

Roos, L.L. & Hall, R.I., 2014. Influence Diagrams and Organizational Power. *Administrative Science Quarterly*, 25(1), pp.57–71.

Scavarda, A.J. et al., 2004. A Review of the Causal Mapping Practice and Research Literature. In *Second World Conference on POM and 15th Annual POM Conference*. pp. 612–624.

Shariah, A.M. & Ecevit, A., 1995. Effect of hot water load temperature on the performance of a thermosyphon solar water heater with auxiliary electric heater. *Energy Conversion and Management*, 36(5), pp.289–296.

Shariah, A.M. & Lof, G.O.G., 1996. The optimisation of tank volume - to - collector area ratio for a thermosyphon solar water heater. *Renewable Energy*, 7(3), pp.289–300.

Shipworth, D., 2006. Modelling UK home energy use using Bayesian Networks. *Reading*.

Shipworth, D., 2005. Synergies and conflicts on the landscape of domestic energy consumption: beyond metaphor. In *ECEEE 2005 Summer Study*. pp. 1381–1391.

The Energy Monitoring Company Ltd, 2001. *Analysis of performance data from four active solar water heating installations*, Department of Trade and Industry (DTI), London.

Thur, A., Furbo, S. & Shah, L., 2006. Energy savings for solar heating systems. *Solar Energy*, 80(11), pp.1463–1474.

Turner, C. & Frankel, M., 2008. *Energy Performance of LEED for New Construction Buildings*, Washington.

Viridian Solar, Solar Heating Selector. , pp.1–2. Available at: http://www.viridiansolar.co.uk/Assets/Files/choosing_the_right_solar_panel.pdf [Accessed May 21, 2012].

Washington, S. & Oh, J., 2006. Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: example applied to at grade railroad crossings in Korea. *Accident Analysis and Prevention*, 38(2), pp.234–47.

Zhang, N.L. & Poole, D., 1994. A simple approach to Bayesian network computations. In *Proceedings of the Tenth Canadian Conference of Artificial Intelligence*. pp. 171–178.

Chapter 7

BREDEM-Based Bayesian Network to Predict STS Performance

7.1 Introduction

The following section describes the application of Bayesian methods in parallel with a Monte Carlo approach to incorporate uncertainty into a simplified model for the prediction of solar thermal system performance. The model used is based on BREDEM (Henderson & Hart 2013) and contains deterministic relationships developed from theory (Henderson 2002) as well as regression equations which were developed from data (EST 2008). The deterministic relationships were modelled using Monte Carlo methods to sample the uncertain variables; the regression equations were replaced with probabilistic relationships between dependent variables and regressors using the concept of Bayesian inference. However, some regression equations were modelled using Monte Carlo methods due to the use of intermediate variables in these equations. The entire BREDEM model is graphically represented as a Bayesian network with nodes representing all variables in the model. The Bayesian network was constructed in Norsys Netica which facilitates the use of Monte Carlo methods for deterministic nodes and Bayesian inference for probabilistic nodes. The input variable nodes were populated with data obtained from several data sources (EST 2008; EST 2011; Forward & Roberts 2008; The Energy Monitoring Company Ltd 2001; EST 2001; Ayompe et al. 2011; Mondol et al. 2006). Uncertainty was propagated through the model using Monte Carlo methods and Bayesian inference to allow quantification of the uncertainty in the model output which is the annual solar yield of the system. It was shown that the Bayesian network displayed aleatory uncertainty associated with a single system (also considered as a group of homogenous systems) as well as the epistemic uncertainty associated with a heterogeneous group of systems represented by the un-marginalised network; in this way the Bayesian model can be used to quantify uncertainty on both the large scale and small scale.

The reasons for utilising the BREDEM model are threefold:

- BREDEM-based estimates of performance are expected to be used in the RHI deeming phase (Crowther et al. 2010; MCS 2013) for solar thermal installations, which will not be monitored under the scheme (Ofgem 2014);

- The national calculation method (NCM) (SAP) is based on equations and assumptions used in BREDEM; as such, exemplifying the incorporation of uncertainty in the model using real data in a Bayesian network model can demonstrate potential improvements that could be made to the NCM in the future to overcome issues with discrepancies between actual and estimated building energy performance;
- Using an established predictive model as the foundation of the BN would overcome the limitations of the knowledge-based BN associated with small datasets of questionable quality (Chapter 6).

Several modifications to the BREDEM model are in existence including BREHOMES (Shorrock & Dunster 1997), the Johnston model (Johnston 2003), UKDCM (Boardman et al. 2005), DECarb (Natarajan & Levermore 2007), CDEM (Firth et al. 2010); however only CDEM evaluates the uncertainty in model predictions due to uncertainty in the model inputs. CDEM achieved uncertainty evaluation using sensitivity analysis in which input parameters were altered to assess the impact on the output.

7.2 Incorporating uncertainty into the BREDEM solar thermal system yield prediction using Bayesian networks

BREDEM is a simplified building model used to estimate the energy performance of buildings and their subsystems. The model is modular in its construction and allows for separate estimations of the performance of heating subsystems, including solar thermal systems, to be made. The model contains regression based relationships relating hot water usage, weather conditions and system parameters to the annual solar yield of the system. The solar thermal model and its equations are described more fully in Chapter 4.

Some of the main limitations of the BREDEM model highlighted in Chapter 4 are the assumptions required by the modeller in lieu of measured data and the inability of the model to consider uncertainty in the input parameters and annual solar yield prediction. This chapter demonstrates the role Bayesian networks can play in overcoming these limitations by providing a method by which predictions of solar thermal system performance can be made without prior knowledge about the system being known by inferring a possible distribution of values from an evidence base. Furthermore the method is able to incorporate uncertainty into the final estimation of solar yield using Monte Carlo methods for deterministic nodes and Bayesian inference for probabilistic nodes.

7.3 Developing the structure of the network

Converting the BREDEM model into a Bayesian network is relatively straightforward; each of the input parameters in the model is represented by a node in the network. The primary input parameters in the BREDEM model are those which are required for the calculation of secondary inputs and outputs; they may be assumed values or measured and include: roof pitch; orientation; latitude; number of occupants; shower type; collector performance parameters (η_0, k_1, k_2); collector area; amount of shading; tank configuration; and tank volume. The secondary input parameters can be calculated from the primary inputs using regression equations or obtained from measured data, these include: annual solar energy; DHW volume; DHW energy

content; and volume of solar storage. These are calculated in the BREDEM model using regression coefficients obtained from tables in the model notes and the primary input parameters.

7.4 Developing the CPTs

Three case files for the BN were constructed using annualised data as required by the BREDEM calculations from 219 UK-based households to describe the three subsystem elements: user behaviour; operating environment; and system configuration. Data for the households was obtained from the following field trials (EST 2008; EST 2011; Forward & Roberts 2008; The Energy Monitoring Company Ltd 2001; EST 2001; Ayompe et al. 2011; Mondol et al. 2006). The data in the case files provides prior distributions and probabilistic links for the input variables. Due to the presence of missing data values the EM algorithm was used to populate the CPTs (Chapter 6).

7.5 Calculating solar radiation

In the BREDEM model, the solar radiation is calculated using a regression equation containing the roof pitch, latitude and parameters obtained from tables corresponding to the orientation of the collector (Chapter 4). In the Bayesian network the complex regression relationships relating pitch, latitude and orientation are replaced by input nodes for each of these parameters which are linked with directed arcs to the node for solar radiation (Figure 7.1). The relationship is probabilistic in nature.

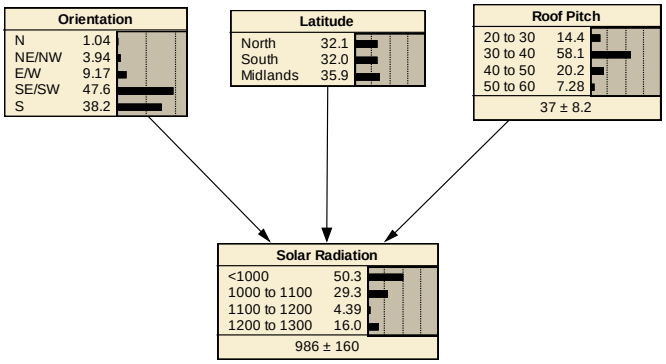

Figure 7.1: BREDEM BN nodes for estimating solar radiation

The relationship between these nodes is probabilistic in nature due to stochastic variation in climatic factors year on year and factors such as the presence of shading objects introducing system-to-system variation; these factors are not incorporated in the standard BREDEM prediction of solar radiation.

Using the BN model, it is unnecessary for the modeller to look up regression coefficients and to calculate pitch factor or orientation factors; instead the modeller simply selects the appropriate values for these primary inputs to update the distribution of annual solar energy. The BN also means that the modeller does not have to make assumptions about these parameters if they are unknown in order to obtain a value for solar energy. However if the modeller knows the annual solar energy from measured data then this can be directly inputted into the model.

7.6 Calculating DHW usage

In the BREDEM model, the DHW volumetric and energetic consumption is calculated based on the number of occupants and assumptions/knowledge made about the bathing and water usage habits of the occupants. The number of baths, type of shower and number of occupants are applied to regression equations to give the volumetric consumption of hot water; this figure is then used with average temperature increase of the hot water to give the energy content. In the Bayesian network model the measured DHW consumption is probabilistically linked to the number occupants using the case file and EM algorithm (Figure 7.2). The advantages of this are:

- Assumptions about bathing habits to make estimations of hot water volumetric consumption are made redundant;
- 100% of the variation in the volumetric consumption of DHW is presented in the model thus overcoming the issues of unexplained variation associated with regression equations;
- The energy content of the water is calculated using the measured average annual temperature rise of hot water for 219 systems as opposed to using monthly scale factors derived from 113 systems (EST 2008). The measured average annual temperature rise is presented as a continuous distribution which is sampled 1000 times thus introducing variation in the deterministic calculation of hot water energy content as a result of uncertainty related to the temperature rise and DHW consumption volume.

A further improvement to the BREDEM model made by the BN method is the ability to combine qualitative and quantitative data. With respect to DHW consumption volume, estimations of the amount of hot water used on average each day is dependent on the type of household as well as the number of occupants. For example a retired couple and a working couple may have different levels of hot water consumption, which can be readily incorporated into the Bayesian inference.

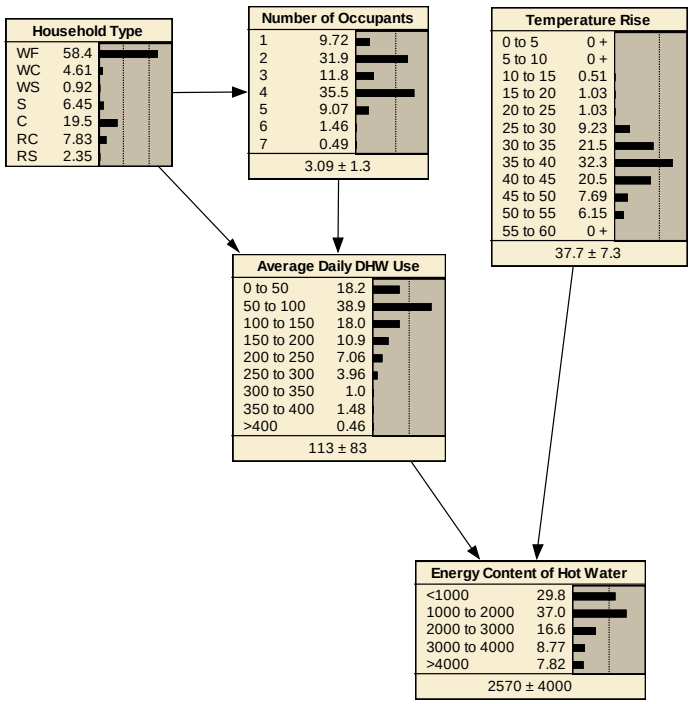

Figure 7.2: BREDEM BN nodes for estimating DHW consumption volume and hot water energy content

7.7 Calculating annual solar yield

7.7.1 Utilisation factor

BREDEM uses the resulting solar radiation and DHW energy predictions to calculate a parameter called the load ratio; this parameter is used in a regression equation to determine another parameter called the utilisation factor, which represents losses due to usage. Figure 7.3 shows the nodes in the Bayesian network BREDEM model involved in the calculation of the utilisation factor.

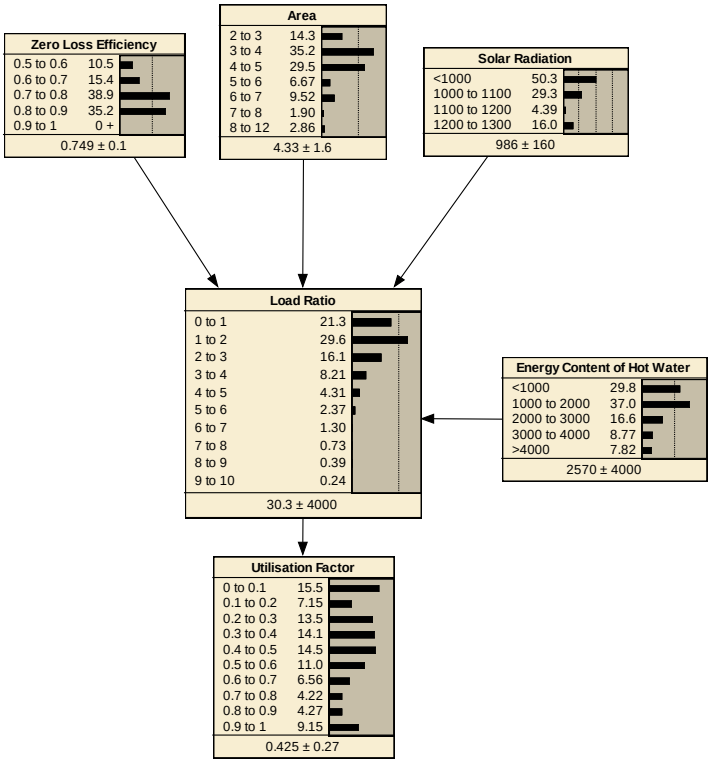

Figure 7.3: Utilisation factor calculation in the BREDEM BN

The input variables to the deterministic equations needed to calculate the utilisation factor are sampled 1000 times. Uncertainty is therefore incorporated in the BREDEM equations using simple Monte Carlo methods.

7.7.2 Performance factor

The collector performance factor is determined based on a regression equation with the variable a^*/η_0 which is the ratio of the collector heat loss coefficient (a^*) to the zero loss efficiency (η_0); the collector performance factor represents losses from the collector due to thermal and optical losses. Figure 7.4 shows the nodes involved in the calculation of the performance factor. The heat loss parameters and zero loss efficiency are probabilistically determined given the type of collector being modelled.

The use of actual collector performance data in the BN provides prior distributions of these parameters which are updated given evidence about the type of collector. If the parameters are unknown to the modeller then this introduces further uncertainty into the performance of the system due to a distribution of possible collector parameters being possible. The modeller is not

forced to use assumed values for collector performance (as in the standard BREDEM approach) in this scenario thus the uncertainty is captured fully and propagated through to the model output.

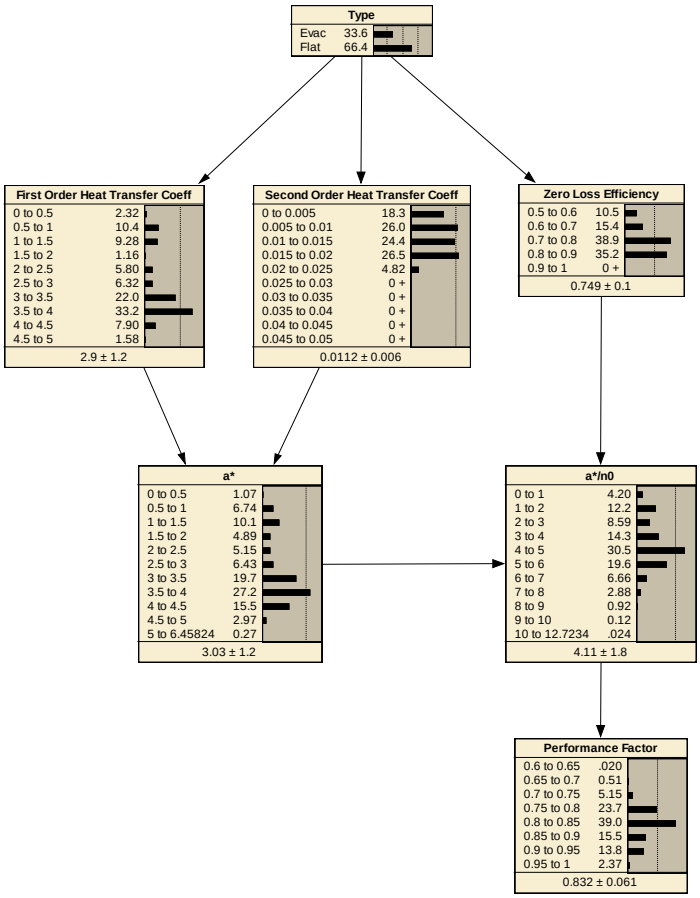

Figure 7.4: Performance factor calculated in the BREDEM BN

7.7.3 Solar storage factor

Storage losses are represented by the solar storage factor determined using the effective solar volume and average daily volumetric consumption of DHW. The nodes involved in determining the solar storage factor in the Bayesian BREDEM model are shown in Figure 7.5.

7.7.4 Annual solar yield

Figure 7.6 shows the nodes used by BREDEM to calculate the annual solar yield of a solar thermal system. By performing the calculation 1000 times a distribution of performance is generated. Figure 7.6 shows the prior distribution of Q_{sol} when no information about the system is known, this is updated as more and more evidence about the system characterisation is known to the modeller.

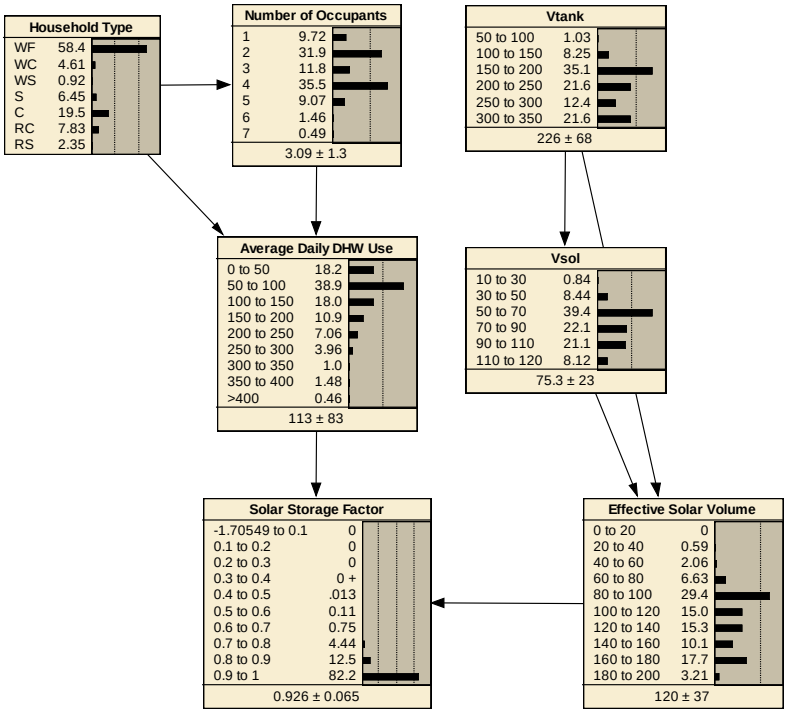

Figure 7.5: Storage factor calculated in BREDEM BN

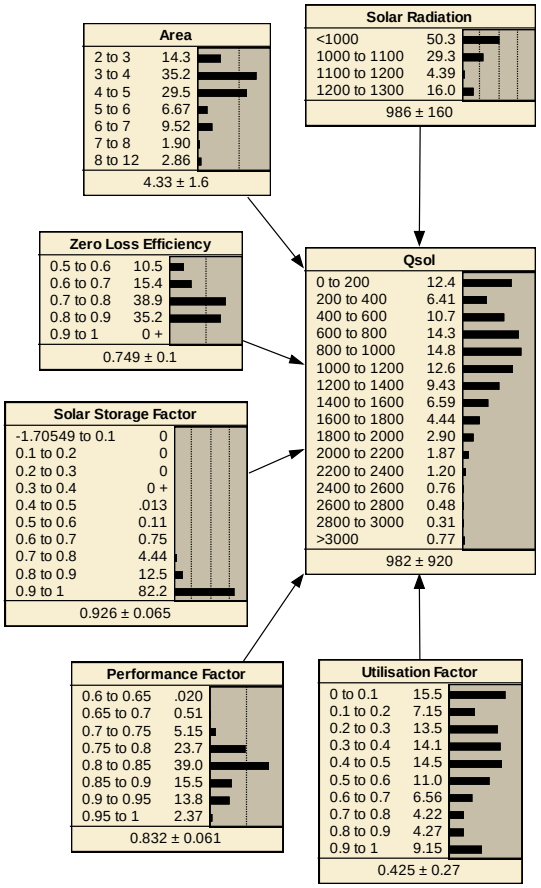

Figure 7.6: Calculation of Q_{sol} using the BREDEM Bayesian network

Detailed information about the equations used in the BREDEM model is provided in Chapter 4.

7.8 Results from the BREDEM model and Bayesian BREDEM

In the Bayesian network model of the BREDEM method a single resulting value for the annual solar yield a system might generate is replaced by a distribution of possible values. This distribution is a result of the incorporation of uncertainty into the model by introducing distributions for each of the parent nodes obtained from data; replacing regression equations with probabilistic relationships using the CPTs; and by performing Monte Carlo-based simulations for the remaining deterministic nodes in the network. Furthermore the assumptions that must be made in BREDEM (obtained from tables) following a lack of evidence about the system characterisation are made redundant by the use of distributions obtained from real data about existing systems.

A comparative evaluation of these two models is presented in this section. Firstly the standard BREDEM model was used to obtain predictions about the annual solar yield for 22 systems taken from the EST solar thermal field trial for which a full year's worth of data was available; these predictions are compared to the measured values. Following this the same systems were modelled using the Bayesian BREDEM model to obtain a distribution of likely annual solar yield. Results from this are presented for two systems shown to exhibit large gaps between actual and predicted performance.

7.8.1 Standard BREDEM results

Two predicted values for the annual solar yield were obtained using the BREDEM model. The first result makes use of the primary data of each system used to predict the values of DHW volume and solar radiation; the second result substitutes measured values for DHW volume and annual solar radiation, in this way the discrepancy between measured and predicted solar radiation and DHW volume is removed from the final calculation of solar yield.

Figure 7.7 shows the predicted versus measured annual solar yield; the black line indicates the case where predicted and measured values are equal.

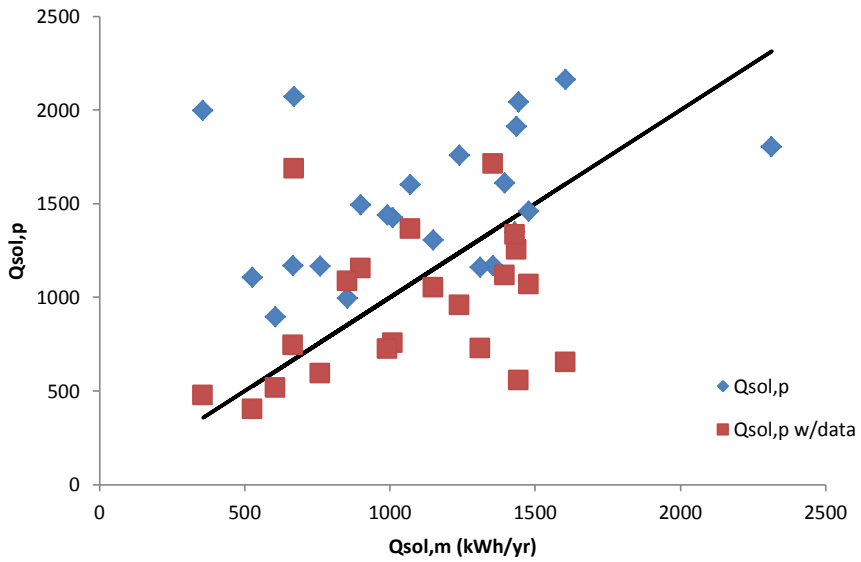

Figure 7.7: Predicted versus measured annual solar yield: red markers show Q_{sol} predictions made with measured data for DHW volume and solar radiation substituted into the model

The blue markers show the predictions made using BREDEM without the use of measured data for DHW volume and solar radiation, these were calculated using the equations detailed in Chapter 4. As indicated by the tendency of the blue markers to be above the black 'y = x line', the BREDEM model over predicts the amount of heat that is generated by the majority of the systems modelled (18/22 systems). The average predicted solar yield is 1504kWh/yr compared to the measured average of 1117kWh/yr. The standard deviations of the predicted and measured yields show that there is greater variation in the measured data 445kWh/yr versus 375kWh/yr.

Predictions are improved when measured data for solar radiation and DHW volumetric consumption are substituted into the model indicated by increased clustering of red markers around the black line. The predicted average solar yield becomes 950kWh/yr with standard deviation of 381kWh/yr, which indicates an average under prediction of solar yield.

Figure 7.8 shows the predicted versus measured annual solar radiation; the average predicted and measured solar radiation values are 1063kWh/m^2/yr and 1015kWh/m^2/yr respectively, which is an over prediction of 5%. The standard deviations for predicted and measured radiation are 103kWh/m^2/yr and 138kWh/m^2/yr suggesting greater variation in solar radiation in reality which will contribute to the increased variation in Q_{sol} values compared to the estimation.

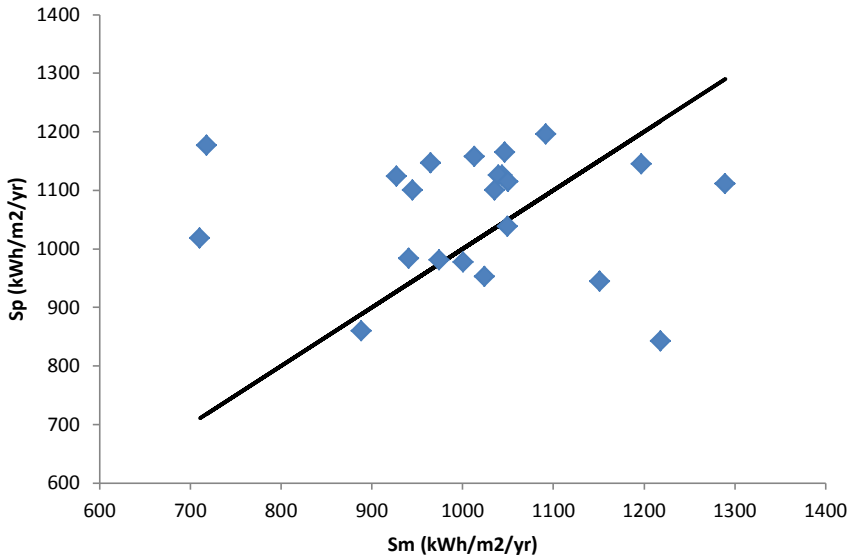

Figure 7.8: Predicted versus measured solar radiation

Figure 7.9 shows the predicted versus measured average daily volumetric consumption of DHW; DHW volume predicted in BREDEM is shown to be over predicted in all but one systems modelled. The average predicted and measured DHW consumption for the systems is 159.86L/day and 97.31L/day respectively, which is a 64% over prediction; this is likely to contribute to the over predicted annual solar yield since increasing the DHW volume increases the solar yield in the BREDEM model.

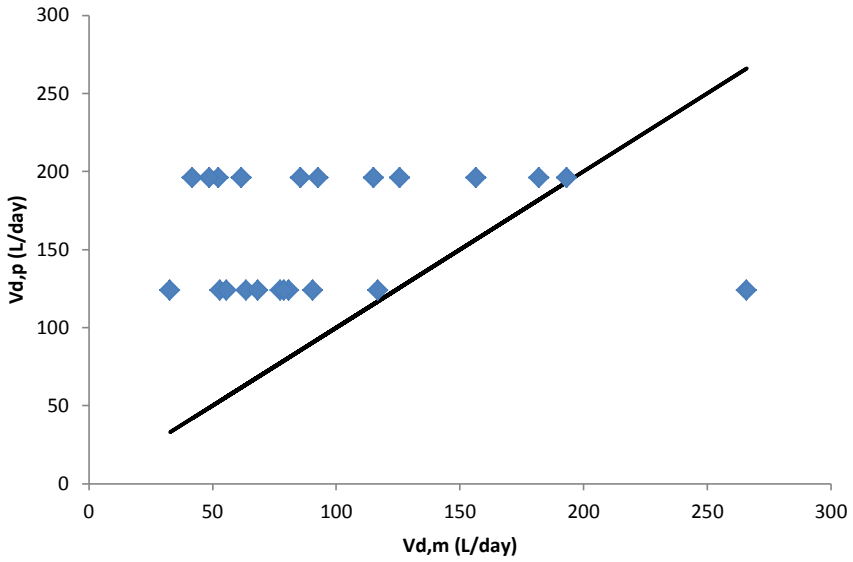

Figure 7.9: Predicted versus measured average daily DHW volumetric consumption

Figure 7.10 shows that the BREDEM model calculates only two different DHW volumes based on the number of occupants in the house (only two occupancy levels existed in the systems modelled, 2 and 4 people); in actual fact for each of these occupancies the DHW volume is far more variable with predicted and measured standard deviations being 36.82L/day and 57.54L/day respectively. This increased variation in real DHW volume values is likely to introduce greater variation in the annual solar yield values.

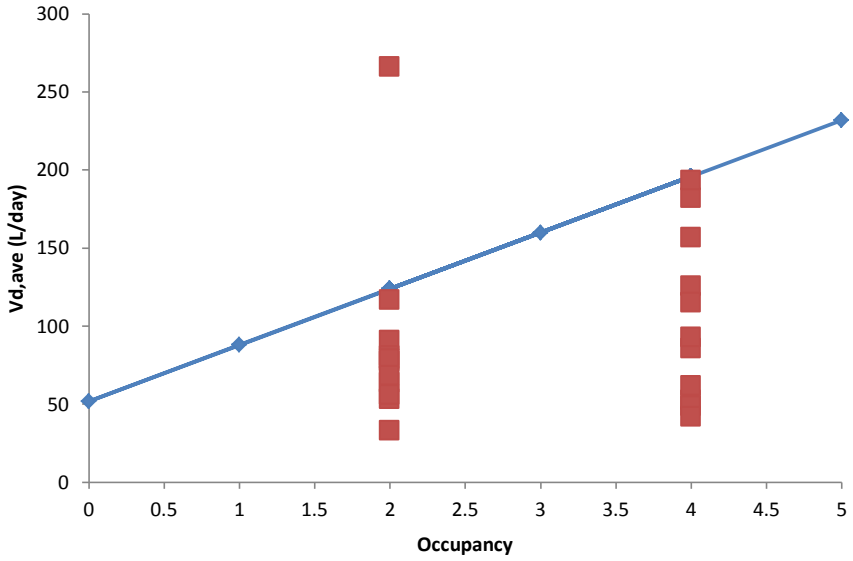

Figure 7.10: Predicted (blue line) and measured (red markers) daily average DHW volume versus occupancy

The red markers show the variation in daily average DHW volume for the two different occupancy levels in the systems modelled; the blue line shows the predicted DHW volume given occupancy produced by BREDEM. Not only is there a tendency to over predict the amount of hot water consumed, but the variation in DHW volume given occupancy means that more occupants does not necessarily mean higher levels of hot water consumption and vice versa.

The resulting performance gaps for the systems are presented in Figure 7.11.

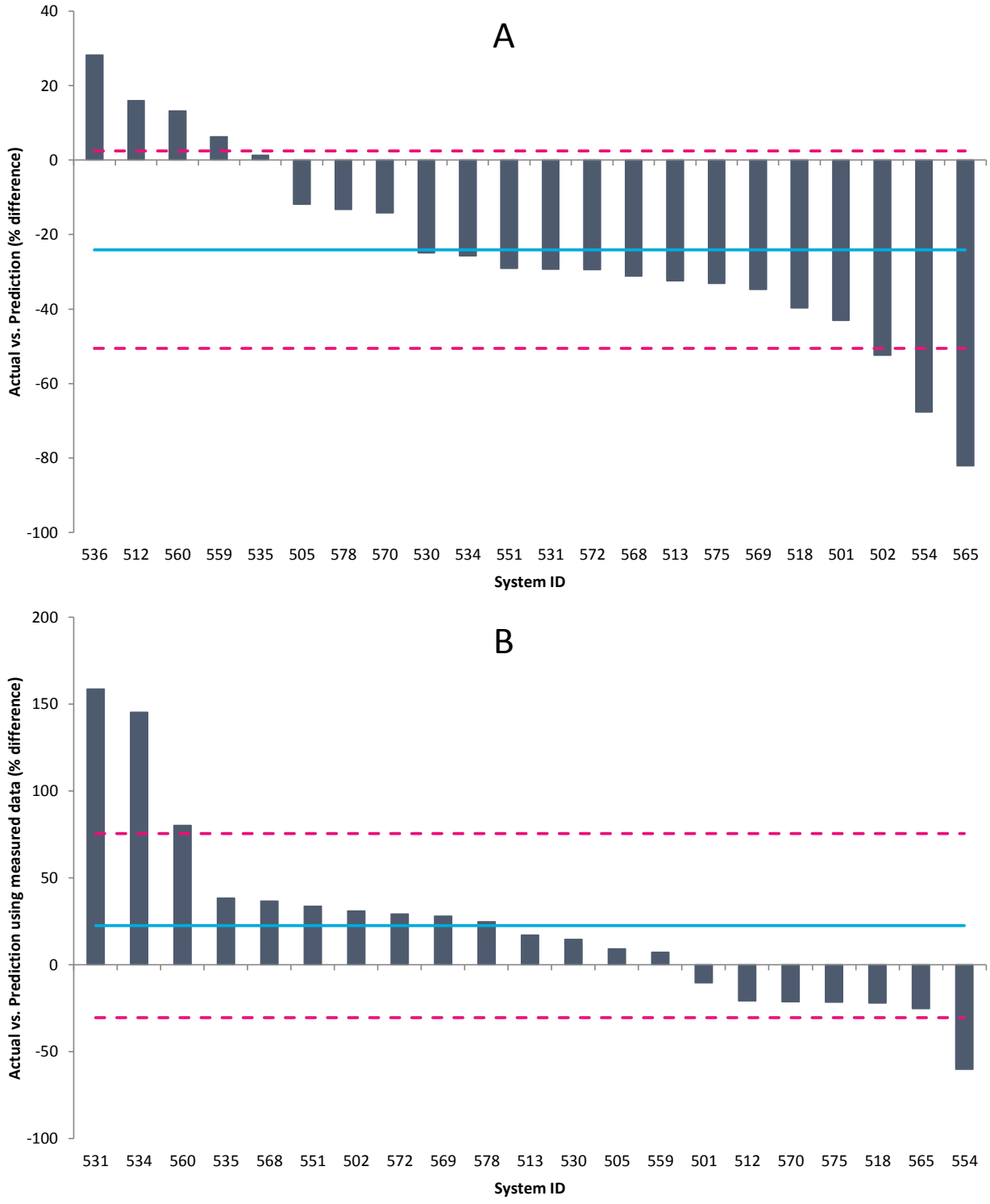

Figure 7.11: Percentage difference between measured and predicted annual solar yield a) standard BREDEM prediction; b) BREDEM prediction with substituted data for DHW volume and solar radiation

Figure 7.11a indicates that for the majority of systems, the actual annual yield is less than the predicted performance in BREDEM with a mean percentage difference of -24.05% and standard deviation of 26.48%.

When real data is substituted into the BREDEM model (Figure 7.11b) the actual annual yield tends to be higher than the prediction with a percentage difference of 22.48% and standard deviation 53%. It is interesting to note that when measured data are substituted into the BREDEM model the predictions of some systems is made worse when compared to the measured performance; this may be due to complex system interactions influencing performance in a way not considered by the BREDEM method.

7.8.2 Bayesian BREDEM results

Two systems have been modelled using the Bayesian BREDEM model; these systems have performance gaps of -39.7% and -43%. The system characterisation is summarised in Table 7.1.

	System 501	System 518
Occupants	2 (working couple)	2 (retired couple)
Location	Kendal, Cumbria	Aberdeenshire
Orientation	S	SE/SW
Pitch	35.5°	23°
Collector area	2.58m^2	5m^2
Collector type	Evacuated tube	Flat plate
η_0	0.775	0.791
k_1	1.476W/m.K	?
k_2	0.0075W/m^2.K^2	?
Tank volume	175L	300L

Table 7.1: Description of systems 501 and 518 modelled in the Bayesian BREDEM model

Measured and predicted average daily DHW volume, annual solar radiation and annual solar yield are summarised in Table 7.2.

	System 501		System 518	
	Predicted	**Measured**	**Predicted**	**Measured**
DHW volume	124L/day	77L/day	124L/day	81L/day
Solar radiation	1196kWh/m^2/yr	1092kWh/m^2/yr	1147kWh/m^2/yr	965kWh/m^2/yr
Solar yield	1170kWh/yr	667kWh/yr	1494kWh/yr	900kWh/yr

Table 7.2: Predicted (standard BREDEM) and measured values for system 501 and 518

From Table 7.2 it can be seen that BREDEM over-predicts DHW consumption and annual solar radiation for both systems, which is likely to be the main reason for the over-prediction of system performance in terms of annual yield.

These systems clearly show the disadvantage of using regression equations to predict highly variable quantities such as DHW consumption volume: The BREDEM model assumes that because the occupancies are the same for both systems then the average daily consumption is also the same; however measured data reveals that this is not a valid assumption. Regressing DHW consumption to the single variable, occupancy explains around 12-13% of the variation in this parameter (EST 2008). This disadvantage is overcome by the Bayesian BREDEM model as a distribution of measured hot water consumption values given occupancy and household type is provided and sampled; the distribution presents 100% of the variation in this parameter given these two variables.

Figure 7.12 shows how the distributions of daily average DHW volume change depending on the type of couple occupying the house (EST 2008; EST 2011); these probabilities are obtained from real data and explicitly captures the variation in this parameter in a way that regression models do not, whilst readily incorporating qualitative descriptions of the household type.

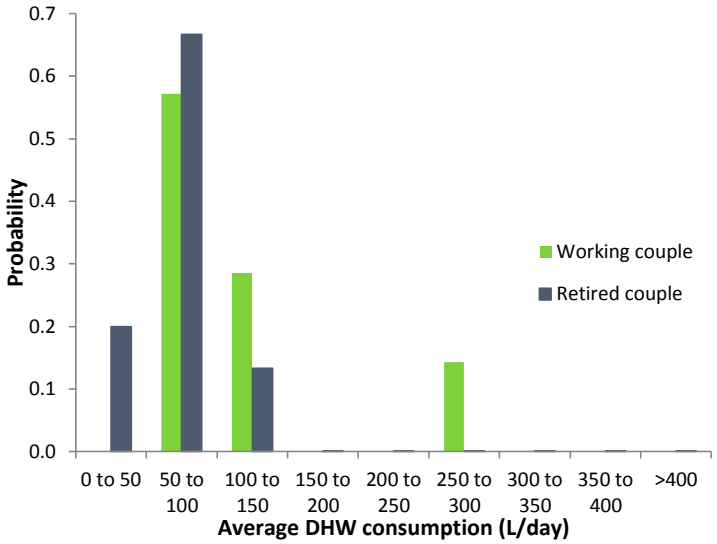

Figure 7.12: Probability distribution of daily average DHW volume for two household types

The figure shows that the predicted mean average daily DHW usage for working couples (system 501) is 118L/day with a standard deviation of 69L/day. The predicted mean average daily consumption for retired couples (system 518) is 71.7L/day with a standard deviation of 32L/day. The indications of this are that different types of household have different DHW requirements with working couples using on average more hot water than retired couples, which may have implications on the performance of solar thermal systems associated with different types of household. However, more data is required in order to increase the representativeness of the distributions.

Further uncertainty is introduced due to the temperature rise of the DHW which affects the energy content of the hot water. The BREDEM model uses fixed values for average temperature rise obtained from data collected in a hot water field trial (EST 2008). The Bayesian model uses this data and appends new data from the solar thermal field trial to provide a distribution of average temperature rises which is sampled for the calculation of energy content.

Although the variation in average daily consumption over a year is incorporated into the Bayesian predictions it is still susceptible to the "flaw of averages". Consider that this quantity is the average daily consumption over an entire year for a sample of systems then the daily variation between consumption volumes is neglected; this is a limitation of the BREDEM methodology rather than the Bayesian approach.

A summary of the predictions from the standard and Bayesian BREDEM models versus the measured annual solar yield of the two systems is presented in Table 7.3.

	Standard BREDEM	Bayesian BREDEM	Measured
System 501	1170kWh/yr	840kWh/yr (SD = 390kWh/yr)	667kWh/yr
System 518	1494kWh/yr	752kWh/yr (SD = 460kWh/yr)	900kWh/yr

Table 7.3: Predicted and measured annual solar yield for systems 501 and 518

The Bayesian model suggests that the most likely annual yield for system 501 is between 600-800kWh/yr with a probability of occurrence of 23.5% (Figure 7.13a). The probability of the system performing as the standard BREDEM model estimates is 14.9%.

The Bayesian model suggests that the most likely annual yield for system 518 is between 800-1000kWh/yr with a probability of occurrence of 17.1% (Figure 7.13b). The probability of the system performing as the standard BREDEM model estimates is 3.7%.

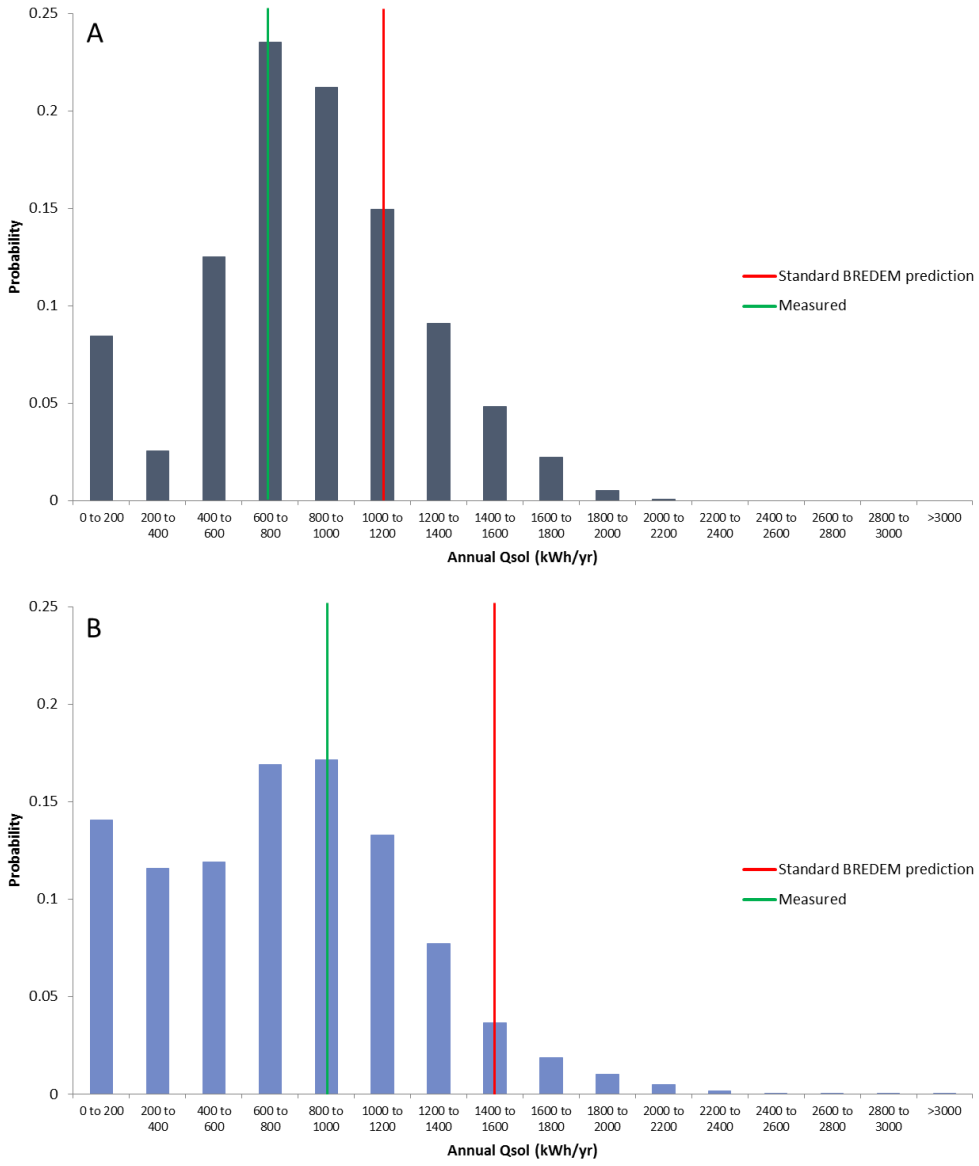

Figure 7.13: Bayesian BREDEM predicted distribution of annual solar yield for a) system 501; and b) system 518; the red line indicates the point estimate from the standard BREDEM model; the green line represents the actual measured annual solar yield

Combining the BREDEM and BN methods effectively closes the performance gap associated with actual versus predicted annual solar yield. Instead, probabilities are provided that can be used to determine the likely yield of a system or group of systems. The distributions can be used to quantify the risk of not obtaining a designed-for yield in a way that point estimate models such as BREDEM do not offer. Continued monitoring of systems would improve the predictions by reducing the proportion of missing data records and increasing the representativeness of the data.

The figures above indicate that because carbon performance of solar thermal systems is linked to the annual yield, it too is subject to annual and system-to-system variability. This is an important consideration when current point estimates of energy and carbon performance may be quite different from the actual situation at any given time.

The incorporation of uncertainty drastically changes the way in which investment decisions are made by more accurately quantifying risk (Booth & Choudhary 2013). For example performing an NPV analysis on solar thermal systems based on the standard BREDEM model prediction yields very different results compared to the use of a probability distribution of performance given by the Bayesian model. To demonstrate how incorporating uncertainty into the performance estimation of solar thermal systems allows for risk-conscious decision making in a financial sense, a financial analysis of two of the EST solar thermal systems is presented in the next section.

7.9 Discounted Cash Flow (DCF) analysis

The domestic renewable heat incentive (RHI) is a mechanism developed by the UK government to increase the uptake of renewable heating technologies, such as solar thermal systems, by homeowners. The homeowner receives payment per kilowatt hour of heat generated paid over 7 years. The amount of heat generated is deemed (estimated) although householders are encouraged to install a monitoring platform. The estimated heat generated by a solar thermal system is obtained using a SAP-based model with some modifications made to address the issues with the SAP 2009 which arose due to its purpose for compliance rather than a prediction tool; these include entering the number of occupants rather than estimating based on total floor area, and regional estimations of solar radiation (MCS 2013).

In this section a financial analysis is performed comparing the returns on investment of the two case study solar thermal systems detailed above. The value of the investments will be evaluated using DCF analysis in which the net present value (NPV) is calculated. The estimated heat generated by the standard BREDEM model and the Bayesian BREDEM model are applied to a NPV calculation; it highlights the limitations of the RHI in making solar thermal systems a financially viable option for domestic users and shows how the annual variation in performance can impact on the NPV of the investment which can lead to overpayment. Despite this overpayment neither of the two systems will see a return within 20 years based on standard and Bayesian BREDEM estimations of performance.

7.10 The NPV method and assumptions

Net present value is a widely accepted method of determining the financial viability of an investment (Abu-Bakar et al. 2014)and is given by:

$$NPV = C_0 + \sum_{t=1}^{20} \frac{I_t}{(1+r)^t} \quad (7.1)$$

Here C_0 is the capital cost of the system and is a negative value; t is the number of years after the installation, a lifetime of 20 years is assumed which is the average lifespan of a solar thermal system (Abu-Bakar et al. 2014); r is the discount rate assumed to be 3.5% (DECC 2013); and I_t is the net income of the investment in the year t.

The income I is a function of the performance of the solar thermal system, the RHI tariff, savings made by offsetting fossil fuel consumption, and the cost to maintain the system:

$$I_t = (Q_{sol,t} \times R_{RHI,t}) + \left(Q_{fs,t} \times R_{f,t}\right) - (C_{m,t}) \quad (7.2)$$

The performance of the solar thermal system is estimated using the standard and Bayesian BREDEM models. For the standard BREDEM estimation $Q_{sol,t}$ is the same for each year in the 20 year period; for the Bayesian BREDEM model 20 points are sampled from the distribution of annual performance. $R_{RHI,t}$ is the tariff rate paid by the RHI scheme which is currently 19.2p/kWh; this increases each year linked to the retail price index (RPI) assumed to be 3% (Cherrington et al. 2013). The amount paid by the RHI tariff is based on the standard BREDEM estimated yield to reflect the fact that solar systems will not be monitored. The savings on fossil fuel use (gas in the case of the system 501 and oil for system 518) is given by $Q_{fs,t}$ and is equal to the amount generated by the solar thermal system; $R_{f,t}$ is the price of gas (or oil) in year t assumed to be 4.9p/kWh (oil = 5.36p/kWh (EST 2014)) in the first year and increasing 8% each year (Abu-Bakar et al. 2014; Cherrington et al. 2013). The cost to maintain is assumed to be £64 in the first year (Abu-Bakar et al. 2014) and is also index linked. Figures used in the analysis are summarised in Table 7.4.

Discount rate, r	3.5%
RHI tariff rate, R_{RHI}	19.2p/kWh
RPI	3%
Fuel cost, R_f	Gas = 4.9p/kWh Oil = 5.36p/kWh Electricity = 14.39p/kWh
Fuel price increase	8%/yr
Capital cost, C_0	Flat plate = £700/m^2 Evac. tube = £1000/m^2
Initial maintenance cost, C_m	£64/yr

Table 7.4: Summary of figures used in NPV calculations

In order to perform the NPV analysis using the Bayesian BREDEM performance estimations the performance distribution was sampled 20 times to obtain 20 years' worth of Q_{sol} values and applied to the NPV calculation; this procedure was performed 1000 times in order to build up a distribution of NPV values and RHI revenues.

Figure 7.14 shows the distribution of NPV for the two systems 501 and 518 with the NPVs determined using the standard BREDEM estimation shown by a solid line.

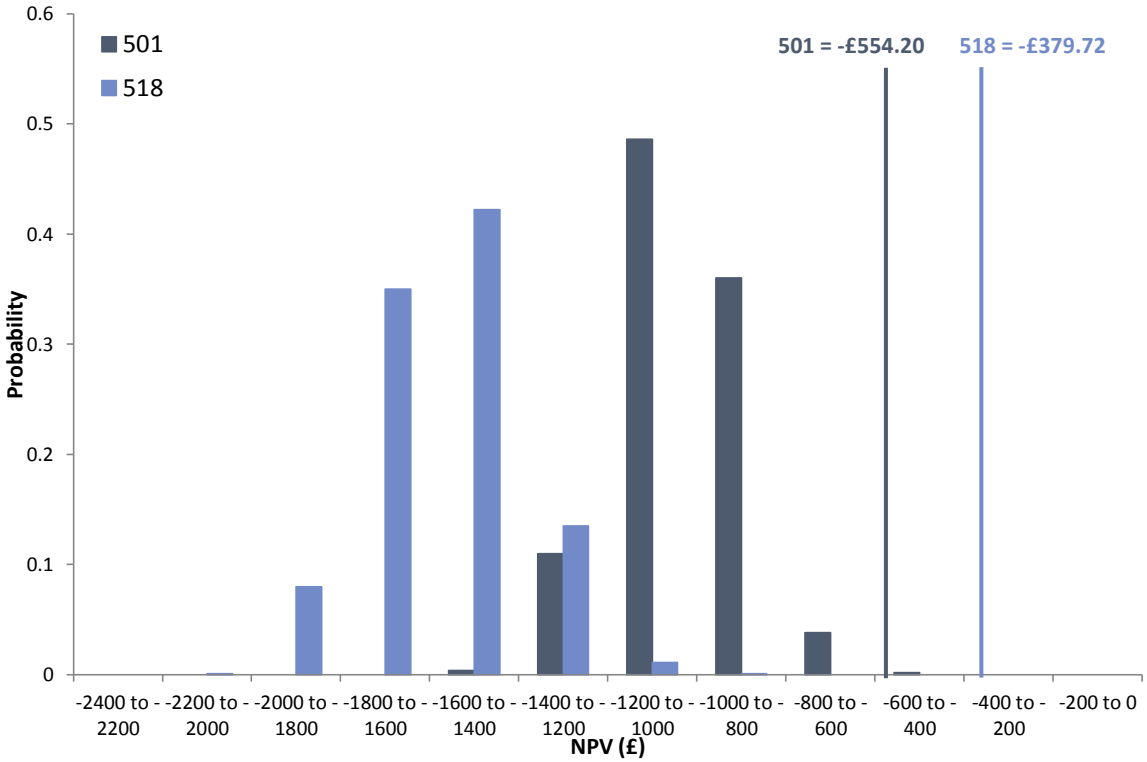

Figure 7.14: Distribution of NPV for system 501; and 518 (lines show the NPV calculated using the standard BREDEM annual Q_{sol} prediction)

The figure shows that using the standard BREDEM prediction gives a more optimistic value of the investment and is greater than any of the values in the NPV distribution for system 518 and in the top 5% of values for system 501. Despite this over estimation of NPV, neither of the systems appears to be financially viable as neither obtain a positive NPV. System 501 is a better investment than 518 with an average NPV of -£1033.81 (SD = £135.74) versus -£1570.56 (SD = £163.23), but as mentioned previously no system is financially viable under the current RHI scheme.

Figure 7.15 shows the distributions of total revenue generated by the RHI scheme for the two systems if it was paid based on measured performance. The solid lines show the RHI revenue generated using the standard BREDEM prediction of performance. The results show that if RHI payments were based on measured performance system 501 would have a 3% probability of obtaining the revenue predicted by the standard BREDEM model whereas system 518 would be highly unlikely to achieve the predicted RHI payment. The measured performance based revenue generated by system 501 is on average higher than that of system 518 being £1882 (SD = £211) versus £1102 (SD = £248); this is due to an average higher performance as a result of a higher DHW consumption and annual solar radiation.

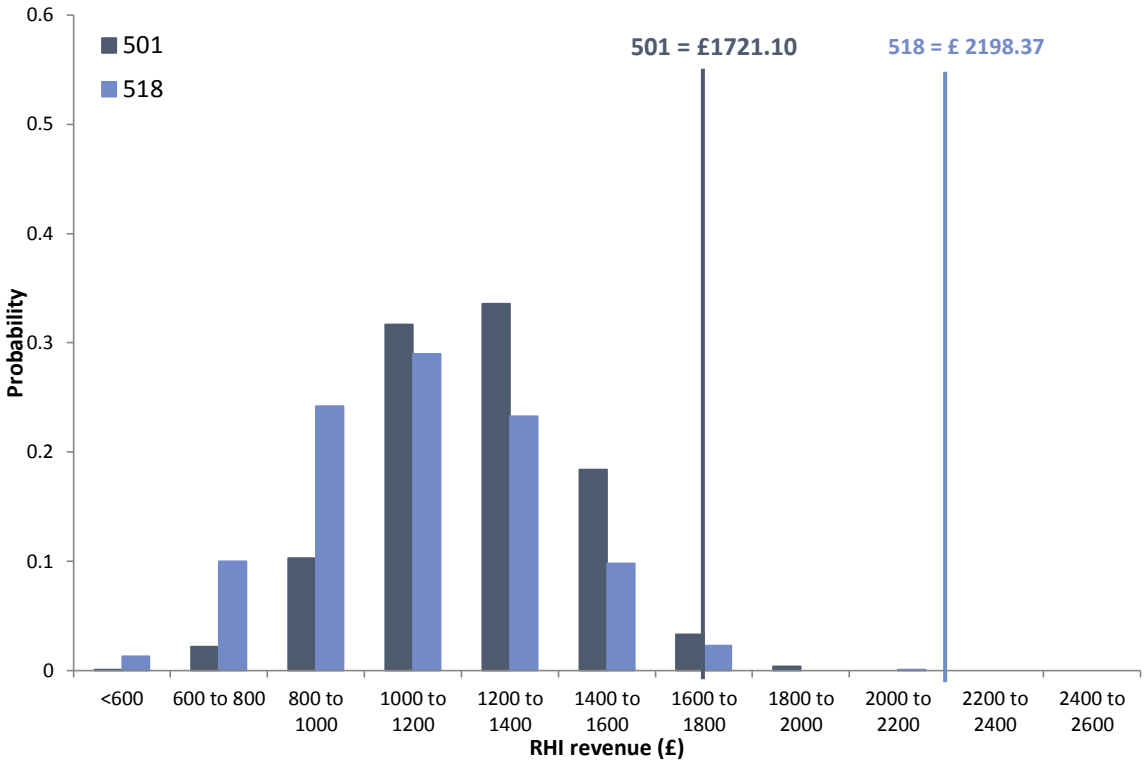

Figure 7.15: Distribution of revenue generated by RHI for system 501 and 518 based on measured performance (lines shows the RHI revenue calculated using the standard BREDEM annual Q_{sol} prediction)

Neither of the above systems makes financial sense according to the results of the NPV analysis; this is due to a number of factors:

- The cost of gas and oil is low (4.9p/kWh and 5.36p/kWh), which minimises the savings that are made especially considering variable performance from year-to-year;
- The RHI tariff is not high enough and does not cover the initial cost of the systems assumed to be £2580 and £3500 (systems 501 and 518 respectively), or maintenance costs;
- The duration for which RHI payments are made is too short again leading to failure to cover the initial and maintenance costs of the systems;

Abu-Bakar et al. (2014) show from a similar NPV analysis that the current RHI scheme does not make solar thermal systems financially viable and suggest that the tariff should be increased to 32p/kWh or should run for a minimum of 17 years rather than just 7; furthermore it is shown that solar thermal systems make better financial sense when they are used to offset consumption of a more expensive fuel such as electricity at 14.39p/kWh (Abu-Bakar et al. 2014). However this study assumes the performance of the system is consistent from year-to-year with only degradation being taken into account; there is no allowance for variability in performance which as shown above increases the financial risk of installing solar thermal systems. This risk is shared by the homeowner and the policy makers: the savings the homeowner makes is dependent on the performance of the system and therefore will remain with them for the life of the system – there will be some years where the cost to maintain the system is greater than the savings made due to variable performance. In addition the current tariff value of 19.2p/kWh is not enough to cover the cost of the installation or its maintenance. The risk to the policy maker

is that the installation will receive higher revenue from the RHI than the likely performance deserves due to over prediction of performance and neglect of variability (Figure 7.15). In addition it will be uncertain as to whether carbon targets are being met in actuality; Figure 7.13b suggests that the probability of failure to meet the carbon emissions predicted by the standard BREDEM model are upwards of 93%.

Three scenarios are explored for system 518 (the worse performing system of the two under consideration): increasing the RHI tariff from 19.2p/kWh to 32p/kWh following the suggestion by (Abu-Bakar et al. 2014); increasing the duration of the RHI payments to cover the assumed life of the system (20years); and offsetting of electricity as the main method of heating as opposed to oil. Figure 7.16 shows the effect of these three scenarios on the NPV of the investment. Increasing the tariff rate does very little to improve the financial viability of the investment with only a 3.8% chance of receiving a positive return with an average NPV is -£295.56 (SD = £163.23). Increasing the duration of the RHI payments to cover the assumed lifetime of the system provides a 100% chance of seeing a positive return on investment. This is because the total value of the RHI payments is highest in this scenario with a total of £7709.16 over the lifetime of the system, £2489.46 more than the sum of the initial and total maintenance costs of the system; this scenario ensures that even if the system produces zero solar yield in a year the revenue from the RHI is enough to cover the loss of savings and annual maintenance. The average NPV in the lifetime RHI scenario is £1813.73 (SD = £163.23). If the primary source of hot water heating is electricity as opposed to oil an 83.8% probability of a positive investment is given. This is due to the higher cost of electricity; the householder could save 62p/kWh electricity by the 20th year. There is a possibility of failure to see a return on investment in this scenario because the savings are linked to the actual performance of the system, which is variable; this is in contrast to the previous scenario where tariff payments are based on a point estimate of annual yield made by BREDEM. The average NPV for this scenario is £448.57 (SD = £438.23).

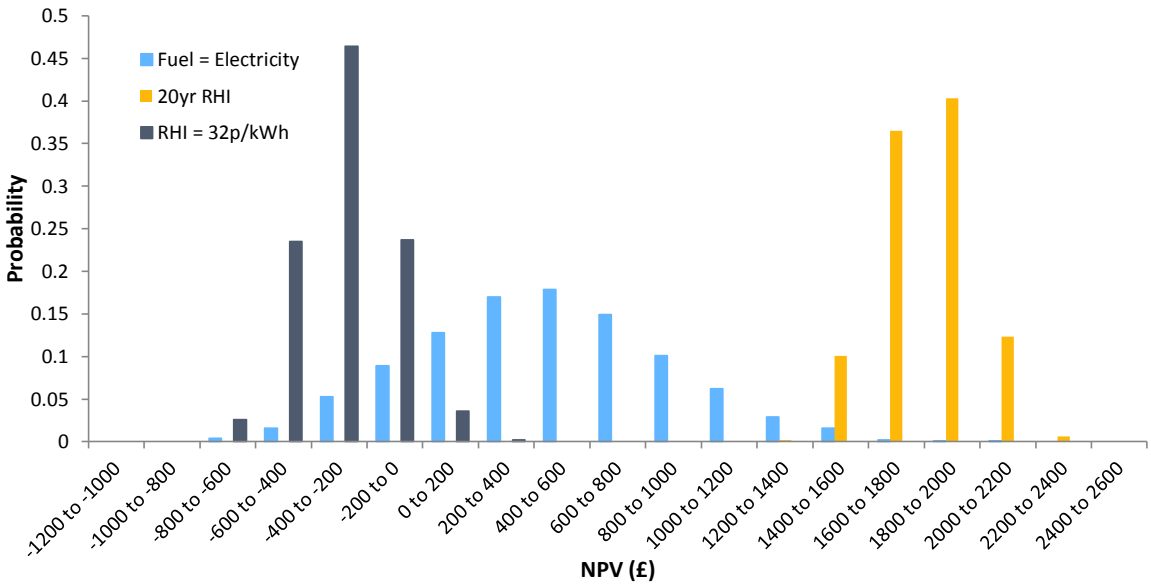

Figure 7.16: Distribution of NPV for system 518 under three scenarios: increase in RHI tariff to 32p/kWh; increase in duration of RHI to 20yrs (assumed life of system); and offsetting of electricity as main fuel

The standard BREDEM model prediction indicates NPVs of £895.28, £3004.56, and £3645.61 for a tariff rate of 32p/kWh, RHI duration of 20years and replacement of electricity as the main heating method respectively.

The results indicate that in order for the RHI to offer adequate financial support for the homeowner with an installed solar thermal system, the scheme should either run for a period of time longer than the current 7 years. Alternatively solar thermal systems should be targeted to those homes currently relying on mains electricity to provide their DHW needs. However it is important to address technical issues that may lead to solar thermal system dysfunction to avoid funding poorly performing systems with publicly funded tariff schemes.

7.11 Conclusions

The variation in annual system performance is a product of the variation in the annual solar radiation, average daily DHW consumption volume and the temperature rise of the hot water. By applying a probabilistic methodology to the prediction of solar thermal system performance this variation can be represented. The above methodology involves representing the BREDEM solar thermal prediction model in a Bayesian network. This method allows variation in the input variables to be incorporated into the calculations and for some regression equations to be replaced with probabilistic relationships where data for the relevant variables is known such as DHW consumption given occupancy and household type; and solar irradiation given location, orientation and roof pitch. The remaining regression equations that make up the BREDEM model are maintained and variables sampled from the input distributions. It is necessary to keep these regression equations because they make use of intermediate variables that are not directly measurable and therefore have no data for probabilistic relationships to be developed. The result is a distribution of the annual solar yield.

The standard BREDEM model was applied to 22 systems from the EST solar thermal field trial. The model was found to over predict the annual yield generated by the majority of systems with an average performance difference between actual and predicted annual yield of -24.05%. The Bayesian BREDEM results eliminate the discrepancy between predicted and measured values and instead offers a probability of obtaining a given yield. The BN model was applied to two systems shown to exhibit performance gaps of more than 40%. The results demonstrated a variation in annual performance as a result of variable operating conditions and enabled the probability of obtaining the measured and standard BREDEM estimated performance. The risk of failing to achieve the BREDEM point estimate could also be quantified and was found to be 68.3% and 92.6% for systems 501 and 518 respectively.

From these results it can be suggested that the carbon performance of these solar thermal systems is highly variable year on year and lower than expected based on a BREDEM estimate. The consequence of this is that true state of carbon emissions associated with domestic buildings making use of solar thermal systems installed under the RHI is likely to be lower than anticipated. This translates to uncertainty associated with actual carbon emissions of the entire UK housing stock that install solar thermal systems to address the issue of CO_2.

Producing a distribution of possible annual yields means that uncertainty can be incorporated into investment decisions; to demonstrate this net present value analysis was performed on the two systems using the standard BREDEM and the Bayesian BREDEM predictions for annual yield. The results show that the RHI does not provide adequate financial support for the two case study systems analysed, nor for the majority of systems. This finding was based on DCF analysis for the BREDEM point estimate of annual yield as well as for the BN distribution of performance. Improvements to the RHI scheme have been stated as lengthened duration of payment or targeting of solar thermal systems to "all-electric" homes.

Even with these improvements to the RHI scheme, basing investment decisions on NPV analyses made using point estimates of performance presents serious issues related to the expected financial return. It has been shown that NPV analysis based on realistic, variable performance would have led to lower than expected financial returns based on BREDEM point estimates and in some scenarios a negative return may have resulted.

7.12 References

Abu-Bakar, S.H. et al., 2014. Financial analysis on the proposed renewable heat incentive for residential houses in the United Kingdom: A case study on the solar thermal system. *Energy Policy*, 65, pp.552–561.

Ayompe, L.M. et al., 2011. Comparative field performance study of flat plate and heat pipe evacuated tube collectors (ETCs) for domestic water heating systems in a temperate climate. *Energy*, 36(5), pp.3370–3378.

Boardman, B. et al., 2005. *40% House*, University of Oxford: Environmental Change Institute.

Booth, a. T. & Choudhary, R., 2013. Decision making under uncertainty in the retrofit analysis of the UK housing stock: Implications for the Green Deal. *Energy and Buildings*, 64, pp.292–308.

Cherrington, R. et al., 2013. The feed-in tariff in the UK: A case study focus on domestic photovoltaic systems. *Renewable Energy*, 50, pp.421–426.

Crowther, M. et al., 2010. *Report to DECC on Heat metering for the RHI*, London.

DECC, 2013. *Renewable Heat Incentive: Impact Assessment*, London.

EST, 2011. *Here comes the sun: a field trial of solar water heating systems*, London.

EST, 2008. *Measurement of Domestic Hot Water Consumption in Dwellings*, London.

EST, 2014. Our calculations. Available at: http://www.energysavingtrust.org.uk/content/our-calculations [Accessed April 14, 2015].

EST, 2001. Solar hot water systems in new housing. *GIR88*.

Firth, S.K., Lomas, K.J. & Wright, A.J., 2010. Targeting household energy-efficiency measures using sensitivity analysis. *Building Research & Information*, 38(1), pp.25–41.

Forward, D. & Roberts, C., 2008. *Viridian Solar - Clearline Solar Thermal Field Trial*, Watford.

Henderson, J., 2002. *Updating the solar water heating procedure in BREDEM*, Building Research Establishment.

Henderson, J. & Hart, J., 2013. BREDEM 2012 – A technical description of the BRE Domestic Energy Model.

Johnston, D., 2003. *A Physically-Based Energy and Carbon Dioxide Emission Model of the UK Housing Stock*. Leeds Metropolitan University.

MCS, 2013. Microgeneration Installation Standard MCS 024: Solar Domestic Hot Water Energy Calculation. , (1.1), pp.1–11.

Mondol, J. et al., 2006. Long term performance analysis of a grid connected photovoltaic system in Northern Ireland. *Energy Conversion and Management*, 47(18-19), pp.2925–2947.

Natarajan, S. & Levermore, G.J., 2007. Predicting future UK housing stock and carbon emissions. *Energy Policy*, 35(11), pp.5719–5727.

Ofgem, 2014. *Do I need metering for the Domestic RHI?*, Available at: https://www.ofgem.gov.uk/publications-and-updates/do-i-need-metering-domestic-renewable-heat-incentive-rhi.

Shorrock, L.D. & Dunster, J.E., 1997. The physically-based model BREHOMES and its use in deriving scenarios for the energy use and carbon dioxide emissions of the UK housing stock. *Energy Policy*, 25(12), pp.1027–1037.

The Energy Monitoring Company Ltd, 2001. *Analysis of performance data from four active solar water heating installations*, Department of Trade and Industry (DTI), London.

Chapter 8

Conclusions

8.1 Introduction

The underlying theme of this research is uncertainty in the performance of domestic building energy subsystems with a particular focus on solar thermal systems. In this context uncertainty is related to whether predicted system performance is achieved in practice, as well as uncertainty related to the actual performance of domestic installations as a result of variable operating conditions.

The research began with an exploration of discrepancies between actual and designed performance of domestic buildings, termed the "performance gap". A brief review of the literature made in Chapter 1 revealed that the performance gap is due to many issues that arise throughout the life-cycle of the building. These issues mean that the design assumptions relating to the installation/construction quality of subsystems, expected performance of subsystems, and operating conditions (including occupancy levels and user interaction with subsystems) made at the modelling stage of the project are rarely met in reality leading to modelled performance being quite different from that as-built (CarbonBuzz 2014; Bordass et al. 2004; Majcen et al. 2013; Kelly et al. 2012; Branco et al. 2004; Hens 2010; Cayre et al. 2011; Carbon Trust 2011; Fokaides et al. 2011; Turner & Frankel 2008).

The consequence of whole-building performance gaps is the risk that predicted energy consumption and related sectoral carbon emission reduction targets will not be met in reality. This leads to serious doubt about the levels of true carbon emissions of the overall UK housing stock subjected to low carbon designs and retrofits (Majcen et al. 2013). Furthermore financial risk as a result of performance uncertainty of renewable energy technologies (RET) is a real issue facing stakeholders being encouraged to install such systems under the Feed in Tariff (FiT) and Renewable Heat Incentive (RHI) schemes (Abu-Bakar et al. 2014). One output of the research showed that under the current RHI tariff mechanism the greatest financial return in the lifetime of a solar thermal system would be as a result of savings on energy bills; this is directly linked to the cost of the fuel being saved and the performance of the system, unlike the tariff payments which are based on a point estimate of annual yield and cease after 7 years. However it has been

shown in this work that the annual yield of a solar thermal system is highly variable and would therefore impact on the energy, carbon and financial savings that can be made.

Quantifying the performance risk of building subsystems would allow design and investment decisions to be made with greater confidence in the presence of uncertainty. A brief overview of the common risk management techniques used in the field of engineering revealed that these methods do not quantify risk in a meaningful way and are largely based on the product of subjective values for the probability and consequence of a risk event happening (Fenton & Neil 2013; INCOSE 2011). In contrast, describing the performance uncertainty using a probability distribution enables the risk of failing to meet a performance benchmark to be readily quantified (Fenton & Neil 2006); this realisation led to the following research aim:

…to evaluate probabilistic approaches for quantifying performance uncertainty associated with renewable energy technologies.

The research objectives were thus:

- *Quantify the contribution made by building subsystems to the "performance gap" of a case study dwelling;*
- *Identify causal factors contributing to discrepancies between predicted and measured performance of a specific RET;*
- *Develop a flexible methodology for evaluating the effect of uncertainty related to system elements on the performance estimates of the case study RET;*
- *Apply this method to improve the predictions of solar thermal yields using a current UK compliance model.*

The result of the research was a novel probabilistic approach to modelling solar thermal system performance. The novel modelling approach took the form of a Bayesian network which responded to the aforementioned performance uncertainty issues by:

- Incorporating performance data of real, in-situ systems thus eliminating the assumptions made at the design stage about installation quality, subsystem performance and operating conditions;
- Identifying causal links between system elements, thus describing the problem/system domain in more thorough detail;
- Propagating uncertainty related to system elements via modelled system performance using Bayesian inference thus producing a distribution of annual solar yield facilitating the quantification of performance risk;
- Evaluating large-scale and small-scale uncertainty in system performance based on evidence.

This concluding chapter will be structured as follows:

- The key empirical findings from each chapter will be briefly restated and discussed and provide a summary of the flow of theories and research that led to the development of the novel Bayesian network model;

- The theoretical implications of the research will be discussed in which contributions made by this research to the current understanding and application of knowledge will be put forth;
- A discussion on the implications of the research on current policies such as the RHI will be made;
- Areas of further work and future possibilities of the BN model will be presented;
- Final concluding remarks will be made.

8.2 Empirical findings

Chapter 3 presented evidence of performance in a group of low-carbon, retrofitted domestic buildings. The buildings were retrofitted as part of the TSB's Retrofit for the Future (R4F) competition with the target to reduce a dwelling's carbon emissions by 80% based on pre-retrofit levels. Many buildings failed to meet the designed carbon emissions with an average percentage difference between measured and designed carbon performance of 60%, with standard deviation of 70%.

From the initial literature review made in Chapter 1 related to performance gaps in buildings it was clear that whole-building performance discrepancies arise from uncertainties related to the building subsystems. A systems engineering perspective was taken that allowed a whole-building to be represented as a complex system of systems (Hensen 2002) reproduced graphically in Figure 8.1.

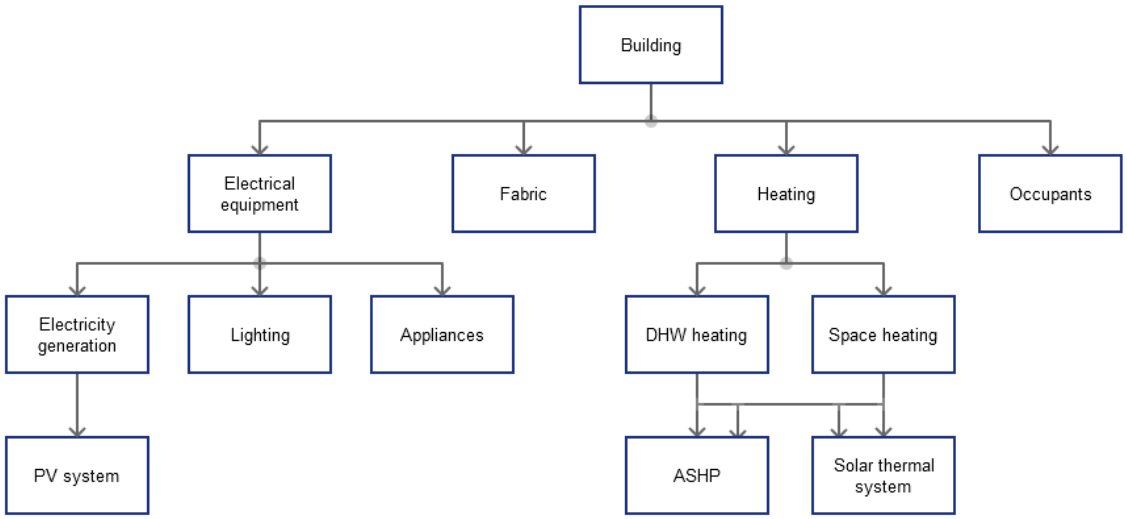

Figure 8.1: Example system hierarchy of a low-carbon retrofitted dwelling

Following the systems engineering approach an evaluation into the performance gap of the R4F dwellings categorised by the type of renewable energy technologies installed suggested that properties with ASHPs tended to have larger discrepancies between actual and designed carbon emissions than those without.

The case study dwelling selected for in-depth analysis revealed that the combined ASHP/solar thermal heating system was the largest contributor to the whole-building performance gap. The ASHP component of the combined heating system consumed almost five times the amount of

electricity as predicted, whilst the solar thermal system generated 39% less thermal energy than predicted. Differences between modelled and actual performance were due to: system dysfunction, thus resulting in worse performance than expected; and differences in the model parameters related to subsystem performance, usage behaviour, weather conditions, and building fabric (influencing the space heating requirements). Potential causes for system underperformance included complex interactions between the CO_2 heat pump and the solar thermal system; oversizing of the heat pump for the small, highly insulated property with low space heating requirements; and low DHW requirements limiting the solar thermal yield.

The overall conclusion of this chapter was that the performance gap is a common issue amongst low-carbon retrofit dwellings and that the cause is related to the actual performance of building subsystems as well as limitations associated with current modelling techniques in accurately representing real operation. Uncertainty in each of the system elements in Figure 8.1 contribute to the overall uncertainty in building performance, which is often not considered in prediction models.

A critique of the common modelling approaches used to predict solar thermal system performance was made in Chapter 4. Modelling approaches were categorised as follows: simplified physics based models; statistical models; and dynamic simulations. The advantages and disadvantages of each category of model were discussed in detail, however common weaknesses of all these modelling techniques include:

- All models produce a point estimate of solar thermal yield, which by its very nature excludes variation in performance and so uncertainty exists as to whether this performance will be met in reality (performance gap) and/or the actual performance of the system (performance variation);
- All require assumptions to be made about the performance of system components such as the collector and storage tank; the weather conditions over the course of a year; and hot water usage patterns (including volumetric consumption, temperature and time of draw). These assumptions are uncertain quantities and may vary from day to day, year to year and system to system thus contributing to the overall uncertainty in the prediction of annual yield.

The common limitations of the above modelling techniques revolve around the lack of incorporation of uncertainty related to the model input parameters and the solar yield prediction. In Chapter 5 two families of probabilistic modelling were discussed: Monte Carlo methods; and Bayesian methods. Monte Carlo methods can be applied to all three categories of predictive model (Mathioulakis et al. 2012; Domínguez-Muñoz et al. 2012) however this can be a computationally intensive endeavour (Lee et al. 2013). Furthermore the limitations of each of the modelling approaches still hold true even when Monte Carlo methods are applied. Bayesian networks were found to offer a potential solution to the limitations of the current modelling techniques whilst being able to introduce uncertainty into predictions of solar thermal yield.

The knowledge-based BN approach was advantageous in that it graphically represents and describes the system domain. Through careful analysis of expert literature on the performance of solar thermal systems a map of the causal factors influencing system performance could be made thus addressing the second research objective related to understanding the problem

domain in greater detail. Through the use of real in-situ performance data of domestic solar thermal systems, the causal map was transformed into a Bayesian network following the approach outlined in Chapter 6 (Nadkarni & Shenoy 2001; Nadkarni & Shenoy 2004). The advantages of the knowledge-based BN were thus:

- Graphical representation of the causal factors influencing solar thermal system performance;
- Use of qualitative and quantitative data sources thus overcoming a limitation of regression-based models including those that form part of a simplified physics based model;
- The use of real data associated with the system parameters and performance means assumptions about unknown parameter values do not have to be made in order for solar yield to be predicted. Any unknown values are represented by a prior distribution of values obtained from real life systems, an improvement over the single value approximations provided in models such as BREDEM;
- The performance of system components is probabilistically linked to the performance of the overall system. This has two advantages: The effect of installation quality on component performance and variations in component performance as a result of manufacture are implicitly included in the probabilistic relationship; secondly the issues related to the performance of system components tested in isolation under standard conditions versus actual in-situ performance (Zero Carbon Hub 2013) are overcome because this parameter is probabilistically (not deterministically) linked to overall system performance;
- Uncertainty is propagated from probability distributions assigned to each model parameter through to the prediction for solar yield. In this way 100% of the variation in the data is captured by the model. This overcomes a limitation of regression based models which may require many variables to fit the data exactly and explain 100% of the variation in the data; these complex models are more susceptible to noise. With respect to dynamic simulations, the uncertainty in a BN is rapidly propagated through the network and does not require multiple simulation runs to be performed to produce a distribution of performance unlike Monte Carlo methods applied to dynamic simulations.
- Uncertainty can be propagated from cause to effect and effect to cause facilitating both prognostic and diagnostic functions useful for design and investment decisions as well as fault finding;
- Performance uncertainty of all measured solar thermal installations in the UK is represented as readily as the performance uncertainty of a homogenous group of systems or a single system.

Despite these advantages over the traditional point-estimate models, there were some limitations with a knowledge-based BN. These related to the requirement of sufficient amounts of good quality data. The data sources used had many missing values affecting the efficacy of the EM algorithm used to develop the probabilistic links between nodes. Furthermore due to the number of parent nodes feeding into the solar yield node the amount of data was potentially not enough to represent all node state combinations adequately; continual measurement of these systems and additional systems would go some way to improve this.

To overcome the issue related to knowledge-based BNs, a second BN was produced based on the BREDEM model. This BN sought to improve the standard BREDEM model by:

- Replacing regression equations with probabilistic relationships;
- Introducing uncertainty to deterministic variables through Monte Carlo methods;
- Making assumptions about system performance parameters redundant by providing an evidence-based prior distribution of zero loss efficiency and heat loss parameters;
- Providing an easy to use model that calculates a probability distribution of annual solar yield.

The BREDEM BN addressed the third research objective by incorporating uncertainty related to hot water usage behaviour, weather conditions and system characteristics, and through Bayesian inference propagated this uncertainty throughout the network to generate a probabilistic prediction of annual solar yield. An evaluation of a subset of solar thermal systems in the EST field trial revealed an average difference between actual and BREDEM predicted performance of -24%, with a standard deviation of 26.48% indicating that most systems performed worse in reality compared to the BREDEM estimation. The BREDEM BN was shown to provide closer approximations of measured performance based on mean and modal predicted annual yield for two case study systems; however the main strength of the BREDEM BN is the prediction of the performance gap by way of a probability distribution that quantifies the performance uncertainty.

The research demonstrated how a model currently used in the construction industry (BREDEM) can be modified to incorporate uncertainty from input to output using a Bayesian network. The newly modified model has done away with regression based equations where possible and sampled prior distributions obtained from real data for inclusion of uncertainty in deterministic calculations. The model maintains the core calculations of the BREDEM model, therefore maintaining its validity as a prediction tool used in the building industry. The simple graphical user interface of the BN allows for rapid predictions to be made that display all the uncertainty related to the system parameters. The following section discusses the theoretical implications of such a model.

Financial analysis of two solar thermal systems revealed that the current RHI tariff mechanism is unlikely to provide a positive return on investment unless a review of the scheme is made (Abu-Bakar et al. 2014).

8.3 Theoretical implications

The performance gap is a common issue amongst low-carbon buildings and represents uncertainty about the actual energy consumption and carbon emissions versus the design targets. The research provided quantifiable evidence of a gap between designed and actual carbon emissions associated with a group of low-carbon retrofitted domestic buildings. In particular the results suggested that the dwellings that relied upon ASHPs for space and water heating had lower reductions in carbon than expected compared to those dwellings using solar thermal and/or PV. This supports the evidence that ASHP systems in the UK do not perform as well as European systems (Boait et al. 2011) and that possible causes for this stem from smaller

UK dwelling sizes and a tendency to reduce space heating requirements through fabric improvements, both of which result in commercially available heat pumps being oversized for UK heating requirements (Boait et al. 2011).

Deeper analysis into a single case study dwelling revealed that the ASHP consumed 4.7 times the amount of electricity as predicted. In addition the solar thermal system generated 39% less heat, whilst the PV system generated 10% more electricity than expected. These findings confirmed that whole-building performance uncertainty is a direct result of uncertainty related to the performance of the building subsystems. Furthermore, the uncertainty related to building subsystem performance was found to be a result of discrepancies between design and actual model input parameters associated with:

- Heating requirements: measured space heating loads were lower than predicted as were DHW requirements. Assumptions about the DHW draw profile and water temperature differed from the assumptions made in the model for the combined heating system;
- Subsystem performance parameters: different values for fabric performance (U-values) were used in the model compared to the actual construction. The efficiency of the solar thermal and ASHP systems differed from the manufacturers' data which provided performance figures based on standard test conditions and isolated testing of the systems;
- System dysfunction: the combined heating system performed worse than expected due to complex interactions between the solar thermal and ASHP components;
- Weather conditions: the amount of solar radiation measured for the year 2012 was 10% higher than that used in the PV model resulting in a 10% difference between measured and predicted yield.

A comparison between estimated and measured annual yield generated by a group of solar thermal systems also indicated performance gaps associated with this technology as a result of uncertainty related to the input parameters in BREDEM.

The above findings support the hypothesis that whole-building performance uncertainty is a result of subsystem performance uncertainty and that this can be traced back to uncertainty related to the system parameters. A system-of-systems perspective enabled the causal factors influencing solar thermal system performance to be mapped and so provided a holistic view of the system domain. This acts to synthesise the research elicited from expert literature into a single network of causal factors.

Bayesian networks were shown to offer a means by which uncertainty in the performance of solar thermal systems can be managed. Bayesian networks can be developed from causal maps and system diagrams facilitating the propagation of uncertainty throughout the system domain from input nodes through to a prediction of annual yield. The object oriented nature of BNs means that the approach can be developed further to include other building subsystems and so describe a low-carbon building, quantifying the annual energy consumption and carbon emissions with a probability distribution. This would describe the uncertainty in whole-building performance as a result of subsystem performance uncertainty and eliminate the performance gap allowing the probability of obtaining a design target for annual energy consumption or carbon emissions to be quantified; in this way performance risk is described in a more

meaningful way than *risk = probability x consequence*. Underperformance of the dwelling can be investigated using the backward propagation of uncertainty through the network to identify the most likely causes of higher than anticipated energy consumption. In addition, BNs may be produced for other renewable energy technology systems such as heat pumps and PV systems which may be combined to assess the socio-economic impacts of renewables in the community (Leicester et al. 2013).

However one limitation of the Bayesian network method is the requirement for high quantities of good quality data to ensure that all node state combinations are represented in the network by the data. This is not a strict requirement since BNs are resilient to missing data records and prior distributions can be elicited from experts when no data exists (Shipworth 2005); however for meaningful information to be displayed about real, in-situ performance it is advisable to have as much data to hand as possible (Onisko et al. 2001). The requirement of performance data is not peculiar to BNs; the feedback of real performance data into building models is an aspiration shared by others (Turner & Frankel 2008; Menezes et al. 2012; Diamond et al. 2006; Lowe & Oreszczyn 2008). However data is often used to calibrate an existing model in order to make the prediction fit the data and therefore obtain closer estimations of a building that has already been modelled (Reddy 2006). Knowledge-based Bayesian networks use the data directly to develop the relationships between system variables and so are not fitted to the data but display uncertainty in the data. A BN model of building performance or subsystem performance can be modified as new information is available and the probabilistic relationships are improved with each additional data record. With continual addition of data from UK buildings and RETs a BN continually reflects the variation and uncertainty in energy consumption across the entire housing stock, or energy generation of a particular RET in the UK. Furthermore evidence inputted into the network updates the probability distributions to provide a description of the uncertainty related to a subset of buildings or systems in the UK. Using real data to calibrate an existing energy model for a particular building may make the prediction more accurate, but cannot achieve the large scale and small scale evaluation of performance uncertainty that a BN can.

8.4 Policy implications

The UK has a commitment to reduce carbon emissions by 80% by 2050 based on 1990 levels. This commitment formed the basis of the R4F competition target in which low-carbon retrofit designs aimed to reduce a dwelling's emissions by 80%. Unfortunately, none of the retrofitted buildings were documented as to having achieved the competition target. This is indicative of the difficulty in achieving such ambitious targets in the presence of technical issues that make the energy and carbon performance of a building uncertain. Lowe & Oreszczyn (2008) suggest a strategic review of the carbon targets, policies and technologies for UK housing over the next 50 years. In light of the high levels of uncertainty surrounding building subsystem performance as a result of technical and human issues prevalent throughout the lifecycle of a building, a rethink of targets, policies and technologies might well be beneficial, especially considering that carbon targets have been shown in the research to be missed on a regular basis.

Technical issues associated with building subsystems are common, and in some cases are a major contributor to the performance uncertainty of a building. These issues need to be

addressed with a serious commitment from the construction and RET industries to ensure high quality constructions and installations. This could possibly be achieved in the form of a performance review of the system after a period of 1 year, including benchmarking the energy consumption/generation against a minimum guaranteed level agreed upon at the design specification stage by the stakeholders. As a necessity, some form of monitoring would be required which should also be made readily accessible by researchers, a view shared by (Lowe & Oreszczyn 2008). Communication and sharing of information between designers, installers/builders, homeowners and researchers would improve the performance of buildings and their subsystems by revealing faults, installation malpractice, solutions to technical issues, information regarding optimum energy usage behaviour, and a database of energy usage and performance data to be used by researchers, policy makers and designers.

Policies such as the Green Deal, FiTs and RHI should adopt a process of "due diligence" in which risk of an investment in a RET or energy efficiency improvements can be quantified to enable the investor to make an informed decision in the presence of uncertainty. The research indicates a great deal of uncertainty in the actual performance of different RETs and retrofitted buildings versus the point estimate. Furthermore there exists variation in system-to-system performance and annual performance of a given system not reflected by point estimates. This performance uncertainty demonstrates investment risk in financial and carbon reduction terms and it should be the duty of the policy provider to explain this to the investor. The Bayesian network approach demonstrates a method by which this can be achieved. In particular the BREDEM-based BN for solar thermal systems is readily adoptable by MCS installers responsible for deeming the amount of energy generated by the proposed installation. This modified model provides a distribution of performance which can be used to provide the probability of obtaining a positive return on investment using financial analysis techniques such as net present value.

Financial analysis of two case study solar thermal systems revealed that the current RHI tariff scheme is unlikely to result in a positive return on investment; these findings are corroborated by (Abu-Bakar et al. 2014). A review of the scheme for solar thermal systems is required to readdress the duration of the payments which was shown to demonstrate a higher probability of positive returns compared to increased tariff rate. The RHI scheme has been quoted as being targeted towards off-gas households (DECC 2013), which would maximise savings on energy bills given that gas is currently the lowest cost fuel source. However, at least with respect to solar thermal systems, the analysis suggests that "all-electric" households would benefit more than those using cheaper off-gas fuel sources such as oil.

8.5 Further work

Future work should focus on making improvements to the knowledge-based BN as well as expanding the currently utilisable BREDEM-based BN:

8.5.1 Knowledge-based BN

- Simplification of the knowledge-based BN to remove specific nodes, based on a sensitivity analysis and available data. Insensitive parent nodes to the daily solar yield node will be eliminated first. Data considerations that influence which nodes to maintain include which system variables are typically measured during performance monitoring of

solar thermal systems. It is likely, therefore, that the stratification node will be removed from the BN. Simplification of the network by removing nodes reduces the dimensionality of the solar yield CPT thus resulting in more data records representing each node state combination.

- Additional data for currently monitored systems would be beneficial in supplying year-to-year variation in performance thus improving the prediction of annual yield based on the daily yield distribution. Data from additional systems would increase the number of data records for each V:A ratio and collector performance factor node states, thus providing a more representative distribution of daily yield given these parameters.
- Research into a method of calibrating the predicted annual yield based on measured annual yield distribution in a more robust way. Currently the distribution of daily yield is used to probabilistically inform the specific annual yield distribution from measured data; however applying a Monte Carlo based calculation of specific annual by multiplying daily yield by the number of days in a year and then calibrating the result using the actual distribution of specific annual yield may improve the prediction of annual yield.
- Including data for pump failure could be included in order to attempt to explain low solar yield on days with optimum ambient conditions.

8.5.2 BREDEM-based BN

In terms of future work on the Bayesian BREDEM model, expansion of the network to include other modules in the standard BREDEM model and, therefore, allow for whole-building performance uncertainty to be quantified. This would require further data to be collected to obtain performance uncertainty for additional building subsystem performance as well as uncertainty related to additional model input parameters.

Large and small scale financial analysis would be beneficial for stakeholders of individual systems such as homeowners as well as local authorities that may have many systems installed across a stock of social housing. To facilitate this kind of financial analysis additional nodes that allow the calculation of NPV can be added to the existing Bayesian BREDEM model; alternatively a new Bayesian network that can receive inputs from the BREDEM BN for solar thermal modelling can be developed. The financial analysis network would be able to produce a distribution of NPV and RHI returns for a group of solar thermal systems owned by a local authority, or for a single installation owned by an individual homeowner. Furthermore, uncertainty related to NPV variables including the capital cost of the system, discount rate, price of fossil fuels and retail price index would also be incorporated into the financial analysis. Inclusion of performance uncertainty estimates of different renewable energy technology strategies (such as PV and heat pumps) in the financial analysis would allow stakeholders in RETs to make informed decisions about which strategy to employ based on a probabilistic prediction of the financial risk.

8.5.3 Collaborative work

Development of additional BN objects for heat pumps and PV systems initially for collaboration with researchers working on BNs that explore the socio-economic impact of small scale renewables in the community (Leicester et al. 2013); this could be used as a decision support tool for selecting the most appropriate RET for achieving socio-economic goals such as alleviating fuel poverty.

8.6 References

Abu-Bakar, S.H. et al., 2014. Financial analysis on the proposed renewable heat incentive for residential houses in the United Kingdom: A case study on the solar thermal system. *Energy Policy*, 65, pp.552–561.

Boait, P.J., Fan, D. & Stafford, a., 2011. Performance and control of domestic ground-source heat pumps in retrofit installations. *Energy and Buildings*, 43(8), pp.1968–1976.

Bordass, B., Cohen, R. & Field, J., 2004. Energy Performance of Non-Domestic Buildings: Closing the Credibility Gap. In *Building Performance Congress*.

Branco, G. et al., 2004. Predicted versus observed heat consumption of a low energy multifamily complex in Switzerland based on long-term experimental data. *Energy and Buildings*, 36(6), pp.543–555.

Carbon Trust, 2011. Closing the Gap: Lessons learned on realising the potential of low carbon building design.

CarbonBuzz, 2014. CarbonBuzz an RIBA CIBSE platform. Available at: http://www.carbonbuzz.org/index.jsp [Accessed February 9, 2015].

Cayre, E. et al., 2011. There are people in the house! how the results of purely technical analysis of residential energy consumption are misleading for energy policies. In *European Council for an Energy Efficient Economy (ECEEE) Summer School*. pp. 1675–1683.

DECC, 2013. *Domestic Renewable Heat Incentive: the first step to transforming the way we heat our homes*, London: Department of Energy and Climate Change.

Diamond, R. et al., 2006. Evaluating the energy\nperformance of the first generation of LEED-certified commercial buildings.

Domínguez-Muñoz, F. et al., 2012. Design of solar thermal systems under uncertainty. *Energy and Buildings*, 47, pp.474–484.

Fenton, N. & Neil, M., 2006. Measuring your Risks. *www.agenarisk.com*, pp.1–6.

Fenton, N. & Neil, M., 2013. *Risk Assessment and Decision Analysis with Bayesian Networks* First., CRC Press.

Fokaides, P. a. et al., 2011. Comparison between measured and calculated energy performance for dwellings in a summer dominant environment. *Energy and Buildings*, 43(11), pp.3099–3105.

Hens, H., 2010. Energy efficient retrofit of an end of the row house: Confronting predictions with long-term measurements. *Energy and Buildings*, 42(10), pp.1939–1947.

Hensen, J., 2002. Hensen, J.L.M. (2002). Simulation for performance based building and systems design: some issues and solution directions. Proceedings of the 6th International

Conference on Design and Decisions Support Systems in Architecture and Urban Planning, 7-10 July. , pp.7–10.

INCOSE, 2011. *Systems Engineering Handbook*, San Diego.

Kelly, S., Crawford-Brown, D. & Pollitt, M.G., 2012. Building performance evaluation and certification in the UK: Is SAP fit for purpose? *Renewable and Sustainable Energy Reviews*, 16(9), pp.6861–6878.

Lee, P. et al., 2013. Probabilistic risk assessment of the energy saving shortfall in energy performance contracting projects–A case study. *Energy and Buildings*, 66, pp.353–363.

Leicester, P.A., Goodier, C.I. & Rowley, P., 2013. Using a bayesian network to evaluate the social, economic and environmental impacts of community deployed renewable energy. In *Proceedings of CISBAT, Clean Technology for Smart Cities and Buildings*.

Lowe, R. & Oreszczyn, T., 2008. Regulatory standards and barriers to improved performance for housing. *Energy Policy*, 36(12), pp.4475–4481.

Majcen, D., Itard, L.C.M. & Visscher, H., 2013. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications. *Energy Policy*, 54, pp.125–136.

Mathioulakis, E., Panaras, G. & Belessiotis, V., 2012. Uncertainty in estimating the performance of solar thermal systems. *Solar Energy*, 86(11), pp.3450–3459.

Menezes, A.C. et al., 2012. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97, pp.355–364.

Nadkarni, S. & Shenoy, P.P., 2001. A Bayesian network approach to making inferences in causal maps. *European Journal of Operational Research*, 128(3), pp.479–498.

Nadkarni, S. & Shenoy, P.P., 2004. A causal mapping approach to constructing Bayesian networks. *Decision Support Systems*, 38(2), pp.259–281.

Onisko, A., Druzdzel, M.J. & Wasyluk, H., 2001. Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. *International Journal of Approximate Reasoning*, 27, pp.165–182.

Reddy, T.A., 2006. Liteterature Review on Calibration of Building Energy Simulation Programs: Uses, Problems, Procedures, Uncertainty, and Tools. *ASHRAE Transactions*, 112(1), pp.226–240.

Shipworth, D., 2005. Synergies and conflicts on the landscape of domestic energy consumption: beyond metaphor. In *ECEEE 2005 Summer Study*. pp. 1381–1391.

Turner, C. & Frankel, M., 2008. *Energy Performance of LEED for New Construction Buildings*, Washington.

Zero Carbon Hub, 2013. *Closing the Gap Between Design & As-Built Performance*, London.

Appendix A

Coded Causal Map

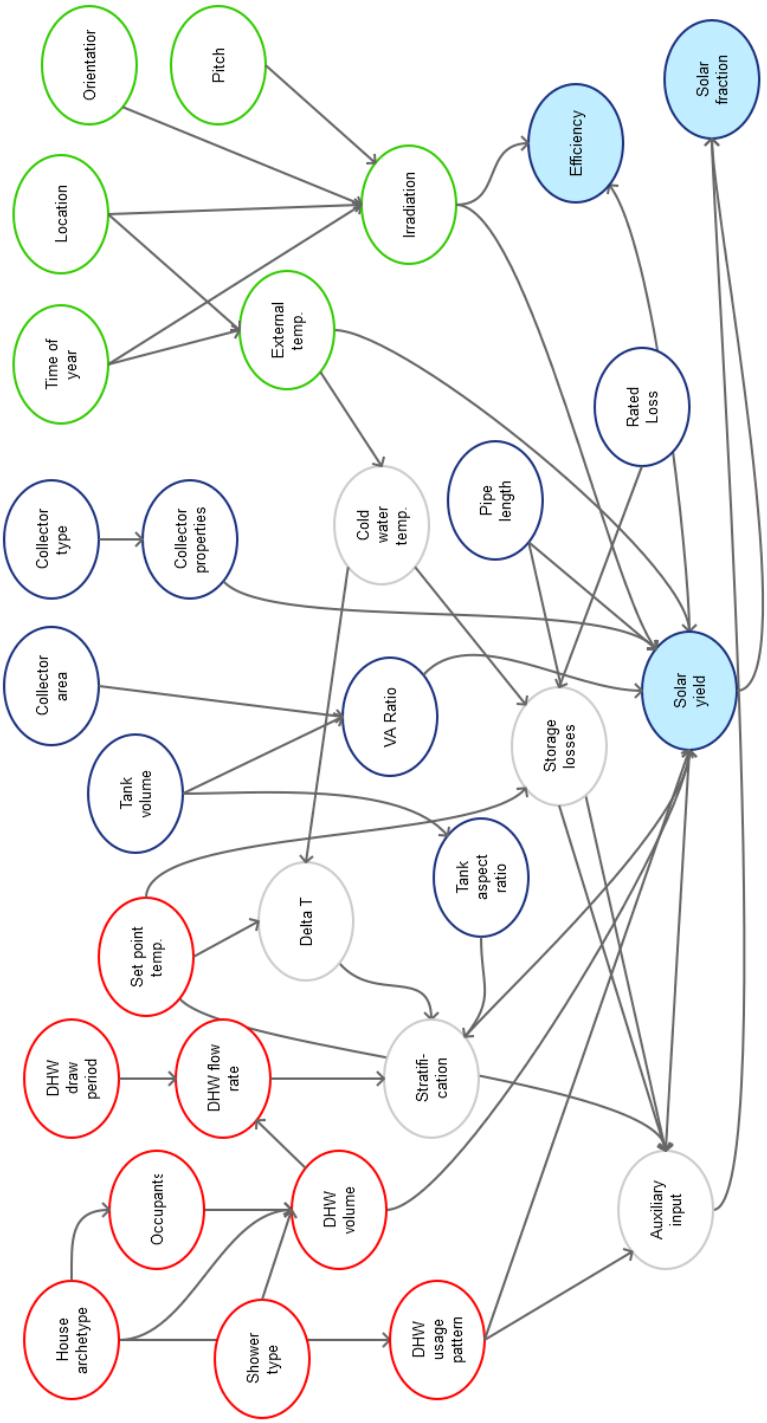

Figure A.1: Coded causal map for the knowledge-based BN

Appendix B

Knowledge-Based BN Node Descriptions

B.1 User Behaviour Object

1. *Household type*: Uses qualitative states to describe seven different types of household: WF = working family; WC = working couple; WS = working single; S = unspecified single (unknown whether working, retired or unemployed); C = unspecified couple (unknown whether working, retired or unemployed); RC = retired couple; RS = retired single. This node is parent to *Occupants* and *DHW Volume*. Data was taken from several sources to produce the conditional probability table for (Occupants | Household type) using a separate case file containing just these two variables; the CPT that links Household type to DHW volume was established from the object case file.
2. *Occupants*: Is a discrete node with seven states (1-7) representing the number of occupants in a household. The parent node is *Household type* and the case file for these nodes was developed separately using several data sets. It is a parent to *DHW volume* the conditional probabilities for which are contained within the object case file.
3. *DHW volume*: Is a continuous node with two parents, *Household type* and *Occupants*. The conditional probabilities between these three nodes are contained within the object case file. The DHW volume represents the daily volumetric consumption of hot water and varies according to the probability distribution obtained from data in the EST solar thermal field trial. The CPT for this node updates the distribution depending on the number of occupants in the household as well as the type of household; for example 4 occupants have a higher average daily DHW consumption than 2 and a retired couple uses a different amount to a working couple. It is a parent to *Daily* Q_{sol}*, DHW flow rate,* and *DHW duration*.
4. *DHW duration*: This is the length of time a draw of hot water takes to complete. It is probabilistically inferred from the amount of hot water consumed in a day – larger daily consumptions increase the chances of having increased draw times and so change the distribution of *DHW duration* to reflect this.
5. *DHW flow rate*: This is probabilistically determined from the DHW duration and DHW volume. A probabilistic relationship was used to avoid unrealistic flow rates being obtained through deterministic calculation of flow rate after sampling the distributions

for *DHW duration* and *DHW volume.* In this way the distribution of flow rate is representative of the measured data. Low volumes drawn over a long duration produce a higher chance of a lower flow rate; conversely high volumes with small durations produce increased chances of high flow rates. *DHW flow rate* is an input node for the system configuration object and is used to calculate the Reynolds number.

6. *Set-point temperature*: This represents the average daily temperature of the water in the top of the storage tank. It was not measured by the EST but could be inferred from the temperature of the water being drawn from the tank (which was measured). It has no parents and is an input node for the system configuration object where it is used to determine the temperature rise of the water (*Delta T*) and the parameter v_{hot}.
7. *DHW usage*: The DHW usage refers to the time of day at which the hot water is used by the occupant. Data pertaining to this parameter could not be obtained from the measured data sets and so the effect on the solar thermal system performance was determined using simulations. The data from these simulations was incorporated by scaling the probabilities in the CPT for *Daily* Q_{sol} – a description of how this was done is presented later in the chapter.

B.2 Solar resource object

1. *Time of year*: This node has 12 discrete and qualitative states representing each month in the year. It is a parent node of *Daily irradiation* and *External ambient temperature*.
2. *External ambient temperature*: Provides probability distributions of the external ambient daytime temperature conditioned on the time of year and location of the system; the parent of *Cold water temperature*.
3. *Cold water temperature*: The temperature of the cold water entering the storage tank is influenced indirectly by the time of year and geographical location through the node *External ambient temperature* which affects the temperature of the mains water. This output node is an input for the system configuration object where it is used to determine *Delta T* and v_{cold}.
4. *Location*: This is a discrete, qualitative node and describes the geographical location of the system. In the current BN there are only two states, North and South, as a result of a limited number of systems for which location data is known and can be used to condition the CPTs of the child nodes. The advantage of describing the location states qualitatively rather than using latitudes is that sometimes the exact latitude is unknown especially in design scenarios. It is easier for the modeller to enter a rough approximation of location rather than be forced into assuming latitude. A similar approach is taken in the most recent SAP in which the UK is divided into 12 regions and annual irradiation determined based on this (BRE 2013). The location influences the external ambient temperature and the daily irradiation.
5. *Pitch*: This represents the angle of inclination at which the collector is installed; in many cases this will be equal to the pitch of the roof. It is a continuous node and influences the daily irradiation incident on the collector.
6. *Orientation*: Represents the direction which the collector faces. It is a qualitative node with discrete values which are easily understood by the modeller compared to the

different numerical notations for orientations. This is a parent node of the daily irradiation.

7. *Daily irradiation*: This is the amount of solar radiation incident on each square metre of collector area measured in kWh/m^2/day. The distribution of this parameter is conditional on the time of year, location, pitch, and orientation of the collector. These parameters have been incorporated into deterministic relationships in BREDEM and dynamic modelling packages but uncertainty is not considered. The Bayesian approach uses measured data from the EST solar thermal field trial to produce the solar resource object case file which is exclusively used to produce the CPTs contained within this object. As more data becomes available from future studies the CPTs of the object or a subset of nodes in the object can be updated. Daily irradiation is the second output node of the solar resource object. It is a parent to *Daily* Q_{sol}.

B.3 System configuration object

1. *Tank diameter*: The data for this node was not contained within the EST solar thermal field trial data set; however the tank volume was known. In order to build up a data set for this node a range of storage tank manufacturers' technical specifications were reviewed and the tank dimensions collated. *Tank diameter* is a parent node of *Aspect ratio* and a separate case file was produced that linked *Tank diameter*, *Tank volume*, and *Aspect ratio*.
2. *Tank volume*: Data for the tank volume was available for the systems in the EST solar thermal field trial data set and so could be included in the system configuration object case file. The tank volume influences the aspect ratio of the tank by influencing the tank height for a given diameter. It is a parent of *VA ratio* and a child of *Collector area*.
3. *Aspect ratio*: The aspect ratio (AR) of the tank is the ratio of the height to the diameter and so is a unit-less quantity. Due to different models and manufacturers of storage tanks it is possible for a tank of a given volume to have a distribution of possible aspect ratios due to variation in tank height and diameter. From the perspective of the designer/modeller of a solar thermal system it is perhaps more common to know the available foot print for the tank and so the diameter can be selected based on this; if the volume is as yet to be determined then the tank height, and therefore aspect ratio, is also unknown. Consequently the aspect ratio is conditional on the tank diameter and the tank volume which are quantities more likely to be known to the modeller. A range of aspect ratios was calculated based on the data collected from the manufacturers' data sheets for different tanks; this allowed the P(*Aspect ratio* | *Tank diameter, Tank volume*) to be determined from the case file. The advantage of this over calculating deterministically is that tank height does not have to be specified and extreme values caused by sampling low tank heights and high tank diameters (and vice versa) are eliminated. Because the actual aspect ratio for each of the systems in the overall object case file was unknown it was calculated by regression using tank volume as the regressor. This is explained in more detail further on in the chapter. This was necessary in order to calculate the stratification factor - *Stratification* is a child node of *Aspect ratio*.
4. *Collector area*: The collector area data was taken from several data sources and incorporated into a separate case file. The collector area was known for the majority of

the systems used to populate the overall object case file. Within the node itself there is no distinction between gross area, aperture area or absorber area - the node states are bin widths and the value of any of these areas for a given system will be within the same bin; however if the modeller happens to know exactly which area should be used and the value of that area this can be input as a numeric value. The advantage of bin widths in terms of specifying the collector area is that either gross, aperture or absorber area can be used since they will typically fall within the same bin width; furthermore variation in the manufacturing process will be incorporated within the width of the bin. The collector area is a parent node of *Tank volume* because it is assumed that the designer will size the tank in accordance with the collector area to provide an optimal VA ratio (German Solar Energy Society 2010; Viridian Solar n.d.). *Collector area* is used to scale up the specific annual yield (kWh/m^2/yr) prediction to an overall annual yield (kWh/yr); it is also used in the calculation of solar fraction which requires the total energy delivered by the solar thermal system in kWh/day.

5. *VA ratio*: This is the ratio of the tank volume to the collector area; this node is a deterministic node. The distribution for *Collector area* and *Tank volume* are sampled 1000 times to give a distribution of 1000 VA ratio calculations. Since the tank volume is influenced by the collector area, extreme values of VA ratio due to sampling are avoided; however it also shows how some of the systems in the EST field trial had less than optimal VA ratios, this is discussed further on in the chapter. It is a parent to *Daily* Q_{sol}.
6. *Collector type*: A qualitative node with two states representing the two types of collector on the market: flat plates and evacuated tubes. This node influences the values of the collector parameters: zero loss efficiency; and first and second order heat loss parameters. Therefore if the type of collector is known but the performance parameters are not the conditional probabilities obtained from the object case file provide a distribution of the values based on real installations. This is an advantage over assumptions made by designers using dynamic based models or suggested in BREDEM based models; furthermore variation in performance due to manufacturing and testing differences are incorporated within the bin widths of these parameters.
7. *Zero loss efficiency*: The majority of the systems in the EST field trial had collector make and models provided; these were used to find the relevant data sheet and obtain data for zero loss efficiency, and first order heat loss and second order heat loss parameters. CPT for this node was obtained from the overall object case file. It is used to calculate the collector performance factor and is influence by the type of collector installed.
8. *First order heat loss*: The majority of the systems in the EST field trial had collector make and models provided; these were used to find the relevant data sheet and obtain data for zero loss efficiency, and first order heat loss and second order heat loss parameters. CPT for this node was obtained from the overall object case file. It is used to calculate the collector performance factor and is influence by the type of collector installed.
9. *Second order heat loss*: The majority of the systems in the EST field trial had collector make and models provided; these were used to find the relevant data sheet and obtain data for zero loss efficiency, and first order heat loss and second order heat loss parameters. CPT for this node was obtained from the overall object case file. It is used to calculate the collector performance factor and is influence by the type of collector installed.

10. *Collector performance factor*: This is a deterministic node given by $\frac{0.892(k_1+(0.45k_2))}{\eta_0}$ where k_1, k_2, and η_0 are the first order heat loss, second order heat loss and zero loss efficiency respectively. The formula is taken from BREDEM (Henderson & Hart 2013) and is used to reduce the number of nodes feeding into *Daily* Q_{sol} in an attempt to reduce the number of conditional probabilities in that node table and the processing time of the EM algorithm. The collector performance factor is calculated for each of the systems in the performance case file to give P(*Daily* Q_{sol} | *Collector performance factor*).
11. v_{hot}: Deterministic, intermediate node calculated using set point temperature from the user behaviour object. It represents the kinematic viscosity of the hot water in the storage tank – higher temperatures give lower viscosities. It is a parent node to *Grashof* and is used to calculate the stratification factor of the tank. How the kinematic viscosity was determined from the water temperature is discussed later in the chapter.
12. *Beta*: Deterministic, intermediate node calculated using set point temperature from the user behaviour object. It represents the volumetric thermal expansion coefficient of the hot water in the storage tank – higher temperatures give greater expansion. It is a parent node to *Grashof* and is used to calculate the stratification factor of the tank. How the thermal expansion coefficient was determined from the water temperature is discussed later in the chapter.
13. *Delta T*: This is the temperature rise of the water in the storage tank and is the difference between the set point temperature and the cold water temperature. Although it is deterministically calculated, it is a probabilistic node in this network with parents from the user behaviour and solar resource objects; it was calculated within the object case file to give a conditional probability distribution of *Delta T* given *Set point temperature* and *Cold water temperature*. In this was extreme values are avoided caused by sampling the probability distributions of *Set point temperature* and *Cold water temperature* - sampling could potentially produce negative delta T values if the cold water temperature sampled is higher than the set point temperature; this is not possible since the lowest delta T has to be zero in which case the tank is uniformly mixed this can occur on periods where no heat is delivered to the tank which cools to the temperature of the cold water. *Delta T* is a parent node of *Grashof*.
14. v_{cold}: Deterministic, intermediate node calculated using cold water temperature from the solar resource object. It represents the kinematic viscosity of the cold water flowing into the storage tank – colder temperatures give higher viscosities. It is a parent node to *Reynolds* and is used to calculate the stratification factor of the tank. How the kinematic viscosity was determined from the water temperature is discussed later in the chapter.
15. *Grashof:* This node represents the Grashof number which is the ratio of buoyancy forces to viscous forces acting on a fluid (Incropera et al. 2007). Within the overall object case file the tank diameter was assumed to be 570mm for all systems because this was the average diameter for all tank volumes in the collated dataset for tank dimensions.
16. *Reynolds*: This node represents the Reynolds number which is the ratio of inertial forces to viscous forces acting on a fluid (Incropera et al. 2007). It is a function of the viscosity of the fluid (in this case the cold water) given by v_{cold} and the flow rate given by the output node from the user behaviour object, *DHW flow rate*. More viscous fluids with lower flow rates produce lower Reynolds numbers indicating a less turbulent flow. The

Reynolds number (Re) is given by $Re = \frac{\rho V l}{\mu}$ where ρ is the fluid density, V is the velocity, l s the characteristic length, and μ is the dynamic viscosity of the fluid. Following (Lavan & Thompson 1976) this can be rewritten to include parameters that are more readily available to the modeller of a solar thermal system: $Re = \frac{4Q}{\pi \nu d}$ where Q is the volumetric flow rate (not to be confused with Q_{sol} which denotes energy; note that: $Q = V\frac{\pi d^2}{4}$), ν is the kinematic viscosity ($\nu = \frac{\mu}{\rho}$) in this case the kinematic viscosity of the cold water, and d is the diameter of the inlet pipe taken to be 22mm.

17. *Stratification*: Represents the stratification factor of the storage tank. In the network it is a deterministic node with parents *Grashof*, *Reynolds* and *Aspect ratio.* This node gives a prior for the stratification factor. The stratification factor is calculated in the performance case file to give the conditional probability P(*Daily* Q_{sol} | *Stratification*) - it is a parent of *Daily* Q_{sol}. Lavan and Thompson (1979) provide a method of calculating the stratification of the tank given Gr, Re and AR: $1 - (e^{-0.067Re^{-0.55}Gr^{0.35}AR^{0.58}})$ (Lavan & Thompson 1976).

B.3.1 Calculating the aspect ratio

The aspect ratio is the ratio of tank height to tank diameter however for the systems in the EST solar thermal field trial these dimensions are unknown. Geometric data for a range of tank manufacturers and models was used to provide a prior distribution of tank diameter; the tank diameter is most likely to be known to the modeller along with the volume and so the conditional probability of AR given tank diameter and tank volume was incorporated in the BN. The second function of the manufacturers' data was to aid in the calculation of AR based on the tank volume for the systems used in the BN case files. This was necessary for the stratification factor of each of the systems to be calculated using the equation in Lavan and Thompson (1979).

The maximum tank volume of the systems used in the BN was 330L and so the data used in the regression model to find AR were limited to cases where the tank volume was 330L; this makes the regression model more accurate over the range required. Figure B.1 shows the aspect ratios from the manufacturers' data versus tank volume along with the regression line given by:

$$AR = 0.0075V_{tank} + 1.2673 \quad \text{(B.1)}$$

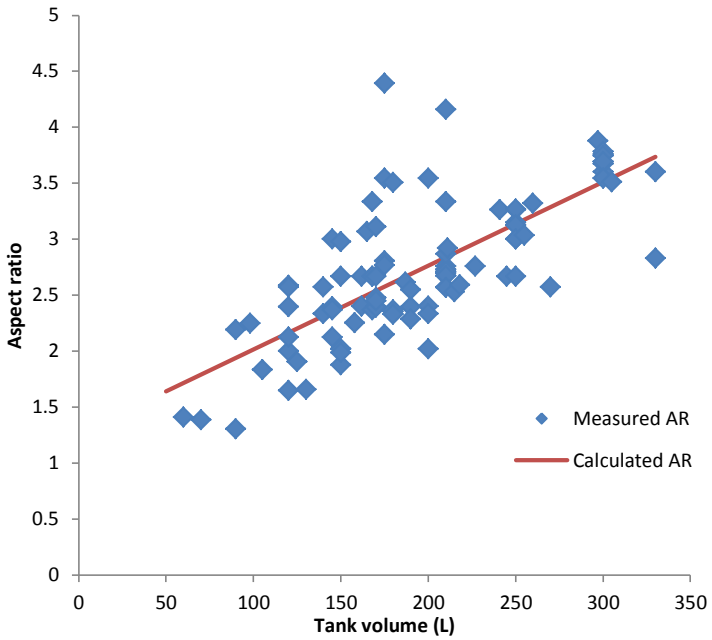

Figure B.1: Correlation between aspect ratio and tank volume

B.3.2 Calculating the kinematic viscosity and thermal expansion coefficient

The kinematic viscosity of water is a function of the temperature since both dynamic viscosity and density are also a function of temperature. Therefore to calculate the kinematic viscosity given the water temperature a regression equation was developed. The thermal expansion of water also changes with temperature; again a deterministic relationship was used to calculate this parameter given the water temperature.

Temperature (°C)	Viscosity (x10-6) (m2/s)	Thermal expansion (1/°C)
4		0
10	1.308	0.000088
20	1.002	0.000207
30	0.7978	0.000303
40	0.6531	0.000385
50	0.5471	0.000457
60	0.4658	0.000522
70	0.4044	0.000582
80	0.355	0.00064
90	0.315	0.000695
100	0.2822	

Table B.1: Kinematic viscosity and thermal expansion of water at different temperatures (The Engineering ToolBox n.d.)

This data was used to produce the following regression equations where θ denotes the temperature of the water:

$$\nu = (-4 \times 10^{-8}\theta) + (4.41 \times 10^{-10}\theta^2) + (-1.8 \times 10^{-12}\theta^3) + 1.65 \times 10^{-6} \quad \text{(B.2)}$$

$$\beta = (1.55 \times 10^{-5}\theta) + (-1.3 \times 10^{-7}\theta^2) + (6.04 \times 10^{-10}\theta^3) - 5.75 \times 10^{-5} \quad \text{(B.3)}$$

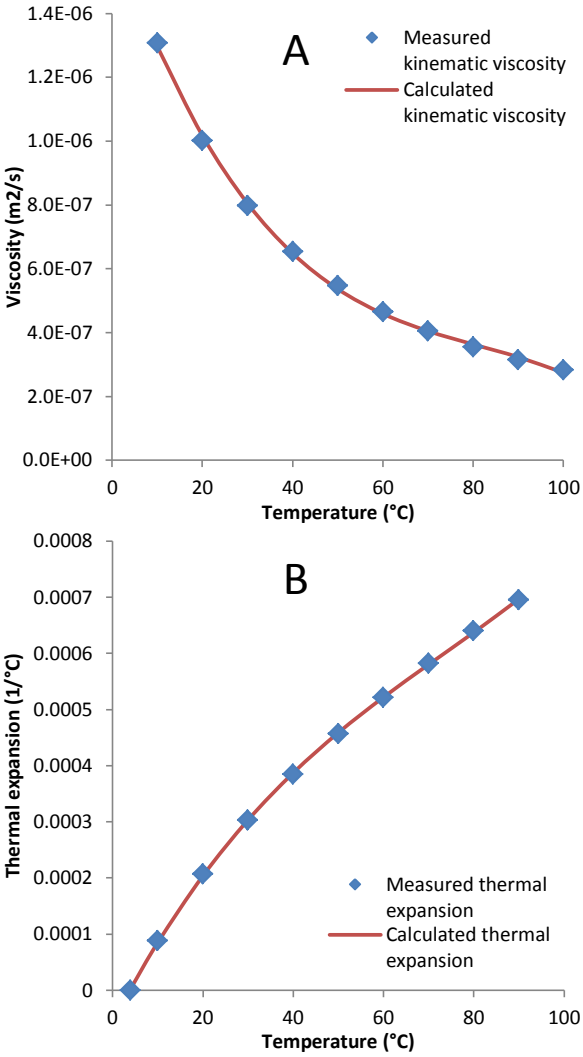

Figure B.2: Measured and calculated values for a) kinematic viscosity; b) thermal expansion of water

Figure B.2 shows the closeness of the calculated parameter with the measured value. The maximum relative error for kinematic viscosity is 2.8% and occurs when θ is 100°C; for thermal expansion the maximum error is 3.2% and occurs when θ is 10°C.

B.4 Performance nodes

1. *Daily* Q_{sol}: This is the daily solar yield, which is the thermal energy measured in kWh/m^2/day delivered to the storage tank; it is a specific value meaning it is the energy delivered per unit area of collector. The parents to this node are the input and output nodes from the Bayesian objects including: *Daily DHW volume*; *DHW usage*; *Daily irradiation*; *VA ratio*; *Collector performance factor*; and *Stratification*. It is used to infer *Daily* Q_{aux}; *Daily solar fraction*; *Daily system efficiency*; and *Specific annual yield*. The conditional probabilities were contained within the performance case file as described above.
2. *Daily* Q_{aux}: This represents the daily auxiliary input required to meet the thermal demands of the occupant. This data was measured by the EST in the solar thermal field

trial and is used in the performance case file to elicit the conditional probability of *Daily* Q_{aux} | *Daily* Q_{sol}. It is used to calculate the daily solar fraction.

3. *Daily solar fraction*: This is a deterministic node and provides a distribution of the fraction of thermal energy delivered to the storage tank by the solar thermal system in a day. Calculated using: $SF = 100 \times \left[\frac{(Daily\ Q_{sol} \times Collector\ area)}{(Daily\ Q_{sol} \times Collector\ area) + Daily\ Q_{aux}}\right]$. The distributions of *Daily* Q_{sol}, *Daily* Q_{aux} and *Collector area* are sampled 1000 times to produce a distribution for solar fraction.
4. *Daily system efficiency*: This is a deterministic node and provides a distribution of daily system efficiency given by: $\eta_{sys} = \frac{Daily\ Q_{sol}}{Daily\ irradiation}$. Since *Daily* Q_{sol} and *Daily irradiation* are specific values *Collector area* is not a parent node to *Daily system efficiency*. The distributions of *Daily* Q_{sol} and *Daily irradiation* are sampled 1000 times to produce a distribution for system efficiency; if the sampled Q_{sol} is greater than the irradiation then this is given a zero value because it is impossible for this to happen.
5. *Specific annual* Q_{sol}: This node is probabilistic; it infers the annual solar yield given the distribution of the daily yield based on measured data. The reason why it is a probabilistic node is to allow measured annual specific yields to be incorporated into the network. When Monte Carlo analysis is performed sampling from the distribution for daily solar yield the variation in annual solar yield is much less than that measured, shown in Figure B.3; therefore to incorporate the true variation in specific annual yield its probability is conditional on the daily solar thermal yield. In addition the MCA performed makes backward inference in the BN much more difficult as it is necessary in this case to obtain the distribution of *Daily* Q_{sol} given *Annual* Q_{sol}.
6. *Annual* Q_{sol}: This node is the total solar yield obtained over the year; it is a deterministic node calculated by multiplying the collector area by the specific annual yield. This output is useful for making investment decisions where the annual performance is required to determine the annual return.

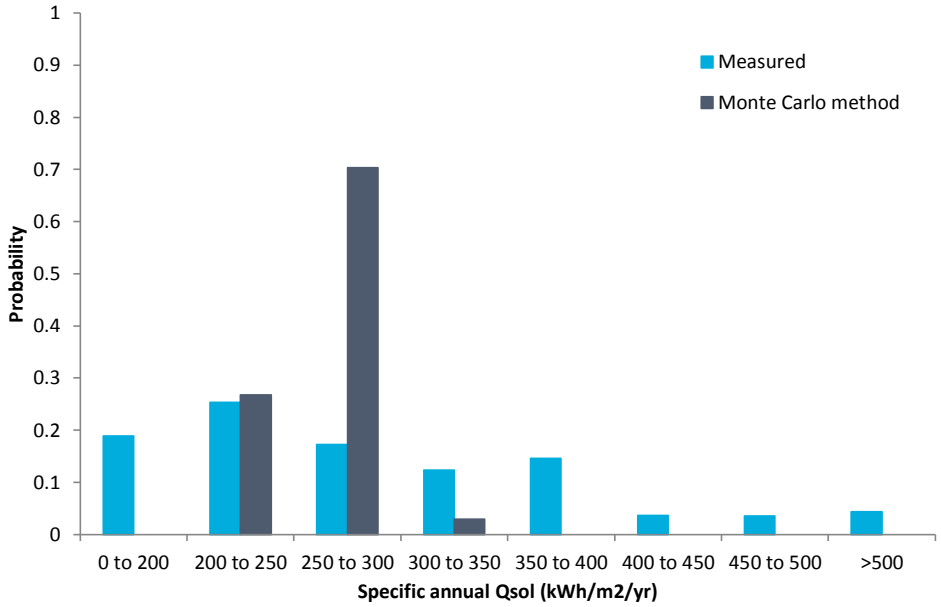

Figure B.3: Monte Carlo sampling applied to *Daily* Q_{sol} distribution obtained from BN versus measured distribution of specific annual yield

B.5 References

BRE, 2013. SAP 2012 The Government's Standard Assessment Procedure for Energy Rating of Dwellings.

German Solar Energy Society, 2010. *Planning and Installing Solar Thermal Systems* Second., London: Earthscan.

Henderson, J. & Hart, J., 2013. BREDEM 2012 – A technical description of the BRE Domestic Energy Model.

Incropera, F.P. et al., 2007. *Fundamentals of Heat and Mass Transfer* Sixth Edit., John Wiley & Sons Inc.

Lavan, Z. & Thompson, J., 1976. Experimental study of thermally stratified hot water storage tanks. *Solar Energy*, 19, pp.519–524.

The Engineering ToolBox, Volumetric - Cubic - Thermal Expansion. Available at: http://www.engineeringtoolbox.com/volumetric-temperature-expansion-d_315.html.

Viridian Solar, Solar Heating Selector. , pp.1–2. Available at: http://www.viridiansolar.co.uk/Assets/Files/choosing_the_right_solar_panel.pdf [Accessed May 21, 2012].

Appendix C

Publications

The work described in this thesis has contributed to the following publications:

Rowley, P., Gough R., Doylend, N., Thirkill, A. & Leicester, P., 2013. From smart homes to smart communities: Advanced data acquisition and analysis for improved sustainability and decision making. In *International Conference on Information Society (i-Society)*. pp. 263–268.

Thirkill, A. & Rowley, P., 2013. Analysis of the performance gap and performance variation for solar thermal systems. In *Proceedings of CISBAT, Clean Technology for Smart Cities and Buildings*. pp. 853–858.

Thirkill, A. & Rowley, P., 2013. Probabilistic Analysis of Renewable Heat Technologies. In *Proceedings of the 20th Advances in Risk and Reliability Technology Symposium (AR2TS)*. pp. 120–131.