Modelling and Analysis of Energy Demand Variation and Uncertainty in Small-Scale Domestic Energy Systems

PhD Thesis

Graeme Flett

Energy Systems Research Unit

Mechanical and Aerospace Engineering Department

University of Strathclyde, Glasgow

May 4, 2017

Abstract

A range of different scenarios have been predicted for future UK energy supply. While there is significant uncertainty, all expect an increase in small-scale distributed generation integrated in constrained or independent networks and with predominantly domestic consumers. This reduction in system scale has not, however, driven a significant change in design practices, with deterministic models and rules-of-thumb prevalent. Little consideration has been given to how the specific household characteristics and the size of system impact on demand level and timing, the degree of uncertainty in any demand prediction, and how design practices should change to reflect this. The main contribution of the presented work has been to address this.

To allow the variation and uncertainty to be quantified; a highly differentiated, probabilistic, bottom-up demand model has been developed for electrical and hot water use. The 1-minute resolution model incorporates an enhanced Markov chain occupancy model and is based on a newly developed discrete-event approach for occupant-initiated demands. Utilising realistic factoring for appliance ownership, income, occupancy, and random energy-use behaviours, the model has been shown to capture the range of potential household demands. Assessment that the developed model, and any existing model calibrated using group data, tended to rapidly converge to the group average basis, prompted further method development to improve the model's performance in capturing individual household demand behaviours.

Analysis of both existing data and the demand model output has shown that energy system demand can vary significantly based on socio-economic characteristics and the types of households supplied. It also highlights that demand uncertainty for individual households can exceed an order of magnitude, even if household characteristics are known. As the system scale is increased, the level of overall demand uncertainty remains significant to at least 200 household systems. A method has therefore been developed that allows multiple runs of the probabilistic model to be reduced to a representative subset, which can be used to analyse potential energy system performance scenarios probabilistically using existing optimisation tools.

Contents

List of Figures

List of Tables

Preface/Acknowledgements

I would particularly like to acknowledge the support of my supervisor, Dr Nick Kelly, for his support and guidance, and extensive editing at times, during the project. I would also like to thank Professor Joe Clarke, Dr Paul Strachan, Luke Tait, Katalin Svehla, John Allison, and Andrew Cowie for additional help, support, contacts, data, and software guidance.

Thanks also go to my family and friends for their patience as I do tend to get a little (over) focused and have maybe used PhD work as a convenient excuse once or twice. I should be marginally less grumpy now.

Finally, I would like to thank the BRE Trust for their financial support.

Chapter 1

Introduction

1.1 Chapter Overview

The following chapter provides introduces the overarching themes, the specific research gaps identified, and the aims of the presented work. The energy system concepts of distributed generation and increasing electrification of demand are introduced, placed in context in terms of current and future energy use, and the associated technical challenges outlined. There is particular focus on the need for improved high resolution domestic demand prediction methods, focusing on household and community type influenced demand variations and the residual prediction uncertainty as a result of different energy use behaviours, to enhance the accuracy and effectiveness of energy system analysis for small-scale systems.

1.2 Background

Energy systems globally have changed significantly over the last thirty years. The risk of irreversible climate change from fossil fuel energy sources driving the development of new methods of generation and, in parallel, significant improvements in the energy efficiency of buildings and appliances, changing occupancy patterns, and evolving types and levels of demand.

This transition has given rise to a several design and operating challenges. Critically, these include increasing energy system complexity with smaller-scale, dispersed generation units, often incorporating renewable sources with variable output, being considered and implemented as an alternative to a small number of large sources [1] and interacting with energy networks in a manner for which they were not originally designed. For example, decentralising electrical power generation changes the dynamics of the national grid, which can have detrimental effects on performance [2]. Additional challenges include, an increase in the risk of overheating from heat gains because of improved building thermal performance and specific user behaviours, and identifying methods to alter demand patterns to better match supply and demand. This requires a detailed understanding of a number of elements, including the time dependency of energy supply, building occupancy, and specific energy demands [3]. All these aspects, and a significant number of others, have reduced the margin for error.

The complexity is likely to increase in most countries in the short-term ([4], [5], [6]). In addition, the predicted increase in electric car use has the potential to significantly alter the electricity demand profile with the associated potential and risk [7].

1.3 UK Energy Generation

A consequence of the current and proposed developments, and associated challenges, is to reduce the scale of interest in energy analysis. The operation of regional and national grids supplied by centralised generation requires the ability to predict demand in the immediate future to allow supply and demand to be balanced. Whilst this requires complex analysis of the various demand drivers, at this scale industrial and commercial consumption is significant and the influence of individual households or different types of communities is not. Increasing use of distributed generation, either at the household level, connected to predominantly domestic sub-sections of the grid, or designed to supply communities independently, has resulted in an increasing focus on domestic consumption, and the potential deviation from average demand behaviour at small scales. As will be discussed, it is not clear that design practices have kept pace with the changing focus.

Lifestyles for a large proportion of people have changed in the same time period with consequences for domestic energy consumption patterns. Changes in occupancy patterns have been primarily driven by changes in employment with more flexible working hours and weeks ([8] and [9]), and more people working from home [10]. The result for energy system planning and design is less predictable occupant behaviour at a time when the influence of individual households needs to be better understood.

Recent years have also seen an increase in the number and type of appliances owned

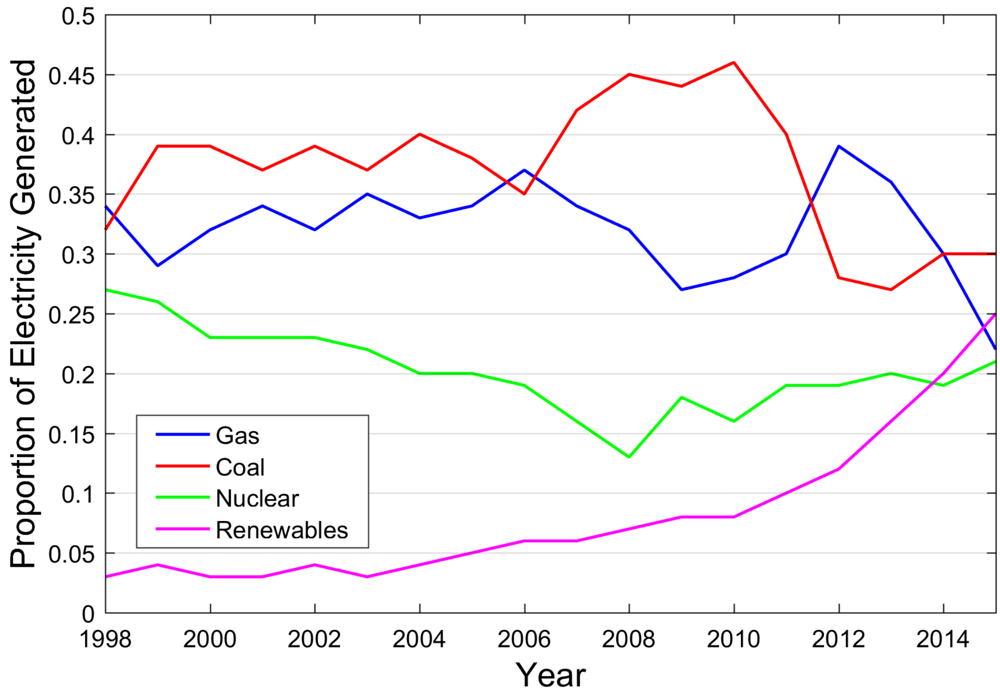


Figure 1.1. UK electricity generation by fuel type. Data from [13].

([11] for UK data) and changes in how they are used, in parallel with the improvements in energy efficiency. Understanding the combined impact of these potentially conflicting influences is critical for a detailed understanding of household energy use [12].

UK electricity generation, as shown by Figure 1.1, has seen a significant reduction in coal use and an increase in renewable sources since 1998 [13]. The primary driver for change has been the need to decarbonise generation.

Heat and hot water use in the UK remains predominantly from gas. The integration of renewable sources has been directly from biofuel sources and indirectly using renewables-generated electricity for heating and hot water generation from heat pumps, but remains low at 5% based on 2013 UK data [14].

The drive for further decarbonisation is generally considered to require multiple actions [15], combining both reductions in demand and carbon emissions per unit generated, including: replacement of fossil-fuel generation with nuclear power; increased appliance energy efficiency; demand management; and, using low-carbon sources.

1.3.1 Distributed Generation

1.3.1.1 Concept

Low carbon source systems are typically characterised by smaller, modular individual generating units and either the ability or technical requirement to be located close to consumers. These systems are known generally as 'distributed generation', and range in scale from individual household supply to generation equivalent to demand for significant areas.

There are many different types and potential combinations of distributed generation. These include: CHP systems that can supply all three primary energy inputs; wind turbines or solar panels for electricity; heat pumps that use both electricity and heat from the environment to provide heating and hot water; solar thermal for heating and hot water; and gas or biomass boilers for heating and hot water supply.

As outlined, an implication of their integration has been a reduction in analysis scale from national grids to smaller sub-systems, which were not originally designed for supply connections, and in extreme cases can be either self-contained or minimally grid-connected 'microgrids' [16]. As an example of the different scales, the UK Power

Network authority identifies four levels of electrical distributed generation [17] in terms of the impact on the national grid. Single household systems (roof-top PV, small wind turbines), multiple connections up to 11kW, larger systems up to 50kW, and over 50kW. For electricity generation, in particular, the degree to which the generation system, or the sub-system to which it is connected, interacts with the overall grid is a significant consideration, with the potential time dependency of both supply and demand influencing the degree, timing, and even direction, of the current flows in each part of the system.

1.3.1.2 Role and Potential

UK National Grid data shows that 'distributed generation' (DG) electricity capacity (i.e. systems less than 100MW) was c.14GW in 2015 [1]. Of this 3.5GW is connected to the low voltage network and a further 4.3GW to 11kV subsystems [18], which represent the proportion most likely to require detailed analysis for grid interaction impact or potential for grid independence.

Future electricity DG potential in the UK has been assessed by the National Grid based on four potential future energy strategies as outlined above [5]. The two most optimistic cases predict the capacity increasing to either 20 or 33GW by 2030 [1]. This suggests both significant potential and uncertainty.

District heating (DH) systems, 'distributed generation' primarily for heating and hot water supply (including CHP which also provides electricity), accounts for c.1% of UK heat demand (of which c.73% is domestic) [19]. There is potential for DH to account for 9% of total demand by 2030 and 18% by 2050 [19], but there are significant barriers to this level of implementation.

A number of consistent benefits have been attributed to DG by various sources ([20], [21], [22], [23]), which include: lower emissions; reduction of transmission losses; shorter build times; security of supply; improved competition; consumer and community involvement; network balancing; and resource efficiency. The short build time potential is increasingly becoming a focus in the UK. Aging and polluting centralised energy generation and slow decision-making on alternatives, including nuclear power options, may require fast, localised solutions to fill the supply gap, in addition to any climate change drivers [15].

Distributed generation does, however, have potential problems. For electricity systems, major technical issues include maintaining stable grid voltages and power quality [17]. All system types have inherent economic issues, including economies of scale compared to large-scale generation, supply agreement complexity and instability [23], and imbalanced stakeholder benefits [24]. Other issues include: intermittency, with several renewable sources not having constant supply characteristics; and reliability, with the need for either a grid connection or redundant generating units for back-up.

The primary barrier to the introduction of DG in the UK is currently political. Recent reduction in subsidies and an environment that favours large, commercial projects [25] has limited recent uptake. Rebalancing of incentives and drivers is required to achieve an integrated decarbonisation strategy [15].

Whilst this research project has focused primarily on the UK, the potential for distributed generation is global [26]. For the developed world, the drivers for grid decarbonisation are universal. DG is also a potential solution for areas without existing centralised infrastructure ([27], [28]), which typically include rural areas in the developed world and more widely in the developing world.

1.4 UK Domestic Energy Use

1.4.1 Current Status

In the UK in 2014, domestic energy consumption was 27% of overall energy use [29]. This comprises 17.0% for electricity, 17.3% for hot water, and 65.7% for space heating, based on end-use consumption. The balance of these three primary uses has been stable since 2000, with the total falling by c.10% from the peak in 2004.

Electricity use peaked in 2007 following a 100% increase in consumption through the 80's and 90's, primarily driven by increased use of consumer electronics and computer equipment. Since 2007, the impact of improvements in appliance and lighting energy efficiency have predominated.

Hot water energy consumption has been falling steadily since 1970 as the result of more efficient boilers offsetting any lifestyle changes [30], and, in particular, changing bathing habits [31].

Despite significant improvements in building thermal efficiency in the same period, heating energy increased steadily by 68% from 1970 to 2000 as a result of a 6°C increase in internal temperature from greater central heating availability and a 40% increase in the number of households [30]. Since 2004 it has fallen by 22% from the peak.

1.4.2 Future Drivers

Projections for future UK energy use are highly variable [5]. The uncertainty reflects that demand is driven by a wide range of influences, including economic growth, speed and scale of energy efficiency initiatives, and the willingness of consumers to change habits. Of four potential 2030 scenarios mooted by the UK National Grid [1], with significant variations in projected economic growth and decarbonisation speed, the proportion of energy from renewable sources varies from 11% to 30%, annual electricity demand from 332 to 362TWh, and gas demand from 200 to 300TWh [5].

There are three primary drivers for future changes in domestic energy use. Major disruptors, such as smart meter-driven demand shifting and electric cars, that individually have the potential to change overall demand profiles significantly. In addition, building design and appliance technology changes provide a constantly changing dimension which can range from small incremental changes to immediately apparent effects, driven by a diverse range of factors, such as new technologies, legislated energy efficiency targets, and changing trends. And finally, changes in occupant lifestyle, driven primarily by changes in employment patterns, average age, and evolving habits, influences both the frequency and timing of specific demands.

1.4.2.1 Major Disruptors

Household smart energy meters linked to a wider ‘Smart Grid’ will allow increased monitoring and control of energy use facilitating better balancing of supply and demand [32]. Household monitoring of use has the potential to reduce overall demand, although initial results suggest that the impact is small (<2.5%) in overall terms [33]. Shifting of demand through timers, remote activation, and increased use of time-based tariffs, is therefore the potentially more significant outcome. These elements are typically known collectively as demand management.

Electric car registrations in the UK have increased from an average of 0.3% of total sales for the first six months of 2014 to 1.3% for the first six months of 2016 [34]. Projections for future penetration, both in the UK [35] and globally [36], are highly variable but all indicate a high growth rate. This has implications for domestic electricity demand with potentially a significant number of, typically off-peak, hours of charging required and the potential to use the car battery for storage ([37], [38]). Understanding how this can be integrated with the timing of other energy use, both at the household and grid scales, will be required.

Other potential disruptors include the electrification of heating demand through use of heat pumps, and the development of energy storage technologies that make large-scale storage a feasible option, which would require a significant improvement on current storage technology [39]. The integration of some or all of these major disruptors requires detailed modelling of their impact, both individually and in combination [7].

1.4.2.2 Changing Technologies

Less dramatic but no less significant changes in demand are driven by the continuous evolution in appliance-level demand because of changing trends (e.g. use of internet-linked devices), technological improvements (e.g. flat-screen televisions replacing CRT units), and the drive to reduce energy consumption (e.g. increasing use of low energy lightbulbs, condensing boilers).

As outlined, until 2007, the overall impact of this evolution was to increase domestic electricity use, but since then has driven a reduction. Projecting the overall influence into the future is difficult with the conflicting requirements of a more technology-driven society and the need for an overall reduction in energy use. It is, however, at least safe to conclude that the effect is unlikely to be neutral and requires the means to analyse the impact.

Changes in the energy efficiency requirements of building construction standards has driven a significant reduction in the heating energy used by new-build housing. For example, a 2016-built UK home requires approximately 20% of the heating energy of a 1976-built home for the same thermal performance [40]. The overall effect is reduced by the slow replacement rate for housing, with an average net increase in stock of 0.2% but a demolition rate of only 0.12% between 2011 and 2015 in Scotland [41]

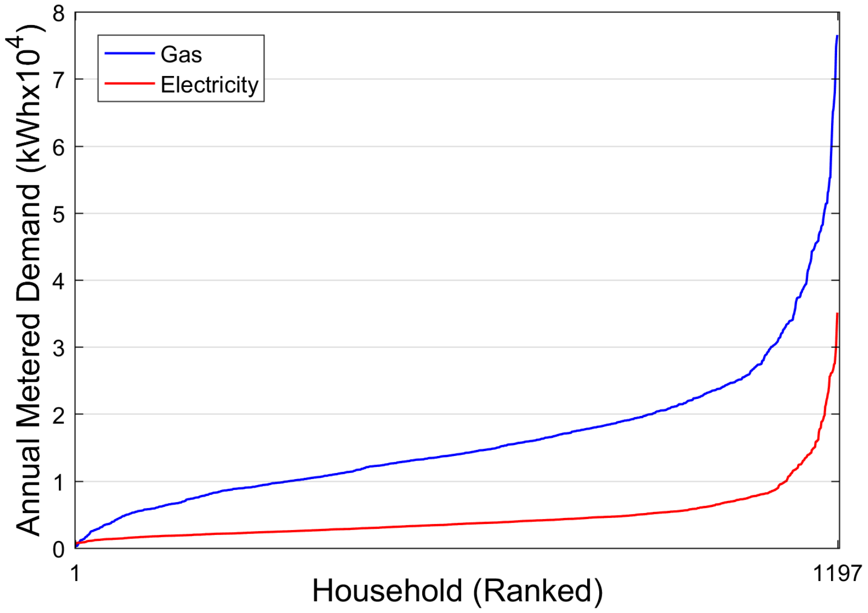


Figure 1.2. Ranked distribution of annual gas and electricity demand per household from analysis of the EFUS dataset [43].

and equivalent values of 0.6% and 0.06% for England [42]. However, building thermal efficiency improvements remain a significant driver for energy consumption reduction in the long-term.

1.4.2.3 Lifestyle Changes

The recent, primarily employment-driven, changes in UK lifestyles identified are predicted to continue and potentially accelerate [44]. This has implications for occupancy prediction with fewer people having traditional working patterns and increased homeworking. Prevalence of frequent homeworking increased from 11.1% to 13.9% of the working population between 1998 and 2014 [10].

Another future energy driver will be an ageing population, as both historical demographic patterns and improved health care take effect. As a minimum, this will influence occupancy patterns and household spending power, both of which impact energy use.

1.5 Demand Variation and Uncertainty

Reducing the scale of analysis increases the influence of individual household behaviours on system demand, significantly for household-scale analysis and with a diminishing influence as the scale increases. There are two main determinants of demand variation and uncertainty for any group of households; scale and type. In addition, the level of known information about the households at the design phase is another potential source. The degree of influence is also dependent on whether average or time-dependent demand is considered.

1.5.1 Scale-Dependent Uncertainty

Average energy demand per household varies significantly and is driven by a number of factors, including household size, income, and occupant behaviours, which are reviewed in detail in 2.4. This is demonstrated by the range of average annual electricity and gas use per gas-connected household from the nationally representative 1345-household UK EFUS dataset [43] as shown in Figure 1.2. As the scale of analysis is reduced the impact of these individual variations becomes increasingly important.

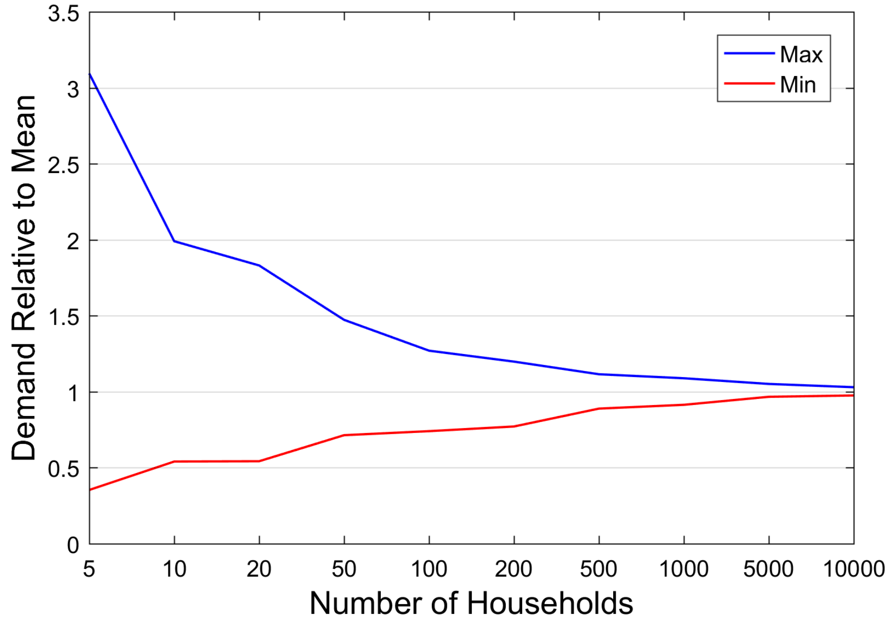


Figure 1.3. Maximum and minimum annual electricity demand from randomly selected combinations from the 1345-household EFUS dataset [43] relative to the mean value.

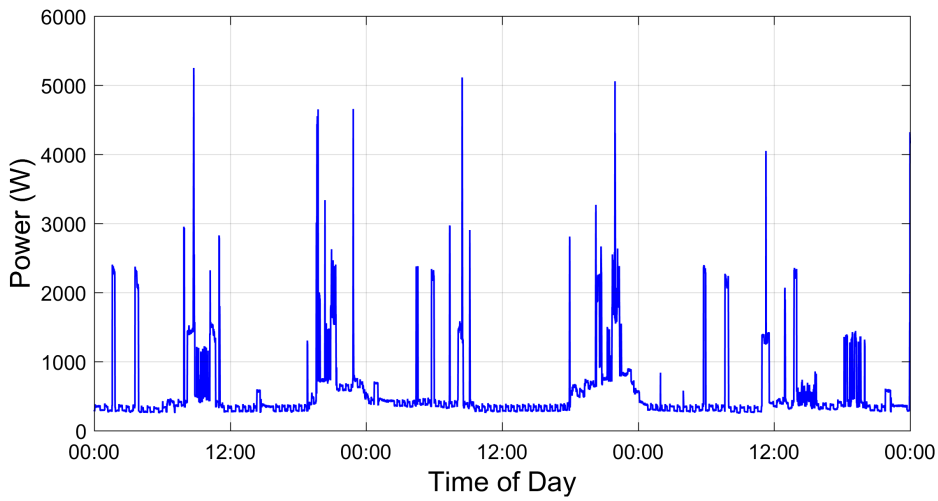


Figure 1.4. Example 3-day electricity demand profile for a single household from the REFIT dataset [45].

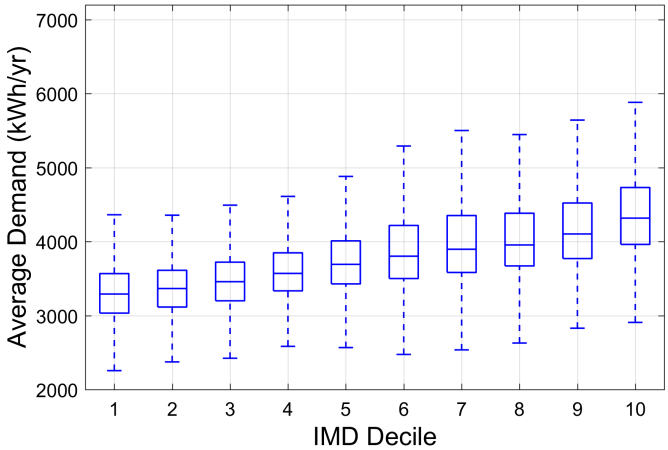
As an indication, the electricity distribution shown in Figure 1.2 was extrapolated and 1000 random combinations of different fixed numbers of households selected. The maximum and minimum ratios to the mean are shown in Figure 1.3. Whilst the generated groups of households will be more diverse than would be expected for many localised systems, and the extent to which the uncertainty is important is dependent on the type of analysis, scale-dependent average demand uncertainty is potentially significant to at least the 200-500 household range.

The scale influence on temporal demand is more difficult to assess due to the lack of large datasets with high resolution demand data for individual households. However, time-dependent electricity and hot water use per household is highly stochastic. As shown in Figure 1.4, electricity demand patterns for individual households are characterised by a baseline demand with intermittent spikes associated with high power, short cycle appliances, and the pattern varies from day to day. From this it can be inferred that data is required for a significant number of households on a single day, or for several days for a single household, for a stable average demand profile to be discernible, and therefore to any extent predictable.

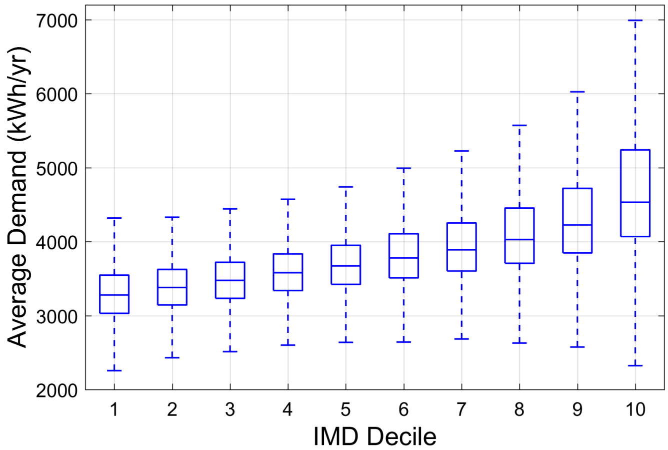
1.5.2 Type-Dependent Variation

Published average electricity [46] and gas [47] demand data for England and Wales for areas of between 600 and 1000 households shows significant variation in average energy use between areas (data at this resolution is not currently available for Scotland). Each area has a unique set of socio-economic characteristics with significant variations in age, household size, employment, and income; factors that have been identified as having an influence on relative energy use as reviewed in detail in 2.4.

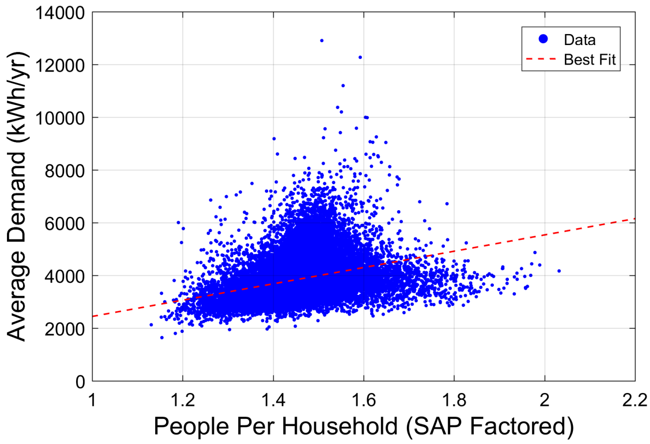
Figure 1.5(a) shows the distribution of electricity use differentiated by deciles of the combined Index of Multiple Deprivation (IMD), the official UK area deprivation measure that captures several socio-economic criteria, including income, employment probability, and living environment quality (the box represents the 25-50-75% percentile values and the limits of the dashed lines represent the 0.7 and 99.3% points in each data distribution). This shows that the average demand varies by c.30% between the lowest and highest deciles. Considering employment probability deciles specifically,



(a) Index of Multiple Deprivation (IMD) Decile



(b) Employment Probability



(c) Average Household Size

Figure 1.5. Relationship between average area annual electricity demand and three area defining characteristics. Data for analysis from [46]. (SAP Factored = (Number of Occupants)$^{0.4714}$ [48])

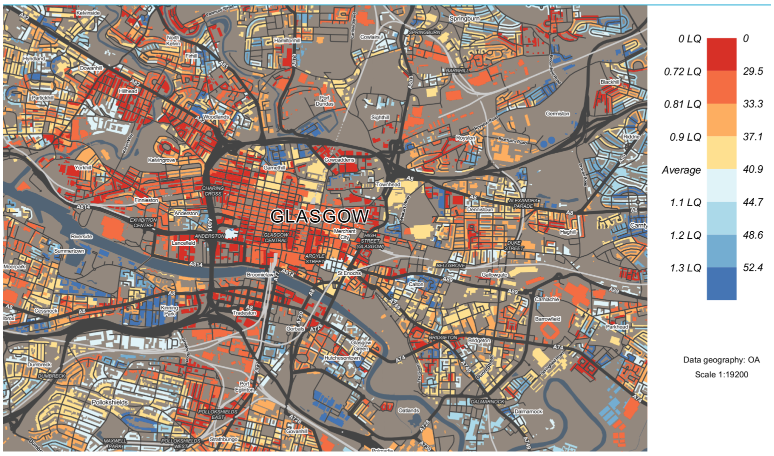
there is a similar pattern as shown by Figure 1.5(b). However, the relationship is more complex with the minimum demand levels rising then falling as employment increases, suggesting that employment also drives longer and more frequent absences from the dwelling. Figure 1.5(c) highlights that larger households drive typically higher demand (see best-fit line) as would be expected. The influence of household age profile is more complex and is reviewed further in Chapter 2.

Graphical representations of UK census data (based on areas of 50 households on average) from Datashine Scotland [49], shown in Figure 1.6 for a high population density, city centre area of c.4 miles by 3 miles, demonstrates that these factors can be highly variable over small areas. This demonstrates both why there is considerable demand variation between adjacent districts, and that, as the scale is reduced, there is an increase in potential for localised populations to deviate significantly from the national average. This indicates that demand prediction at the district scale (<1000 households is used as the definition for this project) requires an understanding of the area socio-economic characteristics and their potential influence.

As for scale-driven uncertainty, the influence of area characteristics on temporal demand is also complex. Time-dependency of demand is strongly linked to occupancy ([50], [51]) and therefore the variations in employment probability in particular are likely to influence relative demand timing.

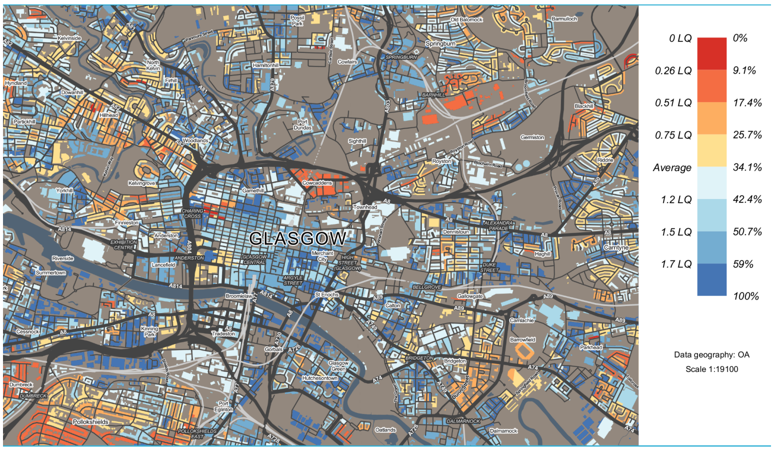
1.5.3 Known Household Data Uncertainty

During the design phase, the level of known data about the households will vary. For new-build schemes, the level of data that can be assumed is reduced. Beyond the socio-economics of the location, and the tenure and size of the house, no other information may be available. In contrast, for retro-fit schemes for existing housing, a significant level of information, including measured demand, may be available. However, little work has been done to determine the degree to which demand can be predicted from known information and the additional demand uncertainty that needs to be considered for increasing levels of unknown information.



(a) Average Age

(b) Economic Activity (Percentage Active)



(c) Household Size (Proportion of single-person households)

Figure 1.6. Distribution of three energy use determining average household characteristics at a c.50 household resolution for a 4-mile by 3-mile city centre area (Glasgow) [49].

1.6 Small-Scale Energy System Analysis

There are several different types of energy system analysis where the influence of individual household and small area demand patterns need to be considered. One principal example is the assessment and design of 'Distributed Generation' systems, the concept and potential of which was outlined in 1.3.1. In addition, planning for future demand patterns, demand management potential, and detailed building thermal performance simulation all require high resolution, small-scale demand prediction.

1.6.1 Distributed Generation

As outlined, unlike large-scale power plants or household-scale instantaneous generation systems (e.g. boilers), distributed generation (DG) technologies often do not produce energy consistently and at times that consistently match natural (i.e. prior to any incentivised demand shifting) domestic demand patterns [52]. As shown in 1.5, demand is itself highly variable both day-to-day within households and between different households. This overall variability and the on-demand expectation from consumers makes the design and implementation of certain types of DG systems challenging [16].

Where distributed generation is connected to a capacity-constrained section of the national grid, is designed to primarily supply a designated area with minimised wider grid interaction, or, in extreme cases, designed to be integrated within an independent small-area grid, several design issues must be addressed. The primary concern being supply and demand optimisation in the absence of effective storage technology ([53], [54], [55], and others). Five key assessment areas to address for systems that aim for a degree of grid independence ('microgrids') as identified by Abu-Sharkh et al [16] are as follows:

- *Sizing and Balance* - Understanding the balance of supply and demand at different time scales is important. At short time scales (milliseconds to hours) for control design and circuit sizing and at longer time scales (days to years) for economic analysis. In addition, the sizing of the generation equipment and any connected storage (see below) is also determined by detailed, time-dependent analysis.

- *Grid Connection* - Where national grid connection is feasible, which is primarily driven by location, it needs to be determined whether the connection is required. This will depend on the potential for supply shortfalls, the feasibility of exporting to the grid during periods of over-supply, and the associated economic analysis, including connection costs. This is primarily an issue for electricity grids but can also apply to district heating and cooling systems supplied from non-standard sources (e.g. industrial waste heat). For electricity grids, the impact of the import and export dynamics on grid voltage stability also must be considered.

- *Storage* - The ability to store excess supply at times when it exceeds demand, and to supply the stored energy at times of high demand, is a key component of any energy system.

- *Demand Management* - Shifting demand from periods of low supply or high demand to improve matching and reduce storage requirements.

- *Seasonal Matching* - Energy demand varies throughout the year, particularly for heating and cooling demand. Supply from renewable sources can also vary by season. The impact on system dynamics needs to be understood.

Energy system optimisation tools analyse the interaction between energy supply systems, distribution grids, and the consumers. This can include supply and demand matching, grid sizing and operation, and economic performance.

Existing optimisation software, such as *Homer*, *Retscreen*, and *Merit*, allow sophisticated representations of supplied-side equipment and environmental conditions to be simulated. However, they require either the user's own occupancy and demand data or the use of simplistic integrated archetypal schedules.

As the scale of the system for analysis reduces, the influence of individual consumer behaviour increases. There is evidence, however, that current design methods have not adequately accounted for the increasing uncertainty introduced at smaller scales, with an overreliance on inflexible generic design rules.

As an example, one of the principal design criteria in the design of any multi-consumer DG system, particularly those that seek to be largely independent of the national grid, is estimation of peak demand. The diversity of peak demand for multiple dwelling systems is often expressed as a coincidence or simultaneity factor in terms of

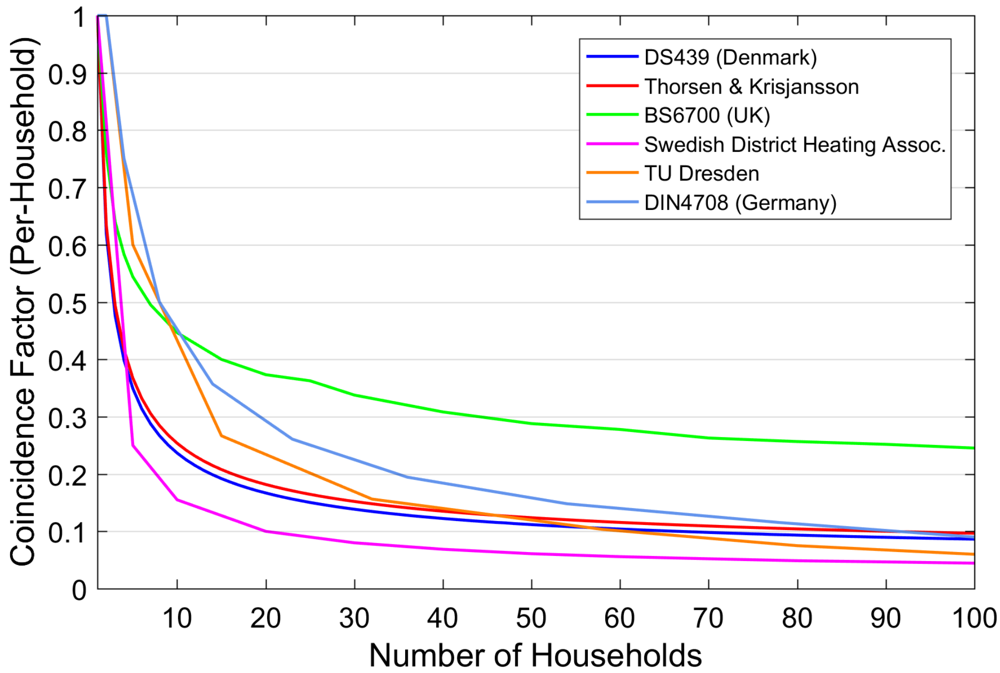


Figure 1.7. Coincidence factors for instantaneous domestic hot water systems by number of dwellings for multiple defined Standards. Data from [56], [59], [60], [61], and [62].

the proportion of the total connected load [56] or as a contribution per household ([57], [58]). This type of factor acknowledges that as the size of the system grows the number of consumers using the system at any time as a proportion of the total number falls. Accurate prediction of this factor impacts all five of the areas identified above.

For district heating networks, hot water diversity prediction is a key requirement for distribution pipe sizing. Several studies of hot water use diversity have been undertaken, particularly in Scandinavia and Germany where high historical use of district heating has driven the development of a significant body of standards and design tools. Figure 1.7 shows the diversity distributions from a number of sources, with CIBSE in the UK [63] recently endorsing the Danish DS439 standard [56] to replace the conservative BS6700 basis [59]. Even though the distributions range upwards from very small numbers of households, they do not account for any influence of the type of households included in the network or acknowledge any degree of uncertainty. Similar fixed design rules are also often used for electricity distribution and heating system sizing ([64], [65]). What are effectively 'rules-of-thumb' are therefore being used for significant design decisions without acknowledgement of the degree of uncertainty or range of applicability.

More generally, there is evidence that a lack of detailed demand prediction models for each of the three main overall demands (electricity, hot water, and space heating), and therefore an inability to accurately address the five design issues identified above and specific determinations like diversity, limits the potential for distributed generation in the UK [66].

1.6.2 Future Planning and Demand Management

A key function of energy system modelling is to understand the potential influence of future changes in specific demands and behaviours, and to investigate if the time dependency of certain demands can be altered to improve the balance between supply and demand (i.e. demand management).

An example of both future planning and demand management potential investigation is the integration of electric car charging while the vehicle is parked at the dwelling. Increasing use of electric cars will drive a significant increase in domestic electricity demand. In addition, the expected predominance of off-peak, overnight car charging, and potential to use the car battery as a flexible, short-duration energy store for the household, offers the potential for demand management. This requires both a detailed assessment of existing demands and predicted car charging patterns.

Other relevant areas where further detailed analysis at the grid sub-system scale is required include a number of the impacts identified in 1.4.2, including: the increasing use of electric heat pumps; changing building fabric standards on dwelling heat gains; and, changing working patterns, including further increases in homeworking.

1.6.3 Building Performance Simulation

Building performance simulation covers a range of typically computer-based methods to predict how specified building designs perform under different environmental and operating conditions. These can include energy performance assessment for the specified building, incorporating behavioural sub-models for energy use, and the assessment of thermal comfort over time as a result of external and internal heat gains.

Building performance simulation packages, such as *Energy-Plus* and *Esp-r*, are characterised by highly detailed representations of the building physics, thermal performance, and building systems (e.g. HVAC, blinds, etc.), but significantly less realistic representations of how the building is populated and used by the occupants [67].

The lack of detailed domestic occupancy behaviour and energy demand input data for thermal models also impacts other aspects of building performance simulation. As an example, the development of low-carbon building design, which typically drive buildings to be both air-tight and highly insulated, requires a careful assessment of the impact of occupant-driven heat gains, both casual and from appliance use, that potentially lead to overheating under certain conditions.

1.7 High Resolution Domestic Demand Modelling

1.7.1 Current Status

Building occupancy and occupant behaviour are acknowledged as key determinants of energy demand level and timing. Occupancy behaviour in both domestic and nondomestic buildings is currently the subject of the IEA EBC Annex 66 project [67], which has the aim of defining both the current state-of-the-art and the future research direction.

The identified link between building occupancy and demand drives the need for occupant behaviour simulation for demand modelling to comprise either separate building occupancy and demand behaviour sub-models or a single integrated occupancy-driven model. It is a subject area with few existing commercial or open-source examples.

The stochastic and country-specific nature of occupant behaviours makes this a challenging element to model. Domestic occupant behaviour and behavioural uncertainty prediction is a key problem area for which the overall impact on demand and system performance at the sub-1000 household scale is poorly understood and simulated, as recognised by the Annex 66 initiative.

Much of the key recent work in this area has been achieved by a variety of research teams, in particular [68], [69], [70] and [71]. However, the applicability of existing methods and overall models to comprehensively model occupancy-driven demand, allowing demand for small-scale systems to be assessed probabilistically, and with sufficient accuracy and differentiation to address the identified uncertainties driven by scale, household type, and level of known information, has yet to be fully assessed. This is required to allow the use of high resolution energy models to be extended from the production of generic, short-duration outputs to longer duration analysis that captures individual household behaviours, and is therefore more appropriate for detailed distributed generation, future planning, and building performance assessments.

1.7.2 Research Gaps and Aims

The most significant research gap with respect to accurate demand prediction analysis of small-scale domestic energy systems is the understanding of occupant behaviour, its influence on patterns of demand, and the resultant uncertainty. Whilst the influence of household characteristics and behaviours on energy demand is acknowledged, limited work has been done to quantify the influence and integrate it within a predictive demand model. The focus of this research project was therefore to review and utilise existing approaches where shown to be effective in capturing the occupant behaviour influence on demand, and the development of new or improved data analysis methods and simulation tools to address identified limitations.

The primary focus was to understand both the influence of the types of households supplied and energy system scale on system demand. Rather than take a deterministic approach, the aim was to utilise or develop probabilistic methods that would generate similarly probabilistic results, allowing the uncertainty in the predicted demand to be quantified. This should allow for improved design decisions by encouraging 'stress-testing' over a range of potential operating scenarios and reduce any poor performance that might be caused by households or communities that deviate significantly from average behaviours.

Supply and demand matching, in particular, requires high time resolution input data to be accurate with a sub-hourly resolution required as a minimum [72], and significant benefit of using at least 1-minute resolution data [73] to capture high frequency supply and demand variations. Sub-minute data is rare, however there is some evidence that energy systems analysis improves when sub-minute models are used, with specific benefits for wind and PV optimisation [74], and peak demand analysis [75]. The highest resolution for existing models is typically a 1-minute resolution ([69], [70]). The aim was to incorporate this level of time resolution as a minimum, with higher resolutions considered within the limitations of the available calibration data.

To enhance the probabilistic assessment of peak demand, and also to allow seasonal variations to be analysed, extended periods models of at least six months' duration were also a specific aim, although the twin aims of high resolution and extended period modelling required further review with the parallel requirement of effective computational speed for community-scale analysis.

To mirror the globally consistent goals of Annex 66, prioritisation was given the development of a flexible, bottom-up approach using commonly available data to allow the developed methods to be applicable to a wide variety of energy design problems and only country-specific in relation to the examples used. This would also ensure that elements of the overall model could be updated on the release of updated or more comprehensive UK data in the future, or to account for the expected behaviour changes associated with significant penetration of disruptive technologies, such as smart meters and electric cars.

1.7.3 Project Aims Summary

In summary, the aims of the project were as follows:

- Development of a domestic energy demand model:
  - Review existing methods and incorporate where shown to be effective to allow focus on areas with the highest potential for improved method development.
  - Capture individual household variation based on characteristics and behaviours, including time-dependent occupancy-driven variations, probabilistically and at the appliance-level.
  - Model time resolution basis to be highest possible using the available calibration data.
  - Extended period model basis, with target of a minimum of six-month duration simulations to allow seasonal influences to be analysed and to generate data suitable for probabilistic analysis, such as peak demand prediction.
  - Computation speed that allows practical modelling of community-scale (min. 100-200 households) energy systems.
- Use the developed model to quantify demand uncertainty for different scales and types of communities.
- Analyse potential impact of the identified demand uncertainty on performance of a distributed generation system, with specific focus on connections to either constrained grid sub-systems or independent 'microgrids'.

The presented analysis has been limited to electricity and hot water demand modelling. The lack of detailed heating data for calibration and validation has not allowed this element to be investigated fully at this time.

1.8 Chapter Summary

This chapter detailed the overall themes, research gaps to be targeted, and the specific aims of the work. The chapter highlights are as follows:

- Distributed generation has significant potential as a solution to the energy supply needs and challenges for the UK and globally. Although there is also significant uncertainty in the degree of future development.

- Overall domestic energy demand, and its time dependency, will continue to evolve, driven by influences such as the integration of electric cars, demand management, and incremental technology change.

- Variation and uncertainty in energy demand in small-scale (<1000 households) energy systems is a complex interaction of the type of households included in the system, the scale of the system, and the degree to which the household characteristics are known in advance. Occupancy and occupant behaviour differences are a significant factor in determining variations in energy demand between households, as acknowledged by the ongoing IEA Annex 66 project.

- Distributed generation integration requires a detailed understanding of both demand and supply to allow effective matching, estimation of the degree, timing and impact of national grid importing and exporting, and accurate prediction of peak supply periods for system sizing. In addition, detailed domestic demand analysis is required for future planning, demand management, and building performance assessments.

- The key research aim was to utilise and develop methods to predict energy demand at the small-scale and with high time resolution. By incorporating the influence of both household characteristics and random behaviours within a probabilistic modelling framework, the target was a similarly probabilistic model output, allowing the degree of uncertainty in the results to be assessed and robust energy system design decisions to be made in response.

Chapter 2

Household Characteristics, Occupancy, and Energy Demand

2.1 Chapter Overview

In Chapter 1, it was shown that energy demand varies significantly between areas with different socio-economic characteristics and that average demand between randomly selected groups of households can also vary significantly. This chapter reviews in detail the relationship between household characteristics and both occupancy and demand, identifying those which have a strong influence and therefore must be incorporated in any high-resolution demand model. Further to this, the influence of behavioural variations on demand differences between households with similar identifiable characteristics, and its impact on potential demand prediction accuracy, is also explored.

2.2 Energy Demand Prediction Background

To account for the observed variations in household and district energy demand shown in Chapter 1, various household characteristics have been identified that are correlated with domestic energy consumption. Studies by Kreutzer and Knight [76], Yohanis et al [77], McLoughlin et al [78], Haldi and Robinson [79], and Kelly et al [80] have shown that these include, but are not limited to; household size, number of children, floor area, bedroom number, occupant age, income, employment status, tenure, and location (i.e. urban/suburban/rural etc.).

What is less clear is whether the influence of these characteristics is direct or indirect, and, therefore, whether they are effective in predicting demand. For example, is the apparent correlation between bedroom number and demand, a direct result of the number of rooms and therefore increased floor area, a result of the relationship between number of bedrooms and household size, or a complex interaction with both direct and indirect effects?

The degree of influence also varies with the time resolution of analysis and, increasingly with higher resolution, the extent to which the characteristics predict occupancy patterns. For example, working-age and retired single-person households have similar average electricity demand but differences in the timing (see Figure 2.6(a) and (b)).

Most electrical appliance initiation and lighting use, almost all hot water use (if not necessarily when it is heated), and when a house is most likely to be heated, are all driven by occupant presence. Of the identified characteristics, the influence of age, employment status, and household size on energy demand would be expected to be at least partially driven by the resultant occupancy variations. Specific consideration of occupancy duration is therefore of importance for all demand models, with variations in occupancy timing relevant to models with a resolution of less than a day.

Existing research has demonstrated that only a proportion of the differences in energy demand between households can be attributed to household characteristics. The remaining element that determines household energy demand is a less tangible behavioural element. Gill et al [81] have shown that part of the residual variation can be attributed to attitudes to energy use, energy efficiency, environmental concerns, and spending prioritisation. They measured this potential attitude-driven variation to be 37% for electricity, 51% for heating, and 11% for hot water.

Haldi and Robinson [79] investigated the need to move beyond deterministic modelling of occupant behaviours to stochastic models that capture both the extremes and distribution of likely behaviours. They applied this specifically to window and blind use, and found that repeated probabilistic simulation was required to capture real behaviours. From their review of existing work and their own simulation analysis they concluded that occupant behaviour has a significant impact on overall demand, at least by a factor of two, and significantly in excess of this for extreme behaviours.

A key research question is therefore to determine to what extent energy demand can be predicted directly by household characteristics, by household occupancy patterns that result from the household characteristics, and what residual uncertainty remains from occupant behaviours, and how the relative influence of each element can be captured effectively.

In addition, as outlined in Chapter 1, the degree to which household characteristics are known when the simulation is undertaken will vary depending on the type of analysis, from feasibility analysis and research to detailed evaluation of schemes for existing households. The impact of this additional potential uncertainty on modelling accuracy also needs to be better understood.

2.3 Occupancy and Demand Datasets

Occupancy and demand modelling is limited to a significant degree by available data. Calibration of high-resolution models requires equivalent high-resolution input data to be effective. The following section identifies the UK datasets that are available for different aspects of high-resolution modelling, from low-resolution data that allows an overall assessment of the influence of individual household characteristics, to household and appliance-level monitoring data that allows specific actions to be isolated and converted to statistical representations for detailed modelling.

2.3.1 Occupancy Data - Time-Use Surveys

A significant proportion of occupancy- and demand-related analytical studies and modelling methods use time-use survey data ([82], [51]). They represent the only open-source occupancy datasets which include large-scale and representative data for the country surveyed.

Time-use data typically comprises self-compiled, single day diaries for individuals that captures both location and activity detail at a 10-15 minute resolution. The 2000 UK Time-Use Survey (TUS) dataset [83], for example, defines 146 separate activities that consolidate all potential occupant activities into appropriately linked groups at a 10-minute resolution. These include 'Sleep', 'Work', 'Laundry', and 'Food Preparation' which consolidates all cooking and meal preparation activities. The activities have also been used by some existing demand models ([70], [69], [71]) to infer appliance use (e.g. cooker use is inferred when the 'Food Preparation' activity is predicted).

Many country-specific time-use surveys follow similar basic guidelines allowing methods developed using data from one country to be transferable. The Multinational Time-Use Survey (MTUS) project [84] was initiated in the 1970s and now includes 60 surveys from 25 countries that have a common basic structure for activities and time resolution. Recent major European studies have been compiled according to a common guideline, 'Harmonised European time-use surveys' (Hetus) [85]. Fifteen countries have contributed Hetus-compliant studies at time of writing, all of which are also compliant with the MTUS requirements.

The 2000 survey is the most recent full UK survey for which data is currently available. This includes one weekday and one weekend day activity diary completed by 10490 individuals from 2490 households. A smaller, less detailed UK study was completed in 2005 [86], with 4941 single-day diaries from one person per household. A new, as yet unpublished, full UK survey was completed in 2015.

Few time-use surveys include diaries of longer periods than 24 hours, which is a significant limitation to determining the consistency of individual and household occupancy patterns. One exception is the Dutch 'Tijdsbestedingsonderzoek (TBO)' time-use survey, which includes 1-week diaries and has been completed at approximately 5 year intervals since 1990. The 2005 TBO TUS dataset [87] is the latest that has open access and includes 15428 diaries from one person per household.

Longer term occupancy studies are typically restricted to a small number of households that are unlikely to represent the full range of potential occupancy patterns and are typically not open-source.

2.3.2 Demand Data

At the lowest resolution applicable to small area analysis, total annual electricity and gas demand data is published annually by the UK Government for two area sizes for England and Wales (Medium-layer Super Output Areas (MSOAs) with between 2000 and 6000 households and Lower-layer Super Output Areas (LSOAs) with between 400 and 1200 households), and one area size for Scotland (Intermediate Government Zones (IGZs) with between 1000 and 2500 households [88]). The LSOA-level data allows assessment of the influence of different differentiating characteristics (e.g. income, age, household size, etc.) on average demand to be assessed without distortion from any

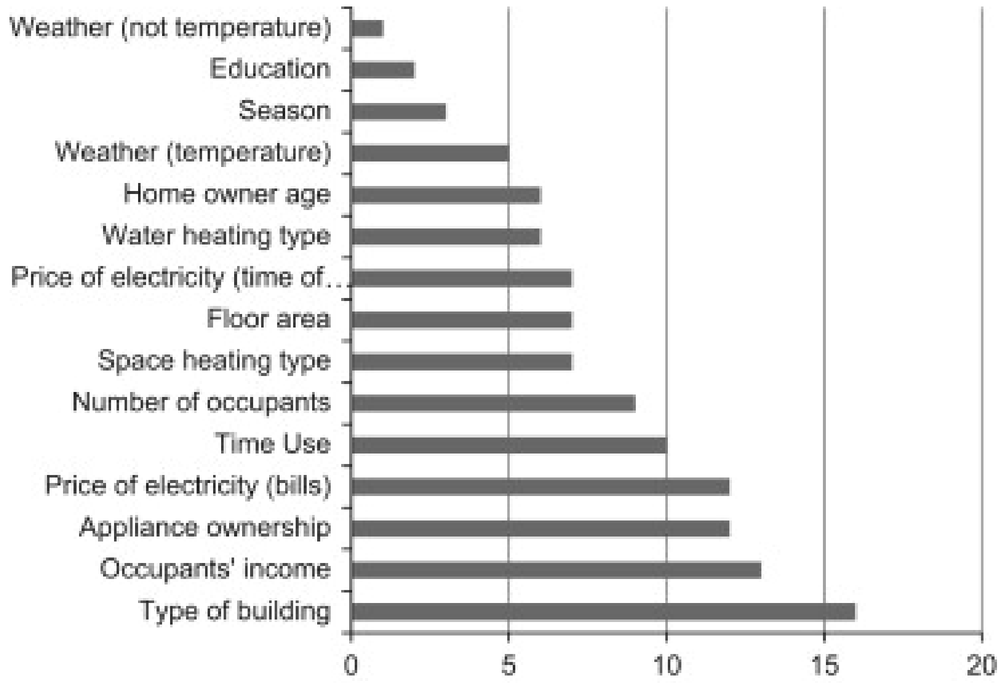


Figure 2.1. Data used in domestic electricity demand models by number of citations. Reprinted from [51]. (Image reproduced with permission of the rights holder, Elsevier.)

individual household behavioural influences.

Monitoring campaigns of sub-hourly energy usage, either for total or specific demands, in individual households represents the most detailed available data. For UK electricity use, the Government-commissioned Household Electricity Survey (HES) [89] includes detailed usage data of at least one-month duration for individual appliances in 251 UK households and corresponding household characteristics data. Smaller open-source datasets compiled by research teams include: the 22-household total electricity demand study by Richardson et al [69] with a 1-minute resolution over a two-year period; and the REFIT study [45] of total electricity and 6-8 specific appliance demands from 21 households over 1-year with a 6-second resolution. Both smaller studies have limited household characteristics data.

Other intermediate resolution dataset types include annual total demand from individual households (e.g. EFUS [43]). This level of data is primarily useful for model validation. (During the validation phase of this work the EFUS household characteristics data was not available and it was therefore not used, however, access to household detail can now be applied for.)

For UK hot water use the most comprehensive recent study was undertaken by the Energy Savings Trust (EST) in 2006/07, with 10-minute resolution total volume and temperature data for 107 households [90] for a 1-year duration, which is available on request to researchers.

2.4 Household Characteristics and Demand

As outlined in 2.2, the following characteristics have been identified by a number of authors ([76], [77], [78], [79], [80]) as being influential on energy consumption; household size, floor area, bedroom number, occupant age, income, children, employment status and tenure.

Analysis by Torriti [51] of the frequency that different types of differentiating data have been used by electricity demand models is reproduced in Figure 2.1. As Torriti states, the frequency may not reflect effectiveness but the model type and the specific data availability, however, it does give an indication of potentially useful model calibrating factors.

Of the factors identified few are completely independent and the interdependencies can be complex. For example, income varies with age, employment status, and household size and type, and itself influences appliance ownership and potentially also frequency and duration of appliance use.

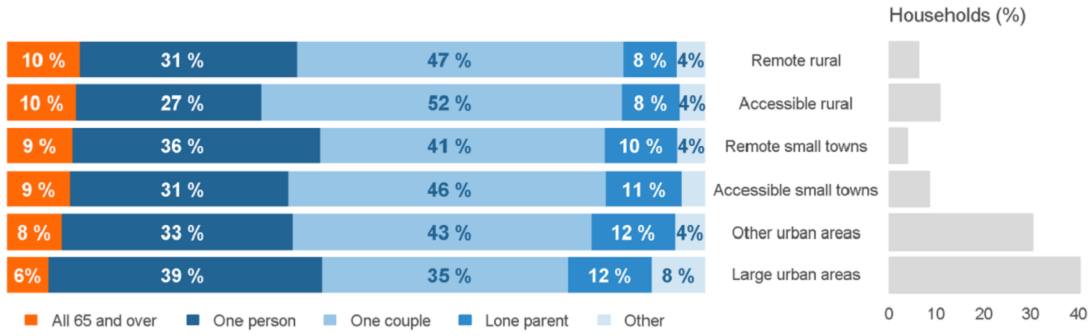
The analysis by Yohanis et al [77] reviewed each potential factor in turn but did not consider their independence. Whilst this is useful to indicate some level of correlation, it does not allow specific judgements to be made or indicate if and how each factor should be incorporated in a demand model.

The calculation steps for any model will also be determined by the level of household data that is available. The level of input data will be variable, from simple location information to detailed socio-economic data per household. Consequently, a demand model that aims to capture household-level variation and uncertainty should be flexible enough to accept different levels of known characteristics, and to provide a means to probabilistically generate realistic values for unknown variables (e.g. predicting income from household type and employment status).

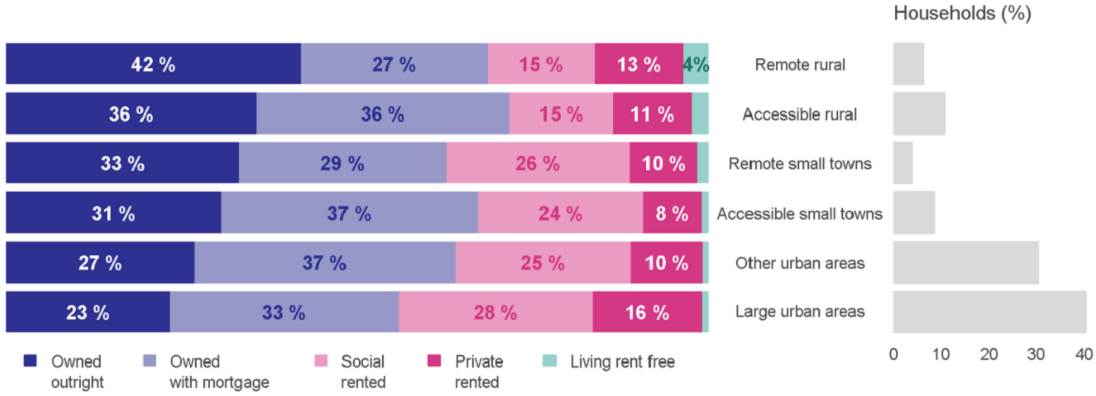
It is therefore necessary to review each potential model-calibrating characteristic factor, how they both influence and are influenced by other factors, and the degree to which they directly influence demand. This will indicate the necessary calculations steps required within a model to step from known characteristics to a demand prediction. For the review, it was assumed that the location and number of households was the minimum level of known information for any analysis.

With consideration of both presumed and identified influence on other factors, an increasing likelihood of being a direct influence on energy demand, and a decreasing likelihood of being a known input prior to any modelling exercise. The determined calculation sequence was: location; house tenure, type and size; household composition and age; employment; income; floor area; appliance ownership; energy ratings and power; occupancy and time-use; occupant behaviour; and specific demand use frequency and duration. The following section reviews each potential characteristic factor in turn, except for occupancy and time-use, and occupant behaviour, which are reviewed in 2.5 and 2.6 respectively.

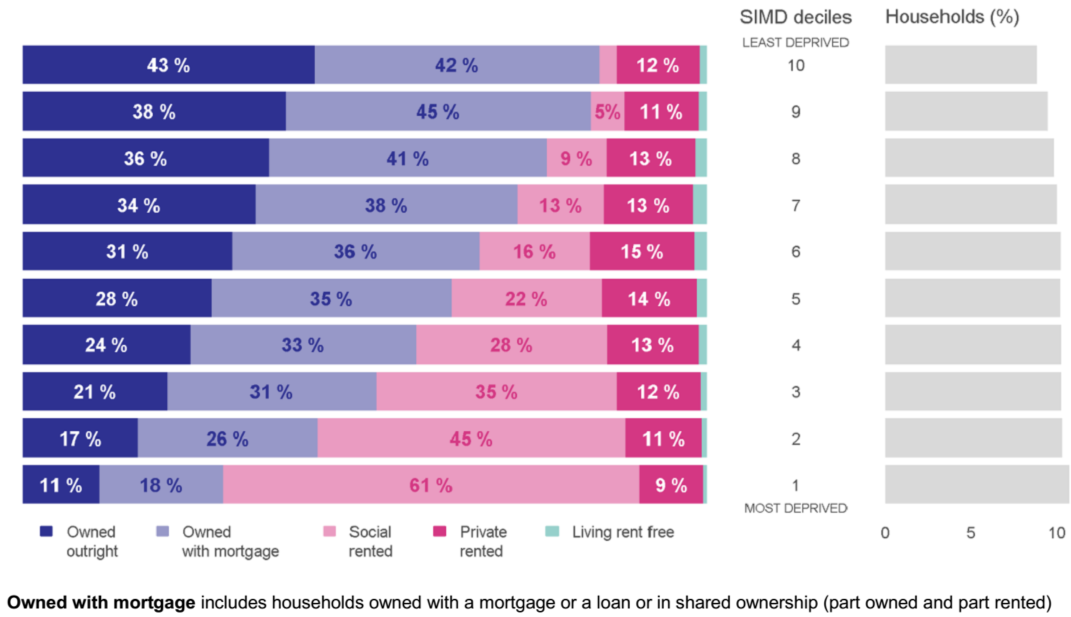
The review of each potential characteristic factor presented includes data taken directly from sources and further analysis of census and housing survey data by the



(a) Household type by area classification



(b) Tenure by area classification



(c) Tenure by deprivation decile

Figure 2.2. Variation by location type of key energy use determining household characteristics for Scotland. Reprinted from [91].

author where suitable data was not directly available. For clarity, data from both types of sources are presented together, with data compiled from further analysis identified.

2.4.1 Demand-Influencing Area and Household Characteristics

2.4.1.1 Location

As shown by Figure 1.6, location drives the probabilities for several of the identified influencing characteristics, such as age, employment, and household size.

A number of different criteria can be used to classify a location. The Rural/Urban classification, which in Scotland covers six area types from 'Remote Rural' to 'Large Urban' areas, defines proximity to large population centres and population density. In the UK, the socio-economic status of an area is defined by an Index of Multiple Deprivation (IMD), which is based on analysis of 38 indicators in 7 areas: income; employment; health; education, skills and training; housing; geographic access; and crime. The area of analysis for IMD typically covers 400 to 1200 households, and areas are typically grouped into IMD (deprivation) deciles.

Analysis of the Scottish 2011 Census data [91] confirmed that location, as defined by both classification (urban, rural etc.) and deprivation decile, has an influence on the population age range, the proportions of each household type and size, and the distribution of dwelling tenure, type (house/flat) and size (bedrooms). For example, urban populations have smaller and younger households (see Figure 2.2(a)) and a larger proportion of tenant households (both social and private) (see Figure 2.2(b)). An increasing (less deprived) area deprivation decile reduces the proportion of rental households (see Figure 2.2(c)).

Figure 1.6 showed that employment status is strongly correlated with location and Figure 2.3 demonstrates the link between employment and area deprivation decile using Scottish Household Survey data [92].

Analysis of the census data shows that while location alone cannot be used to directly predict all household characteristics and therefore energy demand, it is a significant determinant of many of the more directly influencing characteristic factors identified.

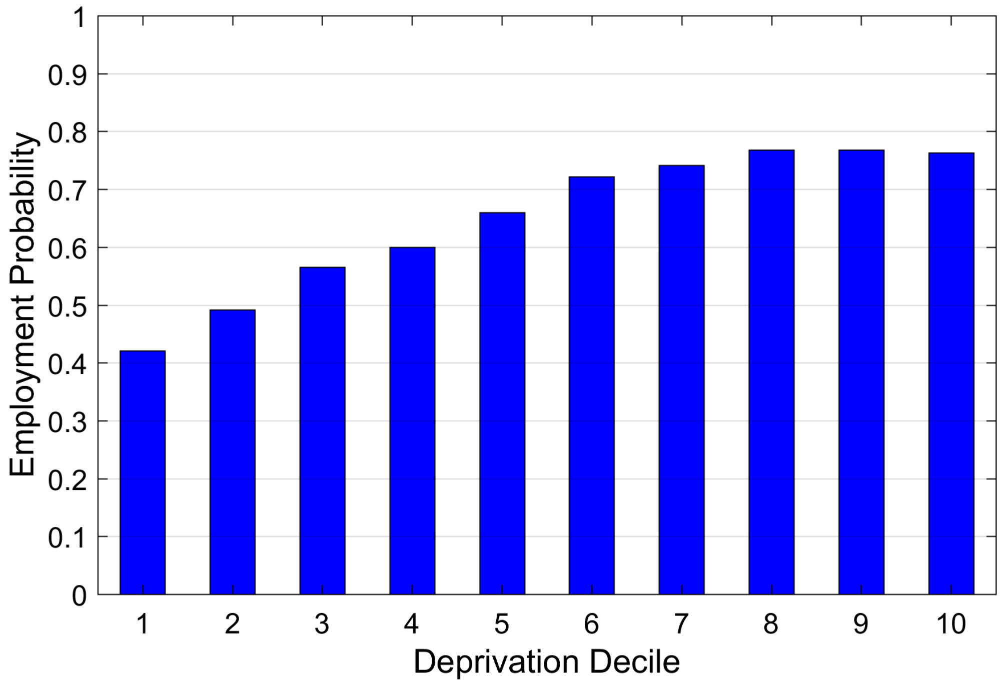


Figure 2.3. Employment probability per-adult by area deprivation (IMD) decile for Scotland. Data for analysis from the 2014 Scottish Household Survey [93].

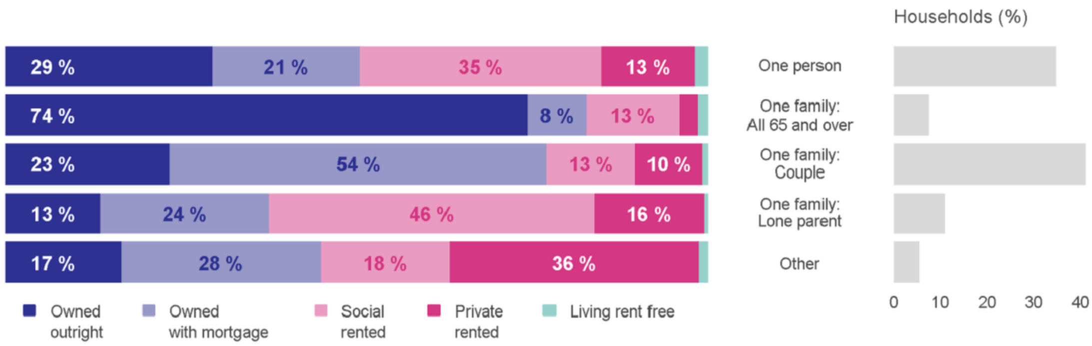


Figure 2.4. Household type by tenure for Scotland. Reprinted from [91].

2.4.1.2 House Tenure/Type/Size

The energy system location determines the mix of tenures (private, rented, social housing), house types (flats or houses), and house sizes (typically defined by number of bedrooms). This type of data can be derived from census data, by direct survey for existing housing, or from project plans for new-build schemes.

Tenure is strongly linked to household composition and age [91] (see Figure 2.4), and employment/education status [94]. In terms of energy use, potentially the most significant impact of tenure on household composition is the degree to which a house is under- or over-occupied, as measured by people per available bedroom space (allowing for reasonable sharing). According to the 2014 Scottish Housing Survey (SHS) [95], 74% of privately owned households had one or more spare bedrooms with 2% being classified as over-occupied, compared to 43% and 5% respectively for social-rented households. The values for England were similar [96].

In addition, tenure is a strong influence on employment probability. In Scotland, 82.1% of private, 33.5% of social-rented, and 64.0% of private-rented, working age householders are in employment [92].

Kreutzer and Knight [76] compared electricity demand in 69 social-rented houses to the UK average and determined that the demand was c.5-8% lower on average, and over 10% lower during the winter period, driven principally by lower ownership of higher-demand appliances. This is consistent with Figure 2.2(c), as social-rented housing is predominantly in lower deprivation (IMD) decile areas.

House type, and principally whether a house or flat, has a significant influence on the household type. From analysis of the 2014 SHS data [95], the relative proportion of younger occupants is higher in flats than in houses. For example, the probability that a 1-bed flat has a single occupant of working age is 14% higher than the average for all house types, while for a 1-bed house the probability is 35% lower. House type also impacts floor area, which influences heating and lighting demand, with flats being typically smaller than houses with the same number of rooms, especially for new-build homes [97].

House size, in terms of number of rooms, has a direct influence on number and type of occupants. Number of bedrooms, for example, as a minimum, places a limit on the maximum number of people that would be typically expected, but as discussed above, household space utilisation can vary from significant underuse to overcrowding. Housing survey analysis highlights that number of rooms also has a significant influence on floor area ([95], [96]).

Detailed analysis of the SHS data [95] shows that tenure, house type, and size, particularly when combined, have a significant impact on the probability of a household having particular composition, age, income, and employment characteristics, and for determining the house floor area. However, as with location, while they influence probabilities for characteristics, in most cases they do not limit the potential options. They are therefore important intermediate factors in determining energy use potential, but not directly predictive for high-resolution modelling.

2.4.1.3 Household Composition/Age

Household composition and age profile has a significant influence on a number of other household characteristics: employment/education probability; income; floor area; appliance ownership; appliance energy rating and power; occupancy; and specific demand use frequency and duration. The influence of household composition and age on occupancy characteristics is reviewed in detail in 2.5. Both were found to directly influence active occupancy potential.

Age defines the likelihood that an occupant is at school, is in higher education, is employed, or is retired. More detailed analysis of 2008 SHS data [92] also confirmed that within the working age group, employment probability reduced with age. For example, for single-person households, 73% of 18-33 year olds were in full-time employment, falling to 65%, 62%, 57%, and 45% for the 33-40, 41-47, 48-55, and 56-65 age ranges respectively. This, however, conflicts slightly with more recent UK-wide data for all types of employment, including part-time. For people of working age, employment probability peaks for the 25-49 age group (81.7% for 25-34 and 83.7% for 35-49) and is lower for both the 16-24 age group (75.4% for those not in full-time education) and the older 50-64 age group (70%) [98]. The indicates that the influence of age is both complex and closely linked with prevailing economic conditions.

Data provided by the UK Office of National Statistics (ONS) [99] also provides the proportion of each common household composition in each household income decile.

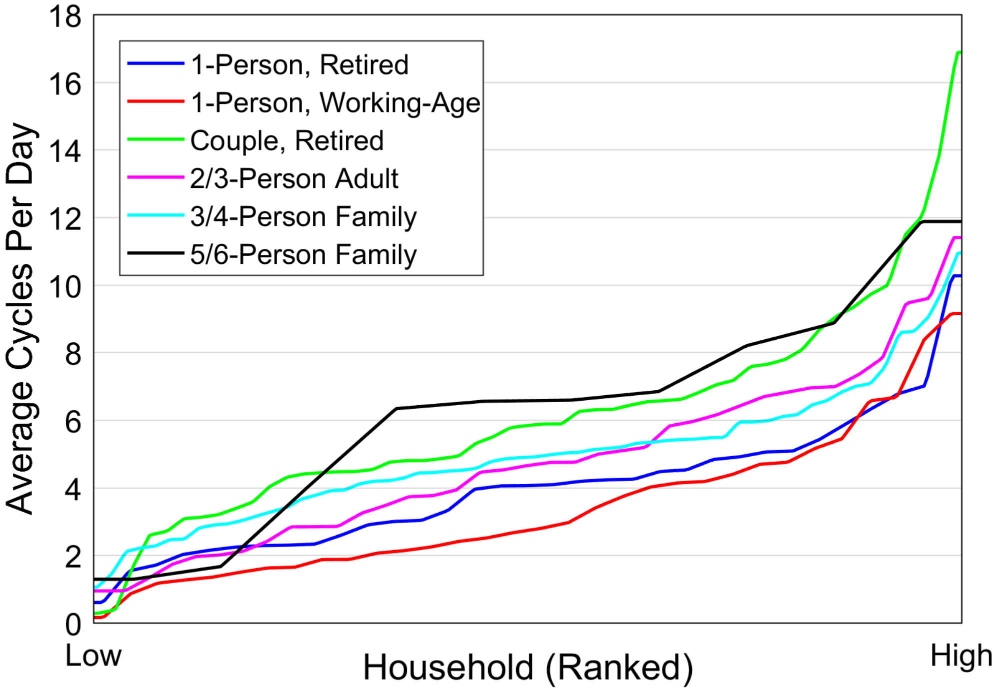


Figure 2.5. Ranked per-household distribution of average number of daily kettle cycles (use events) by household composition type. Data for analysis from the HES dataset [89].

As outlined below, this is strongly linked to employment potential per occupant.

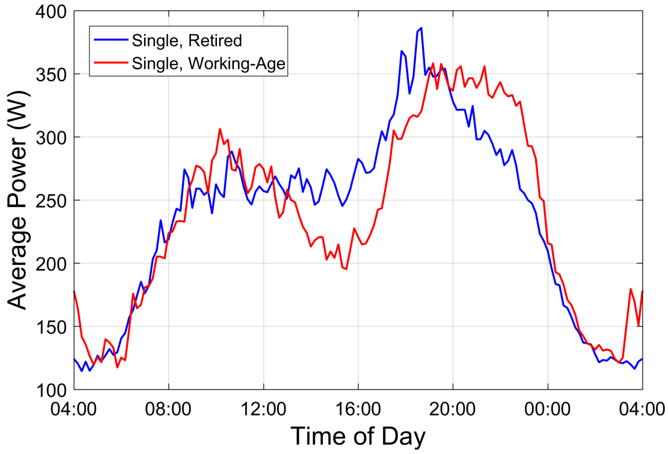
Additional ONS data [100] shows that household composition and age profile influences ownership potential for key high power appliances, such as dishwashers and tumble dryers. For example, dishwasher ownership in 2014 ranged from 17% for retired single-person households to 59% for households with two adults and two children. Ownership analysis of the Household Electricity Survey (HES) dataset [89] also showed that there was a strong link between household size and the number of televisions and cold appliances owned. For example, the average number of cold appliances is 1.51 for single-person households, 1.94 for 2-person, 2.03 for 3-person and 2.22 for all larger households.

In addition to simple ownership probability, analysis of the HES dataset by Palmer et al [101] determined that the age of appliances was correlated with household composition and age profile. In general, appliance age increases and energy efficiency decreases with increasing age of the main householder, decreasing income, and decreasing household size.

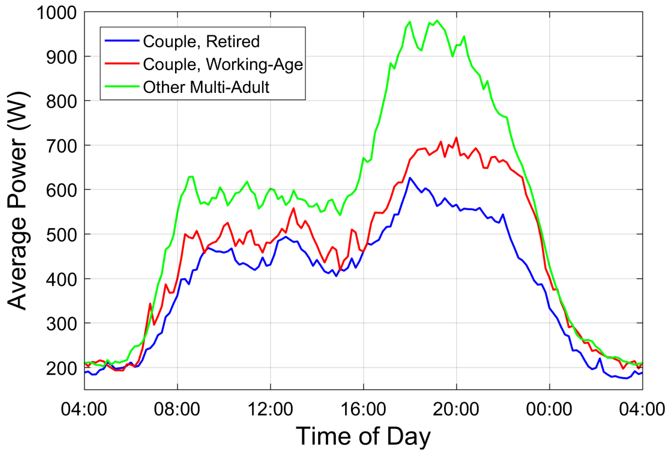
Linking absolute power capacity or size of appliances with household composition and age profile is more difficult due to lack of data and conflicting influences (unit age, type etc.). Specific data can be discerned from analysis of the HES dataset, such as the relationship between both cold appliance volume and television power, and household composition (increases with increasing household size), but this cannot be easily replicated for most appliances.

UK 2011 Census data shows that number of rooms, and by extension floor area, is influenced by household occupant number [102], age profile, and overall composition [103]. Potential distortions, as outlined above, include underoccupied households, such as older people remaining in what was the family home once their children have left [104].

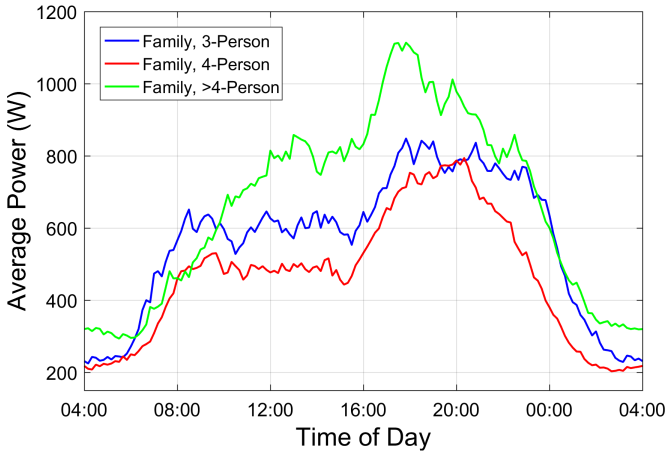
Analysis of the frequency and duration of specific demands differentiated by household composition and age profile determined that there was a strong influence on average use and also that there was significant variation within each household composition group. For example, the distribution of average kettle cycles per day for each household composition type (extrapolated to equal length distributions) is shown in Figure



(a) Single-person households



(b) Multi-adult households



(c) Family households

Figure 2.6. Average electricity demand per 10-minute timestep for different household composition types. Data for analysis from the HES dataset [89].

2.5. Data for other appliances shows similar characteristics. (A 'cycle' is defined as a distinct use event)

Demand data is not available with the same depth and household characteristics detail as occupancy data. However, it remains possible to assess differences in demand between different household composition types. For initial analysis of the HES dataset the households were divided into eight specific types of single-person, multi-adult, and family households, with the average electricity demand profile for each type shown in Figure 2.6.

Whilst there are clear differences in average power use for each distinct household composition type, the variation within each type (see Figure 2.7) suggests that composition and age alone are not sufficient to determine overall or specific demands with any accuracy as also shown by Figure 2.5. Similar analysis for hot water use again shows that the number of occupants influences the average demand but that the variation within each occupant number group exceeds the variation between groups.

The overall conclusion that can be drawn, therefore, is that household composition and age profile is a key determinant of several potential energy use characteristics and sets a baseline average energy use per appliance and per household, but is only directly correlated with average rather than individual household demand.

2.4.1.4 Employment

Employment has a direct influence on household income. As shown by Jenkins [105], equivalised income (i.e. factored for number of people in the household to normalise spending power (see below)) ranges from c.£250/week for households where the main householder is unemployed to c.£590/week for single-person and couple households all in full-time employment.

The most significant direct influence of employment on energy demand is the impact on the occupancy profile for individuals and the overall household (see Figure 2.14(b)), and consequently the relative timing of energy demand. This is reviewed in detail in 2.5.

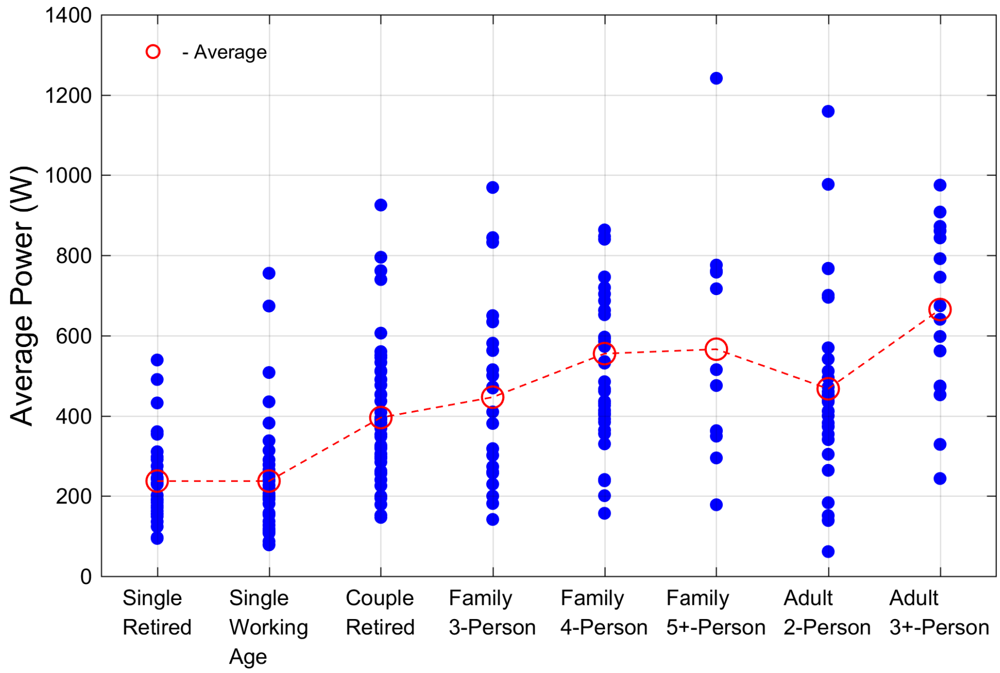


Figure 2.7. Average per-household electricity demand by household composition type. Data for analysis from the HES dataset [89].

2.4.1.5 Income

Income influences appliance ownership [100], household floor area (per person and overall) [106], appliance age (i.e. older, less energy efficient appliances vs. newer) [101], and both appliance-level and overall use behaviours [107]. Income is therefore an important direct and indirect influence on household energy use.

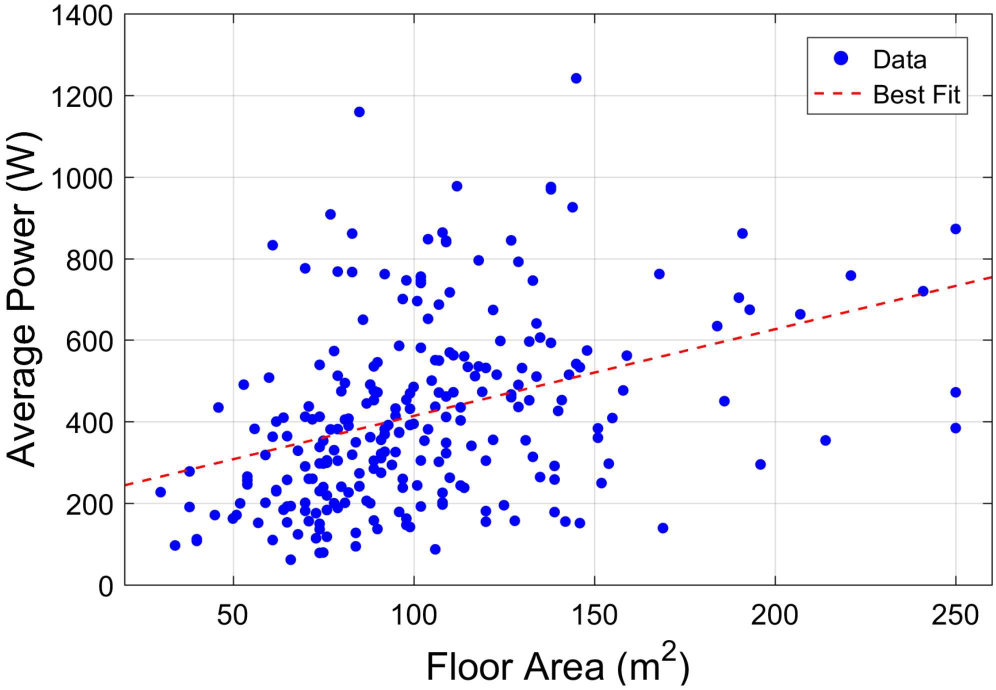
The overall relationship to energy demand is complex. White et al [107] identified households with high energy use despite low incomes. This was partly attributable to influences such as lower dwelling thermal efficiency but also suggests significant variations in energy spending prioritisation between households. Whilst there are highly variable behaviours for households with similar income characteristics, there is a clear influence of income on average behaviour. Jamasb and Meier [108] have calculated the income-specific elasticity in energy demand to be 0.06 (log of energy expenditure per log of income).

The household income effect is also dependent on the number and type of occupants. To determine a comparative spending power multiplier to account for household size, an approach has been developed by the OECD [109] and widely used. Person 1 is assigned a factor of 0.58, Person 2 and subsequent adults are each assigned a factor of 0.42, and each child is assigned an additional factor of 0.3. Household income is divided by the sum of occupant OECD factors to give an equivalised income for the household.

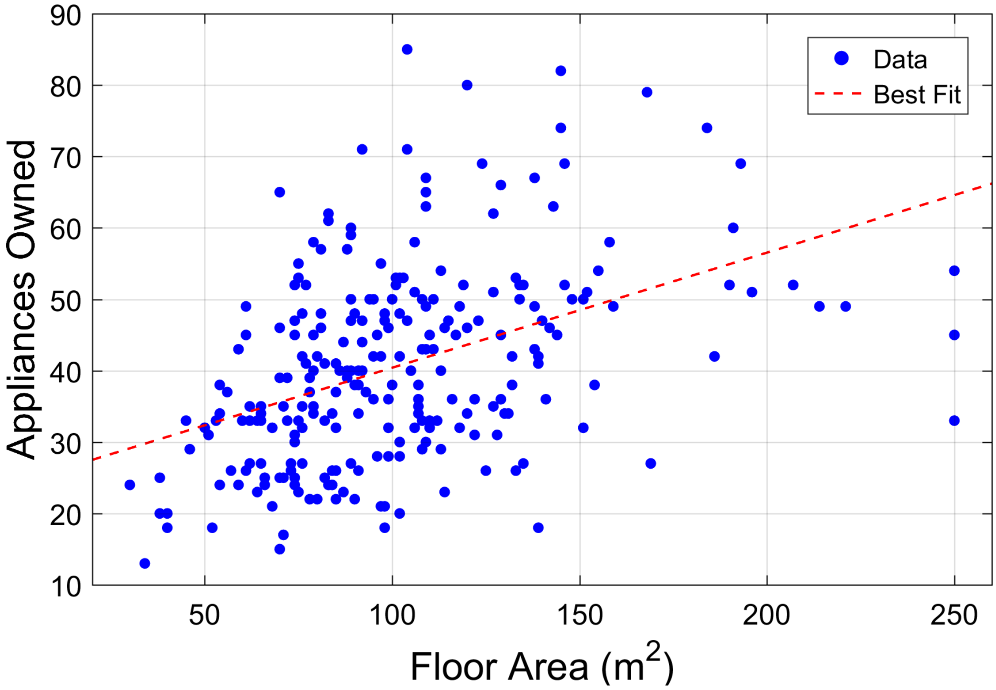
2.4.1.6 Floor Area

Floor area is a more accurate representation of the size of a house than number of rooms, particularly for heating and lighting demand prediction. However, as this data is more difficult to obtain, number of habitable rooms or bedrooms is often used as a proxy [110]. Floor area is used as a specific input for the BREDEM low resolution, housing-stock demand model [111] (the other input is occupant number), and as a unit of analysis for the UK National Energy Efficiency Data-Framework reporting [112].

Analysis of the HES dataset shown in Figure 2.8(a) indicates there is a relationship between floor area and demand, but it is relatively weak with high variation from the best-fit line shown. Figure 2.8(b) shows that there is also a weak relationship between floor area and the number of owned appliances. Floor area also potentially



(a) Average electricity demand



(b) Number of owned appliances

Figure 2.8. Average per-household electricity demand and number of owned appliances by household floor area. Data for analysis from the HES dataset [89].

impacts on the size, and therefore power use, of individual appliances, but there is insufficient data to allow this to be determined accurately. In addition, analysis by Yohanis et al [77] showed that demand per bedroom (used as a proxy for floor area) fell more significantly as the number of bedrooms increases than the increase in area would predict. The lighting analysis of the HES dataset by Terry et al [113] shows a clear relationship between lighting demand and floor area, although the strength of the relationship is difficult to quantify due to differences in monitoring periods and bulb efficiencies. The weaker overall correlations suggest that floor area is unlikely to be an effective defining characteristic for overall electricity demand, but a partial correlation with lighting demand is assumed.

2.4.1.7 Appliance Ownership

The number of appliances owned by a household is correlated with final energy demand. This was demonstrated by Jones and Firth [114], by Kreutzer and Knight [76] in relation to tenure-related demand differences, and by direct analysis of the HES dataset (see Figure 2.9). The relationship is complex with ownership only determining a potential for additional demand but not frequency of use or total energy used.

The types of appliances owned can also have a significant influence on energy use, particularly for households connected to the gas network. For these households, two of the key energy use appliances, cookers and showers, can be either electricity or gas-supplied, with cookers further complicated by the potential to have mixed units with gas hobs and an electric oven. Analysis of the EFUS dataset [43] determined that 42.6% of households with mains gas have an electric shower and 50.3% for those without. From the same dataset, 34.1% of households with mains gas have an all-electric cooker and 30.3% have an electric oven with gas hobs.

2.4.1.8 Energy Rating/Power

There is significant variation in power used by units within the same appliance type. This is driven by size and age of the appliances. Newer appliances tend to be more energy efficient, driven by technology improvements and energy efficiency legislation. The improvement can be gradual through incremental but cumulatively significant

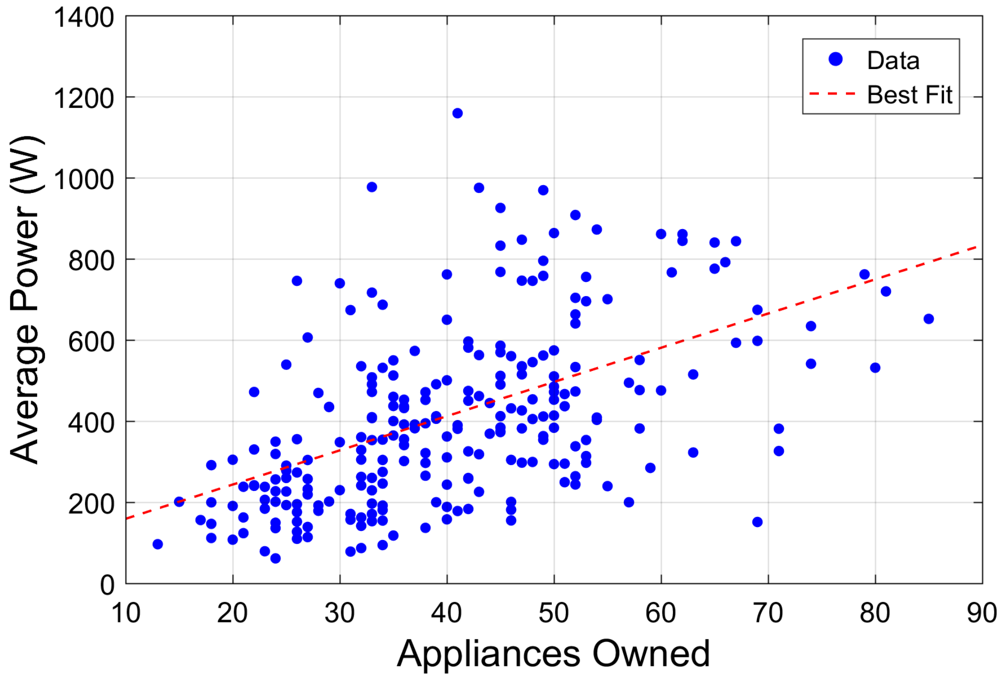


Figure 2.9. Average per-household electricity demand by number of owned appliances. Data for analysis from the HES dataset [89].

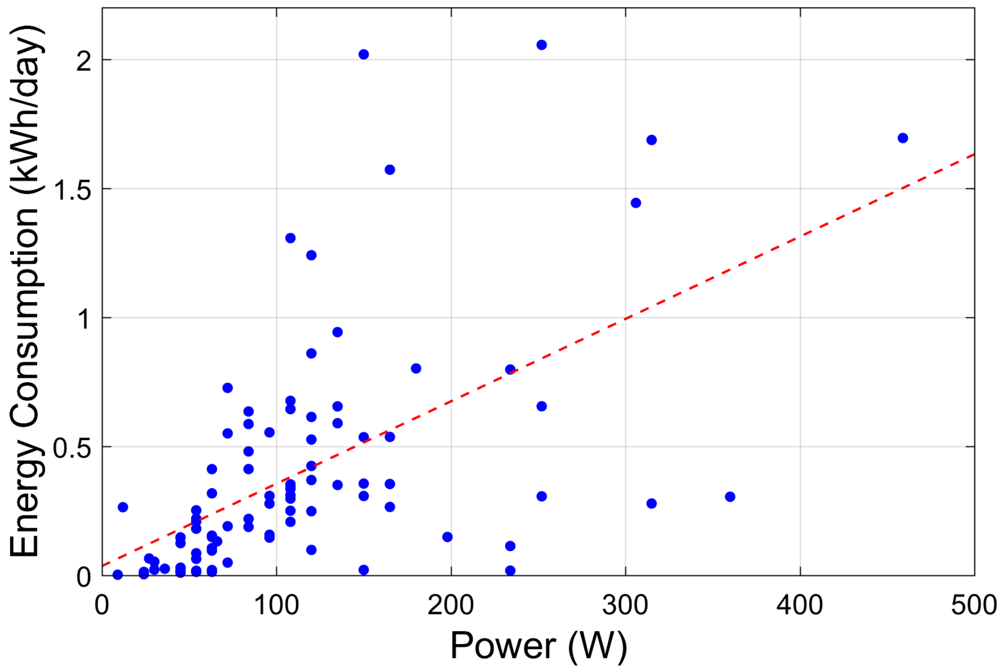


Figure 2.10. Relationship between TV maximum unit power and overall unit energy consumption. Data for analysis from the HES dataset [89].

improvements of existing technologies (e.g. water use reduction in washing machines), or more immediately significant where a new technology rapidly replaces an existing design (e.g. flat-screen televisions replacing CRT units using approximately one-third of the power in use [115]).

Analysis of the HES dataset determined that for each appliance type, there was a significant range of maximum and average power values. For example, the peak power for the main household television varied from 24W to 1584W, with an average of 151W.

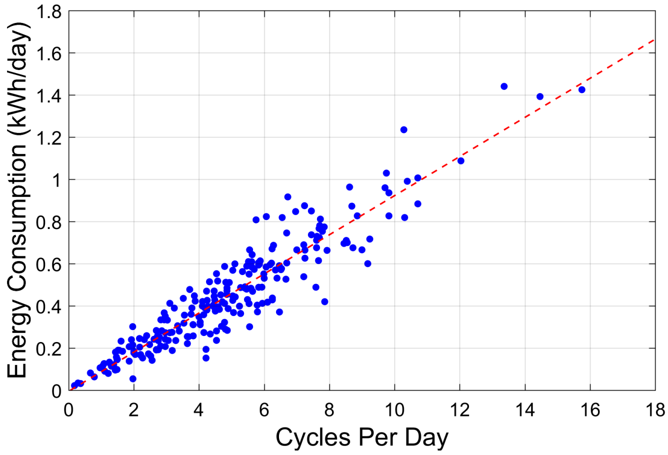
In Europe, several of the major power consuming household appliances (ovens, washing machines, dishwashers, cold appliances, televisions, and tumble dryers) and lightbulbs are defined by an energy rating. This is a European Union initiative to grade appliances based on relative energy efficiency. For example, washing machines are rated from A+++ to D, in descending order of efficiency, with each higher grade representing an average 10-15% decrease in total energy required for the same type of use [116].

The specific impact of the energy rating is difficult to discern for some appliances due to multiple potential operating cycles, particularly for laundry appliances and dishwashers. As detailed further in 5.7.3.1, for these appliances individual appliances cycles must be analysed. Televisions are the most frequently used constant power appliance, and the relationship between maximum unit power and average daily demand in kWh/day is shown in Figure 2.10. Whilst there is clearly a correlation between unit power and demand, it is relatively weak. However, in combination with the 'Use Frequency' factor introduced below, unit power is a primary characteristic in determining energy demand in a household.

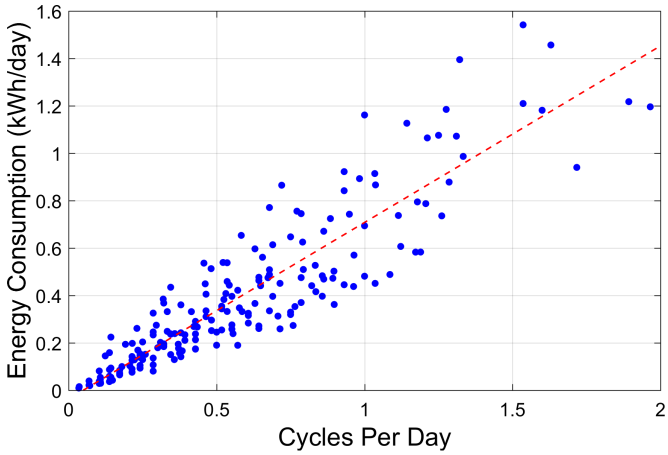
2.4.1.9 Use Frequency

The frequency with which each specific demand is used is a primary factor is determining energy demand and, as outlined above, is impacted by household composition and age profile, income, occupancy, and also the random behavioural element detailed in 2.6.

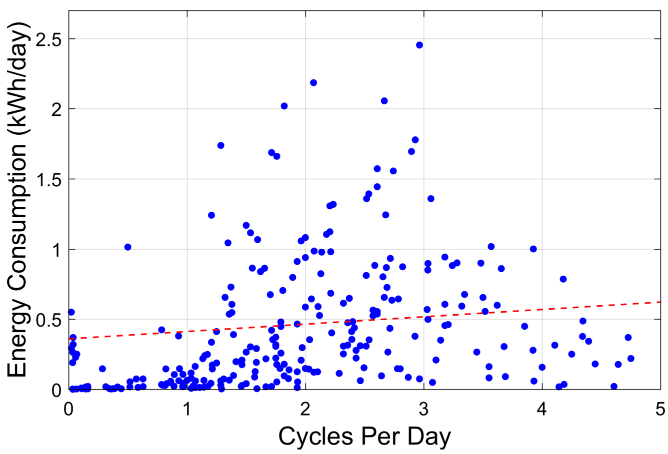
The direct influence of this factor on demand is use-specific. HES dataset analysis, specifically for electrical appliances, identified that demand for appliances with limited duration use cycles, such as kettles and microwaves, have a close correlation between



(a) Kettle



(b) Washing Machine



(c) TV

Figure 2.11. Relationship between cycles (use events) per day and daily unit energy consumption for three appliances. Data for analysis from the HES dataset [89].

daily cycle number and energy consumption, as shown by Figure 2.11(a) for kettle use. For appliances with fixed or typical cycle power profiles, such as washing machines and dishwashers, there is a degree of correlation but this is complicated by different cycle types and durations, and different appliance energy ratings. This is shown by Figure 2.11(b) for washing machine use which shows a less distinct correlation than for kettle use. Appliances with extended use cycles only limited by user behaviour, such as TVs, computers, dryers, and cookers, have a more complex relationship between cycles and consumption, with cycle frequency, duration, and unit power to be considered. Figure 2.11(c) for TV use, for example, shows only a weak correlation between daily cycle number and energy consumption. For hot water use, overall demand is similarly characterised by both use frequency and volume per use.

The appliance-specific relationships between use frequency, duration, and unit power are analysed further in Chapter 5, with distinct modelling methods developed to account for the different relationships demonstrated in Figure 2.11.

2.4.1.10 Demand Factors Summary

The analysis of the contribution of each identified household characteristic to demand prediction highlights that individually each one has only a weak correlation with demand, and therefore predicting demand requires the combined and relative influence of all characteristics and their inter-relationships to be accounted for.

Time-specific household demand is determined directly by frequency of use ('Use Frequency' factor), the power characteristics of the appliance ('Energy Ratings/Power'), and occupancy ('Occupancy/Time-Use'), with floor area also an important determinant specifically for lighting use. The determined sequence of inter-relationships between all identified household characteristics and these directly influencing factors, and therefore the required calculations steps to predict household demand from often limited household information, such as only location, and house type and size, is shown in Table 2.1.

Depending on the level of known household information at the time of analysis, several probabilistic assessments may be required to predict a number of demand-influencing household characteristics, such as age and employment status, prior to a prediction of the directly influencing factors and demand itself. The model development

Table 2.1
Relationships between identified energy demand-influencing household characteristics.

Characteristic	Influenced By	Has Impact On
Location		House Tenure/Type/Size Household Composition/Age Employment/Education
House Tenure/Type/Size	Location	Household Composition/Age Employment/Education Floor Area
Household Composition/Age	Location House Tenure/Type/Size	Employment/Education Income Appliance Ownership Energy Ratings/Power Occupancy/Time-Use Use Frequency
Employment/Education	Location House Tenure/Type/Size Household Composition/Age	Income Occupancy/Time-Use
Income	Household Composition/Age Employment/Education	Floor Area Appliance Ownership Energy Ratings/Power Use Frequency
Floor Area	House Tenure/Type/Size Income	Appliance Ownership Energy Ratings/Power **Demand**
Appliance Ownership	Household Composition/Age Income Floor Area	Use Frequency
Energy Ratings/Power	Household Composition/Age Income Floor Area	**Demand**
Occupant Behaviour		Use Frequency
Occupancy/Time-Use	Household Composition/Age Employment/Education	Use Frequency **Demand**
Use Frequency	Household Composition/Age Income Appliance Ownership Occupant Behaviour Occupancy/Time-Use	**Demand**
Demand	Floor Area Energy Ratings/Power Occupancy/Time-Use Use Frequency	

to account for the defined calculation steps is outlined in 4.2 and is detailed in Chapters 4, 5, and 6.

2.5 Occupancy and Demand

2.5.1 Background

The specific influence of occupancy on temporal demand for households has been well documented. Yao and Steemers [50] concluded that “both behavioural determinants and physical determinants related energy-consumption are more or less influenced by peoples occupancy pattern”, and that employment related daytime absences were the most significant occupancy effect.

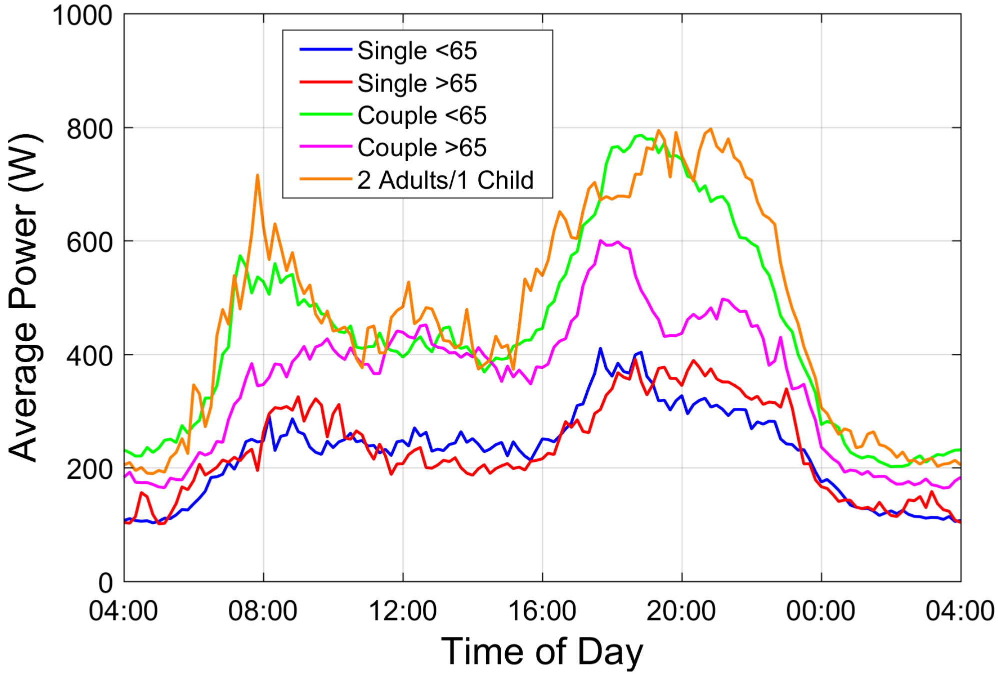
The influence of occupancy on domestic demand modelling can be determined from the review of demand models undertaken by Grandjean et al [82]. Over half the identified models, and 6 of the 9 detailed bottom-up models ([117], [118], [119], [120], [70], and [69]), incorporate standalone occupancy sub-models calibrated from time-use survey data. The output is used to directly influence when certain demands occur.

The relationship between occupancy and demand, and the variety of household characteristics that influence either or both, is complex. As an example, a higher income household potentially has lower than average occupancy (higher employment potential), but additionally, higher appliance ownership, larger appliances, more energy efficient appliances, and a tendency for higher usage behaviours. Understanding and accurately capturing the inter-relationships is critical for high-resolution modelling.

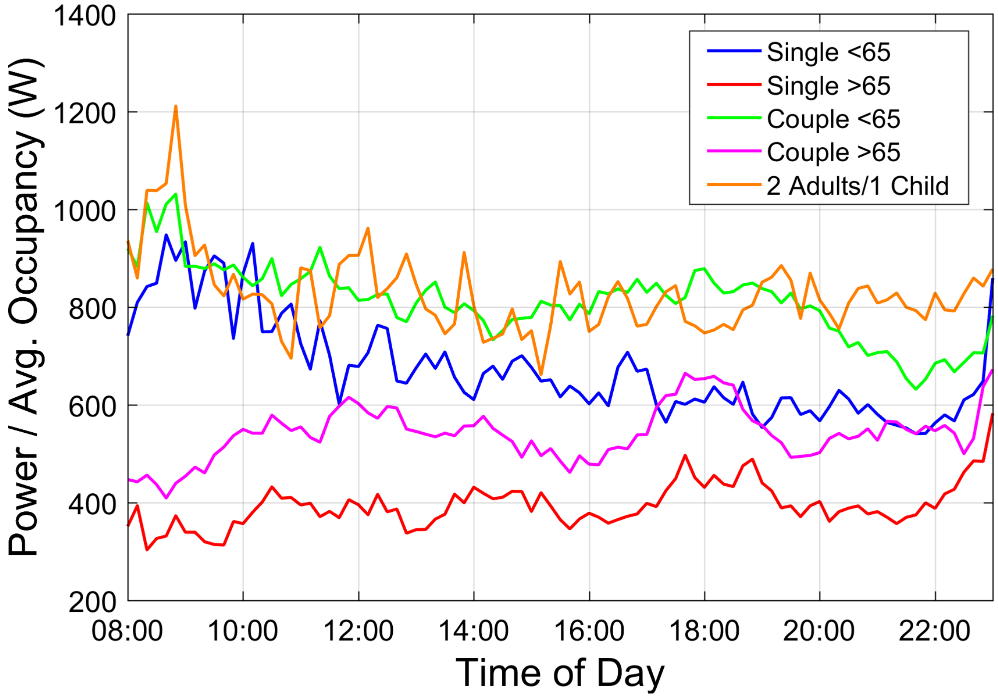
2.5.1.1 Electricity Demand

For electrical appliance demand, the overall household demand is an aggregate of a number of both continuous and occupant-initiated elements. Analysis of the HES dataset shows that the proportion of electricity used for typically user-initiated appliances is 78%, which would strongly indicate a strong correlation between occupancy and demand.

A review of the literature linking time-use behaviour and electricity demand was performed by Torriti [121], stating that “residential electricity demand profiles are



(a) Average electricity demand



(b) Electricity demand relative to occupancy

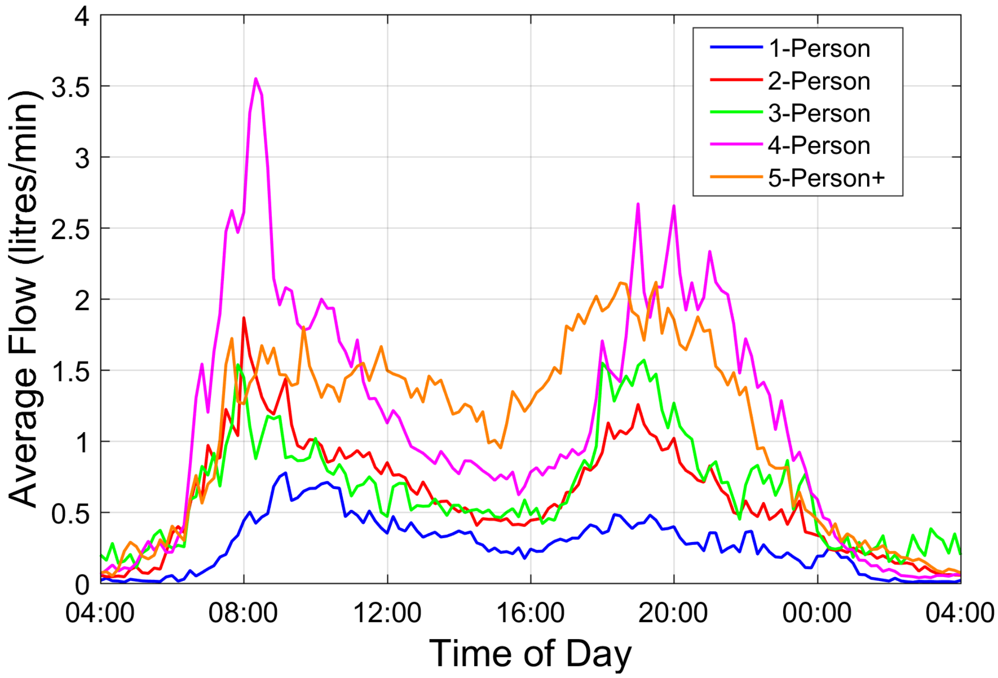
Figure 2.12. Average time-dependent electricity demand and demand relative to average active occupancy probability for different smaller household composition types. Electrical demand data for analysis from [89] and occupancy data for analysis from [83].

highly correlated with timing of active occupancy, i.e. when consumers are at home and awake". The relationship can be demonstrated by analysis of the HES dataset. When the average demand profile for a particular household composition type (see Figure 2.12(a)) is divided by the average occupancy probability profile derived from the equivalent UK 2000 TUS population, the factored relative profiles for the daytime period have a broadly linear shape (see Figure 2.12(b)), suggesting a strong correlation. (Smaller households are shown but the relationship is similar for all households.). However, there is also evidence of time-specific activities, such as cooking, showering etc., and the variable relative impact of continuously powered appliances, principally cold appliances, which increases during periods of lower active occupancy probability.

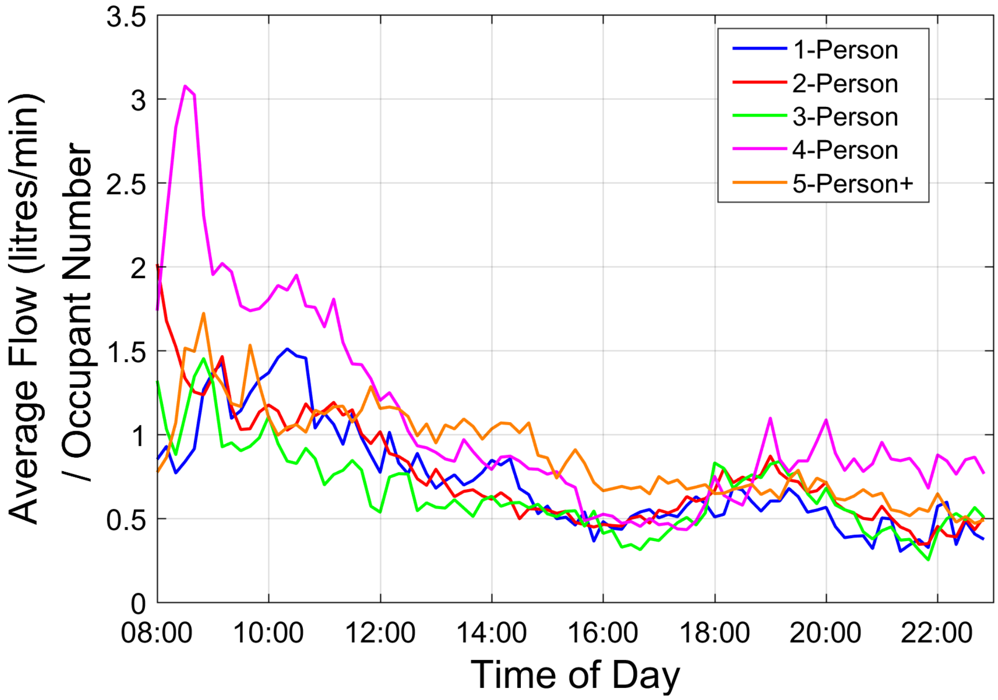
The relationship between occupancy and electricity demand is further complicated by the degree to which occupancy determines frequency of use. For example, the need to use appliances such as washing machines and irons is not strongly influenced by occupancy duration, and therefore occupancy will primarily drive when they are used not how frequently. Conversely, use frequency of smaller cooking appliances, such as kettles and microwaves, are likely to be much more closely correlated with total occupancy duration. The degree to which occupancy is related to both frequency and timing of use therefore needs to be understood for each specific demand, although this is currently limited by the lack of datasets which capture both occupancy and demand.

Evidence of Demand Management The link between occupancy and demand for occupant-initiated appliances is potentially complicated by demand management initiatives. As detailed in 1.4.2.1, demand management is where a user is incentivised to shift demand to a higher supply or lower demand period to help balance the network. In the UK, the principal current mechanisms are the 'Economy' tariffs (principally '7' which offers a single lower nighttime tariff, and also '10' which has three lower tariff periods throughout the day, both of which are predominantly focused on households with electric storage heating systems).

Analysis of the HES dataset demonstrates that only dishwashers exhibit any usage patterns that can be directly attributed to unattended, tariff-driven use (i.e. nightime period use at fixed times). Timer use probability can be indirectly assessed by the distribution of cycle (use event) start times. Assuming that most timers would be set for specific 'rounded' times (i.e. x:00, x:30 etc.), extensive timer use would result in



(a) Hot water demand



(b) Hot water demand relative to occupancy

Figure 2.13. Average time-dependent hot water demand and demand relative to average active occupancy probability for all household sizes. Hot water demand data from [90] and occupancy data from [83].

cycle start time distributions that are skewed. This can be shown for heating use, but there is no such evidence from the HES data for any appliances except dishwashers. For most electrical appliances, therefore, an occupant-initiated cycle start can be assumed.

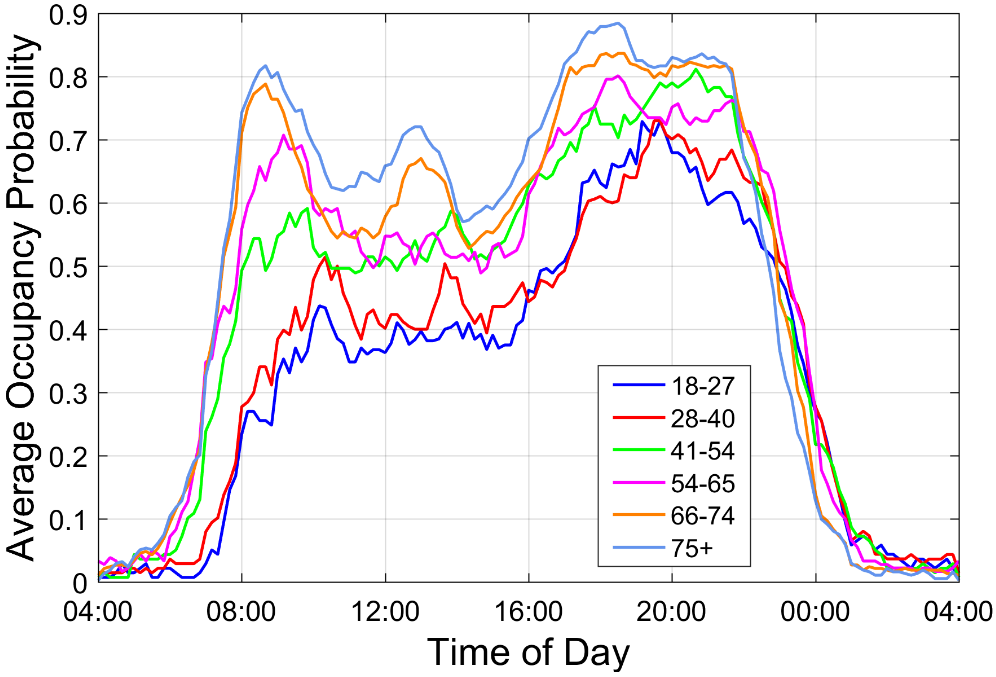
Whilst use of timers for heating systems are relatively common (c.60% of households [80]), timers for electrical appliances are less so. Only 13.6% of UK households currently have time-of-use tariffs [122], which may at least partially explain the low prevalence of shifting behaviour. There is also some evidence of a specific reluctance for people to allow electrical appliance to run unattended, due to fear of fires, or specific concerns such as, for example, water leaks or damp clothing issues associated with washing machine use [123].

The conclusion that can be drawn is that the current level of demand shifting is low. The implications for any electricity demand models calibrated using existing data of future behavioural changes caused by the use of remotely-activated appliances and more widespread use of demand shifting tariffs will need to be monitored and updated as required. Current data can, however, be considered a direct reflection of when appliances are likely to be used without forced behaviours, and used as a baseline for analysis of future demand shifting potential and impact.

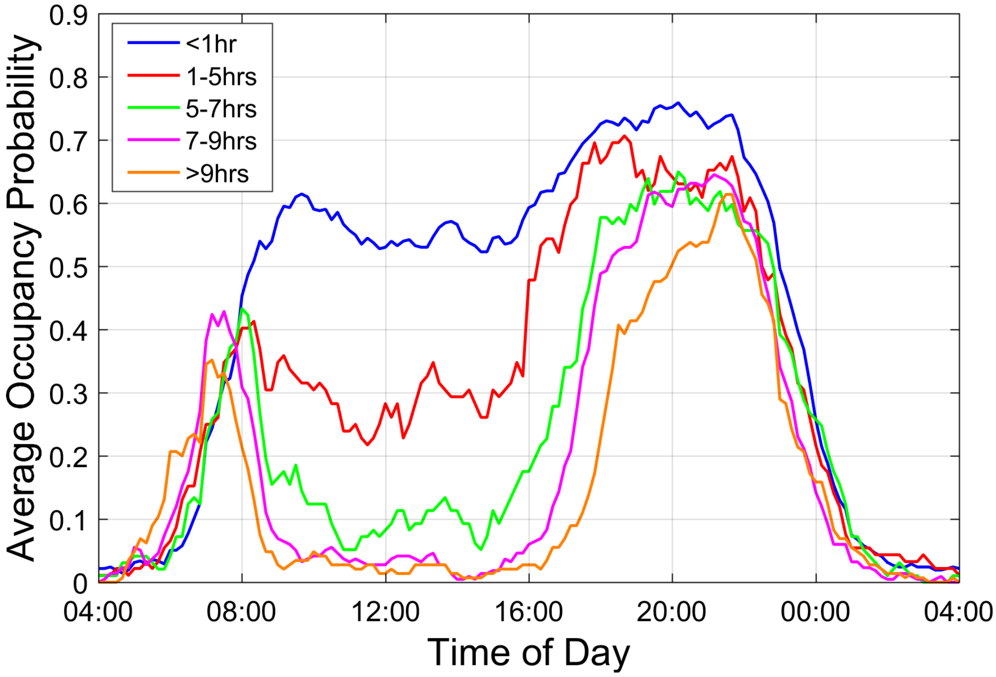
2.5.1.2 Hot Water Demand

For hot water use, a high overall correlation with occupancy would be expected, particularly as most hot water-driven appliances that could potentially be timer-controlled (i.e. washing machines, dishwashers) are typically cold water supplied. All main hot water use activities, such as showers, baths, manual dishwashing, and handwashing, all require active occupant presence.

The average point-of-use demand across all households in the Energy Savings Trust (EST) dataset (see 2.3) shows two distinct peaks, one at 7.30am, and the other at 7.30pm (see Figure 2.13(a)). Whilst it can be assumed that most hot water cycles are occupant-driven, the relationship with overall active occupancy is less distinct than for electricity use (see Figure 2.13(b)). This suggests that there are strong time-dependent activity drivers beyond basic occupancy (e.g. morning bathing, post-dinner dishwashing), particularly for higher volume hot water cycles.



(a) Age



(b) Working hours

Figure 2.14. Average active occupancy profiles from the UK 2000 TUS dataset [83] by age for non-working days and working hours' ranges.

2.5.2 Occupancy Data Analysis

The analysis in 2.4 concluded that several of the household characteristics that influenced demand did so indirectly because of their influence on occupancy patterns, namely; household composition, age, and employment/education status. Further review is therefore required to determine how significantly each potential factor influences occupancy.

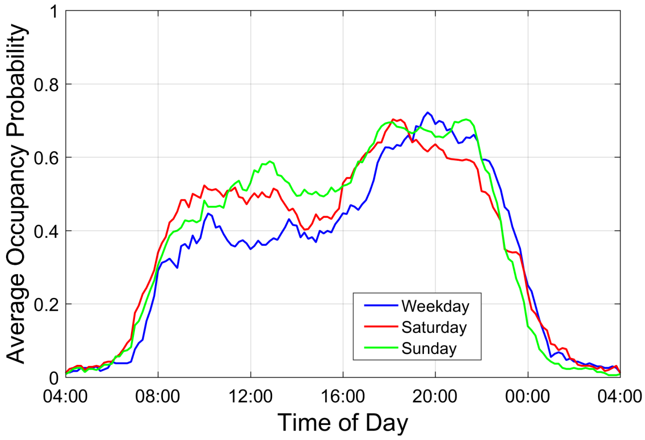
UK 2000 TUS occupancy data was analysed in three ways to evaluate the relative significance of a variety of potential differentiating occupant characteristics. Firstly, the influence of individual factors, such as age, gender, and employment status, was assessed. Secondly, different household composition types were analysed separately to determine if distinct patterns were observable. Finally, the degree of individual household variation within groups with similar basic characteristics was assessed. This was required to determine the most effective way to use existing data to develop a statistically robust, occupant-differentiated occupancy model.

Table 2.2
Results of regression analysis for occupant characteristic correlation with average active occupancy.

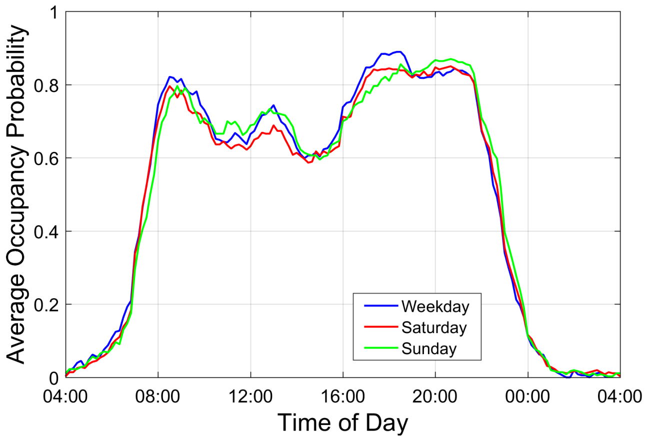
Correlation	Characteristic
Strong	Working Hours (p=2.47x10^{-47})
	Age (p=1.96x10^{-35})
Weak	Employment Status (p=0.016)
	Income (p=0.026)
	Day Type (p=0.048)
None	Location (p=0.632)
	Tenure (p=0.647)
	Gender (p=0.798)

2.5.2.1 Relative Influence of Occupant Characteristics on Occupancy

As outlined in 2.3.1, TUS data provides single-day occupancy data for a large number of individuals and, in some cases, overall households. The typical single-day diary is a limitation to understanding behavioural consistency, although the depth of data does allow an overall assessment of the influence on occupancy for a number of occupant characteristics.



(a) 18-40 age range



(b) 78+ age range

Figure 2.15. Average active occupancy profiles on non-working days by day type. Data for analysis from the UK 2000 TUS dataset [83].

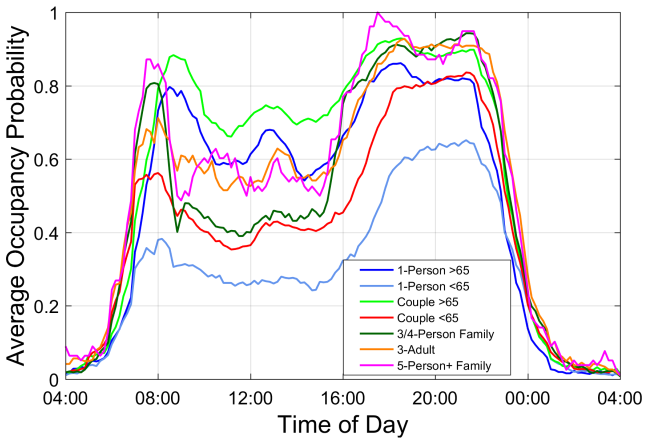


Figure 2.16. Average overall active occupancy profiles for different household composition types. Data for analysis from the UK 2000 TUS dataset [83].

An initial regression analysis of the UK 2000 TUS dataset for average occupancy included the following potential factors; day type (weekday/Sat/Sun), diary day working hours, employment status (yes/no), age, gender, tenure (own/rent), income (10 bands), and location population density (as a proxy for location type (i.e. urban/rural etc.)). Regression was used to allow the relative influence of each factor to be evaluated, and as it was expected that the factors selected were not independent and could not be effectively analysed individually. The analysis was limited to the single-person household population to remove potentially distorting interactions with other occupants. The regression results for correlation with average occupancy are shown in Table 2.2, ranked by degree of correlation.

Age and working hours on diary day were by far the strongest correlated characteristics with occupancy with very low p-values. This is further confirmed by Figure 2.14, which shows the average occupancy profiles for, (a) non-working weekdays by age range, and (b) for all working age people on weekdays by working hours' range.

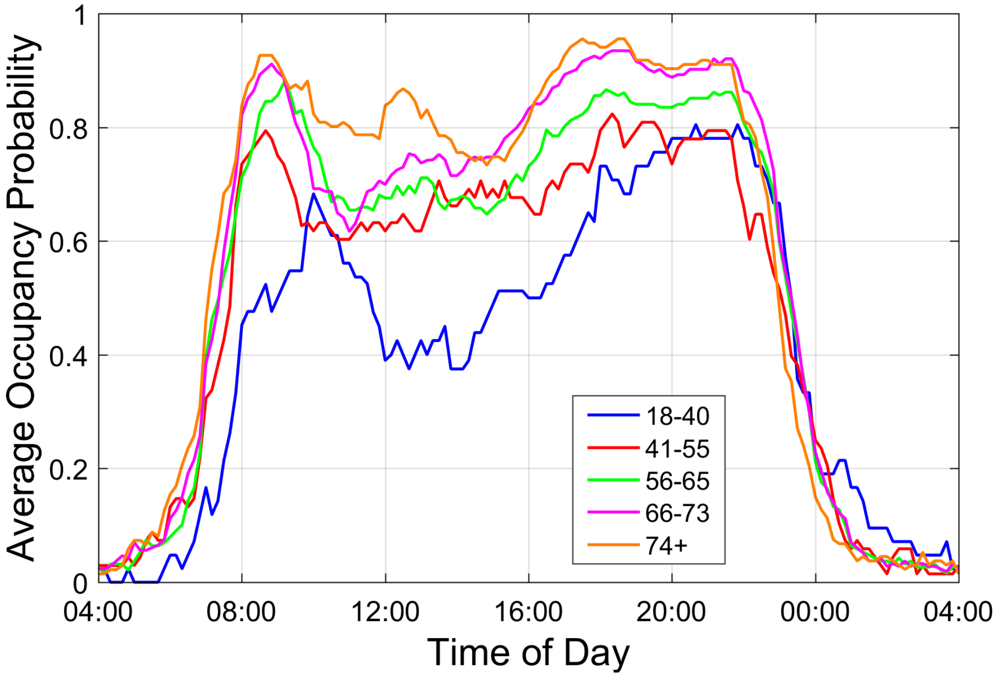
The results for both gender and day type were less significant than expected. Analysis of the dataset split by gender shows a significant average occupancy difference (male - 29.1% vs. female - 34.3%), but this is primarily driven by higher levels of diary day employment and worked hours in the male dataset. Once working days are removed, the difference is much less significant (42.4% vs. 43.6%).

The time dependency of day type on occupancy is more significant than the influence on average occupancy, therefore differentiation by day type requires further consideration. As shown in Figure 2.15, however, the day type influence reduces with age for non-working days, with the variation reducing for each successive age range analysed.

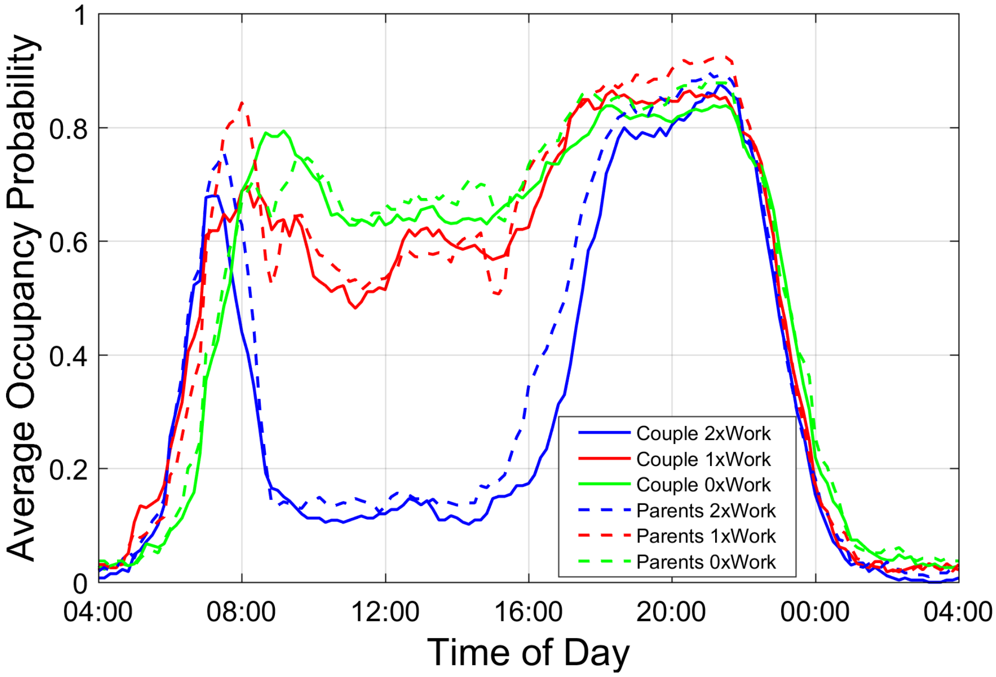
2.5.2.2 Household Composition Occupancy Characteristics

Further analysis was undertaken to determine the time-dependent variations in overall occupancy (i.e. periods with at least one person present and awake) between different household composition types. The households have been broadly characterised as single-person, couple, family, and multi-adult.

Figure 2.16 shows the time-dependent average active occupancy for a variety of different households, based only on whether at least one person is present. The results



(a) Couples on non-working weekday by age



(b) Couples and parents by employment status

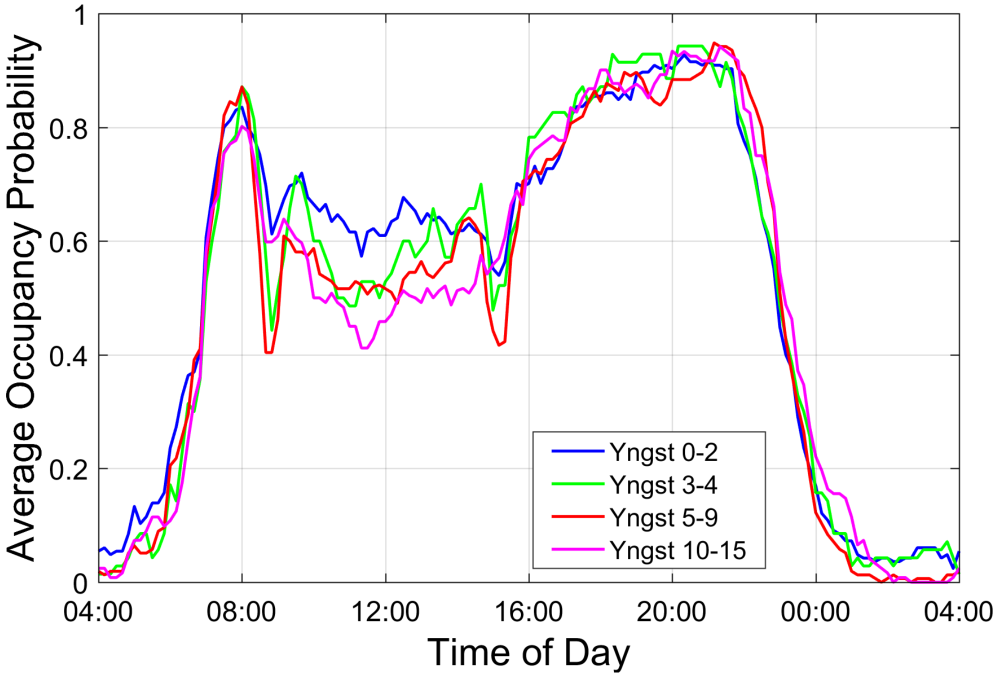
Figure 2.17. Combined average active occupancy profiles for couples by age for non-working days, and couples and parents by employment status. Data from the UK 2000 TUS dataset [83].

highlight some key characteristics:

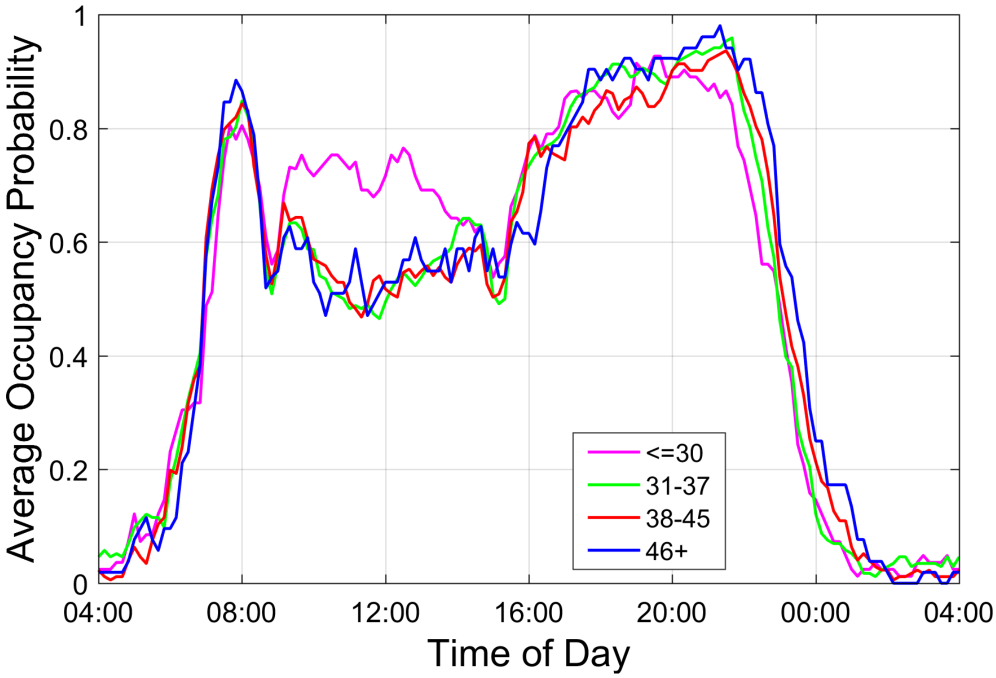
- Average timing of waking and sleeping are very similar on an average basis. Distinct differences only become evident when differentiated by diary day working status.
- Age remains a key differentiator for multi-adult households. For example, average active occupancy for a retired couple household is the highest of all identified types.
- Family households with school age children on school days have a short period of high occupancy probability in the early morning period.
- Family households show an earlier and more consistent late afternoon occupancy increase than non-child households with similarly aged adult occupants.

More detailed analysis highlighted that the significant differences detailed for single-person households above, driven by age and employment, are also shown for the other multi-person household types. This analysis was focused on basic active occupancy (i.e. at least one person present and awake) rather than number of occupants. The main conclusions, based on a minimum of 100 couples/parent-pairs per identified population, were as follows:

- Figure 2.17(a) shows that age remains a key factor, particularly for households without children. (Data is based on 2-person households on non-working days to remove employment-specific influences.)
- Figure 2.17(b) shows the significant influence of different working combinations on couple and parent combined occupancy, with markedly lower occupancy where both work on diary day.
- Figure 2.18 shows that youngest child age is a more significant determinant of parent occupancy than parent average age, particularly in the daytime period (based on family households with one non-working parent to reduce multiple occupant effects). Youngest child age was also shown to be a better differentiator than other potential factors, including youngest and oldest parent age, and oldest child age.



(a) Youngest child age



(b) Parent average age

Figure 2.18. Combined average active occupancy profiles for parents by two age parameters for days with one parent working. Data from the UK 2000 TUS dataset [83].

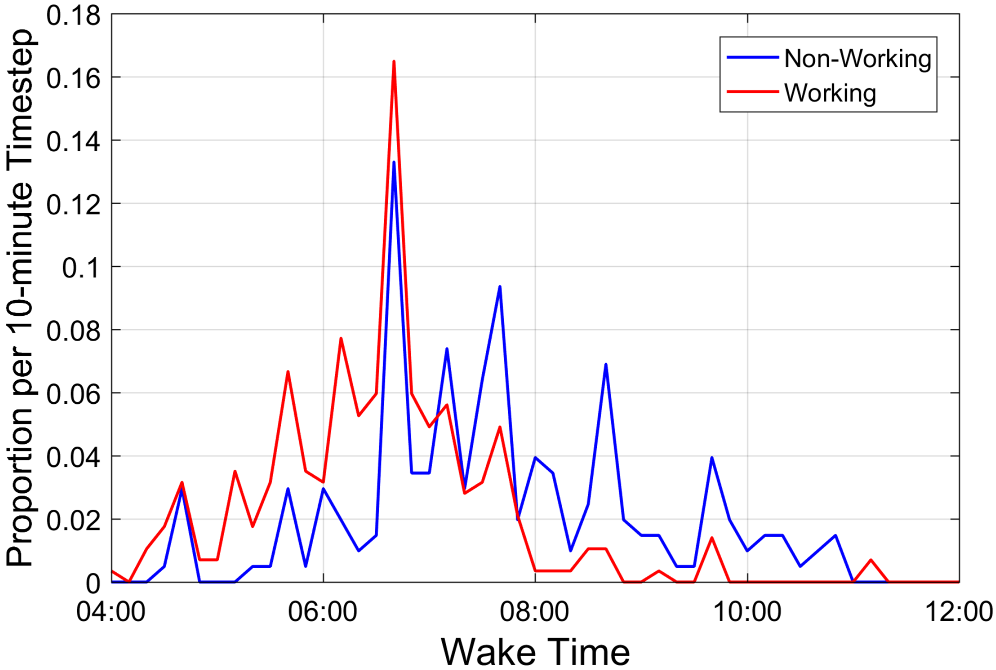
There are two main overall conclusions from the analysis of the relative influence of occupancy-determining characteristics. One is that there are three primary differentiating occupancy profile characteristics that are evident at the household composition type level and have the potential to influence demand patterns significantly; daytime and evening occupancy probability, and the timing of the late afternoon occupancy increase. The other is that distinct differences in occupancy related to age and employment characteristics are at least as significant as differences between the basic household composition types (single-person, couple, family, multi-adult). Therefore, multiple differentiating occupant characteristics will be required for more realistic modelling of occupancy variations between households. The development of an occupancy model with significant differentiation by occupant and day types is detailed in Chapter 4.

2.5.2.3 Household-Specific Occupancy Variations

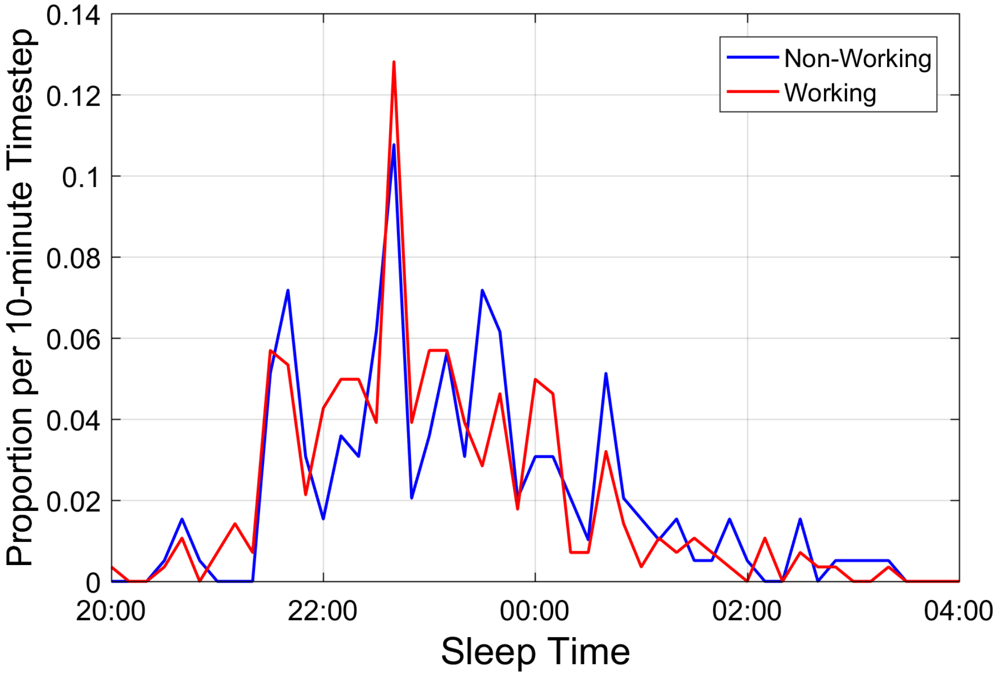
Analysis of individual diaries within distinct populations (based on household composition type, age profile, employment, and day type) shows that there remains significant in-group variability. Figure 2.19 shows the distribution of the identified waking and sleep timing for each individual in the 28-56 years old age range, single-person household working and non-working populations on weekdays [83]. Similar results are seen for all equivalently differentiated populations.

The European standard practice of single day diaries for time-use surveys [85], followed by the UK 2000 TUS survey used as the primary reference for this project, significantly restricts both the calibration and validation of multi-day occupancy models as it is impossible to discern occupancy consistency over time. One open-source exception is the Dutch 2005 TBO TUS dataset [87] which has one week diaries. Whilst specific occupancy characteristics may differ from the UK population, it is assumed that the dataset can at least provide information on the relative occupancy variations between identical occupant types and consistency for individual occupants, that is broadly applicable. Unfortunately, this dataset does not include multiple people within the same household, which does not allow the consistency of occupancy interactions within households to be analysed.

Analysis of this dataset for both mean and standard deviation of wake time, sleep time, and sleep duration, for both working and non-working days for individuals with



(a) Wake time



(b) Sleep time

Figure 2.19. Wake and sleep time distributions for single-person households in the 28-56 years old age range on working and non-working weekdays. Data for analysis from the UK 2000 TUS dataset [83].

at least four qualifying entries, shows significant variation in both measures. There is a wide range of behaviours from early to late and consistent to erratic (from low to high standard deviation).

The standard deviations for the wake and sleep times shown in Figure 2.19 for the UK single-person population groups were 75 and 64 minutes for the working population and 81 and 83 minutes for the non-working population. The equivalent Dutch 2005 TBO TUS population data (see Figure 2.20) shows that the variance for individuals is typically lower, much so in the case of working people with an average wake and sleep time standard deviation of 28.0 and 20.6 minutes respectively. Therefore, calibration data based on population behaviour remains significantly more variable than would be expected for individuals.

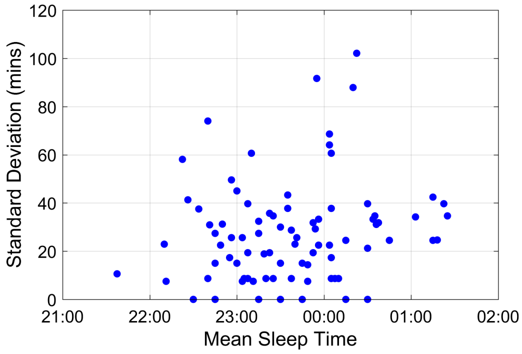
Sleep and waking times were chosen for analysis as they typically occur daily for most people and within a clearly defined time range. Similar analysis for other key, but less time-specific, occupancy transitions, such as 'first leave' and 'last return' timing, show similar, if less distinct, characteristics.

The conclusion that can be drawn is that while there are broad household composition type characteristics for occupancy, individual variations both between and within households are potentially significant for either individual household or smaller district demand analysis. It is necessary to capture this influence to meet the stated aims of the developed demand model. The specific influence of individual occupancy behaviours, and modelling methods to capture the within-type variation, are reviewed in Chapter 7.

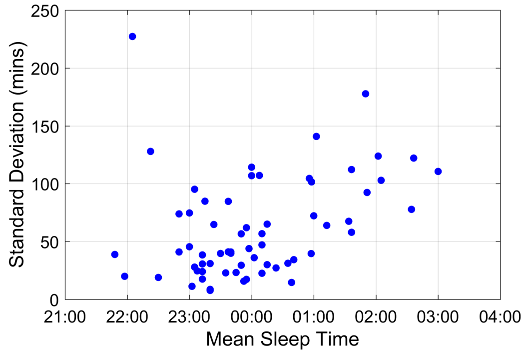
2.6 Occupant Behaviour Uncertainty and Demand

2.6.1 Background

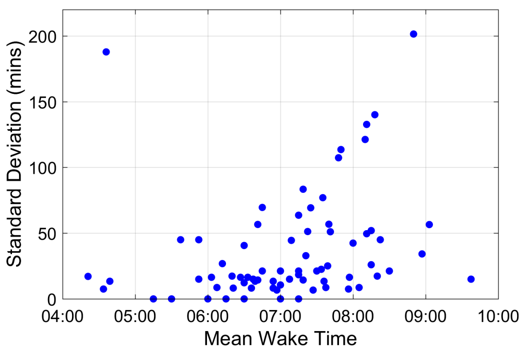
Uncertainty analysis in building simulation is a broad subject area that addresses potential modelling inaccuracy in different areas. Uncertainty analysis was originally defined as the potential variation in model output that results from expected variation in input parameters [124]. Input parameter variations have a variety of sources, from errors in physical properties to the accuracy of the algorithm used. More recently, the



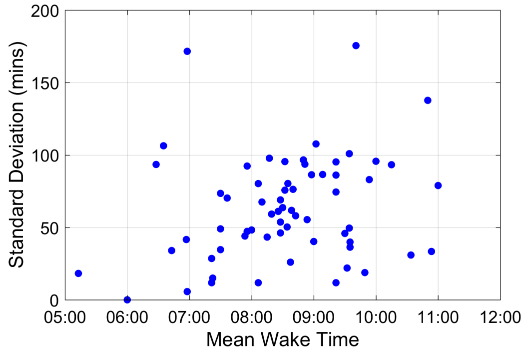
(a) Sleep time - working day



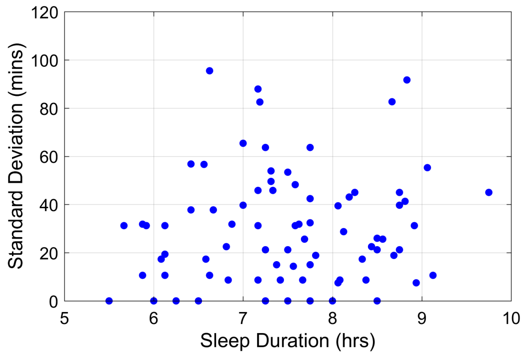
(b) Sleep time - non-working day



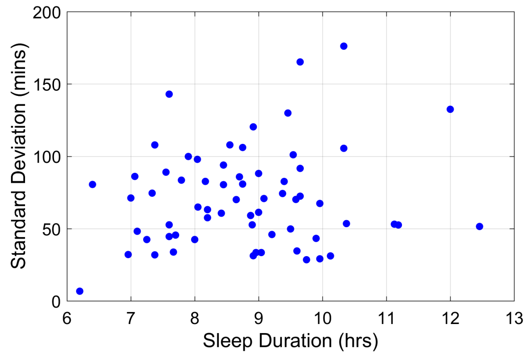
(c) Wake time - working day



(d) Wake time - non-working day



(e) Sleep duration - working day



(f) Sleep duration - non-working day

Figure 2.20. Mean and standard deviation of sleep-related timing parameters for the Dutch 2005 TBO TUS 28-56 years old age range population on working and non-working days. Data for analysis from [87].

term has been used to address many aspects of modelling output inaccuracy, including the impact of user behaviour ([125], [126], [127]).

For models that seek to capture the influence of occupants, such as demand modelling, user behavioural variation is a significant source of output uncertainty and is, arguably, the least well understood and simulated. The user influence on building energy use is significant, and current modelling methods, in general, poorly address the complex relationship between occupants and energy demand. The degree to which this uncertainty needs to be considered and captured probabilistically is model-specific, dependent on both scale (e.g. national, district, household) and time resolution.

As addressed by Mahdevi and Tahmesbi [128], for occupancy modelling as an input to building simulation models, there is no single modelling method that is applicable for all scales and resolutions. They determined that at a low time resolution (e.g. monthly or annual demand assessment), non-probabilistic approaches can be appropriate and model accuracy is more strongly associated with the baseline calibration accuracy than probabilistic variations. At higher time resolutions (e.g. hourly or less) using temporally differentiated models, probabilistic approaches become more appropriate. By extension it would also be expected that the same logic applies to all aspects of demand modelling, not merely occupancy influences. Decreasing the system scale and increasing the time resolution of analysis would require an increase in the degree to which the stochastic nature of specific energy consumption behaviours is captured.

Leytens and Kurvers [129] introduced the concept of robustness in building design. A robust building design is defined as one that has low sensitivity to errors in design assumptions and which meets the overall operating criteria under the full range of conditions. Hoes et al [130] extended this principle to include expected variability in user behaviour and changes that can be expected over time (e.g. cultural changes, household turnover etc.), and anticipate that the integration of more detailed user behaviour models into the design process will aid this.

In relation to demand prediction for constrained and small-scale energy schemes, a robust analysis would be one that considers the full range of potential demand scenarios, with the predicted range capturing both the stochastic impact of behavioural uncertainty and predictable variations resulting from household characteristics. Similarly, a robust system design would be one that can accommodate the predicted range of scenarios. Currently this has not been addressed in a way that can directly influence the design process.

2.6.2 Demand Modelling Uncertainty

As outlined in 2.2, while a degree of household demand can be predicted from known household characteristics, there is a residual uncertainty that results from unknown household characteristics and from behavioural variation that is either independent or only weakly correlated with characteristics.

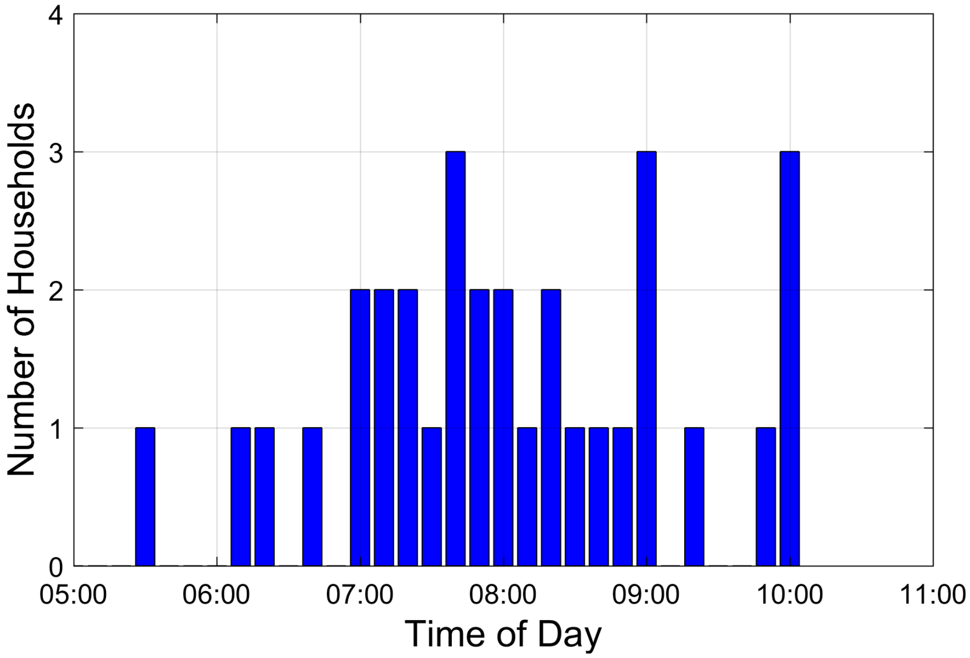
There are two distinct elements of occupant behaviour uncertainty for high time resolution models. The first ('Baseline Behaviour'), addresses differences in average energy use behaviour between households with the same characteristics, as driven by both overall attitudes towards energy use (i.e. climate concerns, spending prioritisation) and appliance-specific behaviours. The second ('Temporal Behaviour'), addresses variations in occupancy and appliance use behaviours in relation to the timing and the degree to which behaviour within households is consistent day-to-day. In essence, the first addresses *if* energy is likely to used and the second *when*.

The potential impact of unknown characteristics and these two identified behavioural elements are discussed below.

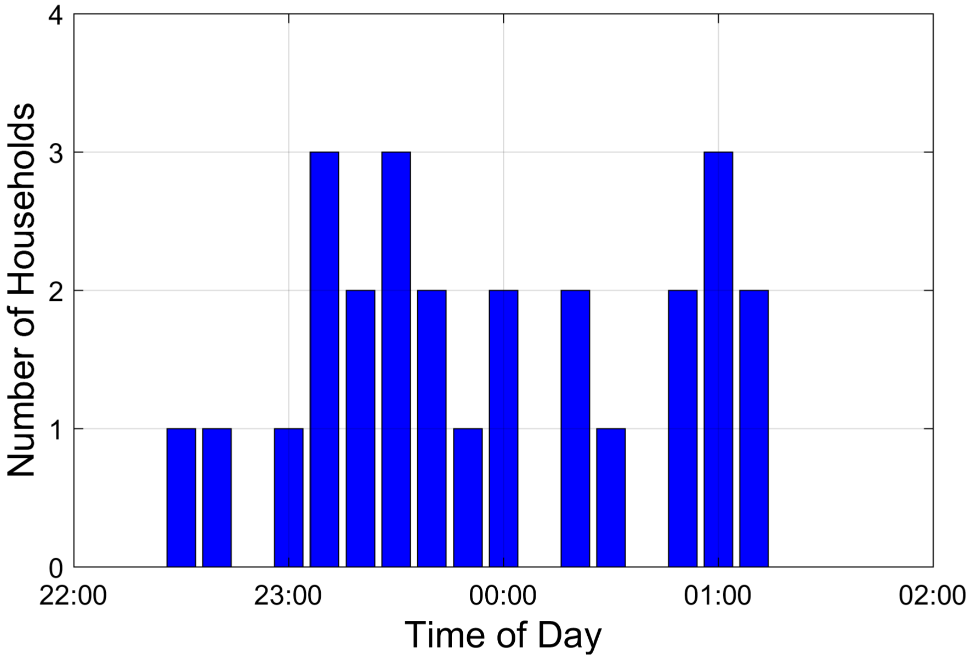
2.6.2.1 Unknown Characteristics

That typical energy use per household is determined by a complex sequence of interdependent household characteristic relationships was introduced in 2.4. The analysis demonstrated that for each element that is unknown, there is a probabilistic relationship with the known characteristics. Therefore, the degree of uncertainty between known and actual household characteristics, and by extension the ability to predict baseline energy use for the household, is determined by the number of unknown elements.

As an example, the design of an energy system for a new-build housing scheme will have a lower level of known household characteristics (typically only location, tenure, house type and size) compared to a system design for existing households where significant details are available for current residents. Robust system design should be able to account for this uncertainty.



(a) First timestep



(b) Last timestep

Figure 2.21. First and last 10-minute timestep at which average timestep electricity demand exceeds the overall household average demand for working age, single-person households. Data for analysis from the HES dataset [89].

2.6.2.2 'Baseline' Behaviour Uncertainty

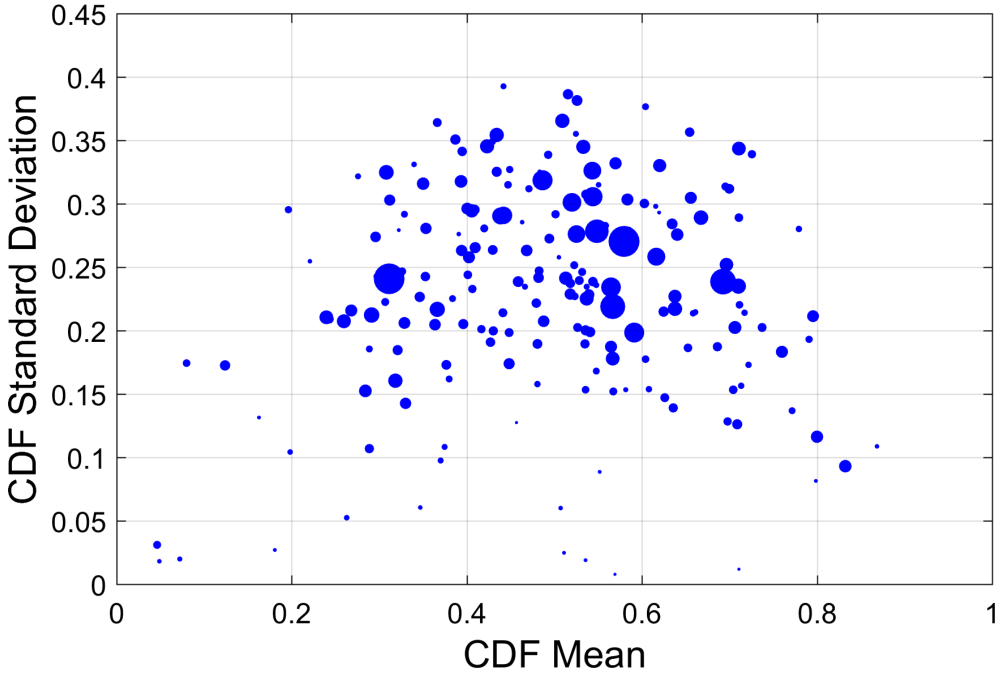
Existing research has demonstrated that only a proportion of the differences in energy demand between households can be attributed to household characteristics and occupancy. As outlined in 2.2, Gill et al [81] evaluated the residual variation, that can be attributed to attitudes to energy use, energy efficiency, environmental concerns, and spending prioritisation, to be 37% for electricity, 51% for heating, and 11% for hot water.

In addition to this overall behavioural element, there is also appliance-level use variation within specific household composition type groups. Analysis of the HES dataset has shown that there is no clear overall correlation between the relative use of each owned appliance per household, beyond that which would be predicted by the Gill et al overall use variation. This suggests that specific demand use frequency cannot be easily characterised by overall household energy use characteristics.

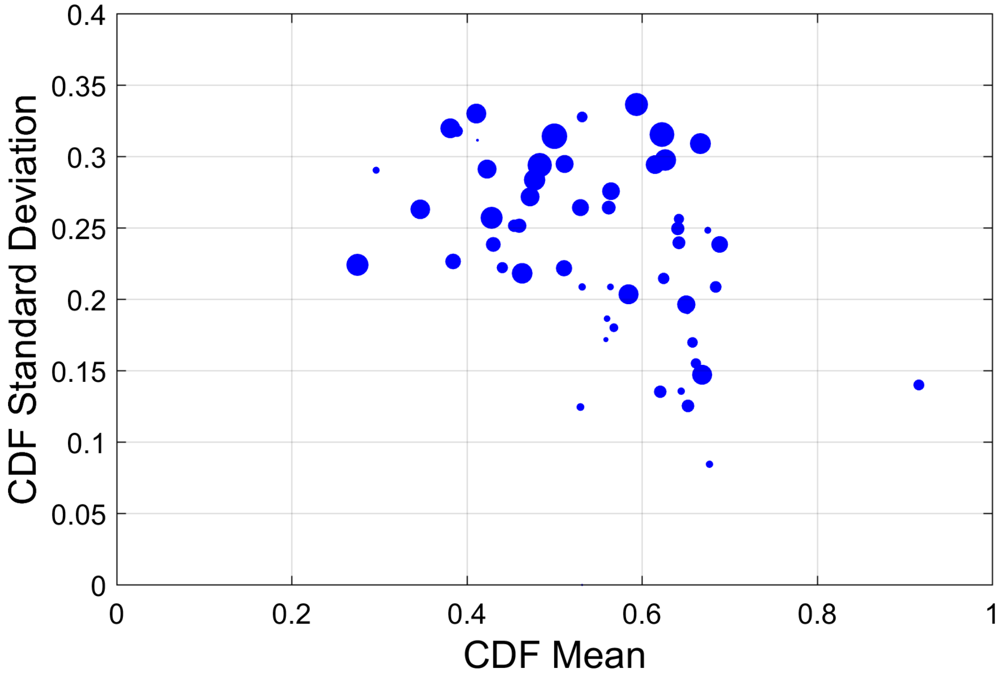
2.6.2.3 'Temporal' Behaviour Uncertainty

Comparison of the average intra-day demand profiles for individual households relative to the household type equivalent average profile allows the relative timing of the key demand transition periods (i.e. waking, evening return, sleep) and peak use periods to be captured. Significant variation is seen for households with similar characteristics. For example, Figure 2.21 shows the distribution of times when the average 10-minute timestep electricity demand for each working age single-person household exceeds the overall household average for the first and last time in a 24-hour period from 4am. The variation is significant and similar to that shown for wake and sleep times in Figure 2.20. Whilst the lack of comparative occupancy and demand data does not allow a direct link to be inferred, the similarity indicates at least a partial link between the observed occupancy variations outlined in 2.5 and similar variation in key demand transitions.

For specific demands, significant behavioural differences in terms of use timing are observed for each household. For some, use timing is habitual with little day-to-day variation, others use appliances typically earlier or later than average but with a degree of variation, and for the remainder use is effectively random. This can be demonstrated by the mean and standard deviation of the timing of washing machine use per HES



(a) Washing machine



(b) >30 litre hot water cycles

Figure 2.22. Variation in mean and standard deviation of per-household cycle start time cdf equivalent values for washing machines and >30 litre hot water cycles. Data for analysis from the HES dataset [89].

dataset household shown in Figure 2.22, where a significant variation in use behaviour is observed (for further definition of the relationship between the *cdf* value and timing see 5.9.1). Significant variation in the mean from 0.5 shows use that is earlier or later than average, and the lower the standard deviation the more consistent, and therefore habitual, the timing of use. The size of the circle is proportional to the number of use events in the HES dataset (maximum=173), and indicates that significant variation remains even for households with a high number of monitored events.

The overall conclusion that can be drawn is that the time dependency of individual household demand is a complex combination of occupancy and usage behaviours, although, as stated, the relative impact is difficult to assess due to the lack of comparative occupancy and demand data. The analysis strongly suggests that a bottom-up approach, capturing specific demand behaviours and their probabilistic distribution across all households, should be considered for increasingly realistic, high-resolution demand modelling.

2.6.3 Implications of Uncertainty for Small-Scale Modelling

Analysis of occupancy and energy demand for individual occupants and households demonstrates that at a high level of detail each is unique. However, distinct patterns exist that apply at the household composition type level and broader patterns can be observed for the majority of households; low nighttime use, morning increase on waking, daytime use linked to occupancy probability, evening increase, tapering demand until retiring.

Calibration data, programming, and computational speed constraints, determine to what degree any energy demand model captures specific behaviours or is a composite of averaged behaviours. A degree of behavioural averaging in any model used to predict occupancy and demand is not necessarily of limited use [128], particularly when applied to integrated energy systems with multiple connected households.

Demand prediction for small districts should therefore permit for a degree of behaviour averaging, with the tolerance increasing as the size increases. Even for individual household analysis, this type of model output will provide useful baseline demand profiles, which can be used for feasibility or typical performance analysis. However, it is not currently well understood how accurately existing modelling methods represent reality, how this varies for different scales and types of analysis, and whether results are sufficiently accurate and comprehensive to ensure a robust energy system design. This is reviewed in Chapter 3.

2.7 Characteristics, Occupancy, and Demand Analysis Summary

The analysis presented in 2.4 and 2.5 has shown that there are distinct occupancy and demand patterns associated with different types of occupants and households. The conclusion that can be drawn from this is that, as a minimum, significant household composition differentiation is required for accurate modelling of occupancy and demand.

A number of household characteristics have been identified that have the potential to either directly or indirectly influence both total and time-dependent demand. In addition, the characteristics have been assessed for the degree to which they are interrelated and also likely to be known. The degree of demand prediction uncertainty is therefore dependent on the level of known information compared to that which must be predicted.

What is currently not well understood is at what scale of system detailed knowledge of the household characteristics is sufficient to allow deterministic demand prediction and how significant the influence of individual household behaviours (see 2.6) becomes as the scale is reduced. At the individual household level, confirmation is required of the assessment of Haldi and Robinson [79] that the level of behavioural variation equals or exceeds a factor of two, and the implications this has for small-scale system design determined, including for individual household energy systems.

The link between occupancy and demand has been clearly demonstrated by the work presented and by others. Prediction of occupancy and identification of occupancy variations between households is therefore a necessary precursor to improve the accuracy of demand modelling.

2.8 Chapter Summary

This chapter reviewed the influence of different household characteristics on other key characteristics and ultimately on occupancy and demand. This review allowed the most effective sequence of calculation steps required for a determination of probabilistic household characteristics to be identified. The impact of household behavioural variations that cannot be directly predicted from household characteristics was also explored. The chapter highlights are as follows:

- In order of decreasing likelihood of being known and increasing direct influence on energy demand, the following were identified as the key household characteristics for determining energy demand: location; house tenure, type, and size; household composition and age profile; employment and education status; income; floor area; appliance ownership; appliance energy ratings and power; occupant behaviour; occupancy and time-use; and use frequency.

- Occupancy was shown to be a key determinant for both overall and the time dependency of demand. Occupant type, age, and day-specific working hours, were shown to strongly influence occupancy behaviour.

- Whilst distinct average occupancy patterns can be shown for different occupant and day types, significant additional variation in the timing of key occupancy transitions (i.e. waking, sleep, first leave, last return) is also evident within each type group.

- Uncertainty in demand modelling results from variations in total energy use, and from the timing of individual uses between different households, that cannot be predicted from knowledge of household characteristics. Additional uncertainty stems from situations where household characteristics are unknown, requiring a probabilistic assessment of the unknown characteristics. Robust small-scale energy system design requires that all these identified uncertainties are accounted for in the system design process.

Chapter 3

Domestic Occupancy and Demand Modelling Methods

3.1 Chapter Overview

In Chapter 2, it was shown that occupancy was a key determinant of household energy consumption, and, in particular, its timing. Household type, occupant age(s), and working hours were determined to be the most significant influences on relative occupancy. The following chapter includes a critical assessment of current occupancy and demand modelling methods and potential areas for improvement for high-resolution modelling.

For occupancy modelling, existing models are primarily based on the first-order Markov chain method. Limitations of these existing models are shown to include, (1) weaker occupancy state duration prediction as an inherent feature of first-order models, (2) inaccurate replication of occupancy interactions between couples, and parents and children, using models that treat each occupant independently, (3) little of no differentiation based on age, employment or day type to account for distinct differences in occupancy behaviour, and (4) over-use of time-use activities within the primary occupancy model, which are shown to be both weak predictors of appliance use and therefore energy demand, and a significant limitation to additional occupant differentiation as a result of the increased calibration data required.

The review of existing, high-resolution, bottom-up demand models, identified as the type most relevant to the stated aims of the project, identified that virtually all split electrical appliances into various sub-groups with distinct modelling methods for each group. The main conclusions drawn were, in addition to the weak correlation between time-use survey activities and demand demonstrated in this chapter, that insufficient factoring was incorporated to allow for overall and appliance-specific variations in use between similar households, and that the per-timestep probability methods used by the majority of existing demand models generate unrealistic use patterns.

3.2 Basic Model Types

Energy demand models can be characterised by several descriptive, and often interrelated, binary terms; top-down or bottom-up, deterministic or probabilistic, system or agent-based, time or event-based, high or low resolution. Given the complexity of detailed energy models, an overall model may not be purely one type but located on a continuum between both options.

The following section describes each type in relation to domestic demand modelling:

- ***Top-Down/Bottom-Up*** - 'Top-Down' models take the final target variable for the model, and use analytical methods, such as regression, to determine the individual factors that influence the final variable and the degree to which they do so. 'Bottom-Up' models aggregate results from multiple sub-models of individual elements to determine the overall result. The elements can range from individual households to specific demands in each household. 'Bottom-Up' models typically allow for higher time resolution, greater household differentiation, and a more refined assessment of the probabilistic variation, but have greater calibration data requirements and computational complexity.

- ***Deterministic/Probabilistic*** - 'Deterministic' models assume that the action or output can be directly and fully predicted by the input conditions. 'Probabilistic' models reflect the likelihood of a particular action or output based on the same input conditions. The likelihood can be a fixed probability, have time dependency, or be based on the preceding state(s) using techniques such as Markov chains.

  Human behaviour is inherently probabilistic, with tendencies towards different, distinct behaviours within populations evident [131]. While probabilistic models are most effective for capturing detailed occupant behaviours, they are significantly more complex to develop and calibrate. Computing speed, data, and development time constraints potentially limit the ability of probabilistic models to capture individual occupant or household behaviours at a high time resolution. For example, calibration data for the probability model may be based on the composite behaviour of a large, undifferentiated population, resulting in an output that reflects that average behaviour of the overall group and not distinct behaviours within the population.

- ***System/Agent*** - There is no accepted general definition for system and agent models, but in this case it defines the extent to which individual behaviours are identified. 'System' models attempt to define complex systems using overarching rules or equations, and are therefore typically top-down and deterministic in nature. 'Agent' models are a specific example of a bottom-up model, where individual actors are modelled separately with unique behaviours and aggregated to represent the overall population. They vary by the degree of agent differentiation and accuracy of the individual behaviours modelled. 'Agents' can be modelled using individual behaviour calibration, but, as a result of data and computational limitations, more typically by composites of multiple individual behaviours, with each composite group differentiated by agent type.

- ***Time/Event*** - Discrete-time models step sequentially by the model-defined timestep to determine if a change of state occurs. Discrete-event models determine, as a single calculation, either the timing of each state change directly or duration between such events sequentially, and the type of change if there are multiple options [132]. The decision whether to use discrete-time or -event models is dependent on several factors [133]. Event models are typically more computationally efficient, particularly for models with significant periods of no change. Time-based models are typically simpler to understand and calibrate, and allow any time-dependency to be captured explicitly.

- ***High/Low Resolution*** - Existing demand models vary in the resolution of the results. The lowest resolution predict demand at the monthly to annual level. These are typically used for housing stock assessments, and include BREDEM in the UK [111]. At the opposite extreme, models exist that aim to capture sub-hourly variations in demand, with the highest resolution being in the 1-5 minute range. These are typically used to model elements such as peak demand or supply and demand matching, where the resolution is critical to capture natural variability [73].

In general, however, models can be split into two broad types: 'baseline' models, which have a resolution of greater than a day, focus on average demand, and do not consider time-of-day influences, and which are typically top-down, deterministic, time-based, system-based, and, by definition, low-resolution; and 'temporal' models, which also aim to account for inter-day variations at least at the hourly level, and which are typically bottom-up, probabilistic, either time- or event-based, agent-based, and high-resolution.

The aim of this project to account for household behaviour-driven demand variations at a resolution applicable for constrained and small-scale system analysis, strongly suggested the use or development of 'temporal'-type methods. The focus for the remainder of the review of existing models is therefore this type.

3.3 Existing Occupancy and Activity Models

Energy demand models can also be defined by how the occupant presence influence is captured. Occupancy variation can either be inferred by factoring within the demand model or can be modelled as a separate sub-model and the results integrated with the demand model. The existing 'temporal' type models typically model occupancy separately, with each 'agent', either occupant or overall household, modelled separately.

For individual 'agent' occupancy models, a variety of statistical techniques have been developed to translate occupancy data (typically from time-use surveys) to probability data that is used for model calibration. These can be broadly categorised as being either discrete-time or -event based, and then by the time resolution and degree of agent differentiation.

The extensive use of time-use data, which includes significant detail on the specific activities being undertaken, allows for both modelling of basic occupancy states or more detailed modelling of specific activities using the same methods. Models therefore vary by the level of detail to which occupancy is defined from simple occupied/unoccupied

State	1	2	..	*i*	..	*n*
1	$p_{1\to 1}$	$p_{1\to 2}$	..	$p_{1\to i}$	..	$p_{1\to n}$
2	$p_{2\to 1}$	$p_{2\to 2}$	..	$p_{2\to i}$	..	$p_{2\to n}$
..	..	..	..	..	..	..
i	$p_{i\to 1}$	$p_{i\to 2}$	..	$p_{i\to i}$	..	$p_{i\to n}$
..	..	..	..	..	..	..
n	$p_{n\to 1}$	$p_{n\to 2}$	..	$p_{n\to i}$	..	$p_{n\to n}$

Figure 3.1. Discrete-time Markov chain transition probability matrix (TPM) structure.

models to tracking multiple occupancy-related activities (i.e. sleeping, eating, cooking, laundry etc. plus absence).

3.3.1 Discrete-Time Based Methods

A major subset of recent high-resolution occupancy and activity models, all calibrated using time-use data ([134], [135], [70], [136], [137], and [138]), use the Discrete-Time Markov chain (DTMC) method to produce sequences of occupancy-related states. The DTMC approach is a commonly used method for modelling stochastic processes [139].

DTMC models allow the occupancy state at a timestep, t, to be determined based on the state at the previous timestep, t-Δt. The basis for any DTMC model is transition matrices (see Figure 3.1). These hold the probability of transition from one state a to another state b ($\mathrm{p}_{a \to b}$). The size of this matrix is determined by the number of independent states to be modelled. For a model with n states, an n x n matrix is required. A row in this matrix therefore contains the probabilities of a transition from some state i to all n possible states (including no change from state i) and all entries per row should sum to 1.

To calculate a sequence of states over several timesteps, a random number R between 0 and 1 is generated for each modelled timestep and the new state is determined by systematically comparing the generated random number with the cumulative probabilities in the appropriate row i of the matrix. For example, if a state i persists at timestep t-Δt then k, the next state at time t, is the first cumulative probability $\sum_{j=1}^{j=k} p_{i \to j}$ that exceeds R.

For a first-order DTMC model, only the state at the preceding timestep is considered. A second-order model considers the two preceding states, third-order three states, etc. Higher-order models consider the duration of the existing state at each modelled timestep. First-order models assume that the system being modelled has the Markov property, which requires that the system state at time t+Δt is only dependent on the state at time t and not the sequence of preceding states or the duration of the current state. All existing DTMC models are first-order.

Richardson et al [134] developed a two-state (active-inactive) model to define the number of occupants present using separate modules differentiated by total number of occupants per household. Widen et al [135] first used a DTMC method to define a three-state (active-inactive-absent) occupancy model for lighting use, and subsequently Widen and Wackelgard [70] refined the process to incorporate nine occupant activities (absent, sleeping, six core time-use survey (TUS) activities plus 'Other'). Both developed models simulated occupancy patterns for individuals, which were then combined for multi-person households with no interaction assumed. None of the three models differentiated between different types of occupant.

Others have used variations on the same first-order DTMC method. These include: Muratori et al [136] using a method similar to Widen and Wackelgard [70] with 9 activity and location related states, with five archetypal occupants considered (working/non-working male and female plus children); Meidani and Ghanem [137] reviewed the multi-activity Markov method performance, including incorporation of additional random transition factoring; Collin et al [140] developed an eleven activity model similar to [70], which included a method for identification of shared activities.

A variation of the basic Markov chain method was proposed by Baptista et al [138] to incorporate interaction effects between individuals in a household. This approach simplifies the general interactive Markov chain approach proposed by Conlisk [141], which determines individual transition probabilities from the distribution of all agent states. This simplification is achieved by randomly fixing one individual as the 'leader' agent, using a standard first-order Markov chain model for that individual, and then determining appropriate first-order transition matrices of other agents based on the determined state of the 'leader' agent.

Examples of non-Markov chain discrete-time models are rare. One example is from Baptista et al [142] using a 'Nearest Neighbour' approach. This aims to find the closest match in the TUS data to the current modelled activity profile of the household and then to determine the next transition based on the observed transition from the closest TUS match. Whilst the method was shown to have similar performance to the first-order DTMC approach, it is also significantly more complex to calibrate and computationally intensive to model.

3.3.2 Discrete-Event Based Methods

The first example of a high-resolution model using time-use data was developed by Tanimoto [68]. Rather than a sequential model, activities are placed within the day based on the probability of an activity taking place at a particular timestep, and with a duration based on the mean and standard deviation of the actual durations from the time-use data converted to a logarithmic Gaussian distribution. Thirty-three distinct activities were identified, which is a level of detail that would not be practical with a Markov chain model.

This method is limited by the need to define each individual activity and therefore does not effectively replicate activities which happen either multiple times in a day or with a daily probability of significantly less than one. It also needs to be run multiple times to achieve one realistic profile, as results which are not between 23 and 25 hours in total duration are discarded. Further analysis and development of this basic method by Yamaguchi et al [143] compared it with the Markov chain method and concluded that, while it could be useful in situations where data was too limited for Markov chains to be used and could be set up to stop unrealistic activity repetitions, it was not as accurate as the DTMC method for either number of transitions or activity durations. The number of transitions is higher than for the measured data confirming that forcing daily occurrences of behaviours leads to unrealistic models.

Wilke et al [71] used a similar approach to Tanimoto but using a sequential method that identifies the type of state transition and duration of the subsequent state probabilistically. The Markov property assumption is incorporated for the state transition as each subsequent state is only dependent on the preceding state, although in this case a higher-order basis for state duration prediction is achieved by direct probabilistic determination at each transition. The primary motivation for this approach was to improve on the ability of first-order Markov chain methods to capture the duration distribution of each occupancy-related activity state. An additional benefit is that is it significantly more computationally efficient, with two calculations per transition (new state plus duration) required rather than a single calculation per timestep.

Wilke et al generated probability data for 24 distinct occupant parameters by using differentiators for age, day, level of education, income, employment/education status, gender, location (urban/rural), etc. The model then uses a statistical approach (multinomial logit) to combine the relevant parameters for each modelled individual to capture the combined influence on occupancy. Twenty different occupancy-related states and activities were identified in the occupancy model. The model was shown to perform well for simulation of the overall population but with some discrepancies for analysed sub-populations. The authors suggested that this may have been caused by an excessive number of modelled activities or incorporating too many differentiating parameters.

This method has also been used by Aerts et al [144] with differentiation based on different clusters of similar behaviours. This method is restricted by the need to identify clusters from highly variable data with the inherent level of subjectivity involved, and that it is not easy to determine from existing single-day or one-week diary data the sequences of day types that are appropriate for specific individuals or occupant types.

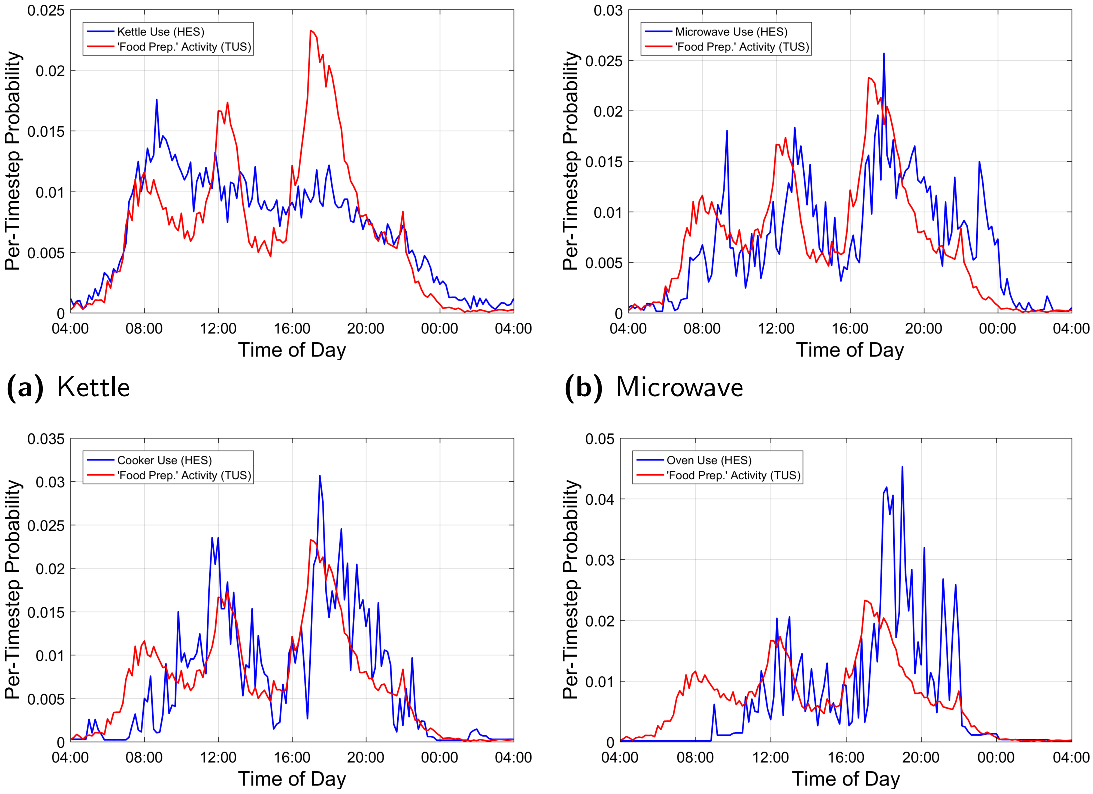
3.3.3 Existing Occupancy Model Evaluation

Validation performed on existing discrete-time and -event based methods have shown that they perform well in achieving occupancy and activity sequences that simulate the average behaviours of each calibration population used. It is less clear from the presented analysis how well the developed methods represent individual occupants and households, and the within calibration population variation.

For all the methods outlined, there are a variety of model structure decisions that were taken, either implicitly or explicitly, including; number of occupancy or activity states, occupant differentiation, state sequence or duration 'memory', and occupant interactions, and, in addition, inherent limitations that need to be either accepted or accounted for. The following section reviews each element in turn.

3.3.3.1 Number of Occupancy States / TUS Activity Model Evaluation

The activity level detail in time-use surveys has resulted in models that vary from simple two-state occupancy models [69] to 20+ occupancy and activity state models ([68], [71]). There has been little discussion regarding the most appropriate number or selection of states, particularly when the goal is household energy demand or heat gain



(a) Kettle

(b) Microwave

(c) Cooker

(d) Oven

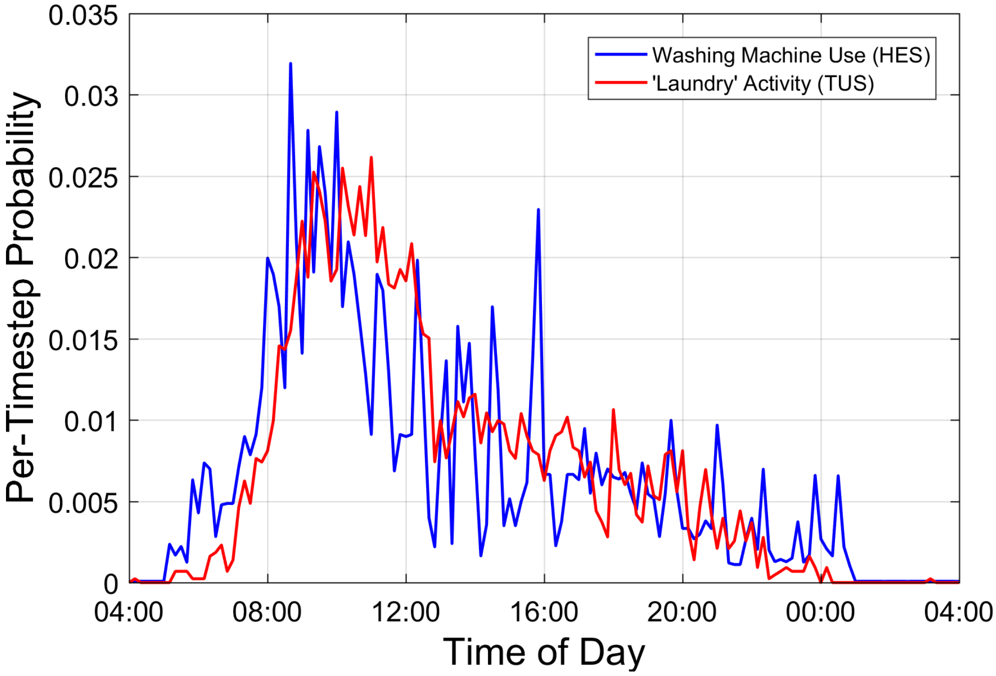
Figure 3.2. Average per-timestep use probability for various 'cooking' appliance in comparison with the UK 2000 TUS dataset [83] 'Food Prep.' activity probability. Electrical demand data for analysis from [89].

prediction. For demand prediction, modelling of states can only be useful where there is a good correlation between an activity and energy demand. Torriti [51] states that this is dependent on "how appliance-specific the diary entry is", with clear inferences possible for the 'TV' activity but much less so for 'Household Upkeep', for example. For Markov chain models there is also a limit on the number of states (n) that are practical for the creation of n^2 time-dependent transition matrices with sufficient depth of calibration data for effective modelling.

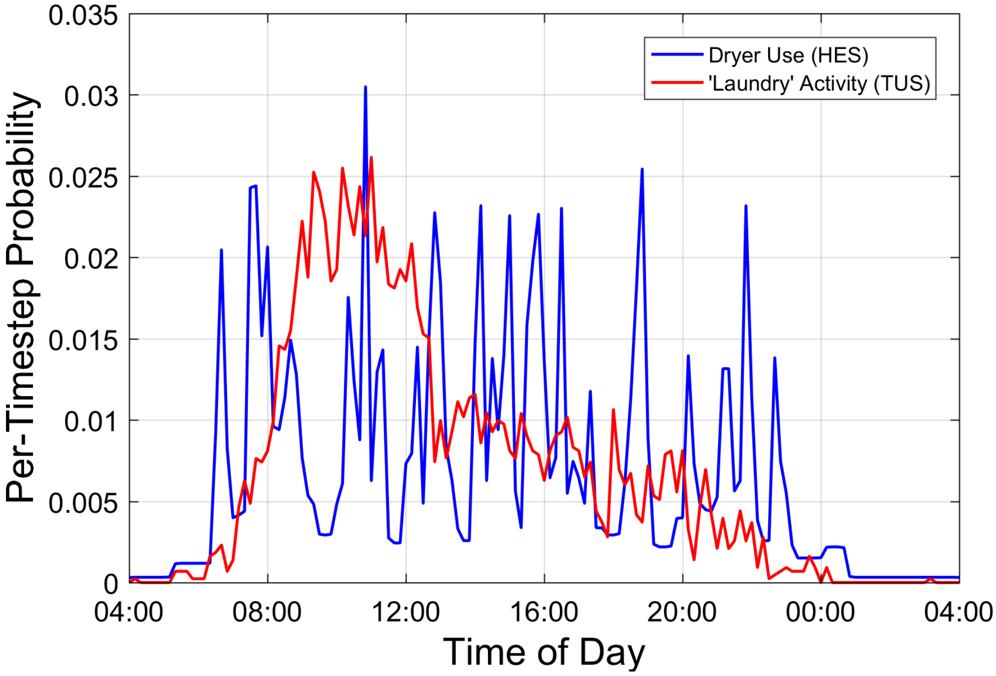
The availability of detailed appliance-level demand data [89] allows the relationship between TUS activities and appliance use to be determined, and, if weak, potentially allows the occupancy sub-model to be restricted to basic occupancy states, with associated demand determined from a separate sub-model directly calibrated from the demand data. For the key activities of 'Food Prep.', 'Laundry' and 'TV', the correlation between TUS Activity and appliance demand was investigated. The analysis assumes that the HES dataset is sufficiently large and representative to capture typical UK appliance use timing with good accuracy.

Figure 3.2 shows the comparison between the 'Food Prep.' activity from the UK 2000 TUS dataset [83] and the start-time distributions for four relevant appliances for all 1-person households. Whilst there is some consistency between activity and start time, particularly for the mid-day and early evening peak periods for both microwaves and cookers, there are significant discrepancies resulting in weak overall correlation at the appliance-level. For example, kettle use is generally poorly represented with an underestimation in the morning period and no clear correlation between use and the main meal-times captured by the TUS activity. For both microwave and cooker use, the potential for use is overestimated in the morning period and underestimated in the mid-evening period. Oven use is primarily in the early evening period, a behaviour that is not captured by the activity.

For cooker and oven use, in particular, it would be expected that the activity peaks would be later than the start time peaks if there was a strong correlation as the cooking activity has the potential to be an extended activity with appliance starts occurring early in the activity. For both appliances in the evening period the opposite occurs suggesting the relationship between the activity and the use of the appliance is more complex. This discrepancy and the general lag between the activity and microwave use



(a) Washing Machine



(b) Dryer

Figure 3.3. Average per-timestep use probability for laundry appliances in comparison with the UK 2000 TUS dataset [83] 'Laundry' activity probability. Electrical demand data for analysis from [89].

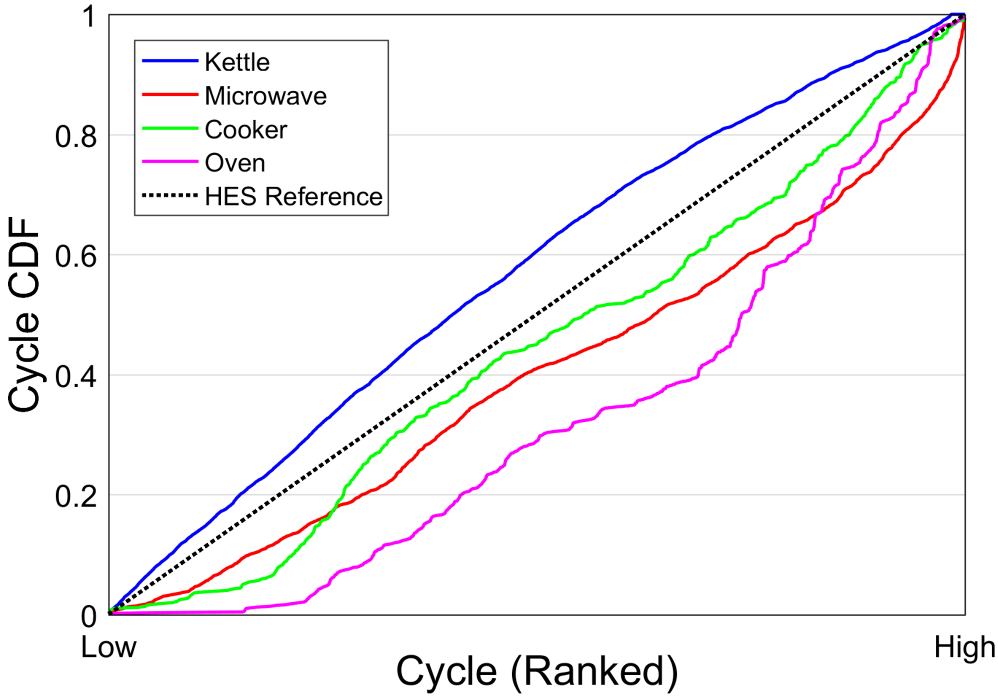
suggests that the 'Food Prep' activity includes significant non-appliance use elements, such as ingredient preparation, that have the potential to distort the results if a direct relationship between activity and appliance use is assumed. Except for kettle use, the activity-based model predicts earlier appliance use than seen in the HES dataset (see Figure 3.4 below), suggesting that the probability that the broadly defined 'Food Prep.' activity involves the use of a powered appliance is not consistent throughout the day.

Similar analysis for the TUS 'Laundry' activity with respect to the use of washing machines and dryers in the HES dataset is shown in Figure 3.3. The correlation between washing machine use and the 'Laundry' activity is generally consistent, with some specific periods of weaker correlation. For example, use in the late morning period is overestimated, which suggests that a proportion of the occupants undertaking this activity in this period are completing rather than starting the activity (e.g. removing washing and hanging to dry). Dryer use has no clear correlation with the activity.

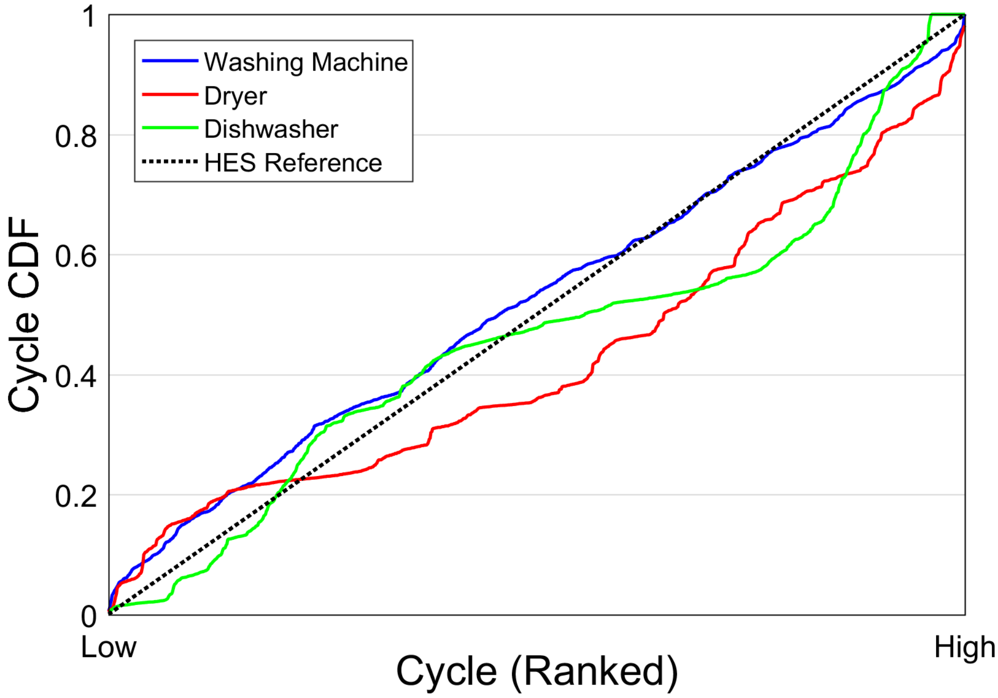
The presented analysis, and equivalent analysis for other activity-appliance combinations, shows that while there are variations in the degree of correlation between activity and appliances, with some (e.g. washing machines) significantly stronger than others (e.g. kettles, ovens), using this relationship for all intermittently-used appliances is likely to only weakly predict actual use timing. Only for 'TV' use can a direct relationship be inferred, which is reviewed in 5.10.

Further analysis was undertaken of the output from a TUS-activity based demand model to confirm the potential for weak prediction accuracy. Individual appliance cycle start time prediction using the same model basis as Richardson et al [69], calibrated by linking UK 2000 TUS activity data to specific appliance use, determined that the average and range of predicted cycle start times varied significantly from start time distributions in the HES dataset for most appliances. The evaluation was undertaken by generating appliance start time cumulative probability distributions from the HES data and then determining the appropriate cumulative distribution function (cdf) value for each modelled cycle start time.

Figure 3.4(a) shows the results for the four main 'cooking' appliances, where the dotted line represents the linear distribution of *cdf* results from the HES dataset and the model results are the ranked distribution of HES-equivalent *cdf* values for each modelled cycle. For all but the kettle distribution, the average modelled cycle start



(a) Cooking Appliances



(b) Washing Appliances

Figure 3.4. Comparison of TUS-activity calibrated model appliance timing prediction to the HES dataset [89] analysed basis for various cooking and washing appliances.

time is significantly earlier (lower *cdf* value) than for the measured data, and the oven and microwave distributions, in particular, indicate a poor overall match between model and measured data. Similar results are showing for the washing appliances in Figure 3.4(b), with only the washing machine model output showing good correlation with the measured data. In all cases the results are consistent with the direct comparison between activity and use time detailed above, and the relative probabilities shown in Figures 3.2 and 3.3. This confirms that the identified weak correlation between TUS activities and specific appliance use limits demand model accuracy where the activity is used directly to predict appliance use.

The conclusion is that TUS activities are, in general, a weak proxy for time-dependent appliance use and therefore for high resolution demand prediction. If alternative, appliance-specific, data is available it should either be used directly, or used to calibrate the TUS activities for variances in the per-timestep ratio between power and activity probability. As this analysis was based solely on UK data, it would need to be determined if the lack of correlation holds for other countries.

3.3.3.2 Occupant Differentiation

Existing models have incorporated a variety of methods to distinguish between individual occupancy patterns. Few have presented a logic for the differentiation, other than the base assumption that different occupant types will have distinct occupancy traits. The models also vary in the extent to which they are comprehensive, that is allowing the majority of all occupants and households to be simulated.

Of the existing methods, only the Richardson et al model [134] both incorporates an element of differentiation (household size) and allows most households to be modelled. However, distinguishing based solely on number of occupants does not capture the potential differences based on factors such as age and employment status.

The models that use specific but limited occupant and household archetypes ([68], [136]) are suitable for investigating the potential for differences between dissimilar households but do not allow for comprehensive modelling. Selecting representative archetypes from existing short duration diary datasets is also difficult. Aerts et al [144] has attempted to solve this by first identifying archetypal behaviours then linking them proportionally to occupant types, which can then be combined to produce more realistic

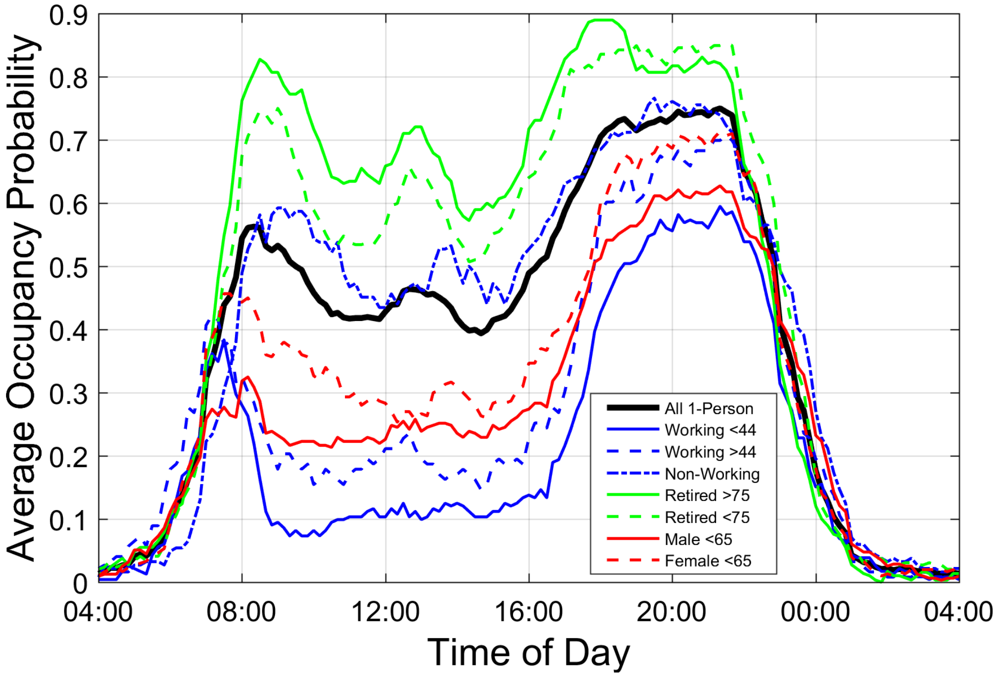


Figure 3.5. Average weekday active occupancy probability for different single-person householder types. Data for analysis from the UK 2000 TUS dataset [83].

profiles. In this case, the model can be comprehensive but, as stated, difficulties stem both from identifying useful archetypes from clustering analysis of well distributed data and allocating realistic sequences of archetypal days to individuals.

Yamaguchi et al [143] and Wilke et al [71] present the most comprehensive models in terms of the variety of different occupant types used. Yamaguchi et al split the dataset into 25 different occupant types, aided by the large size of the Japanese TUS dataset. As outlined, Wilke et al attempted to isolate the impact of each individual occupant trait and day type, and combine the influence statistically for each modelled individual.

Figure 3.5 shows the impact of the basic differentiators of age, employment, and gender on single-person householder average occupancy. From this, and the analysis of [143] and [71], it can be concluded that a significant level of differentiation is appropriate to capture the fundamental differences.

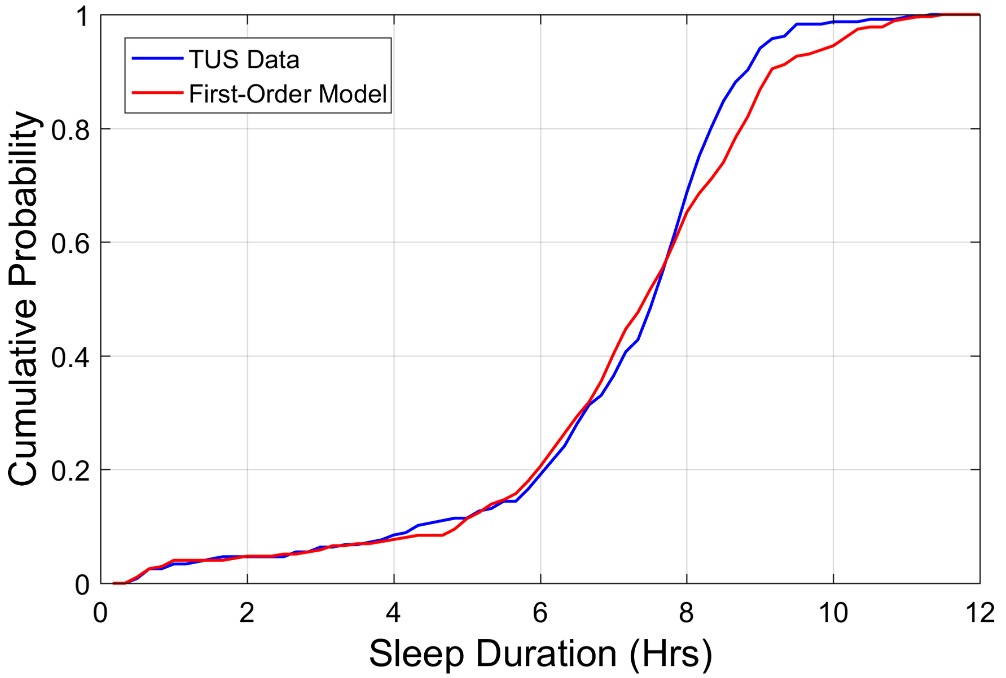
There is an obvious balance to be struck between differentiation for specific occupancy states and maintaining statistical robustness with an effective depth of probability data from larger calibration populations. It needs to be determined therefore what the minimum population size is for the selected method and which particular differentiators are most significant.

3.3.3.3 'Memory'

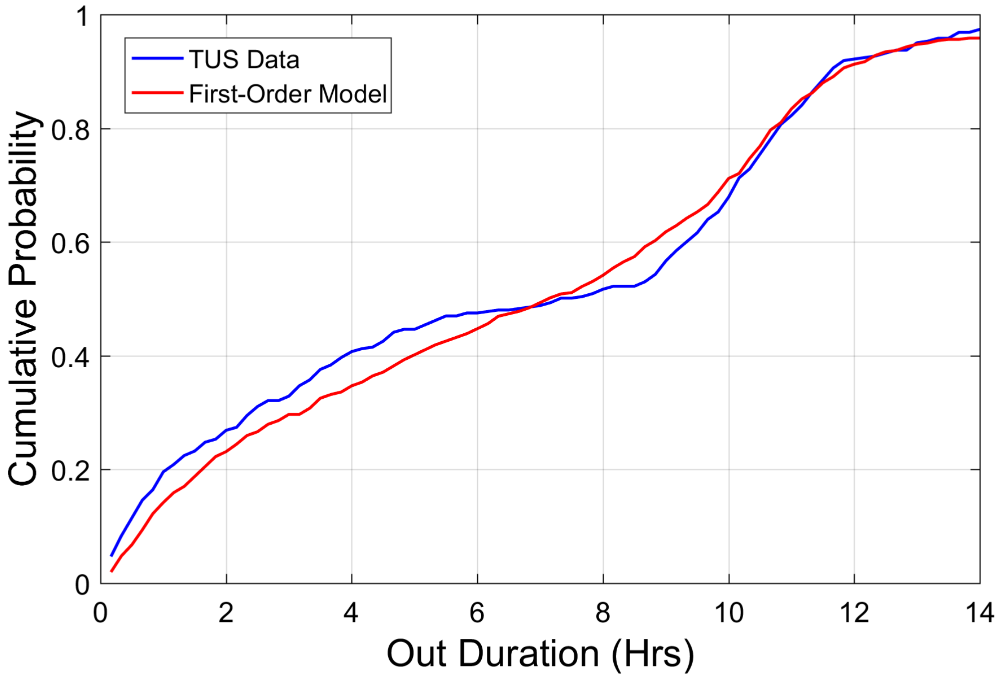
As stated, a process with the Markov property is defined as one in which the state at time $t+\Delta t$, is only dependent on the state at time, t. The concept of 'memory' in time-dependent probability modelling is the degree to which the prior sequence of states or duration of the current state influences the transition probabilities. The extent to which any model requires 'memory' is therefore dependent on whether the Markov property can be safely assumed.

For domestic occupancy or occupancy-activity models this property has been implicitly assumed to be correct by a number of first-order models ([134], [135], [136]). The event-based models ([68], [71]) are higher-order only for state duration and assumed to be Markov property-compliant with respect to the state transition, which remains a first-order process.

For models which focus solely on basic occupancy parameters (i.e. sleep/active/out)



(a) 'Sleep' state



(b) 'Out' state

Figure 3.6. TUS data and first-order occupancy model comparison for 'sleep' and 'out' state duration distributions for single-person householders in the 28-56 years old age range on working weekdays. TUS data for analysis from the UK 2000 TUS dataset [83].

([134], [135]), this assumption is largely acceptable for the type of transition as transition to 'sleep' and 'out' states are typically from the 'active' state. The assumption cannot be made generally for the prediction of state duration for all models and specifically for models which include multiple TUS activities as detailed below. That state duration is not Markov compliant was shown by Wilke and was a primary driver for the higher-order 'event' method used for the multiple activity model.

Whether the Markov property assumption can lead to poor prediction of state durations in a three-state model can be shown with reference to the 'sleep' and 'out' states. Analysis of the UK 2000 TUS dataset (see Figure 3.6) shows that both have distinct duration distributions; 'sleep' with a high proportion in the 6-9 hour range and 'out' on working days with a distribution where both short (<3 hours) and longer duration (>8 hours) absences predominate. Using a first-order Markov chain model calibrated from the same populations as the TUS data, each simulated duration distribution is less distinct than the equivalent TUS distribution (see Figure 3.6), although the broad characteristics of the distribution are maintained.

The reason for this weak replication can be explained with reference to the significant range of sleep and waking times, and sleep durations, for a specific population (see Figure 2.20). When combined in a first-order Markov chain model, there is no means to account for the significant differences in both timing and duration. The modelled waking time only being driven by the independent probability of a waking event at a particular timestep.

Further analysis is required to determine if a higher-order model can be developed that significantly improves the replication of the TUS dataset duration distributions and whether the degree of improvement with regard to overall modelling accuracy justifies the increase in calibration data complexity.

For multi-activity models, typically based on TUS activities, while the next activity is not strongly correlated with the preceding activity, the probability of a specific activity is strongly dependent with whether it has occurred within the preceding waking period. Comparison of the UK 2000 TUS dataset and output from a first-order eleven activity model similar to that used by [70], based on all one-person households, shows that the model underestimates the probability of an activity non-occurrence during a particular day by 10% and also overestimates the number of 6+ occurrence by over

114%. As with the occupancy duration analysis, the activity repetition for the first-order activity model tends towards the average per-timestep behaviour, which can lead to unrealistic activity sequences if calibrated from multiple individuals with conflicting patterns.

Analysis of the dispersion of activities throughout the day shows that, where specific patterns are observed, the first-order model fails to capture this detail. For example, the model underestimates by over 140% the tendency for people to return to watching TV as the *next+1* activity and overestimates by 80% the probability of 'food preparation' reoccurring within four transitions.

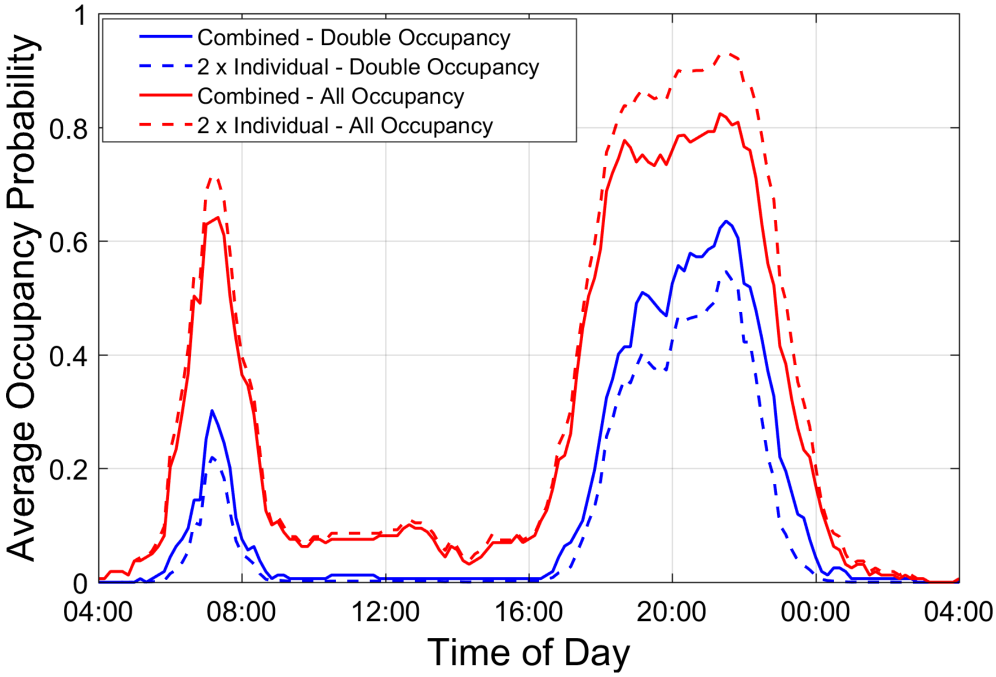
The overall conclusion is therefore that the use of a first-order basis for multiple TUS activity models is not sufficiently accurate primarily as the result of weak replication of intra-day activity frequencies and requires additional assessment for use in basic occupancy state models to determine if weaker state duration replication is significant.

3.3.3.4 Occupant Interactions

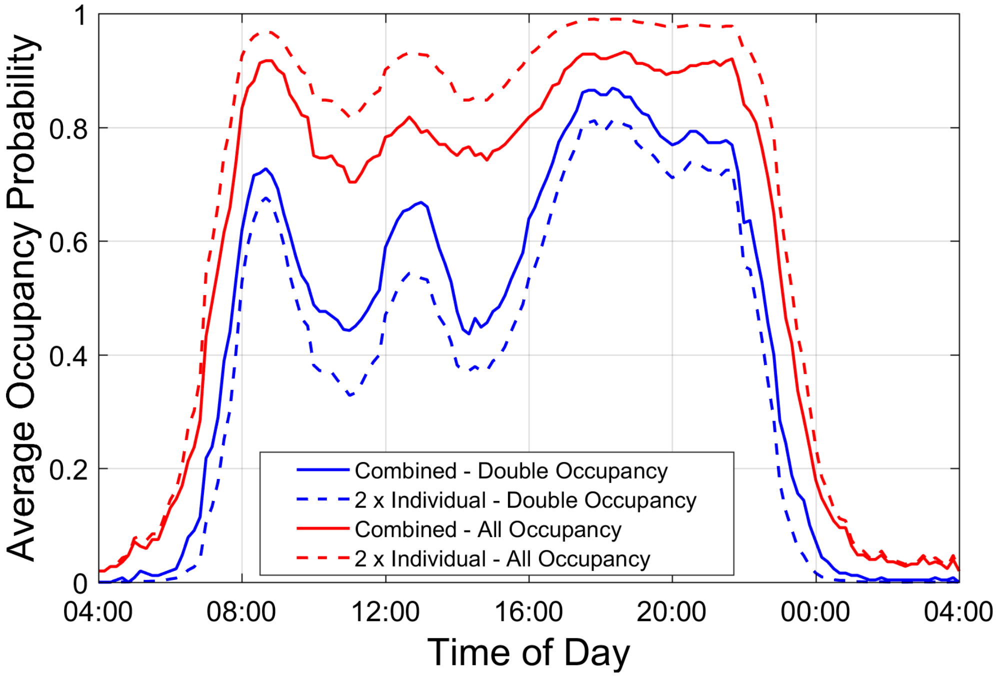
Only the work by Richardson et al [134], Baptista et al [138], and Collin et al [140] have attempted to capture occupancy or activity-related interactions between different members of the same household. Other models have assumed that each occupant can be modelled independently. As stated by Yamaguchi and Shimoda [145], the unit of high-resolution energy simulation is the household not specific individuals, therefore ignoring interaction effects has the potential to result in poor occupancy prediction.

Using number of active occupants as the model occupancy state basis, the Richardson et al model [134] captures the changing probability of multiple occupancy at different times of day. Baptista et al [138] focused on mealtime and bathing interactions, with the timing and interaction for all occupants driven by an identified primary occupant, with the activity state for other individuals determined probabilistically based on the state of the primary occupant. Collin et al [140] used a standard individual occupant first-order Markov chain approach but determined the probability of shared appliance use by analysing the UK 2000 TUS data for the proportion of time where occupants shared an activity and location. This is of particular relevance to the 'TV' activity, where a single unit per location can be assumed.

These three methods identify distinct ways that interactions can be simulated. By



(a) Under-44 years old, both working, weekday



(b) Over-68 years old, weekday

Figure 3.7. Potential combined active occupancy prediction error for two couple average age sub-populations if each individual is modelled independently.

direct model calibration with the interaction term built in, by targeted manipulation when a specific state is identified for one or more occupants, or by a further statistical manipulation when a particular scenario is predicted.

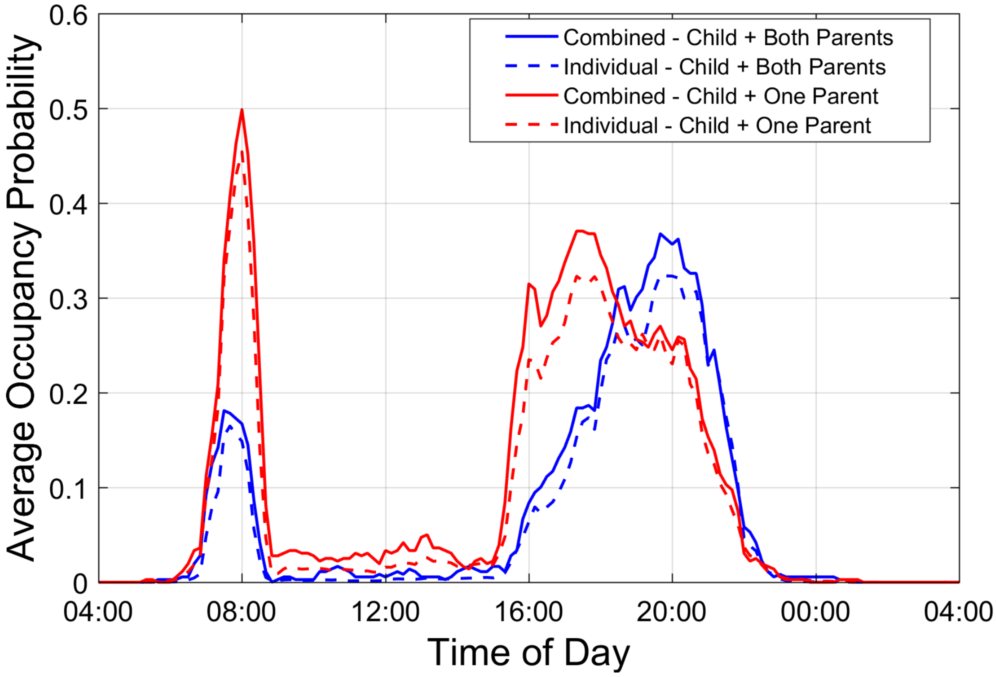
Modelling occupancy interactions adds significant additional complexity, therefore it needs to be determined whether interaction effects are sufficient to warrant inclusion. A variety of different potential occupancy interactions were reviewed for relevance and sufficient data availability using the UK 2000 TUS dataset:

Couples - Co-habiting couple households account for 26.3% (1305) of the UK 2000 TUS dataset households. Analysis of a variety of couple sub-populations, differentiated based on age and working status, showed that if modelled as individuals there would be an underestimate of double occupancy and, therefore, also an overestimate of both overall and single occupancy in all cases. (Double occupancy probability for the individual model basis is determined from the square of the individual occupancy probability). Figure 3.7 shows the weekday analysis for couples with an average age of less than 44-years old with both working and retired couples with an average age of over-68 years old, and confirms that occupancy interactions within couple households are significant.

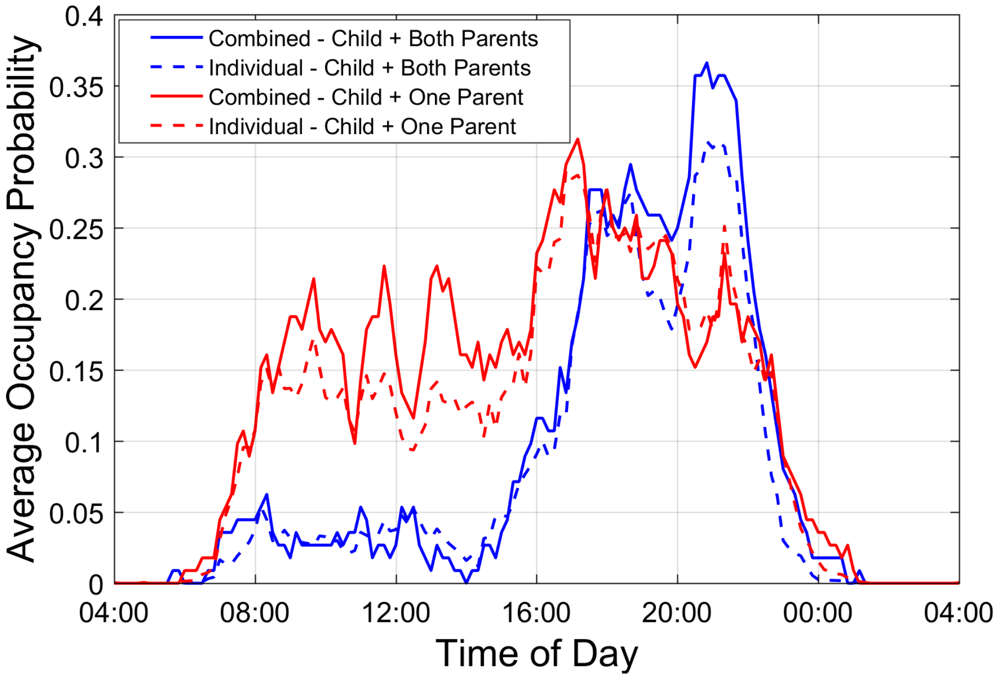
Parents - Family households with two parents account for 25.4% (1258) of UK 2000 TUS dataset households. The results were similar to the couple sub-populations, showing significant occupancy interaction and potential for prediction errors if each parent is modelled separately. Typical parent occupancy patterns differ from equivalently aged couples without children, particularly in the waking and late afternoon periods, and therefore must be considered separately.

Children - The same interaction influence can also be shown for parents and children. Assuming no interaction underestimates the periods of child occupancy with both one and both parents present (see Figure 3.8). This is consistent across all child ages and for different day types (school-term/non-term, weekday/weekend). The same effect can also be shown for single parent families.

Others - Couple and family households represent the most numerous types for which a degree of occupancy interaction could be expected. Other multi-person household types are less common. For example, there are only 43 households with two unrelated or multigenerational related adults with a variety of age combinations, and



(a) 8-11 years old, weekday, term-time



(b) 12-15 years old, weekday, non-term

Figure 3.8. Potential combined active occupancy prediction error for two parent and child populations if modelled independently.

fewer for equivalent larger households. Therefore, other potential multi-adult household interactions have not been considered at this stage due to lack of data.

3.3.3.5 Model Evaluation Analysis Summary

To achieve the aim of a high-resolution, occupancy-driven, demand model that captures specific household behaviours, the following conclusions were drawn from the analysis of existing methods. The identified problems with first-order occupancy-activity models, of low correlation between TUS activities and specific appliance use, and unrealistic repetition and sequencing of activities, render them less effective for demand modelling where appliance-level demand data is available. With access to appliance-level demand data from the Household Electricity Survey (HES) [89] dataset, restricting the number of states to basic occupancy elements, and using alternative means for demand prediction, was considered appropriate.

In addition, using higher-order methods to capture more realistic occupancy state durations, and capturing the occupancy interactions inherent in couple and family households, was also considered to be a fundamental requirement for any model developed to achieve the project aims.

3.4 Existing Bottom-Up Demand Models

3.4.1 Demand Model Types

A review of "bottom-up" demand modelling methods by Swan and Ugursal [146] identified two broad types; statistical and engineering. The 'statistical' type takes a top-down approach for individual households or specific demands which are then aggregated. The 'engineering' type "explicitly account for the energy consumption of end-uses based on power ratings and use of equipment and systems and/or heat transfer and thermodynamic relationships". Only the engineering type accounts for occupancy influence and individual behaviours directly, with the specific 'archetype' and 'sample' sub-methods applicable for individual household or small area analysis.

'Archetype' models identify common household reference types and generate models reflecting the average behaviour of the type. The number of archetypes is typically limited, and this method is therefore not used for comprehensive models. They are most appropriate as a means of comparing broad differences between household types or the impact of specific behaviours for a single type. The 'sample' approach extends the archetype concept to divide all households into multiple different types that reflect distinct behaviours. The method requires higher levels of calibration data and model complexity but offers the most comprehensive and realistic approach.

3.4.2 Existing Demand Models

Grandjean et al [82] performed a detailed review of demand models and, in particular, the ability of different methods and specific models to capture demand diversity. The highest level of performance was attributed to the set of models with individual behaviour factoring defined as 'scripted probability', analogous to the 'sample' approach defined above, all of which were calibrated using time-use data.

The 'scripted probability' models were developed by Walker and Pokoski [117], Capasso [118], Armstrong et al [119], Widen and Wackelgard [70], and Richardson et al [69]. As modelling demand diversity is a key aim of this project the following review focuses on this type of model as the current 'state-of-the-art'. Aspects of other models are included where applicable, in particular the methods developed by Stokes [147] and Wilke [71]. Unless otherwise stated, the detail provided in the work of Walker and Pokoski, and Capasso was insufficient to discern how the model was constructed.

3.4.3 Appliance-Specific Methods

Analysis of the existing bottom-up, high-resolution models identified similarities in the structure and methods used. High consumption appliances, in terms of both ownership and power used, are modelled separately, using different techniques appropriate to their specific characteristics. Remaining low ownership appliances are either ignored or treated as a single element. The different techniques are summarised below, followed by a more detailed review of how they were applied in the existing models:

- ***Cyclic*** - The majority of specific household energy demands are characterised by intermittent cycles of use that are largely occupant initiated and therefore highly dependent on occupancy. The use cycles may have significant time dependence and highly variable behaviours between individual households. This designation applies to most high ownership and usage appliances (except for cold appliances), including washing machines, cookers, TVs, etc., and to hot water use. (A 'cycle' is defined as a distinct use event.)

- ***Constant*** - 'Constant' demands are largely independent of occupancy. The demand may be typically constant (e.g. alarms, telephones, pond pumps) or have a cyclic power demand which is time and/or condition dependent (e.g. cold appliance pumps cycling on and off based on temperature).

- ***Conditional Loads*** - Heating, cooling, and lighting use is driven by a complex set of factors. Primary factors are occupancy and the relevant environmental condition (external and internal lighting levels, temperature, etc.). Fixed timing patterns based on average or typical need may also be used.

- ***Standby*** - 'Standby' power is demand associated with appliances that are not fully switched off while not in use. This is primarily from cooking, audio-visual (e.g. TVs), and IT appliances.

- ***Miscellaneous*** - Appliances with low ownership probability and/or low total consumption are typically not possible or necessary to model individually. Most bottom-up models therefore need a method to capture statistically both the ownership and power demand potential for this group of appliances.

3.4.3.1 Cyclic Demands

Walker and Pokoski [117] combine an availability (occupancy) sub-model based on a normal routine with probabilistic variations in key transition times, with activity-specific proclivity functions (i.e. mealtime, laundry, etc.), to determine a per-timestep probability of a specific appliance use. This was applied in the same manner to both electrical appliance and hot water loads. The work of Capasso [118] was based on the same method.

Armstrong et al [119] used an approach which did not consider occupancy directly but based potential appliance use on analysed time-of-use potential per appliance. For each appliance per household, an annual target for power consumption was allocated from measured data and a use probability per timestep determined to achieve this overall consumption. A sequential per-timestep model then determined when a use occurred using a Monte-Carlo method. Fixed power draws and cycle durations are set per appliance. The method was only applied to electrical loads, but it was stated that it could also have been used for hot water use if the equivalent demand data was available.

Richardson et al [69] used a similar per-timestep probability approach to Armstrong et al but with appliance use first predicted by occupant presence using the output of the occupancy sub-model reviewed in 3.3.1. Per timestep use probability is determined based on occupancy, TUS activity probability, and a fixed number of annual cycles. The same sequential per-timestep probability method as Armstrong et al then determines specific cycle times. This work only considered electric shower cycles and not overall hot water demand.

Widen and Wackelgard [70] linked electrical appliance demand directly to the TUS activity prediction sub-model detailed in 3.3.1. Cyclic loads are characterised by three basic configurations, (1) a constant demand for the activity durations (e.g. for cooking), (2) a variable power cycle commencing at the end of the activity to reflect a pre-use preparation period followed by the appliance start (e.g. for laundry, dishwashing), and (3) as for (1) but with an additional standby load when not 'in-use' (e.g. for TVs and computers). Hot water demand was not considered.

A hot-water specific demand model ('DHWcalc') [148] has been developed by Jordan and Vajen as part of the IEA-SHC Task 26 project [149], and is open-source. The model is focused on identifying demand timing based on user-specified average daily demand (in litres) and overall timing characteristics. The model allows the proportion of total demand in six user-defined periods, and four distinct demand types with different flowrates and durations, to be specified. Alternatively, built-in probability distributions based on analysis by the authors can be used. Per-cycle variation is allowed for by setting both a mean and standard deviation for cycle flowrate. The model also incorporates a sinusoidal seasonal variation element and an allowance for extended absences.

3.4.3.2 Constant Loads

The nature of constant loads allows them to be modelled using relatively simple methods. However, for some applicable demands there are intra-day (diurnal) and seasonal variations to be captured.

Richardson et al [69] included cold appliances as constant loads. They were treated in the same manner as cyclic loads with a fixed cycle time, minimum time between cycles, and number of annual cycles. The probability of a cycle start per timestep is constant, is independent of occupancy, and has no additional factoring for diurnal or seasonal variations. Widen and Wackelgard [70] applied a fixed power per cycle for fridges and freezers. Cycle on and off durations and timing was fixed per household and allocated probabilistically based on analysis of measured data. Armstrong et al [119] used a fixed 70-minute cycle based on measured data from a single unit, that is then scaled to the target annual consumption with no diurnal or seasonal factoring.

Analysis of the cold appliance demand in the HES dataset by Zimmermann et al [150] highlighted a significant seasonal influence of approximately ±15-20% as a result of room temperature variations. Stokes [147] used a sinusoidal function to capture the same effect relative to an annual average power level based on analysis of demand data. This is a common method used to track seasonal variations caused by either temperature or average daily solar levels. Wilke [71] explicitly did not consider seasonal influences but did incorporate a per-timestep intra-day function for each modelled appliance, including cold appliances. Intra-day variations for cold appliances are expected because of the variable probability of door opening resulting from the time dependency of active occupancy.

3.4.3.3 Lighting

Lighting use is driven primarily by a combination of occupancy and external solar levels. Additional factors such as non-occupancy driven uses for security and safety, and the potential for poorly daylit rooms, also need to be considered.

Richardson et al [151] combined the output of an occupancy sub-model with measured external solar levels to determine if there is a lighting demand. The method determines lighting power per timestep based on a probabilistic allocation of individual bulb wattages and an empirical model based on observation data to determine the use probability per bulb and duration of each lighting use.

Widen et al [135] used a similar method to combine occupancy and external solar level. The lighting power level is set based on the difference between the external solar level and a target, with the lighting level assumed to be controllable in a series of defined steps between a minimum and maximum wattage level.

Armstrong et al [119] used three seasonal use probability profiles (winter, summer, shoulder) and five lighting levels. Lighting events were allowed to occur randomly during occupied periods with durations of between 5 and 120 minutes, and cycles could overlap.

As part of a detailed analysis of lighting use in the HES dataset, Terry et al [113] developed a method that used TUS activities to probabilistically place occupants in specific rooms, HES data to determine room lighting characteristics and unoccupied use potential, and external solar data to determine if lighting is required.

As outlined by Robinson et al [152], there is a lack of lighting models based on comprehensive observations of real behaviours, which remains a valid conclusion as all the methods outlined require a degree of extrapolation from indirect data sources to define lighting use. Unfortunately, the lighting data provided by the HES dataset, which is primarily measured at central distribution boxes, does not provide a clear indication of specific room-level lighting behaviours.

3.4.3.4 Standby Loads

Standby power use for appliances is consumption for units that are left powered while not in active use. In particular, televisions, microwaves, and computers are often left in this state [153]. Analysis of the HES data [153] determined that this accounted for 5.1% of electricity used on average, but varied significantly per household.

Armstrong et al [119] applied a constant standby load of 65W based on the average for Canadian households. Richardson et al [69] applied standby power use as a fixed quantity per appliance when not in use for a range of appliances, including TVs, set-top boxes, microwaves, and ovens. Widen and Wackelgard [70] applied a fixed standby load for TVs, stereos, and computers.

As outlined below, Widen and Wackelgard, and Wilke [71], both applied an additional load to account for residual consumption not accounted for by the appliance-specific sub-models, which also included an element of standby power.

3.4.3.5 Miscellaneous Loads

One key limitation of bottom-up models is the need to account for all sources of demand separately. Where an existing model has aimed to account for all demand rather than only to introduce a method for activity-linked or cyclic demands, a variety of techniques have been used to account for low ownership or low total consumption appliances without the necessity to model each individually.

Armstrong et al [119] identified 27 smaller appliances (including TV's, computers, kettles, microwaves etc.) that were allowed to operate randomly based on an overall time-use curve calibrated for a target proportion of 'on' time per appliance. As for lighting, cycles were set randomly at between 5 and 120 minutes, with the potential for use of different appliances to overlap. Duration and power rating were also factored based on a relative energy use distribution determined from measured data.

Widen and Wackelgard [70] primarily focused on TUS activity-linked demands and incorporated a single additional factor to account for undefined loads. The added value was constant and differentiated solely by the number of occupants based on measured data analysis. Wilke [71] subtracted specifically modelled demands from the total and found, for the residual consumption, limited time dependence and a significant range of values per household (0-600W with a mean of c.120W). It is not explicitly stated how this is applied within the overall model.

An alternative approach to capture the stochastic nature of total household electricity demand using a Markov chain model was developed by McLoughlin et al [78]. Measured data was used to calibrate a single transition matrix without any time dependency for each of five households. The Markov chain matrix 'states' were 24 equal sized ranges of measured power levels. This was then used to generate synthetic profiles. The results show that the model was able to replicate the overall characteristics (i.e. mean, variance, etc.) of the measured data but did not capture time-specific details. The lack of any time dependency and equal size ranges are likely causes of the poor detailed performance. Whilst the model may not be suitable to model the overall demand for a household as published, it could potentially be used to account for a proportion of the demand (e.g. miscellaneous appliances) as a single element, with the addition of time-dependent Markov chain calibration if a time dependency can be shown.

3.4.4 Existing Method Evaluation

In addition to the weak correlation between time-use survey activities and appliance-specific demand timing demonstrated in 3.3.3.1, which impacts performance for models that use the activities to determine appliance use timing directly, two further potential areas of weak demand modelling performance have been identified.

3.4.4.1 Household Characteristics Factoring

The identified existing modelling methods have either limited or no factoring of either total or specific demands per household that accounts for differences in household characteristics or energy-use behaviours.

Armstrong et al [119] set total annual demand targets for three distinct household archetypes (low, medium, high) and used the model to determine per-timestep demand profiles. Widen and Wackelgard [70] used fixed power levels linked to the identified TUS activities. The method captures a degree of temporal demand diversity but not diversity in baseline power, which they state could further refine the method. Richardson et al [69] only used undifferentiated appliance ownership probabilities to distinguish households, with a fixed number of cycles per appliance and a fixed power level per appliance, regardless of the number of occupants or other characteristics.

The analysis presented in 2.4 demonstrated that household characteristics and individual household behaviours have a significant influence on overall demand, and for each specific demand. The influence can be shown for a diverse range of factors, from appliance age and size to number of daily cycles and use duration. Ignoring these variations, which can typically be characterised by probability distributions, is likely to result in poor prediction of the range of potential demand profiles in any model output.

3.4.4.2 Realistic Cycle Sequencing

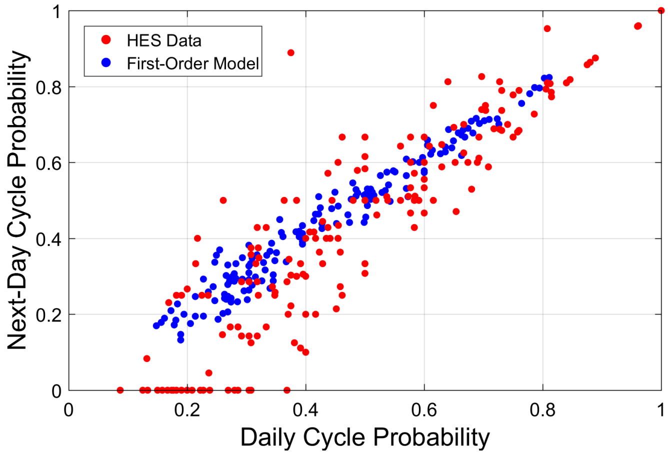
Similar to the Markov property assumption for occupancy modelling (see 3.3.3.3), existing first-order, per-timestep appliance cycle models also have an inherent assumption that future cycle start times are independent of previous sequences or time since the last cycle. As for occupant activities, this assumption ignores distinct use behaviours for certain appliances. For appliances where both daily use and multiple daily cycle probabilities are high, such as kettles and microwaves, the potential is for both the modelled number of cycles per day and the intra-day cycle sequences to be unrealistic if the preceding sequences are ignored.

Kettle data for each of the single-person households in the HES dataset, and the results for a first-order, per-timestep probability model, similar to that used by [69] and [70] and averaged over 20 annual duration runs, were compared for the average absolute difference per day between the actual and average number of cycles relative to the average number of cycles. The average value for each household gives a measure of the typical household daily use variation. For the HES dataset the average is 0.387 and the equivalent for the first-order model was 0.608. This suggests that the first-order method generates excessive day-to-day variability. Similar results were observed for the other high use appliances.

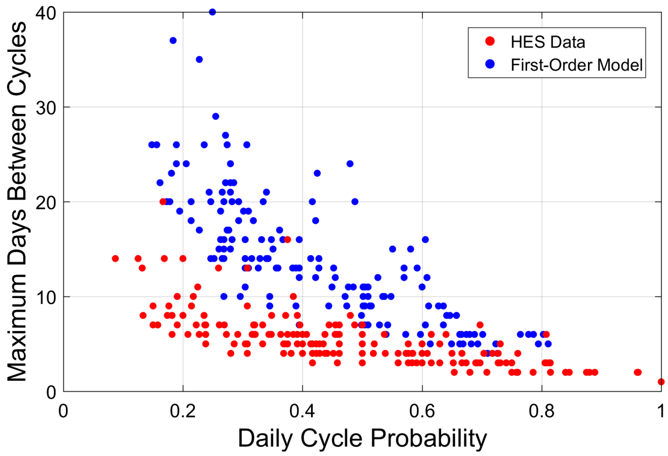
Intra-day cycle sequence performance can be determined by comparing the standard deviation of cycle start times per household from the HES dataset to the same first-order model. This is achieved by converting all cycle start times to minutes after 04.00 (e.g. 08.30=270) and then determining the standard deviation of the converted cycle start times for each day. For a 6-cycle kettle use day, the HES dataset average cycle start time standard deviation is 285 minutes, with most households varying linearly from 200 to 375 minutes. The first-order model output has an average of 346 minutes with an equivalent range from 100 to 600 minutes, suggesting distributions of daily cycle start times that vary significantly from actual behaviours.

For lower use appliances with use patterns that are likely to extend over several days, such as washing machines and dishwashers, the modelling challenge is realistic sequences of use and non-use days, rather than cycle sequencing. For some appliances of this type, the probability of a cycle will also increase with time since the previous cycle as the need (e.g. quantity of clothes or dishes requiring cleaning) accumulates. The sequential per-timestep probability models ([119], [69], [70]) do not reflect this.

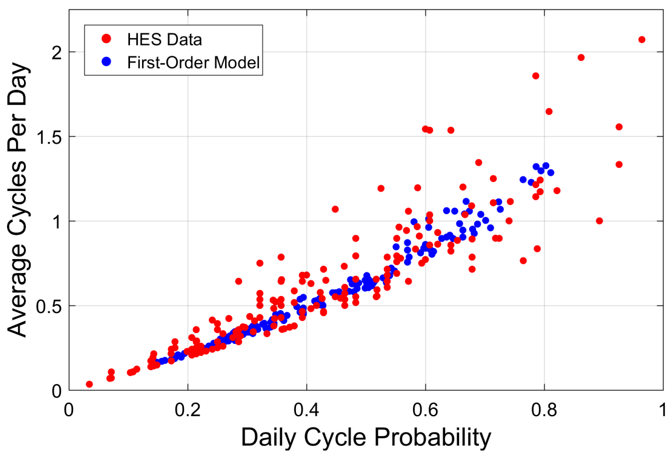
For washing machine use, the HES dataset and the first-order, per-timestep use probability model results for three key statistics have been compared based on daily



(a) Next-day cycle probability



(b) Maximum days between cycles



(c) Average cycles per day

Figure 3.9. Comparison of three key use pattern defining statistics by daily cycle probability for washing machine use from analysis of the HES dataset [89] and a first-order, per-timestep use probability model.

cycle probability as shown in Figure 3.9. Next-day use probability is the probability of use on the day following a cycle, and the first order model has a narrower distribution and higher average than the HES data. For the maximum number of days between cycles, the first-order model shows a significantly higher basis than the data, suggesting that the per-timestep approach, with very low use probability per timestep, overestimates this potential. Finally, the average cycles per day is lower for the first-order model, suggesting that there are multiple-use behaviours not captured by a first-order model. Deeper analysis of the data shows that distribution of days between cycles is more random for the first-order model than the HES dataset, with the first-order model failing to capture need-driven use behaviours.

In contrast to the per-timestep approach, the hot water timing model developed by Jordan and Vajen [149] uses an event-based approach, where timing is determined probabilistically based on overall daily use characteristics that are defined separately. This provides a means to control both the total use per day and relative timing that is not possible with a per-timestep method as shown.

In general, as a means to determine realistic appliance use patterns, and therefore overall demand profiles using a bottom-up approach, current published methods are ineffective for household or small-scale energy system modelling.

3.5 Model Development Options

The selection of the most appropriate modelling methods is dependent on a number of factors; calibration data availability, scale of analysis, computational power, and desired outputs. Bottom-up, probabilistic, agent-based, and high-resolution discrete-time models have the potential to generate the greatest level of detail, but can be inefficient if not justified by the limitations and requirements imposed by each of the four listed factors.

As defined in 2.6, occupancy and demand patterns within all households can be characterised as being highly stochastic with time dependency. Whilst the potential states and values are finite, and broad patterns and individual drivers influence behaviour at a daily, weekly and seasonal level; there is a significant inherent, or at least apparent, randomness in occupancy state or energy demand at any specific time, which increases with increasing time resolution.

The stated aim is to capture the influence of individual household and household-type behaviours, and apply the model to time sensitive analysis such as solar energy matching and diversity. As identified in 3.2 and confirmed in the remainder of this chapter, this strongly suggests any such model should be, in general terms, bottom-up, probabilistic, agent-based, and high-resolution, with consideration for any limits imposed by calibration data availability and resolution, and computational limitations.

The principal model development decisions were therefore; the degree of differentiation of occupant (agent) behaviour, the model time resolution, whether to use a time or event-based method for occupancy and demand timing evaluation, and how to incorporate the influence of occupancy within the overall demand model. The conclusions are detailed in Chapters 4 to 7.

3.6 Chapter Summary

This chapter reviewed the basis and performance of existing high-resolution occupancy and demand models. The chapter highlights are as follows:

- Existing occupancy models can be broadly characterised as either discrete-time, requiring a probability calculation for each timestep, or discrete-event, requiring a single probability calculation to determine each state and its duration. The majority of existing models are discrete-time models using the first-order Markov chain method.

- The use of time-use survey activities in existing occupancy sub-models to determine when specific appliances are used within an associated demand model was shown to be less effective. The time-use activities are not sufficiently specific to allow the use of appliances to be directly inferred and there was shown to be a weak correlation between associated activity and appliance use timing.

- Existing occupancy models are limited by one or more of the following: a lack of differentiation for occupant type, age, and employment status; by weak prediction of occupancy state durations as a result of the first-order methods used; and by ignoring occupant interactions in both couple and family households.

- Existing demand models primarily use a first-order, per-timestep use probability model that can be shown to poorly replicate actual sequences of use within each day and over several days.

- Existing demand models do not account for behavioural variations in total energy or appliance-specific use between household types or between individual households of the same type. This limits their use for high time resolution analysis of the range of potential demands for individual households or small-scale energy systems.

Chapter 4

Model Structure, Characteristics and Occupancy Sub-Model Development

4.1 Chapter Overview

In Chapter 2 it was shown that a number of occupant and household characteristics influence occupancy probability. In particular, age and working hours were shown to be highly correlated. In Chapter 3 several potential improvements to existing occupancy modelling methods were identified, including improving occupant differentiation to capture the age and day type influence. In addition, improved methods were sought to account for occupant interaction effects and occupancy state duration prediction accuracy.

The following chapter first details the development of a sub-model for the prediction of household characteristics based on location and house type. This is used for situations where these are unknown or where a representative population is sufficient.

The development of an enhanced occupancy model is also described. The specific enhancements are as follows: (1) combining couple and parents as single entities within the sub-model to account for occupancy interactions, and linking child occupancy directly to the parent occupancy, (2) improving state duration prediction using a higher-order Markov chain approach based on ranges of duration, (3) significant differentiation based on household type, age, and working hours based on an assessment of the optimal number of sets of individual single-day occupancy data required for effective modelling, and (4) secondary models developed to run in parallel with the

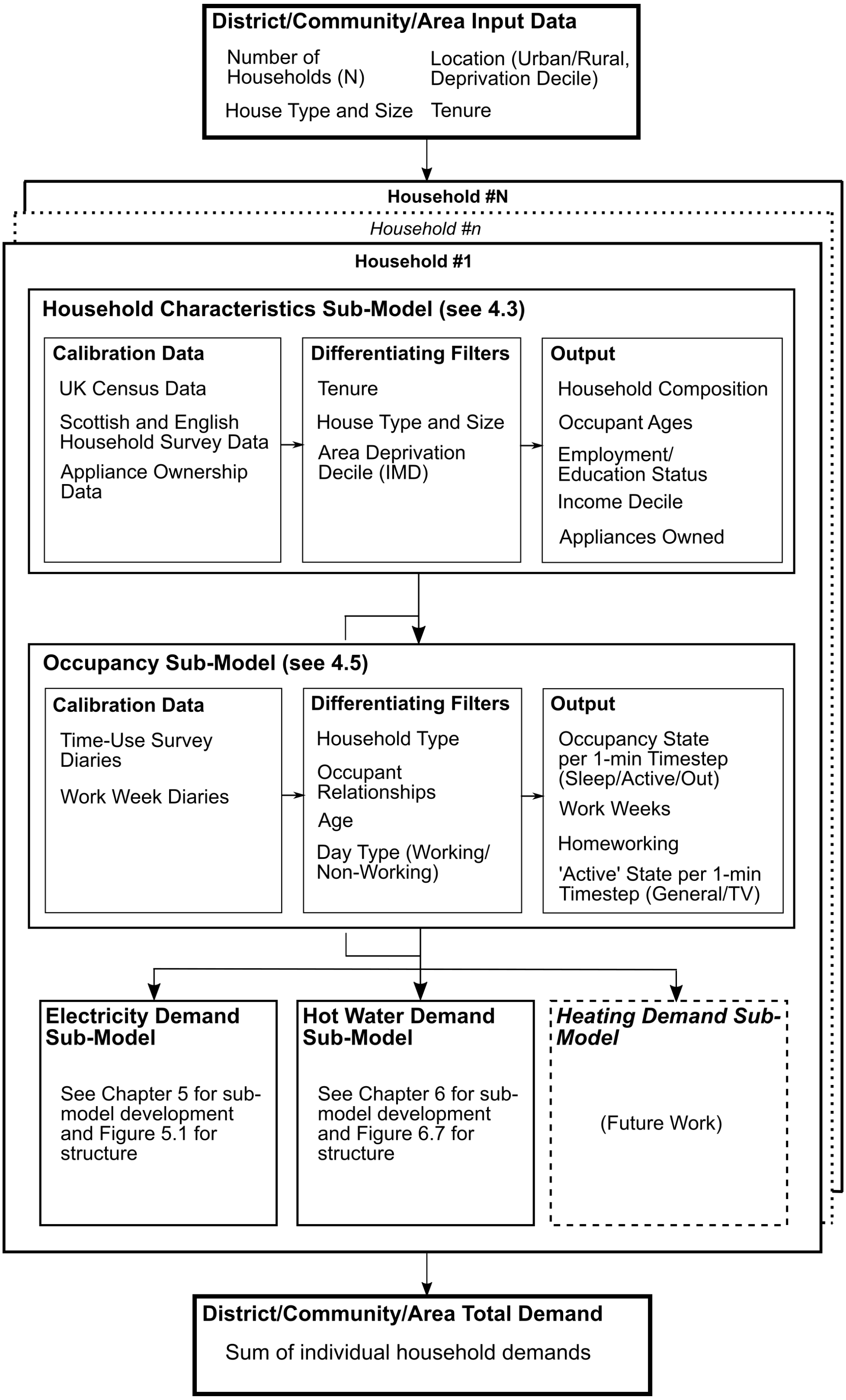


Figure 4.1. Overall demand model structure.

primary three-state occupancy model to account for homeworking and sleeping away potential, and TV use. The developed model limits the number of occupancy states in the primary analysis to the three base states (sleep, active, and out), to account for the weaker demand prediction performance and increased calibration data requirement of models that incorporate multiple time-use activities for the active periods, and the availability of appliance-level demand data.

This chapter also outlines the overall structure of the full occupancy and demand model in the first section. The development of the demand-specific elements is covered in Chapter 5 for electricity and in Chapter 6 for hot water consumption.

4.2 Overall Model Structure

As detailed in 2.4, household demand is a complex interaction of a variety of inter-dependent factors, including household type and size, income, occupancy, employment and education status, and unique behaviours. Analysis of the relationships presented in Table 2.1, defines a sequence of steps required to transition from the minimum required level of household information to a prediction of energy consumption. This sequence is outlined in Table 4.1 and has been split into three distinct sub-models; Household Characteristics (Steps 1-4), Occupancy (Steps 5-6), and Demand (Steps 7-11). This has been used to structure the overall model which is shown graphically in Figure 4.1, including a summary of the principal calibration datasets, differentiating filters, and outputs.

The developed method assumes basic location, tenure, and house size and type information is available as a minimum. Analysis of specific projects will also vary in the extent to which the overall community and individual household characteristics are known. At each stage known information can replace modelled characteristics, if available.

All modelling work presented was implemented in *MATLAB*.

Table 4.1
Overall demand model calculation sequence.

Step	Calculated Factor	Input Factors
1	Household Composition	Location, House Tenure/Type/Size
2	Ages	Location, House Tenure, Household Composition
3	Income	Location, Household Composition, Ages
4	Employment / Education	Location, House Tenure, Ages, Income
5	Work Weeks / Calendar	Ages, Employment/Education
6	Occupancy	Household Composition, Ages, Employment / Education, Work Weeks / Calendar
7	Appliance Ownership	Household Composition, Ages, Income
8	Cycles/Use	Household Composition, Ages, Income, Occupancy, Appliance Ownership
9	Energy Ratings / Appliance Power	Household Composition, Ages, Appliance Ownership
10	Electricity Demand	Cycles/Use, Energy Ratings / Appliance Power, House Tenure/Type/Size, Random Behaviour
11	Hot Water Demand	Cycles/Use, Random Behaviour

4.3 Household Characteristics Sub-Model Development

4.3.1 Minimum Input Data

Location within the sub-model is defined by the deprivation index decile and area classification (urban, town, and rural). As shown in 2.4.1.1, they define the basic socio-economic characteristics of an area, which has a direct influence on potential household composition, income, and employment/education status.

Tenure is defined as private-owned, social-rented (Housing Associations in the UK), and private-rented. This also has a direct influence on the potential household composition per house type and size, and employment status.

House size detail is restricted to number of bedrooms, which has a strong influence on household composition. Other indicators such as floor area or number of habitable rooms could be used but this information is typically harder to acquire in the UK.

4.3.2 Household Composition and Ages

To determine household composition (type and size) and age profile, the sub-model has been calibrated using household survey data. The 2008 Scottish Housing Survey (SHS) data [92] has been used (the English Housing Survey provides the same data and would also be applicable). (The 2008 SHS data was the most recent available at the time the calibration activities were undertaken but has since been superseded by data from 2012-2014).

Analysis of the SHS dataset determined the household composition probability for each house size based on the Index of Multiple Deprivation (IMD) decile and tenure. Further separate probability multipliers were determined based on area classification (urban, town, rural) and house type (flat or house), and combined to further manipulate the size-driven probability as there is insufficient data to allow all factors to be captured in a single level of analysis. Table 4.2 shows an example for a 2-bedroom social-rented house in an urban area with a deprivation (IMD) decile of 3.

Adult ages are determined probabilistically based on the generated household composition and the area deprivation decile. Child ages are linked to the parent age(s). Both derived from SHS analysis.

Table 4.2
Household composition probability for a 2-bedroom social-rented house in an urban area with a deprivation (IMD) decile of 3. Data for analysis from the 2008 Scottish Housing Survey [92]. (A=Adult, R=Retired, C=Child, P=Person)

1A	1R	2A	2R	1A1C	2A1C	1A2C	3A	2A2C	1A3C	2A3C	3A2C	6P+
0.154	0.190	0.113	0.163	0.095	0.080	0.034	0.065	0.062	0.019	0.010	0	0

4.3.3 Income and Employment/Education Status

The calculation sequence to determine both the household income decile (relative to all households) and employment status of each individual is arbitrary as they are closely correlated. The sub-model first determines household income decile based on household composition, deprivation decile, and tenure from analysis of the SHS dataset, and then employment status. Table 4.3 shows an example of the income decile probability of a working age, single-person household in an area with a deprivation decile of 2.

Table 4.3
Income decile probability for a working age, single-person householder in an area with a deprivation decile of 2. Data for analysis from the 2008 Scottish Housing Survey [92].

Income Decile	1	2	3	4	5	6	7	8	9	10
Probability	0.271	0.243	0.236	0.130	0.074	0.028	0.011	0.007	0	0

Employment status is determined based on the household composition and income decile. In couple and parent households, the employment status of both adults is combined and dependent on the overall household income. Adults between 16 and 24 years old in family households are treated independently and assessed for employment and full-time education potential based on age and household income decile. Table 4.4 shows an example of the employment status probability for a single-person householder aged between 49 and 54 years old by income decile. Table 4.5 shows an example of the combined employment status probability for a co-habiting couple with an average age of less than 35 years old by income decile.

Table 4.4
Employment probability for a single-person householder aged between 49 and 54 years old by income decile. Data for analysis from the 2008 Scottish Housing Survey [92].

Income Decile	1	2	3	4	5	6	7	8	9	10
Full-time	0.045	0.134	0.468	0.602	0.706	0.833	0.803	0.720	0.789	0.647
Part-time	0.223	0.128	0.160	0.094	0.144	0.087	0.155	0.260	0.132	0.235
Non-working	0.732	0.738	0.372	0.304	0.150	0.080	0.042	0.020	0.079	0.118

Table 4.5
Combined employment probability for a co-habiting couple with an average age of less than 35 years old by income decile. Data for analysis from the 2008 Scottish Housing Survey [92]. (FT=Full-time, PT=Part-time, NW=Non-working)

Income Decile	1	2	3	4	5	6	7	8	9	10
FT/FT	0.111	0	0.074	0.054	0.176	0.455	0.686	0.798	0.819	0.838
FT/PT	0	0.111	0.074	0.243	0.485	0.348	0.232	0.157	0.146	0.107
FT/NW	0	0.389	0.481	0.243	0.191	0.116	0.041	0.030	0.017	0.028
PT/PT	0.056	0.222	0.185	0.216	0.103	0.071	0.041	0.015	0.017	0.028
PT/NW	0.222	0.056	0.037	0.081	0.029	0	0	0	0	0
NW/NW	0.611	0.222	0.148	0.162	0.015	0.009	0	0	0	0

4.4 Occupancy Model Validation Metrics

Comparison between occupancy model output and actual data is required to allow performance to be assessed and methods compared. Validation of existing occupancy models have used a variety of visual (i.e. charts, graphs) and statistical representations to compare a number of different factors. Table 4.6 summarises the methods used for published models and the validation resolution, and indicates that there is no clear consensus on the best method(s) except for visual assessment of average occupancy profiles.

Table 4.6
Validation factors, resolution, and representation basis used for existing occupancy models.

Factor	Resolution	Visual	Statistical
Household Daily Profile	Per-Timestep	[135]	
Household Average Profile	Per-Timestep	[135], [134], [68], [136]	
Number of Specific Transitions	Per-Timestep	[134]	
Number of All Transitions	Avg., Max., Std. Dev.		[138]
Number of Occupancy Patterns	Avg., Max., Std. Dev.		[138]
Daily Probability	Average	[68]	[138]
Duration	Avg. & Std. Dev.	[68]	
Error	Per-Timestep - Averaged		[68], [71]
Edit Distance	Per-Timestep - Averaged	[144]	
Number of Correct Predictions	Per-Timestep		[71]

Visual methods have the benefit of both simplicity and a lower risk of being misinterpreted due to unrepresentative but statistically distorting outlier results, being typically used to compare model outputs with equivalent data. Statistical methods allow different methods and applications to be directly compared, as per Baptista et al [138] for the comparison of alternative models. Therefore, where possible, visual representations are used for model-to-data comparisons, but the following three statistical metrics have been used to assess relative method performance.

Average Occupancy Metric – determines the average per-timestep error between the time-use survey (TUS) input data and model output for each occupancy state - quantifying the quality of model calibration. Equation 4.1 is based on 144 data points per day (10 minute timesteps). This is equivalent to the 'Error' used by Tanimoto et al [68] and Wilke et al [71], and is a frequently used method for comparing distributions.

$$AO_{state} = \sum_{t=1}^{144} \frac{\left|\bar{P}^{mod}_{state}(t) - \bar{P}^{tus}_{state}(t)\right|}{144} \tag{4.1}$$

where, AO$_{state}$ is the Average Occupancy Metric for state, *state*, $\bar{P}^{mod}_{state}$(t) is the average modelled active occupancy probability for state, *state*, at timestep, *t*, and $\bar{P}^{tus}_{state}$(t) is the average active occupancy probability for state, *state*, at timestep, *t*, derived from the input time-use survey data.

Two means of analysis are possible with this metric:

- First, it can be used to calculate the prediction error for the average per timestep results of multiple profiles generated using an occupancy model. This determines how effectively the model converges to the calibration population average (hereafter referred to as *AO_Conv*).

- Second, it can be used to calculate the prediction error for each individual profile. The mean of this error can be used to determine how closely individual profiles replicate the input data (hereafter referred to as *AO_Var*).

Over multiple profiles, a refined Markov chain model should be consistent with the input data. The two *AO* measures provide a means to assess individual run variation against the overall convergence performance. Within real populations individual occupants will deviate from the population average occupancy, therefore individual profiles should also demonstrate a degree of deviation. A model that tracks broad occupancy characteristics but with some variation about the average behaviour is therefore acceptable within limits.

State Duration Distribution Metric – (hereafter referred to as *DurDist*) is used to assess the ability of a model to generate a realistic range of occupancy state durations. It compares the difference in the cumulative probability function (cdf) values at each 10-minute duration range for the analysed histograms for an occupancy model output and equivalent TUS data in order to determine if the generated occupancy profile replicates the occupancy state durations seen in the TUS data using Equation 4.2. The *DurDist* is the sum of the absolute difference between an occupancy model and TUS data *cdf* values at each duration value for each modelled occupancy state.

This metric is commonly known as the Earth Mover's Distance: a commonly used quantitative histogram similarity measure where the bin values are not independent and cross-bin analysis is required [154].

$$DurDist_{state} = \sum_{d=1}^{144} \left| \sum_{d=1}^{d} \bar{P}^{mod}_{state}(d) - \sum_{d=1}^{d} \bar{P}^{tus}_{state}(d) \right| \tag{4.2}$$

where, $\bar{P}^{mod}_{state}(d)$ is the probability of a modelled state duration of d for state, *state* and $\bar{P}^{tus}_{state}(d)$ is the probability of a state duration of d for state, *state*, derived for the input time-use survey data.

Occupancy Profile Similarity Metric – the process used is generally known as the Levenshtein Edit Distance (LED) method for character string similarity analysis, which is used to compare individual occupancy profiles and is similar to the method used by Aerts et al [144] for clustering analysis. The derived metric is hereafter referred to as *ProfSim*.

The LED method is used to quantify the dissimilarity between two strings by quantifying the measures needed to transform one into the other. A 'cost' of 1 is assigned for each edit (insertions, deletions, and replacements) required in the transformation. For example, transforming 110111 to 001011 would require a minimum edit of a replacement of the first digit, insertion of the second, and deletion of the last digit - a total cost (*ProfSim* metric) of 3. The approach can therefore be applied when comparing two numerical profiles. When two profiles are compared, for clarity, the *ProfSim* metric is converted from a per-timestep to an hour equivalent by dividing the result by the number of timesteps per hour (i.e. six).

The metric can be used in two ways.

- First, it can be used to compare occupancy model output profiles with the input TUS dataset. The smallest *ProfSim* metric per profile, representative of the closest match, is determined, and an average calculated across all modelled days. This is a measure of the average similarity between generated profiles and the closest real profile.

- Second, each profile in either the input dataset or model output dataset can be compared with other profiles in the same dataset quantifying the behavioural similarity within and between each dataset.

There is no clear definition of when an input dataset, in terms of occupancy behaviour, is either overly similar or contains an unrepresentative population. Similarly, there is no clear delineation of the point at which the output results change from overly random to realistic or from realistic to narrowly replicating the input data. The *ProfSim* metric does, however, allow a relative assessment to be made.

4.5 Occupancy Sub-Model Development

The review of existing occupancy models in 3.3.3, determined that there were four main potential areas for improvement: optimising the number of occupancy states modelled; prediction of occupancy state durations; capturing couple, parent, and child occupancy interactions; and differentiation based on identified occupant and day type factors correlated with specific occupancy behaviours. Each are reviewed in the following sections.

4.5.1 Modelled States

The availability of appliance-level demand data for separate appliance-use calibration and the identified weak correlation between time-use survey activities and appliance use (see 3.3.3.3), determined that including time-use activities within the occupancy model, as is the case with several existing models ([120], [71]), would be less effective. To allow significant occupant differentiation also requires limiting the number of modelled occupancy states, to maximise the degree to which the dataset can be split while retaining the necessary depth to calibrate a statistically robust and representative occupancy model.

Models that distinguish between active occupancy and passive occupancy associated with sleeping periods, such as [135], allow for better assessment of occupant heat gains, standby power use, and the use of lighting while people sleep. Therefore, a three-state (sleep-active-out) occupancy model was deemed to be the simplest effective option for integration with a demand model calibrated directly using demand data.

4.5.2 Duration Prediction / Higher-Order Method Development

The fundamentals of Markov chain models and the concept of first- and higher-order methods were reviewed in 3.3. A key aim was to incorporate a higher-order approach if it could be shown to perform better than the basic first-order method, particularly for occupancy state duration prediction. As detailed by Wilke [71], Aerts et al [144], and McKenna and Thomson [155], first-order methods do not predict state durations accurately where the duration probability does not decay exponentially. As shown, sleep and working-day absence duration probabilities have complex distributions (see

	Sleep	Active	Out
Sleep	$P_{S \to S}$	$P_{S \to A}$	$P_{S \to O}$

First-Order

Current State / Duration	Sleep	Active	Out
Sleep / 0-2 hrs	$P_{S2 \to S}$	$P_{S2 \to A}$	$P_{S2 \to O}$
Sleep / 2-4 hrs	$P_{S4 \to S}$	$P_{S4 \to A}$	$P_{S4 \to O}$
Sleep / 4-6 hrs	$P_{S6 \to S}$	$P_{S6 \to A}$	$P_{S6 \to O}$
Sleep / 6-8 hrs	$P_{S8 \to S}$	$P_{S8 \to A}$	$P_{S8 \to O}$
Sleep / 8hrs+	$P_{S8+ \to S}$	$P_{S8+ \to A}$	$P_{S8+ \to O}$

Higher-Order

Figure 4.2. Transition from a first-order to a higher-order Markov chain model ('sleep' state example).

Figure 3.6).

First-order Markov chain methods have been the most commonly used for recent high-resolution occupancy and activity model development ([134], [120]). A first-order occupancy model has been developed for comparison purposes based on the standard method as defined in 3.3.1. This also incorporates the interaction and differentiation developments detailed in 4.5.3 and 4.5.4, and is hereafter referred to as the 'FOM' method.

Two higher-order methods have also been developed and their performance compared. One is an enhanced higher-order version of the Markov chain approach which includes probability data differentiated by ranges of current state duration. The other is similar to the discrete-event method developed by Wilke [71] that identifies the type of state transition and duration of the new state as single calculations, repeating this sequentially for the total required duration. Both are detailed in the following sections, followed by a performance comparison for the three defined methods.

4.5.2.1 Higher-Order Markov Chain ('HOM') Method

A new higher-order Markov chain method was developed where transition probability matrices (TPMs) have been generated according to the duration of the existing state within fixed ranges. This addresses the fundamental problem with first-order models in capturing occupancy states with complex duration probability distributions. This is hereafter referred to as the 'HOM' method.

In this case, each first-order transition probability matrix is replaced with matrices corresponding to, for example, sleep durations of 0-2, 2-4, 4-6, 6-8, and 8+ hours. So, if an occupant has been asleep for 3 hours then the 2-4 hour sleep duration transition probability matrix would be used to determine the next occupancy state. This approach captures the changes in relative probability of waking having slept for different lengths of time.

The difference in the per-timestep transition matrices for the first and higher-order approaches is shown in Figure 4.2. Each row of the first-order matrix transforms into a multi-row matrix. ('P' is the probability of a particular transition, 'S' refers to 'sleep', 'A' is 'active' and 'O' is 'out'). For example, the matrix element on the third row and second column represents the probability that someone who has been asleep for between 4 and 6 hours at a specific timestep will transition to the 'active' state.

Optimum duration ranges vary per day type and state transition based on specific behaviours, and, in particular, those related to sleep and work-related absences. Ranges were defined by examining the duration distribution for each transition and setting the ranges such that significant changes in behaviour are captured separately. For example, working days are characterised by a high probability of shorter (<3 hours) and longer (8-12 hours) absences, with two broad exponentially decaying curves, one from 0-8 hours and one from 8-14 hours (see Figure 3.6), and further distinct behaviours for specific ranges. The upper range bounds selected were 1, 2, 3, 5, 8, 12 and 24 hours to account for both the overarching and distinct absence duration behaviours.

4.5.2.2 Higher-Order Discrete Event ('HDE') Method

The event-based methods developed by Tanimoto et al [68] and Wilke [71] (see 3.3.2) provide an alternative approach that aims to capture state durations based on the analysed distributions from measured data. The Tanimoto et al method was discounted based on the performance issues identified in 3.3.2 and as it is also not clear how it could be developed for a three-state model given that it requires each state/activity period to be defined individually. However, the Wilke basis can be easily converted for any number of modelled states and a similar method was developed to allow method comparison. This is hereafter referred to as the 'HDE' method.

The developed HDE approach is similar but not identical to the Wilke occupancy model. Differentiation is achieved by creating separate probability matrices for each defined person type rather than combining regression factors for each occupant and day type characteristic (age, employment, day of week, etc.). This was done to ensure the approach was consistent with the FOM and HOM methods and due to uncertainty about the effectiveness of the statistical method used by Wilke for a three-state model. Instead of creating Weibull equivalent duration distributions as per Wilke, the actual distributions of durations were used both for statistical simplicity and to prevent excessive smoothing of the data. Apart from these changes, the method follows the basic logic of predicting each change of state and the duration of the subsequent state.

The UK 2000 Time-Use Survey (TUS) dataset [83] was used to derive probabilities

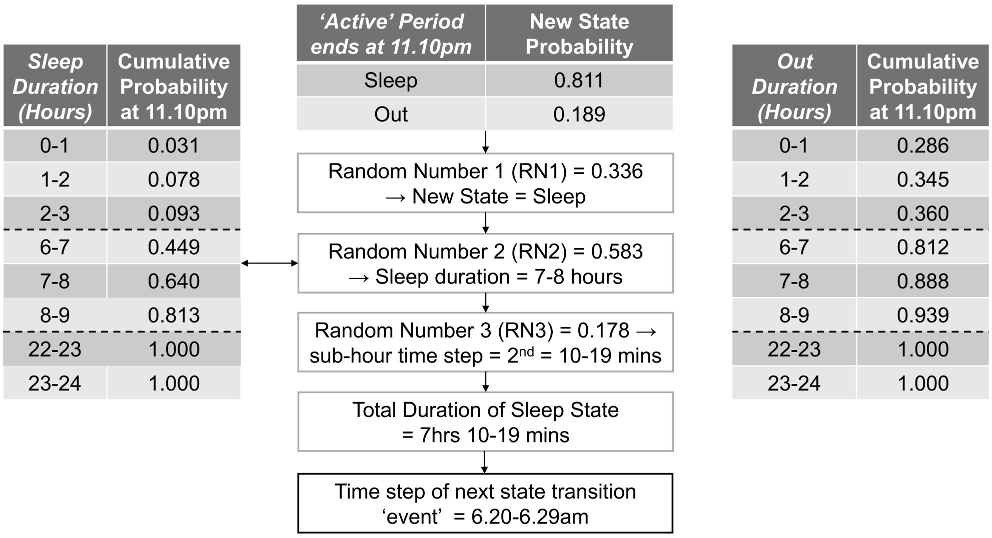


Figure 4.3. Higher-order discrete-event ('HDE') method next state type and duration calculation example.

at each 10-minute timestep for each potential occupancy state transition and the duration of each new occupancy state to the nearest hour. The sparseness of transition data requires that probabilities are averaged over a number of 10-minute timesteps. Therefore, data for each specific timestep is based on the average for that timestep and the three preceding and subsequent timesteps.

As an example of the method, Figure 4.3 shows the transition and duration matrices for an 'active' period that ends at 11.10pm. The HDE algorithm will first generate a random number (RN1) between 0 and 1 to determine if the transition is to 'sleep' or 'out' states. The duration in hours of the new state is determined in the same manner using the duration probability matrix for the new state if starting at 11.10pm. A third random number (RN3) determines with equal probability the exact 10-minute timestep on which the next transition occurs. The same process is repeated for this identified next event. Using this approach, a sequence of occupancy states and their durations with a 10-minute resolution are calculated.

Table 4.7
Occupancy model validation metric comparison for two single-person householder sub-populations on weekdays using different occupancy model methods.

Single Hhld Populations	Model	AO_Conv (x E-3)	AO_Var (x E-3)	DurDist Sleep	DurDist Active	DurDist Out
'Working 18-37'	FOM	0.6	14	2.26	0.79	2.87
'Working 18-37'	HOM	5.5	14	1.42	0.47	1.43
'Working 18-37'	HDE	41.6	45	3.17	0.87	2.87
'Over 76'	FOM	0.5	17	1.64	1.63	1.10
'Over 76'	HOM	5.0	18	1.26	1.15	1.03
'Over 76'	HDE	44.4	47	1.49	1.26	1.41

4.5.2.3 Occupancy Method Comparison

The performance of each of the three defined methods was compared using the validation metrics introduced in 4.4. Initial comparison was completed for single-person householders and the results for two sub-populations on weekdays are shown in Table 4.7. The 'Working 18-37' sub-population represents those between 18 and 37 years old working full-time and the 'Over 76' sub-population those over 76 years old.

The results show that the statistically simpler first-order method (FOM) converges more closely to the overall calibration data average (*AO_Conv* metric) in comparison

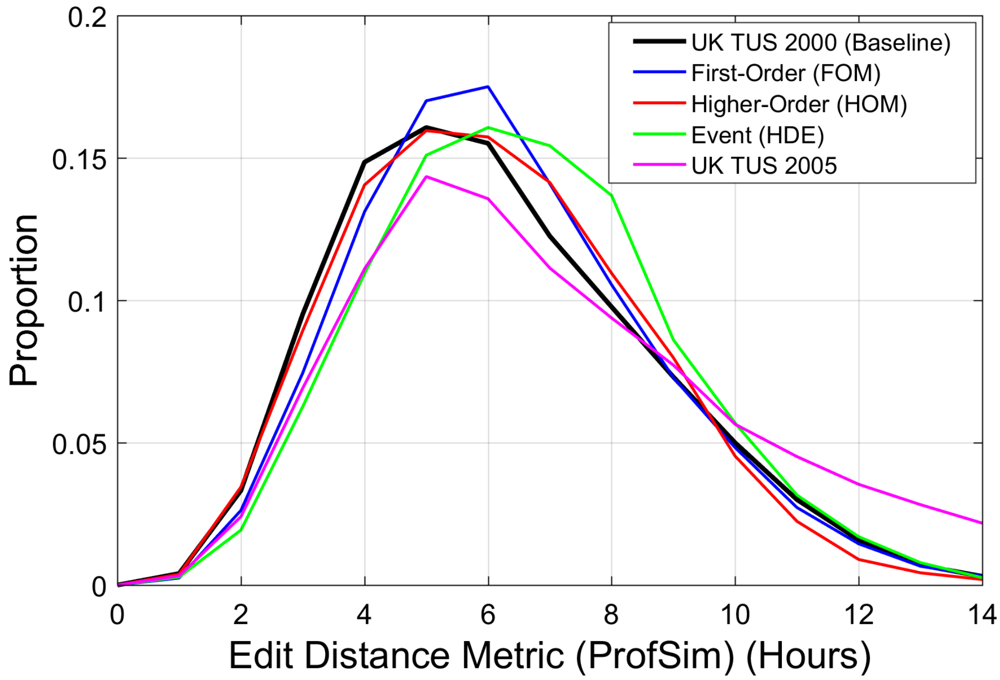


Figure 4.4. Edit distance (*ProfSim*) distributions for the TUS input datasets and each modelling method output for the 'Working 18-37' single-person householder sub-population.

with the higher-order method (HOM), which has additional statistical complexity resulting from the introduction of duration ranges. However, the weaker convergence of the HOM approach is not statistically significant, particularly when considered against the potential for variation between the TUS calibration data and real population occupancy characteristics (see Figure 4.10 and Table 4.14 for variation between the UK 2000 and 2005 [86] TUS datasets for equivalent populations). The *AO_Var* results, which only allow a comparative not absolute assessment of individual run consistency, do not show any significant increase for the HOM method, showing that individual annual FOM and HOM method outputs vary to a similar degree. Further analysis with extended period occupancy data is required to determine if this variance is realistic.

The HDE method shows significantly poorer performance for both *AO* measures, with an order of magnitude difference for *AO_Conv* compared to the HOM approach. That the *AO_Conv* and *AO_Var* metrics are similar magnitudes is a specific demonstration of poor performance, indicating a method that is generally poorly replicating the calibration dataset. Using metric *DurDist*, the HOM method shows better performance, suggesting it is better able to replicate actual distributions of occupancy state durations than either the HDE or FOM methods.

The occupancy profile similarity metric (*ProfSim*) was used for single-day profile analysis to determine the closest match between each TUS profile and the equivalent model output. This is a measure of how closely the range of actual profiles are replicated. For the 'Working 18-37' sub-population the average minimum *ProfSim* result for the HDE method was 1.98 hours, which compares poorly with 1.53 and 1.75 hours for the HOM and FOM methods respectively. The equivalent values for the 'Over 76' sub-population are 2.21(HDE), 1.92(HOM), and 1.98(FOM) hours, confirming the weaker performance of the HDE method, and indicating that the benefit of the HOM method is greater for populations with more distinct behaviours.

To indicate why the HDE method is less effective, Figure 4.4 shows the histogram distribution of edit distances when each profile is compared to the other profiles in the same dataset. Method performance can be gauged by how closely each method matches the TUS dataset distribution. Overall profile similarity is lower for the HDE method in comparison to both FOM and HOM methods. This can be inferred from the greater rightward shift from the TUS-derived target distribution, indicating higher

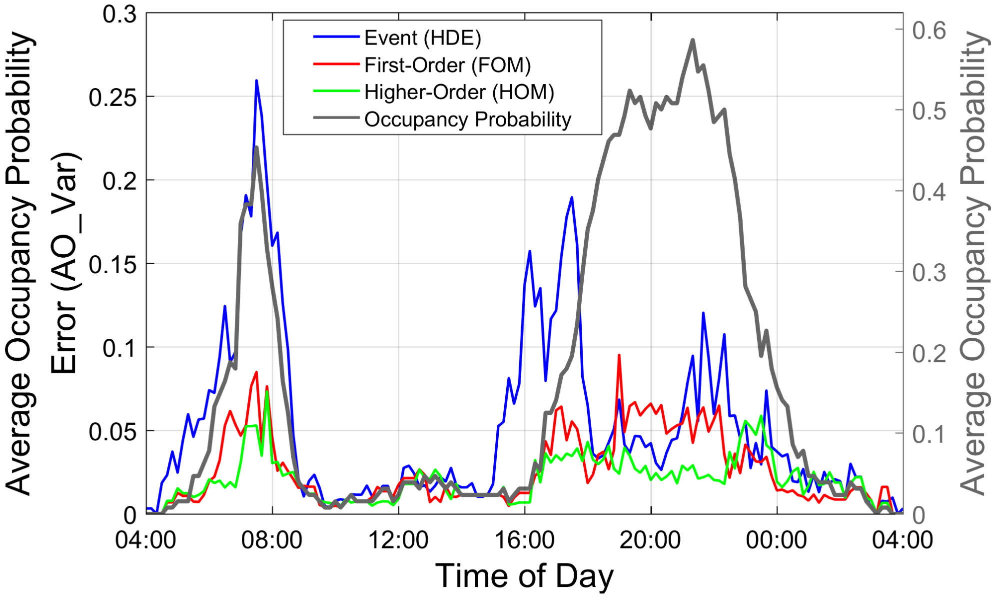


Figure 4.5. *AO_Var* error time dependency comparison for each modelling method for the ' Working 18-37' single-person householder sub-population. Occupancy data for analysis from the UK 2000 TUS dataset [83].

overall *ProfSim* values (and lower similarity).

That the HOM method performed better than the FOM method for duration prediction and individual profile replication but not for overall convergence was expected but the poor performance of the HDE method in comparison to both Markov chain approaches was not. Further investigation of this result was required.

Per-timestep *AO_Var* results were compared to determine if the source of the poor HDE method performance could be attributed to specific periods. Figure 4.5 shows the average *AO_Var* error calculated from 1000 annual model runs for the three methods for the 'Working 18-37' single-householder sub-population, with the majority of the error for all models concentrated in the morning and early evening periods. Both periods correspond to significant transitions in occupancy probability, which the HDE approach fails to capture as effectively as demonstrated by larger error peaks.

Comparative analysis was undertaken with the HDE method to determine if the weaker performance was a result of the number of adjacent timesteps or the duration ranges selected. Analysis with five and seven adjacent timesteps, and 20 and 30-minute duration ranges, showed no significant change, therefore the poor performance is inherent to the basic method.

One possible explanation for these results is that the HDE method does not have the self-correcting nature of a per-timestep probability model. The balance of this method is too focused on state duration prediction at the expense of time-specific state probability. Furthermore, not effectively tracking time-specific behaviour also compromises the duration prediction as demonstrated by poor duration (*DurDist*) and occupancy profile similarity (*ProfSim*) metric results. This is also illustrated by a detailed visual review of occupancy model outputs, which show an increased tendency for the HDE method to produce unusual behaviours (e.g. no daily sleep period, less distinct work-related absences, etc.).

Following this assessment, the HDE method was no longer considered and the following sections reviewing occupant interactions and differentiation only consider the performance of the FOM and HOM methods.

4.5.3 Occupant Interactions

As shown in 3.3.3.4, there are occupancy interactions between household occupants that are not properly captured by methods that combine two or more independent individual occupancy model outputs. A method was therefore required to account for this to avoid overestimating occupied period duration while underestimating multiple occupancy probability. This was focused on interactions within couple and family households due to the lack of data for other multi-person household types.

The Richardson et al occupancy model [134], which is based on number of occupants, broadly captures interactions but is not differentiated enough to capture specific relationships. The Baptista et al [138] model, using an interactive Markov chain approach, requires that a primary occupant be defined and modelled using a standard Markov chain model, with the transition probabilities for the second person determined by the state of the primary occupant. The problem with this approach is that it is difficult to define who the primary occupant is at each timestep and it does not consider whether the primary occupant has changed state which would impact the probability of a change of state of the secondary occupant.

4.5.3.1 Couples and Parents

A new method has been developed specifically for co-habiting couples (with and without children) that models each couple as a single entity, having a single combined occupancy state derived from both individual states. To minimise the data requirement, and assuming, as gender was not highly correlated with occupancy (see 2.5.2.1), that tracking specific individuals is not critical, the individual states are unassigned (e.g. sleep/active combines sleep/active and active/sleep etc.).

For couples without resident children, the average age of the couple was used as the age parameter as it was shown by analysis of the UK 2000 TUS dataset to be a better differentiator for occupancy than youngest or oldest individual age. Days with both individuals, one individual, and neither working were modelled separately.

This combined method was also applied separately to parents with resident children as they have distinct, child-influenced occupancy characteristics. It was also determined that the age of the youngest child was the strongest determinant of parent occupancy by comparing the relative influence of a variety of factors (average parent age, age of oldest parent, age of youngest parent, average child age, age of eldest child) on overall occupancy probability.

4.5.3.2 Children

The outlined primacy problem with the primary/secondary interactive Markov chain approach of Baptista et al for adult couple interactions does not hold for the parent-child relationship. Parent occupancy can be the primary element and the child model can be simplified and linked to the parent model output. The developed child model is Markov chain based, but is first-order and only tracks whether the child is active or inactive. For a child, sleep and out distinctions for the inactive state can be inferred based on time-of-day as children generally have highly consistent diurnal wake-sleep patterns.

Table 4.8
Relationship between adult and child occupancy transitions. Data for analysis from the UK 2000 TUS dataset [83].

Child Age/Day Type	Data Timesteps	All Adult Transitions	All Child Transitions	Adult & Child Transitions	Linked Transition (%)
All / All	183372	8251	6742	1402	20.8
8-11 / Non-Term	13080	559	437	109	24.9
8-11 / Term	48725	2248	1832	477	26.0
10-13 / Non-Term	14948	656	497	102	20.5
10-13 / Term	48824	2193	1837	372	20.3
12-15 / Non-Term	14192	633	491	72	14.7
12-15 / Term	43603	1962	1646	270	16.4

The method is similar to that defined by Baptista et al but the secondary (child) model transition probability is based on the parent state transition at the timestep rather than the updated parent state. Data analysis shows that there is a strong correlation between a parent occupancy transition and a simultaneous child transition, with a decreasing likelihood as the child age increases and evidence of an increased probability during school-term days (see Table 4.8). Overall, 20.8% of child occupancy transitions are linked to an adult transition which is higher than would be expected if random.

The 'primary' parent occupancy model is run first to determine the parent state at the next timestep. Then if, for example, at timestep, t-Δt, parent occupancy is

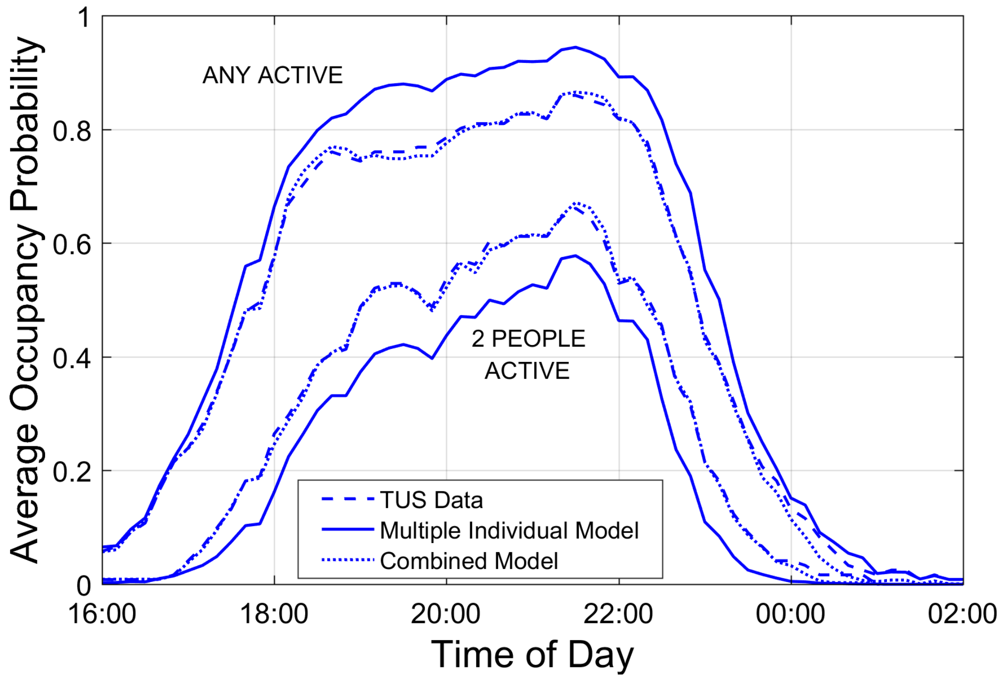


Figure 4.6. Impact of combined couple model basis on individual and overall active occupancy prediction for the 18-43 years old average age population with both individuals working.

active/inactive and the child is inactive and, at timestep, t, parent occupancy becomes both active, the selected transition probability matrix (TPM) for the child model is the one that determines whether the child remains inactive or becomes active if a second parent becomes active. Similar TPMs are available for all potential parent occupancy transitions (including no change), for both potential t-Δt child states (active/inactive).

4.5.3.3 Interaction Method Performance

Couple Households

Output from the combined couple occupancy model was compared to that from two individual couple householder models with the same characteristics. As an example, Figure 4.6 shows the evening period results for the 18-43 years old average age range population if both are working. The combined model average output was significantly closer to the equivalent TUS dataset results for combined occupancy. Similar results can be shown for the prediction of periods where one person is active and in the dwelling, and for other sub-populations. The results shown are based on the proven FOM basis, however, the HOM method results show equivalent performance.

Table 4.9
Average active occupancy prediction and occupancy profile analysis (*ProfSim*) comparison between combined and multiple individual model options for the 18-43 years old average age couple population with both individuals working.

Model	AO_Conv (x E-3) (Any Occ)	AO_Conv (x E-3) (Occ Num)	ProfSim (Hours)
2 x Individual First-Order	47.3	105.8	3.88
2 x Individual Higher-Order	42.2	99.4	3.38
'Combined' First-Order	14.2	18.6	3.28
'Combined' Higher-Order	15.3	24.1	2.89

Analysis with the average occupancy prediction (AO) metrics (see 4.4) is less straightforward for multi-person models as either simple active occupancy ($Any\ Occ$) or the specific occupant number ($Occ\ Num$) can be analysed. Table 4.9 shows the results for the average active occupancy variation metric (AO_Conv) analysis, considering both options, for working couples with an average age up to 43 years old. For the specific occupant number, the total error is the sum of the errors for single and double occupancy prediction compared to the input TUS dataset.

The results demonstrate both the improvement switching from independent to com-

Table 4.10
State duration analysis (*DurDist*) comparison between combined and multiple individual models for the 18-43 years old average age population with both individuals working. ('S'=Sleep, 'A'=Active, 'O'=Out)

Model	S-S	S-A	S-O	A-A	A-O	O-O
2 x Individual First-Order	3.53	1.33	0.85	1.54	1.44	2.97
2 x Individual Higher-Order	2.59	1.45	0.75	1.84	0.88	2.12
'Combined' First-Order	0.99	0.37	0.84	0.65	0.30	1.67
'Combined' Higher-Order	0.97	0.29	0.88	0.50	0.30	1.36

bined occupancy models for related adults, and, as also shown for the single-person householder models, that the HOM approach has weaker overall convergence to the dataset population average occupancy in comparison with the FOM approach. However, in this case the convergence performance difference is significantly smaller and again unlikely to be statistically significant.

The status duration comparison metric (*DurDist*) for the 'Working Couple 18-43' model (see Table 4.10) also shows a significant improvement using the combined model approach, and a more limited additional benefit from using the HOM method. In Table 4.9 results for occupancy profile similarity analysis (*ProfSim*) of the same population show similar relative performance.

Considering all results, both quantitative and visual representations, the combined, higher-order approach provides an improved method for predicting active occupancy for co-habiting couple households. The higher-order method performs better than the first-order for both duration prediction and similarity measures, and, as outlined, the weaker overall convergence to the average population behaviour is not statistically significant in relation to overall occupancy model accuracy.

Family Households

As outlined, the two-parent family occupancy model combines the method for co-habiting couples with a simple child model linking child occupancy directly with parent occupancy. The parent model exhibits similar metric performance as shown for couple models. Figure 4.7 shows that the parent model tracks the average total occupancy in all one-child households with good accuracy, with similar performance seen for other family sizes. As also shown, the child model tracks the input data reasonably well with some short periods of relatively weaker agreement (late afternoon, mid-evening). For single parent families, a specific individual occupancy model is used for the parent, and

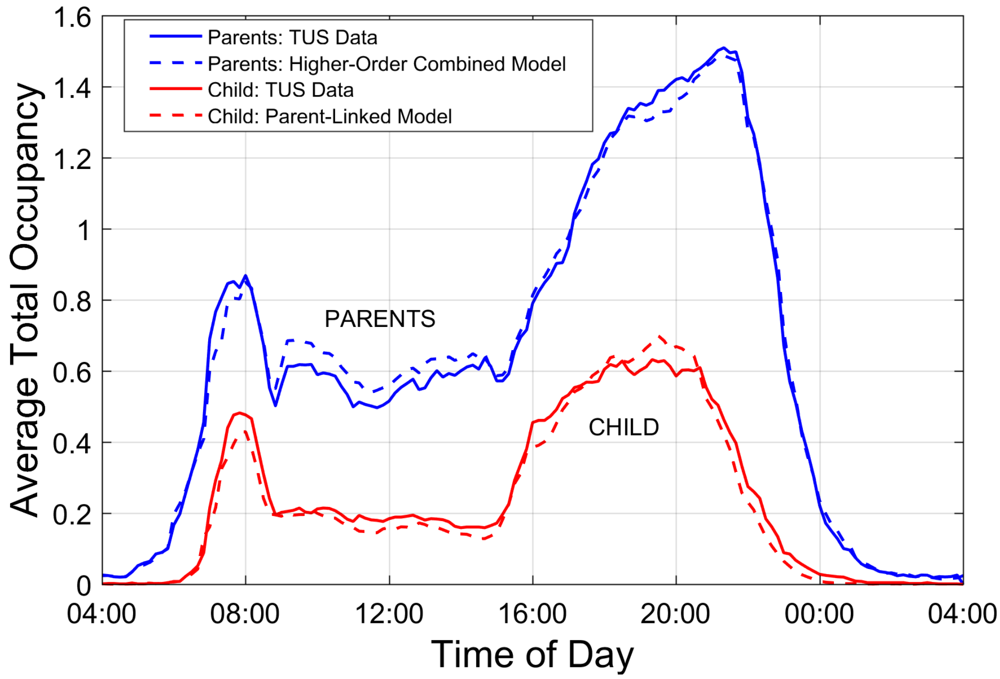


Figure 4.7. TUS data and higher-order combined model output comparison of average parent and child total active occupancy for all one-child households. TUS data for analysis from the UK 2000 TUS dataset [83].

the child model is linked to parent occupancy in the same manner as for two-parent families.

4.5.4 Differentiation

As reviewed in 3.3.3.2, existing occupancy models vary in the extent to which they differentiate between occupant and day types; with the majority incorporating only very limited differentiation. With little discussion and no consensus on the optimum approach, and the degree of effective differentiation being related both to the size of the calibration dataset and country-specific cultural influences, a review of the UK 2000 Time-Use Survey dataset [83] was undertaken as detailed in 2.5.2.1. To summarise: diary day working hours and age had a strong correlation with occupancy; employment status on non-working days, income and day type were weakly correlated; and location (population density), tenure, and gender showed no correlation.

Diary day working hours, age, and day type were selected for further review as primary differentiating factors. Whilst the day type correlation for average occupancy was weaker than expected, there are time-dependent differences which the average occupancy analysis did not capture. For example, in younger, single-person householder populations, non-working Saturday evening and Sunday daytime occupancy is lower than the equivalent periods on non-working weekdays (see Figure 2.15(a)).

Several other occupancy-related lifestyle characteristics were also identified as potential differentiators: significant variation in work weeks and working hours per work day in the TUS work-week diaries; extended absences, primarily due to vacations; sleeping away from home on TUS diary day; unusual work and sleep timing for a minority of TUS respondents; and that a significant number of people work from home on some or all working days [10].

Determination of the most effective level of differentiation required a series of assessments to determine the optimum group size to isolate occupancy behaviours while retaining sufficient data depth for calibration. Firstly, each highly correlated characteristic (occupant type, day type, working hours and age) and lifestyle characteristic (work-weeks, extended absences, home-working, and sleeping away) were reviewed in turn. Then, occupancy model performance using different levels of differentiation was compared. Finally, an assessment of the minimum population size for effective calibration and modelling was performed.

4.5.4.1 Occupant Type

The outlined occupant interaction method requires that couples, parents, and children be analysed and modelled separately. Single parents have different occupancy characteristics to equivalently aged single-person householders and were also modelled separately with the same linked child method as for dual-parent/child households (see 4.5.3). Within family households there is also a distinct population of older children between 16 and 24, either in education, working or non-working, that have occupancy patterns that are not closely linked to the parent occupancy. This population is therefore modelled separately as individuals within the overall family occupancy model.

Most of the remaining TUS diaries are for single-person householders, with the others associated with multi-adult households comprising unrelated adults or multi-generational family members. With too few of the multi-adult type for effective modelling, all adults in single-person and multi-adult households were modelled independently using the single-person householder population occupancy model basis.

Most individuals in the TUS dataset have normal diurnal waking patterns but there is a small subset with significant variations, particularly those with working hours in the evening and night period. These occupants must be differentiated within the occupancy model, primarily to avoid distorting the calibration of typical behaviour groups.

There is a range of different behaviours within this 'nightworker' group but only 330 diaries that fit the criteria, preventing this group from being further split. This group are assumed to be a mix of consistent nightworkers and those that work nights as part of a shift pattern. Based on analysis by Weston [156], 22% of men and 14% of women work shifts "most of the time". Of these, approximately 40% incorporate some element of night work ([157] and [156]). Four distinct night shift patterns have been identified as being most common; all nights (24.9%), nights every third week (28.4%), nights randomly 50% of time (39.8%), and nights 3-4 times in a two-week period (6.9%) [157]. There is also both an income and age correlation with nightworking probability, with lower values in both cases increasing the likelihood [156]. These working pattern and relationships were incorporated within the occupancy model.

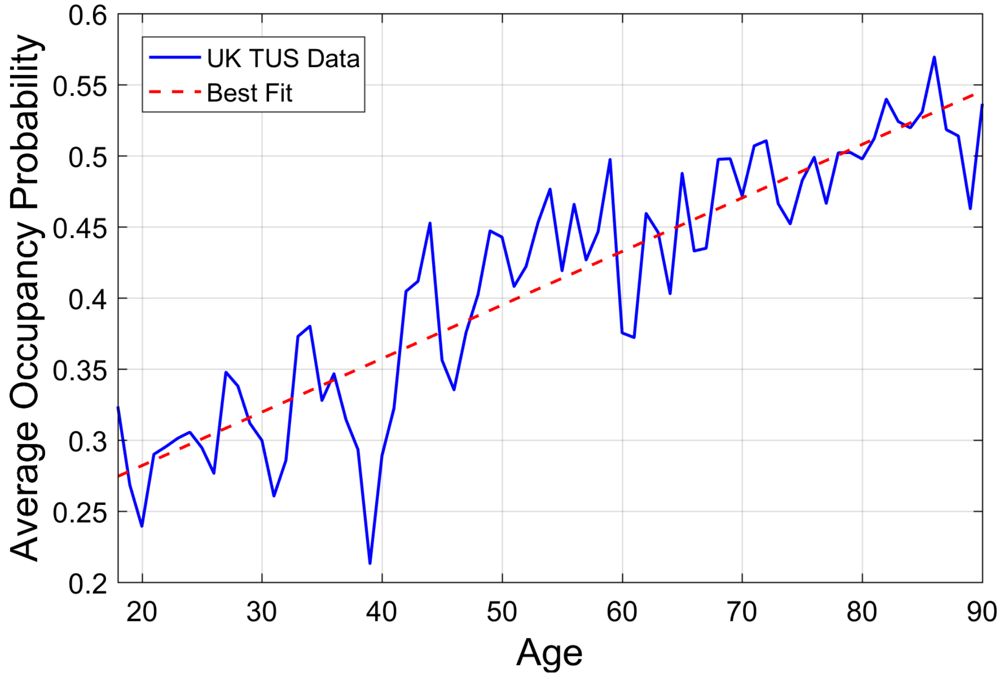


Figure 4.8. Variation of average active occupancy with occupant age for all single-person householders. Data for analysis from the UK 2000 TUS dataset [83].

4.5.4.2 Age

Figure 4.8 shows the relationship between average occupancy and age for all single-person householders on non-working days. It demonstrates that there is a consistent increase in average occupancy with age and also more variation in younger populations.

4.5.4.3 Day Types

The influence of working hours on occupancy shown in 2.5.2.1 determined that day types should, as a minimum, be split by diary day employment status. The overall and time-dependent effects associated with specific days of the week also need to be considered.

The extent to which existing occupancy models differentiate for day types varies. Widen et al [135] and Richardson et al [69] used separate models for weekdays and weekend days. Wilke et al [71] identified each day separately within the regression based factoring for occupant and day type variances, while acknowledging that this could be simplified.

The potential options for differentiating by day type are also driven by the UK 2000 TUS dataset basis. With one weekday and one weekend diary per person, the data is not evenly distributed. Analysis showed no meaningful variation in occupancy between Monday and Thursday. Friday evening has a different characteristic to other weekday evenings but there is insufficient data to capture this behaviour separately. All weekday data is therefore combined. Saturday and Sunday both have distinct overall patterns and due to the proportionally greater number of diaries can be modelled separately for non-working occupants.

4.5.4.4 Working Hours

Given the range of employment durations it is not straightforward to differentiate by working hours. Demarcation using terms such as ‘full-time’ and ‘part-time’ is overly simplistic. Therefore, average active occupancy for all working-age single-person householders was compared with working hours to determine if there was either a clear correlation or step changes in active occupancy associated with specific durations.

Figure 4.9 shows that people with greater than zero and up to 6 hours working

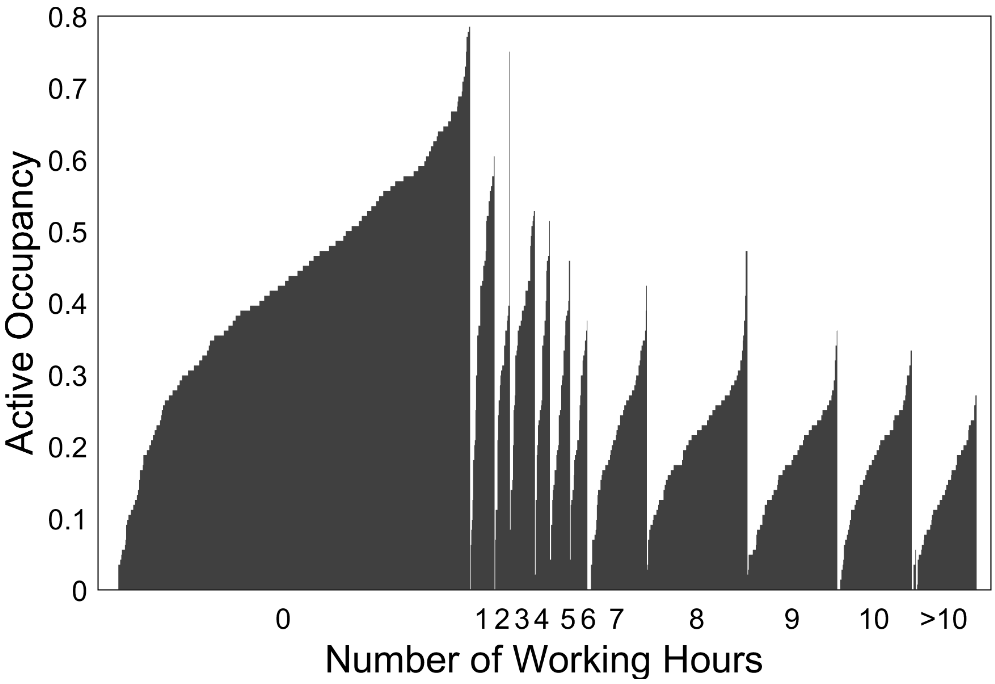


Figure 4.9. Diary day active occupancy distribution by number of working hours for all working age, single-person householders. Data for analysis from the UK 2000 TUS dataset [83].

time are relatively rare, active occupancy falls most markedly between 0 and 4 worked hours, and is relatively consistent as working hours increase beyond 4 hours. Only days with greater than 5 working hours are therefore defined within the developed occupancy model as 'working days'. Days with fewer working hours are defined as 'non-working' as this has a limited influence on occupancy probability, and there are too many conflicting working patterns and too few relevant diaries for effective modelling of specific behaviours. There is therefore a distinction made between 'part-time' workers that work a small number of >5 hour days and those that work a small number of hours on most days. The former are modelled as 'working' on each >5 hour working day, the latter as 'non-working' on each <5 hour working day. The probability of the two part-time options was determined from analysis of the TUS work diaries as detailed in the following section.

For each adult, the occupancy model determines if they work full-time, part-time, or are non-working on each modelled day. 'Full-time' is defined as a minimum of 4 'working' days per week and 'part-time' as 1-3 'working' days. For couples and parents, all possible two-person combinations of the three employment options are incorporated as shown in Table 4.5. Work weeks are then defined probabilistically for each individual.

4.5.4.5 Work Weeks

The UK 2000/1 Time-Use Survey includes one-week duration working diaries for each applicable individual which include the number of hours worked per day. This has allowed the distribution of typical working weeks to be determined and probabilistically allocated to each modelled occupant.

As outlined, 'full-time' workers are defined as having a minimum of 4 working days with a minimum of 5 hours worked on each 'working' day, and for workers designated as 'part-time', only those days with more than 5 hours worked are modelled as 'working' days. Fewer than 10% of part-time workers in the TUS dataset did not have at least one 'working' day. The overall distribution of working weeks for full-time workers is shown in Table 4.11, with similar data for part-time workers for days with at least 5 hours worked also compiled.

Table 4.11
Full-time (minimum of 4 worked days) work week probabilities in the UK 2000 TUS dataset work-week diaries. Data for analysis from the UK 2000 TUS dataset [83].

Work Week	Probability(%)	Work Week	Probability(%)
4 Weekdays	9.4	4 Weekdays + Sat	4.5
5 Weekdays	46.2	4 Weekdays + Sun	2.7
5 Weekdays + Sat	13.2	4 Weekdays + Sat/Sun	4.3
5 Weekdays + Sun	4.3	3 Weekdays + Sat	1.5
All Days	9.5	3 Weekdays + Sun	0.8
		3 Weekdays + Sat/Sun	3.5

4.5.4.6 Extended Absences

The single-day diary basis of time-use surveys do not allow them to be used to determine the frequency and duration of extended absences. The majority of extended absences are assumed to be related to vacations, with employment related absences potentially significant in single-person households. Data on the latter is scarce, therefore the occupancy model is only currently calibrated for vacation absences.

The probability of taking a vacation is determined based on regression factors generated by Mergoupis and Steuer [158] for UK-specific behaviour as part of a Europe-wide study. A vacation in this study is defined as being an absence of at least four days. The factors account for adult and child ages, number of occupants, relationships, education, and income. The study does not identify the potential for multiple vacations per household, therefore the model currently assumes only one annual vacation per household, which potentially underestimates total absences in some households. In addition, the lack of data prevents the frequency of shorter absences than four days from being simulated, which is also likely to contribute to an overestimate of total occupancy for extended period models.

The timing (4-periods: Dec-Feb, Mar-May, June-Aug, Sep-Nov) and duration of the vacation-related absences was determined from the BDRC 2014 Holiday Trends report [159]. As an example, for family households, there is a 54.9% probability that the main vacation is in the June-August period and a 23% probability it is between 11 and 14 nights.

4.5.4.7 Homeworking

The number of homeworkers in the UK increased from 2.9 million in 1998 to 4.2 million in 2014 [10]. TUS diaries do not explicitly state working location but it can potentially be determined from the location in adjacent time periods. Where occupants transition directly from home-based activities to work and vice versa, they can be assumed to be working at home. However, as the number of identified homeworkers is low, they cannot be modelled as a separate population. Homeworking probability is therefore assigned based on UK government statistics.

UK Population Survey data [160] identifies homeworkers, and whether the home is the main working location or used as an occasional base. From this data it was determined that 4.9% of workers work mainly from home (of which 3.2% are self-employed and 1.7% are employees) and 8.8% are based at home but work in different locations. In addition, it has been determined that an additional 8.3% of people occasionally work from home [161].

Within the model the homeworking probability and type per individual is assigned on the above basis with further manipulation based on income decile using data from Bloom et al [162]. The probability that a working day is spent at home has been set arbitrarily per group, with the self-employed group set randomly between 60 and 100%, the employee group between 40 and 80%, the 'base' group between 0 and 40%, and the occasional group between 0 and 30%, based on assumed typical patterns. More detailed homeworking behaviour data would be useful to improve the calibration of this element.

The homeworking element is incorporated as a secondary model as detailed in 4.7.3. On designated homeworking days, the occupancy model converts any 'out' periods immediately pre- and post-working periods to 'active' periods. Whilst it would be expected that homeworking would impact the use potential of certain appliances (e.g. computers), this cannot be determined from existing data.

4.5.4.8 Sleeping Away

The TUS dataset includes individuals that transition directly from an 'out' state to 'sleep'. A proportion are assumed to be sleeping away from home. McKenna et al [163] identified this potential and modelled ‘sleeping away’ as a primary state in a four-state Markov chain occupancy model.

As for work periods, the location of the ‘sleep’ activity is not explicit in the TUS data. A similar location determination is therefore required based on the location in adjacent periods. As per McKenna et al, for each individual an assessment was made based on the final pre-sleep activity. For example, ‘travelling’ was assumed to be returning to home and an indoor activity at another location was assumed to reflect someone sleeping away. From this analysis, a probability is determined for whether a direct ‘out’ to ‘sleep’ transition reflected sleeping away rather than simply going directly to sleep on returning home.

Unlike McKenna et al, the developed model incorporates this element as a secondary function to the primary three-state model. Sleep location is assumed not to change during a sleep period and a single probability determination is therefore sufficient. The potential for sleeping away is included as a fixed probability per differentiated occupant and day type. Whilst it is assumed that the potential for ‘sleeping away’ will vary within each occupant type group, there is currently insufficient data to factor the potential at this level.

4.5.4.9 Differentiation Performance Comparison

To analyse the impact of using smaller, differentiated occupant populations for occupancy model calibration, an overall single-person householder occupant model (representative of the models developed by Richardson et al [134] and Widen and Wackelgard [70]) is compared to a model calibrated using two smaller, single-person householder sub-populations from the TUS dataset: ’Working 18–37’ – working individuals between 18 and 37 years of age, and ’Over 76’ – retired individuals over 76 years of age. A first-order Markov chain method was used, with 100 1-year duration, 10-minute timestep occupancy state sequences generated for each case. The first-order method was chosen as it is both a proven method and allows direct comparison with the equivalent existing models.

The results were analysed using the metrics identified in 4.4. The average active occupancy variation metrics (AO_Conv and AO_Var) were used to determine the overall and per-sequence convergence by comparing the mean error between the pre-

Table 4.12
Active occupancy and state duration (*DurDist*) metric analysis for different first-order model calibration and validation populations.

Calibration Pop.	Validation Pop.	AO_Conv (x E-3)	AO_Var (x E-3)	DurDist Sleep	DurDist Active	DurDist Out
All single-person	'Working 18-37'	222.8	223.9	8.20	9.72	22.43
All single-person	'Over 76'	133.3	134.4	4.13	7.11	8.93
All single-person	All single-person	0.52	20.0	1.73	1.13	4.04
'Working 18-37'	'Working 18-37'	0.36	14.3	2.26	0.79	2.87
'Over 76'	'Over 76'	0.57	17.2	1.64	1.63	1.10

dicted active occupancy per modelled annual sequence and that found in the input dataset (see Table 4.12). They confirm that the first-order Markov chain model, calibrated using the smaller, differentiated datasets, produces occupancy behaviour that is more representative of those sub-groups, as opposed to the models calibrated using the wider population datasets. For the differentiated datasets, the per-sequence measure (AO_Var) is significantly higher than the overall measure (AO_Conv), indicative of a method that produces some variation between individual runs but convergence overall. Where the wider datasets are used, both metrics are significantly higher and a similar magnitude, indicative of a method where neither individual runs nor the overall average output are representative of smaller groups within the population.

The state duration prediction comparison (*DurDist*) between the models calibrated using the same populations is also shown in Table 4.12. Significant improvements are again demonstrated where both the occupancy model input data and comparison TUS data are from the same population. There is also a further general improvement in this metric for the differentiated populations. More importantly, the results show that the overall single-person householder population significantly fails to properly replicate the range of durations for the two smaller sub-populations.

The *ProfSim* metric was used to identify the closest match (lowest edit distance) for each day in the model output from the differentiated and wider population models in the input TUS data. This allows an assessment of the model's ability to generate realistic profiles. The average minimum edit distance (expressed as a time) for the 'All single-person' model compared to the 'Working 18-37' TUS dataset is 4.35 hours. The result when the 'Working 18-37' specific model is used is 1.75 hours. The equivalent improvement for the 'Over 76' sub-population was from 2.71 to 1.98 hours. This suggests that the expected improvement from differentiation reduces for populations with less distinct behaviour patterns, but is significant in all cases.

Overall, there is a significant improvement in the first-order Markov chain model's ability to replicate observed behaviour using smaller sub-populations. The degree of improvement is dependent on the deviation of each sub-population from the overall population average. Similar analysis for the higher-order (HOM) method confirmed similar benefits resulting from increased differentiation of the calibration populations.

4.5.4.10 Minimum Population Size

Prior to determining the final basis for occupant differentiation, it is necessary to determine the minimum size of population necessary for effective model calibration and then to compare this with the typical population sizes resulting from differentiation by the identified occupancy-correlated factors. Age is the main differentiator that is not driven by other elements of the occupancy model development and is uniquely not a simple binary determinant, therefore the analysis focuses on performance in relation to age range selection.

A variety of methods were used to attempt to identify the minimum population size required to produce a robust statistical model. The higher-order Markov chain (HOM) method described in 4.5.2.1 was used for the analysis as it is the most sensitive to population size due to the incorporation of multiple duration ranges for each state.

Two single-person householder sub-populations were selected from the UK 2000 TUS dataset for analysis (ranked by age with youngest first); the working age population on non-working days and the retired population. Datasets of 5, 20, 50, 100, 150, 200, and 400 were selected from the top of the ranked list (i.e. 1-5, 1-20, 1-50 etc.). For the working age population, the final person in each dataset was 19, 19, 21, 23, 29, 37, and 60 years old, and for the retired population was 65, 65, 66, 68, 70, 72, and 78 years old, respectively.

Edit distance (*ProfSim*) analysis allows individual results within datasets to be compared for similarity. The expectation was that very small calibration datasets would result in overfitting to the input data and that as dataset size increases improved performance from increasing probability data depth would gradually be offset by the age-related differences in behaviour shown in Figure 4.8. The analysis was completed for each of the age-ranked TUS datasets using occupancy models calibrated with the relevant dataset. In this case, each dataset and model output profile was analysed against profiles in the same set of data, with performance assessed based on the similarity in the distribution of results for each dataset and associated model output.

For both populations, up to the 50-person dataset, the model output generates a significantly higher number of closely matched profiles (to within three 10-minute timesteps) than the actual data. Above 50-person groups, the performance for each population diverges. For the retired households, the closest match between dataset and model characteristics is for the 150-person model, although the performance remains broadly consistent between 150 and 400. For the working age population, there is a close match for 100 persons and a better result for 400 persons, but the overall results are more erratic. As shown by Figure 4.8, there is significantly more variation in active occupancy by age for the working age than the retired population, and also a wider range of ages per ranked group, therefore more consistent results for the retired groups would be expected.

The results indicate that with a minimum of 100 input TUS diaries, the model output is broadly consistent with the input data without overfitting. Optimum size on this measure may, however, be population dependent, with a minimum size of 100-150 diaries required to ensure no overfitting, but with larger datasets being preferred for populations with less consistent behaviour as the benefits of capturing the overall behaviour range outweighs any age-specific behaviours captured from smaller populations. Further assessment when the UK 2015 TUS survey is released will allow this to be further investigated.

Two further assessments were used to determine the potential for producing statistically robust models. One was to review the number of elements in the generated Transition Probability Matrices (TPMs) (see 3.3.1) with a fractional value. A zero value indicates that there was no individual with that specific state transition and a value of one is typically associated with the behaviour of one person (and is therefore not necessarily representative of wider behaviour). A fractional value requires multiple people to be represented and the number of such elements can be used as a proxy for probability data quality, and it is assumed, by extension, model stability.

The other was to review the number of times an annual higher-order model had a state and duration range that did not have associated probability data and required a recovery function to be used. This can occur due to the use of duration ranges rather than specific durations in the model calibration resulting in scenarios not seen in the input data. This is also an indirect measure of data quality as reducing the use of the recovery function requires an increasing likelihood of probability data for transitions in adjacent duration ranges. Example results for both measures for the same working age population as above are shown in Table 4.13.

Table 4.13
Fractional TPM elements and recovery function use frequency for different transition probability input datasets sizes (working age, single-person householders on non-working days).

Households in Dataset	Fractional TPM Probability Elements (out of 9072)*	Timesteps Recovery Function Required (x10^{-3} %)
50	531	1.81
100	834	0.93
150	1084	0.64
200	1394	0.42
400	1891	0.31

* There are a large number of unlikely transitions therefore the number is low compared to total elements.

For both measures there is a diminishing performance improvement as the number of input diaries is increased, with the most significant improvement up to c. 200-300 diaries. The target number of diaries per calibration population was therefore set at 200-300 diaries, with a minimum of 150 where justified by distinct behaviour differences, as these performance benefits outweigh the potential loss of specific behaviour replication and given that the results of the *ProfSim* analysis for larger calibration groups did not indicate a significant loss of replication performance. The potential for larger population sizes, covering wider age ranges, to generate overly similar output is partially addressed by using overlapping age ranges to increase the number of distinct calibration populations as detailed in the following section. The residual influence of the selected population size on model convergence to average behaviours is further discussed in 4.6.3 and Chapter 7.

4.5.5 Occupancy Sub-Model Structure Summary

4.5.5.1 Occupant and Day Type Modules

The developed model integrates the three basic modules; individual, couple/parent, and child, outlined in the preceding sections, with further differentiation first by day type (working/non-working) and then by age range.

To make effective use of the available data in achieving optimum calibration population sizes of 200-300 diaries, the UK 2000 TUS dataset was further split into overlapping age ranges for all modules. For working single-person households, for example, the 18-37 years old range TUS population was used for the 18-33 year old range model data, the 28-44 population for the 34-40 model, etc. This increases the number of diaries per population group, while also allowing for a larger number of specific age ranges to capture different average behaviours within the constraints of the 200-300 diary target. It also recognises that the age-related behaviour changes are gradual (see Figure 4.8) and single day diaries may not adequately capture the range of behaviours within groups.

The individual module is used for single-person householders, and for individuals in multiple unrelated adult households or households with related adults of different generations (e.g. adult children). It has seven age ranges from 18-33 to 80+ and two further modules for young adults living in a family household; 16-18 year olds in education and a general 16-24 age group, with both working/term and non-working/term day models.

The couple/parent module has separate probability data for couples and parents. The couple dataset has seven age ranges based on average age. The parent dataset has four ranges based on the youngest child's age. The child module has five age ranges (5-7, 8-9, 10-11, 12-13, and 14-15) and also differentiates for school-term and non-term days. Under 5's are not modelled due to the lack of TUS data for infants, with infant occupancy assumed to track that of the parents.

Different module combinations can be used to replicate actual household types. For example, a family household with one adult child and one under-16 child requires a parent module, an 'individual' module for the adult child, and a child module linked to the parent module output.

Separate transition probability matrices (TPMs) have been generated for each defined occupant type, age range, for each day type (weekday, Saturday and Sunday), and whether the occupant is working or non-working. For couples and parents, there are three options; both working, one working and neither working. As weekend working is less common, both Saturday and Sunday data for multiple age ranges were combined to generate sufficiently sized calibration populations.

For a full list of all modelled populations and the associated calibration populations used, see Appendix B.

4.5.5.2 Occupant Calendars

Calendars are defined for each modelled individual, couple/parent, and child to reflect the sequence of day types through the modelling period. The model selects the appropriate TPM for the required day type as determined by the calendar. Workers are allocated typical working weeks as detailed in 4.5.4.5. For school-age children the occupancy model includes typical term dates. The model can therefore clearly distinguish different characteristic occupancy behaviours for each occupant type (full-time workers, stay-at-home parents, students, school children etc.), that is a key precursor to demand prediction for each household type.

4.5.5.3 Time Resolution

The UK 2000 TUS dataset diaries are completed for each 10-minute period. This dataset resolution restricts the model calibration resolution to the same 10-minute basis. In order to allow a 1-minute resolution for the overall demand model to be achieved, the occupancy model output is converted to a 1-minute basis by assuming that the start and end of an occupancy state occurs randomly within the identified 10-minute period. As the HES dataset demand data does not show any significant correlation between appliance use and each 2-minute segment per 10-minute period, the assumption of random transitions in the occupancy model does not significantly influence the demand model accuracy. The decision to limit the demand model to a 1-minute resolution is detailed in 5.2.3.

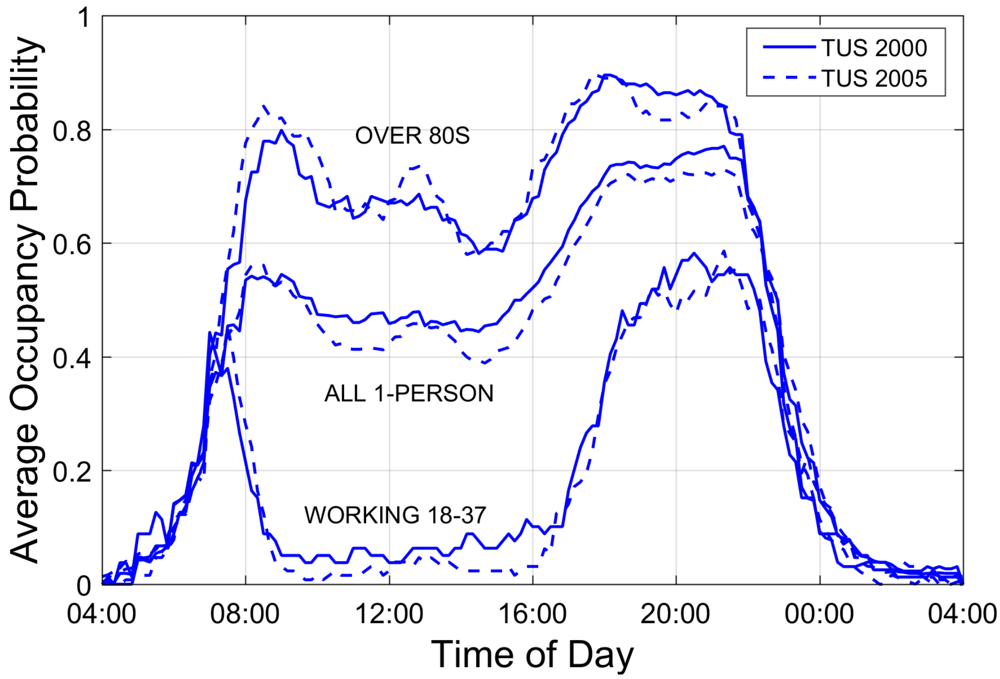


Figure 4.10. Single-person householder average active occupancy comparison between the UK 2000 [83] and 2005 [86] TUS datasets.

4.6 Final Method Selection

4.6.1 Independent Dataset Performance Comparison

For further comparison and validation of the first- (FOM) and higher-order Markov (HOM) methods (see 4.5.2), the results were compared with occupancy profiles from the smaller UK 2005 TUS survey [86]. This dataset also captures the required occupancy data at a 10-minute resolution, with 4941 diaries compared to 20981 for the UK 2000 TUS survey.

For validation purposes, both TUS datasets should capture similar occupancy behaviour. Figure 4.10 demonstrates that the average weekday profile for the overall single-person householder population and two smaller sub-populations (under 37 years old on working days and over 80 years old) are broadly consistent. This confirms that there are occupancy traits that are inherent to the TUS sub-populations, which was also confirmed for other occupant and household types.

If the developed occupant modules are representative of both the calibration dataset and overall occupant behaviour, then there should not be a significant difference in performance, as measured by the identified metrics (see 4.4), between the results for the model output when compared to each TUS dataset and the equivalent metrics for the comparison between the two TUS datasets.

Table 4.14
Occupancy model validation metric results for the UK 2000 TUS dataset [83] and Markov chain model methods compared to the UK 2005 TUS dataset [86]. (FOM=First-Order Markov, HOM=Higher-Order Markov)

Population	Dataset 1	Dataset 2	AO_Var	ProfSim (Hours)	DurDist Sleep	DurDist Active	DurDist Out
'Working 18-37'	TUS 2000	TUS 2005	4.69	11.9	2.55	1.27	4.20
'Working 18-37'	FOM	TUS 2005	4.75	12.1	2.67	1.23	3.25
'Working 18-37'	HOM	TUS 2005	5.15	11.7	2.17	1.23	4.19
'Over 76'	TUS 2000	TUS 2005	4.15	12.4	4.50	3.44	3.76
'Over 76'	FOM	TUS 2005	5.21	12.3	4.40	3.35	3.49
'Over 76'	HOM	TUS 2005	4.91	12.3	4.72	3.58	3.83

The results in Table 4.14 show analysis of the average active occupancy (AO), duration prediction (*DurDist*), and profile similarity (*ProfSim*) metrics. The results for both the first- and higher-order methods compared to the UK 2005 TUS dataset

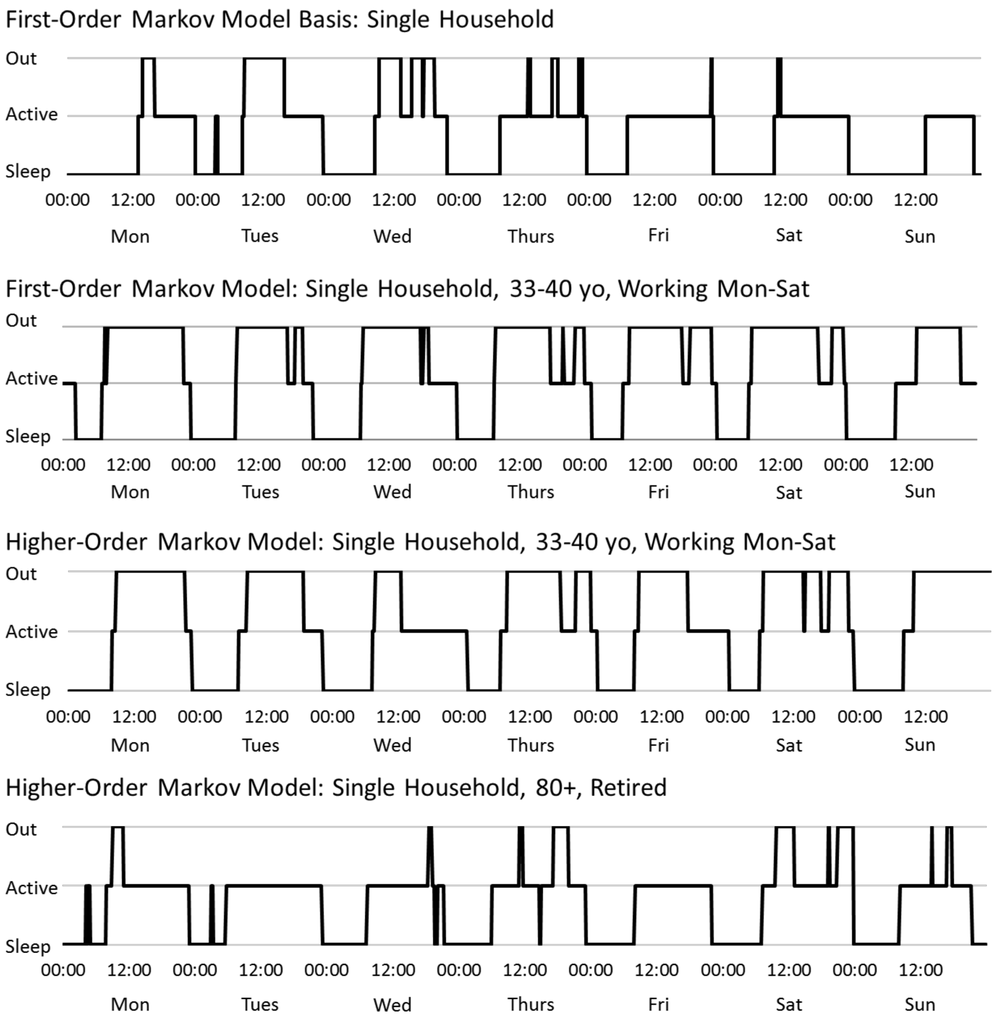


Figure 4.11. Example one-week individual occupancy state profiles for various modelled single-person householders.

are broadly consistent with the results between the two TUS datasets. This demonstrates that the developed occupancy model is, as a minimum, no worse at predicting occupancy for independent datasets than the UK 2000 TUS dataset.

The results are less conclusive regarding the performance of the higher-order method relative to the first-order approach when compared with the 2005 dataset. Both methods perform slightly better on some measures, and worse on others. The 2000 dataset may be too small to produce wholly representative data for the sub-populations. Alternatively, there may be an inherent weakness in the metrics used to differentiate relative performance at this level of similarity. Further analysis with the larger 2015 TUS dataset will be required for a better judgement of the higher-order model benefit relative to real occupancy variations rather than merely for replication of the calibration data.

4.6.2 Method Performance Analysis Summary

4.6.2.1 Markov Chain Method Performance

The overall benefits of the higher-order method (HOM) model compared to the first-order method (FOM) are not yet conclusive. There is a measurable improvement in the metrics for duration prediction and similarity to actual TUS profiles, especially for groups with consistent patterns of behaviour (e.g. workers). In comparison with the independent TUS dataset the results were less clear. However, there is sufficient justification to use the higher-order method for further development as there is evidence that the residual issues are related to current data availability rather than the basic method.

4.6.2.2 Differentiated Model Performance

The primary output from the developed occupancy model is a per-timestep sequence of occupant states. Whilst the validation metrics used allow the differences between profiles to be quantified, visual analysis of actual profiles can also demonstrate model effectiveness.

Figure 4.11 shows results from randomly selected occupancy model runs. The results compare a Mon-Sun sequence for a single-person householder using an undifferentiated

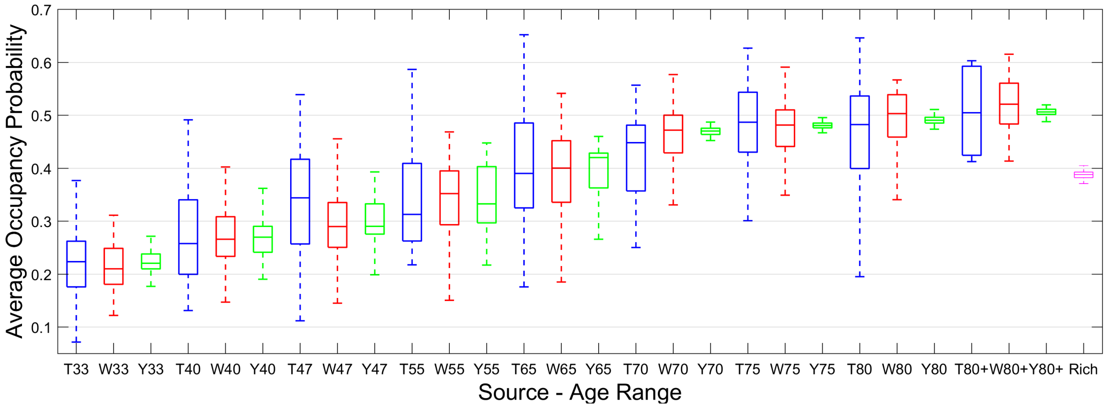


Figure 4.12. Per-individual active occupancy range comparison for the 1-week Dutch 2005 TBO TUS dataset [87] and equivalent higher-order Markov chain occupancy model output. (T=TUS, W=1-Week Model, Y=1-Year Model, Number=Upper Age, 'Rich'=Richardson et al) (Box=25%/50%/75%, Limits=0.3% and 99.7%)

first-order Markov chain approach (similar to Richardson et al [134]), a first-order approach for a younger working householder, and a younger working and an older retired single-person householder using the developed higher-order Markov chain method. The smaller populations were deliberately selected for a strong likelihood of longer 'out' and 'active' periods respectively.

The undifferentiated, larger population model shows no overall consistency between modelled days. This gives credence to the assertion that this type of model generates profiles that are an unrepresentative composite of multiple, conflicting behaviours. The developed differentiated higher-order model more consistently models sleep durations within the most likely duration range, shows daily 'out' periods consistent with a working person, and long 'active' periods consistent with an older retired person. The differentiated first-order model shows the more consistent occupancy pattern associated with a working person, but there is evidence of more erratic sleep durations and very short duration absences under close inspection.

4.6.3 Applications and Limitations

Any TUS-based model has inherent limitations as a result of the typical single-day per individual basis of the diaries used for calibration. Over multiple annual model runs, the developed higher-order Markov chain method does provide some degree of variability in overall average active occupancy (e.g. +/- c.10% for working, single-person householder models). However, the degree of variation is likely to be less than in reality because of the tendency for probabilistic models to converge to the calibration data average. This is shown in Figure 4.12, where the variation in average occupancy per individual is compared between the Dutch 2005 TBO TUS dataset [87] with 1-week duration occupant diaries, and the equivalent 1-week and 1-year duration model output, for each single-person householder model population. The model output, for the equivalent 1-week timescale, shows significantly less variation, and over a 1-year timescale, extreme convergence in some cases.

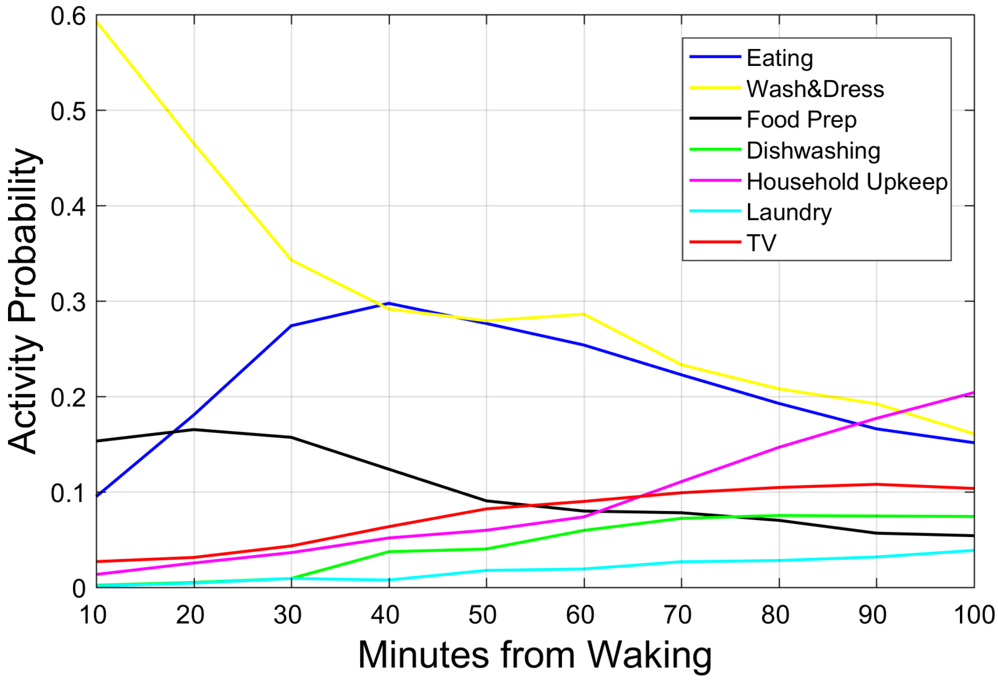
What cannot yet be determined is to what extent the 1-year behaviour of real people converges from the presumably more erratic nature of 1-week data. However, the model convergence is especially marked for the retired populations and it would be expected that real behaviour is significantly more varied than shown. This also highlights that the variation in working age households is predominantly the result of different day types and work weeks, with the convergence of each retired population the result of all occupants being modelled with the same set of day types. Further analysis has shown that the convergence for each working age day type model is similar to that shown for the retired households.

The analysis presented here, and by others ([144], [71]), has clearly demonstrated that there are broad occupancy patterns related to identifiable occupant types. Existing models have generally taken a cautious approach to differentiation with calibration data depth for model stability taking precedence over identifying specific occupant traits. Reviewing model performance for different sizes of calibration populations has shown that more aggressive differentiation does not result in significant reduction in model stability. Whilst Figure 4.12 shows a residual underestimation of in-group variation as a result of convergence to average behaviour within each occupant type module, it does show that the developed model tracks the between-group variation that is absent from most existing undifferentiated population occupancy models. This is highlighted by the Richardson et al model result ('Rich') based on annual output from the undifferentiated single-person householder module which shows tight convergence to the overall average for all such occupants.

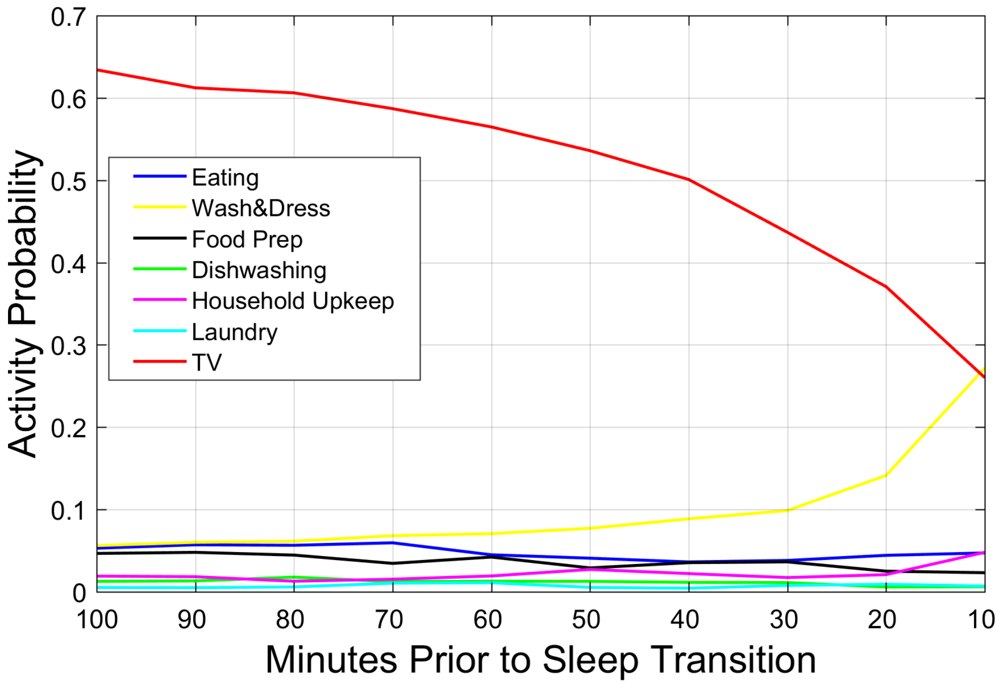
The occupancy model has been developed primarily to generate input occupancy data for a high time resolution, occupancy-driven energy demand model, with the aim to identify specific demand patterns for homogenous communities (e.g. retirement, social housing, commuter). For multi-household systems, the impact of averaging and inaccuracies associated with individual profiles will be reduced. For analysis of individual households, the group-calibrated model has some applicability but with significant qualification as addressed by further occupancy model development addressed in Chapter 7 and recommended further work detailed in Chapter 9.

4.7 Secondary Activity Models

The majority of existing methods incorporating Markov chain models for occupancy assessment use a single-level of probability calculation to capture all occupancy and



(a) Waking period



(b) Pre-sleep period

Figure 4.13. TUS activity probability per-timestep from waking and prior to sleep. Data for analysis from the UK 2000 TUS dataset [83].

activity states. Where only fundamental occupancy states (i.e. sleep, active, out) are captured, the single matrix approach is effective, but when one or more of these fundamental states is further split into more detailed activities (e.g. the active period is defined by multiple TUS activities), the increase in probability data required for each additional 'state' is significant.

An alternative approach to determine specific sub-states of each fundamental state is to incorporate conditional secondary probability functions linked to the three-state primary occupancy model. This can be achieved in different ways depending on the type of secondary detail analysed: (1) a one-time probability function to determine secondary state j based on a previously determined primary state i that continues unchanged until a change in the primary state; (2) assigning a Markov chain transition matrix for a secondary state j based on a previously determined primary state i; (3) assigning a Markov chain transition matrix for a secondary state j based on previously determined primary state i transition at the current timestep.

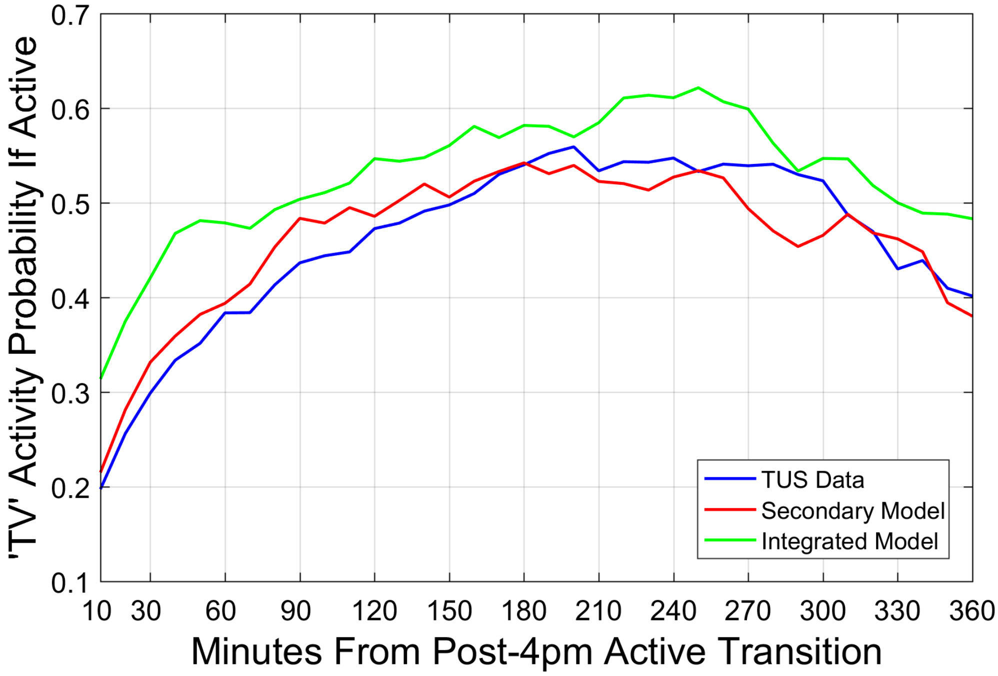
The use of secondary activity models was considered for three elements; TV use and homeworking as detailed below, and for sleeping away as detailed in 4.5.4.8.

4.7.1 State, Transition, and Activity Correlation

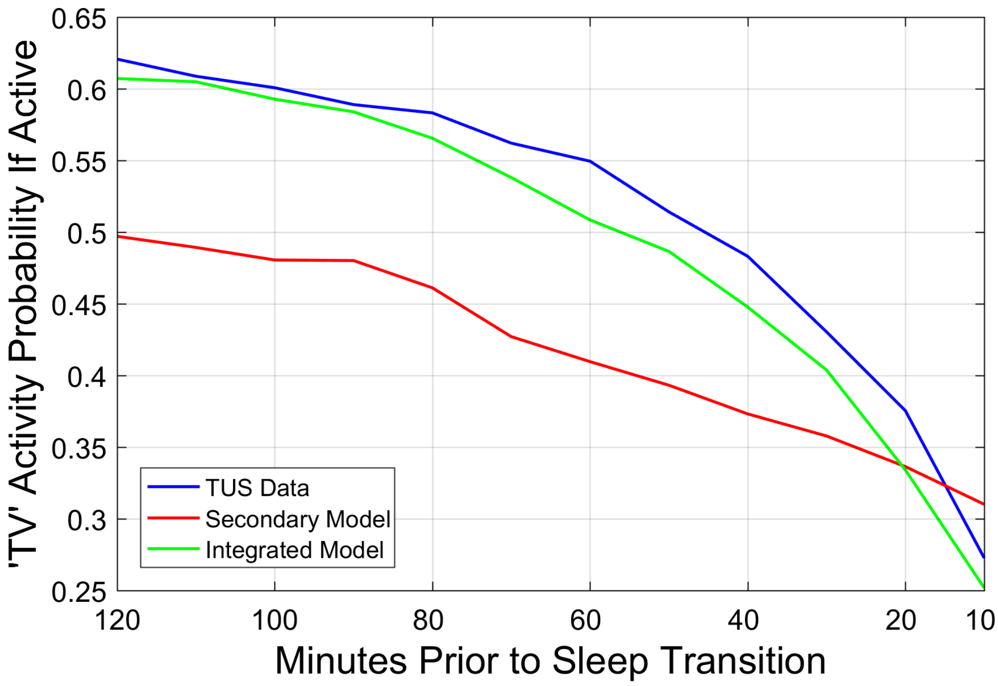
How accurate a secondary model approach is for specific sub-states is dependent on the degree with which the sub-states are correlated with specific primary state transitions. For example, analysis of UK 2000 Time-Use Survey data [83] for activities immediately following a transition to and from the 'active' state shows that for some specific energy-linked activities the probability changes with proximity.

For example, Figure 4.13(a) shows that, for all single-person householders, on waking 'Wash&Dress' and 'Food Prep' predominate, 'Eating' increases and then falls, and the 'Household Upkeep', 'Laundry', 'Dishwashing', and 'TV' activity probability increases steadily with time since waking based on continuous occupancy. Similar patterns can be seen for return from a daytime absence for all except 'Wash&Dress' (low, flat probability) and 'Household Upkeep' (high, flat probability), and for all other populations.

In the final period before the transition to 'Sleep', most people are watching TV



(a) Post-4pm to-active transition



(b) To-sleep transition

Figure 4.14. TV activity probability (if active) comparison for integrated and secondary Markov chain modelling methods based on time from/to two occupancy transitions. TUS activity data for analysis from the UK 2000 TUS dataset [83].

at the start of the period with the number falling sharply as the transition approaches with a parallel rise in the 'Wash&Dress' activity and relatively low levels of other energy-related activities (see Figure 4.13(b)).

Similar behaviour is observed for the equivalent 'active-out' transitions associated with leaving and returning to the dwelling.

4.7.2 Secondary TV Activity Model

Capturing activities within the primary model is more likely to capture the specific transition influenced behaviours shown above. However, given the weak correlation between the majority of TUS activities and the use of associated appliances (see 3.3.3.1), there is limited benefit in modelling specific activities and therefore from this increased accuracy in most cases. However, TV use can be directly attributed from TUS activities, which was the method selected for final demand model development (see 5.10), and a decision was therefore required whether TV use was included as a fourth state in the primary model or could be modelled using a secondary model with the inherent benefits of fewer primary states.

The secondary model option determines whether a person is either 'generally active' or watching TV if the primary occupancy model predicts 'active' occupancy. For the first 'active' timestep, the initial secondary state is determined using a simple time-dependent probability function. For continuing 'active' occupancy, the TV-use probability is dependent on the secondary state at the previous timestep.

Using equivalent models for both options ('integrated' (four-state) and 'secondary' (three-state plus secondary)), a direct performance comparison was made between the model output for 100 representative single-person householders of all ages and the TV activity probability from the equivalent TUS population. The results show that the secondary model performs better than the integrated model for 'post-4pm' return transitions (see Figure 4.14(a)). This is to be expected as the secondary model includes a probability term to determine the starting state based on directly assessed probability data rather than indirectly via the transition probability, and following the first timestep after a transition neither model is directly influenced by the transition. The secondary model, however, does not replicate the significant reduction in TV use prior

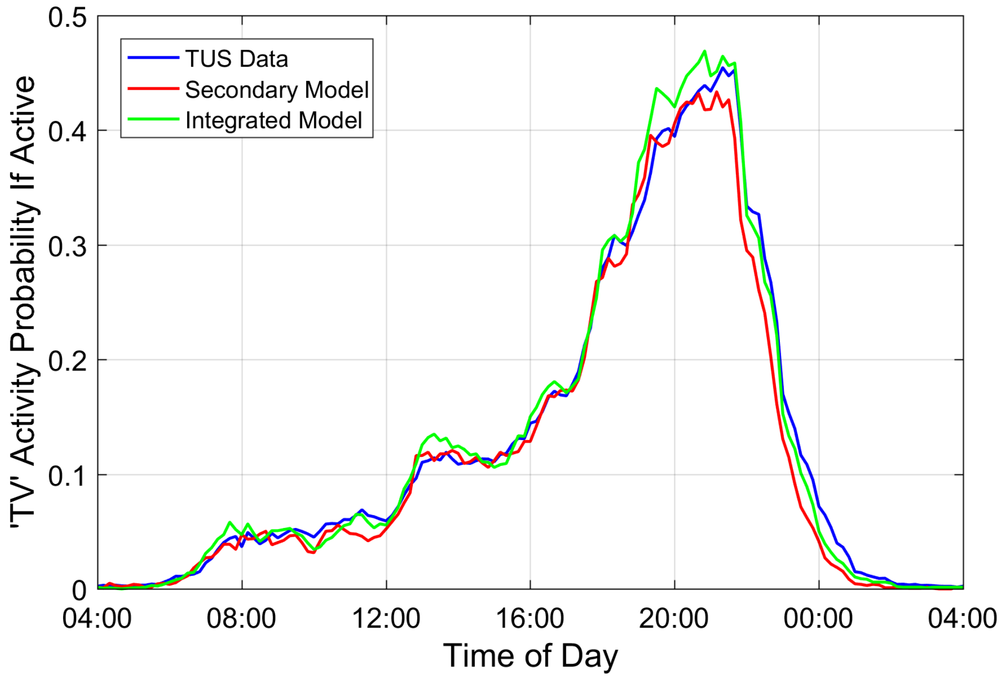


Figure 4.15. TV activity probability (if active) comparison for integrated and secondary Markov chain modelling methods per timestep. TUS activity data for analysis from the UK 2000 TUS dataset [83].

to the sleep transition (see Figure 4.14(b)). Use of the secondary model will therefore generate unrealistic patterns in the immediate pre-sleep period, with higher than actual TV use predicted. Similar performance was seen in periods immediately prior to an 'active-out' transition.

The overall performance for average TV use is similar in both cases (see Figure 4.15). The integrated model slightly overestimates TV use for an equivalent population while the secondary model approach slightly underestimates. The underestimation and relative performance is the result of the lower secondary model probability of a direct transition from TV-use to either sleep or out states in comparison with the non-TV active state in the integrated model.

The conclusion from this analysis was that the use of the three-state primary model plus secondary activity model for TV use provided significant benefits in terms of probability data requirements and the ability to use smaller, more differentiated populations, and that the specific performance benefits of the four-state primary model (i.e. in close proximity to transitions from the active to inactive states) were only apparent at levels of detail that did not reflect the overall accuracy or uses of the overall demand model. The four-state model would be appropriate only for models where TV use was the specific subject of interest.

4.7.3 Secondary Homeworking Model

A similar method to that used for TV use is used to determine whether a person is working when they are in the 'out' state. UK 2000 TUS data does not distinguish a location for working periods, therefore all working periods are set as 'out' periods for the initial calibration. The 'work' state is used solely for the 'homeworker' module to determine when working periods are predicted, and therefore whether these periods, and a number of preceding and subsequent 'out' periods to account for unnecessary travel periods, need to be reset to active occupancy.

A comparison of the secondary model performance to a four-state primary model incorporating 'work' as a separate state was performed in the same manner as the TV-use model. Similar conclusions were drawn, with some minor discrepancies identified close to 'out-active' transitions, but generally good replication of 'work' state behaviour.

4.8 Chapter Summary

This chapter detailed the development process for the sub-models used to define household characteristics and simulate occupancy for different types of occupants and households, for integration in an overall energy demand model. The chapter highlights are as follows:

- The overall demand model structure and calculation sequence has been defined.
- Housing survey data has been used to define the household composition, income, and employment status probabilities for households based on location and house type. This has been used to develop a household characteristics sub-model to allow undefined household data to be predicted probabilistically.
- An assessment of two existing occupancy modelling methods (first-order Markov chain and high-order discrete-event) with a newly developed higher-order Markov chain method, determined that the higher-order Markov chain method performed better overall, with the discrete-event approach performing less effectively in terms of both active occupancy and state duration prediction. A basic three-state (sleep, active, out) occupancy model has been used as it was shown in Chapter 3 that incorporating additional time-use survey activities was less effective.
- The occupancy characteristics of related individuals (couples, parents, and children) were shown to have significantly more periods of combined occupancy and simultaneous occupancy transitions than if each individual were considered independently. Two modelling improvements were developed: the first treats couples (and parents) as single, combined entities within the Markov chain model; and the second links child occupancy directly to the parent occupancy transition at the equivalent timestep using a simple, first-order Markov chain approach.
- Assessment of the occupancy characteristics of different types of households has been limited by a lack of occupant differentiation in existing models. The optimum calibration population size to retain sufficiently robust and varied input data but allow distinct occupancy characteristics to be captured was determined to be 200-300 time-use diaries. Each identified occupancy model population type

(single-person (including unrelated adults in multi-person households), couples, parents, and children) was further split by age range and day type (working/non-working, school-term/non-term) to generate the differentiated groups for occupancy model calibration.

- Additional occupancy model improvements included statistically representative secondary probability models linked to the main three-state occupancy model to account for sleeping away and homeworking potential, and TV use, and extended vacation absences, based on survey data.

Chapter 5

Electricity Demand Sub-Model Development

5.1 Chapter Overview

In Chapter 1, two main aims for the developed overall demand model were stated. To effectively capture individual household demand variations for detailed analysis of small (<1000 household) energy systems, and to be sufficiently differentiated and probabilistic to allow demand uncertainty for different types and sizes of energy system to be determined.

In Chapter 2, a number of household characteristics were shown to both directly influence overall demand and have an indirect effect on both overall and time-dependent demand by changing occupancy patterns. In addition, it was shown that a significant proportion of the demand differences between households cannot be accounted for by characteristics alone. To achieve the stated aims, therefore, the influence of characteristics, occupancy, and behavioural differences must all be captured.

The review of existing demand modelling methods identified several potential areas for improvement, particularly in relation to intermittently used appliances (e.g. kettles, cookers, washing machines), including differentiation by household characteristics and individual behaviours, and realistic sequencing of demand events. The review also determined that models calibrated using the activities identified in time-use surveys to determine appliance use timing would lead to poor demand prediction.

New discrete-event type methods were developed for modelling the use of the intermittently used appliances and enhanced versions of existing methods developed for TV-use, lighting, constantly-used appliances (e.g. fridges, freezers), and other mis-

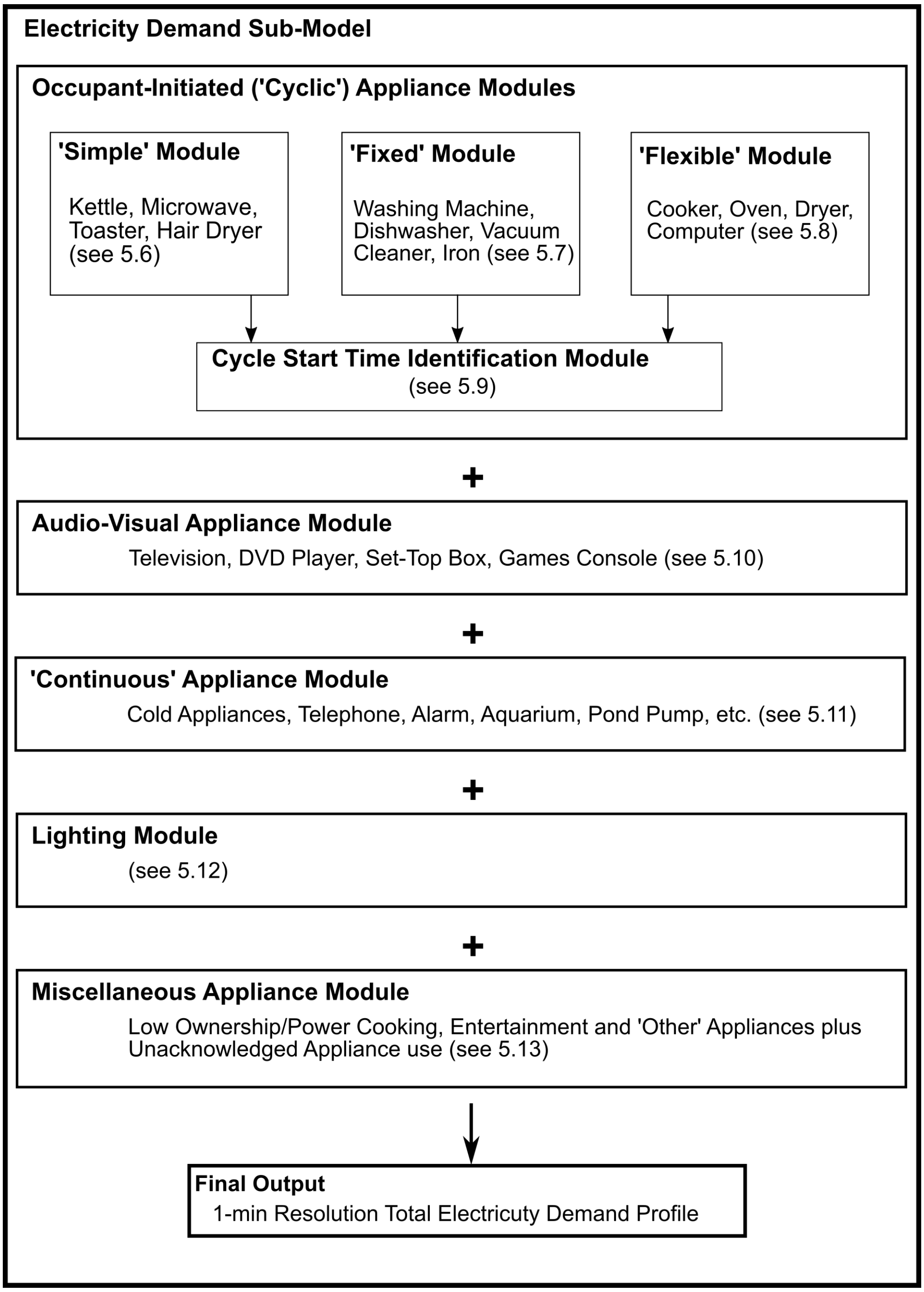


Figure 5.1. Electricity demand sub-model module structure.

cellaneous low-ownership or low-power appliances. To account for the influence of occupancy, the output of the occupancy sub-model detailed in Chapter 4 has been integrated as an input to the electricity demand sub-model. The primary source of calibration data for the demand sub-model development was the Household Electricity Survey (HES) [89] described in detail in 2.3.

The overall demand model output has been validated against both the calibration and independent datasets. The range of results generated has been analysed to confirm the effectiveness of the differentiated and probabilistic elements, and a newly developed similarity assessment method derived from existing edit distance techniques has also been used to determine if realistic individual household demand profiles are generated. Appliance-level analysis has also shown that the model is broadly effective at this resolution.

The overall conclusion from the validation was that the integrated model showed good performance relative to existing methods and good replication of both calibration and independent datasets. However, a degree of convergence to the average behaviours of the calibration groups used for both the occupancy and appliance use timing sub-models was identified, which limits the demand model use at the individual household level. Further model development to account for this is detailed in Chapter 7.

5.2 Sub-Model Structure

Table 4.1 detailed the demand sub-model calculation sequence required to translate the output from the household characteristics and occupancy sub-models into demand prediction. Figure 4.1 shows how the electricity demand sub-model is integrated within the overall demand module and Figure 5.1 shows the overall modular structure of the sub-model. The following section outlines the development of each module that comprises the electricity demand sub-model.

5.2.1 Appliance and Specific Demand Sub-Groups

The review of existing models detailed in 3.4 determined that, for bottom-up models that simulated demands individually, a smaller number of distinct modules were used for demands with similar characteristics. These were typically; 'cyclic', for appliances that are used intermittently and generally occupant-driven; 'continuous', for demands that are usually left in a powered state at all times; 'conditional', for demands such as lighting where use potential is a combination of external drivers, occupancy, and behaviours; 'standby', demand for 'cyclic' appliances that are left in a low-power mode when not in use; and 'miscellaneous', for remaining loads typically associated with low-power and low ownership appliances not captured separately.

Analysis of the HES demand data for specific demands confirmed that, while each has unique patterns of use related to frequency, timing, and duration, broad groups with similar use characteristics that could be modelled in the same manner were discernible. The specific groups identified were as follows.

5.2.1.1 'Cyclic' Appliances

There are fourteen primary occupant-initiated, intermittent use ('cyclic') electrical appliances that are commonly owned and for which a significant depth of cycle (use event) data is available in the HES dataset, justifying separate analysis and modelling; kettles, microwaves, toasters, cookers, ovens, washing machines, dryers, dishwashers, laptop computers, desktop computers, irons, vacuum cleaners, hair dryers, and televisions. Remaining 'cyclic' appliances were modelled using three separate grouped 'miscellaneous' modules as detailed below and in 5.13.

Detailed analysis of the fourteen primary 'cyclic' appliances determined that they could be grouped into distinct sub-type groups based on similar use characteristics and potential for being modelled using the same approach. The four identified sub-types are:

- *Simple* - 'Simple' appliances have a high daily usage probability, usage that is generally independent of use on previous days, and a limited range of potential cycle durations. This definition is applied to kettles, microwaves, toasters, and hair dryers. (see 5.6)

- *Fixed* - 'Fixed' appliances have a low multiple use per day potential, a use probability that is related to time since the previous use, and a typical range of fixed or limited cycle durations. This definition is applied to washing machines, dishwashers, irons, and vacuum cleaners. (see 5.7)

- *Flexible* - 'Flexible' appliances have significant variation in daily use probability and frequency, and duration if used, and therefore require more complex probabilistic analysis than either 'Simple' or 'Fixed' appliances. This definition is applied to dryers, cookers, ovens, and home computers. (see 5.8)

- *Audio-Visual(AV)* - The 'AV' sub-type applies primarily to TV use. Use of other appliances that typically require a TV, such as DVD players, set-top-boxes, and games consoles, are also linked to TV use. TV use is distinguished from other 'cyclic' appliance use as it is explicitly captured by time-use survey activity diaries. In addition, TV-use cycles have highly variable durations and are therefore difficult to capture using other methods. (see 5.10)

To capture individual appliance characteristics, standby use has been defined for each applicable appliance, rather than the approach taken by existing models of setting a fixed demand for this element (see 3.4.3.4).

5.2.1.2 Other Demand Types

Of the remaining types of demands identified by existing models, a similar set of distinct modules has been developed, as follows:

- *Continuous* - 'Continuous' appliances are those that are typically always on and for which there is only a small occupant-driven influence on power use. This definition mainly applies to cold appliances, such as fridges and freezers, but also includes items such as alarms, doorbells, and pond pumps. (see 5.11)

- *Lighting* - Lighting use is driven primarily by external lighting intensity, occupancy, specific occupant location, and house size. (see 5.12)

- *Miscellaneous* - As outlined, in addition to the identified principal appliances and demands, most households also have additional lower ownership or low power appliances. In the HES dataset, these include items such as bread makers, tabletop grills, and radios. For these appliances, there is either insufficient data or impact on overall power consumption to justify modelling separately. Three separate grouped demand modules for cooking, entertainment, and other miscellaneous appliances were therefore developed. (see 5.13)

5.2.2 'Cyclic' Module Structure

The development of enhanced occupancy modelling methods and access to detailed appliance-level data in the HES dataset allowed for significant focus on the modelling of intermittently used ('cyclic') appliances that are typically occupant initiated. As outlined, the AV appliances will be modelled using TUS activity prediction within the occupancy model, the following therefore applies only to the identified 'Simple', 'Fixed', and 'Flexible' appliances.

Based on dataset analysis, the required calculation sequence for a differentiated, probabilistic, bottom-up electricity demand sub-model for the identified key occupant-initiated appliances was determined. Five distinct sequential elements were identified:

- *Household Occupancy and Behaviour Factors* - To account for the observed intra-household-type total demand variation detailed in 2.4.1, in addition to appliance-ownership which is captured separately, the following were identified as potential causes of household-level demand behaviour differences; income-driven behaviour, relative active occupancy probability (particularly as a result of employment), and random energy-use behaviour variations that cannot be directly attributed to household characteristics. The sub-model incorporates the influence of each as a combined multiplier as detailed in 5.3.1.

- *Appliance Ownership* - An initial determination is required to identify if a household owns a particular appliance. This can either be user-specified or determined probabilistically from national survey data based on household type and income.

- *Appliance-Level Variance Factors* - To reflect the significant variation in individual appliance use frequency between similar households, and only a weak correlation between overall household demand and individual appliance use behaviour, use of each appliance per household is independently allocated a relative use multiplier based on HES dataset analysis. The relative likelihood of use based on occupancy duration and timing is also captured. These elements are detailed in 5.5.1.

- *Daily Use Determination* - A discrete-event based approach has been employed to separate the determination of the number of appliance cycles per day and the timing. A cycle is defined as a separate demand event for appliances that are not in constant use. Average daily number of cycles per household is determined, and then cycle number on individual days based on the average with factoring for occupancy and random variation. The average use is also manipulated by both the household- and appliance-level behaviour factors identified. The basis for each of the three appliance sub-types identified in 5.2.1.1 is described in 5.6, 5.7, and 5.8.

- *Cycle Timing Determination* - The start time of each individual use (cycle) is determined based on the identified occupied periods and the probability distribution of start times for a specific cycle (i.e. $\#x$ of total y). This differs from existing models which tend to incorporate a single per-timestep probability calculation sequence calibrated to achieve an average use frequency. The new method was developed primarily to address the problems of unrealistic cycle sequencing and timing (see 3.4.4.2), where previously developed approaches have not addressed the link between the number of cycles in a day and their timing, and is described in 5.9.2. A performance comparison of the per-timestep and newly developed discrete-event method is detailed in 5.14.1.

5.2.3 Time Resolution

The stated aim of the project was for a model with the highest possible time resolution that could be justified by both calibration data availability and computational speed. As detailed in 4.5.5.3, the occupancy model converts an initial 10-minute resolution output, based on the 10-minute calibration data resolution, to a 1-minute output based on the assumption that state transitions occur randomly within each 10-minute period. This assumption was made as there was no clear correlation between demand timing for intermittently-used appliances and specific 2-minute segments within each 10-minute period in the HES dataset and for each 6-second segment in the smaller, but higher resolution REFIT dataset [45].

An initial 1-minute basis was also assumed for the demand model pending final confirmation of computational speed. The 1-minute basis was achieved from the 2-minute resolution calibration data by assuming an equal probability or level of use within each 2-minute period. The lack of any clear timing behaviours at the sub-10-minute level for most appliances suggests that this random within-timestep use timing assumption could be extended to sub-1-minute analysis without introducing significant additional inaccuracy.

Cycle durations can be estimated to a better than 2-minute accuracy for most appliances based on total energy used divided by a nominal baseline power for appliances with relatively constant power use and by statistical analysis of the first and last 2-minute timestep of a use cycle for those with more variable power profiles. In the future, the latter could be extended to sub-2-minute start time analysis if this level of accuracy is considered necessary.

Following completion of the model development, the computation speed for a 1-minute model basis was reviewed and deemed acceptable at c.10 households per minute for a complete annual occupancy, electricity, and hot water analysis on a standard 2013 Quadcore desktop computer. Higher resolution analysis could be implemented with the same developed methods in the future and the relative performance of different resolutions analysed.

5.3 Household Occupancy and Behavioural Factors

5.3.1 Household Behaviour Factor

As outlined, differences in overall household demand behaviour for households with similar characteristics are assumed to be at least partially accounted for by variations in income, occupancy, and attitudes to energy use. A single household behaviour factor ($EHBF$) is determined by combining an income ($EIBF$), overall relative occupancy ($OROF$), and random energy-use behaviour ($ERBF$) factor as shown in Equation 5.1. ($OccUse\_H$ and $OccUse\_T$ are appliance-specific factors for each household (H) and household-type (T) combining both occupancy and relative appliance use timing probability as defined in 5.5.1.1). This factor is applied either to daily cycle (use) number, cycle probability, or total daily usage duration as a single multiplier depending on the appliance sub-type (see 5.5.1.2). The following sections outline how each factor was identified and is determined for each modelled household.

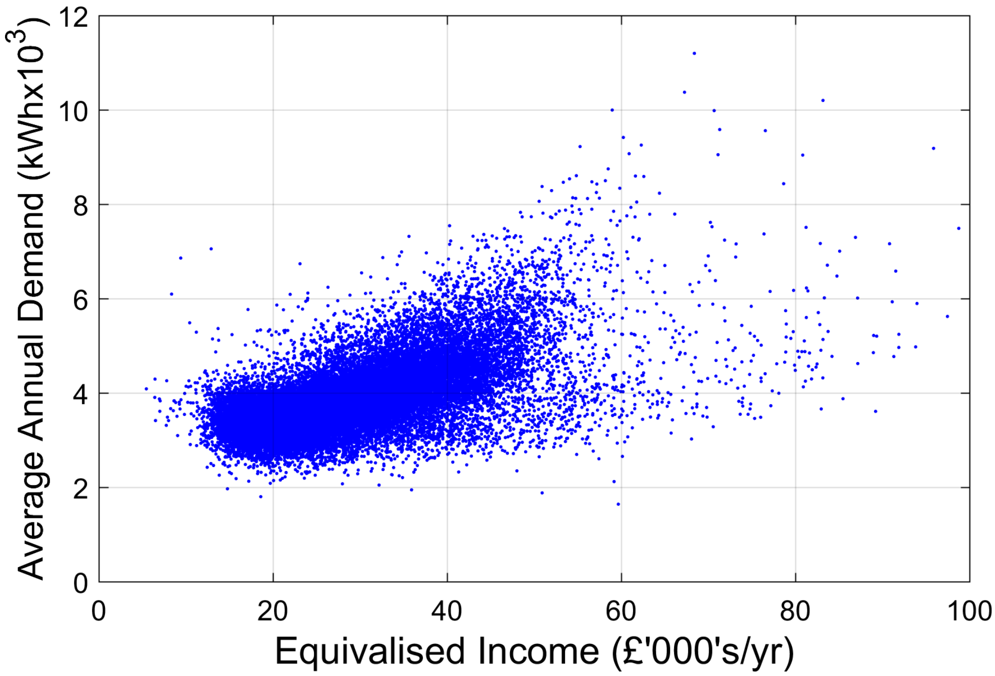


Figure 5.2. Average household annual electricity consumption by average equivalised income per English Lower-Layer Super Output Area (LSOA). Data from [46].

$$EHBF = EIBF \times ERBF \times (OccUse\_H/OccUse\_T)^{OROF} \tag{5.1}$$

5.3.2 Effect of Income on Electricity Demand: Income Behaviour Factor

Analysis of the specific behavioural effect of income on energy use is rare. Jamasb and Meier [108] and White et al [107] determined the overall income effect on electricity demand. However, it is not solely behavioural but also influenced by household type and size, occupancy, and appliance ownership, which are modelled separately. The plot of electricity demand against "equivalised" income (defined below) shown in Figure 5.2 indicates that there is a general increase in demand with income; however, there are clearly one or more additional factors that influence demand, particularly at higher income levels. A regression analysis was therefore undertaken that allowed these other income-related factors to be accounted for separately and the residual behavioural effect of income on electricity demand determined.

Published UK 2011 Census data has been separated into different sizes of area for comparative analysis. Area annual electricity demand data [46] is available down to the Lower Layer Super Output Area (LSOA) level for England; this corresponds to areas of typically between 600 and 1000 households. This is assumed to be sufficiently small that each area has distinct characteristics for comparative analysis but is large enough to ensure that any random household-level behavioural effects are negated.

All London boroughs were removed from the analysis as there was clearly a different cost of living basis and relationship between several factors (particularly income) and energy usage that distorted the results for the rest of the country. For example, the London boroughs have an average annual electricity demand to "equivalised" income of 0.117 kWh/£, with the remainder of England having a value of 0.147. The number of LSOAs included in the analysis was 28,203.

Using UK 2011 Census data various factors were determined for each LSOA to represent factors modelled separately. The following were found to generate the most accurate regression model based on F-Stat and RMSE analysis:

- *"Equivalised" Income* – Average gross income (2011 basis) factored by household size, with a reducing marginal impact of additional people on spending power based on the OECD method [109] (1st Person=0.58, 2nd Person=0.42, 3rd Person+=0.3). This was shown to be better correlated with demand than the unfactored gross income.

- *SAP-Factored Occupant Number* – Number of people factored using the SAP $N^{0.4714}$ basis for demand relative to the household size (N) [111]. This was shown to be better correlated with demand than the unfactored occupant number.

- *Rooms per Person* – Average number of habitable rooms per person. Habitable rooms are living rooms, kitchens, bedrooms etc., but not toilets and hallways. This was shown to be better correlated with demand than room number.

- *Owned Appliance Power (Relative)* – Estimate of relative power use related to ownership level of appliances based on LSOA household type and income mix.

- *Occupancy (Relative)* – Estimated average active occupancy for the LSOA based on household type mix.

The RMSE was minimised with the "equivalised" income term ($IncE$) raised to a power of 4.2 (with all other factors set to a power of 1 to reflect the model basis). This demonstrates that electricity demand increases rapidly and disproportionally with increasing disposable income.

Substituting average values for the non-income factors and rebasing the regression output to the LSOA average demand of 3863.3kWh/yr reduces the regression output to Equation 5.2 to express the income-behaviour impact (based on 2011 "equivalised" income ($IncE$) values) as a relative Income Behaviour Factor ($EIBF$).

$$EIBF = 0.9852 + (0.440/3863.3) \times (IncE/10000)^{4.2} \tag{5.2}$$

This factor is used as a multiplier, first incorporated in an overall household behaviour factor (see 5.3.1), and then used to manipulate the appliance-specific primary demand parameter (daily cycle number, probability or total duration) used for each modelled appliance (see 5.5.1.2). The HES dataset does not include sufficient income data or include enough households to allow the relative income effect for each potential use to be determined, it is therefore applied to all appliances and demands equally. In reality, the income behaviour influence may be greater for higher demand, lower-use appliances.

5.3.3 Effect of Occupancy on Electricity Demand: Relative Occupancy Factor

A limitation of the HES dataset is that it does not include occupancy data. Without integrated occupancy and appliance usage data an explicit assessment of the impact of occupancy duration and timing on appliance usage is difficult. Model calibration, however, requires two relative occupancy manipulations; first, that average usage is determined for each household type and then adjusted for each modelled household based on household occupancy relative to the household type average ($OROF$); and second, that the determined average usage per-household is further adjusted for daily occupancy relative to the household average occupancy ($DROF$). Here, occupancy is defined as the proportion of time the house is actively occupied (i.e. not including sleep) by at least one person, further investigation is required to determine if there are more complex relationships linked to number of occupants present.

To overcome the integrated data problem, the relative impact of occupancy on appliance usage has been determined from analysis of usage differences between retired and working age households in the HES dataset. These populations show a marked difference in active occupancy characteristics, which can be extracted from UK 2000 Time-Use Survey (TUS) data [83].

The relationship between the ratio of individual household to household type average occupancy and appliance use is assumed to be a power law (see Equation 5.1), with an exponent between 0 and 1. A value of 1 would represent an appliance with a usage that was directly proportional to active occupancy. In reality, the value is much lower than 1 for all appliances. The exponents are determined by comparing the ratios of average number of uses and average active occupancy for the working age and retired populations as shown in Equation 5.3. The exponent quantifies the proportional impact on relative appliance use frequency of a change in relative occupancy based on the average appliance use behaviours of the two populations with distinctly different average active occupancy levels.

$$ROF = log(Occ_{ret}/Occ_{work})/log(Cyc_{ret}/Cyc_{work}) \tag{5.3}$$

For higher use appliances (i.e. kettle, microwave, toaster, computer), a single factor is used for both the overall household relative occupancy factor ($OROF$) and daily occupancy adjustment ($DROF$) as use is assumed to be highly correlated with occupancy.

For less frequently used appliances (i.e. washing machine, dryer, dishwasher, cooker, and oven), the influence of relative average occupancy on the probability of use is expected to be stronger with regard to daily use probability ($DROF$) than overall use probability ($OROF$), as overall use is driven by other factors, including basic need. For these appliances, a separate value for daily occupancy impact ($DROF$) has been estimated prior to combined occupancy and use data becoming available. The 'Overall' factor ($OROF$) was determined in the same manner as for the higher use appliances. Table 5.1 shows the overall ($OROF$) and daily ($DROF$) relative occupancy factors used.

Table 5.1
Overall ($OROF$) and daily ($DROF$) relative occupancy factors for different appliances.

	Kettle	Microwave	Toaster	Computers
Overall/Daily	0.3	0.1	0.05	0.5
	Washing Machine	Dryer	Dishwasher	Cooker/Oven
Overall	0	0	0.5	0.1
Daily	0.3	0.5	0.67	0.2

To reflect the fact that appliance use is assumed to be impacted by the timing of the occupied period relative to when specific appliances are typically used in addition to basic occupancy duration, the factors are applied to a modified occupancy-use factor as detailed in 5.5.1.1.

5.3.3.1 Effect of Behaviour on Electricity Demand: Random Energy-Use Behaviour Factor

Gill et al [81] determined that 37% of total household electricity use can be attributed to behaviour independent of identifiable household characteristics related primarily to attitudes to and prioritisation of energy use. While the Gill et al analysis is a small-scale study of 11 households, it represents the only study that has attempted to quantify purely behavioural differences between households under UK conditions. An additional 'random behaviour' factor (*ERBF*) is therefore applied; selected randomly between 0.77 and 1.23 (equivalent to a 37% variation). Assuming both the wider applicability of the small-scale Gill et al analysis and a linear distribution are likely to be over-simplifications, however, this basis represents a reasonable initial assessment of household-level behavioural variations pending better data.

As with the Income Behaviour Factor, the Random Behaviour Factor (*ERBF*) is first incorporated in an overall household behaviour factor (see 5.3.1), and then used to manipulate relative appliance demand per household (see 5.5.1.2).

5.4 Appliance Ownership

The literature review by McLoughlin et al [164] determined that appliance ownership was the second most cited influence on electricity demand. Figure 2.9 showed the significant relationship between the number of appliances owned and electricity demand in the HES dataset. As defined in 2.4.1.7, ownership is primarily driven by household composition and income.

The UK Office of National Statistics (ONS) Family Spending survey [100] provides annually updated ownership data for several key appliances (washing machines, dryers, microwaves, computers, and dishwashers). Ownership data is provided separately by household composition and income decile. A separate dataset in the same survey provides the proportion of each household type and size within each income decile. Combining the data allows ownership probability to be determined for each household based on composition and income decile. This analysis, for example, determined that dishwasher ownership ranged from 9% in both working-age and retired single-person households in the lowest income decile to 90% in family households and near universal ownership in larger multi-adult households in the highest income decile.

TVs and cold appliances are characterised by near universal ownership and a significant potential for multiple units per household. The HES ownership data was used as the source for the model with probability data for the number of units per household differentiated by number of occupants. Cooker ownership is also very high and assumed to be universal. The main cooker variable is the fuel source used for each element. Analysis of the EFUS dataset [43] determined that for houses with mains gas there was 34.1% probability of a gas hob and electric oven and 30.3% probability of a fully electric cooker. Ownership probability of the remaining key appliances, irons, vacuum cleaners, and hair dryers, was taken from the HES dataset based on household type.

5.4.1 Appliance Energy Ratings

For a number of appliances, the typical power use can be inferred by an assigned energy rating. This applies in particular to cold appliances, washing machines, dishwashers and dryers. In the European Union, ratings are assigned based on EU Directive 2010/30/EU [165].

In the HES dataset, appliance ownership data also includes the unit EU energy rating in some cases. UK ONS data is available detailing the numbers of each appliance per energy rating owned in the UK [166]. As defined in the relevant demand module sections below, in some cases the demand and energy rating data has been combined to define power profiles for modelled appliances.

5.5 'Cyclic' Appliance Modules

Separate modules have been developed for the four sub-types of 'cyclic' appliances identified in 5.2.1.1. A similar discrete-event based approach with distinct characteristics per sub-type, calibrated directly using demand data as detailed below, has been employed for the 'Simple' (see 5.6), 'Fixed' (see 5.7), and 'Flexible' (see 5.8) sub-types to first determine the average number of use cycles per day for each modelled household. An identical method is then used to determine the cycle timing within each modelled day (see 5.9). As addressed in 5.2.1.1, this type of method is employed to address identified areas of poor performance for existing discrete-time models. A performance comparison between the existing and developed approaches is detailed in 5.14.1. As outlined, TV and other AV appliance use is modelled with a different approach based on the relevant time-use survey activity (see 5.10).

Each developed module incorporates the output of the occupancy model to deter-

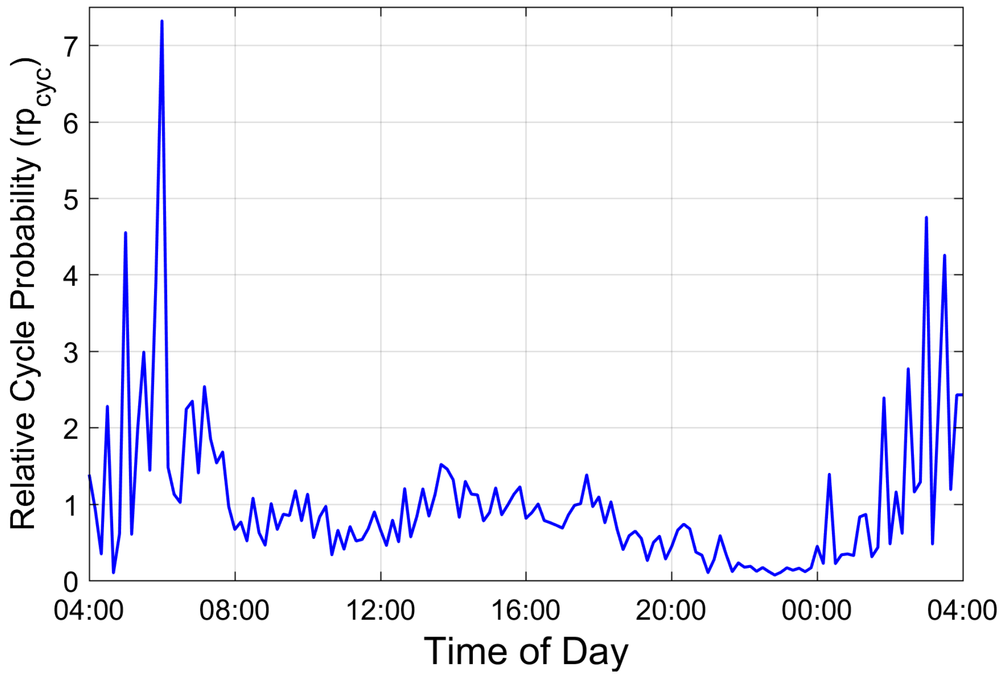


Figure 5.3. Example modified kettle cycle (use) time probability distribution with occupancy influence removed for retired single-person households.

mine use potential. This is achieved directly for the AV model, using the secondary model described in 4.7.2, and for the other modules by limiting when appliances can be used. The household-level behavioural factors identified in 5.3, and additional appliance-level use factors also need to be applied to each module as outlined in the following section.

5.5.1 Appliance-Level Variance Factors

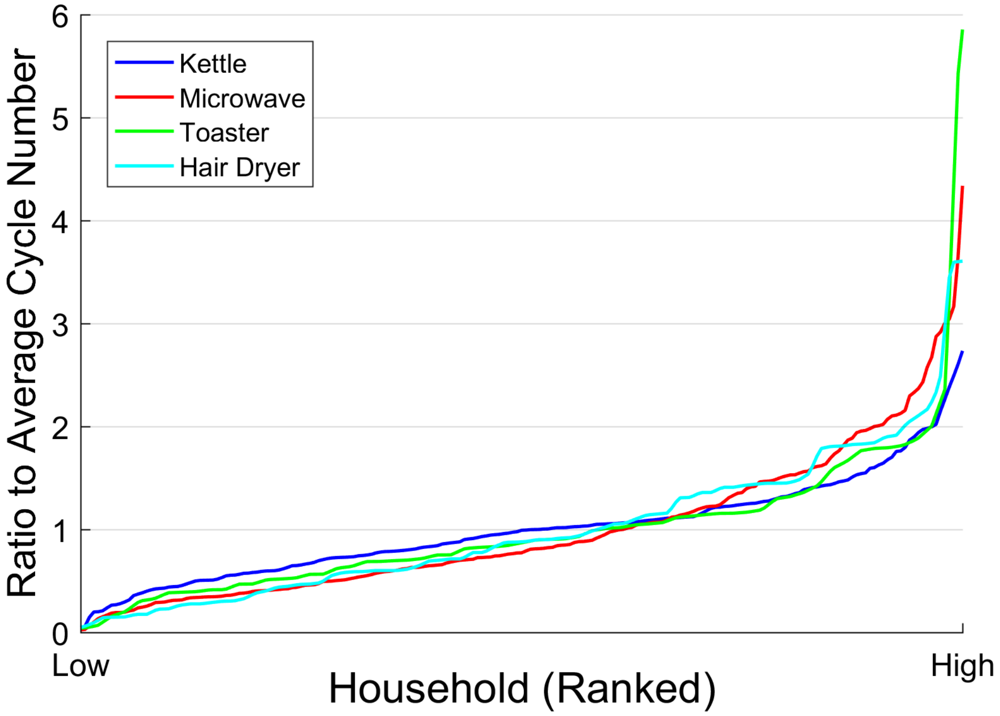
5.5.1.1 Occupancy-Timing-to-Appliance Use Relationship

The relative occupancy factors defined in 5.3.3 are not applied directly to occupancy but to a modified occupancy factor that also reflects relative appliance use probability during the occupied period(s) if the dwelling is occupied. In order to do this, a probability density function (pdf) distribution for cycle start time probability was generated for each appliance and household type from identified start times in the HES dataset. This distribution was then modified to remove the occupancy-driven influence on timing by dividing each timestep pdf value by the occupancy probability at the timestep for the household type and then rebased to an average value of 1 for clarity. The updated relative value per timestep is the element $rp_{cyc}(t)$ in Equation 5.4. As an example, Figure 5.3 shows the $rp_{cyc}(t)$ distribution for kettle use in retired single-person households.

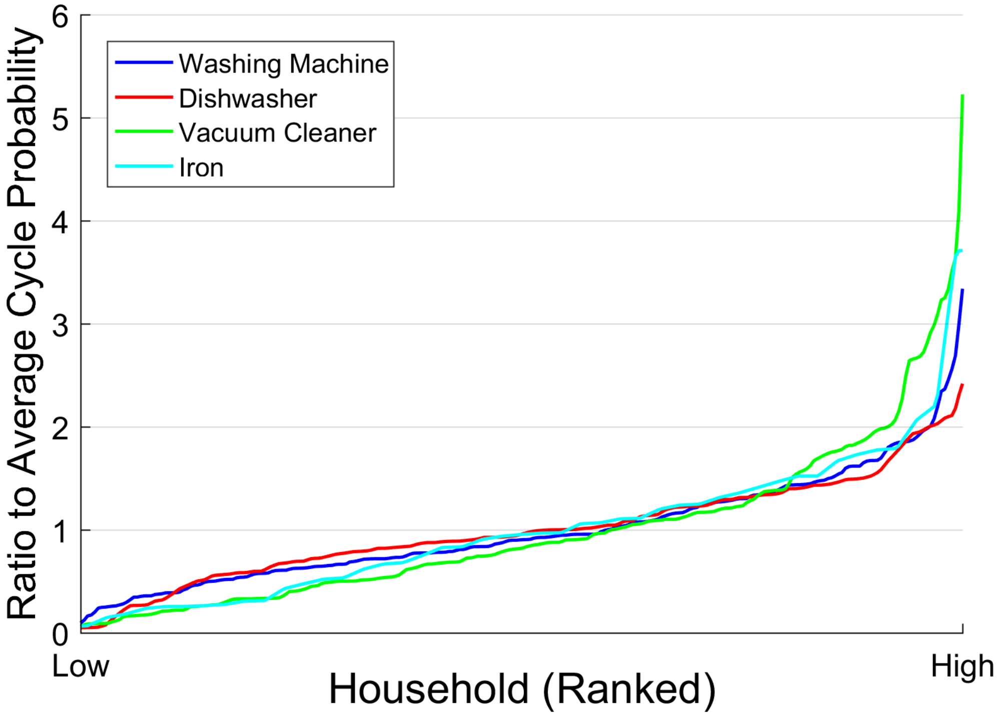
A combined factor ($OccUse$) reflecting both occupancy and use probability summed for each timestep was then generated (see Equation 5.4). The term $p_{occ}(t)$, the occupancy probability at timestep, t, can be either population average, household average, or specific day (0 or 1 per timestep) based depending on the analysis required.

$$OccUse = \sum_{t=1}^{144} p_{occ}(t) \times rp_{cyc}(t) \tag{5.4}$$

The average $OccUse$ value for a household ($OccUse\_H$) compared to the population average ($OccUse\_T$) gives a basis for determining the potential occupancy-driven variation in average appliance use. Comparing the value for each modelled day ($OccUse\_D$) to the household average also gives a basis for the occupancy-driven variation in daily use. The determined ratios are factored by either the overall ($OROF$) or daily ($DROF$) relative occupancy factors as defined in 5.3.3.



(a) 'Simple' appliances



(b) 'Fixed' appliances

Figure 5.4. Ranked average household daily cycle number to household type mean ratio distributions for all 'Simple' and 'Fixed' appliances. Data for analysis from the HES dataset [89].

5.5.1.2 Appliance Use Factor

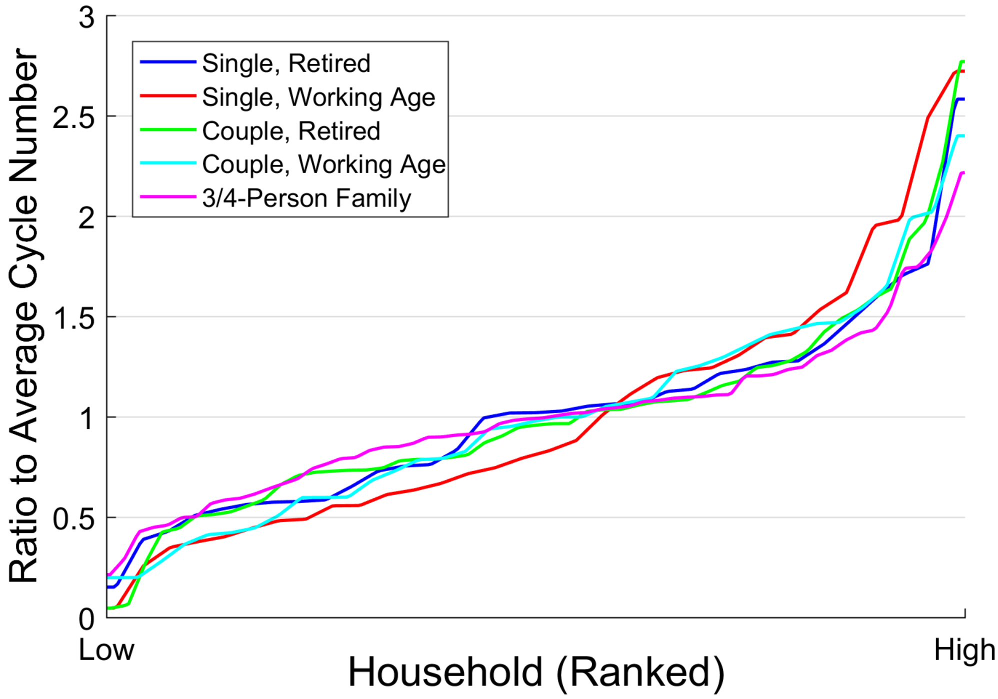
For each appliance, there is significant variation in the primary use defining parameter ('Simple' - number of daily cycles (uses), 'Fixed' - daily cycle probability, or 'Flexible' - daily use duration) for each household relative to the household type mean (see Figure 5.4 for examples).

For each of the appliances the ratio-to-mean distributions for each household type are similar (see Figure 5.5(a) for kettle example), particularly for the larger populations, and with sparse data for the smaller populations. To ensure representative ratio-to-mean distributions for all populations, the data could therefore be consolidated to the single combined distribution of the overall range of ratios to the household type mean shown in Figure 5.4. An example comparison of the average cycle number distributions for kettle use generated from actual data ('actual') with those generated using the consolidated distribution ('consolidated') is shown in Figure 5.5(b) and exhibits close replication.

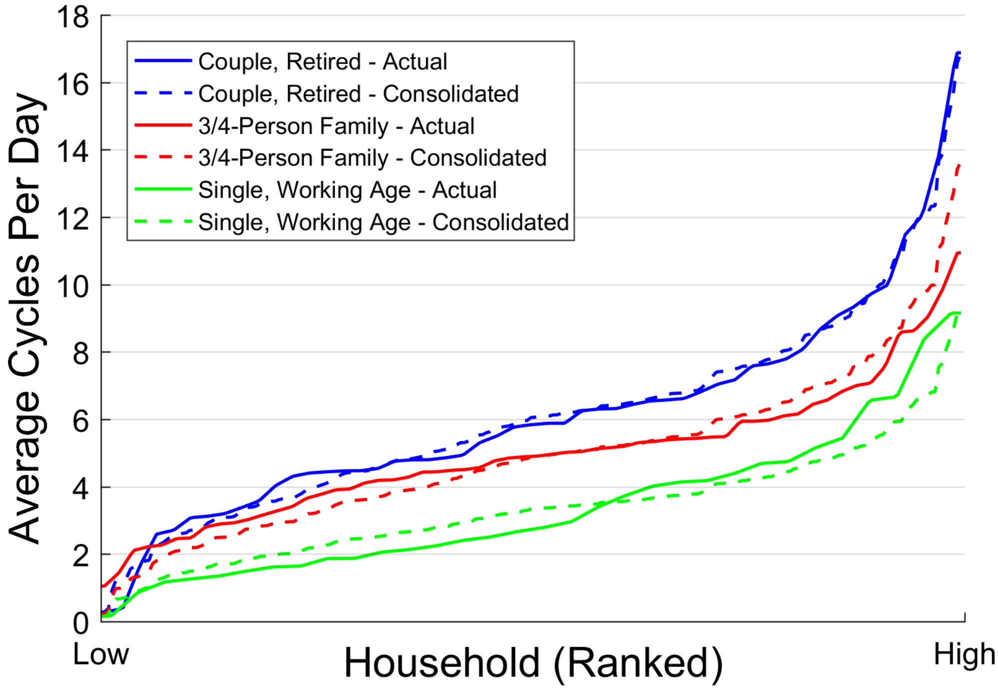
Whilst the observed use variation is partially attributable to household characteristics and occupancy, a significant influence on the distribution is assumed to be random behavioural differences between similar households at the appliance level.

For each modelled appliance, an 'Appliance Use Factor' (AUF) is randomly selected from the appliance-specific ratio-to-mean distribution (see Figure 5.4). This is a multiplier that is combined with the mean value of the appliance primary use defining parameter (see above) for the household type and the household behaviour factor (*EHBF*) (see 5.3.1), as per Equation 5.5, to determine the parameter for each modelled household incorporating all behavioural factoring.

Analysis determined that there was no discernible relationship between the relative ranking in the ratio-to-mean distribution for each appliance per household beyond that which would be predicted by the application of the household-level factoring introduced in 5.3. This suggests that any per-household appliance-specific factoring could be allocated randomly and independently using the generated ratio-to-mean distributions, with additional modification using the *EHBF* to account for relative household-level behaviours.



(a) Kettle daily cycle number distributions by household type



(b) Performance comparison between HES data and consolidated cycle number prediction method

Figure 5.5. Relationship between household daily appliance-specific cycle number and household type mean. Data for analysis from the HES dataset [89].

5.5.2 Cycle Data Analysis

The 'Simple', 'Fixed', and, 'Flexible' modules are calibrated using demand data from the HES dataset. Despite only comprising private households, the range of household types and social classes is nationally consistent. It was therefore assumed that the HES dataset was broadly representative of appliance use for UK households.

The data analysis first required that the data was filtered to allow each separate use cycle to be identified as described below. Once the number of distinct cycles for each household were determined, they could then be further analysed for frequency, timing, and duration, and any inter-relationship between these elements. The module specific results of the analysis are detailed in 5.6, 5.7, and 5.8.

5.5.2.1 Cycle Identification

The HES time-series power data was analysed to identify individual appliance cycles. The same basic method was used for each appliance with additional appliance-specific analysis as detailed.

For some appliances there is a constant or regular low level power draw that is not indicative of a distinct cycle but standby use. The data was therefore first filtered of all periods below a minimum power demand (between 0.2 and 1W depending on appliance minimum in-use power characteristics). The residual low power element was analysed separately to determine if significant enough to be modelled as a fixed constant demand for some households.

From the filtered data, demand in adjacent timesteps is analysed to determine the start and finish time, and total energy use per cycle. For some appliances there are short periods of low or no power that are not necessarily indicative of a separate use but a temporary delay or an appliance in-use characteristic. Uses separated by less than a defined period, typically 10 minutes, are therefore combined in a single 'cycle'. In order that the cycle number results are not distorted by significant numbers of short, negligible energy use cycles, the cycles are further filtered by a minimum cycle energy requirement. For all appliances, a suitable value could be set that removed a significant number of unrepresentative uses but with a less than 1% reduction in overall energy use.

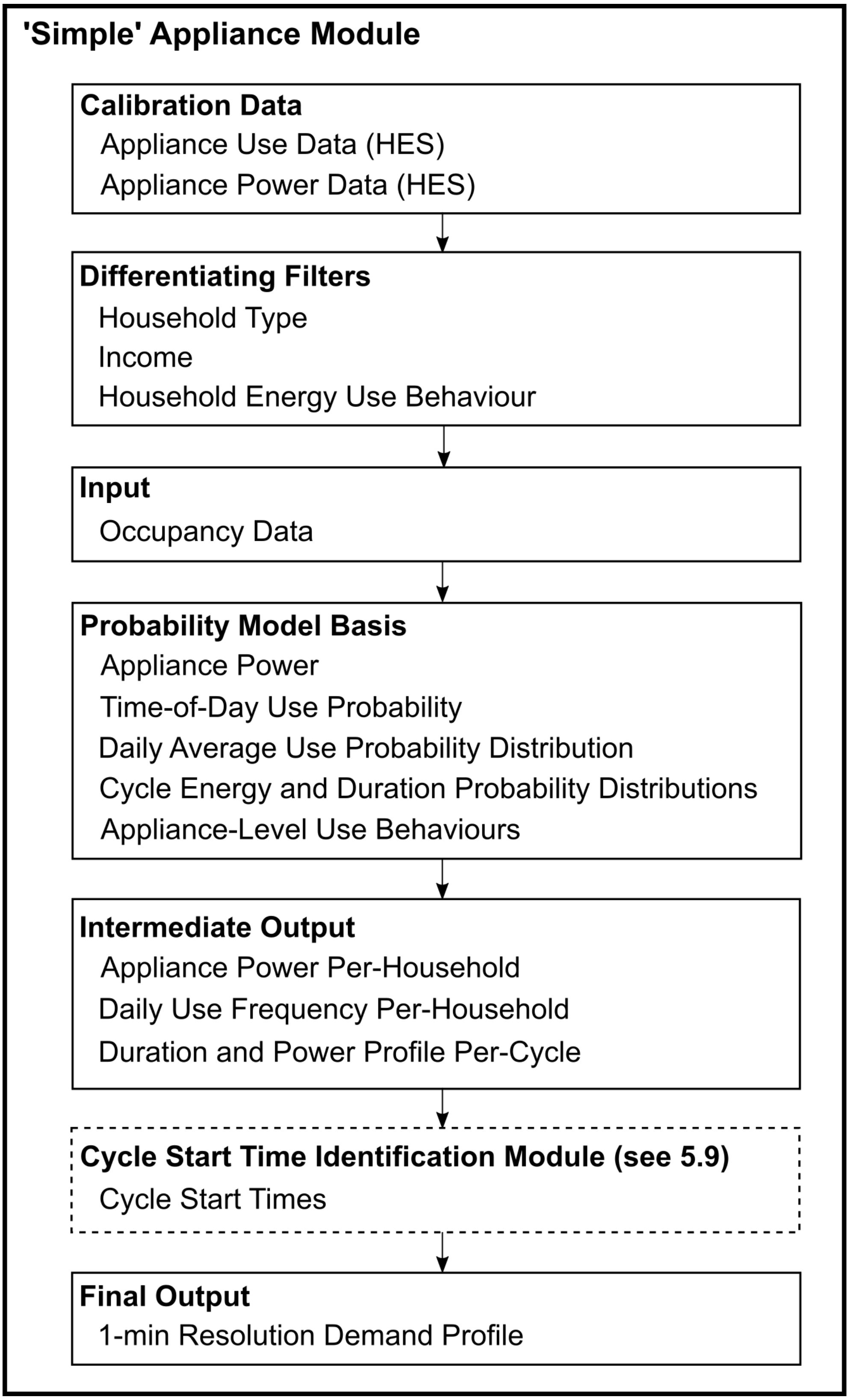


Figure 5.6. 'Simple' appliance module structure.

5.6 'Simple' Module Development

The 'Simple' appliance demand module is applied to kettles, microwaves, toasters, and hair dryers. The overall structure of the module and principal defining elements are shown in Figure 5.6.

5.6.1 Dataset Cycle Identification

Kettle data is characterised by short duration, constant, high power cycles (typically c.2kW) and a low probability of standby power use. There is also the possibility of short 'reboil' cycles when the water is reheated shortly after initial boiling. Data analysis shows that 12.5% of cycles occur within 10 minutes of a preceding cycle and that short 'reboil' cycles are significantly more common than multiple full cycles within a 10-minute period (on average the second cycle uses 48.7% of power of the first and the median value is 31.5% with 69% of second cycles below 50%). Combining cycles that are separated by less than 10 minutes results in 'reboil' cycles being combined with the initial cycle which ensures that they do not distort cycle number counts. Kettle cycles are therefore nominal cycles based on use within a short period following an initial use.

Similarly microwave use is frequently characterised by multiple short cycles within a short period. As for kettles, for both simplicity and to avoid cycle numbers per day being distorted by sequences of short uses in close proximity, these were consolidated with separate cycles distinguished by periods of at least 10 minutes between individual uses.

For toasters and hair dryers, the basic method defined in 5.5.2 was used.

5.6.2 Daily Use Determination

5.6.2.1 Dataset Analysis

For each 'Simple' appliance, the average number of cycles per day was identified for each household type (see Table 5.2).

Table 5.2
Average daily cycle number by household type (*TypeCyc*) for all 'Simple' appliances (combined data due to low ownership is identified with a (*)). Data for analysis from the HES dataset [89].

Household Type	Kettle	Microwave	Toaster	Hair Dryer
Single, working age	3.37	0.86	0.55	0.48*
Single, retired	3.98	0.84	0.50	0.48*
Couple, working age	4.59	1.34	0.92	0.30
Couple, retired	6.10	1.34	0.99	0.39
Three-Adult	6.50	2.14	0.92*	0.74*
3-Person Family	4.94*	1.81	0.92*	0.74*
4-Person Family	4.94*	2.45	1.35*	0.82*
5+-Person Family	6.22	3.09	1.35*	0.82*

5.6.2.2 Module Calculation Sequence

The 'Simple' module assumes the probability of use per day is high and independent of use on preceding and following days. The primary parameter used to define use per household is the average daily cycle number. The steps required to determine this value and use on individual modelled days are as follows:

- *Step 1* - The 'Appliance Use Factor' (AUF) is determined by randomly selecting a value from the appliance-specific ratio-to-mean distribution (see Figure 5.4(a)).
- *Step 2* - The occupancy model output is converted to the combined factor that assesses both occupancy and time-dependent appliance use likelihood ($OccUse$) (see 5.5.1.1) for each individual day to be modelled ($OccUse\_D$) and an overall household average ($OccUse\_H$) for all modelled days.
- *Step 3* - The average number of daily use cycles for the appliance ($HhldCyc$) is determined from the following equation based on the average number of cycles for the household type ($TypeCyc$) (see Table 5.2) and the defined behavioural factors.

$$HhldCyc = TypeCyc \times EHBF \times AUF \tag{5.5}$$

- *Step 4* - For each modelled day, the household average number of daily cycles is used to determine a baseline number accounting for the ratio of the day-specific $OccUse$ factor ($OccUse\_D$) to the household average ($OccUse\_H$), as shown by the following equation. The baseline cycle number is the average predicted number based on the day-specific occupancy characteristics.

$$CycBase = HhldCyc \times (OccUse\_D/OccUse\_H)^{DROF} \tag{5.6}$$

- *Step 5* - The predicted number of cycles for each modelled day is determined using a binomial probability distribution (see below) with the baseline cycle number (*CycBase*) as the average output and a random number generated between 0 and 1 to identify the actual cycle number from the discrete probability distribution.

5.6.2.3 Binomial Cycle Number Probability Method

As outlined in Step 5 above, a binomial probability determination is used to predict the actual number of cycles in relation to the average predicted number of cycles for a specific day. This further manipulation assumes that there is a natural variation in cycle number about the mean predicted value due to random differences in occupant behaviours and external factors, such as weather. Without combined occupancy and use data the extent of this variation is difficult to predict accurately but the binomial basis was shown to replicate day-to-day cycle number variations in the HES dataset better than if not further manipulated and in comparison with other probability distributions, such as Poisson. Future work with a combined source of occupancy and demand data would allow the assumed relationship to be confirmed.

Binomial distributions are characterised by a number of tests, N, and the probability of a success per test, p. In this case, N is the maximum daily cycle number for the household (probabilistically allocated from HES dataset analysis of the relationship between average and maximum daily cycle number values per appliance), and, p, the daily baseline cycle number, *CycBase*, divided by N. An example predicted cycle number distribution for a daily baseline of 3 cycles and a household maximum of 6 cycles is shown in Table 5.3.

Table 5.3
Binomial distribution probabilities for a daily cycle number baseline of 3 and household maximum daily number of cycles of 6.

Predicted Cycle Number	0	1	2	3	4	5	6
Probability (%)	1.6	9.3	23.5	31.1	23.4	9.5	1.5

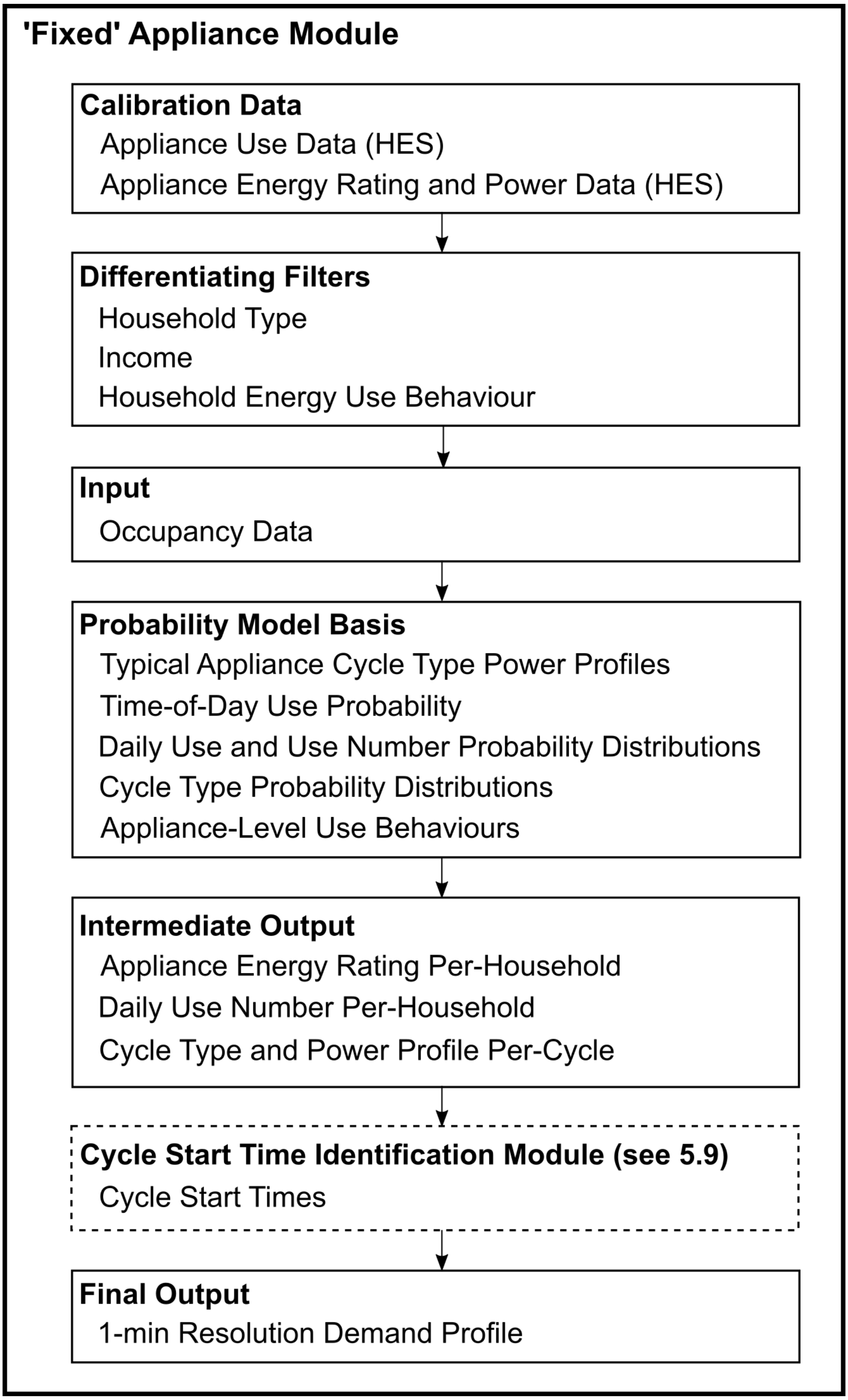


Figure 5.7. 'Fixed' appliance module structure.

5.6.3 Cycle Power and Duration

Each 'Simple' appliance is characterised by short cycle durations, typical appliance power levels, and constant power profiles. As it is difficult to determine exact cycle durations from the 2-minute resolution data and relative appliance performance is unknown, the dataset analysis was simplified to identifying the total energy per cycle, allowing household usage to be compared on an equivalent basis.

Three sets of calibration data were generated from the HES dataset; the average cycle energy per household type; the overall ratio-to-mean distribution of household mean cycle energy to the household type mean ('Household'); and the overall distribution of individual cycle energy to the household mean ('Cycle'). The range of peak unit power values per household was also determined.

Each household is randomly allocated a unit power per owned appliance from the determined range, and an average total energy per cycle based on the household type mean and a randomly selected value from the 'Household' ratio-to-mean distribution. Each individual cycle energy is determined from the household mean value multiplied by a randomly selected value from the 'Cycle' ratio-to-mean distribution. Cycle duration is determined from total cycle energy divided by the allocated unit power.

For microwaves, the process is the same as the other appliances except that there is a defined relationship between average daily cycles and total energy per day that is used to determine the average energy per cycle. The Kernel Density method (see Appendix A) is used to simulate this relationship, with the average daily total energy determined for each household based on the simulated average daily cycle number.

5.7 'Fixed' Module Development

The 'Fixed' appliance cycle module is used for washing machines, dishwashers, vacuum cleaners, and irons. The overall structure of the module and principal defining elements as applied to both washing machine and dishwasher use are shown in Figure 5.7. A simplified version is applied to vacuum cleaners and irons.

5.7.1 Dataset Cycle Identification

Most washing machine cycles are characterised by an initial high power period of between 10 and 30 minutes, a longer lower power period of between 40 and 80 minutes, and a medium power period of between 5 and 10 minutes. Although there is significant variation in total cycle energy, analysis of the data determined that full wash cycles typically ranged from 0.5kWh to 1.5kWh. Below 0.5kWh was associated with simpler cycles (rinse/spin etc.) and above 1.5kWh indicative of two adjacent cycles.

The data was initially filtered of all 'standby' periods of less than 1W. This value was selected using trial and error to minimise very high cycle power events that are indicative of two adjacent full cycles while remaining below minimum in-use power. Power use in adjacent periods, including gaps of less than 10 minutes to account for idle periods mid-cycle, are combined. The data is further filtered of low power use cycles (<0.25 kWh) to ensure a realistic number of full cycles are captured. This method was also used for dishwashers that have a similar characteristic cycle.

For vacuum cleaners and irons the basic method defined in 5.5.2 was used.

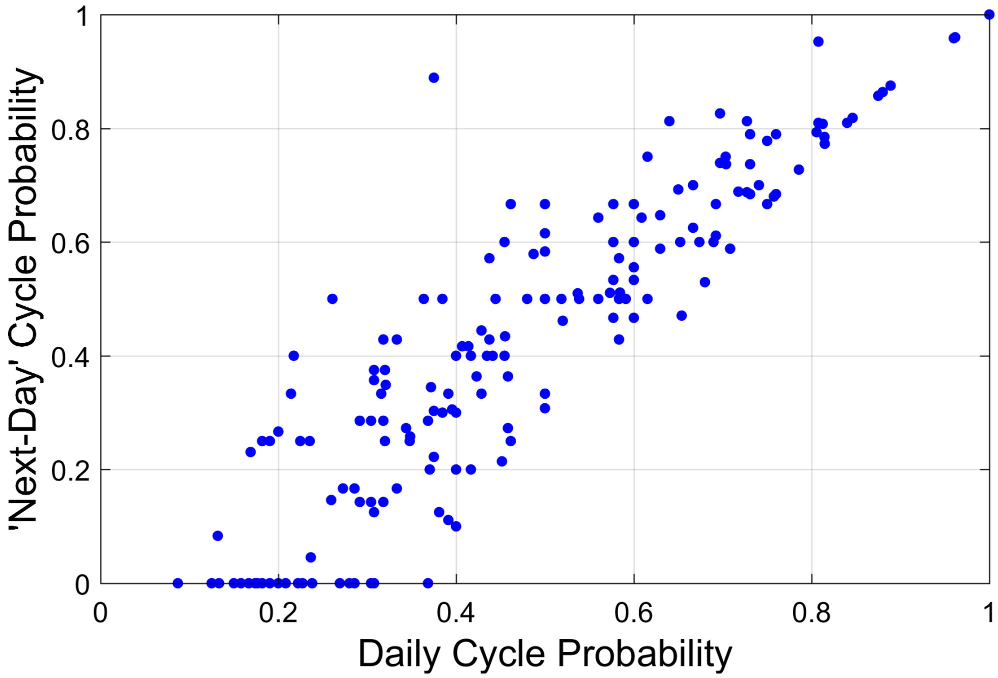
Table 5.4
Average daily cycle number by household type for all 'Fixed' appliances (combined data due to low ownership is identified with a (*)). Data for analysis from the HES dataset [89].

Household Type	Washing Machine	Dishwasher	Iron	Vacuum Cleaner
Single, working age	0.299	0.428	0.258*	0.213
Single, retired	0.275	0.439	0.258*	0.206
Couple, working age	0.598	0.537	0.342*	0.393
Couple, retired	0.397	0.462	0.404	0.324
Three-Adult	0.829*	0.722*	0.342*	0.591
3-Person Family	0.829*	0.722*	0.578*	0.416
4-Person Family	0.829*	0.722*	0.578*	0.392*
5+-Person Family	0.968	0.693*	0.578*	0.392*

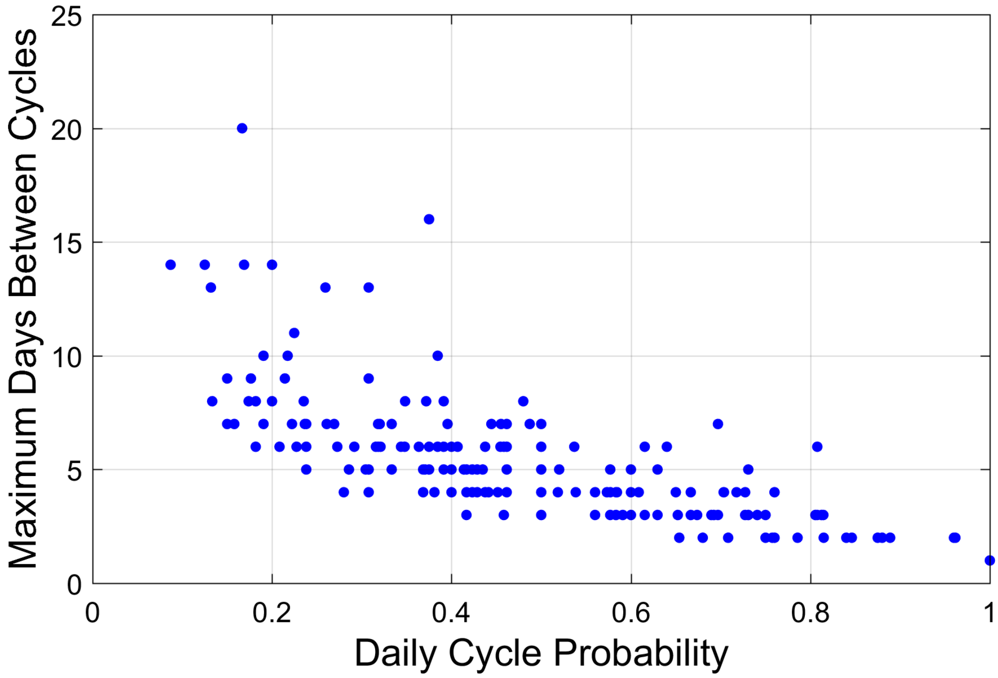
5.7.2 Daily Use Determination

5.7.2.1 Dataset Analysis

'Fixed' appliances have a lower typical daily use probability than the 'Simple' appliances, and are characterised by an increasing need with time. The data analysis focused on determining the average number of cycles per day per household, the relationship



(a) 'Next-day' cycle probability



(b) Maximum days between cycles

Figure 5.8. Relationship between washing machine per-household daily cycle probability and both 'next-day' use probability and maximum days between cycles. Data for analysis from the HES dataset [89].

between this value and the daily use probability, and the sequences of number of days between cycles. The average daily cycle number per household type is shown in Table 5.4.

The sequences of days with cycles varied significantly, particularly for households with a low daily cycle probability. The sequence is therefore assumed to be highly household specific with multiple associated factors, including occupancy and distinct household behaviours. However, there was distinct relationships between the daily cycle probability and both the 'next-day' use probability (use on the day following a cycle) and the maximum time between cycles (see Figure 5.8 for washing machine examples).

5.7.2.2 Module Calculation Sequence

The primary parameter used to define household use of this type of appliance is daily cycle probability. The steps required to determine this value and use on individual modelled days are as follows:

- *Steps 1-3* - Identical to Steps 1-3 for the 'Simple' module except daily cycle probability (*HhldPrb*) replaces average daily cycle number (*HhldCyc*) as the output from Step 3 based on the household type average daily cycle probability.

- *Step 4* - The 'next-day' cycle probability and maximum number of days between cycles are determined from the identified relationship to the daily cycle probability (see Figure 5.8). The Kernel Density method (see Appendix A) has been used to convert the relationships to a probability function to allow the 'next-day' use probability and maximum time between cycles to be probabilistically allocated to each modelled household based on the previously determined daily cycle probability.

- *Step 5* - To determine the number of days until the next cycle, for each household the module generates a cumulative probability function for the number of days until the next cycle randomly but constrained by the 'next-day' probability, identified average, and maximum time between cycles. These probabilities are further factored using the relative *OccUse* factor for each day (see 5.5.1.1).

- *Step 6* - The days in which the appliance is used are identified by using the cumulative probability distribution to determine the sequence of days between cycles for the simulated duration.

- *Step 7* - The number of cycles in a 'use' day is determined from a simple probability determination based on HES dataset analysis of the distribution of cycle numbers per 'use' day for each household type. Household type differentiation is used as multiple cycles are more common for family and multi-adult households.

5.7.3 Cycle Power and Duration

5.7.3.1 Washing Machines and Dishwashers

As outlined, washing machine and dishwasher cycles have broadly similar power profiles associated with distinct processes (e.g. water heating, rinsing etc.), but with variable power levels and durations that are both unit and selected wash type specific, and will also vary with incoming water temperature. This is highlighted by the smooth distribution of observed cycle energy values from the HES dataset and inconsistent energy used per cycle for individual households.

Energy used and cycle durations are strongly influenced by appliance energy rating (see 5.4.1). In 2014, over 80% of both appliance types owned had one of two energy ratings (for washing machines 27.9% A+ and 56.1% A, for dishwashers 79.0% A and 7.5% B) [166]. The data analysis and model calibration was therefore restricted to these ratings.

HES appliance diaries logged cycle type (wash, rinse, spin) and wash temperature over a 1-week period for some households. The HES dataset also provides specific appliance model information that in some cases also identifies the energy rating. Therefore, for each analysed energy rating, three typical full wash cycles (representing a low, medium and high temperature setting) plus a shorter typical spin/rinse cycle were identified for both power profile and duration, and used as the basis for all modelled cycles. Each household was probabilistically assigned an appliance energy rating, and one of the four typical cycles per modelled cycle.

For washing machines, analysis of the data determined that the distribution of full wash cycles was approximately 16% high (>60°C), 56% medium (50-60°C), and 28%

low temperature (<=50°C). Approximately 25% of cycles in the HES dataset have a total cycle power of <0.5kWh, which were assumed to be associated with additional rinse, spin, and very short, low temperature wash cycles. For dishwashers, the equivalent analysis determined that full cycles were approximately 35% high (>60°C), 32% medium (50-60°C), and 33% low temperature (<=50°C), and that 16% of cycles were low power rinse cycles.

There is insufficient washing machine cycle diary data to determine specifically how the probability of each cycle type varies per household. The identified base probabilities for full wash cycles are therefore further factored for each household type to account for observed average cycle power variations and a further ±20% variability is used arbitrarily to at least partially capture the expected cycle type behavioural differences per household. This value was selected based on the observed impact on the overall distribution of average power use per modelled household compared to the calibration data. The probability of 'spin/rinse' cycles is assumed from analysis of the cycle diaries to be at least partially related to the number of daily cycles. To achieve the 25% overall probability, this has been arbitrarily set at 50% of second and subsequent daily cycles (approx. 34% of all cycles) and 15% of first cycles.

For each household the module probabilistically sets a multiplier of between 0.85 and 1.15 to account for the range of cycle powers allowed within each energy rating, which is used to manipulate the base power values for the appropriate archetypal cycle.

For dishwashers, a similar basis for cycle type and power allocation was developed using equivalent data.

5.7.3.2 Irons and Vacuum Cleaners

Durations for iron and vacuum cleaner cycles are difficult to clearly distinguish as they are characterised by intermittent power draws typically at a fixed power level. The HES dataset was therefore analysed to determine a peak power equivalent duration ('peak-equivalent') for each identified cycle based on the total cycle energy divided by the maximum observed power value for the specific appliance. For each cycle a load factor was also determined based on the ratio of peak-equivalent to actual duration. This provides a standardised method for comparing cycles within and between households.

Each modelled household is probabilistically assigned a peak unit power and an

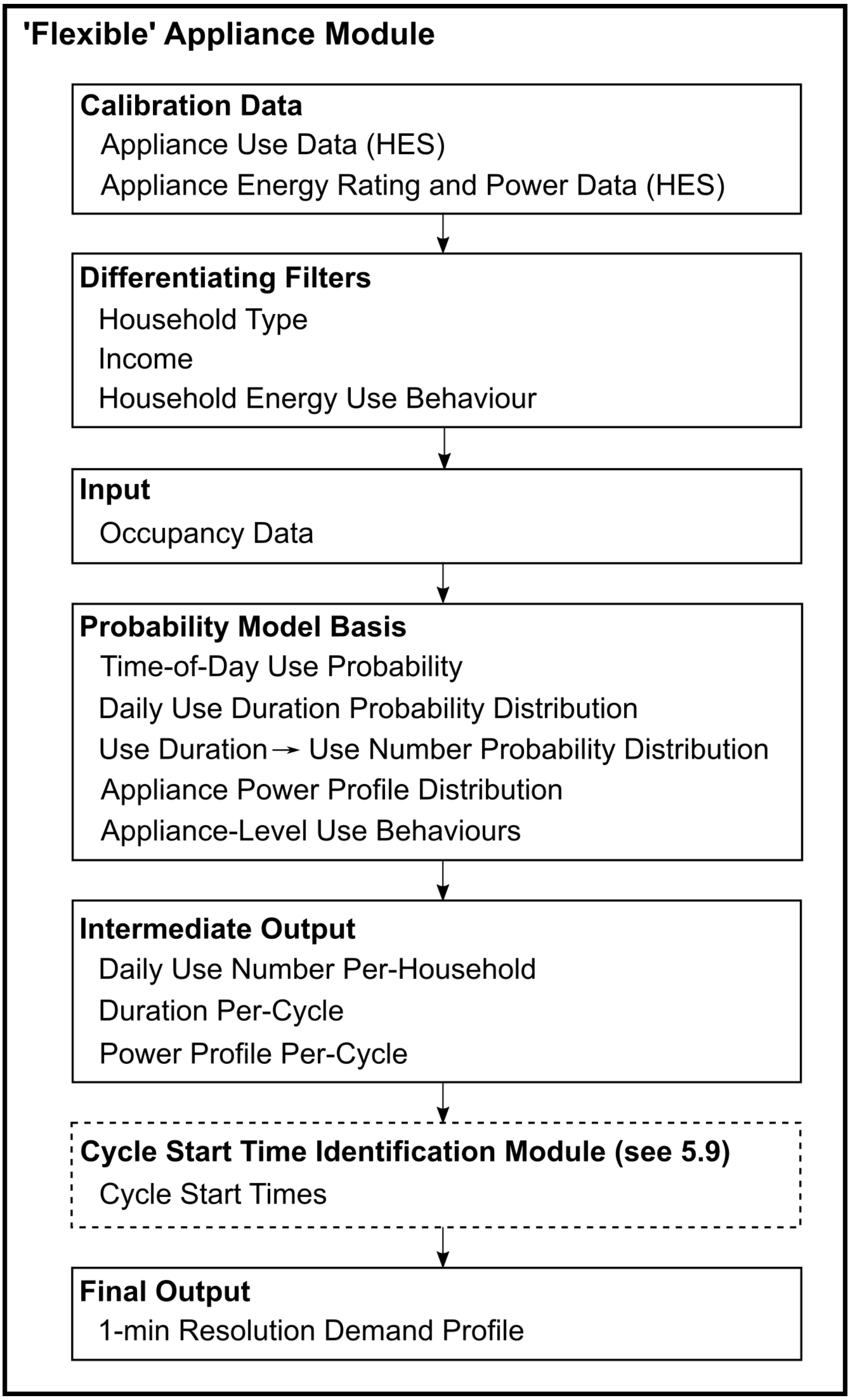


Figure 5.9. 'Flexible' appliance module structure.

average cycle load factor. An average daily peak-equivalent cycle duration is determined probabilistically from the daily use probability using the Kernel Density method (see Appendix A) to define the relationship from analysis of the HES dataset, and from this an average cycle peak-equivalent duration is determined from the average number of daily cycles.

Individual cycle peak-equivalent durations are determined probabilistically from the HES-analysed distribution of individual cycle durations to the household average, and individual cycle load factors determined randomly between 60 and 140% of the household average load factor based on typical variations. To reflect the inherent randomness of the power cycles, the cycle starts with a fixed full power period of 2 minutes and then each subsequent minute can be set probabilistically at either zero power or a power value between 0 and the peak power to achieve the required load factor for the overall cycle duration.

5.8 'Flexible' Module Development

The 'Flexible' appliance cycle module is applied to cookers, ovens, dryers, and computers. The overall structure of the module and principal defining elements as applied to cooker, oven, and computer use are shown in Figure 5.9. A modified version is applied to dryers based on the identified link to washing machine use as detailed in 5.8.1.2.

5.8.1 Daily Use Determination

5.8.1.1 Dataset Analysis

'Flexible' appliances have a complex relationship between number of cycles and the daily use duration. The length of each cycle has a significantly wider range than for the 'Simple' and 'Fixed' appliances and for that reason total duration rather than number of cycles is a more effective determinant of total energy used. Total daily use duration was therefore selected as the primary calibration parameter.

The mean daily use duration for each household type (see Table 5.5) and the overall range of ratios to the type mean were determined. The relationship between total daily duration and number of daily cycles is complex for each appliance with significant

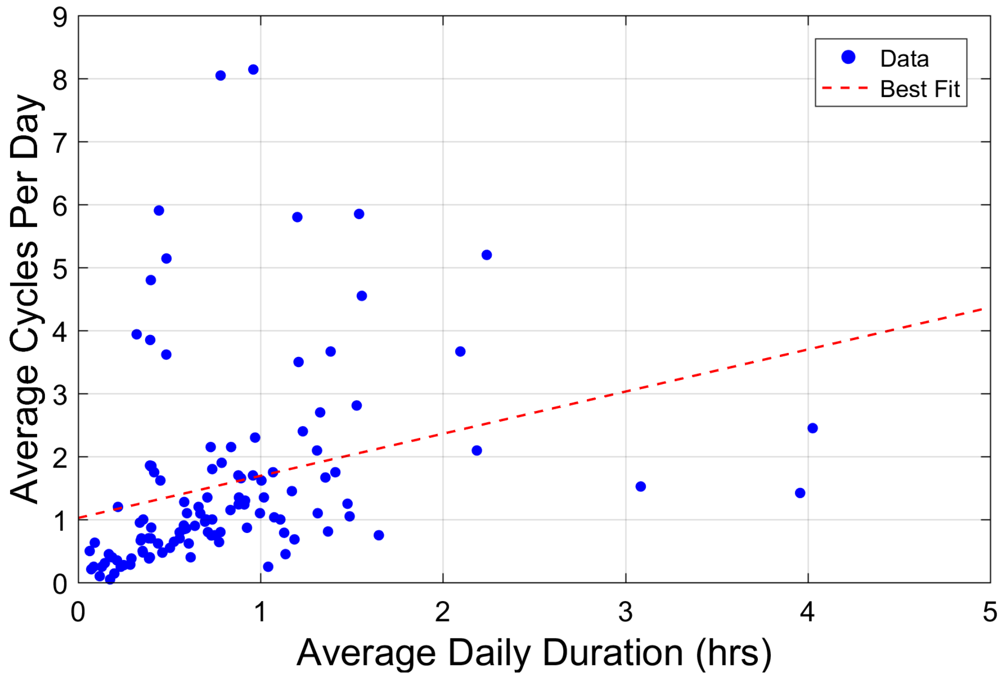


Figure 5.10. Relationship between average daily use duration and cycle number for cooker use per HES household. Data for analysis from the HES dataset [89].

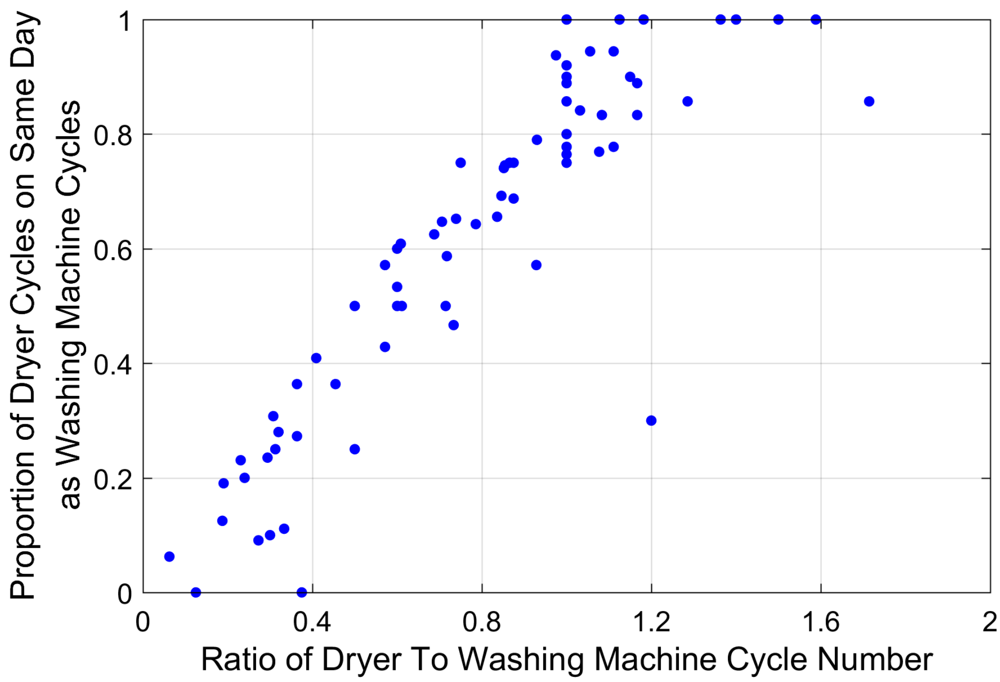


Figure 5.11. Relationship between relative household dryer and washing machine use frequency, and dryer use on the same day as a washing machine cycle. Data for analysis from the HES dataset [89].

variations in behaviour from few long cycles to multiple short cycles (see Figure 5.10 for cooker example).

As detailed below, some desktop computers in the HES dataset are powered continuously (c.14%) and others continuously powered during waking hours (c.7%). These units have been removed from the detailed cycle analysis and treated as constant loads. The remainder were analysed using the defined method, with the data in Table 5.5 for the units not identified as constant loads.

Table 5.5
Average daily use duration (in hours) by household type for all 'Flexible' appliances (combined data due to low ownership is identified with a (*)). Data for analysis from the HES dataset [89].

Household Type	Cooker	Oven	Laptop	Desktop	Dryers
Single, working age	0.75	0.24	1.50	3.60	0.27
Single, retired	0.48	0.23	0.90	2.60	0.27
Couple, working age	0.82	0.61	3.20	5.80	0.96
Couple, retired	1.01	0.41	1.70	4.30	0.59
Three-Adult	0.93	0.694	4.00	6.50	0.82*
3-Person Family	1.12	0.86*	3.00	4.10	0.82*
4-Person Family	1.03*	0.86*	4.40	7.30	1.32*
5+-Person Family	1.03*	0.86*	6.50	5.30	1.32*

5.8.1.2 Dryer/Washing Machine Use Relationship

Analysis of the cycle data for washing machines and dryers confirmed that their use was linked, with a high proportion of dryer cycles closely following washing machine cycles. On a per-household basis there was an average of a 64% probability that dryer cycles would occur on the same day following a washing machine cycle and a 28% probability within 2 hours. On a per-cycle basis, the equivalent values are 77.4% and 31.3%.

Within the module, washing machine use is treated as the primary variable. The relationship between the washing machine and dryer daily use probabilities are modelled using the Kernel Density method (see Appendix A), which is used to probabilistically determine the dryer use probability based on the washing machine probability. The same method is then used to define the proportion of dryer cycles that follow washing machine cycles based on the ratio of dryer and washing machine daily use. Except for several households with low washing machine use probability and higher dryer use, this relationship is broadly consistent (see Figure 5.11).

There is some evidence of other linked appliances, in particular cooker and dishwasher use, but the connection is less distinct than for the laundry appliances. Future work in this area is required to determine if there is sufficient correlation to justify inclusion in the model.

5.8.1.3 Module Calculation Sequence

As defined, the primary parameter used to define use per household for 'Flexible' appliances is average total daily duration of use. The steps required to determine this value and use on individual modelled days are as follows:

- *Steps 1-3* - Identical to Steps 1-3 for the 'Simple' module (see 5.6) except that average daily cycle duration (*HhldDur*) replaces average daily cycle number (*HhldCyc*) as the output from Step 3 based on the household type average duration (see Table 5.5).

- *Step 4* - Using the Kernel Density method (see Appendix A), the probabilistic relationship between average daily duration and average daily cycle number is converted to a cumulative probability matrix. The average number of daily cycles (*HhldCyc*) is determined probabilistically from the distribution appropriate for the determined average daily use duration (*HhldDur*).

- *Step 5* - The average individual cycle duration (*CycDur*) is determined by dividing the average daily use duration (*HhldDur*) by the average daily number of cycles (*HhldCyc*).

$$CycDur = HhldDur/HhldCyc \tag{5.7}$$

- *Step 6* - For each individual day to be modelled a multiplier (*DMR*) is determined from the day-specific *OccUse* factor (*OccUse_D*) compared to the overall average (*OccUse_H*) as shown by the following equation. The relative occupancy factor, *DROF*, is defined in 5.3.3.

$$DMR = (OccUse\_D/OccUse\_H)^{DROF} \tag{5.8}$$

- *Step 7* - The HES data does not allow the relative impact of occupancy on number of cycles and cycle duration to be determined. An equal impact is therefore assumed with the daily baseline number of cycles (*HhldCyc*) and cycle duration (*CycDur*) both being multiplied by the square root of *DMR*.

- *Steps 8-9* - The baseline number of daily uses (CycBase) and predicted actual number of cycles for a specific day is determined in the same manner as the 'Simple' module (see 5.6).

5.8.2 Cycle Power and Duration

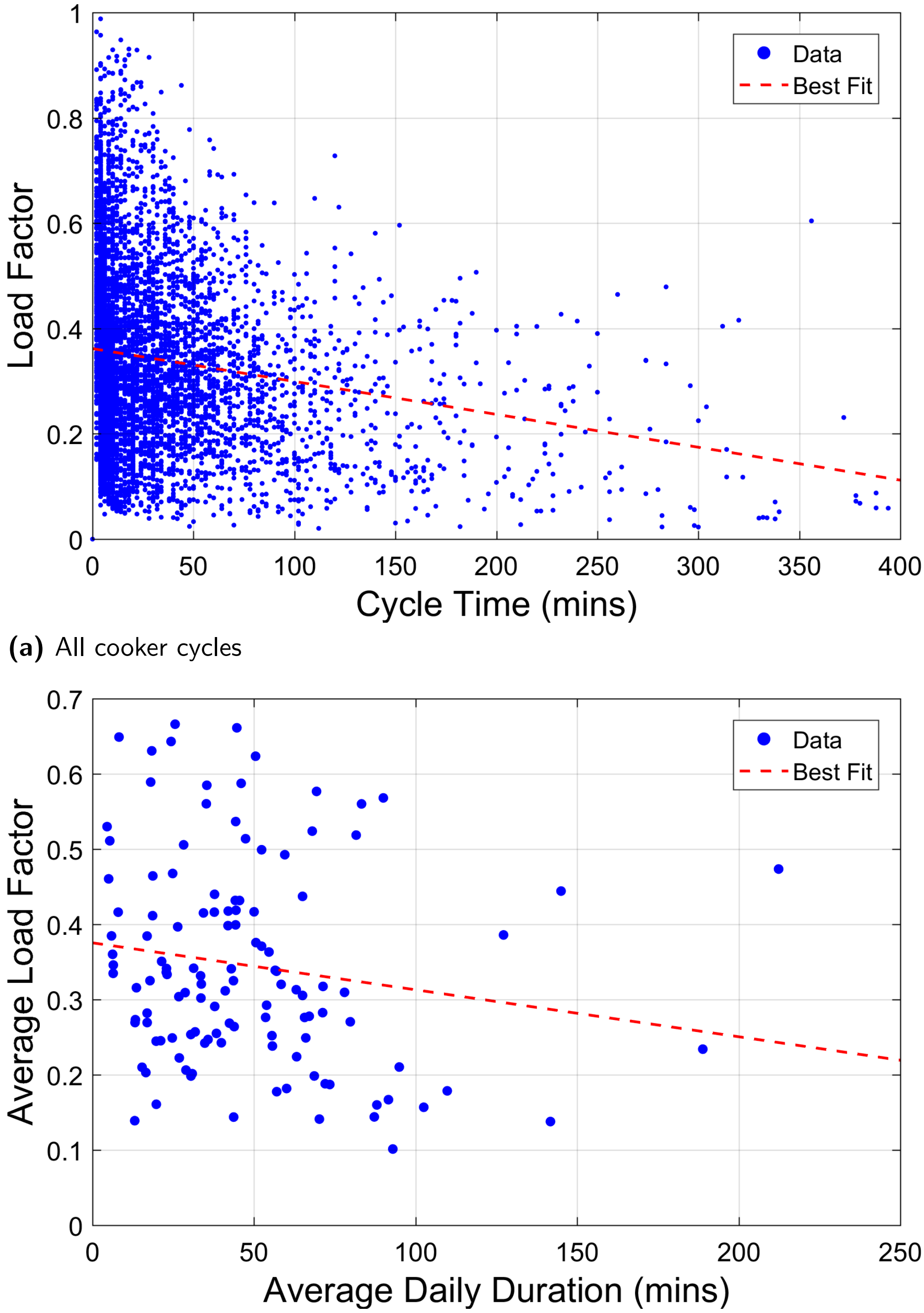
5.8.2.1 Cooker and Ovens

For cooker and oven cycles, the power is typically applied in sequential periods of full and zero power which are timer or thermostat-controlled. Ovens typically allow the temperature to be set across a wide range (60 to 240°C) and electric hobs will have 5-6 different settings. Use therefore has a significant additional variability as a result of user intervention in the settings and the number of elements (hob, oven, grill) used if a fully electric cooker.

The HES dataset power data for electric cookers is logged as a single value so distinguishing specific elements is difficult from the raw data. However, the HES dataset includes 1-week cooker use diaries for 44 households which allows an assessment of typical element use. The majority could be distinguished as oven only, hob only, grill only, and oven and hob combined use. Further analysis determined that the type of cycle was duration dependent, with short cycles more likely to be either hob or grill only use. The results of the analysis are shown in Table 5.6.

Table 5.6
Overall duration ranges and element use probability per duration range for cooker use. Data for analysis from the HES dataset [89].

Duration (mins)	0-10	10-20	20-30	30-40	40-50	50-60	60-90	90-120	120+
Overall	0.331	0.154	0.127	0.101	0.081	0.051	0.078	0.032	0.045
Hobs Only	0.650	0.750	0.620	0.400	0.400	0.300	0.120	0.120	0.05
Grill Only	0.204	0.146	0	0	0	0	0	0	0
Oven Only	0.146	0.100	0.350	0.365	0.400	0.400	0.450	0.450	0.408
Hobs+Oven	0	0.004	0.030	0.235	0.200	0.300	0.430	0.430	0.542



(a) All cooker cycles

(b) Per-household average

Figure 5.12. Per-cycle and per-household relationship between cooker cycle duration and load factor. Data for analysis from the HES dataset [89].

The diaries also identified the number of individual hobs used, which allowed the probability to be assessed. The data showed that there was a 64.3% probability of single hob use, 25.3% for two hobs, 8.3% for three, and 2.1% for four. Proportionally this is consistent with the analysis performed by Mansouri et al [167]. The duration of use for each hob was based on the per-ring data of Mansouri et al, also used by Stokes [147].

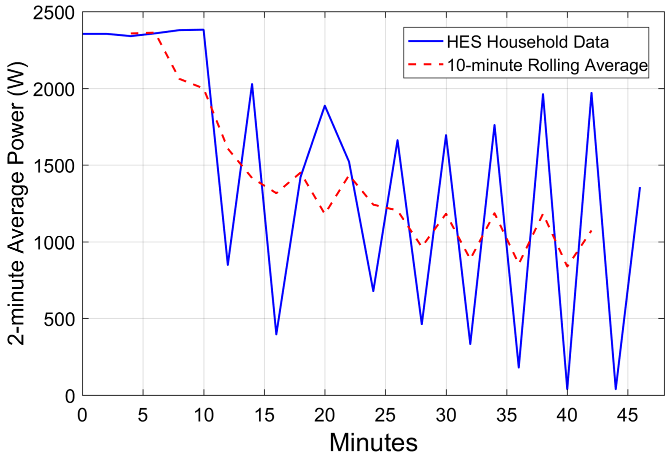
Analysis of the HES dataset did not clearly identify consistent profiles as a result of the 2-minute resolution, impact of user manipulation, and difficulty in distinguishing the elements used and their settings. Most cooker and oven cycles start with a period of several minutes of steady power demand as the initial element(s) warm, followed by highly variable power use. This is demonstrated by Figure 5.13, which shows two typical oven cycle power profiles on a 2-minute average basis from two separate households in the HES dataset.

To allow the overall impact on household power use to be adequately captured within a practical model, each identified cycle in the HES dataset was analysed for duration and a load factor, which is the average cycle energy divided by the observed overall appliance peak power (see Figure 5.12(a)). For each HES household the average load factor was determined and the relationship to the average daily use duration assessed (see Figure 5.12(b)).

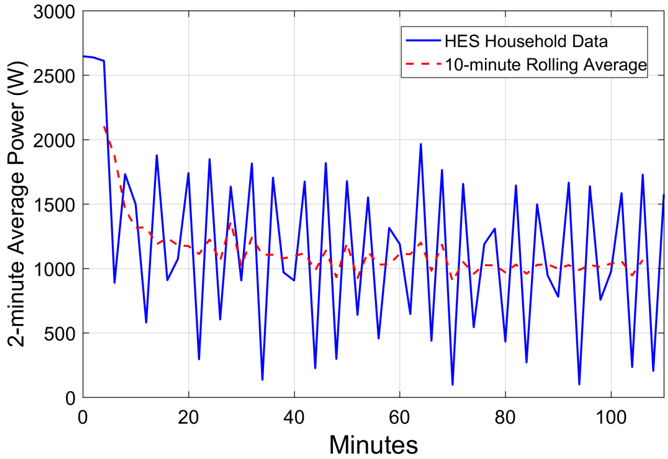
Each modelled household is probabilistically assigned a peak power value and average daily duration. The baseline cycle load factor is determined using the Kernel Density (KD) probability method (see Appendix A) developed from the relationship shown in Figure 5.12(b) based on the assigned average daily duration. A relative-to-mean load factor multiplier for each modelled cycle is determined based on the duration using a KD probability distribution developed from the relationship shown in Figure 5.12(a), with the per-cycle load factor determined from this multiplier and the household baseline load factor.

There is insufficient data to determine if there is a strong relationship between the number of elements used and the load factor. However, multi-element cycles are predominantly longer cycles (see Table 5.6), therefore this should be partially captured by the modelled relationship between cycle duration and load factor.

Cycles are modelled with two phases; a short duration continuous peak power pe-



(a) Household 1



(b) Household 2

Figure 5.13. Typical oven power cycle profiles from two separate households in the HES dataset.

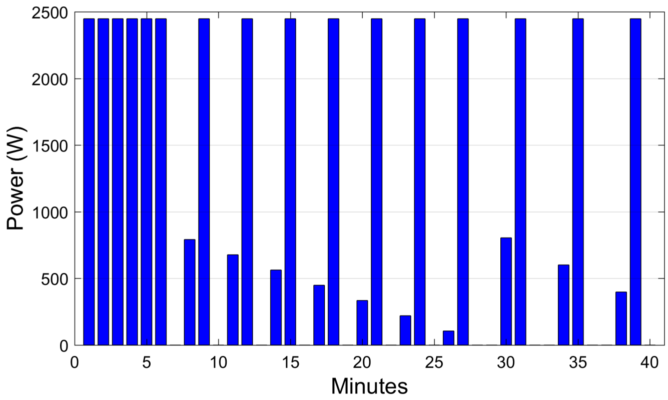


Figure 5.14. Example modelled oven power cycle.

riod, with the duration set between 2 and 10 minutes proportional to the cycle load factor; and then cycles of between 1 and 2 minutes of peak power use sequenced to achieve the target overall cycle load factor over the remaining cycle period. This aims to broadly replicate the typical power profile shown in Figure 5.13.

Where the module predicts the use of multiple elements, the module assumes a fixed residual load factor for all elements once the initial periods at maximum power are accounted for. This mimics the typical on-off control method used by this type of appliance, if not necessarily specific unit characteristics, to achieve a representative distribution of power demand during a cooker/oven cycle, which is considered sufficiently accurate for an overall demand model.

To reflect the observation that average power use tends to fall during the intermittent power phase, the module allows for this by setting a residual load factor multiplier that falls linearly during this phase. The multiplier is set arbitrarily and randomly in a range with a maximum variation of 1.5 to 0.5 and a minimum of 1 to 1.

An example is shown in Figure 5.14 for a 40-minute oven cycle with peak power rating of 2448W, cycle load factor of 0.450, and residual load factor multiplier range of 1.25 to 0.75. A continuous period at full power of 6 minutes is followed by an intermittent phase that has an average load factor of 0.353. The residual load factor starts at 0.441 (0.353 x 1.25), which is effectively modelled with a 3-minute cycle comprising one minute at zero power, one at 32% (representing a 19s period of full power within the 1-minute resolution basis), and one at full power. At the end of the cycle the residual has fallen to 0.264 (0.353 x 0.75).

5.8.2.2 Computers and IT Equipment

The use patterns of IT equipment, and in particular desktop computers and routers, can be simplified to units that are always on, use indicative of units consistently on during waking or occupied hours, and units that exhibit similar cyclic use to other intermittently used appliances. (IT equipment is defined as computers, routers, monitors, and printers for the purposes of the presented analysis)

For desktop computers, 21% were powered for at least 75% of the monitored period and 14% for over 90%. Of these half had low level standby level power use for the majority of the powered periods and peak use with the same characteristics as the intermittently used machines. The remainder exhibited high power use during most or all powered periods but with typically higher daytime levels. The desktop computer module assumes 10.5% are in constant use with a high minimum power use compared to the average and a further 10.5% are in constant use but with low minimum power use associated with periods of standby power. In contrast, no laptop computers were powered continuously reflecting different patterns of both use and charging for portable, battery-powered units.

All laptop and the remaining desktop computers are characterised by the same occupancy-driven use patterns as other appliances and are treated in the same manner. This also applies to active cycles for the constant use desktop units with low minimum power. Based on the HES dataset analysis, each household is probabilistically allocated a peak unit power from the overall distribution and an average cycle load factor (LF) based on total duration of use. Peak power from desktop units typically vary linearly from 50 to 170W with a small number of extreme outliers and laptops from 25 to 45W. Cycle periods are not limited by active occupancy, however, the LF is arbitrarily set higher for periods with active occupancy to reflect a higher probability of active use.

Average cycle duration is determined using the outlined calculation sequence. Durations for specific cycles are then determined based on a randomly selected value from the overall HES population distribution of the ratio of each cycle duration to the specific household average.

Routers are commonly left on for long periods and therefore, despite a typically small power requirement (<15W), can represent a significant base load for a household. Therefore, a simple model is used based on the range of observed power values and four identified typical use patterns (constant use (73%), waking hours (8%), occupied periods (8%), and during computer use (11%)) from HES dataset analysis. For 'constant use' a fixed power value is used for all timesteps, for 'waking hours' the unit is on while at least one person is awake (even if out), for 'occupied periods' the unit is on while at least one person is active in the dwelling, and for the remainder use is assumed for all periods of active computer use. Use variations were observed and the use pattern probabilities adjusted for specific household types, with, for example, single-person households having a significantly lower likelihood of longer periods of use. Routers typically draw a steady power with no variation with duration used and are modelled as fixed loads when in use.

Use of additional IT equipment, such as monitors and printers, can be assessed by comparing the probability of overlapping use. The probability of monitor use with laptop use per-household varies linearly from <1% to 95% with an average of 46%. Monitor use with desktop use is harder to define as in most cases monitor power is not separated and is therefore assumed for all periods of active use. The relationship between computer and printer use is less distinct with only an average of 6% overlapping use for laptops and 9% for desktops and is modelled as arbitrary short periods of additional power use during active computer use periods based on the per-household probability distribution.

5.8.2.3 Dryers

Unlike the other 'Flexible' appliances, dryer use is not primarily characterised by daily average duration as a result of the identified relationship between washing machine and dryer use (see 5.8.1.2). The average daily duration is therefore determined based on the average number of daily cycles in a reverse of the relationship used for cookers, ovens and IT equipment. Specific cycle durations are then determined based on the household average and a randomly selected value from the overall distribution of household cycle durations to the household average value.

Dryer cycles are characterised by periods at constant power with occasional short periods of zero power but with a restart within 10 minutes. Each household with a dryer is probabilistically assigned a fixed unit power and average cycle load factor based on HES dataset analysis. There was no discernible correlation between the two values and they are therefore assigned randomly. The relationship between cycle load factor and duration is similarly random and therefore no relationship is assumed.

The load factor and duration are converted to a representative power profile by first determining the number of heat periods within the cycle randomly with an arbitrary upper limit set by total duration. The durations and start times for each heat period within the overall cycle are randomly assigned, but restricted to achieve the overall load factor.

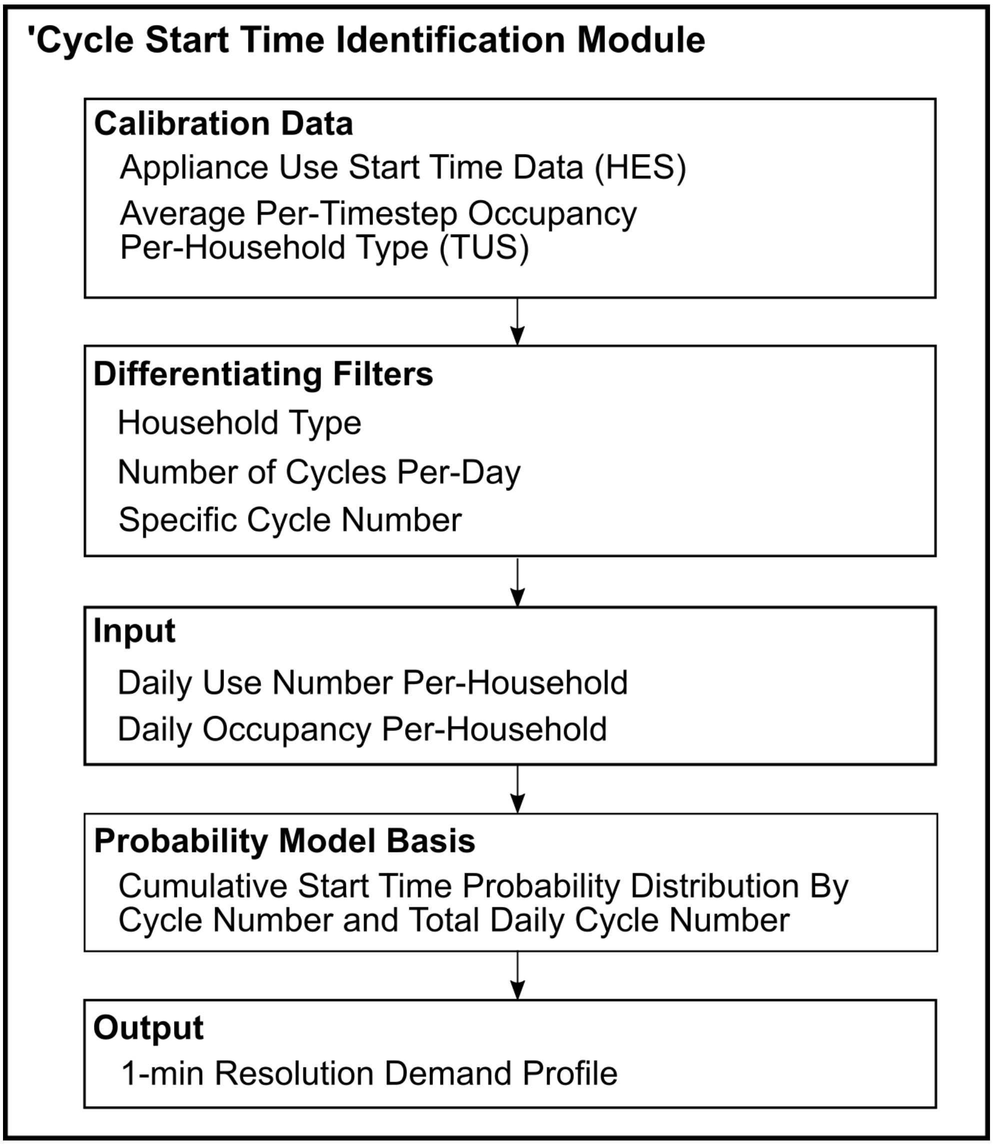


Figure 5.15. 'Cycle Start Time' module structure.

5.9 Cycle Start Time Identification Module

The three 'cyclic' appliance modules ('Simple', 'Fixed', and 'Flexible') use the same method to convert the occupancy model output and the determined number of cycles per day into cycle start times. The method is described below. The overall structure of the module and principal defining elements are shown in Figure 5.15.

5.9.1 Cycle Start Time Probability Distributions

Existing models have used either time-use survey activities or fixed per-timestep probabilities to define appliance use potential, but neither accurately captures both the time-dependency of the use of each specific appliance or how multiple cycles are distributed.

The HES dataset analysis to identify individual use cycles included identification of the start time for each appliance cycle. The overall distribution of these start times gives a probabilistic assessment of when a particular appliance is used. This provides both a simple and more direct means of modelling appliance cycle start times. Analysis of the distribution of cycle start times per appliance showed that timing was dependent on household type, occupancy, number of cycles per day, and time-specific drivers (e.g. meal times).

Separate cycle start time probability distributions were generated for each primary appliance based on household type, total number of daily cycles, and for each specific cycle number. As outlined in 5.2.3, the 2-minute HES dataset resolution was converted to a 1-minute probability distribution resolution by assuming an equal probability of a cycle start within each 2-minute period.

Furthermore, to allow these distributions to be used with the occupancy model output, the occupancy influence on each distribution must be removed. This is achieved by dividing the unmodified cycle start time probability density function (pdf_{raw}) at each timestep by the relative occupancy probability for the household type from time-use survey data (see Equation 5.9) and generating cumulative probability distributions based on the modified pdf values (see Figure 5.16 for unmodified and occupancy-modified

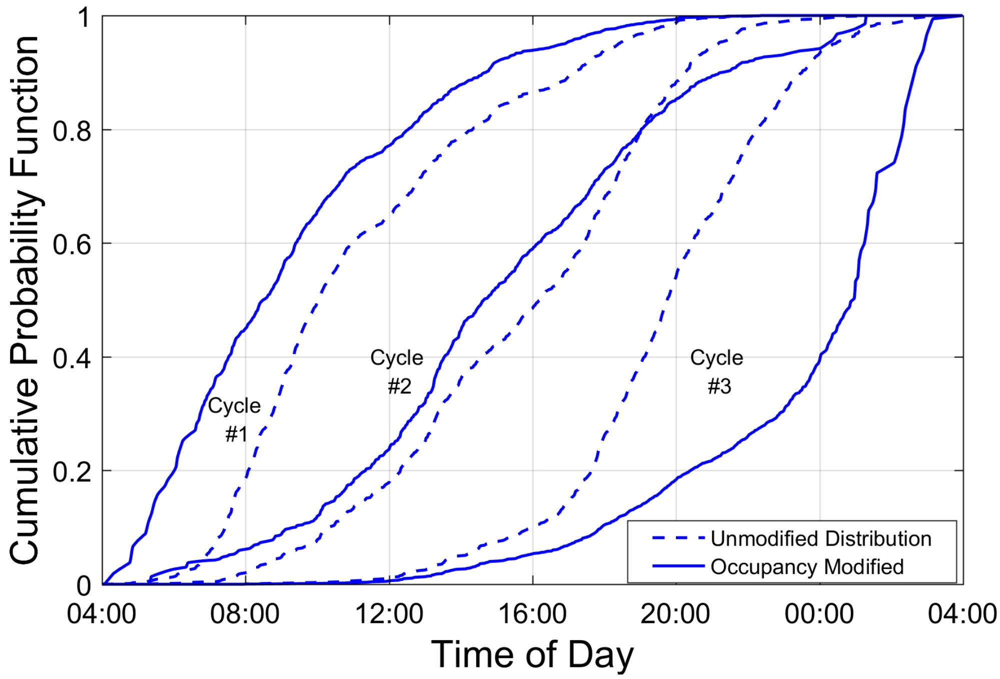


Figure 5.16. Unmodified and occupancy-modified cycle start time cumulative probability distributions for a 3-cycle microwave use day.

distribution examples).

$$pdf_{mod}(t) = pdf_{raw}(t)/(p_{occ}(t)/\bar{p}_{occ}) \tag{5.9}$$

Each household type has unique use cycle patterns beyond what would be predicted by differences in occupancy. Examples include; higher relative use of several appliances in family households in the 8pm-12pm period suggesting that child presence earlier in the day can restrict appliance use by parents; earlier relative use of laundry appliances in older households; higher relative evening use of cooking appliances in working age single-person households and lower use in retired single-person households.

To account for differences in cycle timing based on household type and insufficient cycle data to allow this to be captured for each cycle specific distribution, the overall cycle start time distribution for each household type was determined, and modified for type-specific occupancy using the same method as described above. The same process was also undertaken to generate a single distribution for all households. The probability density function (pdf) value at each timestep for the household type distribution was then divided by the equivalent ‘all household’ pdf value, with the resultant relative use per timestep distribution used to factor the specific cycle distributions for each household type to reflect distinct behavioural patterns. These modified distributions are used as the basis for linking occupancy with appliance use probability as described in the following section.

Table 5.7
Daily event matrix example - microwave - cycle #1 of 3.

Time	07.50	09.05	10.53	13.16	19.10	20.29	22.09	23.55
Occupants	1	0	1	0	1	2	1	0
Availability	1	1	1	1	1	1	1	1
Cumulative Probability	0.438	0.563	0.725	0.844	0.988	0.997	0.999	1.000

5.9.2 Cycle Start Time Identification Method

For each modelled household, the appliance cycle start times are determined based on the active occupancy periods generated by the occupancy model (see Chapter 4), the generated daily cycle number for each appliance (see 5.6, 5.7, and 5.8), and the occupancy-modified cumulative probability curves identified from the HES dataset for

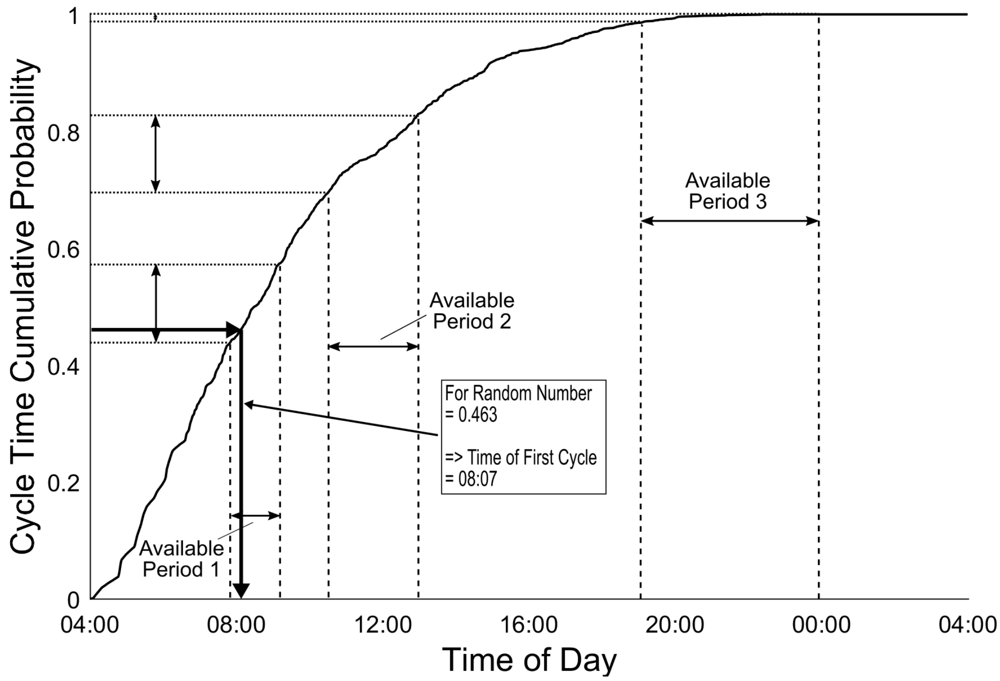


Figure 5.17. Cycle start time prediction method example.

each appliance (see Figure 5.16 for example).

For each modelled day, the occupancy model output is converted into an event matrix that tracks occupancy transitions and appliance availability changes (see Table 5.7). Potential cycle periods require both active occupancy and appliance availability at the start of the period to be greater than zero. Appliance availability tracks whether the appliance is already being used within the defined period.

The appliance cycle start time cumulative probability function values at the start and end of each 'available' period are determined from the appropriate occupancy-modified curve (see Figure 5.16). A generated random number, limited to values within the 'available' periods, is then used to determine the cycle start time (see Figure 5.17). Cycle events are therefore more likely during periods with higher use probability relative to occupancy to accurately reflect realistic appliance-specific behaviours.

For subsequent cycles, the event matrix is updated, with the previously identified cycle periods (including an arbitrary short dead period pre- and post-cycle to ensure adequate cycle separation) set as appliance unavailable (see Table 5.8). The next cycle start time is determined in the same manner using the next cycle start time probability distribution until the total number of daily cycles is reached.

Table 5.8
Updated daily event matrix example for cycle #2 of 3 following a 4-minute microwave cycle at 08.07.

Timestep	07.50	**08.04**	**08.14**	09.05	10.53	13.16	19.10	23.55
Occupants	1	**1**	**1**	0	1	0	1	0
Availability	1	**0**	**1**	1	1	1	1	1
Cumulative Probability	0.059	**0.062**	**0.066**	0.093	0.175	0.354	0.807	0.941
		↑	↑					

5.10 Audio-Visual Appliance Module

As defined in 4.7.2, TV use is predicted per-occupant by a secondary Markov chain model linked to any active period identified by the primary occupancy model. In addition, the following needs to be determined; whether other TV-linked appliances are being used (i.e. DVD/Blu-ray players, set-top boxes, and games consoles), the base power level for each appliance, whether there is shared use, and any standby

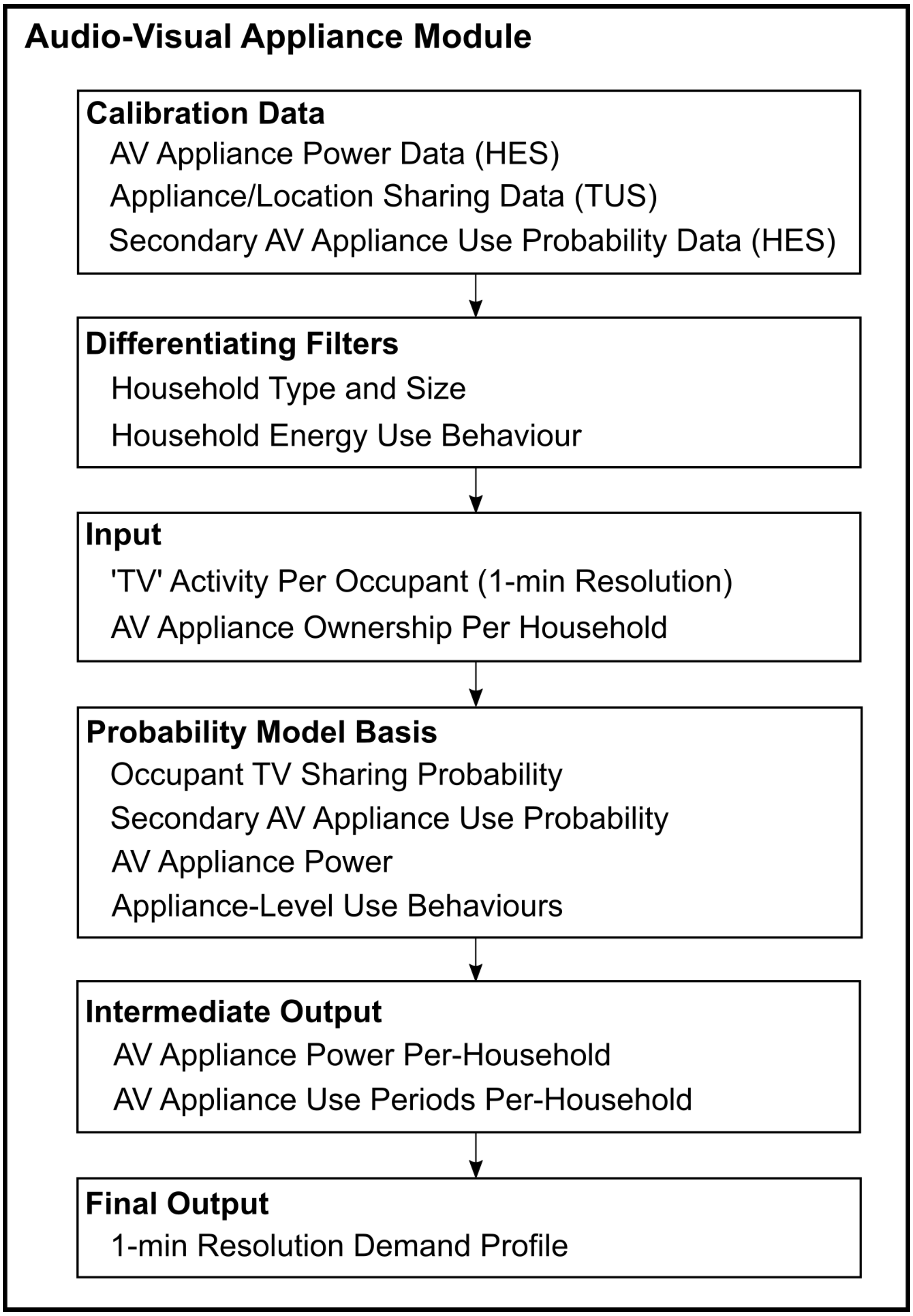


Figure 5.18. 'Audio-Visual' appliance module structure.

power demand. The overall structure of the module and principal defining elements are shown in Figure 5.18.

5.10.1 Dataset Analysis

5.10.1.1 Linked Appliances

The HES dataset was analysed to determine the per-household probability of a secondary AV device (DVD/Blu-ray players, set-top boxes, and games consoles) being used at the same time as a TV. No obvious correlation with either household type or number of owned televisions was discernible, therefore the probability was set randomly for each household based on the overall distribution.

Based on the assigned probability per-household, the probability of linked AV appliance use is determined at each timestep with a change in the number of people watching TV. The simultaneous use of a DVD/Blu-ray player and games consoles linked to the same TV unit is not allowed.

5.10.1.2 Power

All AV appliances have a relatively constant power demand and can be assigned a single value. For each TV in the HES dataset, the average power demand during 'on' periods were determined, and for multiple unit households, the units ranked in descending order. The highest power unit was assigned as 'TV1' and the demand compared for all households. There is a distinct difference in average unit power based on household type, therefore the data was further differentiated on this basis. Each household is assigned a 'TV1' power probabilistically from the household type distribution. This process was repeated for the 2nd, 3rd, all other TV units, DVD/Blu-ray players, set-top boxes, and games consoles, if owned.

5.10.1.3 Shared Use

Modelling TV use in a multi-unit, multi-person household by determining use per-occupant is complicated by the possibility of a single unit being used by more than one person. Collin et al [140] used the UK 2000 TUS survey [83] to determine the shared use probability. The TUS data includes a diary element where it is logged whether the activity was undertaken alone or together with another person. This allows determination on a per-timestep basis of the probability that the activity and location is shared (i.e. a single TV unit is used by more than one person) rather than happening concurrently but with the occupants in different locations.

The work presented by Collin et al is focused on two-person households and does not reflect the different relationships between adults and children or variations in sharing potential based on the size of the house and number of people watching television at a specific time. Each potential adult and child combination for two, three, four, and five-plus person households was therefore differentiated.

There are two main residual issues with this analysis. One is that when four or more people are watching television it is not clear if this is two or more sets of multiple people together or all people sharing an activity and location. However, this situation is extremely rare and is therefore assumed to reflect all people together for simplicity.

The second, more significant issue, is that in larger households some will be unit number constrained but also that in those households the potential for watching television may be lower overall. The UK 2000 TUS dataset includes whether one or more-than-one television is owned by a household. This allows all households with a single unit to be removed from the analysis. Beyond this, however, any potential for the model to underestimate use based on overestimation of sharing from unit-constrained households cannot be accurately determined from the currently available activity and demand data.

The results of the analysis show that sharing potential falls with increasing household size, as would be expected given the increased potential for additional units and rooms. For example, the probability of two adults sharing falls linearly from 77.2% in a two-person household to 51.3% in a five-plus person household. There is also evidence of a lower probability of sharing when children are involved. To illustrate, in a four-person household there is a 62.2% probability of sharing when two adults are watching but only a 20.2% chance of sharing for an adult and child. It should be noted, however, that the probabilities for children sharing TV units is significantly lower than for adults. It is suspected that there may be an under-reporting of location sharing for children in TUS diaries completed by their parents. The overall impact on energy use if child TV-sharing is underestimated is small, therefore the analysed data is used directly. Full sharing probability results are presented in Appendix C.

The potential for time dependence in TV sharing must also be determined. Data has been split between weekday and weekends. There is insufficient data to allow 10-minute timesteps to be used for the analysis, therefore the probabilities have been generated on an hourly basis. Given that television viewing is predominantly an evening activity and some of the modelled situations are rare, only time-specific data with more than 20 examples of the particular base scenario (e.g. two adults watching in a three-person household between 7pm and 8pm) are identified separately to ensure only statistically significant data was used. For time periods with fewer data elements the overall average for the scenario was used.

The analysis of the hourly data does not show any distinct time-dependent patterns, although there is some evidence of lower shared use in the early evening which increases during peak viewing hours (8pm to 10pm). The use of the hourly data has been retained but in most cases the overall average values (see Appendix C) could be used without significant error.

The potential for TV sharing can be modelled in two ways; absolute probability or transitional probability. For the absolute probability approach, the number of units watched is assumed to remain constant if no change in people watching. If a change occurs, the new situation is determined only by the relative probability of all potential new situations. The only influence of the previous state is that the new potential states are limited by the assumption that only the person(s) changing impact the number of units watched.

For example, if, in a 3-person household, 2 adults are watching one unit and a child starts watching television, it is assumed that the new situation can only be one or two units on. For this situation, overall there is a 57.8% probability of 2 units and 26.6% of one unit being watched, which is converted to a 68.5% probability for 2 units and 31.5% for one unit in this particular situation.

The transitional probability method uses a Markov chain approach to incorporate the change probability directly. This has the potential to accurately reflect real transition events and to allow transitions in units watched without a change in people watching to be captured. From direct TUS data analysis, the probabilities for the above example transition were 68.75% (11 of 16) and 31.25%, therefore a Markov chain model for the transition to an additional child watching produces similar results.

Analysis of other high frequency transitions showed a similar correlation between probabilities for the two methods. Examples of changes in units watched but not people watching were also very rare. The sparseness of the transition data requires significant consolidation to allow effective Markov chain models to be developed and does not allow any time dependency to be captured. Given the performance similarity between the two methods and the simpler calibration and computational requirement, the absolute probability method has been used.

5.10.2 AV Module Calculation Sequence

The following outlines the sequence of steps for the 'AV' appliance module:

- *Step 1* - For each owned appliance identified by the appliance ownership module, a power value is assigned probabilistically from the HES-calibrated distribution.
- *Step 2* - Total number of adults and children watching TV per 1-minute timestep is determined from the occupancy model output.
- *Step 3* - Each change in the number of adults and children watching TV is identified.
- *Step 4* - The shared use module determines the number of units on at each identified change event and the unit identifier(s), if more than one unit is owned, to determine the assigned power value and location.
- *Step 5* - At each change event the use of linked appliances (DVD player, set-top box, games console) is also determined from the assigned probability per household.

5.10.3 AV Standby Power

AV equipment is one of the main consumers of standby power. Standby use behaviour varies significantly between households and typical patterns within households are inconsistent. In line with the overall demand model aims, the standby use was assessed statistically and probabilistically rather than attempting to simulate individual unit or occupant behaviours using a more agent-orientated approach.

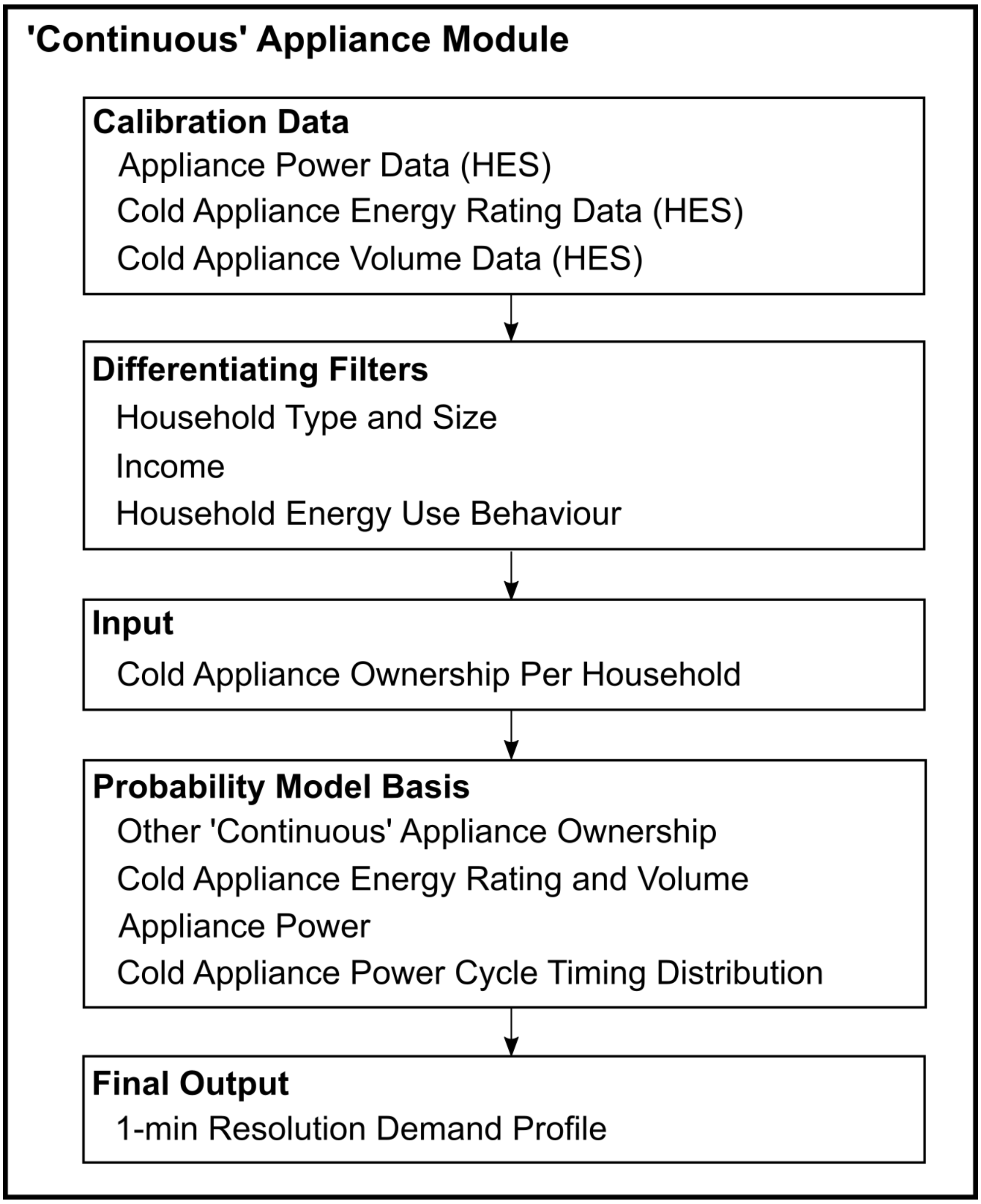


Figure 5.19. 'Continuous' module structure.

Total AV power use was analysed in each household in three periods,' Day' (11.30-13.30), 'Peak' (20.00-00.00) and 'Night' (02.00-04.00). The 75th percentile of the non-zero 'Peak' period power values was set as a baseline for normal use (selected as it captures a typical power level for most households but ignores outlier events). For both the 'Day' and 'Night' periods, the percentile values of the individual measurements were calculated and the ratio to the normal-use baseline determined.

Analysis of the resulting 'Day' and 'Night' percentile distributions shows that for most households there is a significant proportion of the distribution below 0.35 with typically distinct increases between this low level 'standby' demand and values close to and exceeding one (indicative of an 'on' period). The percentile distributions were therefore recalculated only for periods with ratios up to 0.35. The standby power use probability and level during periods with no 'on' cycles were then modelled from these distributions.

'Day' and 'Night' standby distributions for each household are generated probabilistically from the range of distributions for the relevant household type. The percentile distributions are converted to eight probability ranges (1=0, 2=0-0.05, 3=0.05-0.1....8=0.3-0.35). This data is then used to calibrate a Markov chain model to determine percentile probability distributions for modelled households by assessing the probability range at each percentile based on the previous percentile value.

For each modelled household, the 75th percentile of the non-zero AV power demand in the 'Peak' period is calculated as the baseline by running the main AV module first. The ratio of AV standby power to this baseline in relevant 'Day' and 'Night' non-cycle periods is determined by selecting a value probabilistically from the household specific standby distribution generated using the Markov chain model. In the 'Day' period, the value is assumed to be constant between 'on' cycles and, in the 'Night', the value is constant between the last cycle before sleep and the first cycle after waking.

5.11 'Continuous' Appliance Module

A subset of appliances in a household are continuously left in a powered state. This section specifically addresses appliances that are typically always on with relatively small changes in power profile under normal operating conditions. Where continuous use is related to a standby condition or extreme behaviour for a particular appliance (e.g. computers) this has been addressed within the appliance-specific module. The most significant contribution to overall consumption from this type of appliance is from cold appliances. Other applicable appliances include telephones, alarms, aquariums, and pond pumps. The overall structure of the module and principal defining elements are shown in Figure 5.19.

5.11.1 Cold Appliances

5.11.1.1 Dataset Analysis

There are four types of cold appliances commonly owned by households; fridges, upright freezers, fridge-freezers, and chest freezers. The HES dataset includes per-household ownership information such as type, model, volume, and energy rating (see 5.4.1). This allowed ownership, volume per appliance and household type, and typical power cycles per appliance and volume range to be determined.

Energy rating ownership data for the UK is taken from DECC ECUK data tables [29]. The majority of cold appliances owned in the UK have one of four energy ratings (A+, A-C). For example, for fridge freezers the proportions were 18% A+, 52% A, 7% B, and 3% C in 2014. The HES data was therefore differentiated based on these four grades.

Cold appliances typically cycle between periods of zero and full power which are either timer or temperature controlled. Analysis of the HES dataset shows a wide variation in demand patterns with cycles ranging linearly from 25 to 90 minutes. Newer, more efficient appliances tend to cycle more frequently with a lower 'on' power and overall energy consumption.

5.11.1.2 Module Development

The module requires five sequential assessments; number of cold appliances owned, type(s) owned based on number, energy rating for each appliance, volume of each appliance based on appliance and household type, and the specific power cycle based on appliance type, volume, and energy rating.

Number of appliances owned is determined based on the number of occupants, and the type(s) of appliances owned determined based on the appliance number from HES ownership analysis (e.g. one owned appliance is typically a fridge-freezer, the probability of a separate upright or chest freezer increases with number owned, etc.). Energy ratings are set based on the proportions for each appliance from the ECUK dataset, restricted to the four grades with highest ownership.

The relationship between unit type, energy rating, and volume to the average ($AvgP$) and peak power ($PeakP$) requirement, assessed from the HES dataset, is expressed in Equations 5.10 and 5.11, with a fixed and unit volume-dependent (in litres) factor as shown in Table 5.9 for fridge-freezers. These factors were determined by regression.

$$AvgP = C1 + C2 \times Vol \tag{5.10}$$

$$PeakP = C3 + C4 \times Vol \tag{5.11}$$

Table 5.9
Cold appliance average and peak power factors by energy rating - fridge-freezer example. Data for analysis from the HES dataset [89].

	A+	A	B	C
C1	0.504	0.700	0.909	1.189
C2	0.123	0.170	0.221	0.289
C3	80.6	89.2	110.6	121.4
C4	0.075	0.083	0.052	0.056

Based on the typical variation within the dataset from the regression-generated average result, the average was further varied by a factor between 75 and 125% and the peak power by a factor between 50 and 150%. The peak power varies more significantly as it is linked to the power cycle characteristics where the relationship between peak power and proportion of the overall on/off (duty) cycle when power is drawn can vary significantly. A simple linear representation of the variation is used rather than a more detailed probabilistic model as a result of the limited number of units in the HES dataset.

There is insufficient data to accurately set the duty cycle duration ranges per energy rating, therefore nominal ranges of 25-45 minutes for A+, 35-65 for A, 55-85 for B, and 70-90 minutes for C have been used based on analysis of several applicable units per

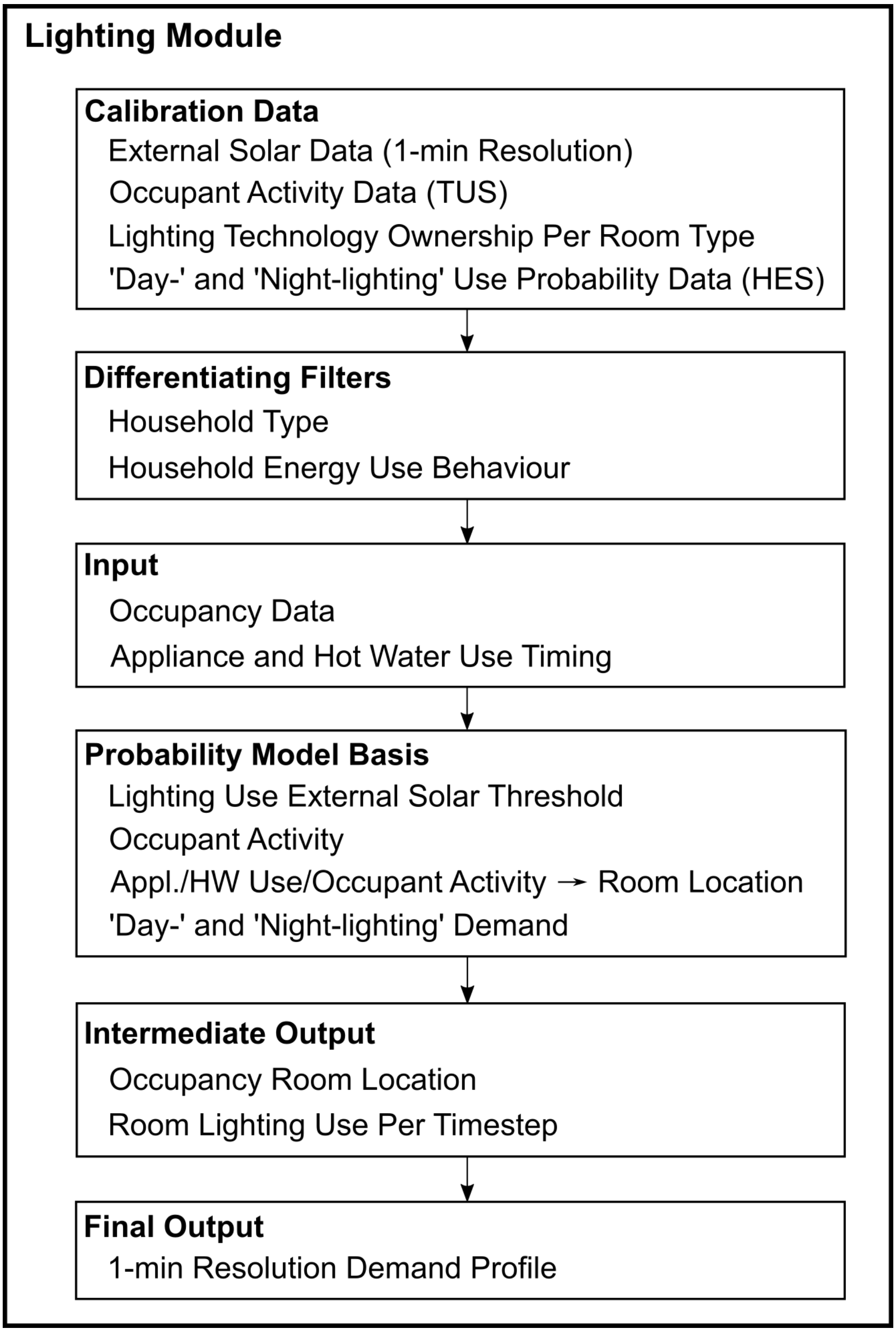


Figure 5.20. 'Lighting' module structure.

type and the typical reduction in duty cycle duration with more efficient appliances.

The time that the appliance is on within the cycle is adjusted for both seasonal and diurnal variations based on HES dataset analysis. The seasonal function is applied to all cold appliance types as type-specific assessments were not possible due to the short duration of analysis (typically 1-month), however, the diurnal function is appliance type specific. The overall seasonal adjustment equation and the appliance-specific time-of-day adjustment equation for fridge-freezers are as follows:

$$SeaAdj = 0.9748 + 0.1715 \times cos\,(0.0154 \times DayNum - 2.990) \tag{5.12}$$

$$DayAdj = 1 + 0.117 \times sin\,((\pi \times (MinuteNum - 752.37))\,/719.26) \tag{5.13}$$

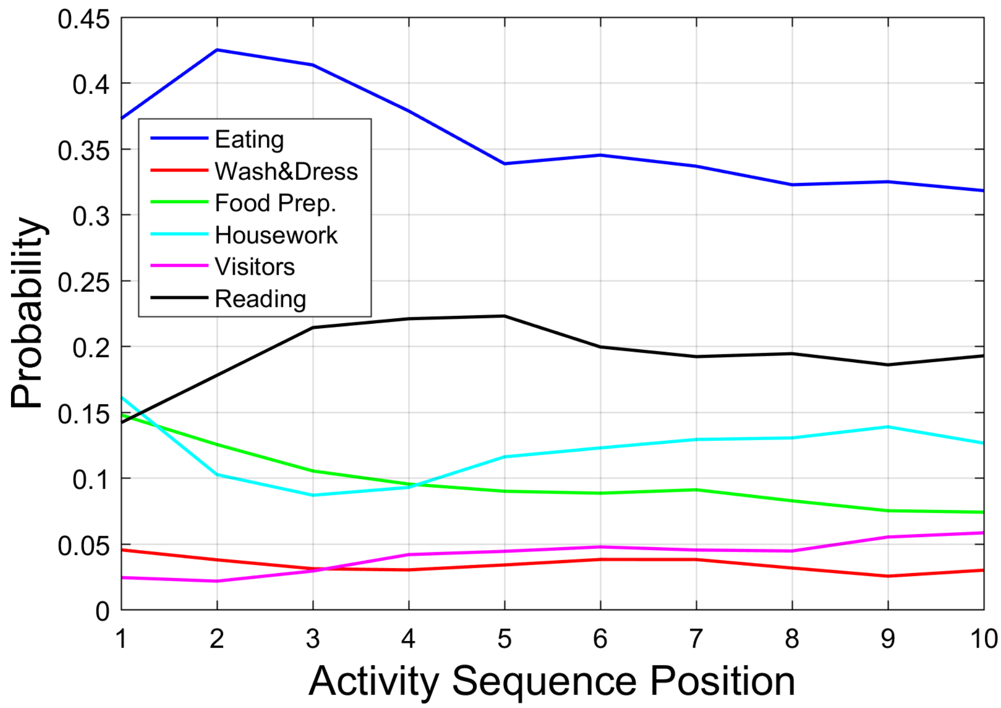
5.11.2 Other Continuously Powered Appliances

The other continuously powered appliances with either high ownership or high overall consumption identified in the HES dataset were cordless telephones, door bells, burglar alarms, pond pumps, aquariums, vivariums, and hot tubs. Each has been modelled as a constant load based on HES dataset analysis of power use, and ownership is based either on national survey data or, if unavailable, HES ownership probability.

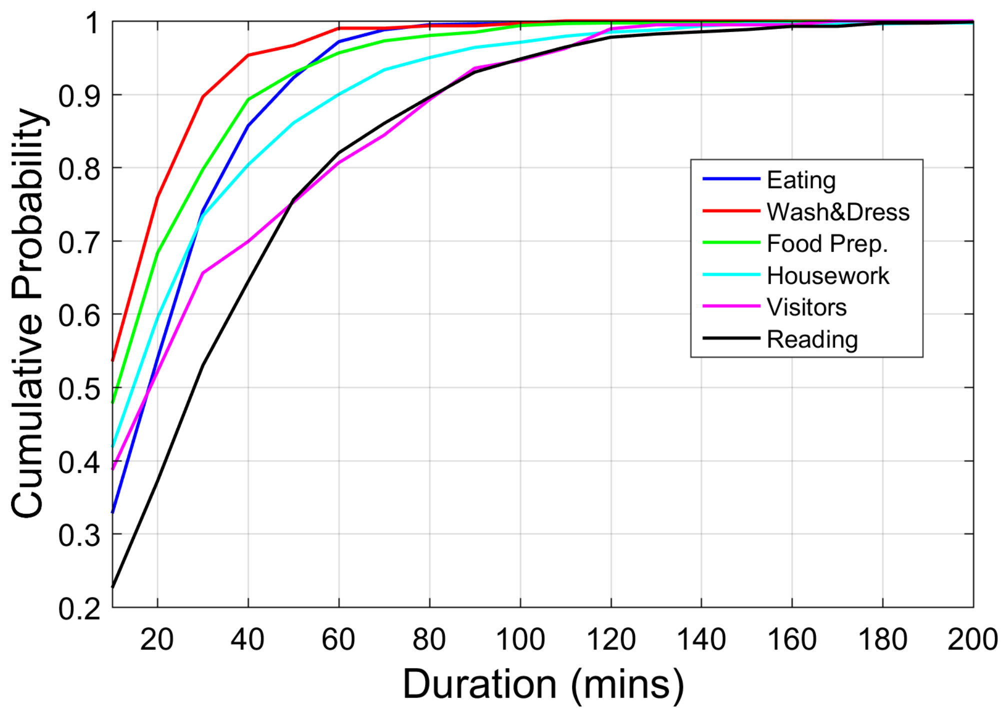
5.12 Lighting Module

The HES dataset has lighting power data from the main distribution boards (typically one per floor) and for individual socket-connected units. Also included are percentages of low energy bulbs per-house and average installed lighting power per room. However, there is no per timestep room-level power data, which limits the potential modelling methods. Analysis of the overall data also highlights considerable variation in both timing and typical power levels. The HES survey was carried out in the 2010-11 period during which the bulb types used for domestic lighting were rapidly changing to low energy variants, but at different rates for individual households.

A lighting module has therefore been developed based on identifying occupant location and using this and the external solar level to determine lighting need. The HES data was used for low-energy bulb proportion and for validation. The overall structure



(a) Activity sequence



(b) Activity duration

Figure 5.21. TUS activity sequence and duration probabilities following a pre-4pm return to the dwelling for a retired couple householder [83].

of the module and principal defining elements are shown in Figure 5.20.

5.12.1 Occupant Activity and Location Module

To ensure a degree of consistency between the appliance and lighting models, the occupant location module is first populated based on the identified appliance cycle periods. Appliance use is attributed randomly to active occupants with no previously allocated location. UK 2000 TUS activity probabilities are used to determine the location for any remaining unallocated occupied periods. The concept is similar to that used by Terry et al [113]. This is primarily a means to simulate the influence of multiple occupants, location sharing and transition likelihood, and therefore realistic lighting levels and level changes. The aim is not a highly realistic agent-based lighting model, but a model that is sufficiently statistically representative for an overall demand model.

For some appliances it is assumed that the occupant is present for the full cycle (e.g. kettle, microwave, toaster, hair dryer) and for others (e.g. cooker, washing machine, dishwasher, computers) only presence in the appliance location for a short duration at the start and end of the cycle period is assumed. The TV module accounts for shared use potential and different locations for specific units.

The TUS activity module that is used to populate the 'active' occupancy time periods which were not determined from appliance use was developed from analysis of TUS activity sequences and durations in six key periods associated with the starting and ending of active occupancy sequences (active period start: waking, return pre-4pm, return post-4pm; active period end: leave pre-4pm, leave post-4pm, pre-sleep).

The TUS activities are filtered to include only the thirteen most common 'active' activities (eating, wash&dress, food prep., dishwashing, housework, laundry/ironing, pet care, receiving visitors, telephone use, resting, reading, watching television, and listening to the radio) to simplify the analysis. The probability that a specific activity is the 1st, 2nd, 3rd, etc. activity undertaken following or preceding a specific occupancy transition is determined. This is then used to determine the sequence of occupant activities. The analysis also determined the distribution of activity durations for each of the thirteen activities, which is used to probabilistically allocate a duration per identified activity. The sequence of activity and duration assessments is repeated until the total duration exceeds the length of the occupied period.

For each individual in each of the five primary identified household types in the TUS dataset (working age single-person; retired single-person, retired couple, family, multi-adult), the sequence of TUS activities was determined from the start of each active period until six 10-minute timesteps from an active period ending. Active periods shorter than six timesteps were ignored. For the pre-sleep and leave periods only a fixed period of six timesteps is considered as there was a clear influence of the transition in the types of activities in this timeframe but not for periods further from the transition. As the model already captures a proportion of activities through appliance and hot water cycle identification, the probabilities for the associated activities have been reduced to account for this (e.g. 'Laundry' reduced by 75% to account for identified washing machine and dryer cycles, 'TV' reduced to zero as all cycles are assumed to be captured by the AV module).

Figure 5.21 shows example activity and cumulative duration probability distributions for a retired couple householder for a return to the dwelling in the pre-4pm period (only the six highest probability of the thirteen potential activities are shown for clarity). For this case, for example, there is a 10.5% probability that the third activity from returning to the dwelling before 4pm is 'Food Prep' and a 47.7% probability that it lasts for a single 10-minute timestep. The location for each activity is assigned using a simple set of common sense assumptions (e.g. for eating there is a 50% probability of the designated occupant being located in the kitchen or living/dining room area).

5.12.2 Lighting Demand Module

5.12.2.1 Installed Lighting

The installed lighting in the house is determined based on the floor area and the percentage of low-energy bulbs installed. Each room has a target average illumination level (in lux) as per Table 5.10 based on industry standard target levels [168]. Specific household values are randomly varied by between 75 and 125% to allow for a degree of variation.

Living rooms, kitchens and bedrooms have two lighting levels, 'main' and 'task', that can be used separately or together. 'Task' lighting represents high intensity sources

Table 5.10
'Main' and 'Task' target lighting levels (in lux) per room type.

Room	Kitchen	Bathroom	Living Room	Bedrooms	Hall/Stairs
'Main'	300	150	300	200	100
'Task'	150	n/a	150	100	n/a

(e.g. lamps) used to illuminate small areas of the rooms for specific tasks (e.g. reading lamps). The lighting intensity (main, task, or both) is dependent on the modelled occupant activity (see 5.12.1).

Table 5.11
Proportion of floor area allocated to each room type by number of bedrooms.

Room	Kitchen	Bathroom	Living Room	Main Bedroom	Other Bedrooms	Hall/Stairs
1-Bedroom	0.2	0.15	0.3	0.25	n/a	0.1
2-Bedrooms	0.15	0.12	0.28	0.2	0.15	0.1
3-Bedrooms	0.15	0.1	0.25	0.16	0.12	0.1
4-Bedrooms	0.15	0.1	0.2	0.15	0.1	0.1

Specific room area is determined from the overall floor area using the representative proportions in Table 5.11. Installed lighting per room in lumens is determined by an approximation based on the allocated target illumination (in lux) multiplied by room floor area (m^2). The proportion of installed lighting used by occupants is selected randomly between 50 and 100%, which allows for variable external lighting level influence, use of dimmer devices, and different behaviours.

Table 5.12
Low energy lighting percentages from HES dataset by household type (2011 basis). Data from the HES dataset [89].

	Low Energy Lighting (%)				
Hhld Type	0-20	20-40	40-60	60-80	80-100
Single, working age	0.345	0.345	0.069	0.207	0.034
Single, retired	0.303	0.152	0.212	0.091	0.242
Couple, retired	0.28	0.24	0.36	0.04	0.08
Family	0.306	0.222	0.208	0.181	0.083
Multi-Adult	0.232	0.333	0.247	0.130	0.058

Based on analysis of the installed lighting carried out by others using the HES dataset households [150], the percentage of low-energy bulbs installed per-household is assessed probabilistically (see Table 5.12). This is primarily to allow the current basis to be validated with both the HES dataset and other datasets observed in the same period. The final model basis will be reconfigured to reflect that the percentage of low-energy bulbs per household has changed significantly since 2011 and will need to be frequently updated for changes in installed technologies.

Based on the assigned percentages each household light source is first defined as being low or high energy and then the specific technology is assigned based on this assessment and the proportion of each technology in each room type [150]. There are six options defined (High Energy - incandescent(12), halogen(14), low-voltage halogen(25); Low Energy - compact fluor.(60), fluorescent(90), and LED(60)). Each has a different lumens/watt value (shown in brackets), based on 2010 data from [169] for the HES-comparison calibration basis.

5.12.2.2 Conditional Lighting Demand

Determination of lighting demand also requires that external lighting levels are assessed. The range of threshold solar levels for switching events was determined by comparing the timing of the main lighting use transitions in the morning and afternoon from the HES data households with average England-wide solar levels. Without occupancy and detailed location data, detailed analysis is difficult as occupancy-driven lighting use cannot be distinguished from solar-driven use and exact sunrise and sunset times are unknown. For most households, the morning decrease and afternoon increase are consistent with external light levels between 20 and 50 W/m. Allowing for some occupancy-driven use that distorts the analysis, baseline lighting use thresholds are set randomly between 30 and 70 W/m for each household. This is a similar basis to the 60 W/m average and 10 W/m standard deviation used by Richardson et al [69].

The HES-population equivalent model specifically uses 1-minute resolution synthetic external light level data from the solar model developed by Bright et al [170]. Nottingham location data is used, which is assumed to be close to the average for the HES population. The same solar model can be used for any UK location by setting latitude, longitude, and height above sea level.

Whether a lighting event (switch on/off, location change, intensity change) occurs is predicted using a set of assumed rules. These are as follows:

Lighting is switched-on if lighting is off and:

- the room is occupied, and external light level is below the lighting use threshold for at least 8 minutes of the 9-minute period centred on the current 1-minute timestep.

- the room becomes occupied and daylighting use is predicted (see below).

Lighting is switched-off if lighting is on and:

- the external light level is above the lighting use threshold for at least 8 minutes of the 9-minute period centred on the current 1-minute timestep.
- the room is not the living room and the room becomes unoccupied.
- there is a 25% probability of a switch off if the room is the living room, and the occupant moves to the bathroom or kitchen.

5.12.2.3 Additional Lighting Demands

Terry et al [113] highlighted levels of day and night use in the HES data that could not be directly attributed to external solar levels and occupancy respectively. These lighting uses therefore need to be modelled separately.

For the analysis of day and night use in each household in the HES dataset, a typical peak use lighting level is determined to set a baseline for the overall lighting use in the household. Percentile analysis of all non-zero lighting periods determined that the 90th percentile represented a suitable measure as it discounted the typically small percentage of extreme use periods and captured typical peak usage. This value is hereafter known as the 'household baseline'.

Daytime Use - For daytime use of lighting, when external solar levels would not predict a requirement (i.e. for rooms with restricted or no external lighting, task-specific lighting, user behaviour), determining household specific behaviour is difficult as the HES data does not track occupancy. Analysis of demand in the 11.30am-1.30pm period for households monitored between April to October determined that 25% of households had no lighting demand. Of the remainder, there was significant variation but generally characterised by significant periods of zero and low power use (<30% of the household baseline) and occasional periods of higher level use. There was no obvious correlation between daytime use and household type beyond that which would be predicted by occupancy variations.

The data analysis determined for each HES household the proportion of time during the 11.30am-1.30pm periods with non-zero lighting use ('Low' probability) and the proportion that it exceeded 30% of the household baseline ('High' probability). For each modelled household, the 'Low' probability is determined probabilistically from the analysed HES household distribution. The 'High' probability is then determined from the analysed relationship between the 'Low' and 'High' factors using the Kernel Density approach (see Appendix A). Daytime use is then modelled in three ways to incorporate each potential driver:

- External solar driven use.
- The 'Low' level non-solar driven use is determined from the assigned 'Low' probability. The probability is further manipulated using a skewed beta distribution random number model based on the household daytime occupancy compared to the overall average (60.2%). At each occupancy state change it is determined if there is a low-level lighting demand from the assigned probability and the level is selected randomly from the observed range up to 30% of the evening use baseline. 'Low' demand can be assigned for both occupied and unoccupied periods to reflect lighting left on for security etc. and is only assigned for periods with no other lighting use.
- The 'High' level non-solar driven use is determined from the assigned 'High' probability in the same manner as solar-driven demand but with the solar threshold replaced by a probabilistic determination using the 'High' factor at each occupant location transition. Daytime 'High' use is restricted to the lower 'Task' level lighting (see 5.12.2.1) based on analysis that average non-zero lighting use in the daytime period is c.60% of the evening period power level.

Night Lighting Use - Data analysis and modelling of night use of lighting is more straightforward than for day use as occupancy and external solar influences are less significant. It can also be assumed that lighting does not change between the last person sleeping and the first person waking.

In a similar manner to day use, the HES dataset lighting use for the 2am-4am period was analysed to determine the power level distribution relative to the household baseline. The distribution was simplified to the probability of lighting demand in five relative power ranges (0%, 0-10%, 10-25%, 25-50%, 50-75%) with values in excess of 75% ignored on the assumption that those are associated with active occupancy.

To capture the significant range of household behaviours, the data is converted to residual percentages (e.g. 25% in the 0% range, 25% in the 0-10% range and 50% in the 50-75% range converts to 75-50-50-0). The Kernel Density probability method (see Appendix A) was used to develop a method to generate realistic variations per modelled household. The percentage in each range is determined from the remaining sum of residuals based on the HES data analysis. Each household is probabilistically allocated a 'sum of residuals' (for the above case the sum is 175), which is then used as the input to the first KD matrix to determine the percentage of zero use. The process is repeated for each range using the remaining sum of residuals (100 in the above case for the 10-25% range calculation) to determine the percentage in the next range.

For each appropriate period, the night-lighting relative range based on the allocated probabilities is determined, and then the exact level based on the distribution of values observed in each range from the HES analysis. This determines the fixed lighting level for each night-time period. The probability of the night-lighting switch on has an equal probability as each occupant transitions to sleep and similarly the switch off has an equal probability as each occupant wakes.

5.13 Miscellaneous Appliance Module

The 'Miscellaneous' module covers two electricity demand elements. One is related to appliances that have low ownership or total energy consumption. The other is unacknowledged appliances in the HES dataset that are identified by the difference between the socket distribution board power measurements and the total at the same timestep for all individually monitored appliances. The same basic method is utilised for both as outlined below. The overall structure of the module and principal defining elements are shown in Figure 5.22.

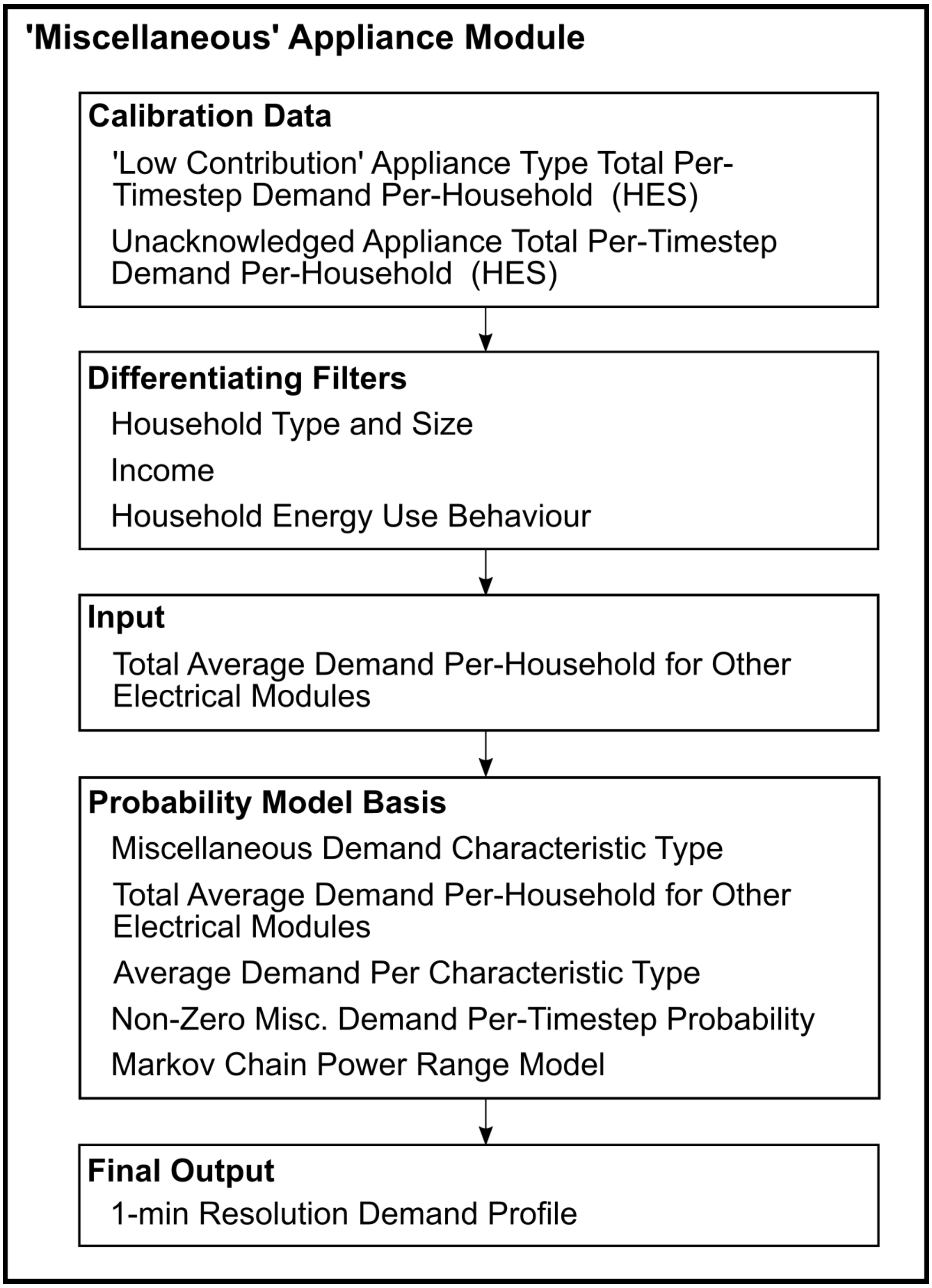


Figure 5.22. 'Miscellaneous' module structure.

5.13.1 Low Contribution Appliance Data Analysis

The HES dataset includes a number of appliances that have low levels of ownership or energy consumption. For these appliances there is both insufficient data and contribution to overall demand to justify simulating each separately. These appliances have therefore been consolidated into three separate groups (‘Cooking’, ‘Entertainment’, and ‘Others’), with each group treated as a single entity for each household. The three groups account for only <1%, <1% and 2.5% of total power respectively.

Each household in the HES dataset was analysed for each group of appliances. The proportion with no applicable appliances was noted and for the remainder the total demand of each group per timestep was determined. Use per group for each household was characterised by three factors: average power, peak power, and the proportion of non-zero timesteps.

Consistent for each appliance group, two distinct patterns of use were observed. Most households had either intermittent use (<10% non-zero timesteps) that would be associated with cyclic appliances only or almost continuous demand (>90% non-zero timesteps) that suggested at least one ‘always-on’ appliance. Each identified miscellaneous group was therefore further differentiated on this basis.

Similar to the ‘Simple’ appliance analysis (see 5.6), for each miscellaneous appliance group, the overall distribution of the ratios of each household mean demand to the household type mean is similar for all household types. The mean power for each household can therefore again be determined by random selection from the overall ratio-to-mean distribution multiplied by the household type mean.

There is a partial correlation between overall average miscellaneous group power and non-zero timesteps, and between the average power for the non-zero timesteps and the ratio of peak power to this value. The Kernel Density method (see Appendix A) has been used to capture these relationships probabilistically, allowing both non-zero timestep percentage and peak power to be modelled from a pre-allocated average appliance group power.

The dataset analysis therefore allows it to be determined probabilistically per appliance group if a household owns one or more relevant appliances, average power use, peak power use, the percentage of time in use, and the average power while in use.

5.13.2 Unacknowledged Appliances Dataset Analysis

The HES dataset provides both total socket demand monitored at the distribution box and individual appliance demands. There is a significant discrepancy between the socket measurement and total appliance demand for many of the households. Overall this shortfall accounts for c.24% of demand (c.100W average). This is presumed to be either unmeasured appliances or mobile appliances not reliably monitored every time. 85% of HES households had this type of unacknowledged demand.

Comparison of the national appliance ownership data with the measured HES appliances show that use data for a significant number of expected appliances is not included. This accounts for 35% of the discrepancy (35W average). The remainder (65W average) is treated in the same manner as the 'Low Ownership' appliances.

For the HES equivalent validation model described in the 'Validation' section, appliance ownership is as per the measured data and the full 100W 'unknown' element is included in the model. For the final demand model, appliance ownership is based on national survey data, therefore the 35% ownership-related shortfall is not included, only the residual 65W element.

Analysis of the per-timestep average 'unacknowledged' demand highlighted four distinct patterns; occupancy correlated (39% of households), continuous plus an occupancy correlated element (18%), continuous with specific peaks (20%), and specific peaks only (8%). The module distinguishes between each type.

As for the 'low contribution' appliances, average and peak power, and non-zero timesteps percentage were determined for each household and used to generate probability models that allow the demand characteristics to be simulated.

5.13.2.1 Module Development

The method selected to simulate the miscellaneous appliance demand is based on a similar first-order Markov chain approach used by existing occupancy models (see 3.3.1 for description). Applied to demand this is similar to the method developed by McLoughlin et al [78], which they applied to the overall household power demand rather than a defined portion and without time dependency.

Analysis of the demand patterns both for individual households and averaged for each specific appliance group (cooking, entertainment, others, and unacknowledged) suggests that the demand is highly stochastic at the household and per-day level but, averaged across all households, follows expected patterns based on occupancy and need (particularly for the cooking and entertainment groups). This suggests that any efforts to develop a household-specific model basis would be both mathematically challenging and potentially unrepresentative when applied to other populations.

In particular, the selected method is limited in that it does not easily allow occupancy to be incorporated in the demand timing and a significant amount of this potential household-specific detail is lost. However, it has the benefit of being mathematically straightforward, capable of handling combinations of cyclic and continuous demands, and able to capture the overall demand behaviour impact.

The 'states' used by the Markov chain model in this case are ranges of demand relative to the non-zero average value. The demand per timestep for each household and appliance group is therefore first converted to a ratio to the non-zero average power value and then each timestep value is converted to an integer between 0 and 9 based on the ratio ranges (i.e. '0'=0, '1'=0-0.2, '2'=0.2-0.4, '3'=0.4-0.6, '4'=0.6-0.8, '5'=0.8-1, '6'=1-2, '7'=2-3, '8'=3-4, '9'=4+).

A 10-minute timestep calibration is used due to the data depth required for a stable 10-element (10x10) Markov chain model. The average power value for each 10-minute period is therefore used for the ratio range integer conversion. At each timestep, the probability for each of the 100 potential transitions is determined to populate the Markov chain transition matrices.

The Markov chain probabilities at each timestep reflect the average non-zero timestep probability for the appliance group and pattern (e.g. cyclic vs 'always-on' for the 'low contribution' appliances). For example, for the 'Cooking' group the values are 6.5% ('cyclic') and 90.0% ('always-on'). To account for household specific patterns, the zero-value probability (range '0') is adjusted by a multiplier that accounts for the ratio between the model allocated non-zero timestep probability and the appliance group average, and all transition probabilities are rebased in order that they sum to one. The factor was determined by running the model for a range of adjusting multipliers and generating equations representing the relationship between the non-zero timestep value and this factor. For the 'cyclic' groups this was best modelled with a power function and for the ‘always-on’ groups with a double exponential function.

The first run of the module sets a baseline sequence for the household based on the ratio ranges calibrated for the average non-zero timestep probability. The deviation from the target average power as a result of the household-specific non-zero timestep percentage is determined and all non-zero values less than the maximum power are rebased in order that the final average power equals the target value for the household. The Markov chain model therefore provides a representative sequence of relative changes that are then manipulated to account for the household-specific characteristics.

5.14 Electricity Demand Model Validation

Validation of this type of highly probabilistic and differentiated model requires long-term, high-resolution (i.e. sub-10 minute) household demand data. This type of data for UK households is rare. The HES dataset is relatively large at 251 households, but remains too small to split effectively into separate calibration and validation datasets. The analysis presented in Chapters 1 and 2 also indicates that the use of small-scale datasets for validation is potentially misleading given the potential variation from the nationally representative mean behaviour at this scale.

Additional validation data was therefore taken from two sources; small-scale, high-resolution data from the Richardson et al [69] and the UK REFIT [45] datasets; and large-scale, low-resolution data from substation, district, and national demand analysis. This was required to determine if the model is both nationally representative and matches distinct demand patterns at smaller scales. (For the analysis: the Richardson et al data was limited to the 2008 data and two households with significant evidence of space or water heating were not used; the REFIT data was limited to 14 households (out of 21) for the same reason.)

Four distinct types of validation of the model have been undertaken.

- A performance comparison of the developed event-based cycle allocation method with models using the per-timestep and TUS-activity calibration approaches of existing models.
- Confirmation that both the overall model, and the combined ‘cyclic’ appliance

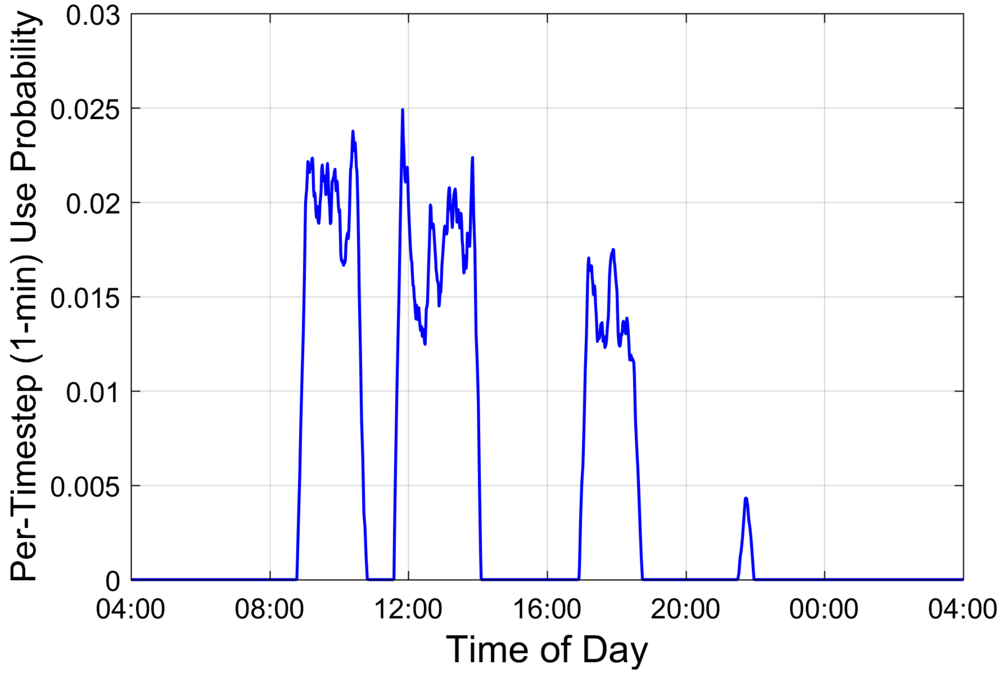


Figure 5.23. 'Timestep' method per-timestep (1-minute) cycle start probability function example for a 6-cycle kettle use day.

modules alone, replicate the average demand profiles from both the HES dataset and validation datasets.

- Confirmation for each specific demand that the developed model replicates the HES dataset calibration data.
- Confirmation that the overall model replicates the average demand profiles for areas with specific characteristics, and replicates national average demand using a representative set of households.

5.14.1 Cycle Method Performance

The aim of the developed discrete-event method was to improve on appliance cycle sequence prediction compared to models that use per-timestep probability calculation methods and the aim of the demand data calibration basis was to improve on existing models that use time-use survey (TUS) activities as a proxy for appliance use. In the following sections, the discrete-event approach ('event') is compared with a per-timestep ('timestep') method using the same calibration data, and the demand data calibrated basis is compared with the time-use calibrated approach used by Richardson et al [69] and others.

The 'timestep' method used for comparison is similar to that used by [147], [69], and [71]. The overall appliance cycle start time distribution is converted from a cumulative to a probability density function (pdf) distribution. As for the 'event' approach, the occupancy model output is used to determine the periods with a non-zero use probability. The distribution is then rebased such that the sum of occupied timestep pdfs is equal to the baseline number of cycles, the baseline identified in the same manner as the 'event' method. For example, a single day pdf distribution for a 6-cycle kettle use day with four occupied periods is shown in Figure 5.23. The methods were compared for both cycle number and timing replication performance.

5.14.1.1 Cycle Number Replication Performance

For the 'event' method, the predicted daily cycle number is identified using the binomial distribution method outlined in 5.6.2.3. For the 'timestep' method, a degree of variance

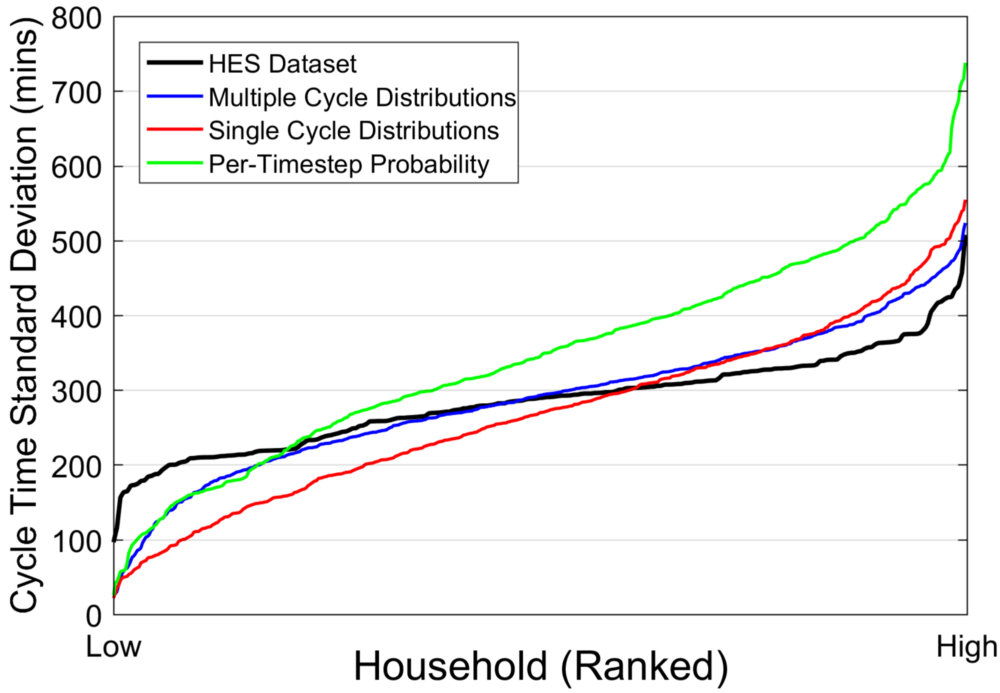


Figure 5.24. Cycle start time standard deviation range for 6-cycle kettle use days for the HES dataset and three different modelling methods. Data for the 'HES Dataset' distribution from [89].

from the average number of cycles to which it has been calibrated is inherent in the process of sequential independent (low) probability calculations.

The kettle data for each of the single-person households in the HES dataset and the equivalent model results for both methods averaged over 20 annual model runs were analysed for the average difference per day between the actual and mean number of cycles divided by the mean number of cycles for each household, and the average for all households determined. For the HES dataset the average household value is 0.387. The equivalent for the 'timestep' model was 0.608 and for the 'event' model was 0.433. Similar results were observed for the other 'Simple' appliances. This suggests that the 'event' method generates daily use number variations that are significantly closer to reality with the 'timestep' method generating excessive variability.

The residual error between the HES data and 'event' model output is a consequence of the lack of linked occupancy and demand data preventing both the occupancy influence on use and natural variation in use for days with identical occupancy to be calibrated more accurately than the relative occupancy factor (see 5.3.3) and binomial method (see 5.6.2.3) currently used.

5.14.1.2 Cycle Start Time Prediction Performance

Both the 'timestep' approach and the developed 'event' approach, if a single overall rather than cycle-specific cycle start time distribution is used, are memoryless with regard to the sequence of cycles within each day. This has the potential to generate unrealistic sequences. To allow the methods to be compared, the standard deviation range (in minutes) for the timing of multiple cycles is compared to the measured data. This is determined by converting each cycle start time to a number of minutes from 4am (e.g. 08.30 = 270) and then determining the standard deviation of each daily start time sequence.

The standard deviation range for 6-cycle kettle days for all single-person HES dataset households is shown in Figure 5.24. There are some outliers but most households are in a range from 200 to 375 minutes with a mean of 285, which is indicative of a significant separation between cycles (for example, cycles at 08.30, 11.52, 14.10, 17.15, 17.42, and 21.58 have a standard deviation of 285).

The equivalent model output results (also shown in Figure 5.24) indicate that the 'event' method with multiple specific cycle distributions more closely approximates the distribution of the actual data, although there remains some discrepancy. Comparison of the two 'event' method options (single and cycle-specific timing distributions) highlights that the cycle-specific approach better captures the broad characteristics of the distribution and the mean value (287 for the use-specific approach and 268 for the single distribution vs. 285 for the HES data). The 'timestep' approach has a significantly higher average standard deviation of 346 and a more linear distribution. This is indicative of results where the modelled daily mean cycle time varies excessively from the mean (c.14.00) because of unrealistic sequencing.

In both cycle number and timing comparisons, the multiple distribution 'event' approach shows a significantly better performance in capturing realistic cycle behaviour. In this case, the residual error is the result of the cycle start time module calibration, which is based on the combined behaviours of multiple households. Further model development is required to manipulate the calibration basis for individual households to account for different behaviours (e.g. typically early or later use) as detailed in Chapter 7.

5.14.1.3 Comparison with a Time-Use Calibrated Model

To confirm the calibration of the cycle start time allocation method (see 5.9), the modelled start time results were converted to the associated cumulative distribution function (cdf) value from the relevant cycle start time probability distribution. The distributions in this case are the unmodified versions prior to accounting for relative occupancy probability (see Figure 5.16).

The overall mean of the modelled *cdf* distribution should be close to 0.5 with a linear variation from 0 to 1. The results are compared with the method utilised by Richardson et al [69], which was based on time-use survey activity probabilities. For the main cooking and washing appliances the mean *cdf* results for both approaches are shown in Table 5.13.

The TUS activity linked model results show significant variation from the 0.5 target, with better performance shown for the demand data calibrated model. Further analysis also demonstrated that the demand data calibrated approach generated a more

Table 5.13
Average cycle start time cumulative probability function results by appliance model calibration method.

Model Calibration	Average Cycle CDF (Target=0.5)							
	Kettle	MW	Toaster	Cooker	Oven	Washing Machine	Dishwasher	Dryer
Time-Use Activity	0.573	0.388	0.648	0.449	0.325	0.531	0.435	0.421
Demand Data	0.521	0.494	0.504	0.499	0.529	0.529	0.515	0.540

consistent overall range of *cdf* results from 0 to 1. This confirms the conclusion from the initial analysis outlined in 3.3.3.1 that broad TUS activities are a weak predictor of specific appliance use. Residual deviation from the 0.5 target for the demand data calibration method is a result of occupancy differences between the HES households and modelled equivalents, and significant variations in the number of cycles per household used for the calibration.

5.14.2 Average Demand Replication

Further validation was undertaken to show that the model converges to the average time-dependent electricity demand profiles and replicates the overall range of demand profiles from the HES dataset, and for the independent 'Richardson' [69] and 'REFIT' [45] datasets (see 2.3), for an equivalent set of households.

For each household in the HES dataset, 500 model runs were generated for the same period. Each household model was set up with the household characteristics (age of respondent, employment status of respondent, appliances owned, etc.) identified by the HES dataset. Similarly, 250 model runs for the 'Richardson' and 'REFIT' datasets were generated in the same manner using the known characteristics. Considering the level of household-level probabilistic factoring used within the model, a significant number of runs were undertaken to generate a representative range of potential results.

Three levels of analysis were performed:

- To confirm that the average model output profile is consistent for each identified HES household type population.
- To confirm that the overall distribution of model results is consistent with the HES and validation dataset distributions.

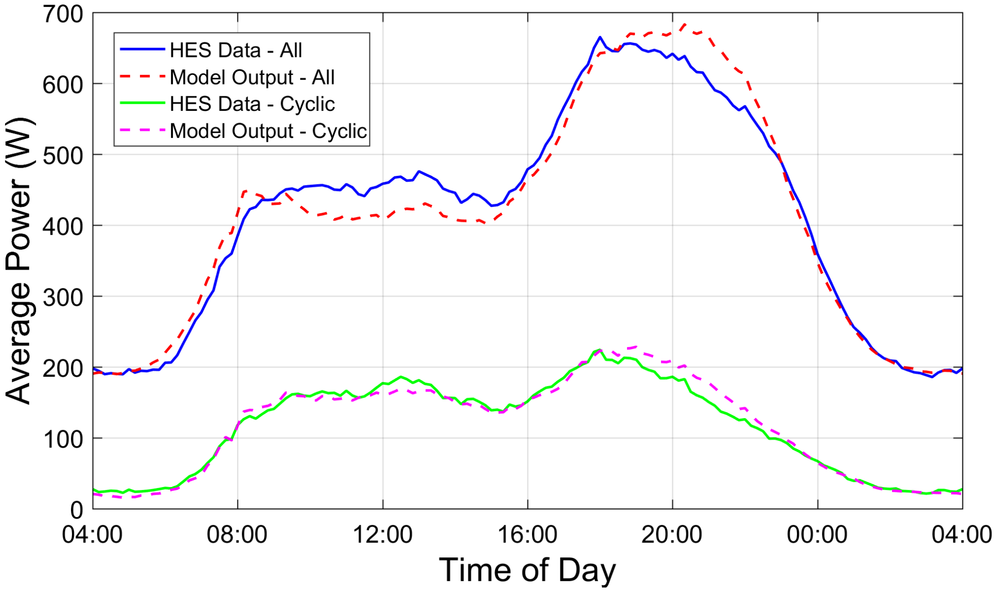


Figure 5.25. Comparison of average power values per 10-minute timestep for the overall HES dataset population and the equivalent modelled population results for the overall electrical demand model and the ‘cyclic’ appliance modules. Data for the ‘HES Data’ distribution from [89].

- To compare individual household measured and modelled average demand profiles over a number of model runs to determine the similarity of the closest match.

5.14.2.1 HES Dataset

The average demand results from the 500 runs both for the overall population and for each of the eight identified household types were analysed to assess if the model converged to the average power consumption and the average time-dependent demand profile from the HES dataset.

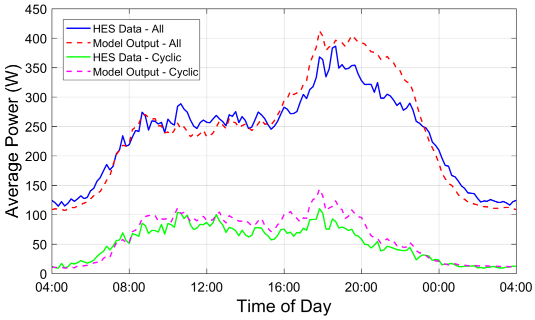
Figure 5.25 shows the results for all households combined and Figure 5.26 the results for each of the eight defined household types. Results for the overall model output and for the combined results of the cyclic appliance modules only are shown separately. In general, there is a good correlation between the model output and the measured calibration data for both assessments suggesting that the underpinning methods are effective. However, there are some discrepancies that need to be further analysed to determine if they are the result of poor calibration, the impact of unrepresentative outliers from the relatively small number of households per household type, or occupancy differences between the actual and modelled populations.

5.14.2.2 Independent Datasets

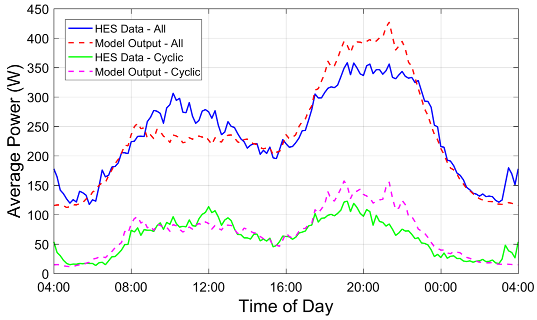
The independent validation datasets are small (20 and 14 households) and therefore the expectation is not that the model results match exactly, particularly given that the household characteristics data is incomplete, but that the results capture the characteristics-driven demand behaviour of each dataset, and no consistent under- or overestimation is observed.

First, the average modelled power values were compared with the equivalent measured values to indicate if the model predicted power baseline is well calibrated. For the 'Richardson' dataset, the averaged measured power is 434W and the modelled equivalent is 463W. For the 'REFIT' dataset, the values were 479W and 491W respectively.

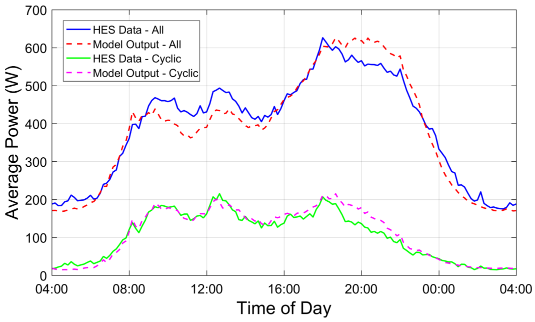
The overestimate of the 'Richardson' equivalent model is primarily the result of significant overestimation for two households and that the ratio of predicted to actual results for the remainder are dispersed equally about parity (see Figure 5.27). The



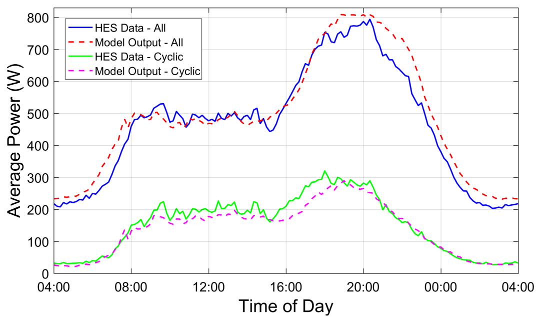
(a) Single-person, retired



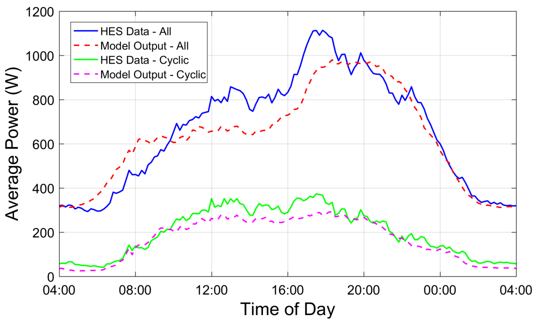
(b) Single-person, working-age



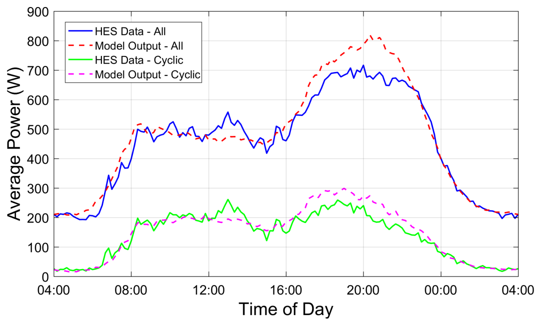
(c) Couple, retired



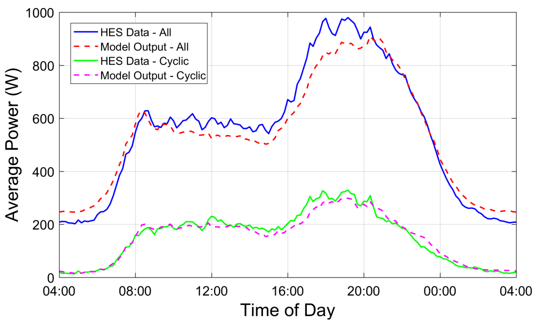
(d) Couple, working-age



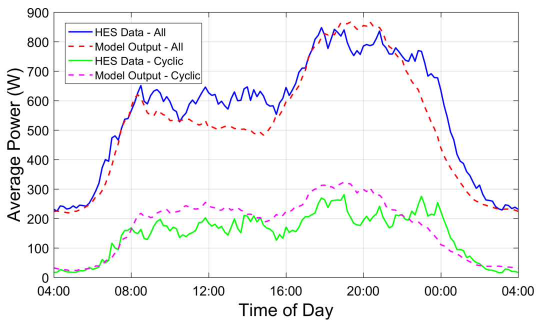
(e) 3-adults+



(f) Family, 3-person



(g) Family, 4-person



(h) Family, 5-person+

Figure 5.26. Comparison of average power values per 10-minute timestep for each HES dataset household type population and the equivalent modelled population results for the overall electrical demand model and the 'cyclic' appliance modules. Data for the 'HES Data' distribution from [89].

household with the worst predicted result has stated ownership of three cold appliances but has an average demand of 333W, which strongly suggest appliances that are owned but unused or a survey error.

There is also a general indication from review of actual and modelled daily average demand values that another potential source of overestimation is that the model currently does not capture the frequency of short absences of 1-3 days and of extended holiday absences seen in the data. Again, the lack of extended duration occupancy data currently limits the ability to more accurately calibrate this element of the model.

Per-timestep analysis of the 'Richardson' and 'REFIT' validation datasets was also undertaken with the results shown in Figure 5.28(a) and (b). Both the average and closest model results from 250 annual-duration runs are shown for comparison. The results show relatively good replication of the measured data overall profile but with some periods of weaker performance.

Significant discrepancies in the early morning and evening periods are associated with distinct demand peaks in individual households. However, the cause of the significantly lower later evening demand and earlier reduction is less clear. The HES model results indicated a tendency to overestimate evening demand but did not display the same timing error. With relatively little household data for each independent dataset, the populations potentially differ significantly from the average predicted by the model. The conclusion, therefore, is that attempting to validate model time-dependent performance with small datasets incorporating limited household data, apart from indicating that general use patterns are replicated, is inconclusive. This is addressed in 5.14.6 where the model output is compared with overall national use data to confirm average timing performance.

5.14.2.3 Profile Replication Performance

As shown, while the model is broadly accurate, there remains some areas of less accurate prediction in certain periods of the day that require further analysis.

For some of the populations analysed, including the independent datasets, the model has a higher mid-evening peak and later evening reduction. This difference is not driven by a single type of demand. There could be several reasons for this. Different occupancy patterns between actual and modelled populations due to specific behaviours

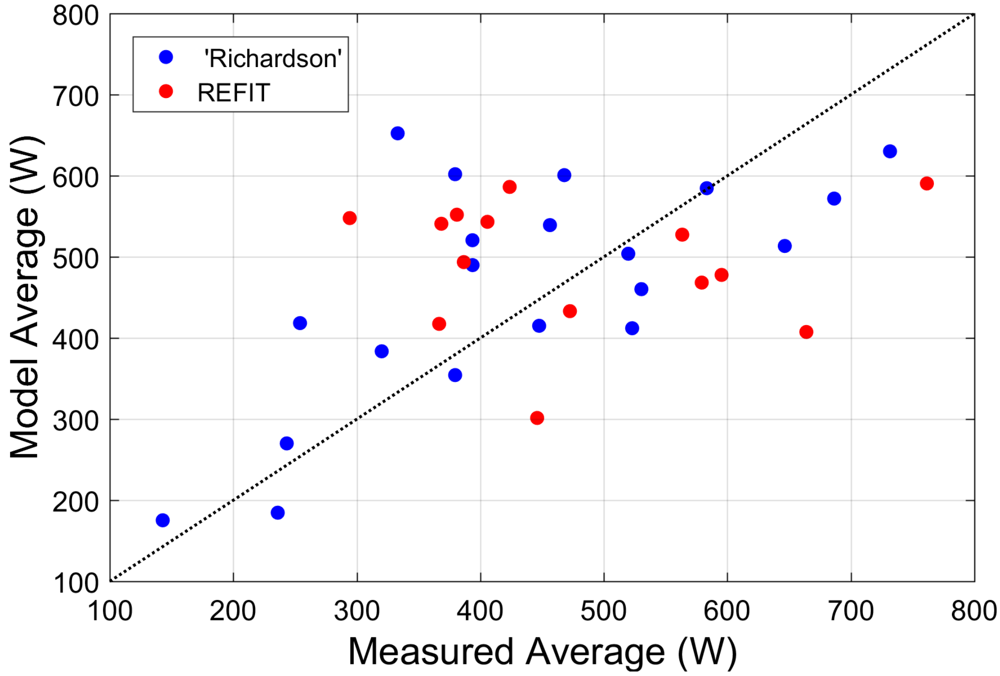
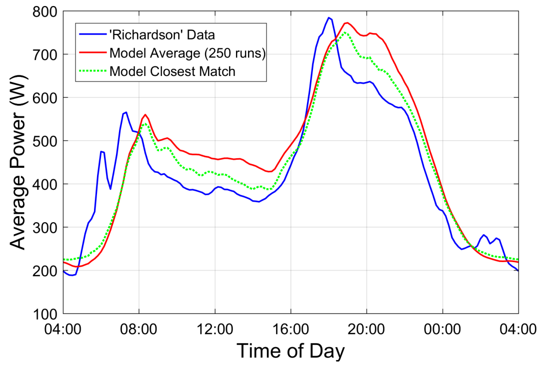
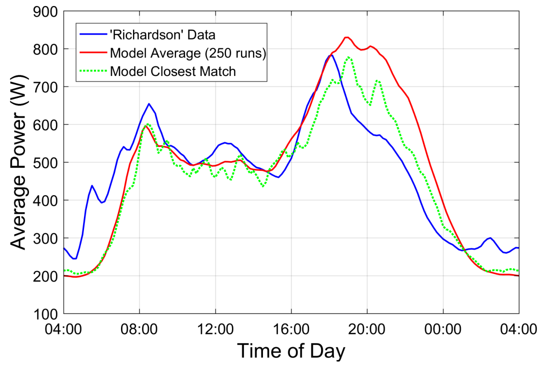


Figure 5.27. Overall distribution of measured and modelled average power values for each independent dataset ('Richardson' and 'REFIT') and equivalent modelled populations. Data for analysis from [69] and [45].



(a) 'Richardson' dataset



(b) 'REFIT' dataset

Figure 5.28. Comparison of average power values per 10-minute timestep for each independent dataset ('Richardson' and 'REFIT') and equivalent modelled populations. Data for 'Data' distributions from [69] and [45]

or differences between predicted and actual household characteristics. The occupancy model also potentially underestimates the variance in sleep transition time, particularly the variation between weekday and weekends, and between co-habiting individuals. Alternatively, there may be a tendency to use lower power cycles and lighting levels in the evening period generally that is not currently captured. For the HES dataset analysis, by contrast, the modelled morning demand increase is significantly more consistent with the actual data suggesting that the basic method is effective but that further calibration is required for specific periods.

For both the family and multi-adult households the model tends to underestimate the difference in demand based on the number of occupants for both overall and cyclic appliance results. Further analysis of the potential impact of number of occupants on typical cycle durations and power requirements, and an increased likelihood of ownership of higher power appliances in larger households is, however, currently limited by data availability.

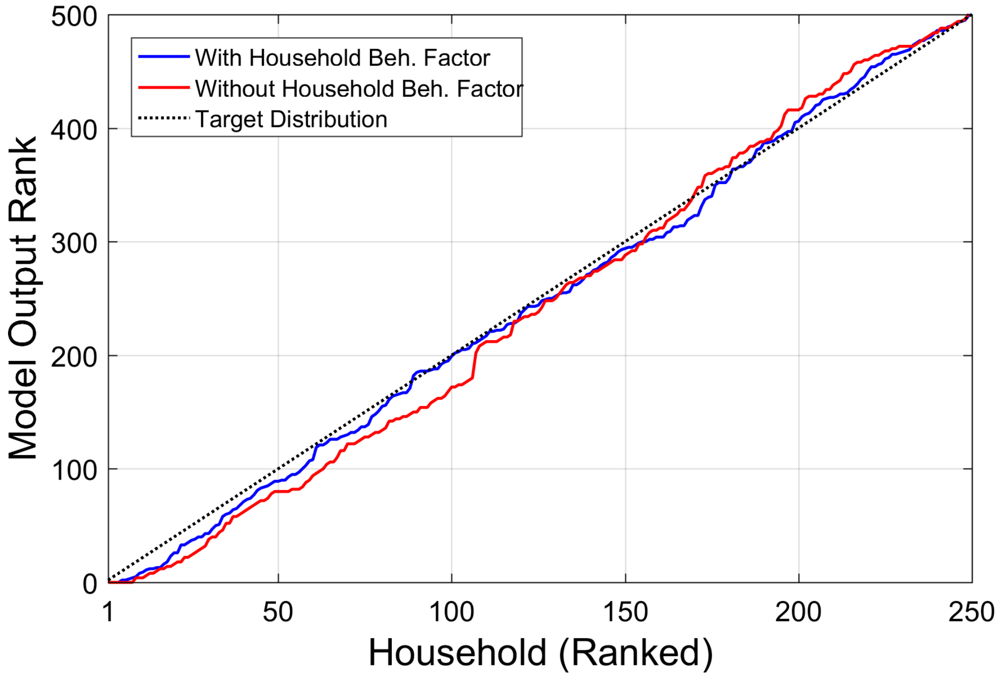
5.14.3 Demand Variance Replication

5.14.3.1 HES Dataset

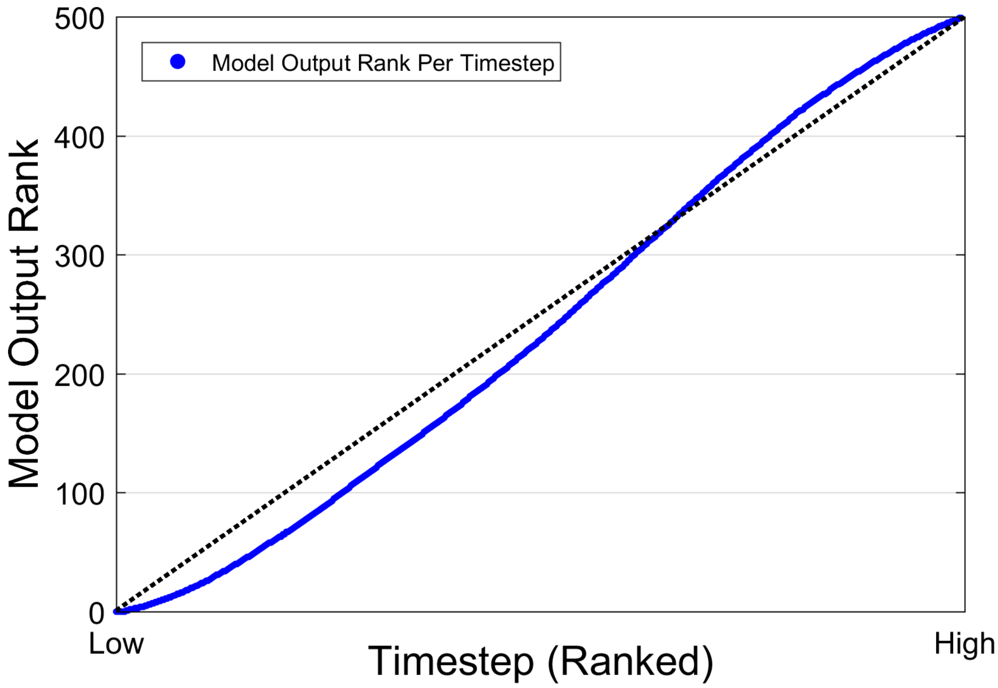
The model results were analysed to confirm that the measured average overall and per-timestep demand for all HES households are evenly distributed relative to the mean value predicted by the model and that the measured values are typically within the range of modelled results. This was done by ranking each HES-equivalent household model output in ascending order (based on 500 runs) and determining the rank of the closest match to the equivalent measured HES household value. To allow the performance of the household behaviour factor (see 5.3.1) to be determined, a further set of 500 runs were generated with the factor set to 1.

For overall average power, Figure 5.29(a) shows a generally linear distribution of model output ranking per actual data point for models with and without household behaviour factoring. This indicates that a significant degree of demand variance between households can be accounted for by differences in appliance ownership, the type of appliances, and appliance-specific behaviours.

However, the model with household-level factoring shows an improved performance,



(a) Per-household



(b) Per-timestep

Figure 5.29. Model output rank (from 500 runs) for HES dataset per-household and per-timestep measured average power.

particularly at the lower end of the demand range, with a 51% reduction in the error to the target distribution. The range of household behaviour factors (*EHBFs*) generated for a single HES dataset equivalent model run is shown in Figure 5.30. The overall influence is to reduce the median demand value (the median *EHBF* is less than 1) and to generate a significant increase in use for a small number of households at the upper end of the range. The impact of this is shown in Figure 5.29(a) where there is an improvement for the factored model in reducing both the overestimate in the low-to-mid range (rank below the target line) and the underestimate at the upper mid-range (rank above the target line).

The results suggest that the 37% random behaviour factor applied as per the analysis of Gill et al [81] is broadly representative of real behaviours in combination with the other generated factors for income, occupancy and appliance use. Similar results are observed when the results for each household type were reviewed, and also when the ranking distribution based on 10-minute timestep results was analysed (see Figure 5.29(b)).

The residual error for the factored method indicates that the household behaviour factor as currently modelled is overly simplistic. Simple statistical relationships have been assumed at this stage due to lack of detailed data, which do not fully capture the range of demand variation that results from income-, occupancy-, and randomly-driven behaviours. The results, however, indicate that household-level behaviour factoring is required to capture the overall demand range, and that reliance on household characteristics (type, size, age profile), appliance ownership, and individual appliance-level variation to account for this is insufficient. Further work is required to better calibrate this input.

The inclusion of the household behaviour factor also reduces the number of HES household demands that are outwith the model output, suggesting that the additional factors are also required to capture more extreme behaviours. Measured data for 4 (1.6%) of the HES households lies outside the range of modelled results for the factored model and 8 (3.2%) for the unfactored. For the factored model, all outliers are at the low end of the range, and, for the unfactored, one outlier is at the upper end.

A significant proportion of the residual error and outlier results are at the lower end of the demand range. This suggests that there may be other drivers of the weaker

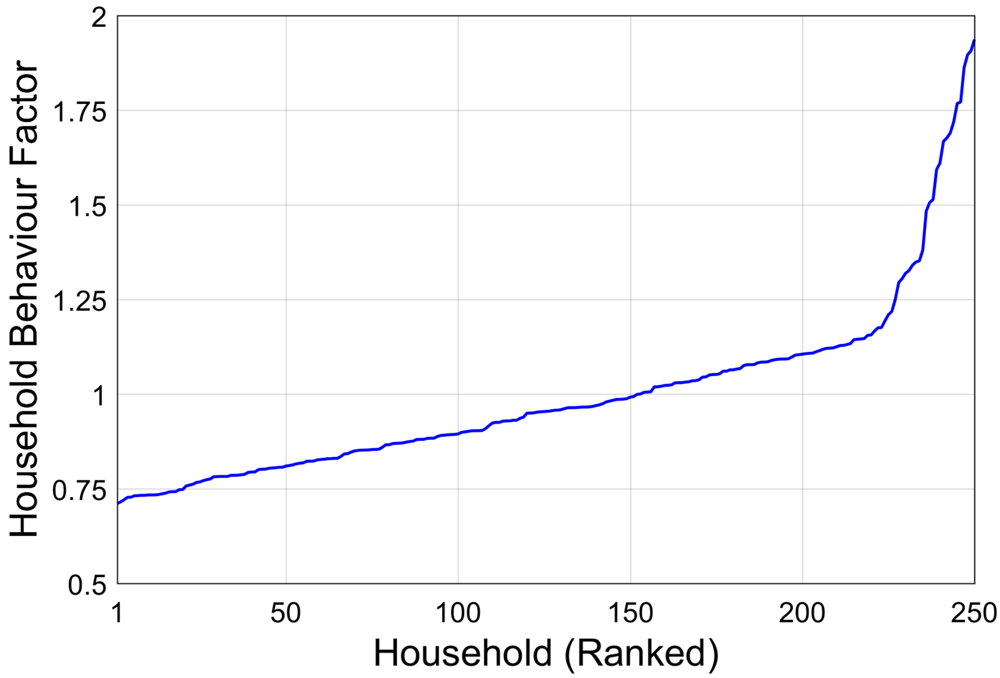


Figure 5.30. Household behaviour factor (*EHBF*) distribution for a single HES-equivalent model run.

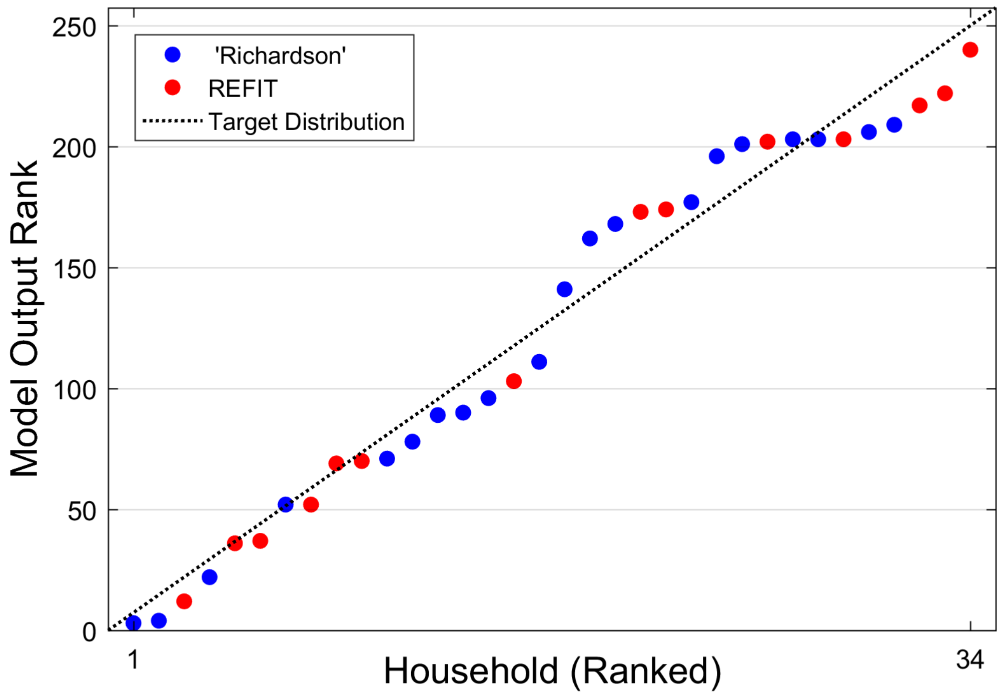


Figure 5.31. Model output rank (from 250 runs) for 'Richardson' and 'REFIT' datasets per-household measured average power.

performance for these households, including; modelled relationships between appliance power ratings and relative use not currently capturing very low demand households; underprediction of households with low relative use for all appliances; and overestimation of minimum occupancy potential as a result of the group-calibrated occupancy model and limited extended absence data. The results, however, suggest that the variety of probabilistic factors applied at different levels of the model broadly captures the typical distribution driven by household characteristic and behavioural factors, without significant under- or overestimation.

5.14.3.2 Independent Datasets

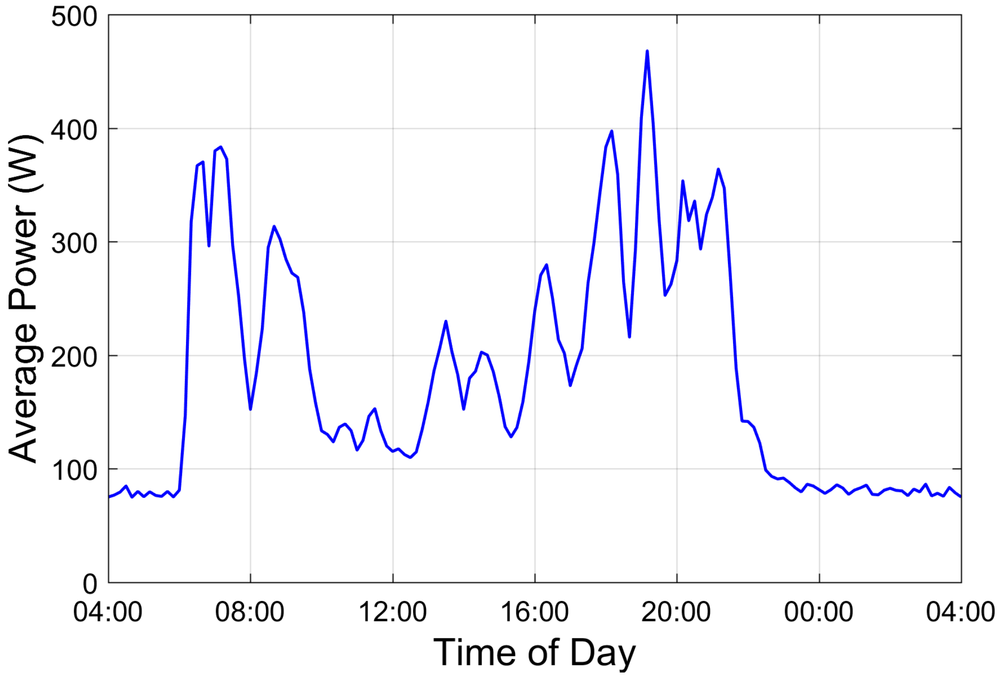
To confirm that the overall range of results generated by the model is consistent with the measured data, similar analysis to that shown in Figure 5.29 was undertaken for both independent datasets. The analysis is less meaningful as the number of analysed households is smaller, however, the distribution of model rankings for the measured data is broadly consistent in comparison with the target distribution (see Figure 5.31) and no measured household average demand is outside the model predicted range.

5.14.4 Individual Household Behaviour Replication

Having confirmed that the model captures average household type behaviours and a representative distribution of average demands, further analysis is required to determine if the model is able to capture household-specific per-timestep demand profiles. The following section outlines a method to determine the similarity between the model output and measured data.

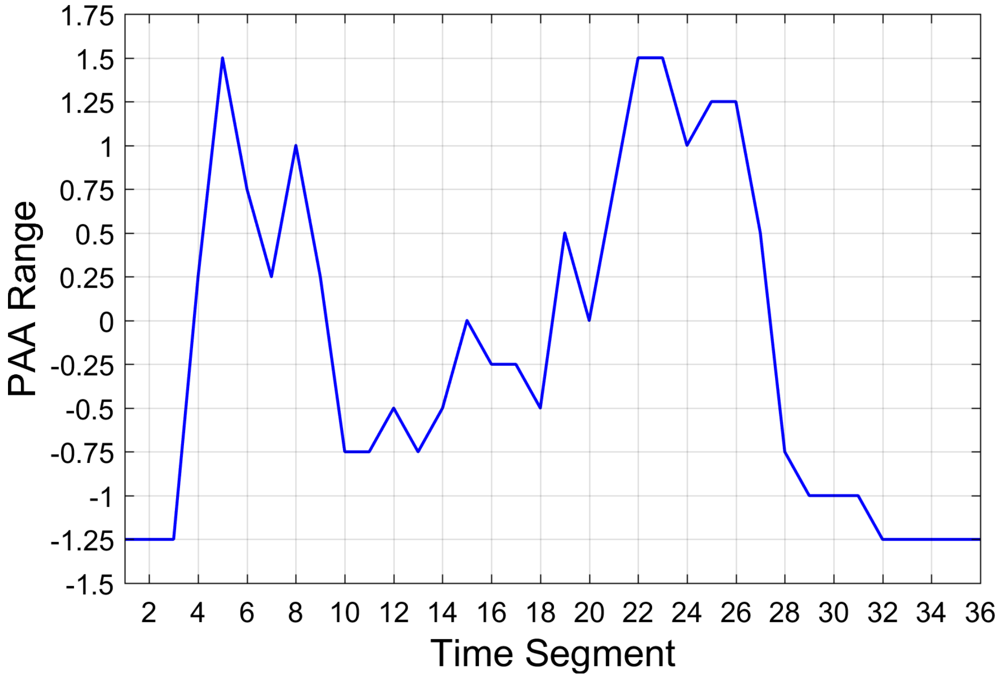
5.14.4.1 Similarity Analysis Method Development

The most commonly used numerical string similarity measure is the Euclidean distance, which is determined by the square root of the sum of the square of the difference ('distance') per string element. Demand profiles at a 10-minute resolution based on one month of either measured or modelled data tend to be erratic and overly influenced by individual high power cycles. Only at longer timescales do the demand profiles become smoother and more consistent. Euclidean distance analysis when used for



(a) Typical 10-minute timestep average household average electrical demand profile

⇓



(b) PAA range 40-minute time segment equivalent profile

Figure 5.32. Average demand profile with 10-minute timesteps to PAA range with 40-minute time segments conversion example.

erratic profiles generates results that do not properly reflect overall similarity and are too dependent on individual per timestep differences. A method is therefore required to reduce the time-series to a smoother profile that remains consistent to the overall profile and household behaviour.

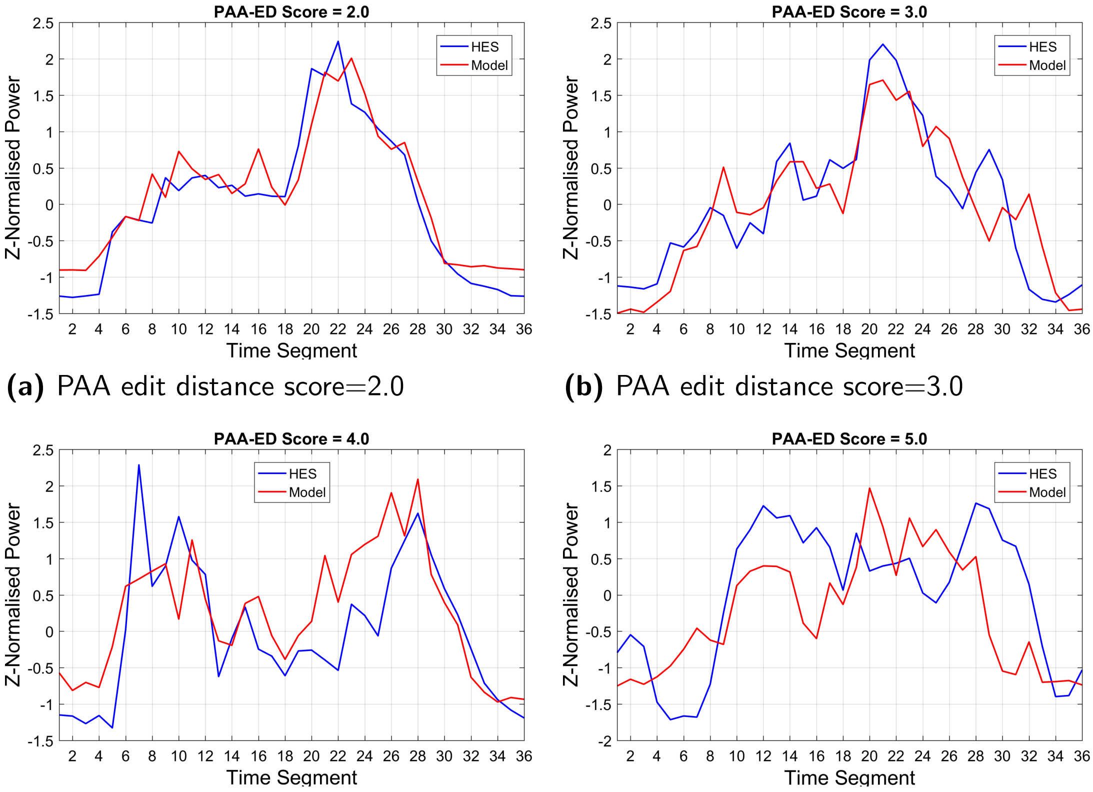
Piecewise Aggregate Approximation (PAA) [171] is a data mining method that allows a time series to be condensed and simplified to its basic structure, reducing the influence of individual data points. The simplified time-series' generated retain the ability to be compared using Euclidean distance. To allow results for different households to be compared on an equal basis, the data must be normalised and analysed based on variance to the time-series mean. For the PAA approach, the time series is z-normalised (difference between actual and mean result divided by the standard deviation). The number of segments to be analysed is then reduced by taking the mean of the 10-minute resolution z-normalised values for each new larger segment.

Further investigation was required to determine the most effective segment size for analysis. Analysis of 20, 30, 40, and 60-minute segment sizes determined that 20 and 30-minute sizes retained a significant degree of the erratic nature of the 10-minute time-series while at 60-minutes too much of the detail is lost. 40-minute segments (36 per 24-hour profile) were therefore determined to be the best compromise between retaining detail and removing the poor performance associated with highly erratic profiles.

Figure 5.32 shows an example conversion from a 10-minute timestep profile to the equivalent 40-minute segment profile based on the defined PAA ranges. It can be seen that the basic demand pattern is retained while reducing the significance of individual peaks and troughs.

To determine the model effectiveness requires two different types of analysis; an 'overall' and a 'timing' comparison. The 'overall' comparison determines the similarity between profiles without any further rescaling. One of the profiles is used to set the mean and standard deviation for the z-normalisation factors used for both measured and modelled profiles. This is a measure of the similarity of both power level and timing.

To remove the influence of variable baseline power levels from the analysis to allow the performance of the cycle start time module to be assessed, a separate 'timing' comparison was also undertaken. This is achieved by allowing each set of profiles to be



(a) PAA edit distance score=2.0 **(b)** PAA edit distance score=3.0

(c) PAA edit distance score=4.0 **(d)** PAA edit distance score=5.0

Figure 5.33. Example HES dataset and model output PAA range profile comparisons for different PAA edit distance (PAA-ED) scores. Data for the 'HES' distributions from [89].

z-normalised individually using factors appropriate to each profile in order that they are rescaled to the same average power basis. In this case, only relative demand timing is assessed.

Z-normalisation was chosen as the overall normalising method as it is an appropriate basis for the 'timing' comparison. For this analysis, any proportional difference is variable and highly time-dependent, which is mitigated by incorporation of the standard deviation in the normalisation. For the 'overall' comparison any normalising method would have been acceptable but using z-normalisation allowed all results to be compared on an equal basis.

5.14.4.2 Similarity Analysis Assessment Basis

For a 36 time-segment Euclidean Distance comparison of two profiles using the PAA method (hereafter known as PAA-ED), a score of 1.5 is equivalent to an average z-normalised PAA value difference of 0.25 per time segment, 3 is equivalent to 0.5 etc. Whether two profiles can be considered similar is, however, a subjective judgement that is best determined from direct visual comparison.

Comparison of results shows that a PAA-ED score of 2.5 or less represents a close overall correlation between the two profiles. A result between 2.5 and 3.5 retains broad similarity with evidence of either some time-shifted offset or failure to simulate specific extreme values. Between 3.5 and 4.5, the basic shape of the profile is typically discernible but the model has not captured specific details. Above 4.5 the model has not captured a significant proportion of the actual demand detail. Results have been grouped into these 4 ranges (<2.5, 2.5-3.5, 3.5-4.5, 4.5+), and the ranges are titled 'High Similarity', 'Good Similarity', 'Some Similarity', and 'Low Similarity' respectively. Example HES dataset and model output PAA range profiles for various PAA-ED scores are shown in Figure 5.33. For clarity, the profiles are included based on the 'overall' comparison PAA-ED score rather than the 'timing' comparison score.

5.14.4.3 Similarity Analysis Results

Using the PAA-ED method outlined, measured profiles for the HES and independent datasets were compared with the model output for equivalent populations. The results

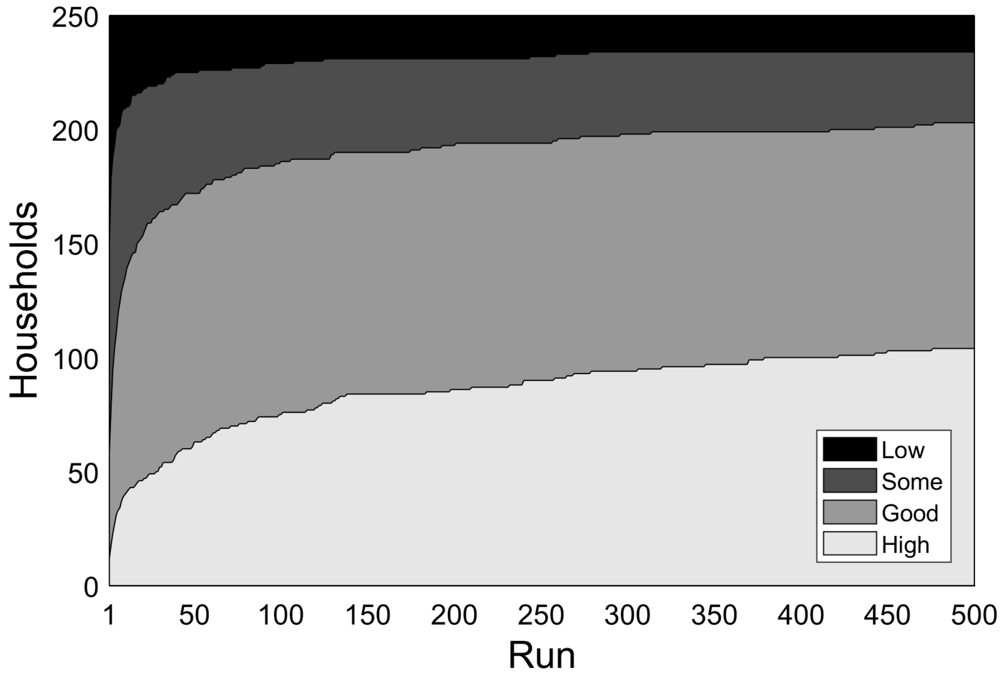


Figure 5.34. Model closest cumulative match similarity analysis range results for the 250 HES household equivalent model over 500 model runs.

have been grouped into the four defined similarity ranges.

HES Dataset

Detailed analysis of the data presented in Figure 5.34 allows the lowest cumulative PAA-ED score for each household after each run to be graded. The most significant improvement occurs during the first 20 runs and with limited further improvement seen beyond 100-150 runs. The final results after 500 runs for both ‘timing’ and ‘overall’ value analysis are shown in Table 5.14.

Table 5.14
Model closest cumulative match similarity analysis range results for the 250 HES household equivalent model after 500 model runs.

Similarity	'High'	'Good'	'Some'	'Low'
‘Timing’	105 (42%)	98 (39%)	31 (12%)	17 (7%)
‘Overall’	49 (20%)	118 (47%)	53 (21%)	31 (12%)

The ‘timing’ results show that 81% of the results are rated 'Good Similarity' or better after 500 runs. The results for the ‘overall’ analysis are lower, as expected, but 67% are within the 'Good Similarity' or better range after 500 runs. The results suggest that the model is able to capture a significant degree of the highly variable nature of demand while producing outputs that are consistent with individual behaviours.

However, there remains a small number of households whose specific demand patterns are not captured by the current model. Further consideration is therefore required as to how the model is calibrated and for potential sources of inaccuracy, with focus on the identified areas where the model is calibrated from composite behaviours as discussed below. Analysis of individual household demand profiles also indicates that there are a small proportion (<10%) that have highly distinct behaviours that are probably outwith the scope of probabilistic model calibrated from currently available data.

To confirm that the overall set of results generated are representative of the actual range of potential demands, and that the closest matches observed are not simply the result of randomly generated profiles, the overall distribution of model results was reviewed.

The overall distribution of PAA-ED scores for the 500 HES-equivalent model runs is compared to the distribution of PAA-ED scores for the Euclidean Distance comparison

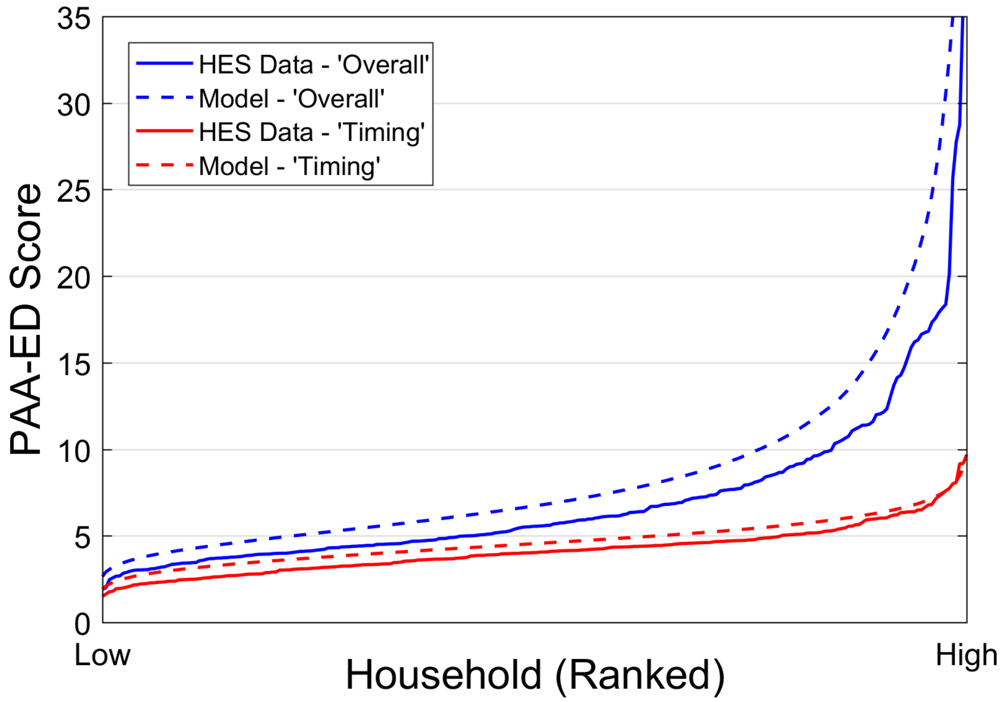


Figure 5.35. Overall PAA-ED score distributions for 500 HES-equivalent model runs and the equivalent scores for the measured HES data compared to the average model output for each HES-equivalent household. Data for the 'HES Data' distributions from [89].

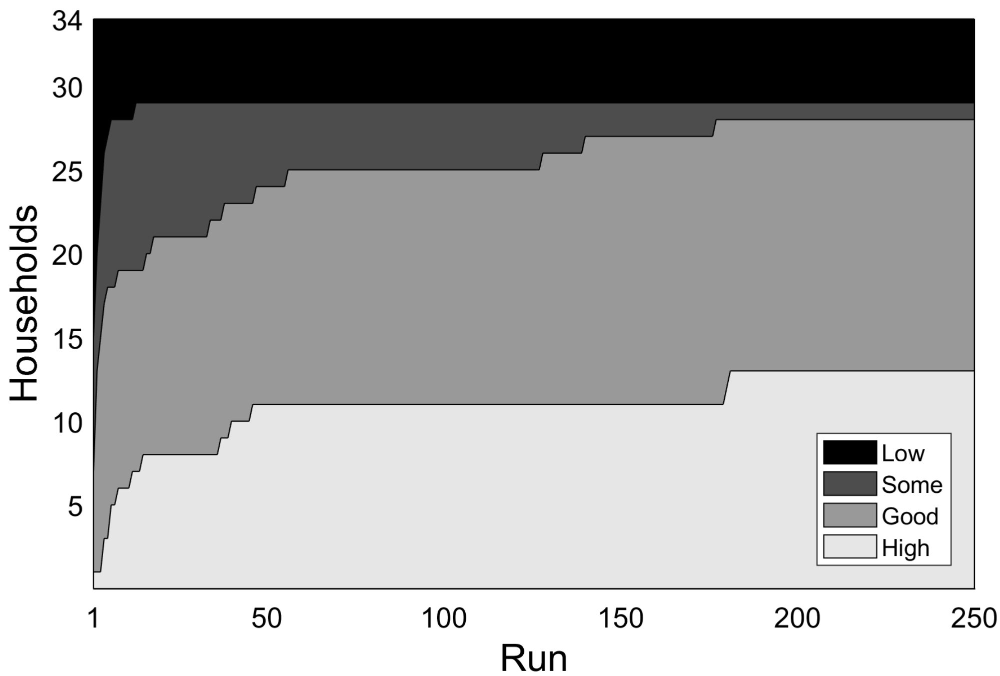


Figure 5.36. Model closest cumulative match similarity analysis range results for the 'Richardson' and 'REFIT' household equivalent models over 250 runs.

between the measured HES data time-dependent profiles and the average profiles generated by the HES-household equivalent model (these are assumed to be equivalent to the expected profile for a household if all behavioural factors are average). If the model is creating matches probabilistically rather than randomly, the shape of the distributions should be similar, with an allowable offset due to model inaccuracy. Figure 5.35 shows that the distributions for both ‘timing’ and ‘overall’ results are consistent with the measured data distributions and that the model is not generating results randomly.

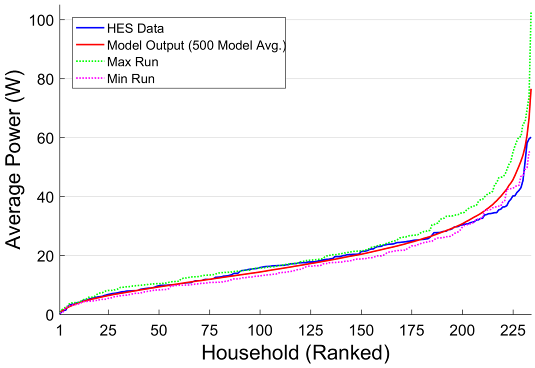
Independent Datasets

The combined results for the independent ‘Richardson’ and REFIT datasets are shown in Figure 5.36 and highlight that the model is able to capture the specific behaviours in each dataset with all but 6 of the 34 household models achieving at least a ’Good’ similarity rating for the closest match. The model performance for this measure is similar to that shown for the HES dataset for an equivalent number of runs which demonstrates that the model is able to capture demand behaviours beyond the calibration dataset. The households that are not captured closely have distinct non-typical patterns (e.g. significant peaks, high night use).

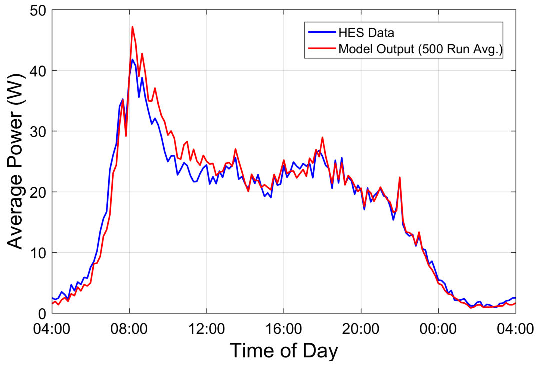
5.14.5 Specific Demand Analysis

A key aim of the developed model is to predict overall demand using a highly detailed, bottom-up approach to allow assessment of demand uncertainty. It is therefore necessary to review the output at the level of specific modelled demands to determine if this has been achieved, and to assess if the model basis is suitable for analysis requiring this resolution, such as the demand management (shifting) potential of individual appliances.

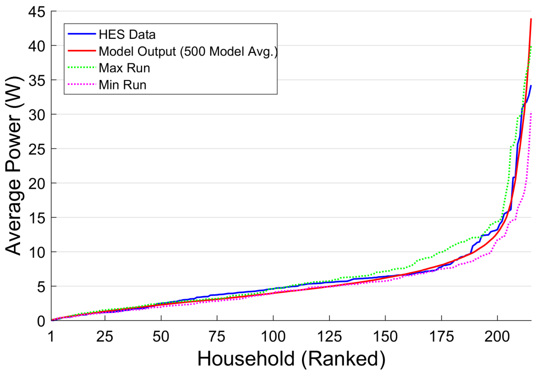
For appliance use, four main comparisons are required. The range of overall demand per household, the temporal demand, the overall distribution of cycle start times, and the distribution of cycle start times per household. This analysis has been undertaken both for the overall population and each household type.



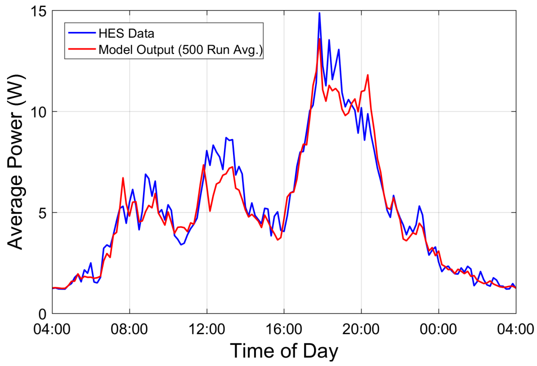
(a) Kettle - Average Demand



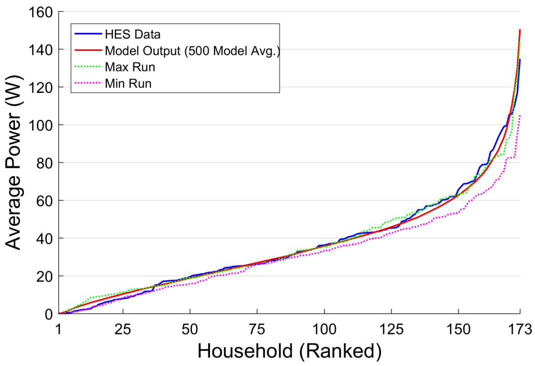
(b) Kettle - Temporal Demand



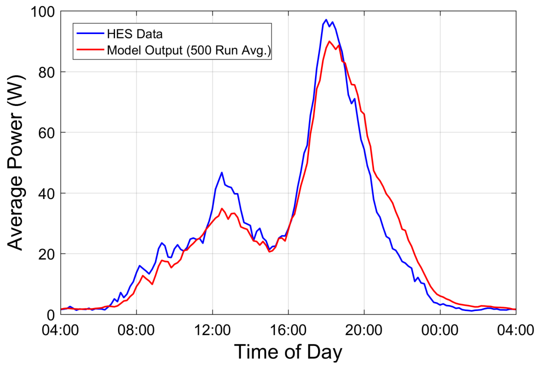
(c) Microwave - Average Demand



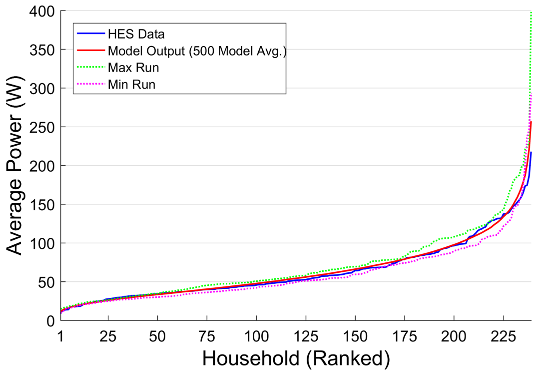
(d) Microwave - Temporal Demand



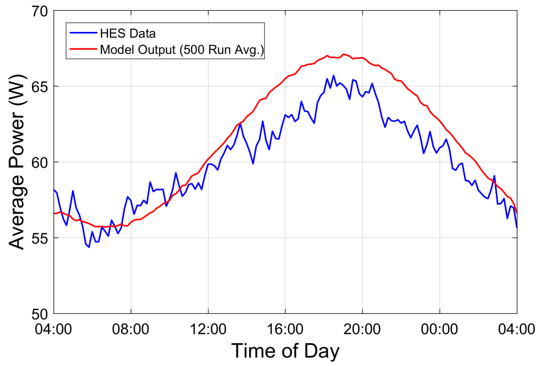
(e) Cooker/Oven - Average Demand



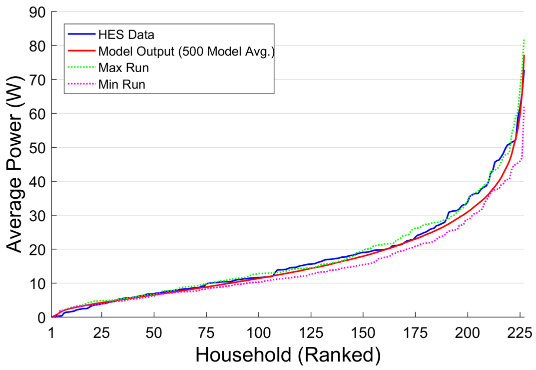
(f) Cooker/Oven - Temporal Demand



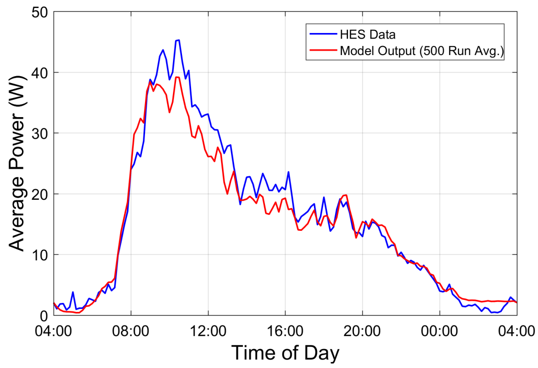
(g) Cold Appliances - Average Demand



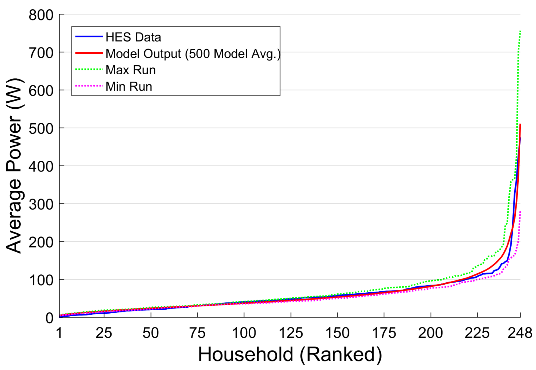
(h) Cold Appliances - Temporal Demand



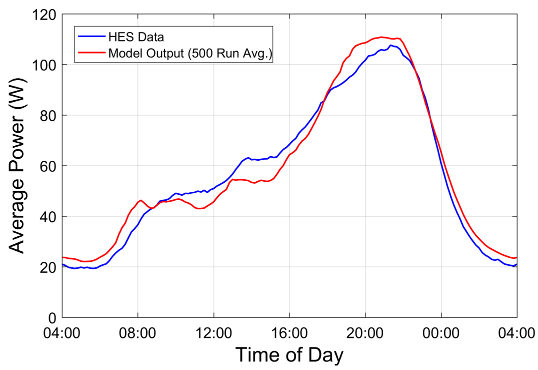
(i) Washing Machine - Average Demand



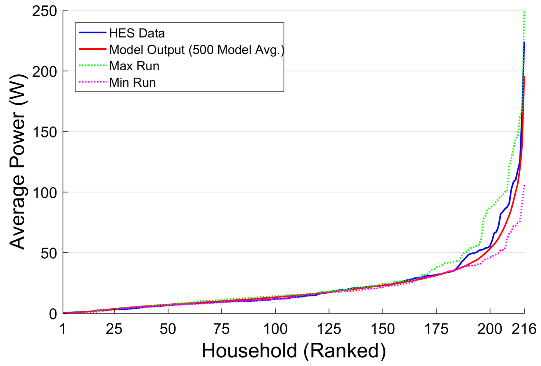
(j) Washing Machine - Temporal Demand



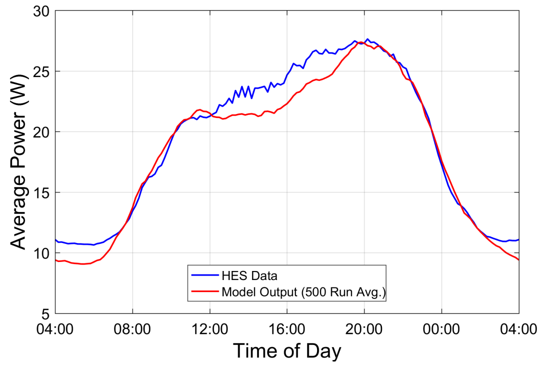
(k) Audio-Visual - Average Demand



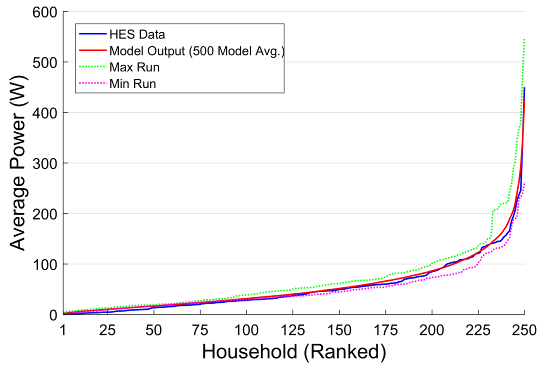
(l) Audio-Visual - Temporal Demand



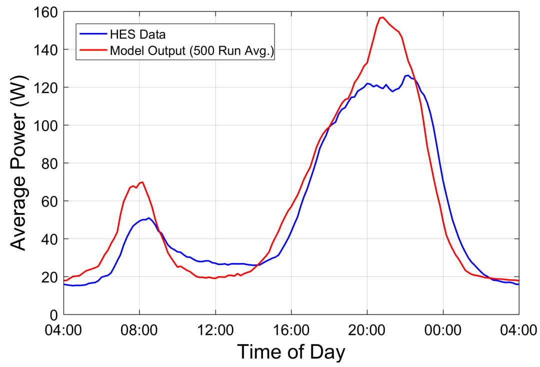
(m) IT - Average Demand



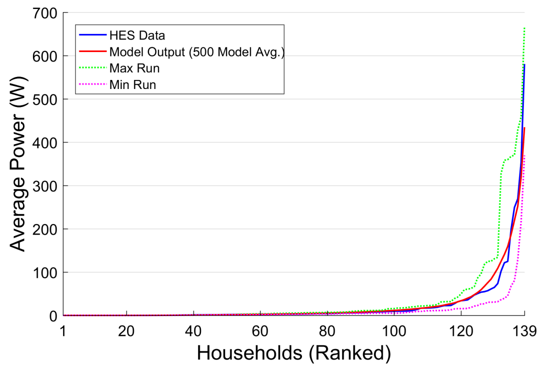
(n) IT - Temporal Demand



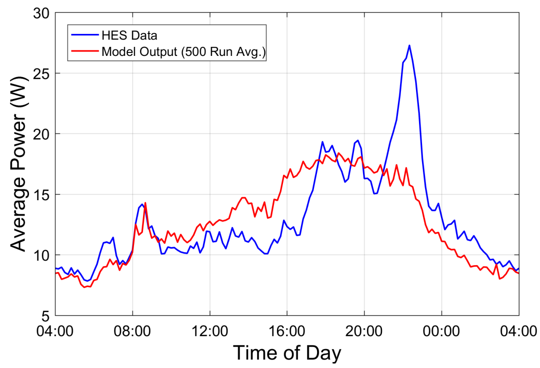
(o) Lighting - Average Demand



(p) Lighting - Temporal Demand



(q) Miscellaneous - Average Demand



(r) Miscellaneous - Temporal Demand

Figure 5.37. Specific demand module average per-household and time-dependent overall average demand profiles in comparison with equivalent HES data. Data for the 'HES Data' distributions from [89].

5.14.5.1 Demand Range

Based on the average results of 500 model runs for the HES-equivalent model, the comparison of the overall predicted demand distributions compared to the HES dataset for all significant cyclic appliances is shown in the left column of Figure 5.37. The model results track the overall distribution of individual household results with good accuracy, particularly for the lower 90% of users. The increased variation at the high end of the range is a direct result of the calibration data characteristics (i.e. more dispersed behaviours), however, further data is required to determine if the output for these households is realistic. Individual model runs show some variation to the mean result as would be expected for a probabilistic model that tracks real behavioural variations.

The primary purpose of this analysis is to assess the effectiveness of the combination of overall and appliance-level factoring. The overall assessment of the results suggests that there is a slight potential to overestimate demand at the upper end of the range. A weighting of one for each behavioural factor identified in 5.3 is likely to be overly simplistic and has the potential to generate overall multipliers that are higher and lower than reality. However, without demand data that includes detailed occupancy and income data, further work to identify additional overall or appliance-specific factors would be ineffective. The results suggest that the approach used is effective with respect to predicting uncertainty in overall demand but results at the extremes of appliance-level distributions need to be treated with care.

5.14.5.2 Temporal Demand

From the same 500 model runs, the right column of Figure 5.37 shows the time-dependent demand profiles in comparison with the HES data. The purpose of this analysis is to determine the overall effectiveness of the cycle start time and duration modules.

The replication of appliances modelled with the 'cyclic' modules show good accuracy. The most significant discrepancies are related to appliances with significant time-of-day influences on cycle behaviour and more extreme variations per household in the calibration dataset, this applies in particular to cookers and ovens. Further analysis with independent data is required to determine if the discrepancies are the re-

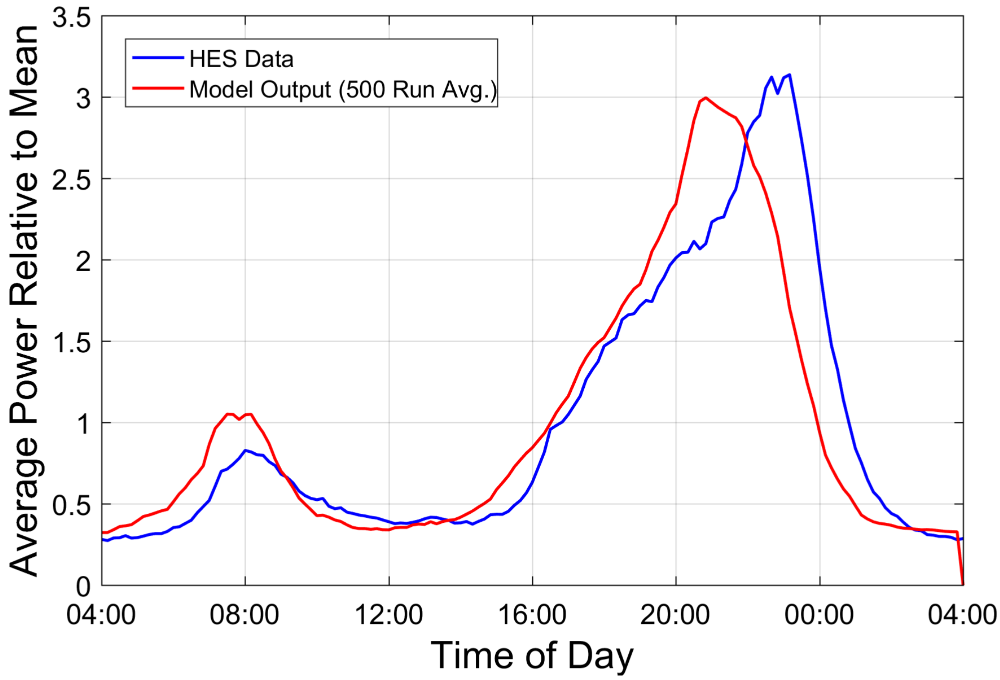


Figure 5.38. Per-timestep lighting power relative to household mean, overall average for HES data and lighting module output. Data for the 'HES Data' distribution from [89].

sult of over-simplistic calibration or if the model is realistically reducing the distorting influence of the extreme behaviours.

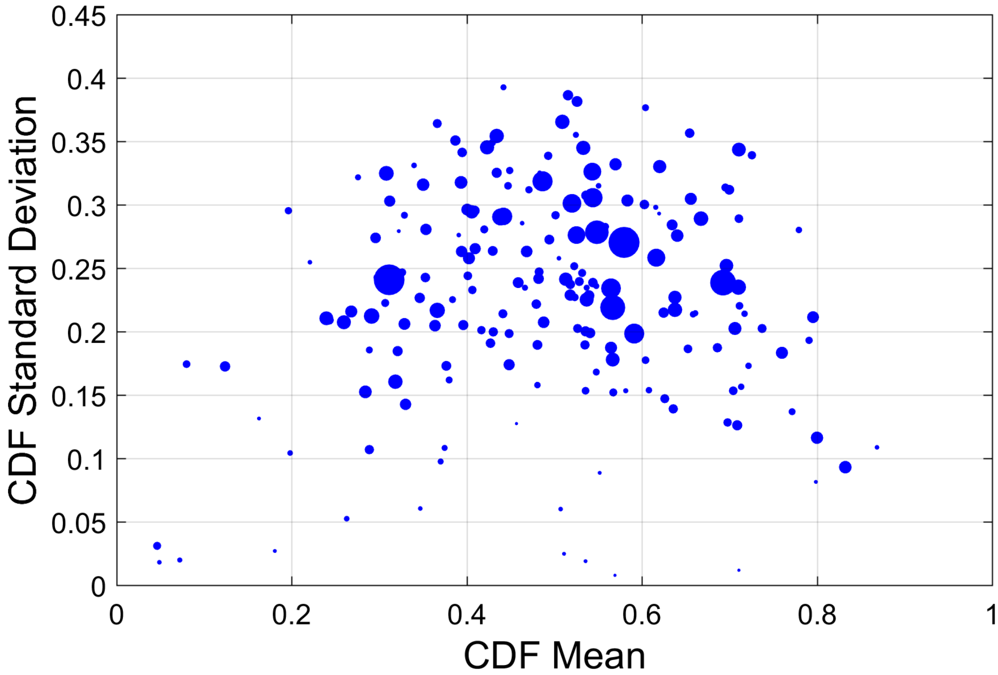
The TUS-activity calibrated 'AV' module shows good replication of the HES data profile with a slightly earlier prediction of the key use transition times. Further analysis with the updated UK 2015 TUS dataset will be required to determine if TV watching habits have changed between 2000 (TUS) and 2011 (HES), or if the HES dataset population has later TV use behaviour than average.

The lighting module tracks the overall profiles with reasonable accuracy but with overestimation in the waking and mid-evening periods. However, the HES data is characterised by a significant range of lighting power levels and is significantly distorted by demand from a small number of households. When relative lighting power to the household mean is considered to partially remove this distortion, the resultant comparison (see Figure 5.38) shows a similar single-peaked evening profile for the HES data, although the timing is not closely replicated. The conclusion drawn is that the lighting module is currently adequate for an overall demand model but needs further work to account for specific time-dependent behaviours, room sharing, and use of lighting per room.

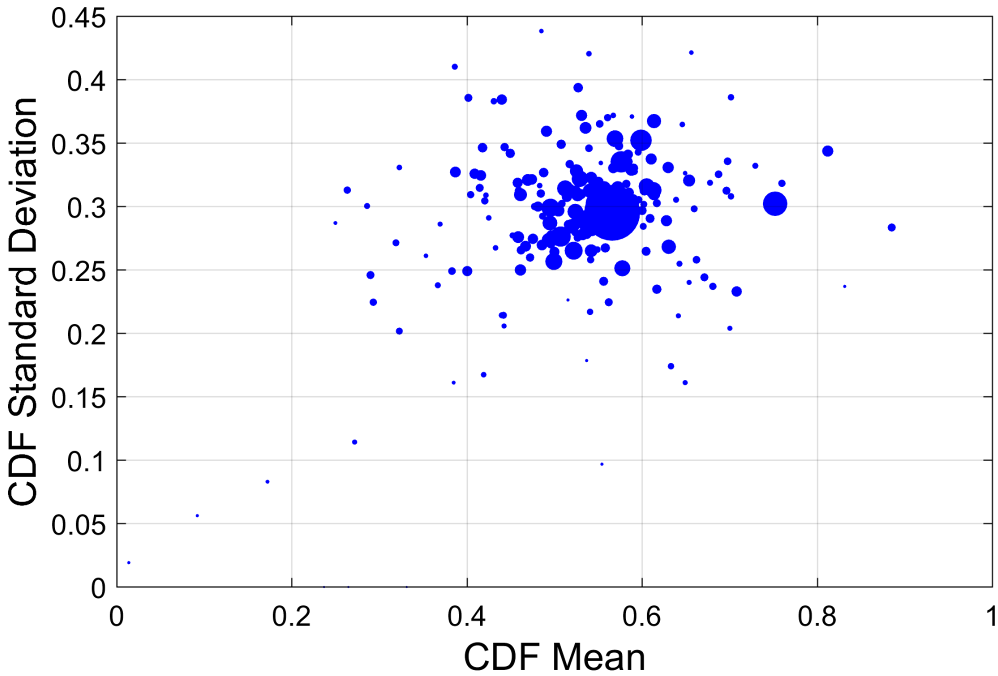
5.14.5.3 Household-Specific Cycle Timing

The average cycle start time cumulative distribution function (cdf) analysis in 5.14.1 confirmed that the developed method was better than existing alternative approaches in terms of the average and overall range of cycle start times. A review of the results for individual households shows that for each appliance the average cycle start time and the range of times (as measured by the standard deviation) per household converges to the mean behaviour, and the degree of convergence increases with an increase in the number of cycles simulated. Figure 5.39 shows a comparison between the HES dataset and equivalent modelled household cycle start time *cdf* mean and standard deviation distributions for washing machine use. The results indicate that the cycle start time identification method calibration based on the composite data from all HES households causes significant and unrealistic convergence of modelled household use behaviours.

Additional analysis of this type for other appliances and specific demands and a proposed method to reduce the unrealistic convergence is detailed in Chapter 7.



(a) Washing Machine - HES Households



(b) Washing Machine - Model

Figure 5.39. Cycle start time cumulative probability function mean and standard deviation for washing machine use per household. Data for the 'HES Households' distribution from [89].

5.14.6 Area Type Validation

To confirm that the model tracks average differences between areas with different socio-economic characteristics, three sets of models were generated for populations representative of both the national average household and social-rented only populations, and also for an area with the lowest deprivation decile (see 2.4.1.1), and the model results compared with available demand data.

In addition, a set of model runs was completed for an identical set of house sizes for each deprivation decile differentiated only by the average tenure per decile. This was then compared to the average demand for the twenty closest match areas to the model population characteristics based on Lower Super Output Area (LSOA) demand data. England is split into 32,844 LSOA areas comprising typically 600 to 1000 households.

Table 5.15
UK national average housing size, type, and tenure. Data from [93] and [96].

	1-Bed House	2-Bed House	3-Bed House	4-Bed House	1-Bed Flat	2-Bed Flat	3-Bed Flat
Private	1	14	28	14	3	7	2
Social-Rent	2	5	5	1	4	6	1
Private-Rent	0	1	1	0	2	2	1

5.14.6.1 National Average

Analysis of Scottish [93] and English Housing Survey [96] data determined that the data shown in Table 5.15 represented the average household mix for the UK in terms of size and tenure, rounded to the nearest percent. Elexon publish electricity demand data that is representative of average UK use [172]. For domestic demand, two sets of data are provided, one for standard tariff customers and the other for customers on the Economy 7 tariff with time-dependent charging, typically used by households with electric storage heating. In this case the standard tariff data is used for comparison.

The Elexon data is provided for five time periods; spring, summer, high summer, autumn, and winter. As the model does not currently include electric secondary heating demand, the model was compared with the combined ‘summer’ and ’high summer’ data representing the period from 14th May to 3rd September only. 100 runs for models based on the national average household tenure and type, with other household factors

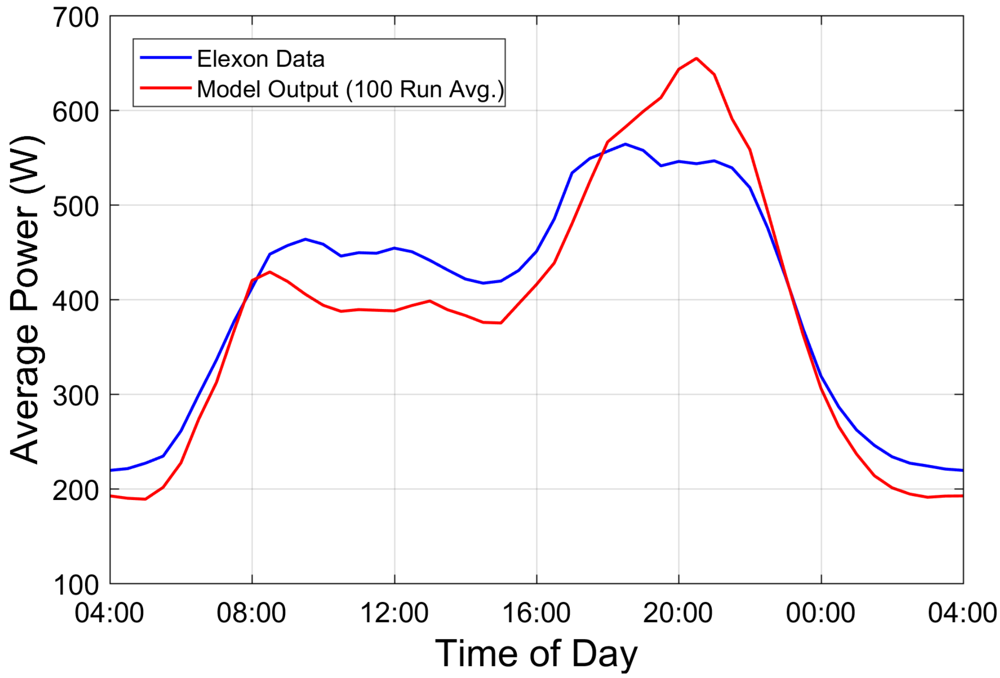


Figure 5.40. National average electricity demand data from Elexon and equivalent model output for combined 'summer' and 'high summer' periods. Data for the 'Elexon Data' distribution from [172].

selected probabilistically for each run, were compared to the Elexon data as shown in Figure 5.40.

The model average demand was 389W compared to 405W for the Elexon data. Considering the potential for some electric space and water heating that is not yet captured by the model, this is a good match. Time-specific demand also shows good accuracy for transition timing but with a specific discrepancy in the later evening period, the cause of which is not clear. The transition timing consistency indicates that the identified discrepancies in the small, independent dataset validation (see 5.14.2) are the result of specific characteristics of those populations, and that the HES-calibration basis is more generally consistent with national behaviours.

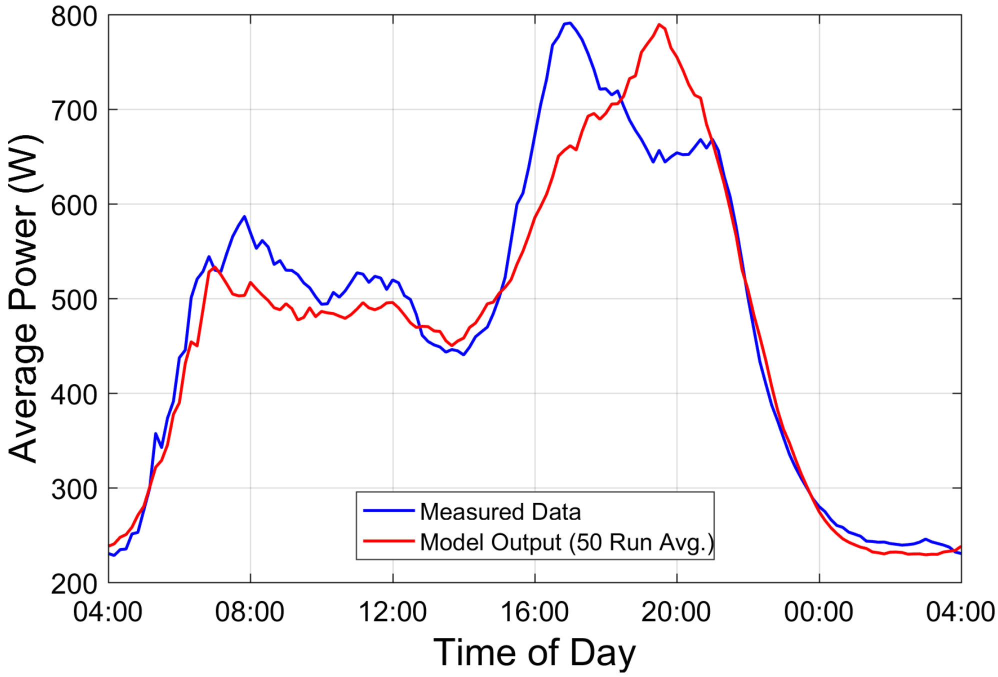
5.14.6.2 Social Housing

As outlined in 2.4.1.2, average electricity demand for social housing has been estimated to be lower than the national average by between 8 and 10% [76]. The national average population comprising 69% private, 24% social-rented, and 7% private-rented households.

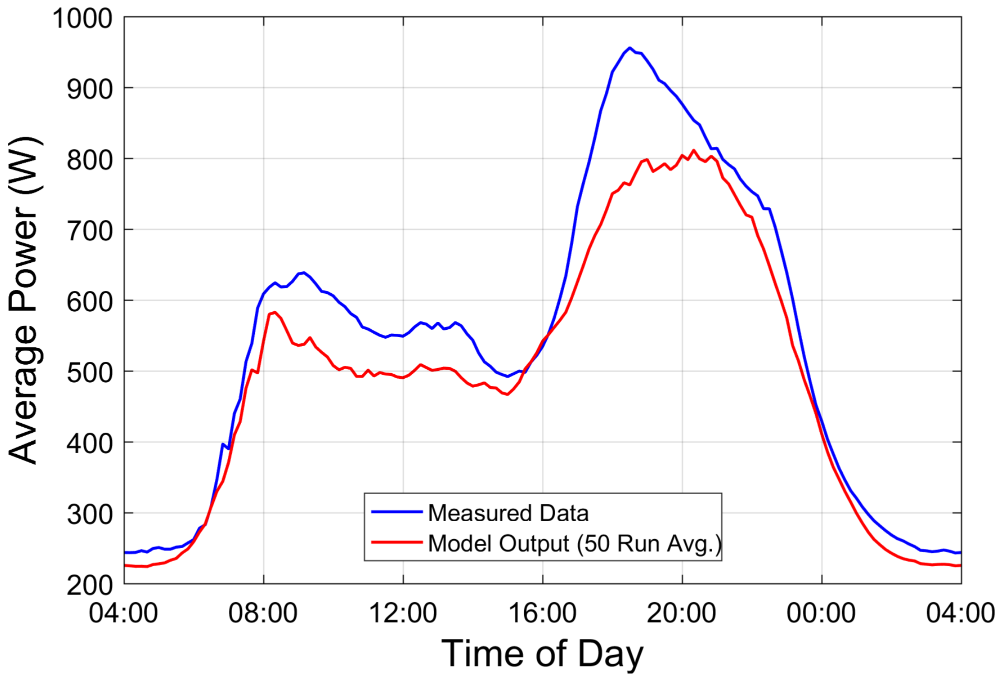
100 annual duration model runs each with 100-household representative populations were run for both the national average set of households and a social-housing only equivalent. The type and size of houses was identical for both populations, with other household characteristics determined probabilistically. The average demand predicted for the national average population was 409W, with a range from 354 to 476W, and for the social-housing population was 378W, 8% lower, with a range from 326 to 434W. The model therefore broadly captures the overall impact of reduced income and appliance ownership expected for a social-rented population.

5.14.6.3 Ashton Hayes (Deprivation Decile=10)

Ashton Hayes is a village in Cheshire with approximately 372 households. Scottish Power Energy Networks performed a substation level monitoring campaign [173] for the village incorporated as part of a wider project to assist the village to become carbon neutral. One of the substations has 75 households and no commercial connections, and is therefore suitable for analysis with the developed electricity demand model.



(a) May to September



(b) Annual

Figure 5.41. Ashton Hayes 75-household data and 100 model run average comparison for two time periods. Data for the 'Measured Data' distribution from [173].

Ashton Hayes is of specific interest for analysis as it is an area with a deprivation decile (IMD) of 10 (lowest deprivation) with larger than average houses sizes, and therefore represents an extreme test of the model's ability to capture the effects of house type and size, household type, income, and employment.

The 75 households comprise one 1-bed, seven 2-bed, thirty 3-bed, and thirty-seven 4-bed houses, which is a significantly larger mix than the national average (3.37 vs. 2.56 bedrooms). Typical floor area has been estimated based on publicly available floorplans from estate agent websites but is a source of potential error. The 2011 UK Census household data [103] for the larger area of 133 households which includes the 75 monitored households has been used to define the household composition and age profile.

LSOA-level gas connection data [174] shows that all households have a gas connection and therefore, it is assumed, all have non-electric main space and hot-water heating. To allow for the possibility of some secondary electric heating, two levels of analysis have been included. One for a full annual simulation basis and a second for only the May to September period when the use of any heating systems should be at a minimum. The latter is the principal performance comparison element.

For the May to September period the average power use for the 75-household area was 482W per household. The average model result over 50 runs was 471W, with a range from 363W to 560W. Given the estimated household floor area and possibility of a small contribution of electric secondary space heating to the measured data, Figure 5.41(a) shows a good correlation between measured and modelled data.

For the full annual period the average power use is 548W with the model predicting 495W. The potential for secondary electric heating use means the results are not directly comparable but comparison of the time-dependent profiles (see Figure 5.41(b)) highlights good correlation between key transition timings and a relatively consistent difference during typical waking hours.

5.14.6.4 Further Area Characteristics Modelling

Over a number of runs, a well-calibrated probabilistic model should converge to the average that would be expected for the chosen input conditions. To further confirm that the model replicates average demand behaviours, a further set of analysis has been

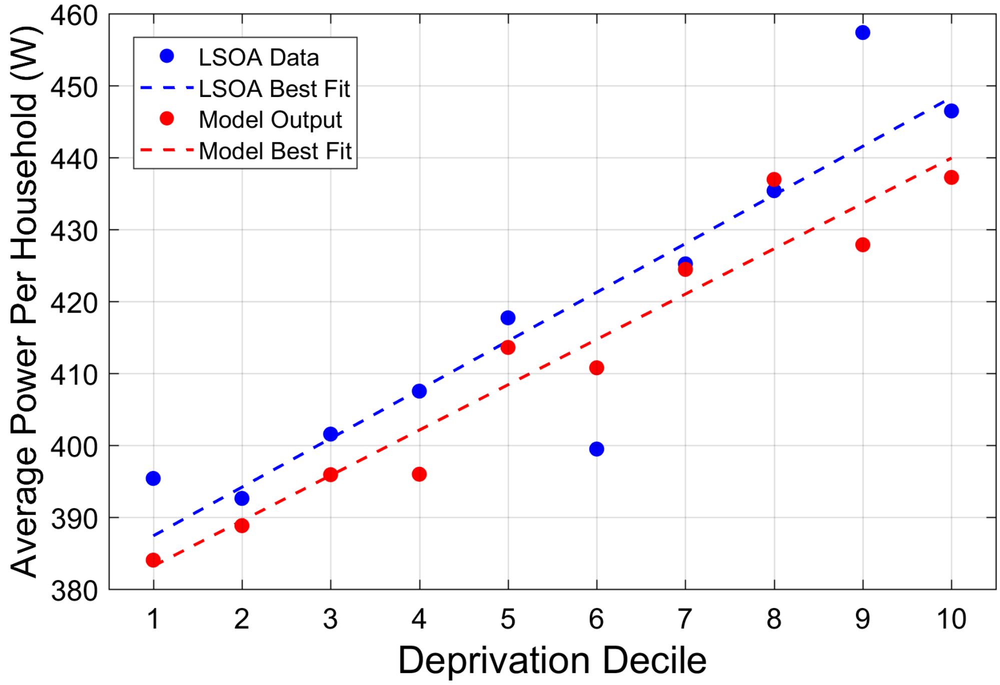


Figure 5.42. Average electricity demand comparison between the model output for average deprivation decile (IMD) household characteristics and the average for the twenty closest LSOA matches by household characteristics. Data for analysis from [46].

performed to determine if the model replicates average electricity demand at the LSOA level (LSOAs are 600-1000 household UK Census geographic areas) based on known area characteristics.

The method used was to run the model 50 times for each deprivation (IMD) decile (see 2.4.1.1) with typical proportions of social housing and house types, with the remainder of the household characteristics determined probabilistically by the developed sub-model (see 4.3). Based on the model results, the twenty closest LSOA matches for the model household characteristics were identified and average annual electricity demand for each LSOA determined from the same LSOA-level electricity dataset [46] used for the income-behaviour regression analysis in 5.3.2.

A Nearest Neighbour approach based on Euclidean distance is used to identify the twenty closest matches based on proportion of social housing, gas connectivity, average number of bedrooms, average number of people per household, average income, and proportion of retired households. The comparison between the average electricity demand for the twenty closest matches and the average from the 50 annual model runs is shown in Figure 5.42.

As the model does not currently capture secondary heating but also potentially underestimates the prevalence of extended absences, it was expected that the model output would be slightly lower than the LSOA data. By restricting the analysis to areas with full gas connectivity it was, however, expected that the secondary heating influence would be minimised.

The results indicate that the model tracks the average expected differential per area deprivation (IMD) decile with good accuracy as indicated by the consistency between best-fit plots, with the expected slightly lower model prediction as outlined. The model to data underestimate increases with deprivation decile which is also consistent with an increase in secondary heating use at higher income levels [175].

5.14.7 Validation Analysis Summary

The overall results of the validation exercise show that the model is broadly capable of replicating general and appliance-specific demand behaviours at the household-type level, average and time-dependent demand within small independent datasets, and av-

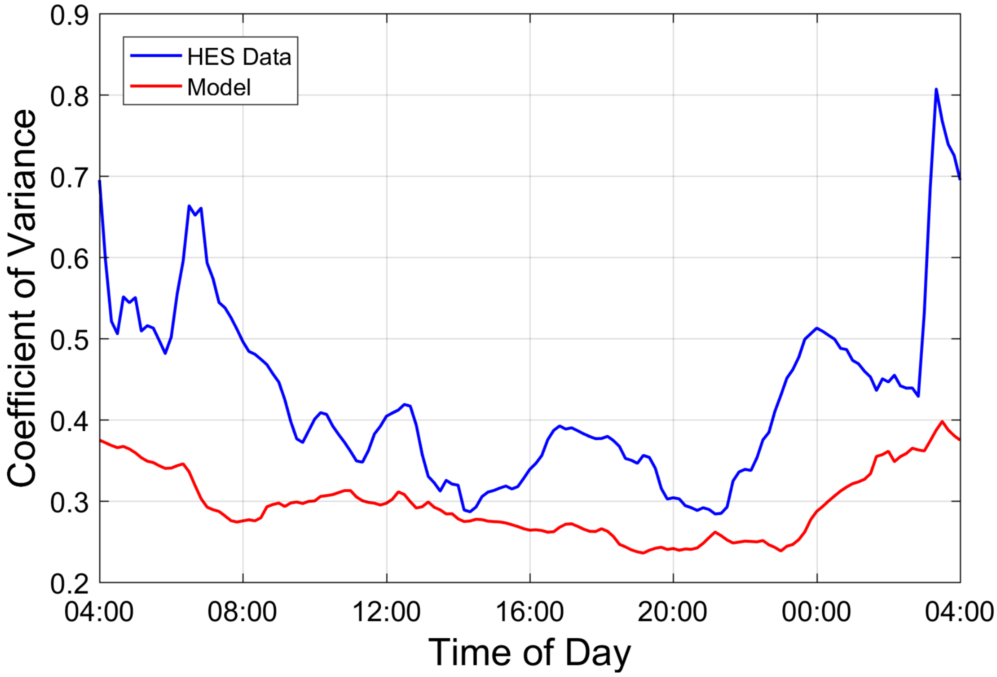


Figure 5.43. Per-timestep coefficient of variance of mean-normalised per-household electricity demand for measured HES data and equivalent modelled results. Data for the 'HES Data' distribution from [89].

erage demand variations for areas with different socio-economic characteristics. Within each household-type group, a significant range of individual household behaviours are also replicated, and in broadly realistic proportions, after multiple runs of the highly probabilistic model, but a residual group of approximately 10-20% are not.

The appliance-level analysis indicated that the cycle start time distribution for each household tends to converge to the calibration population average within a 1-month model period and that this significantly underestimated variations in timing per household for several appliances. Visual comparison of per-timestep occupancy and total demand profiles per household also indicate that the timing of the major occupancy and demand transitions (i.e. early morning, late afternoon, late evening) were overly similar in the simulated households in comparison with the measured data.

The model incorporates a number of factors to differentiate occupancy and demand behaviours. However, the occupancy model basis for each defined occupant and day type are currently based on the average behaviours for each defined group. Similarly, the appliance cycle start time module is calibrated using the combined behaviours of all households, with only minor adjustments for household type behaviours. This use of composite datasets for model calibration has the potential to result in convergence of behaviours, particularly for simulations run over extended periods (i.e. 6-12 months).

5.14.7.1 Residual Behaviour Averaging

To confirm quantitatively that these composite factors result in excessive behaviour averaging, the mean-normalised demand variance per timestep is compared for the HES measured and modelled data (i.e. based on measurement and model periods of 1-6 months). The results are shown in Figure 5.43.

The results indicate that the electricity demand model variance is significantly lower than for the measured data in the key transition periods of 6am to 9am, 4pm to 7pm and 10pm to 1am. The lower variance suggests that the model is not fully capturing individual behaviours in these time periods. Whilst potentially acceptable for larger-scale district modelling, an improved method or calibration basis would be beneficial to allow the model to be more applicable for individual households and smaller districts.

Improving replication of individual occupancy and appliance use behaviours is reviewed in Chapter 7.

5.15 Chapter Summary

This chapter detailed the development process for the electrical demand sub-model and the validation analysis undertaken to assess the model performance. The chapter highlights are as follows:

- Development of a major electrical appliance ownership model based on national survey and electricity demand dataset ownership data, including probability adjustments for household type and composition.

- Analysis of existing models and the Household Electricity Survey appliance-level dataset, determined that seven distinct demand sub-groups were required for effective demand modelling to account for different usage characteristics. As an improvement on existing models that used a single method for all intermittently used ('cyclic') appliances, three separate sub-groups were identified and different methods used. The other sub-groups ('AV', 'Lighting', 'Continuous', and 'Miscellaneous') used enhanced versions of existing methods.

- Identification of additional household-level electricity demand influencing factors to account for income, occupancy, and a further randomly allocated factor associated with attitude to energy use.

- Development of a new discrete-event based method for the 'cyclic' appliances that first determines use per day and then the timing of use on each day. The overall method was shown to reduce the probability of unrealistic cycle sequences compared to the per-timestep probability approaches used by the majority of existing models. Use per day per household was determined based on the distribution of average cycles identified for the specific household type with further factoring based on the identified household-level behaviours and day-specific occupancy, with a final per-day use determination using a binomial probability approach to account for natural variability from average behaviour.

- Development of a common method for the prediction of use times for each 'cyclic' appliance. Based on the predetermined number of cycles per day, cycle start times are allocated within occupied periods based on cumulative probability distributions unique to each daily total and specific cycle number.

- Validation performed showed that the overall developed model and each appliance-specific module replicates the calibration data basis with good accuracy. The highly probabilistic nature of the model, incorporating a significant number of household- and appliance-level use factors, was shown to replicate the range of both overall and per-timestep average results per household. Further validation of the developed electricity demand model showed good performance in replicating detailed independent time-series datasets, and matching overall and time-dependent average use for a variety of different area types.

- Detailed analysis of model output per household has indicated that the variance in demand timing is lower than for real populations. Further investigation of the model calibration, particularly with regard to group-calibrated elements, being required.

Chapter 6

Hot Water Demand Sub-Model Development

6.1 Chapter Overview

In Chapter 3 it was identified that few existing domestic demand models incorporate a hot water element. As hot water accounts for the same proportion of household energy use as electricity (c.17%), a value that will increase as heating demand falls, being able to simulate hot water demand with the same resolution is important. In particular, the design of district heating systems and CHP-driven energy systems requires the timing and potential peak consumption of hot water to be well understood.

Occupancy is not a sufficient basis alone to capture the time dependency of hot water use and multiple TUS activities would be required to capture all uses with uncertainty over the proportional allocation. As a result, in the same manner as the electricity demand sub-model development, direct analysis of demand data rather than inferred use from occupant activities has been employed. The Energy Savings Trust (EST) hot water dataset [90] (see 2.3) was therefore analysed using the same methods utilised for individual 'cyclic' appliances in the electricity dataset (see 5.5.2), to identify individual cycle timing and volume.

The available hot water data is more limited than the electricity equivalent in terms of the number of households, extent of available household characteristics data, and time resolution (10-minute vs. 2-minute). In general, the hot water demand sub-model uses the same basis as the 'Simple' electricity module (see 5.6) and therefore the focus of the analysis presented in this Chapter is on areas specific to the hot water demand sub-model development and to address the impact of the data limitations.

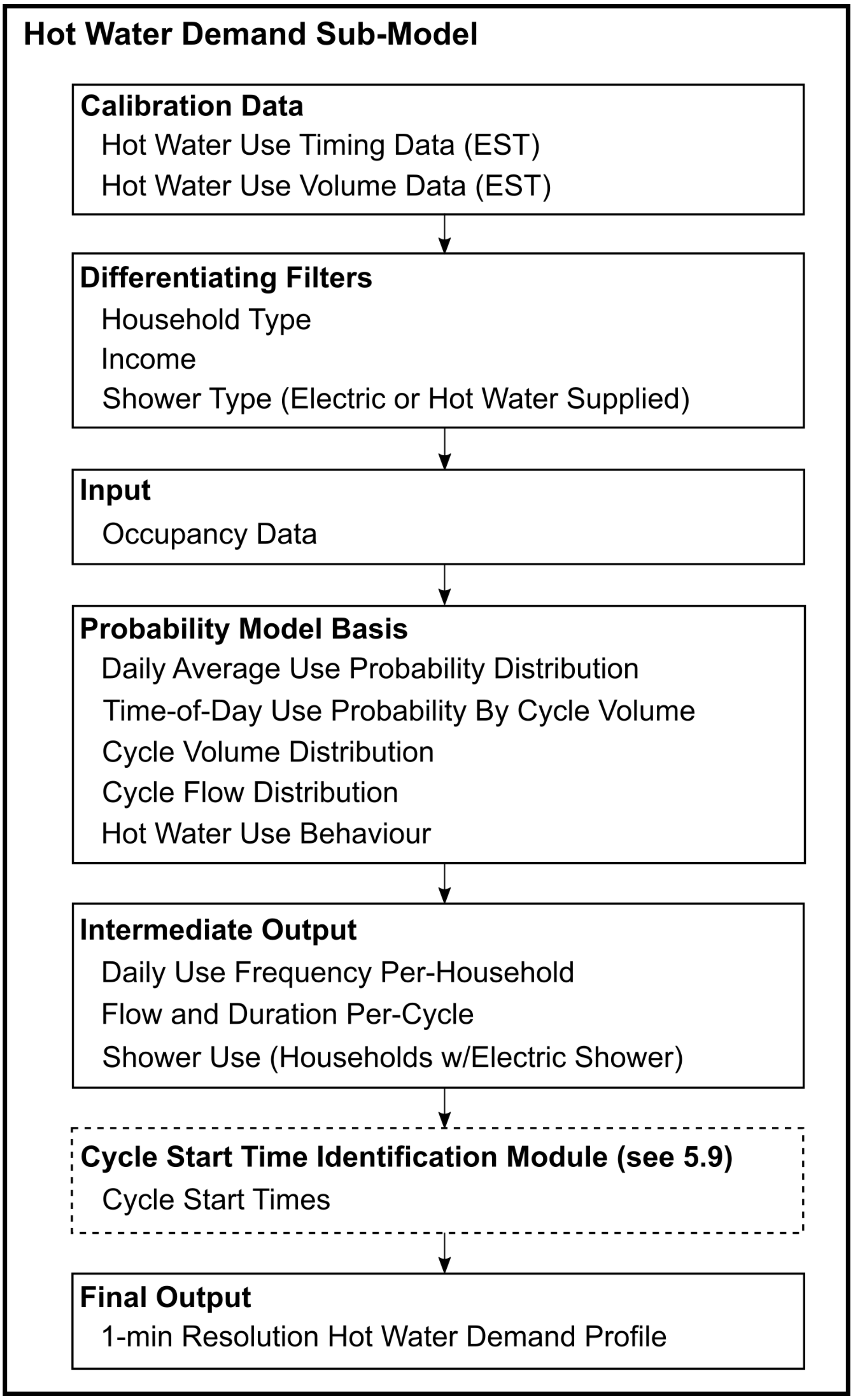


Figure 6.1. Hot water demand sub-model structure.

Initial dataset analysis determined that use characteristics varied based on household size, whether the household had children, and on the volume used per use. These elements were incorporated in the sub-model using differentiation by household size and type, and separate calibration for multiple cycle volume ranges.

6.2 Hot Water Model Basis

Table 4.1 detailed the overall demand model calculation sequence required to translate the outputs from the household characteristics and occupancy sub-models into demand prediction. The model simulates when hot water is used at the point-of-use. The dynamics of the hot water generation system, including storage, were not considered.

Final hot water energy use is also dependent on the incoming water temperature. The developed model uses the monthly temperature adjustment factors from the BREDEM model [111] to adjust final energy use per cycle. As outlined below, the existing hot water data does not allow detailed assessment of output temperature and therefore a more detailed basis for incoming temperature would not significantly improve model effectiveness.

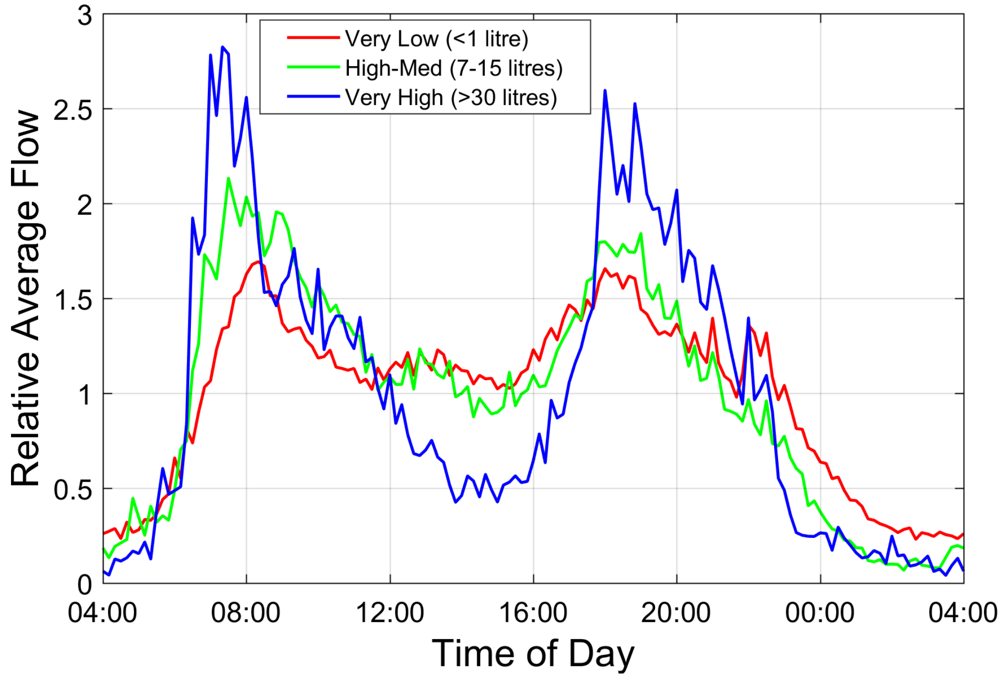
The following sections detail the data analysis, calibration, development, and validation of the hot water demand sub-model. The overall structure of the sub-model and principal defining elements are shown in Figure 6.1.

6.3 Hot Water Data Analysis

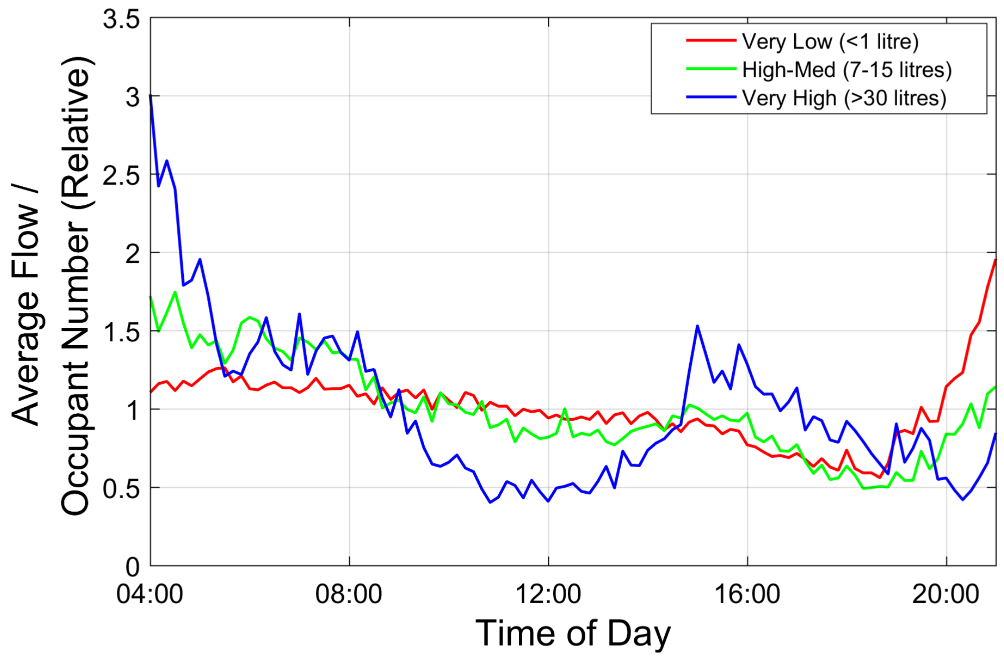
6.3.1 Hot Water Cycle Identification

The EST dataset (see 2.3) used for calibration has 10-minute resolution flow data, therefore identification of individual cycles and volume per cycle within each timestep is not possible. In addition, it cannot be determined if flow in adjacent 10-minute periods was due to a single cycle that overlapped the end of the period or two (or more) separate events. A series of assumptions listed below was therefore required to identify effective cycles from the raw data:

- Total volume used in each 10-minute period is from a single event.



(a) Average flow



(b) Average flow relative to occupancy

Figure 6.2. Average flow profiles for selected cycle volume ranges relative to average from measured hot water use data. Data for analysis from the EST dataset [90].

- There is a defined probability that non-zero use in adjacent time periods is due to a single overlapping cycle. The probability is set arbitrarily at 25% for cycles of less than 3-litres, 50% if between 3 and 15 litres, and 75% if over 15 litres, assuming an increasing likelihood with increasing volume. It has been determined that there is only a small additional error if some individual hot water events in adjacent periods are erroneously combined.

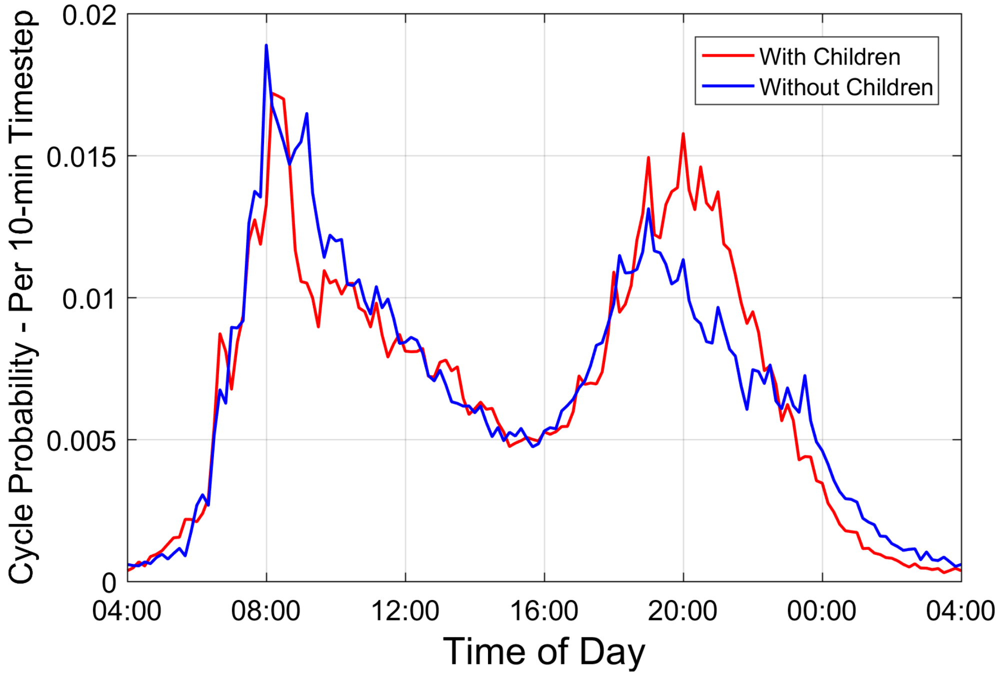
- Differences in hot water output temperature are ignored. This is primarily due to the 10-minute resolution of the temperature measurements making any detailed analysis of energy (volume plus temperature) rather than volume only unreliable. The average output temperature of the overall dataset is 51.9^{o}C and the range has a balanced Gaussian distribution with a $\pm$10^{o}C range. Therefore, any errors introduced for individual events should offset.

6.3.2 Hot Water Use Characteristics

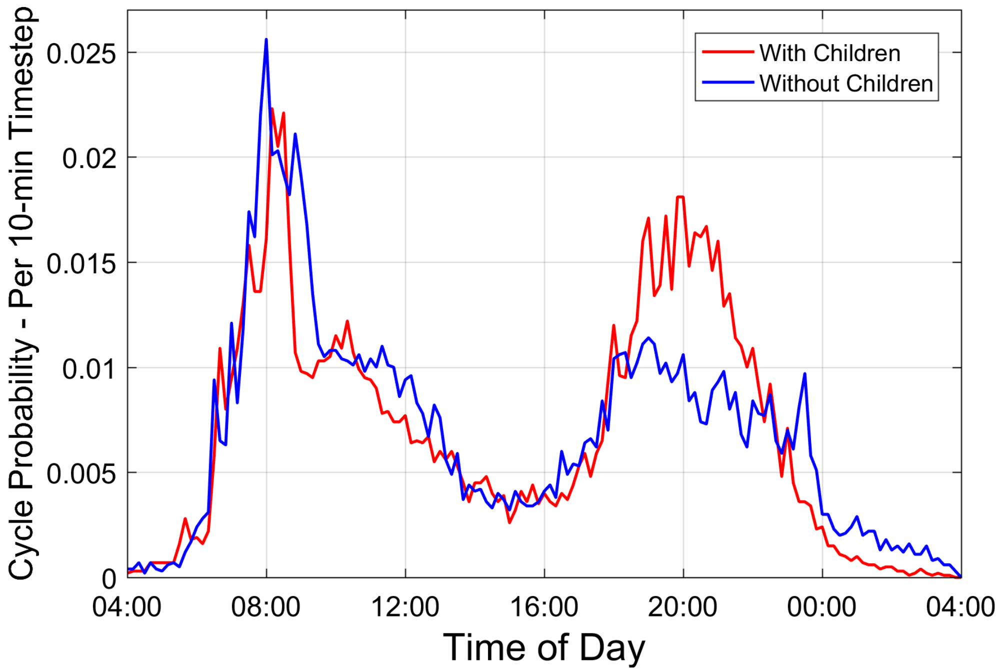
The modified dataset based on the defined assumptions comprises approximately 650000 hot water cycles. Analysis of the distribution of cycle volumes highlighted the predominance of low volume cycles by number (50% are less than 3 litres) but that, in terms of total volume contribution, the full range of cycle volumes are significant.

Further analysis also determined that the time dependency of use varied with cycle volume. Figure 6.2(a) shows the time-dependent profiles for three selected volume ranges relative to the average value. Within each volume range different behaviour patterns are evident. The lower volume ranges show a correlation with occupancy probability (see Figure 6.2(b)), although there are higher relative use periods before and after the typical sleep period. As the cycle volume increases, there is an increasingly distinct dual peak distribution indicating specific time-dependent behaviours.

To capture different behavioural patterns and to distinguish between different typical uses, the data was split into six cycle volume ranges (0-1, 1-3, 3-7, 7-15, 15-30, 30+ litres). The specific choice of ranges is to an extent arbitrary as the changes in use behaviour are gradual but those selected can be aligned with different typical uses. The 0-1 litre range, which accounts for 31% of cycles and 2% of total volume, would be typically associated with short, handwashing events. The 1-3 litre range (19%/5%) is



(a) All Flow Ranges



(b) Very High (30+ litres/cycle)

Figure 6.3. Cycle timing per-timestep relative probability for family and multi-adult households. Data for analysis from the EST dataset [90].

assumed to be longer sink washing events, the 3-7 range (18%/11%) hand dishwashing and other full sink or bucket uses, and the 7-15 range (16%/19%) covers large volume dishwashing up to low volume showers. The 15-30 (8%/18%) and 30+ litre (8%/45%) ranges are assumed to be predominantly baths and showers.

Whilst there may be patterns of behaviour associated with bath and shower use (i.e. an increasing need with time similar to washing machine and dishwasher use), the arbitrary volume ranges used and overlapping influence of multiple individuals in larger households makes this difficult to discern. Analysis of individual household data, even for one-person households, shows that the use day sequences for the higher volume cycles are indistinguishable from random patterns and it is therefore assumed that the influence of any underlying patterns on modelling accuracy are small.

Analysis of the EST dataset by the EST [90] determined that occupant number was the most significant determinant of hot water use rather than household type or child presence. Unlike the HES electricity demand dataset, the socio-economic data does not define occupant ages, restricting the potential for differentiation. Cycle timing (but not total number) was shown to be influenced by child presence. Figure 6.3(a) shows a tendency for family households to have lower morning use and an earlier evening reduction in use potential. In contrast, very high volume cycles in adult-only households are predominantly in the morning period (see Figure 6.3(b)).

Table 6.1
Average total daily hot water cycles by volume range and household size. Data for analysis from the EST dataset [90].

Household Size	Overall	Cycle Volume Range (litres)					
		0-1	1-3	3-7	7-15	15-30	30+
1-Person	10.56	4.38	2.50	2.15	1.10	0.28	0.15
2-Person	12.75	4.24	2.95	2.36	1.83	0.73	0.64
3-Person	15.06	5.66	3.46	2.75	1.85	0.61	0.65
4-person	16.37	5.64	3.38	2.43	2.17	1.28	1.48
5+person	20.44	6.01	3.97	3.74	3.41	1.83	1.48

6.3.3 Hot Water Cycle Number

The overall number of hot water cycles per household size, determined using the method defined in 6.3.1, shows a consistent linear increase with occupant number as shown in Figure 6.4. This allows the relationship to be converted to a simple equation (see

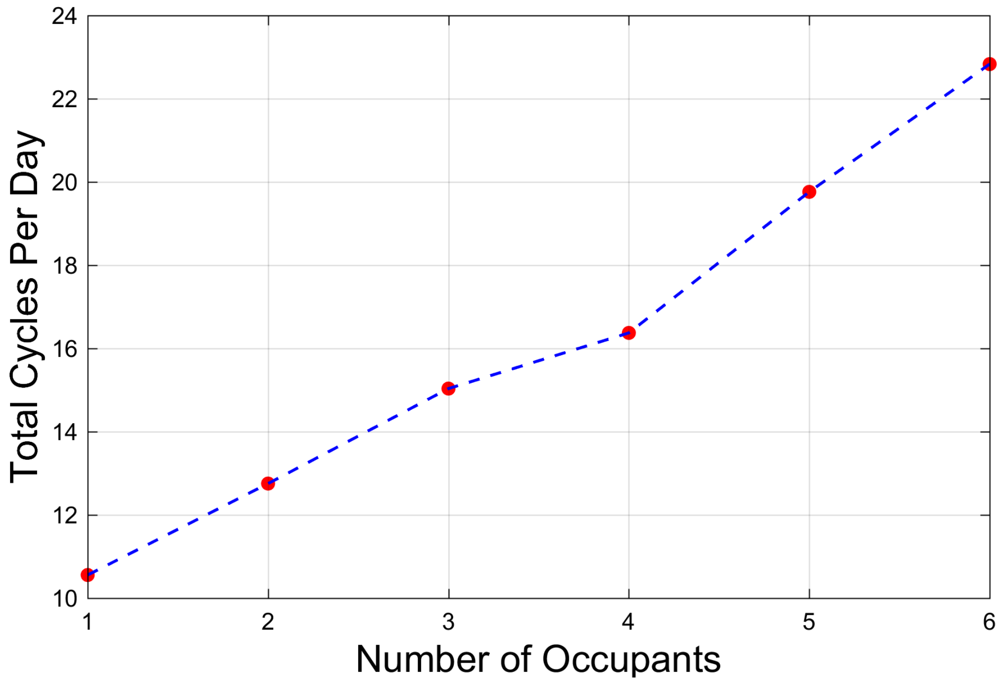


Figure 6.4. Average total daily hot water cycles by occupant number. Data for analysis from the EST dataset [90].

Equation 6.1). More detailed analysis of each identified volume range for average number of cycles per day is shown in Table 6.1.

$$TotalCycles = 7.85 + 2.40 \times OccupantNumber \tag{6.1}$$

6.3.4 Hot Water Cycle Timing

The same cycle start time probability distribution method used for the electricity demand model (see 5.9.2) was also used for the hot water model. The 10-minute EST-derived cycle data was extrapolated to give 1-minute resolution calibration data to allow 1-minute resolution results to be generated. It was therefore assumed, prior to the availability of higher resolution data, that within each 10-minute period the cycle start time is effectively random.

The additional household type manipulation to the cycle start time distributions in the electricity demand model (see 5.9.1) was performed on the hot water distributions for child and non-child households to account for the different use behaviours identified in 6.3.2. Separately calibrated manipulations were carried out for each volume range as the influence was shown to vary by range.

6.4 Hot Water Sub-Model Development

6.4.1 Modelling Basis

Each identified hot water volume range has been modelled in the same manner as a 'Simple' electricity appliance (i.e. cycles on each day considered independently based on an overall average) (see 5.6). Households have been differentiated by occupant number for cycle number, and by occupant number and whether children are present for cycle timing, driven both by the inherent socio-economic data limitations of the EST dataset and the evidence of behavioural differences at this level of differentiation outlined in 6.3.2.

6.4.1.1 Electric Shower Evaluation

The EST dataset was derived from hot water use measured at the tank or boiler outlet (depending on the system type), which does not include use of electric showers if owned and the dataset does not indicate shower type. Analysis of the 'High' and 'Very High' cycles in the dataset determined there were 0.636 cycles per day per person. The expectation, assuming 95% of all 'High' and 'Very High' cycles are either showers or baths, would be 0.857 based on further analysis by EST [176] and SAP [48]. This would suggest c.35% of EST households had an electric shower compared to a 47% national average. The 35% figure is consistent with the number of households with very low levels of measured higher volume water use.

Two separate model versions have therefore been developed: one based on the EST data directly to allow the overall model output to be validated ('EST Equivalent'); and the other with the number of higher volume hot water cycles increased to give a 0.857 cycles per day per person average, and where electric shower ownership can be defined either as zero for district heating systems, to the national average basis for gas and non-gas connected households, or on a user-defined basis. The type of shower does not impact the point-of-use demand assessment but determines how the energy for the proportion of higher volume cycles defined above is allocated (i.e. to electricity supply or hot water generation).

The national average electric shower ownership probability was determined from the Energy Follow Up Study (EFUS) dataset [43], with a 42.3% probability if a house has mains gas and a 50.3% probability if not.

6.4.2 Hot Water Behavioural Factors

As similar combination of behavioural factors to those determined for the electricity demand sub-model (see 5.3) (i.e. income, occupancy, and random behaviour) were incorporated as a single overall behavioural multiplier, the Hot Water Behaviour Factor. These were determined as follows:

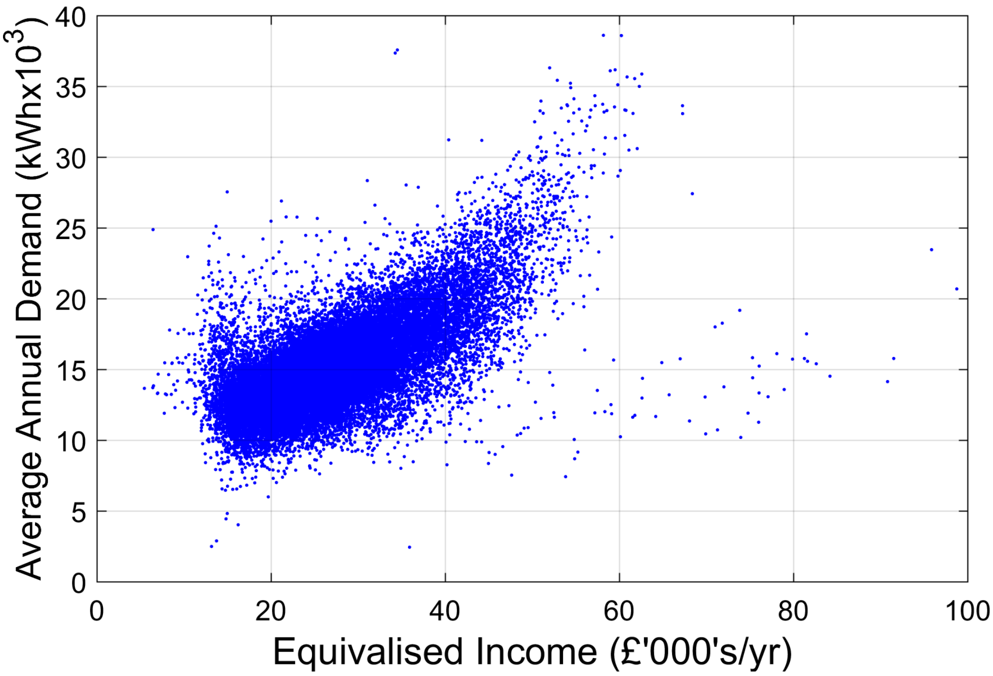


Figure 6.5. Average household annual gas consumption by average equivalised income per English Lower-Layer Super Output Area (LSOA). Data from [47].

6.4.2.1 Effect of Income on Gas Demand: Income Behaviour Factor

Similar analysis was carried out for the income influence on gas demand behaviour as detailed for electricity demand in 5.3.2 using the same set of Lower Super Output geographic areas (LSOAs) [47]. The plot of LSOA-average gas demand against equivalised income (Figure 6.5) demonstrates that the impact of this variable is again complex, with a weaker exponential component than for electricity demand. The same regression factors were used with the exception that 'Owned Appliance Power' was replaced by the following:

- *Building Heat Loss(Relative)* – This factor is based on the multiplication of two separate relative (base=1) factors. One related to the mix of house ages in an area and the other related to the mix of house types. Separate factors are required due to the lack of a single dataset with both ages and types of dwellings detailed. The factors are based on the estimated annual heat demand for an average size dwelling of each age and type based on the original building code and allowing for typical upgrade levels for older dwellings (double glazing, loft insulation etc.). These values are then combined proportionally for each LSOA.

The derived equation for the income-specific multiplier on gas demand (*GIBF*) is as follows (based on an average demand of 15277kWh from the LSOA analysis, IncF = Equivalised Income (2011 Basis) as defined in 5.3.2.):

$$GIBF = 0.8556 + (172.4/15277) \times (IncF/10000)^{2.2} \qquad (6.2)$$

In comparison to electricity demand, gas demand has a lower income exponent but a greater linear variation and a more significant overall income impact as demonstrated by Figure 6.5. As it is not possible with the available data to distinguish between heating and hot water use, the multiplier is used for hot water with the acknowledgement that this may be slightly inaccurate.

Table 6.2
'Relative Occupancy Factor' per hot water cycle volume range.

Range	Very Low	Low	Low-Med	High-Med	High	Very High
Relative Occupancy Factor	0.5	0.4	0.33	0.2	0.05	0.05

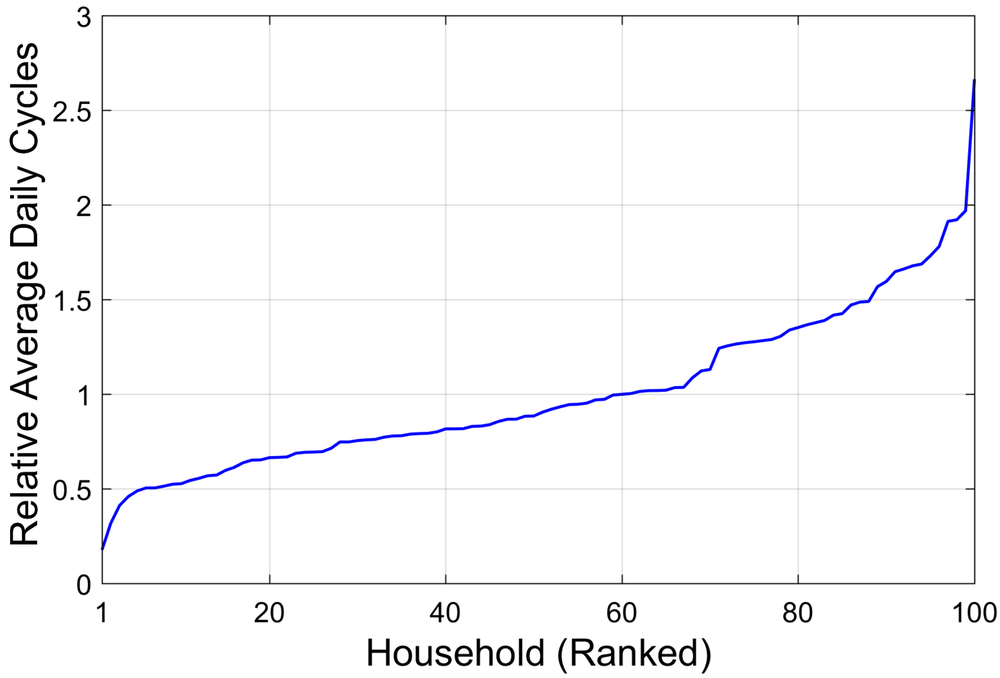


Figure 6.6. Consolidated and ranked average daily total hot water cycles per-household to household size mean cycles ratio distribution for all EST households. Data for analysis from the EST dataset [90].

6.4.2.2 Effect of Occupancy on Hot Water Demand: Relative Occupancy Factor

Without age and employment data in the EST dataset, the relative occupancy influence on hot water use cannot be determined. Equivalent 'Overall' and 'Daily' occupancy factors to those used in the electricity sub-model have therefore been estimated as shown in Table 6.2, with the same value used for both factors per volume range. The 'Overall' factor ($OROF$) determines the relationship between household and calibration population average occupancy and cycle number (see Equation 5.3), and the 'Daily' factor ($DROF$) determines the relationship between household daily and average occupancy and cycle number.

As for the electricity sub-model, the factors are applied to a modified occupancy factor that also accounts for variations in use potential based on time of occupancy. The determination of the *HhldOcc* and *TypeOcc* values used in Equation 6.3 is identical to the method defined in 5.3.1.

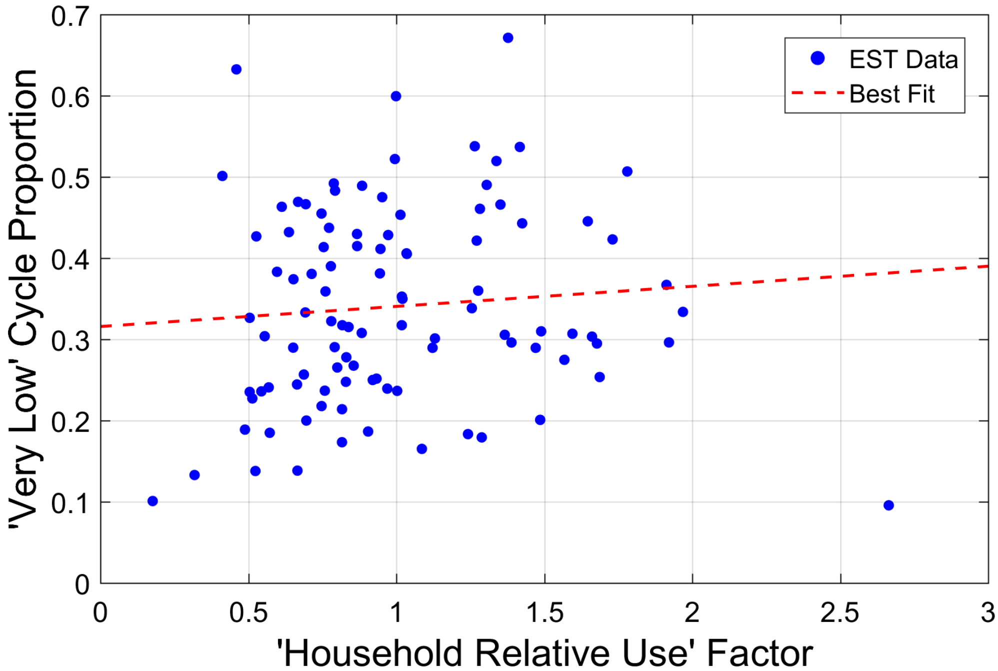
6.4.2.3 Effect of Behaviour on Hot Water Demand: Random Energy-Use Behaviour Factor

The analysis by Gill et al [81] determined that for hot water the random behavioural demand variation between households based on attitudes to energy use that cannot be discerned from household characteristics is 11% (compared to 37% for electricity use). The sub-model captures this variation with a random behavioural multiplier ($WRBF$) selected randomly between 0.94 and 1.06 (equivalent to 11%).

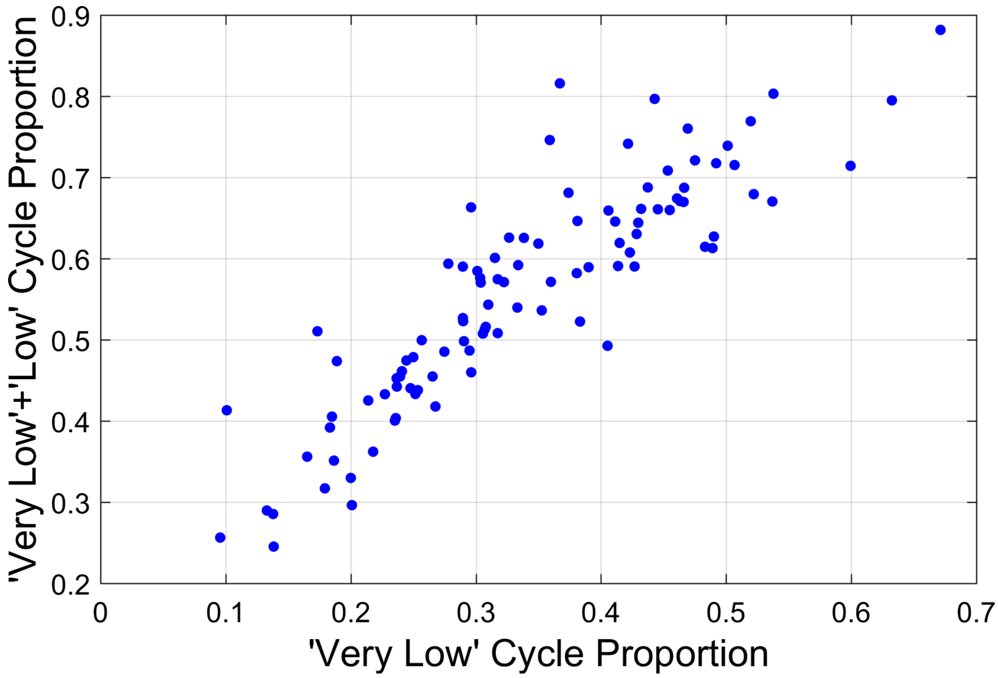
6.4.2.4 Household Hot Water Behaviour Factor

The overall behavioural factor ($HWBF$) applied to each modelled household is determined by the following equation. How the multiplier is used is defined in the following section:

$$HWBF = GIBF \times WRBF \times \ (HhldOcc/TypeOcc)^{OROF} \tag{6.3}$$



(a) 'Household Relative Use' factor to 'Very Low' proportion



(b) 'Very Low' to 'Very Low' + 'Low' proportion

Figure 6.7. Cycle volume range proportion relationships per EST household. Data for analysis from the EST dataset [90].

6.4.2.5 Household Relative Use Factor

In the same manner as the 'Appliance-Use Factor' determined for each individual electrical appliance (see 5.5.1.2), total hot water cycle number variation within household size (occupant number) groups is further differentiated by analysis of the overall variation to the household group mean. Similar to the electricity analysis, the distributions were found to be consistent for all groups and a single combined ratio-to-mean distribution was therefore developed based on the results for all households to their respective group mean values. The overall relationship is shown in Figure 6.6 and the distribution is used to probabilistically allocate a 'Household Relative Use' factor to each modelled household.

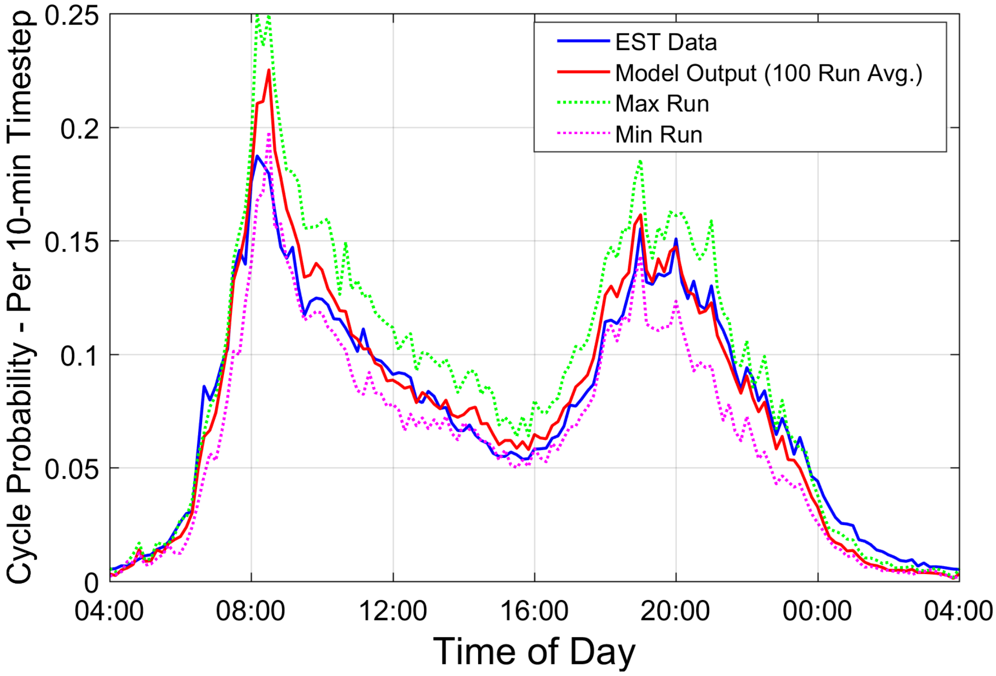
6.4.3 Sub-Model Structure

The analysis of the EST dataset showed that there was significant variation in both the total number of daily hot water cycles and the proportion in each volume range, both between each occupant-number differentiated group (see Table 6.1), and also within each group. A probabilistic method was therefore required to allocate the number of cycles per volume range to each modelled household. The steps are as follows:

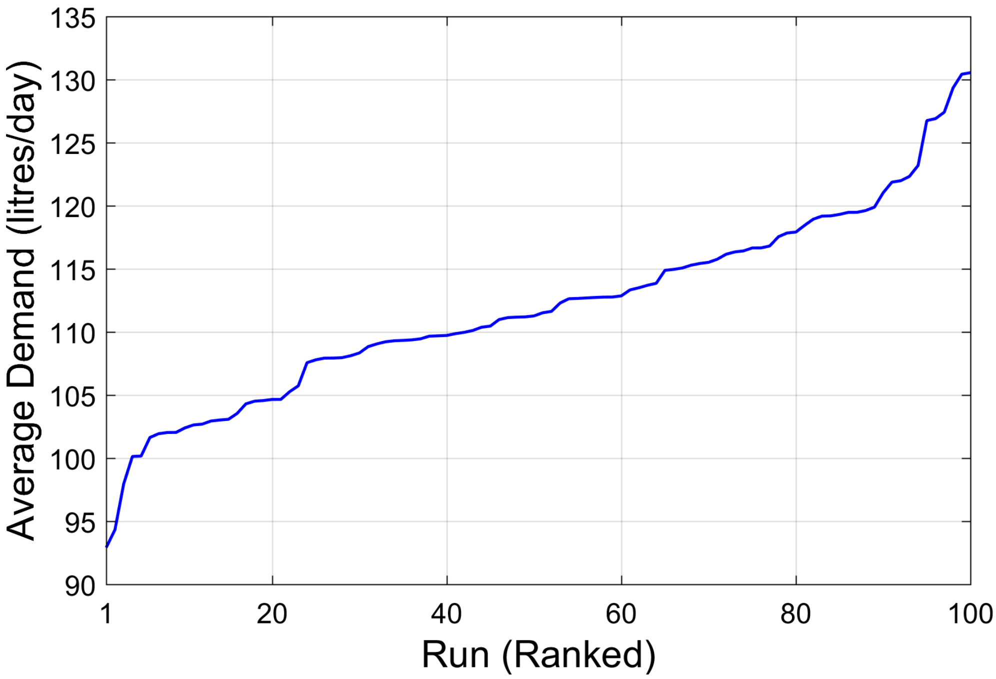
- *Step 1* - A total number of daily hot water cycles is allocated per-household using Equation 6.1 based on the number of occupants with the determined mean value multiplied both by a randomly selected 'Household Relative Use' factor and by the generated Hot Water Behaviour Factor ($HWBF$) for the household, both of which are defined above.

- *Step 2* - A series of Kernel Density (KD) relationships (see Appendix A) are then used to determine the cumulative proportion of cycles in each successive volume range from 'Very Low' to 'High'. This method ensures that the variation in behaviours is captured but restricted to realistic relationships.

  - *Step 2a* - First the Household Relative Use factor is used to determine the proportion of cycles in the 'Very Low' range, based on a discernible relationship as shown in Figure 6.7(a). EST data analysis also determined that the 'Very Low' cycle proportion was related to the number of people (decreases



(a) Comparison of model output with EST calibration data



(b) Range of model-predicted average per-household daily hot water demand

Figure 6.8. Model output and EST dataset comparison for total hot water use - 100 annual model runs average for EST-dataset equivalent households. Data for the 'EST Data' distribution from [90].

with increasing people) and to account for this the random number generation for the KD analysis is restricted to the equivalent of the minimum and maximum proportion values for each occupant number group.

  – *Step 2b* - The determined 'Very Low' proportion is then used to calculate the cumulative proportion of 'Very Low' and 'Low' cycles. The same KD process as Step 2a is used with the 'Very Low' proportion as the input value. Figure 6.7(b) shows the relationship between 'Very Low' and 'Very Low + Low' proportions used to calibrate the KD module. This process is repeated for each subsequent volume range, with the 'Very High' proportion determined by the value that sets the cumulative proportion to 1.

- *Step 3* - The average number of daily hot water cycles in each volume range are determined from the generated total number and proportion in each range.

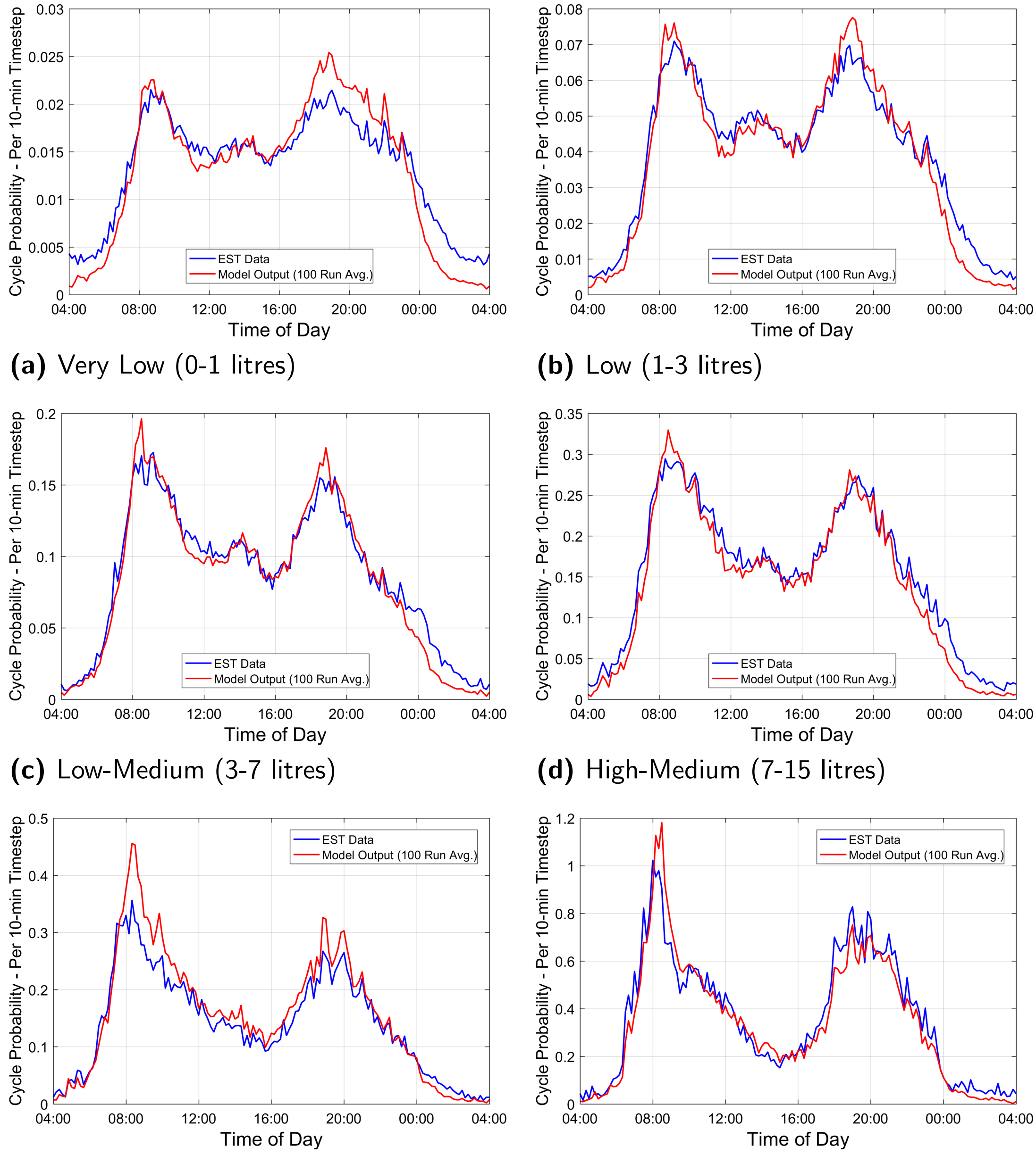
- *Step 4 to end* - The remainder of the hot water sub-model is identical to the 'Simple' electricity module (see 5.6) and the cycle start time identification module detailed in 5.9.2.

6.5 Hot Water Validation

The hot water demand model has been validated for overall average demand, average demand per timestep, distribution of modelled demands, ability to replicate actual profiles over a significant number of runs, and peak demand prediction (i.e. diversity) for multiple household systems. As outlined, the EST dataset used for model calibration is the only large scale UK water use dataset that is freely available. Therefore, validation is restricted to replication of the input dataset and comparison of diversity prediction with current standards.

6.5.1 Total Flow Replication

Comparison of the EST dataset average demand with model results based on equivalent households using known characteristics (number of adults and children) and allowing the model to probabilistically select undefined inputs (e.g. income, employment, etc.), shows that the model converges to the dataset per-household average closely but with



(a) Very Low (0-1 litres)

(b) Low (1-3 litres)

(c) Low-Medium (3-7 litres)

(d) High-Medium (7-15 litres)

(e) High (15-30 litres)

(f) Very High (30+ litres)

Figure 6.9. Model output and EST dataset comparison for each cycle volume range - 100 annual model runs average for EST-dataset equivalent households. Data for the 'EST Data' distribution from [90].

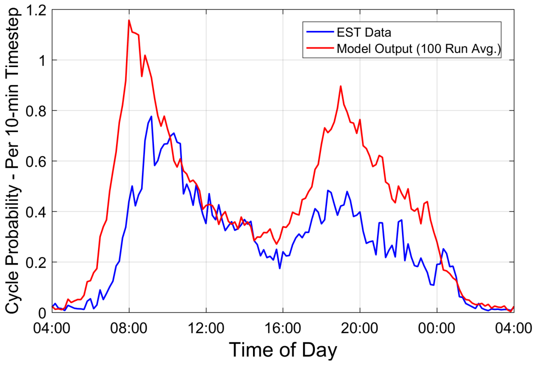
distinct variation between minimum and maximum model runs (see Figure 6.8(a)). This variation is also highlighted by the range of per-household daily average demand for each of the 100 runs (see Figure 6.8(b)). This shows that the basic calibration of the model with regard to linking occupancy and use cycles is effective, but that the model does not force convergence over a small number of runs. More data is required to determine if this modelled variation is realistic.

The major discrepancies between model output and data are in the morning peak period where the model overestimates this peak value and from midnight to 2am where there is an underestimate. The discrepancies are related to specific volume ranges and discussed further in the following section.

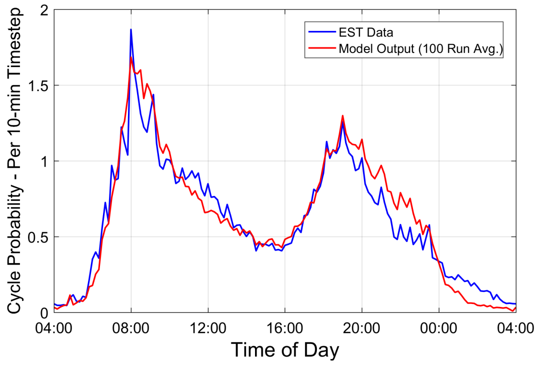
6.5.2 Volume Range Replication

The average model results for each of the six volume ranges (see Figure 6.9) also shows generally good replication of the EST dataset averages with some discrepancies. The overestimate of the total demand in the morning peak period is primarily a result of the higher volume modules. A potential reason for this would be that occupants are more time constrained in this period which reduces volume used per baths and shower use. Further investigation of this time dependence is required. The other significant discrepancies are for the 'Very Low' (<1 litre/cycle) and 'Low' volume (1-3 litres/cycle) ranges in the night period. The model underestimates the frequency of short handwashing events during this period, although the overall impact on model accuracy is small.

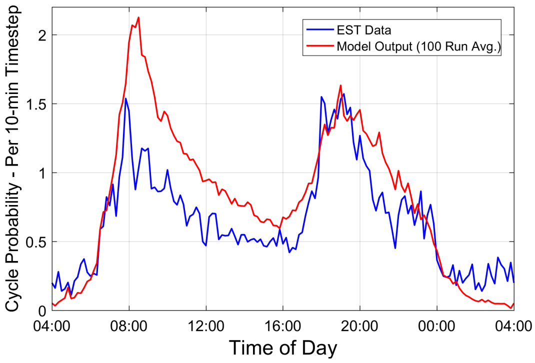
Discrepancies in the sleep transition period (11pm to 1am) for the lower volume ranges are consistent with the results of the electricity demand model. Again, the reason is not immediately obvious and further review of the occupancy model output is required to determine if it is unrepresentative in this period or if there are specific time dependencies for within-range cycle volume per use that are not currently captured.



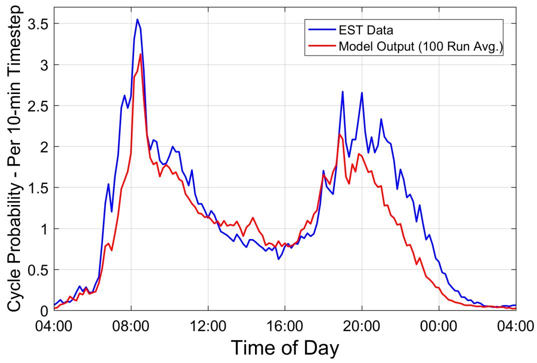
(a) 1-Person



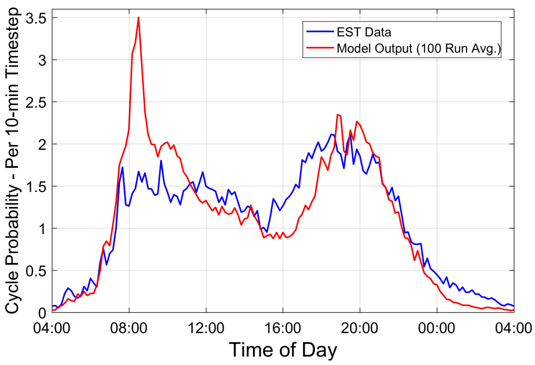
(b) 2-Person



(c) 3-Person



(d) 4-Person



(e) 5+Person

Figure 6.10. Model output and EST dataset comparison for each household size - 100 annual model runs average for EST-dataset equivalent households. Data for the 'EST Data' distribution from [90].

6.5.3 Sub-Population Replication

6.5.3.1 Household Size

Beyond the number of daily cycles and the input from the occupancy model, the hot water sub-model does not make significant differentiation for household size. The small number of each size in the EST dataset, particularly small and large households, and the lack of detailed characteristics data, currently limits effective differentiation.

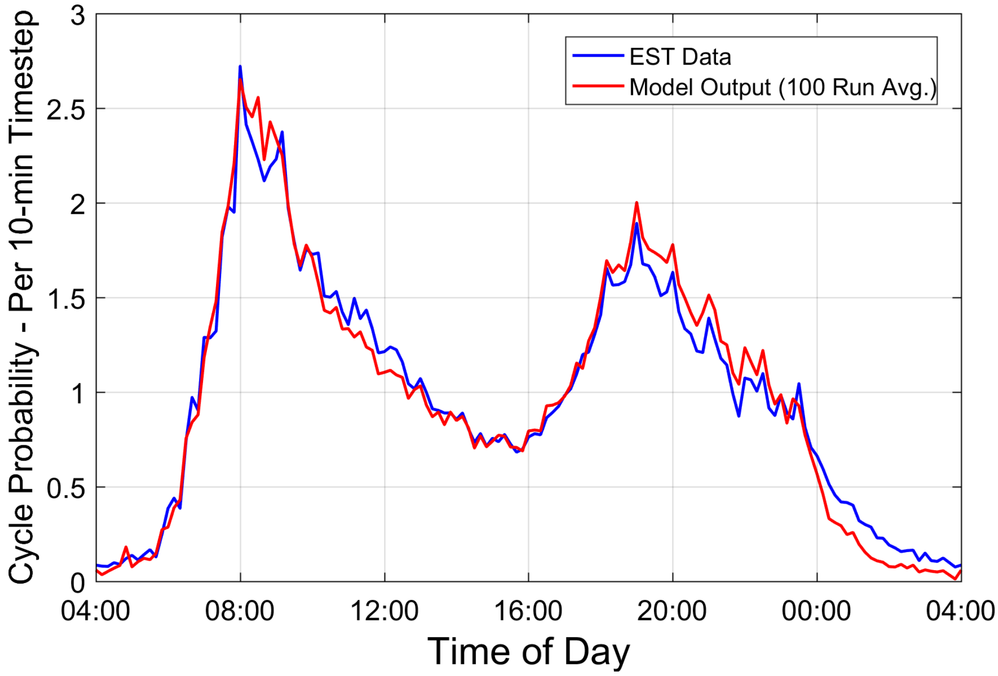
As shown in Figure 6.10, the model output is broadly consistent with the EST input data for the 2- and 4-person households but there are significant discrepancies for the other household sizes. The 2- and 4-person households represent the largest available datasets (34 and 28 households respectively out of 107 in total), and, in the case of 2-person households, are adult-only which removes one potential variable.

Table 6.3 compares the EST dataset and model output volume averages for each household size with that estimated by the BREDEM/SAP model basis of 36+25N litres [177], where N is the number of occupants. The 1- and 3-person EST dataset households, in particular, have a significantly lower actual use than predicted by either the model or BREDEM basis.

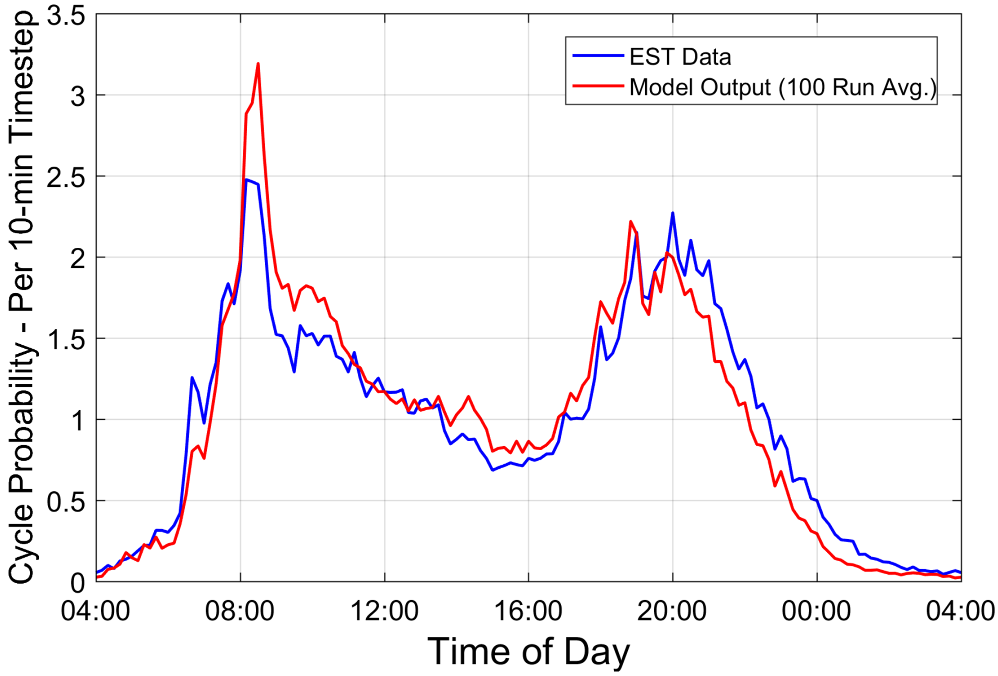
Table 6.3
Average hot water use per household (litres per day) comparison for different methods and datasets. 'EST Data' from [90] and 'BREDEM/SAP Basis' from [177].

Occupant Number	1	2	3	4	5+
EST Data	39.5	88.3	95.3	165.5	156.4
BREDEM/SAP Basis	61	86	111	136	161
Model Output (100 Run Avg.)	60.9	90.3	117.8	138.7	158.2

The average per-cycle volume within each volume range does not vary significantly per household type and the average number of cycles per household size follows a broadly linear increase as shown in Figure 6.4. The primary cause of the variation based on household size is the proportion of each volume range per household. The 1-person and 3-person households have a significantly lower daily probability of 'High' and 'Very High' cycles per person (see Table 6.4), suggesting lower than average use of showers and baths or higher than average ownership of electric showers that are not identified in the EST dataset. The model output closely matches the 'per-person' impact of additional people predicted by the BREDEM basis and is therefore potentially more



(a) Without children



(b) With children

Figure 6.11. Model output and EST dataset comparison for households with and without children - 100 annual model runs average for EST-dataset equivalent households. Data for the 'EST Data' distribution from [90].

representative of overall behaviour than the data for the less numerous EST dataset sub-populations.

Table 6.4
Total 'High' and 'Very High' volume cycles per day per person by household size from the EST dataset. Data for the 'EST Data' distribution from [90].

Occupant Number	1	2	3	4	5+
Cycles Per Day Per Person	0.41	0.78	0.45	0.72	0.62

6.5.3.2 With and Without Children

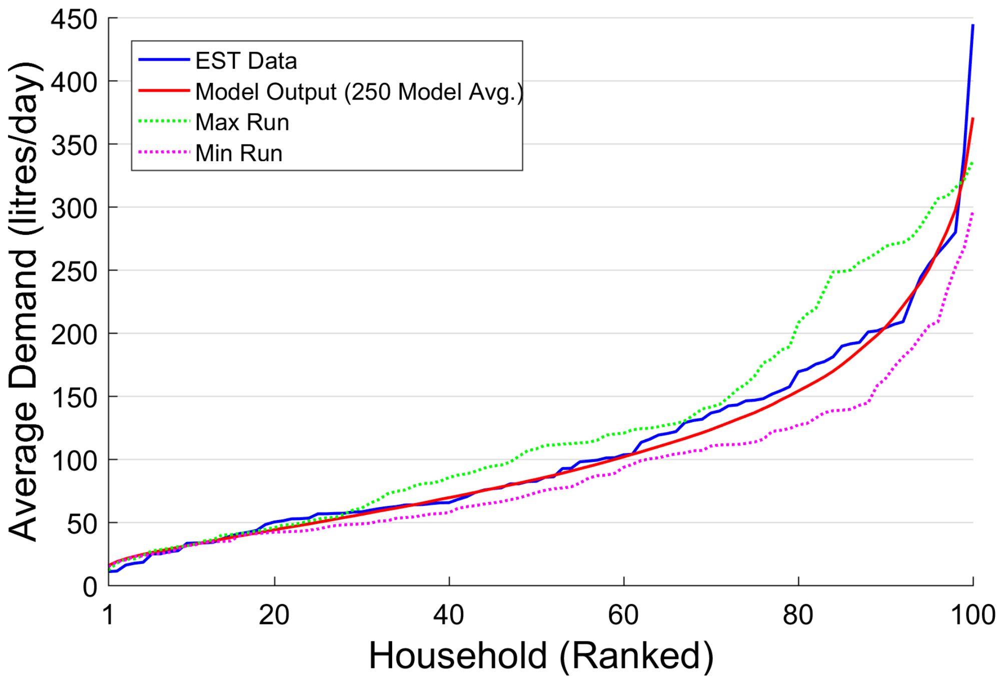
As shown in 6.3.2, there are differences in relative timing of use between households with and without children. Family households show more distinct morning and early evening peak use, and some evidence of higher use in the mid-afternoon period.

Model results (see Figure 6.11) show that these general patterns are captured, particularly for the households without children. The family household results show some variation from the data but given that the age of the children in the EST households is unknown, some level of timing inaccuracy is expected.

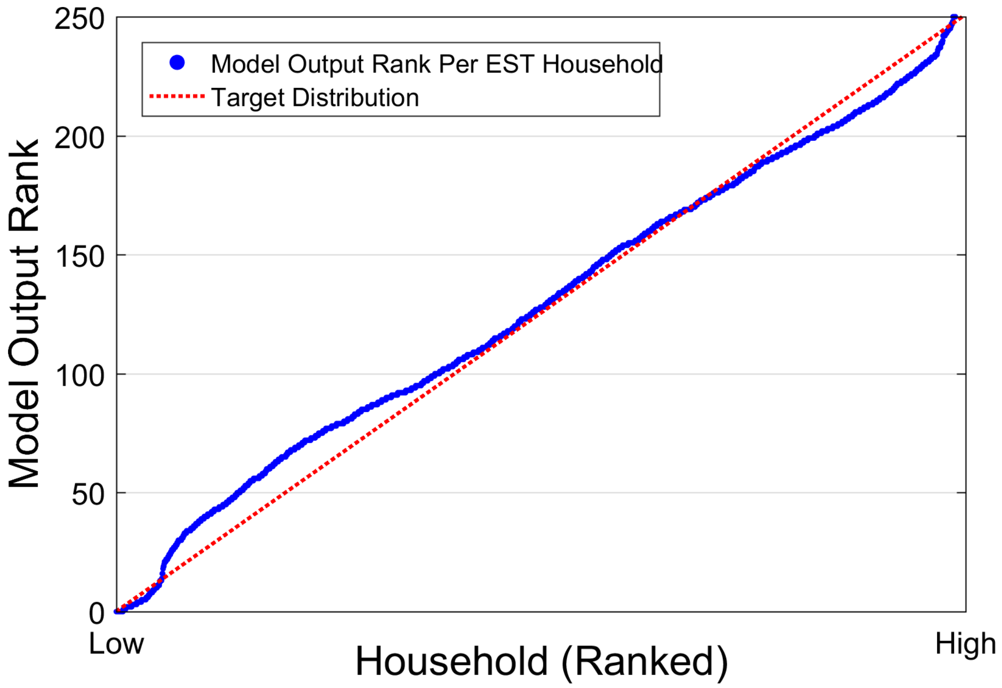
6.5.4 Demand Distribution Analysis

Overall comparison of the range of average daily hot water flow results per household from the EST data against the per-run model output in Figure 6.12(a) shows high similarity. The model results are based on the average ranked results from 250 annual duration simulations. However, as shown in 6.5.1, the range of average flow results per model run ranges from 90.1 to 125.6 litres/day (see also 'Min Run' and 'Max Run' on Figure 6.12(a)) suggesting that convergence is achieved only after a significant number of households are modelled. As with the overall population analysis in 6.5.1, further data would be required to assess if the variation at the household-level is realistic.

Using the same method introduced for electricity model analysis in 5.14.3, each EST household average demand is compared to the ranked results from 250 model runs and the model output rank of the closest match determined (see Figure 6.12(b)). A balanced distribution of model ranks (the target distribution) is indicative of a model that captures a representative range of behaviours. The actual distribution is broadly linear with some discrepancies at the extremes, particularly an inability to predict very low



(a) Model output comparison



(b) Model output rank per EST household

Figure 6.12. Hot water demand model output comparison from 250 runs with measured EST data. Data for the 'EST Data' distribution from [90].

demand levels. This confirms the electricity dataset analysis that determined that the composite nature of some calibration elements, particularly occupancy, underestimates the potential for very low demand households.

Table 6.5
Model closest cumulative match similarity analysis range results for the EST household equivalent model after 250 model runs.

Similarity	'High'	'Good'	'Some'	'Low'
'Timing'	41 (46%)	30 (33%)	11 (12%)	8 (9%)
'Overall'	25 (28%)	37 (41%)	19 (21%)	9 (10%)

6.5.5 Similarity Assessment

The PAA-ED method introduced in 5.14.4 was used to determine the similarity between individual household model outputs and the equivalent households in the EST dataset. The results shown in Table 6.5 indicate that the model is capable of closely matching actual average hot water use profiles, particularly the relative timing of use. This is a stronger measure of the underlying cycle model effectiveness, without the additional impact on similarity of different average flows captured by the 'overall' measure. After 250 runs only a small number of household profiles are not to some extent replicated by both the 'timing' and 'overall' distribution measures, confirming that the 'Simple' module basis used is effective for all but a small number of outlier households with distinct behaviours.

6.5.6 Diversity Assessment

For multi-household hot water system design a key assessment is demand diversity. This is the maximum total demand that can be expected at any time, typically expressed as a multiple of the maximum demand for a single dwelling. As the number of dwellings increases, the ratio of diversity to number of dwellings falls. This is principally used for the sizing of distribution pipework and is therefore required for both the primary network design and smaller subsets supplied via branch pipework.

As outlined in 1.6.1, a number of different hot water diversity standards have been developed, all of which are potentially limited by a lack of household characteristics differentiation. The principal diversity factor standard currently used for multi-household hot water systems is the Danish Standard DS439 [56]. The British Standard BS6700 [59] gives significantly higher values and has been assessed to significantly overestimate peak usage (Galluzzi, 2011). CIBSE have recently endorsed DS439 in preference to BS6700 for use in the UK [63]. Other developed standards include those shown in Figure 1.7, compared to which DS439 is generally more aggressive for smaller numbers of households and more conservative as the number approaches and exceeds 100.

It is not made clear from the published standards whether the stated diversity value is the absolute maximum instantaneous value or maximum demand averaged over a period of time, or whether infrequent short periods at higher flows can be tolerated. However, it is assumed that the timescale of interest is sub-minute. For this reason, the diversity analysis of both the EST data and model output is converted to 1-second resolution data by randomly allocating within-minute start times from the standard 1-minute resolution data. The duration of the cycle is determined based on selecting a flowrate randomly between 2.5 and 5 litres/min for <3 litre volume uses and between 5 and 12 litres/min for other uses. Data on flow per use is limited ([178], [179]), and a more accurate representation of flow per specific use would improve the accuracy of diversity assessment. The analysis then determines diversity values based on the absolute maximum on a 1-second basis, and the maximum values averaged over a 10-second and 60-second basis.

The per-household maximum flow basis for the analysis is based on the 37.5kW maximum demand per household assumption of DS439, which translates to an assumed maximum instantaneous hot water flow of 14.8 litres/min at the average cold-to-hot temperature difference of 37°C in the EST dataset.

As outlined, the EST survey measured hot water use at the tank outlet and therefore households with electric showers did not have this use measured. Therefore, the EST dataset cannot be used as a direct comparison for district heating (DH) design. DH-connected households will typically have all hot water use, including showers, provided by the DH system. Two sets of diversity analysis are therefore required: one to compare the EST dataset diversity with the equivalent model basis, with some households simulated with an unmeasured electric shower, for validation; and the other to compare the diversity standards and the model output for households where all shower use is via the main hot water system.

The 90 households in the EST dataset with consistent extended duration data were combined as an effective energy system to be used as a validation dataset for the diversity values predicted by the model. For the analysis the 10-minute EST flow data is converted to an effective duration in seconds in the same manner as the 1-minute resolution conversion (see above) and randomly positioned within the 10-minute period. The diversity is then analysed on a 1-second, and rolling 10-second and 60-second basis calculated from the total flowrate for all households divided by 14.8 to determine the overall diversity relative to the nominal single household maximum flow. Different timescales are used to confirm model performance under different conditions and to assess the influence of timescale on diversity.

The analysis was carried out for randomly selected sets of 10, 25, 50, and 75 households, and the full 90-household dataset. For each household set number, the process was repeated 1000 times and for the larger subsets (50, 75 and 90 households) each week of data was randomly sequenced within seasonal blocks as the potential household set variation in these cases is limited. The random selection of households was not restricted based either on house or household size, therefore the results for both the EST analysis and model output include a significant range of average house sizes and occupant numbers to ensure the data is representative of a range of potential scenarios.

Table 6.6
EST dataset and equivalent model output diversity comparison.

Households	10	25	50	75	90
EST Avg. (1s)	3.31	4.65	6.10	7.32	7.87
Model Avg. (1s)	3.49	4.72	6.07	7.22	7.75
EST Max. (1s)	5.10	7.23	8.47	10.03	10.81
Model Max. (1s)	5.09	6.83	8.63	10.47	10.97
EST Avg. (10s)	3.11	4.34	5.69	6.83	7.34
Model Avg. (10s)	3.31	4.46	5.70	6.76	7.25
EST Max. (10s)	4.84	6.74	7.54	9.22	10.45
Model Max. (10s)	4.72	6.76	8.04	9.86	10.03
EST Avg. (60s)	2.54	3.50	4.56	5.46	5.91
Model Avg. (60s)	2.51	3.44	4.43	5.28	5.70
EST Max. (60s)	3.98	5.69	6.35	7.70	7.86
Model Max. (60s)	4.05	5.10	6.58	7.58	7.99

The developed model was used to generate equivalent results to the EST dataset (i.e. annual data, identical household sizes). The results are shown in Table 6.6. Repeated 1000-run attempts with a fixed dataset showed that the per-attempt variation could be ±5%, therefore the primary aim of the analysis was no consistent under- or overestimation as opposed to exact replication. On this basis, the developed model shows good accuracy is replicating the characteristics of the dataset. The model therefore provides a means to test diversity standards against the potential variations in use based on household characteristics and unique behaviours. The results also show that the time basis of the analysis is significant.

Equivalent data was then generated for sets of households representing a typical district heating system, with all hot water, including showers, supplied from the system. The total number of people per set of households was restricted to ±10% of the UK average basis of 2.3 people per household, therefore the maximum values are for an average community and do not represent absolute outlier values. The impact of people number and house size is discussed further in Chapter 8.

Table 6.7
Comparison between the 'District Heating' equivalent model output and DS439 diversity estimation. 'DS439' data from [56].

Households	10	25	50	75	100
DS439	2.37	3.76	5.59	7.17	8.64
Avg. (1s)	3.59	4.93	6.44	7.67	8.52
Max. (1s)	5.28	6.63	8.95	9.92	11.65
Avg. (10s)	3.41	4.66	6.08	7.21	8.01
Max. (10s)	4.80	6.63	8.45	9.74	11.22
Avg. (60s)	2.65	3.65	4.80	5.74	6.43
Max. (60s)	3.94	5.36	6.72	7.76	9.53

The results (see Table 6.7) suggest that the DS439 basis has the potential to underestimate diversity in comparison with the average model output to at least 100 households for both the 1s and 10s average bases, with the maximum predicted values significantly in excess. For small sets of households, the underestimate is pronounced, with even the 60s average basis of the model output exceeding the 10-household DS439 prediction. As a minimum the results suggest that the use of DS439 values without understanding the critical timescales for the dynamics of the proposed system could result in poor design.

The impact of household characteristics and system dynamics on diversity prediction is reviewed further in Chapter 8.

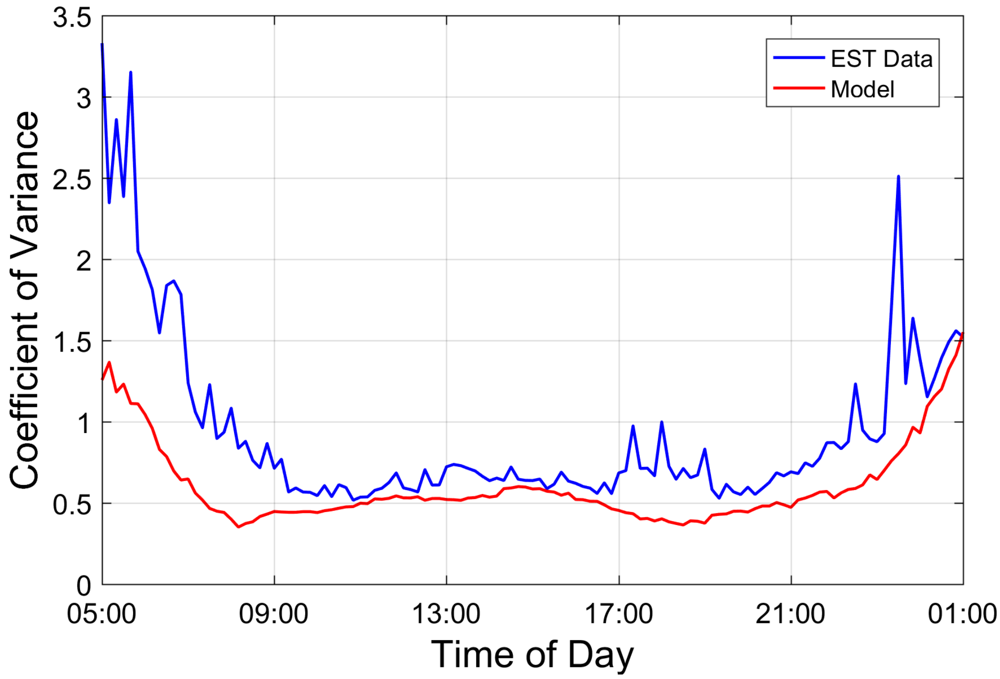


Figure 6.13. Per-timestep coefficient of variance of mean-normalised per-household hot water demand comparison for measured EST data and equivalent modelled results. Data for the 'EST Data' distribution from [90].

6.5.7 Hot Water Validation Analysis Summary

With only 107 households in the dataset and limited household characteristics information, the initial assessment was that there was sufficient data to determine when hot water is typically used but that significant differentiation based on household type would be not be possible. The primary aim of the validation was therefore to determine if the probabilistic methods developed for the electricity demand model could be effectively used for hot water modelling, pending the availability of more comprehensive data.

The model has been shown to replicate the range of average and time-dependent demands from the calibration dataset. The model is also able to capture the time dependency of both use in different ranges of total cycle volume and of overall demand in households with and without children. Given the lack of high-resolution models for hot water demand modelling, the model should provide a useful basis for assessment of larger district heating networks, where any errors associated with limited differentiation would be reduced. As with the electricity demand model, some household demand profiles were not closely replicated and this is discussed below.

6.5.7.1 Residual Behaviour Averaging

In a similar manner to 5.14.7 for the electricity model, the hot water demand model output was analysed to determine if there was evidence of behaviour averaging as a result of the calibration methods used. Figure 6.13 shows that there is a similar underestimate of the level of per-timestep variance in demand, potentially again because of the composite calibration approach for certain aspects of the occupancy and cycle start time identification sub-models. This is also reviewed further in Chapter 7 for hot water use.

6.6 Chapter Summary

This chapter detailed the development process for the hot water demand sub-model and the validation analysis undertaken to assess the model performance. The chapter highlights are as follows:

- Similar methods used for hot water cycle modelling to those developed for frequently used ('Simple') electrical appliances.
- Six cycle volume ranges (from 0-1 litres to 30+ litres) with distinct use behaviours identified from dataset analysis, with each range calibrated and modelled separately.
- Validation confirmed close replication using the developed model of the overall and volume range-specific time-dependent demand. Replication at the household size level was less conclusive but discrepancies could be explained by unrepresentative data for some household sizes.
- Diversity analysis determined that the model output had the same characteristics as the input EST data. Further assessment of the model output against current diversity standards identified a potential for underestimation that is investigated further in Chapter 8.

Chapter 7

Occupant- and Household-Specific Behaviour Modelling

7.1 Chapter Overview

In Chapter 4 for the occupancy sub-model and Chapter 5 and 6 for the demand sub-models, it was shown that elements of each have a tendency for convergence to the calibration data average basis and, consequently, poor replication of individual behaviours. The convergence is most pronounced for elements that use composite data from multiple individuals and households for calibration. For the occupancy sub-model, the convergence is discernible in the output for each occupant-type group, in terms of both transition timings and average occupancy levels. For the demand sub-models, the convergence is related to the timing of use events for intermittent demands. In this chapter, the speed of convergence is shown to be significant for the aims of the developed overall demand model.

Methods to further differentiate behaviours have been developed and are detailed in this chapter. For the occupancy sub-model, the time-basis of the Markov chain models is altered for each individual to account for behaviour variation. For the demand sub-models, relative use timing for individual appliances and hot water events are manipulated for each household based on the dispersion of behaviours seen in actual data. In each case the performance is compared with the unmodified output.

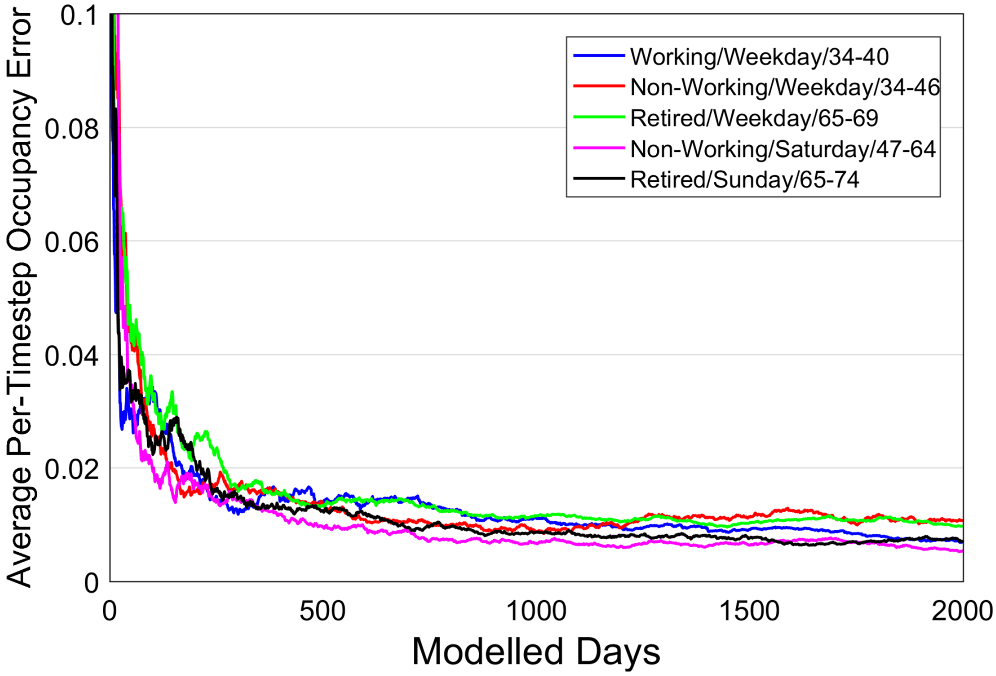


Figure 7.1. Occupancy sub-model convergence - average per-timestep 'error' to calibration dataset average by number of modelled days.

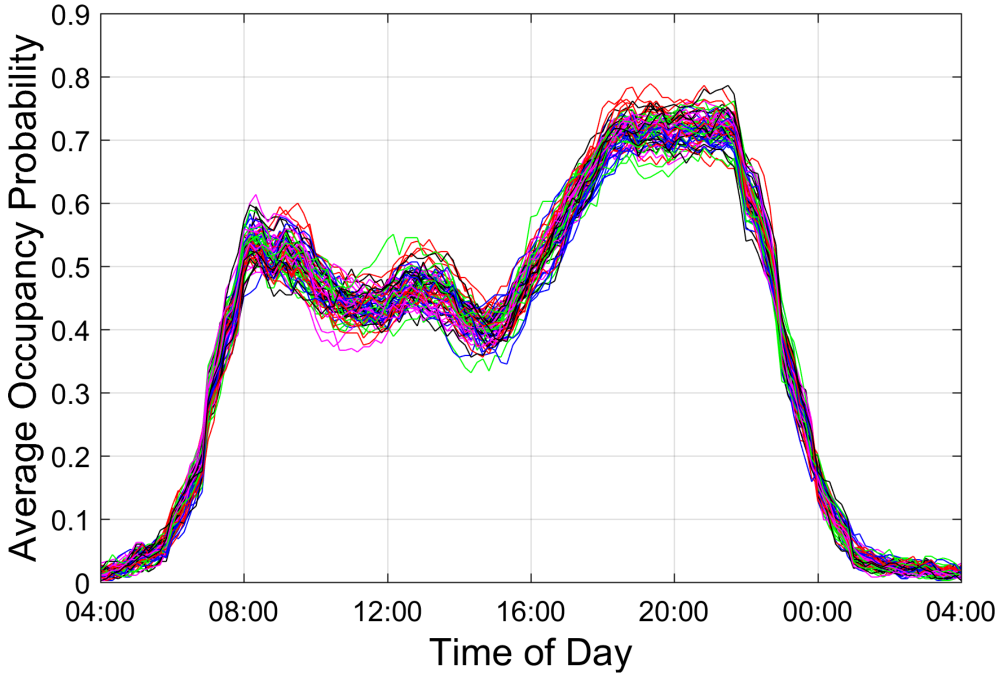
7.2 Model Output Convergence

The overall conclusion from analysis of the developed electricity and hot water demand sub-models detailed in Chapters 5 and 6 was that they are effective for analysis of typical behaviours and for low-resolution average demand analysis but there remains evidence of lower time-dependent variation between individual households than seen in the measured data (see Figures 5.43 and 6.13). Weaker model performance in predicting low energy use households also indicates that the tendency to underestimate the range of average occupancies within each occupant type module (see Figure 4.12) may also result in an underestimate of the range of predicted demand levels.

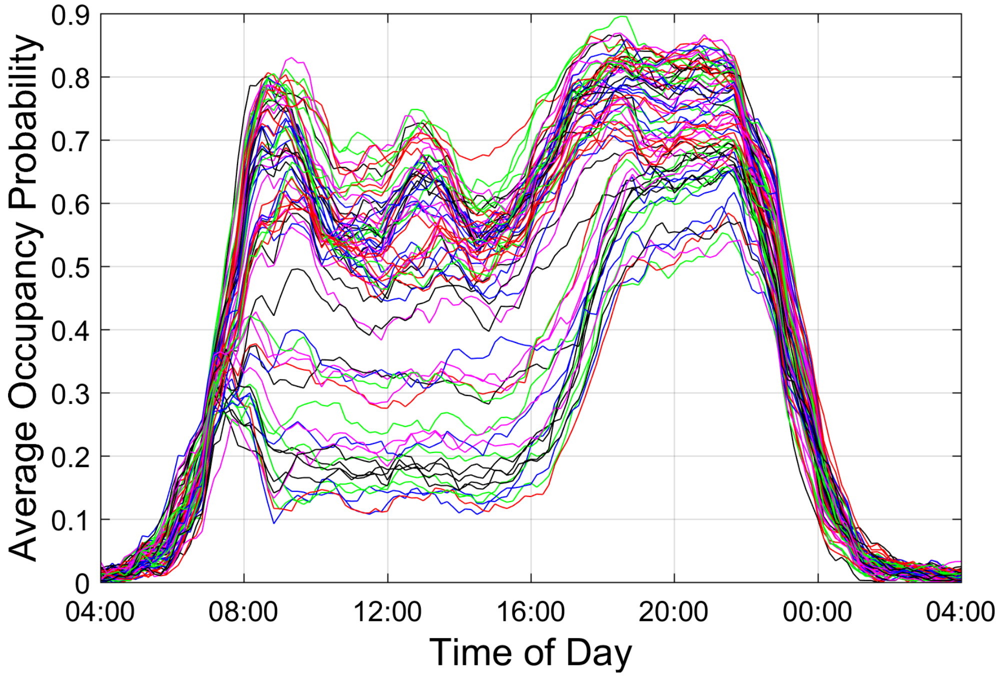
Whilst the overall model has been developed with a number of differentiated calibration datasets and incorporates probabilistically generated factors to account for variations due to household characteristics and individual behaviours, two key time-dependent elements retain calibration data that is a composite of multiple behaviours. They are:

- Occupancy Sub-Model - Each occupant type-specific Markov chain calibration module is generated from the single-day diaries of a large number of individuals (see 4.5.5). The average behaviour for each module output will therefore converge to the average behaviour of the calibration population; an inherent feature of all Markov chain models. Significant differentiation to account for the behaviour differences between occupant types has been incorporated, but each type module remains a composite. Analysis has shown that each module converges rapidly to the type average over approximately 200-300 modelled days as shown in Figure 7.1 for selected examples. The result is that for annual duration output there will be convergence in average occupancy for occupants with similar characteristics. This was shown in Figure 4.12 to be particularly significant for retired occupants, who do not have the variable mix of working and non-working days associated with a large proportion of the working age populations to generate additional variation.

- Cycle Timing Module - The cycle start time identification module (see 5.9) is calibrated using probability distributions generated from the overall population with



(a) 'Richardson' model



(b) 'Composite' model

Figure 7.2. Occupancy sub-model average per-household active occupancy results for 70 single-person households for annual duration model runs.

some further manipulation for specific average occupancy variations per household type. These, however, do not account for any household-specific behaviours related to the time-dependent use of electrical appliances and hot water.

The developed overall demand model that incorporates these composite calibration elements is hereafter known as the 'composite' model. Further model developments to capture individual behaviours detailed in this Chapter are identified by the term 'individualised'.

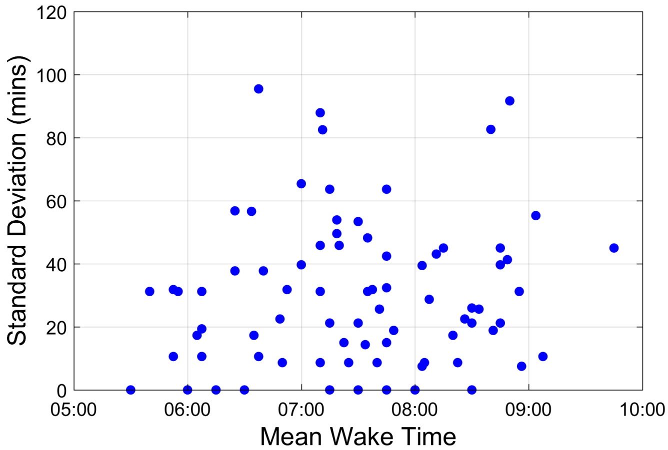
7.2.1 Occupancy Sub-Model Uniformity

Analysis of the output of the developed 'composite' occupancy model (see Chapter 4) and the Richardson et al model [69] (hereafter known as the 'Richardson' model), highlights the convergence problem with group-calibrated probability models. Figure 7.2 shows the average annual occupancy results for 70 single-person households for both methods. The 'composite' model output is based on a nationally representative set of single-person household characteristics. The 'Richardson' model output is based on 70 runs of the single-person household module.

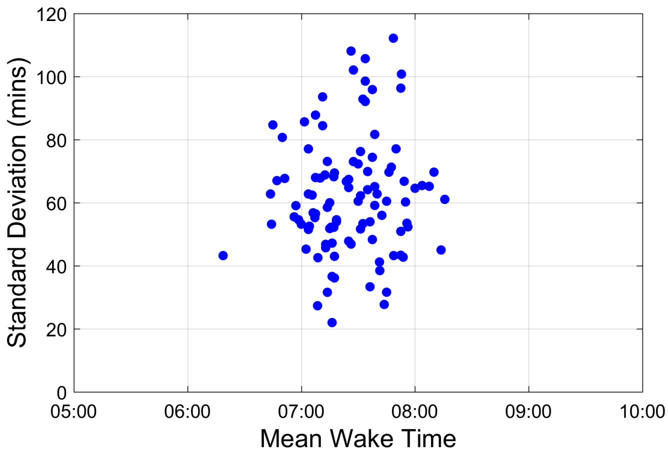
As outlined, for extended period models, predicted occupancy, and therefore to a degree any demand prediction using this output as a basis, converges to the calibration population mean. For the 'Richardson' model, differentiated only by occupant number, all one-person households converge to the same basic occupancy pattern over an annual model run with only minor variation in both average occupancy and transition timings. For the 'composite' model, even with additional differentiation for age and employment status and a representative set of occupants which partially captures realistic variations in average occupancy (vertical variation), the average timing of key transition periods (waking, retiring etc.) (horizontal variation) remains unrealistically consistent.

Analysis of the Dutch 2005 TBO Time-Use Survey (TUS) [87], which includes seven-day diaries, allows this tendency for convergence to be further assessed. It is probable that one-week diaries will exhibit more individual variation than would be expected for longer periods of analysis, but the dataset allows for an initial judgement of individual behaviour variation and model replication of this behaviour.

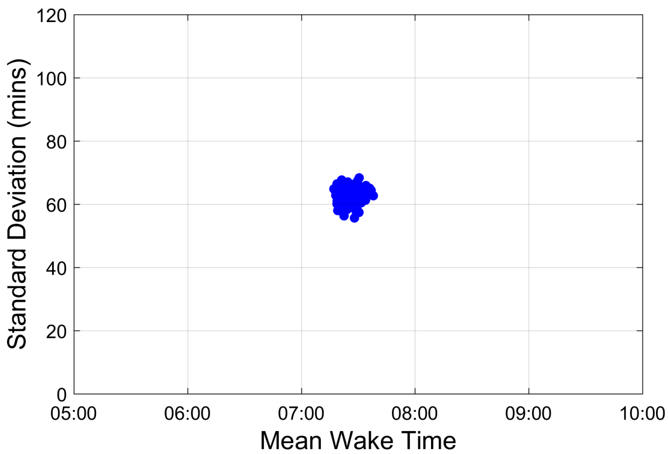
Figure 7.3 shows the waking time mean and standard deviation (in minutes) distri-



(a) Dutch TUS



(b) 'Composite' model - 1 Week



(c) 'Composite' model' - 1 Year

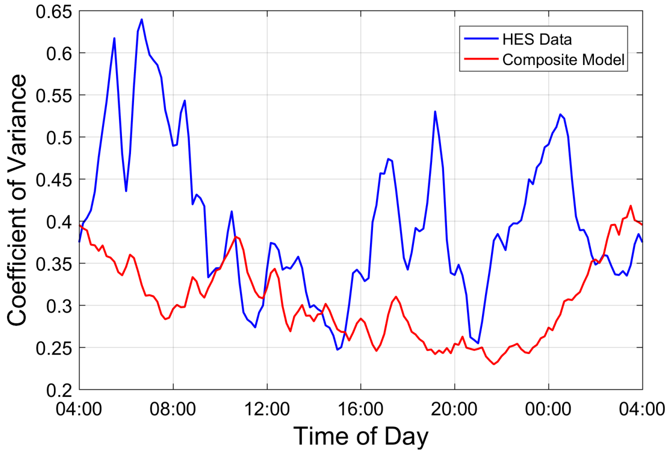
Figure 7.3. Mean and standard deviation of wake time per occupant for single-person, retired households for the Dutch 2005 TBO TUS dataset [87] and 'composite' model output.

bution for retired single-person households from the 15-minute resolution Dutch TBO TUS data and the equivalent for the 10-minute resolution 'composite' model based on 100 one-week and one-year duration runs. The TUS data shows a range of behaviours, including zero standard deviation results. (A zero result implies that the person transitioned within the same 15-minute period on all monitored days.) The equivalent occupancy model results show a more random distribution centred on the average behaviour for the one-week duration model and a highly converged distribution for the one-year duration model, both with a higher average standard deviation than the TUS data. This highlights that the 'composite' model output is highly variable for each individual modelled event but highly convergent overall in comparison with real behaviours. This short-term variation and long-term convergence limits the ability of the model to capture occupancy behaviours of individual households.

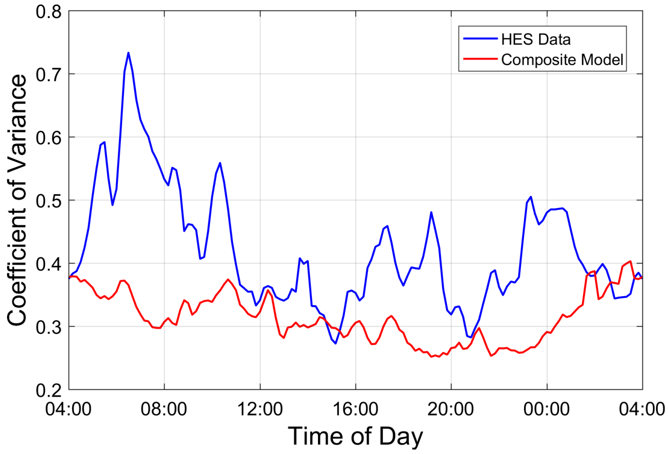
The influence of the 'composite' occupancy sub-model convergence on demand prediction was shown in 5.14.7 and 6.5.7, where the demand sub-model per-timestep relative variance is shown to be significantly lower than for the Household Electricity Survey (HES) [89] and Energy Savings Trust (EST) hot water [90] datasets respectively. Figure 7.4 shows the same electricity demand analysis as Figure 5.43 for three distinct type groups (single-working-age, couple-retired, and multi-adult households), with a higher coefficient of variance indicating a greater divergence in energy use timing. The HES dataset results for each group type are similar, with distinct peaks in variance in the waking period (6am to 10am), late afternoon (4am to 8am), and late evening (10pm to 1am) associated with the main transitional periods of electricity demand. In each case the demand sub-model output again shows a significantly more consistent and lower level of variance, which would be expected for a model that underestimates the timing variance in occupancy and appliance-use behaviours in these periods.

7.2.2 Individual Household Demand Behaviours

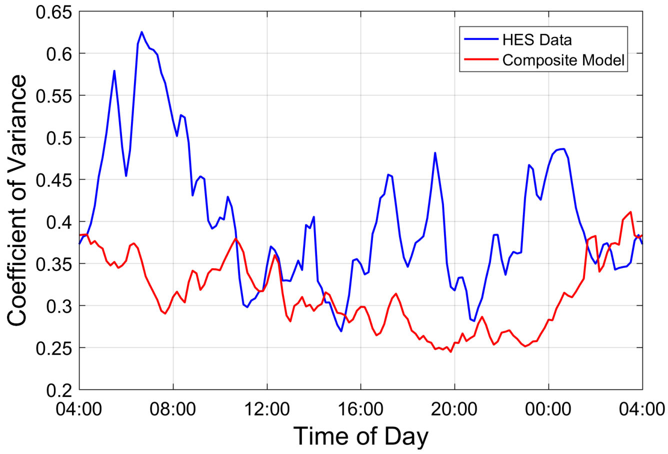
The cycle start time cumulative probability function (cdf) distributions (see Figure 5.16) used within the cycle start time identification module (see 5.9) do not account for any strongly habitual usage behaviours or less distinct household-specific tendencies to use certain appliances or hot water in specific periods. As was shown in Figure 5.39



(a) Single-person, working age



(b) Couple, retired



(c) Multi-adult

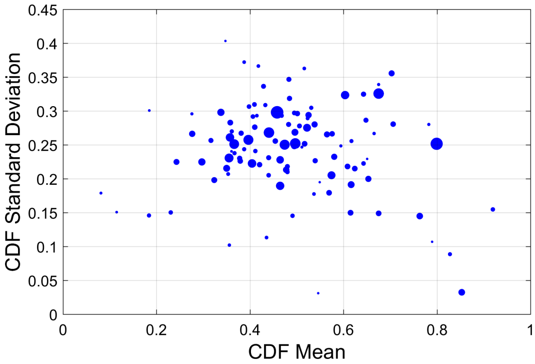
Figure 7.4. Per-timestep coefficient of variance of mean-normalised per-household electricity demand for measured HES data and equivalent modelled results for three household types. Data for the 'HES Data' distribution from [89].

for the washing machine module, the start time *cdf* distributions generated for each appliance allow an assessment to be made of the range of cycle start time *cdf* values for each HES dataset household. The mean and standard deviation of the *cdf* distribution per household gives an indication of the relative timing of usage compared to the average behaviour. For example, a household mean start time *cdf* value significantly lower than 0.5 indicates use that is typically earlier than average. The lower the standard deviation, the more consistent the timing of each use.

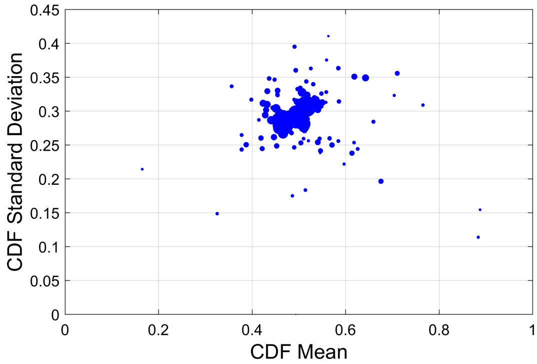
Analysis of the data has shown that significant use timing differences exist for most 'cyclic' demands. However, in terms of both likelihood and overall impact on demand, is most significant for the following; kettles, washing machines, dishwashers, cookers/ovens, and the higher volume (>15 litres) hot water events (i.e. those predominantly indicating bath and shower use). Assuming average behaviour should therefore, to some degree, impact model accuracy. This was confirmed by comparative analysis detailed in 7.4.2.1 and 7.5.1.2.

Figure 7.5 (a), (c), and (e) shows the range of household *cdf* mean and standard deviation values for dishwasher and cooker cycles for all HES dataset households and 'Very High' (30 litres+) hot water cycles for all EST dataset households, indicating a similar wide range of different behaviours as already shown for washing machine use in Figure 5.39. For these distributions, each data point is also only weakly correlated with the number of observed events (indicated by the size of each data point). Figure 7.5 (b), (d) and (f) show the equivalent distributions for households modelled using the 'composite' model, indicating a strong tendency for convergence to the average behaviour. The results also demonstrate that the convergence tends to increase with the number of modelled events, as would be expected for a highly probabilistic model exhibiting an overall tendency to converge. The 'composite' model behaviour is consistent for all the specific demands listed above.

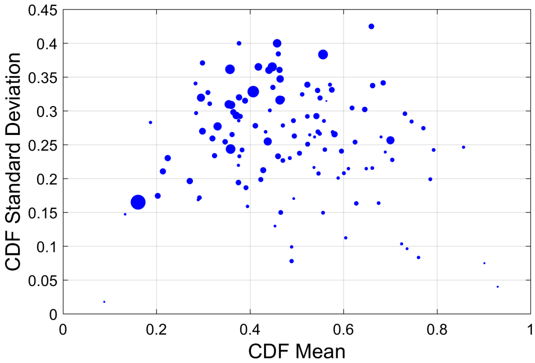
In contrast, sub-1-litre volume hot water cycles, which are likely to be the most occupancy rather than behaviour driven demand, exhibit a significantly lower dispersion of timing behaviour as shown in Figure 7.6(a). The 'composite' model equivalent output shown in Figure 7.6(b) is more converged but not as significantly. As will be reviewed below, a realistic dispersion can be achieved solely with improvements to the occupancy sub-model, confirming the low behavioural influence on timing for certain demands.



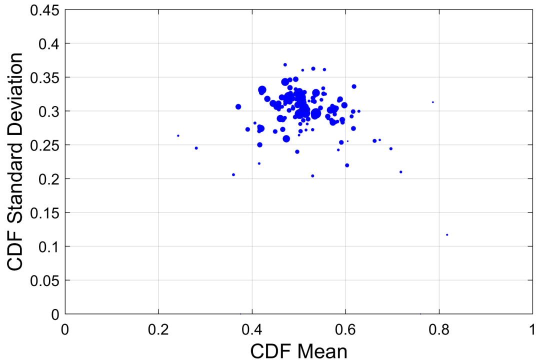
(a) Dishwasher - HES households



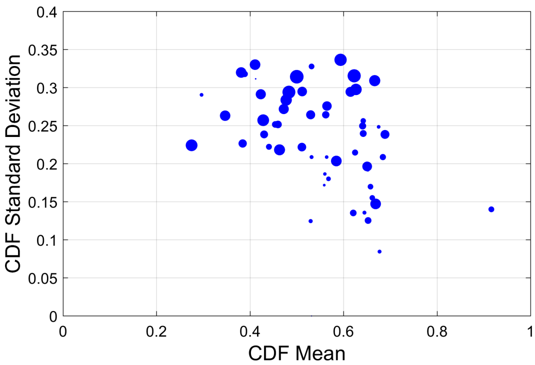
(b) Dishwasher - ‘composite’ model



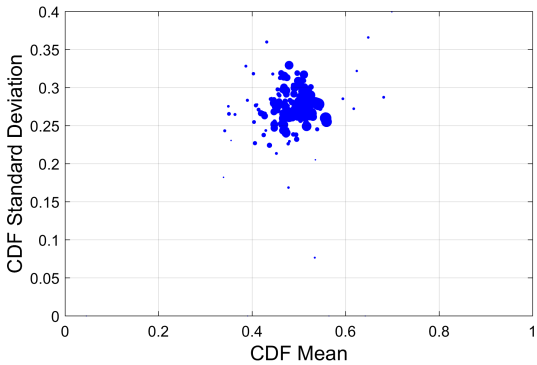
(c) Cooker - HES households



(d) Cooker - ‘composite’ model



(e) ‘Very High’ hot water - EST households



(f) ‘Very High’ hot water - ‘composite’ mdl

Figure 7.5. Cycle start time cumulative probability function mean and standard deviation per household. Data for the ‘HES households’ distributions from [89] and for the ‘EST households’ distribution from [90].

7.3 Individualised Occupancy Sub-Model

Two specific convergence problems with the composite-calibrated occupancy model basis were identified. One related to the timing of specific occupancy transitions and the other to the prediction of average occupancy within each calibration group. Each has been analysed separately.

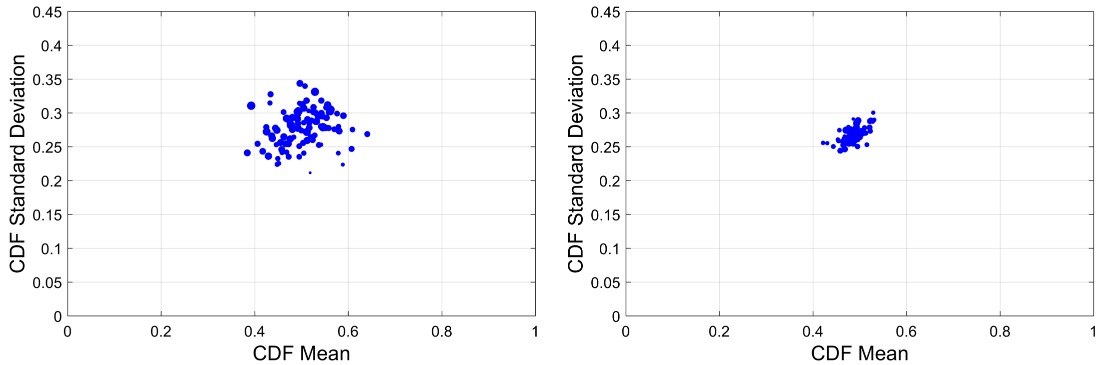
7.3.1 Transition Timing Prediction

Household demand profiles are characterised by four key occupancy-linked transitions; waking, morning leaving, afternoon/evening return and sleep. The developed group-calibrated ('composite') occupancy model has been shown in 7.2.1 to tend to produce overly convergent individual occupancy profiles based on the average behaviour of the calibration group, and for those critical transitions. Therefore, the previously developed occupancy model was modified with a focus on those periods to better capture individual variance from average occupancy behaviour.

The Dutch 2005 TBO TUS dataset, which includes seven-day diaries, has been used to calibrate the model for individual behaviours relative to the average behaviour for each defined calibration group. It is assumed that the variability in this dataset, if not the specific timings, are representative of any developed country population. It is acknowledged, as outlined above, that one-week diaries may not be sufficient to accurately capture long-term behaviours, and that equivalent UK-specific data over a longer period (minimum of one month) would significantly enhance the proposed method. However, the available data does at least allow the method's potential effectiveness to be assessed.

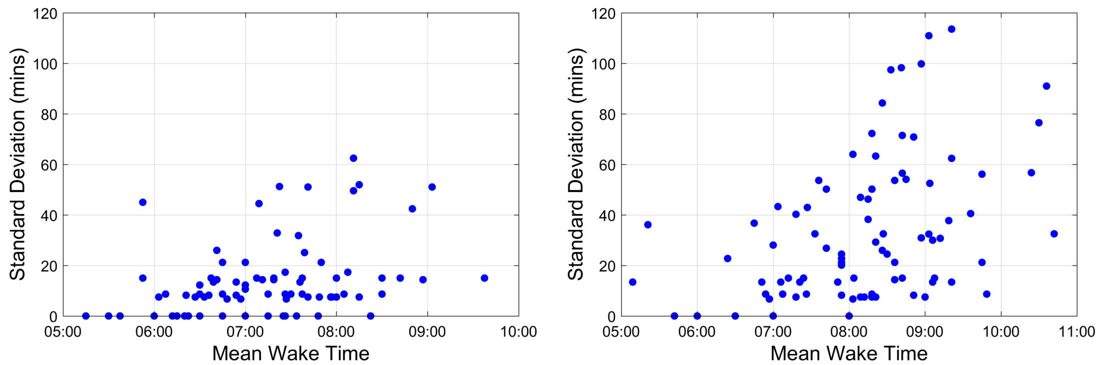
Analysis of the Dutch TBO TUS data has shown that transition time behaviour within each household-type varies significantly between respondents. Some have clear patterns of behaviour; others are more erratic. This was shown for the retired single-person population in Figure 7.3(a), and Figure 7.7 shows the equivalent waking time results for each working age, single-person householder on working and non-working days, which have similarly variable distributions. This variation is replicated across all populations for both wake and sleep transition timing.

For all populations there are clear increases in waking at each hour and smaller peaks at the half-hours as indicated in Figure 7.8 for the single-person households in



(a) EST households **(b)** ‘Composite’ model

Figure 7.6. Cycle start time cumulative probability function mean and standard deviation for ‘Very Low’ volume hot water use per household. Data for the ‘EST households’ distribution from [90].



(a) Working **(b)** Non-working

Figure 7.7. Wake time mean and standard deviation for single-person, working-age householders from the Dutch 2005 TBO TUS dataset [87].

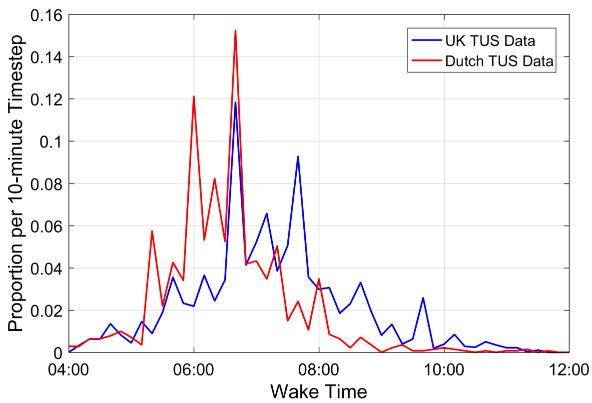


Figure 7.8. Proportion of weekday wake times per 10-minute timestep from the UK 2000 [83] and Dutch 2005 TBO TUS [87] dataset single-person householders.

the UK and Dutch TBO TUS datasets. This suggests that for a significant proportion of households the timing of the first person awake is forced by outside means (i.e. alarms) and is not naturally driven. For occupants with a lower likelihood of a specific outside driver for waking (retired, non-working), the most prevalent wake time remains on the hour, suggesting that this is default waking behaviour for a large proportion of people. Similar, if less distinct, patterns can be seen for the other analysed occupancy transitions.

To capture this behavioural variation, the 'composite' model Markov chain probability matrices are assumed to provide two levels of behavioural information. Existing models rely on the time-specific detail to probabilistically generate statistically consistent stochastic models. The proposed method assumes there is also a higher-level detail inherent in Markov chain occupancy models that captures typical transition patterns that are period- rather than time-specific, and that, for each individual, the specific timing can vary. This potentially allows the 'composite' model to be further manipulated without impacting the overall statistical basis or introducing unrealistic individual occupancy patterns. To this end, the time basis of each occupant type probability model has been altered to reflect the difference between individual behaviours and the group-calibrated average.

7.3.1.1 Method Development

The following section outlines the development of each element in the 'individualised' occupancy model. Figure 7.9 shows the overall process graphically and the individual manipulations required.

Wake time – Dutch TUS analysis shows that there is no significant difference in the average variance per person for wake and sleep times. Wake time was therefore arbitrarily selected as the anchoring statistic for the revised model. Each individual is allocated an average wake time based on the probability distribution for the equivalent UK TUS population. It is assumed that some of the early and late times are outliers for those individuals and comparison of the one-week Dutch data and single-day UK TUS data suggests a 5-10% greater range in the single-day UK data, therefore the potential average times are restricted to those in the middle 90% of the range.

As outlined, wake times, in particular, are proportionally higher in the period fol-

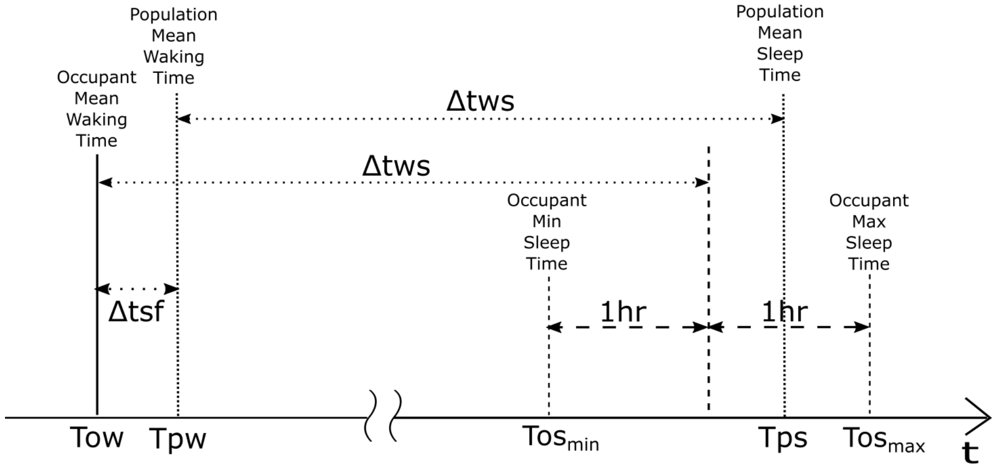


Figure 7.9. Graphical representation of process to convert TUS average wake and sleep times to individual-specific equivalents.

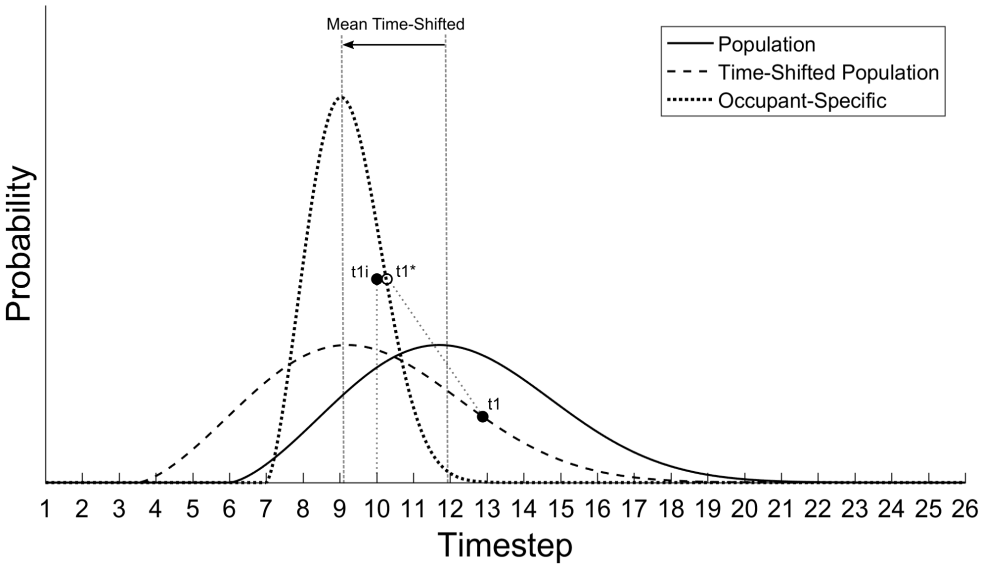


Figure 7.10. Graphical representation of a calibration population distribution conversion to individual time-shifted basis with lower variance.

lowing the hour or half-hour. By allocating an average wake time based on this distribution, the updated model should broadly maintain this pattern, which has been confirmed (see Figure 7.12).

Wake period time-shifting – The 'wake period' is defined as 3am to 10am. For each group calibration module, the average wake time is determined. For each individual, the Markov chain module in this time period is then time-shifted (Δtsf) based on the difference between the individual's determined average wake time and the relevant group average (Tow and Tpw respectively on Figure 7.9).

Sleep period time-shifting – Analysis of the Dutch TUS data shows an average waking duration of 17.25hrs (Δtws), with the majority of individuals being linearly distributed by +/- 1hr of this level. This matches closely with the average waking duration within the unmodified 'composite' model. The sleep-transition period Markov chain module (9pm to 3am) is therefore time-shifted by a randomly selected amount that is +/- 1hr of the wake period time-shift for the individual (i.e. between Tos_{min} and Tos_{max} in Figure 7.9).

Variance factor – The time-shifted probability modules without further modification maintain the overall variance of the composite behaviour, which has been shown to significantly exceed individual variance for almost all individuals. For example, the single-person working day group modules have an average waking time standard deviation of 60.3 minutes and for the retired single-person modules of 73.6 mins. The equivalent averages for each individual in the Dutch TUS dataset are 28.0 mins and 20.6 mins respectively, with only 8% and 4% of each population exceeding the group module variance.

To achieve realistic individual behaviour variance, each modelled individual is probabilistically allocated a standard deviation for each defined transition from the equivalent Dutch TUS dataset occupant-type distribution. The Markov chain module timestep basis is altered based on the difference between the individual-specific standard deviation and the calibration group average. A minimum standard deviation of 2.5 minutes is arbitrarily used to ensure some variance per individual over extended periods within the 10-minute resolution basis (i.e. c.4% of transition events will be in the period preceding or following the assigned average period).

Figure 7.10 demonstrates the rebasing method graphically. The required manipula-

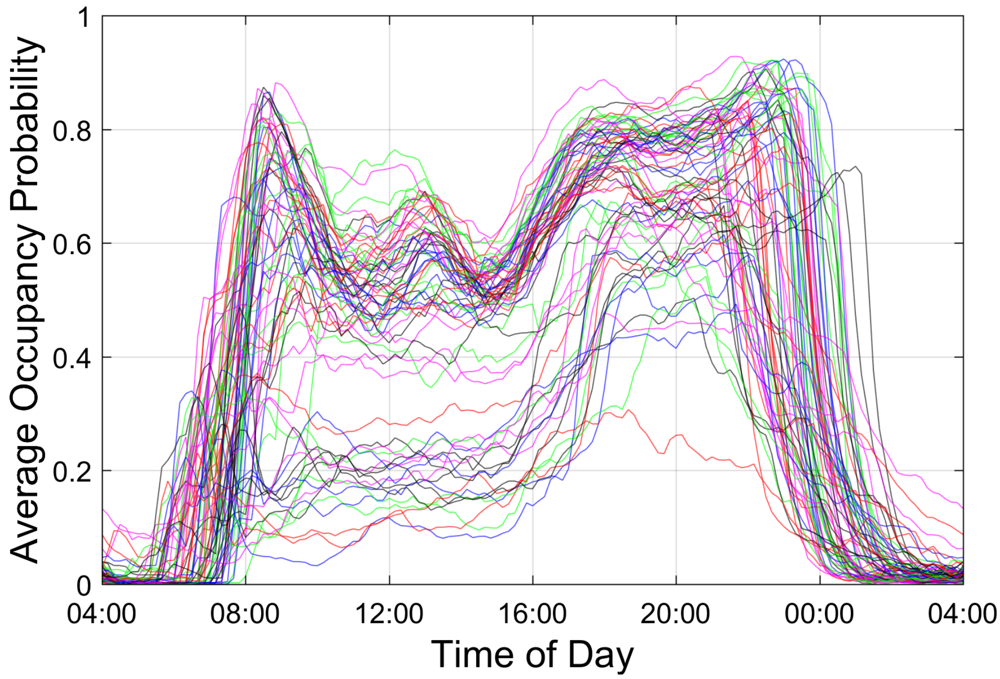


Figure 7.11. Occupancy sub-model average per-household active occupancy results for 70 single-person households for annual duration model runs for the 'individualised' model basis.

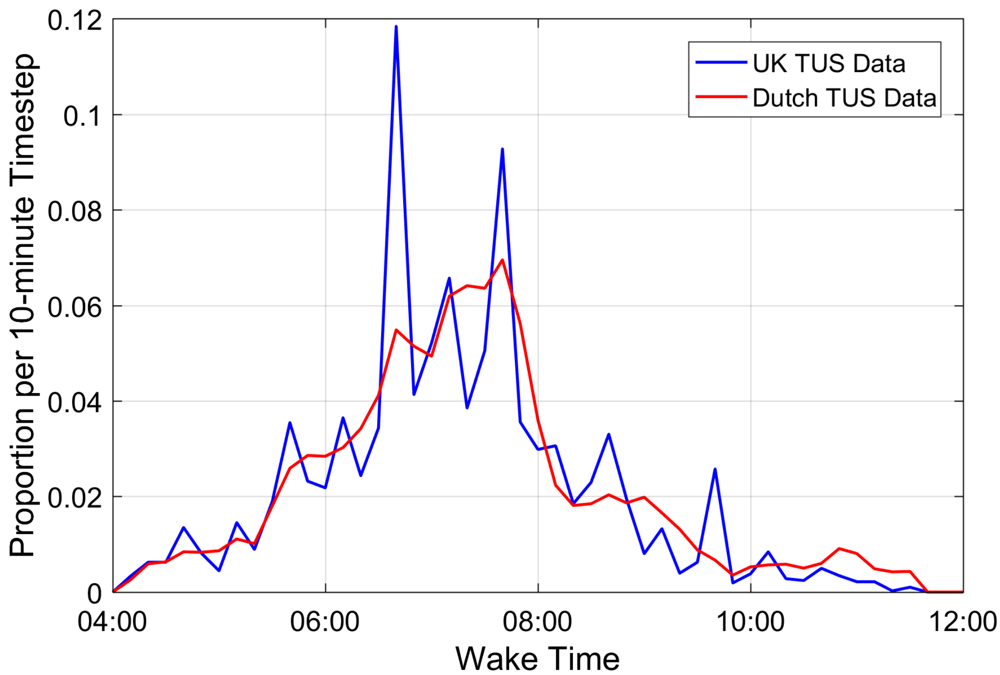


Figure 7.12. Wake time comparison between UK 2000 TUS dataset [89] and 'individualised' occupancy model output for all single-person householders.

tion is achieved by first running the time-shifted Markov chain module normally until a relevant transition is identified. (If no transition is identified the model proceeds as normal to allow for the probability of uncharacteristic behaviours.) The cumulative distribution function (cdf) value for the transition point (*t1*) is determined and the equivalent *cdf* value (*t1**) on the individual-specific variance distribution identified (Figure 7.10 shows an example of an individual with lower variance than the calibration population). The model timestep is then reset to the closest integer timestep (*t1i*) to the identified point *t1**. If *t1i* is before *t1*, the model deletes all modelled timesteps after *t1i* and resets the model timestep to *t1i*. If after, the unmodelled timesteps up to *t1i* are set to the preceding state and the model continues from *t1i*.

Other Key Transition Periods – The same process is also used for the key morning leaving and evening returning transitions, if such a transition is predicted within defined periods (morning leaving - from waking until 2.5 hrs after mean waking time; evening return - $\pm$ 1.5hrs of mean return time).

Average occupancy results using the 'individualised' model basis shown in Figure 7.11 for the same population as the unmodified 'composite' model results shown in Figure 7.2 shows significantly more occupancy variation in the key transition periods. The lack of long-term occupancy data does not allow for a direct comparison between the model output and individual occupant behaviour over an extended period. However, a statistical comparison between the TUS data and the occupancy model output for each occupant-type group is possible and is detailed in the following section.

7.3.1.2 TUS Dataset Replication

To be effective the 'individualised' occupancy method should replicate the timing behaviours of each equivalent TUS population for the identified transitions. Results for the waking period (see Figure 7.12) for all single-person households shows good correlation between the UK TUS distribution and model results. A smoother distribution is to be expected for the model results as they include significantly more data points (25550 vs 1159). Similar correlations are observed for other transitions and occupant-type groups.

The 'individualised' model has been calibrated to reflect the average waking times from the UK TUS dataset, therefore exact replication of the Dutch 2005 TBO TUS

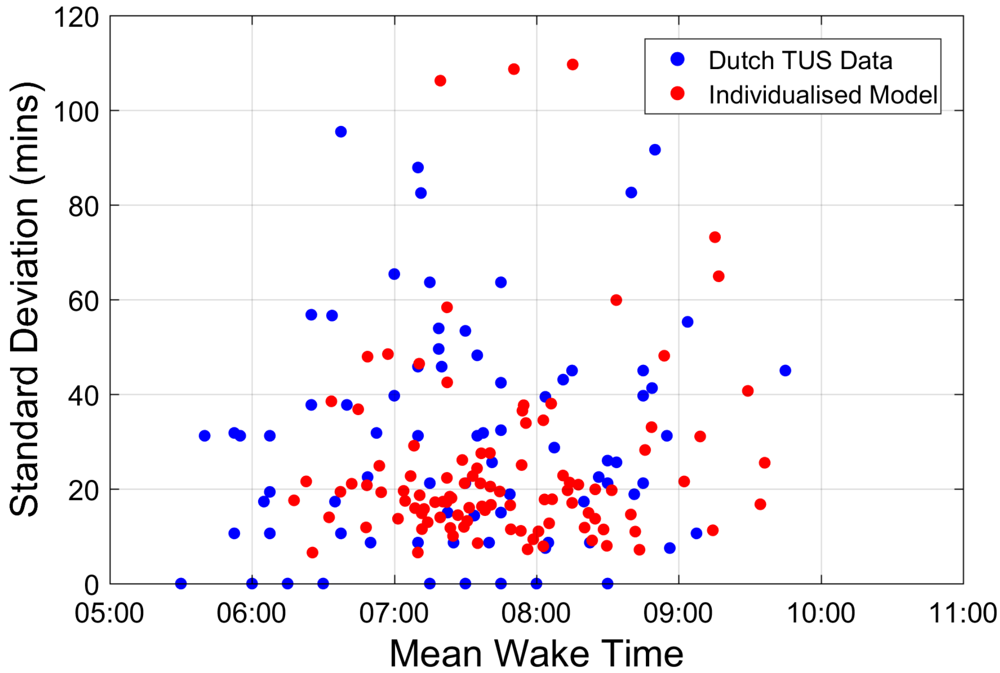


Figure 7.13. Wake time statistical comparison between the Dutch 2005 TBO TUS dataset [87] and 'individualised' occupancy model output for single-person, retired householders.

distributions shown in Figure 7.7 is not expected. However, the results for 100 annual-duration retired household models using the 'individualised' model shown in Figure 7.13 highlights the improvement in replicating the variance in behaviour compared to the tight convergence shown in Figure 7.3(c). The model distribution matches the characteristics of the Dutch TUS data equivalent, with a slightly later average reflecting the overall behaviour difference shown in Figure 7.8.

Further validation of the 'individualised' method was undertaken using the metrics defined in 4.4. This was done for three single-person weekday household model groups (working 34-40 age range, non-working 34-46 age range, and 70-79 age range). The results are shown in Table 7.1.

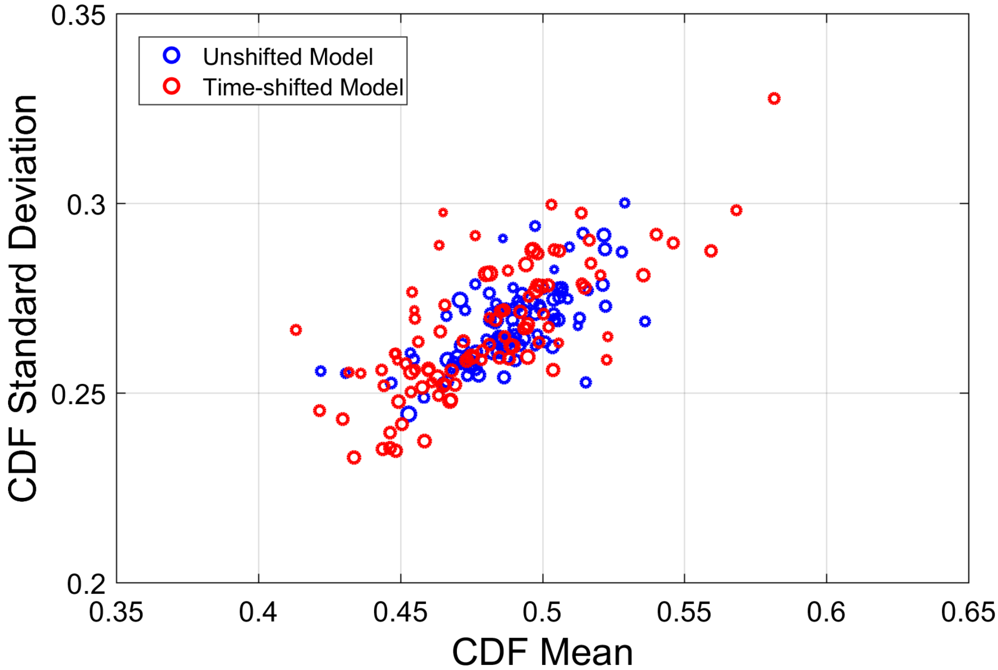
Table 7.1
Occupancy model validation metric comparison for three single-person householder populations for the 'composite' and 'individualised' models.

	AO_Conv (x E-3)	DurDist Sleep	DurDist Active	DurDist Out
Working 34-40 - 'Composite'	12.7	1.51	0.67	2.39
Working 34-40 - 'Individualised'	31.4	3.84	1.57	5.46
Non-working 34-46 - 'Composite'	16.1	1.30	0.86	1.76
Non-working 34-46 - 'Individualised'	37.6	3.47	1.54	5.41
Retired 70-79 - 'Composite'	3.0	0.98	0.34	3.44
Retired 70-79 - 'Individualised'	28.7	3.16	0.47	6.00

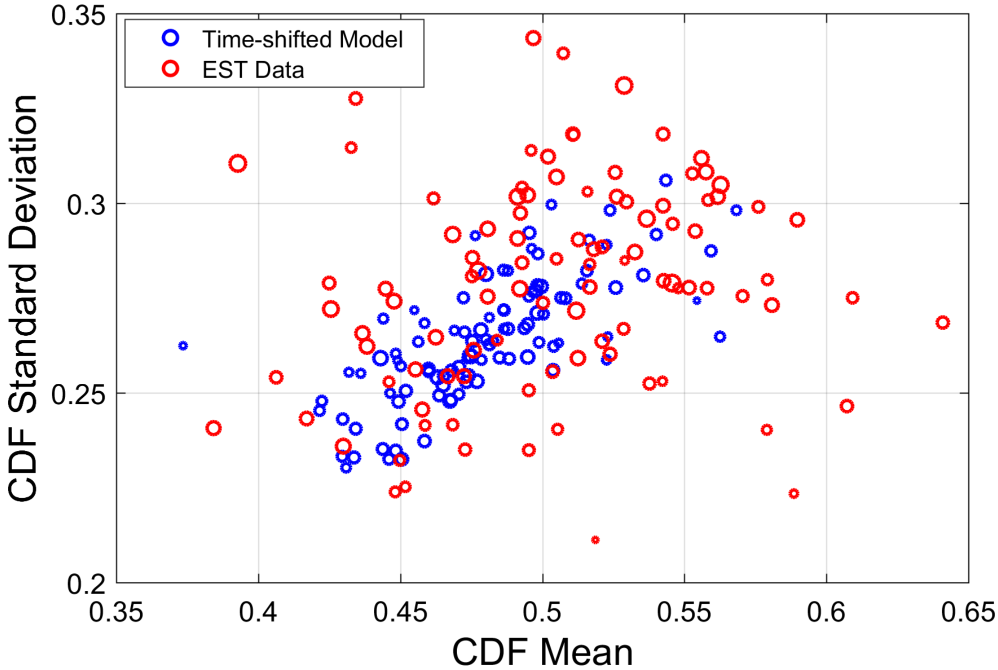
For the *AO_Conv* metric, that measures overall convergence to the calibration dataset average per-timestep occupancy, the performance is less accurate that the 'composite' method. When the 'error' per timestep is analysed the most significant period of weaker replication is in the sleep transition period. This suggests that the simple correlation between wake and sleep time used requires a more complex statistical basis for better accuracy. For the *DurDist* metric, that measures the error in occupancy state duration distribution replication, the performance is also worse for the 'individualised' model by a factor typically of between 2 and 3. This results from the additional forcing of behaviours.

Further analysis of the impact on the overall demand model output is required to determine if the loss of performance in terms of TUS dataset replication is outweighed by an improvement in capturing individual behaviours.

When the distribution of 'Very Low' volume hot water cycles per household is anal-



(a) Unshifted vs. time-shifted



(b) Time-shifted vs. EST dataset

Figure 7.14. Cycle start time cumulative probability function mean and standard deviation for 'Very Low' volume hot water use per household comparison for unshifted and time-shifted model basis. Data for the 'EST Data' distribution from [90].

ysed, which is considered to be the most occupancy-driven and least behaviour-driven demand, Figure 7.14 indicates the distribution using the time-shifted 'individualised' occupancy model is a closer replication of the dispersion of behaviours seen in the EST dataset than the unshifted 'composite' model, indicating the potential for improved demand prediction performance.

7.3.2 Residual Average Occupancy Convergence

The inclusion of significant occupant and day type differentiation (see 4.5.4 and Appendix B) and realistic occupant work weeks (see 4.5.4.5) ensures a degree of variation in average occupancy. However, Figure 4.12 indicates that, while this results in improved performance over existing methods, the occupancy model output remains less variable than real behaviours. The requirement to maintain minimum calibration population sizes for modelling stability (see 4.5.4.10) forcing convergent average occupancy behaviour for significantly sized subsets of modelled occupants.

Addressing average occupancy convergence is less straightforward than for transition timing convergence, as it is an underlying function of the overall probability model rather than related to specific transitions in distinct time periods. However, the main determinant of average occupancy is the balance of 'active' and 'out' periods, with 'sleep' duration variability addressed by the transition timing method identified in the preceding section. This potentially allows the 'active-out' transition probabilities to be manipulated to account for variable in-group behaviours.

A potential improved method identifies periods where an occupant is not asleep and the next timestep has both a non-zero 'active' and 'out' probability. The ratio of the 'active' and 'out' probabilities can be adjusted by a multiplier to increase or decrease 'active' occupancy. Figure 7.15 shows the impact on overall active occupancy for 100 annual simulations for the same 60-year old, non-working single householder as the multiplier is randomly selected between 0.5 and 1.5, in comparison with an unfactored model. The unfactored model results over 100 simulations vary in a tight band between 0.4 and 0.45, with the application of the multiplier generating significantly more variation.

Analysis of the average output from a representative range of multiplier values

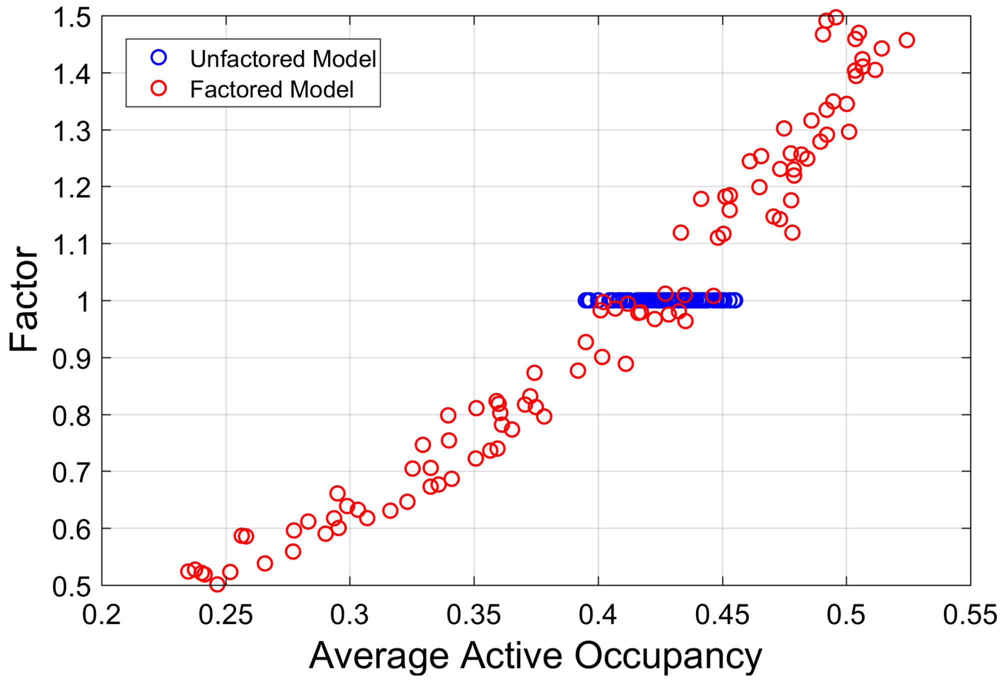


Figure 7.15. Comparison of the range of average active occupancy results for an annual model run generated by unfactored and average occupancy factored occupancy models for a 60-year old, non-working single-person householder.

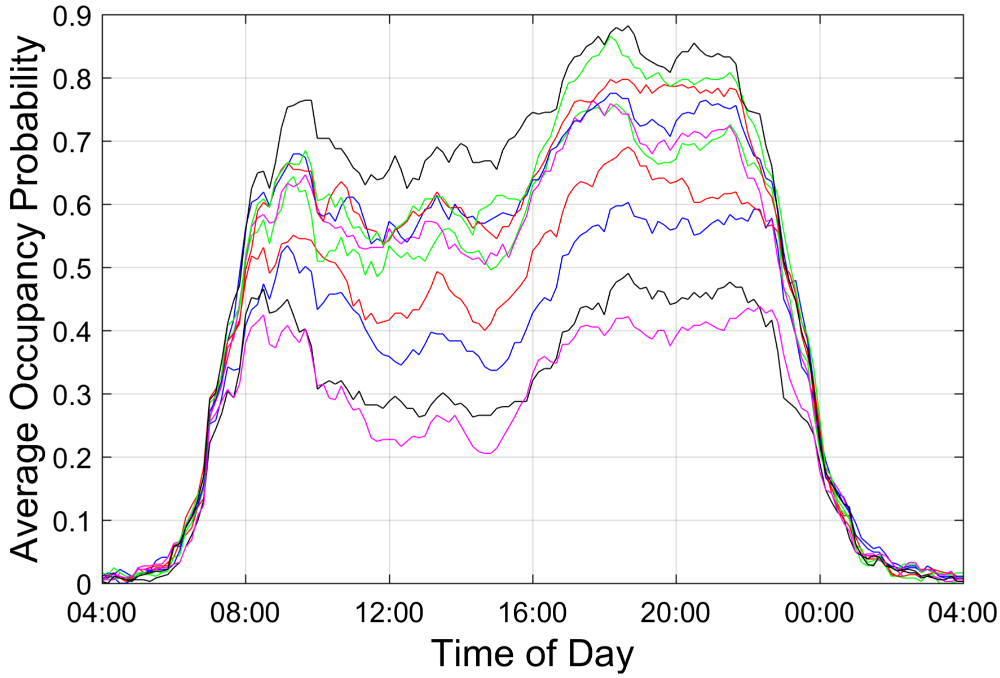


Figure 7.16. Representative range of average active occupancy profiles for a 60-year old, non-working single-person householder for average occupancy factors between 0.5 and 1.5.

shown in Figure 7.16 indicates that the impact is relatively consistent over the waking period. Further analysis of individual day occupancy patterns suggests that the method remains stable with no evidence of an excessive increase in very short or long duration 'active' or 'out' periods. Extension of the method to the combined couple/parent occupancy model is more complex but the same ratio manipulation is possible for the main 'active-out' transitions (i.e. SA/SO and AA/AO/OO, where 'S'=Sleep).

Whilst the identified method allows some basic statistical manipulation, and can be easily calibrated using the data presented in Figure 7.15 to adjust overall average occupancy, as stated it does not account for time-dependent variations but varies occupancy uniformly. For accurate individual behaviour modelling, the multipliers used would need to be both time-period and day type specific. However, with no long-term occupancy data currently available, realistic calibration at this resolution is not possible. Consequently, this method was not developed further at this stage and no average occupancy adjustment was incorporated in the final 'individualised' occupancy model. However, the method was shown to have potential to solve this convergence problem with the availability of suitable calibration data.

7.4 Individualised Appliance Behaviour Module

7.4.1 Module Basis

The appliance cycle start time module, calibrated with start time data from the overall HES population with limited household type adjustments, was shown in 7.2.2 to result in excessive start time distribution convergence for modelled households (see Figure 7.5). The occupancy sub-model with individual behaviour factoring (see 7.3) generates additional variation in typical use times, but this does not account for any individual behavioural traits associated with specific appliances (e.g. typically showering after waking, using the dishwasher prior to sleep, etc.).

As outlined in 7.2.2, a method to identify individual household behaviours was developed based on the cycle start time distributions introduced in 5.9. The distributions used are the 'raw' distributions prior to the occupancy normalisation used for the cycle start time identification module. For each household, cycle start times are converted to

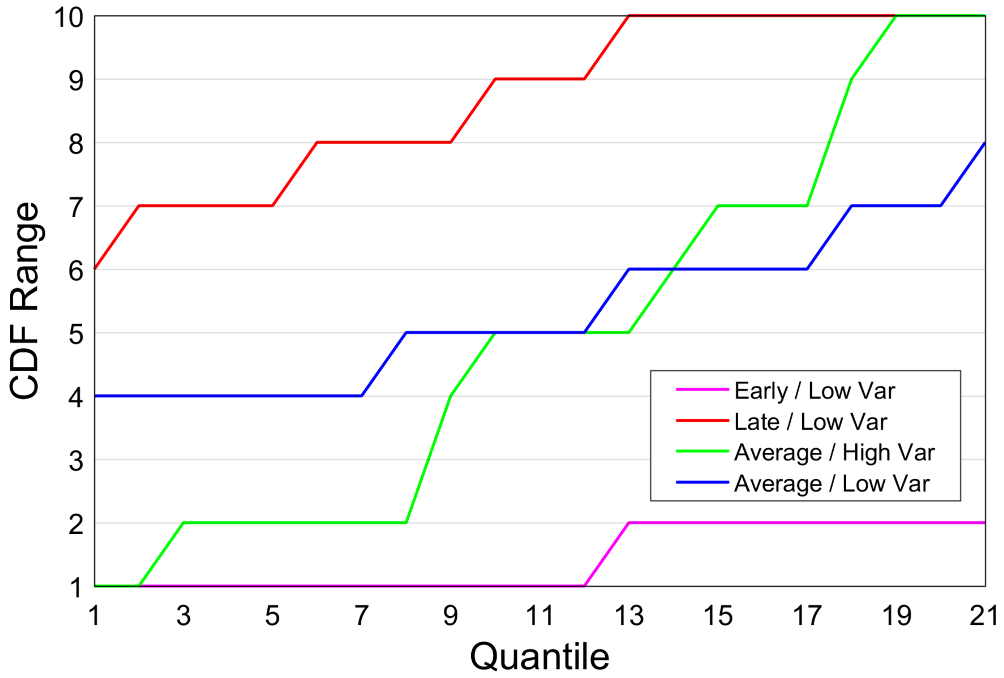


Figure 7.17. Example individual household cycle start time cumulative distribution function (cdf) quantile distributions from developed Markov chain model.

the cumulative distribution function (cdf) value from the appropriate cycle start time distribution. The deviation from a mean value of 0.5 indicates whether a household typically uses the specific demand earlier (<0.5) or later (>0.5) than average. The lower the standard deviation the more habitual the behaviour. Evidence of distinct individual behaviours was identified for kettles, washing machines, dishwashers, showers, cookers/ovens plus 'High' (15-30 litres) and 'Very High' (30 litres+) volume hot water cycles.

7.4.2 Individualised Behaviour Module Development

The 'composite' demand sub-models defined in Chapters 5 and 6 determine the cycle start time from daily occupancy and the appliance cycle-specific cycle start time probability distributions by converting a generated random number to a time based on the relevant start time distribution (see 5.9.2), with the potential times limited to occupied periods. The updated model basis manipulates the random number generation to better replicate the realistic distribution of appliance use timing behaviour shown graphically in Figures 5.39 and 7.5.

The random number manipulation is achieved for each selected appliance and hot water use by interpolating the range of cycle *cdf* values for each household into 21 representative quantiles and converting each quantile value into one of ten *cdf* value ranges (1=0-0.1, 2=0.1-0.2, etc.). The range transitions between each of the quantiles were determined for each HES and EST dataset household and used to calibrate a separate Markov chain model for each specific demand.

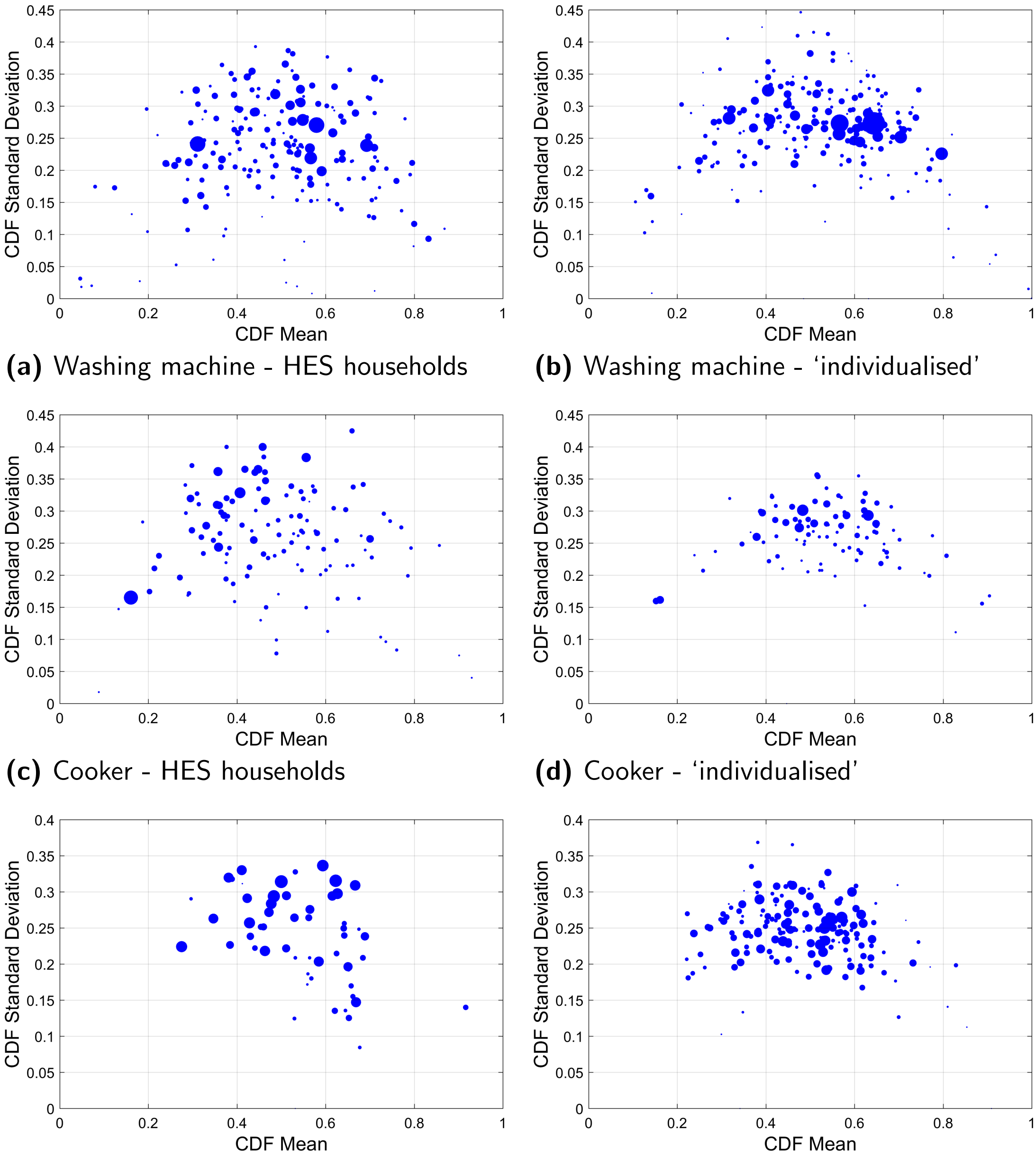
The module generates new distributions for each modelled household based on a probabilistically assigned midpoint (11th) quantile, with the Markov chain model working in both directions from the midpoint to the minimum (1st) and maximum (21st) quantile values to allow the midpoint value to be further factored based on household occupancy timing compared to the average. Within the 'individualised' cycle start time module, a value is selected randomly from the 21-quantile distributions (see Figure 7.17 for examples) and then the actual value used is selected randomly within the range (e.g. a value of 4 randomly selected from the distribution is converted to a number randomly selected between 0.3 and 0.4). The determined value replaces the original 'composite' model 0-1 range random number generation used by to determine the cycle start time (see 5.9.2) to skew the generated start times to reflect individual household behaviours.

The use of a relatively small number of quantiles and ranges ensures that the broad overall pattern of potential behaviours is captured but that there is sufficient variation as a result of randomly selecting values within the ranges to ensure that the model is not forcing close replication of the input data. The method was preferred to other methods of skewing probability distributions as it allows for multiple and well separated periods of higher use probability. Figure 7.17 shows examples from the 'cooker' module of four typical resultant distributions for households that exhibit distinct use behaviours (early/average/late with low variance, and average with high variance).

For the 'High' and 'Very High' volume hot water cycles, there is assumed to be an overarching relationship. Midpoint values are first assigned for the 'High' module and then assigned for the 'Very High' module using the Kernel Density method (see Appendix A) based on the statistical relationship between the two midpoint values identified from the calibration dataset.

A proportion of the observed use timing variation is assumed to be the result of occupancy variations. Without further data to account for occupancy, the mean value for each household is modified by the extent to which the average household occupancy is earlier or later than the average population behaviour. This is a further area where modelling would be improved by an integrated occupancy and demand dataset that allowed the relationship between occupancy and cycle timing for individual households to be better incorporated.

The method as currently implemented does not account for specific daily occupancy patterns. For the original 'composite' method, the potential cycle times are first limited to the occupied periods and then the specific time is determined based on the generated value between 0 and 1 which is used to locate it proportionally within the occupied period. The same process is used for the 'individualised' method, therefore the behaviour is only skewed based on the household-specific distribution and not forced to specific times. Further improvement of this method to account for highly distinct use patterns by linking use to specific time periods is required for better replication of applicable households.



(a) Washing machine - HES households

(b) Washing machine - 'individualised'

(c) Cooker - HES households

(d) Cooker - 'individualised'

(e) 'Very High' hot water - EST households

(f) 'Very High' hot water - 'individualised'

Figure 7.18. Cycle start time cumulative distribution function mean and standard deviation per household. Comparison for measured data and 'individualised' model. Data for the 'EST households' distributions from [90].

7.4.2.1 Results Analysis

Analysis of the *cdf* mean and standard deviation values from the 'individualised' model results in Figure 7.18 shows an overall distribution that is significant closer to the measured data than the unmodified 'composite' model results shown in Figure 7.5. The results also indicate that there is no evidence of greater convergence to the mean behaviour for households with a higher number of use events, which was a critical performance problem for the 'composite' model.

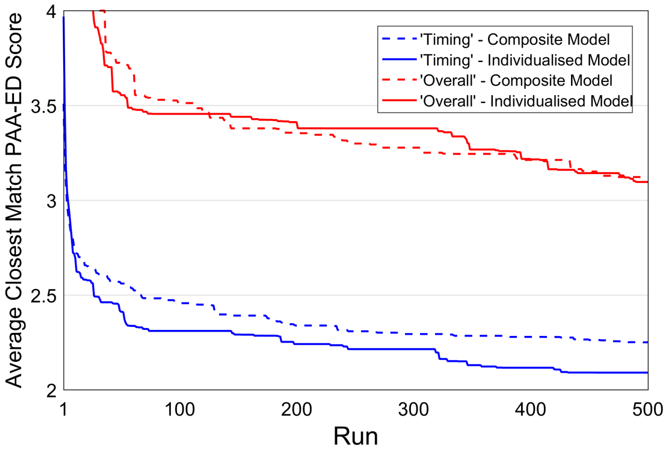
As a statistical measure of the degree to which each sub-model results replicate the HES and EST dataset characteristics, the average *distance* of each mean and standard deviation data point from the average is calculated. *Distance* is defined by Equation 7.1 for a number of data points (households), *H*.

$$Distance = \sum_{h=1}^{H} \frac{((cdf_{mean}(h) - cdf_{mean})^2 + (cdf_{sd}(h) - cdf_{sd})^2)^{0.5}}{H} \tag{7.1}$$

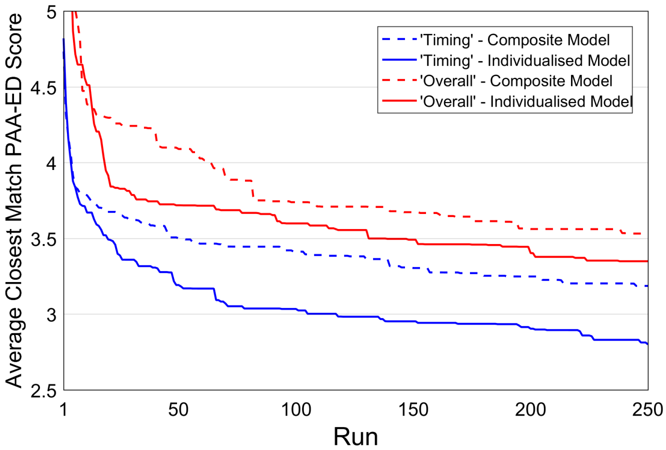
Analysis of results from 10 model runs for an equivalent set of households in Table 7.2 shows that the *Distance* measure for the 'individualised' model basis is significantly closer to the equivalent measure for the actual data for all relevant specific demands. However, as indicated visually in Figure 7.18 and from the *Distance* measure, the dispersion of modelled results is lower than the actual data, particularly for the standard deviation measure. This results from the lack of direct integration between the occupancy and cycle start time sub-models, as outlined above, that does not easily allow highly habitual (i.e. low standard deviation) behaviour in particular to be closely replicated, and also potentially from the size of the *cdf* ranges currently used in the Markov chain model. Further work in this area is required as and when integrated occupancy and demand data is available.

Table 7.2
Average 'Distance' measure for each household from mean cycle start time behaviour for different specific demands. Dataset data for analysis from [89] and [90].

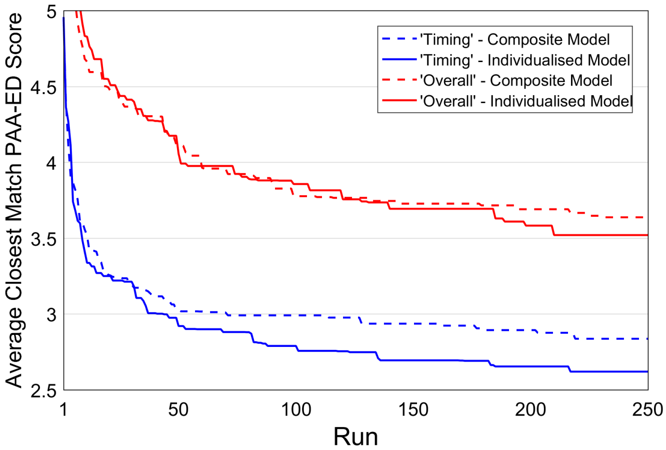
	Cooker	Washing Machine	Hot Water 'High'	Hot Water 'Very High'
Dataset (HES/EST)	0.153	0.175	0.110	0.139
'Composite' Model	0.090	0.102	0.079	0.076
'Individualised' Model	0.126	0.155	0.121	0.139



(a) HES dataset



(b) 'Richardson'



(c) REFIT

Figure 7.19. Cumulative closest match PAA-ED score average per run comparison between 'composite' and 'individualised' models for different dataset-equivalent electricity demand models.

Whilst the individual cycle start time method can be further improved, it has been shown to be a significant improvement on the 'composite' model basis for each specific demand. Further analysis is required to determine if the method improves the overall demand model performance and to determine how applicable it is for energy system development. This analysis is presented in the following section.

7.5 Individualised Demand Model Performance Assessment

7.5.1 Individual Household Similarity

7.5.1.1 Electricity Demand

The PAA-ED similarity method introduced in 5.14.4 can be used to compare the electricity demand sub-model output with and without the defined individualised occupancy and cycle behaviour modifications. The PAA-ED method simplifies each 144-timestep average demand profile to a 36-time segment approximation based on ranges of demand, which can be compared using a standard Euclidean distance similarity measure.

Five hundred HES dataset household equivalent model runs for both 'composite' and 'individualised' methods were compared. After each run and for each of the households, the cumulative lowest PAA-ED value for all runs completed is determined and the overall average for all households calculated. This average value is a simple measure of the ability of the model to generate demand profiles consistent with the range of behaviours observed in the measured data and the progression of average results with runs performed gives an indication of the speed with which each highly probabilistic method identifies real and representative patterns of behaviour.

The results for the full 250-household HES-equivalent model are inconclusive, with similar results for both methods. However, for the 26 households which were monitored for longer than 28 days (between 61 and 249 days with an average of 125 days), the average for both methods is significantly lower, and there is an improvement in the 'timing' value after 500 runs for the 'individualised' model with an average of 2.09 compared to 2.25 for the 'composite' model (see Figure 7.19(a)). This suggests that the

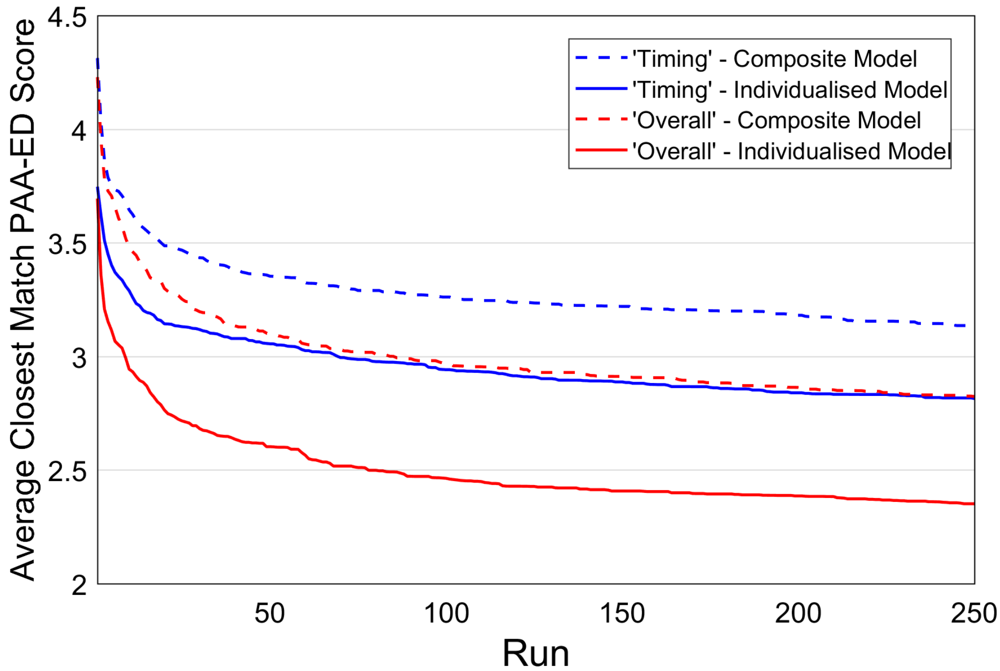


Figure 7.20. Cumulative closest match PAA-ED score average per run comparison between 'composite' and 'individualised' models for EST dataset-equivalent hot water demand model.

length of the analysis period is also important, with 28-day profiles being significantly more erratic than longer duration profiles and therefore more difficult to replicate.

Similar analysis for the Richardson [69] and REFIT [45] datasets (see 2.3), and equivalent models, which are all of a 1-year duration, show a clear performance improvement for the 'individualised' method basis (see Figure 7.19(b) and (c)) with 'individualised' and 'composite' method results of 2.80 and 3.19, and 2.62 and 2.84 respectively for the 'timing' basis, and 3.34 and 3.53, and 3.51 and 3.64 for the 'overall' basis.

7.5.1.2 Hot Water Demand

Similar PAA-ED analysis was undertaken for the hot water demand sub-model comparing the Energy Savings Trust (EST) dataset (see 2.3) with the dataset-equivalent model population.

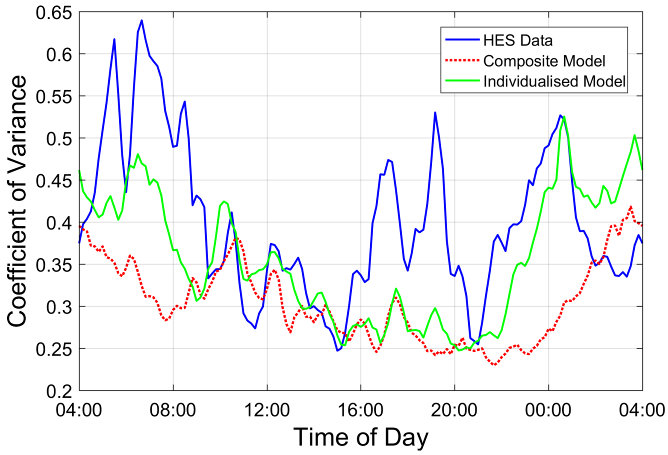
For both the 'timing' and 'overall' assessments (see Figure 7.20), the 'individualised' method shows significantly better performance than the 'composite' method. The improvement is also greater and more consistent than for the electricity demand sub-model. The similarity assessment of individual results also shows a distinct improvement in the number of 'High' and 'Low' similarity results as shown in Table 7.3.

Table 7.3
Hot water model closest cumulative match similarity analysis range results for the 'composite' and 'individualised' EST household equivalent models after 250 model runs.

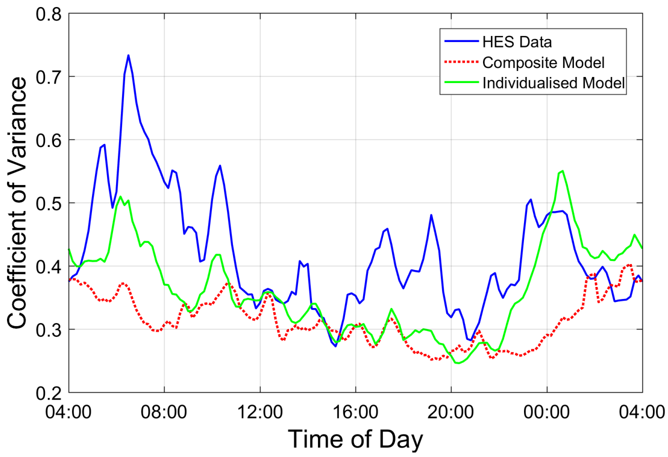
Similarity	'High'	'Good'	'Some'	'Low'
'Timing' - 'Individualised'	62 (62%)	31 (31%)	7 (7%)	0 (0%)
'Timing' - 'Composite'	46 (46%)	33 (33%)	12 (12%)	9 (9%)
'Overall' - 'Individualised'	37 (37%)	40 (40%)	21 (21%)	2 (2%)
'Overall' - 'Composite'	28 (28%)	41 (41%)	21 (21%)	10 (10%)

7.5.1.3 Similarity Analysis Summary

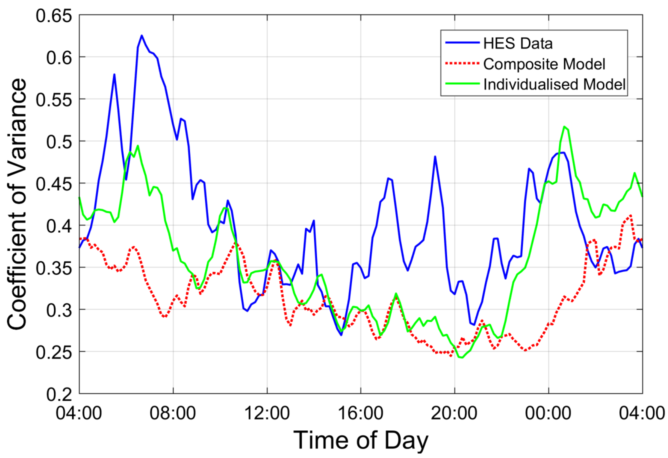
The results of the demand model analysis for the developed individual-calibrated ('individualised') method show that it performs better that the group-calibrated ('composite') method, particularly where annual data is available for comparison. The identified weaker replication of the group-average occupancy characteristics with the addition of the individual occupancy transition timing adjustments is therefore outweighed by the



(a) Single, working age



(b) Couple, retired



(c) Multi-adult

Figure 7.21. Per-timestep coefficient of variance of mean-normalised per-household electricity demand for measured HES data and equivalent 'composite' and 'individualised' model results for three household types. Data for the 'HES Data' distributions from [89].

improvement in capturing individual demand behaviours. Particularly as further improvement of the 'individualised' occupancy method is possible with an improvement in the sleep-transition time element which is the main source of the current weaker performance.

The significant performance improvement seen for the hot water demand sub-model, with virtually no individual household from the input dataset not replicated to a reasonable degree and a high proportion closely replicated, suggests that the impact of the individualisation method can be significant. Applied to the 'High' and 'Very High' cycle volume ranges, the 'individualised' method impacts on 63% of hot water use with a single behaviour adjustment per household. The electricity analysis is more complex, with multiple appliances with different behaviours and power profiles which account for a far smaller proportion of overall demand and are monitored for shorter periods. The benefit of the 'individualised' method is therefore potentially more significant for individual appliance behaviour replication than overall demand. The hot water analysis suggests that this method could be beneficial for demand shifting analysis for individual electrical appliances but this is difficult to confirm with the available demand data, although Figure 7.6 does clearly demonstrate a distinct improvement in start time distribution replication.

7.5.2 Overall Variance

In 7.2.1 it was shown that the 'composite' model basis did not generate the same per-timestep variance in demand per household seen in the demand datasets (see Figures 7.4 and 6.13). Repeating the analysis for the 'individualised' model as shown in Figure 7.21, demonstrates that it improves on the replication of the typical variance per household, particularly in the post-waking and pre-sleep periods on which the occupancy model modification is primarily focused. There remains a significant difference in variance between the dataset and model in the late afternoon/early evening period which requires additional analysis that is discussed further in Chapter 9.

7.5.3 Applicability for Individual Household Modelling

The average demand analysis, both for electricity and hot water consumption, determined that the developed demand model captures the overall range at this resolution. The similarity assessment results suggest that there is a small but significant number of households whose time-dependent demand behaviours, at least for the more complex electricity demand element, are not captured. Analysis of these households shows highly distinctive use patterns, which are either characterised by unusual overall timing or very specific and consistent periods of very high demand. Capturing both types within a probabilistic model is limited both by the uniqueness of the behaviour and the size of the available calibration datasets.

The conclusion is therefore that the model captures the time-dependent demand behaviour variation of between 80 and 90% of households for electricity demand (see Table 5.14) and at a slightly higher level for hot water (see Table 7.3). The majority of the remainder are likely outwith the scope of a probabilistic model. For these outlier households, the use of actual data is likely to remain the best method to capture extreme outlier behaviours. However, the model can be used to capture a significant range of potential behaviours for a comprehensive range of households, and is therefore a distinct improvement on existing methods for assessing individual households.

7.5.4 Minimum Number of Households for Comprehensive Analysis

Further analysis was undertaken to determine the number of households required for the impact of the poorly replicated, outlier households on overall system demand to be negated. This determines the minimum size of energy system, in terms of number of households, for which the demand model can be used to comprehensively capture all potential demand scenarios with sufficient accuracy.

As stated in Chapter 1, sufficient accuracy, that is the degree to which both the real and modelled time-dependent behaviours of a multi-household population has converged for the purposes of system design, is both subjective and situation-specific. Accuracy was therefore analysed quantitatively using the same PAA-ED similarity method introduced in 5.14.4 for the comparison of individual households, but in this case for overall system demand. The relative timing ('timing') analysis was used as it is a better measure of model timing prediction without the distorting influence of variable baseline power levels that the model should account for with sufficient runs.

Table 7.4
Average and maximum per-run combined PAA-ED similarity score for different numbers of households from the HES dataset and HES-equivalent 'individualised' model output.

Households	1	5	10	20	30	50	100
Average	4.10	2.40	1.81	1.46	1.34	1.19	1.08
Maximum	9.38	6.40	4.68	3.25	2.42	2.15	1.64

Comparing the average model output over 500 runs for the HES-equivalent 'individualised' electricity demand model shows that the worst-case individual household closest match PAA-ED score is 9.38 with an average of 4.10. Table 7.4 shows the average and maximum closest match scores from 500 random combinations of the stated number of households based on similarity analysis of each of the 500 model run results for each combination. At five households, the average closest match is less than the '2.5' threshold for 'High Similarity', however, the worst results have poor similarity with 20% exceeding the '3.5' threshold for 'Good Similarity'. At between 15 and 20 households, no result exceeds the '3.5' threshold, and at around 30 households no result exceeds the '2.5' threshold. At 100 households, the maximum value approaches 1.5, which indicates that the model is able to closely match relative demand timing in all cases with a sufficient number of runs. There is an insufficient number of households (250) to make meaningful conclusions for higher numbers of households.

The conclusion is therefore that the influence of outliers is significantly reduced for systems in excess of 15-20 households, and that the results can be used with good confidence for 30-household systems and above, with low risk of poor prediction resulting from the influence of outliers not captured by the model. As outlined, most of the HES household demand data is of a one-month duration which generates demand profiles that are erratic and therefore difficult to predict. The actual minimum number of households for comprehensive modelling may therefore be less than these figures where replication of more consistent annual demand profiles is required.

7.5.5 Minimum Number of Runs

In general, the similarity analysis has shown that between 50 and 100 runs, as a minimum, are required for the per-run improvement in matching specific profiles to diminish to a steady baseline level for both 'composite' and 'individualised' options, and is largely independent of the number of households. Further analysis was required, however, to determine the minimum number of runs required for the probabilistic models to generate representative results and sufficiently close matches for real datasets.

Detailed analysis for multi-household combined system results has shown that the probabilistic nature of the model generates unique results over a high number of runs (>1000). Detailed statistical analysis of the similarity results per run using measures such as the Komolgonov-Smirnov, t, and F tests show that subsets of up to 1000 runs from 10000 runs samples do not converge to a 5% confidence that they are from the same distribution; therefore, the highly probabilistic model, both for the 'composite' and 'individualised' versions, continues to generate statistically unique results over a large number of runs.

For most types of analysis, and to limit the computational time required, a match that approaches the absolute closest match is sufficient. A value within 25% of the minimum similarity result for 10000 runs was arbitrarily selected as a measure of a sufficiently close result. The results were analysed for 1000 combinations of different numbers of households from the HES dataset, and equivalent model results, to determine the number of model runs required to achieve this target value. The results from the 1000 combinations have a long tail and therefore both the median and average results were determined. The 95% percentile is shown as a measure of how many runs were required to capture all but low probability outlier results. The results are shown in Table 7.5.

Table 7.5
Median, average, and 95th percentile number of runs to achieve a combined PAA-ED similarity result within 25% of the 10000-run minimum value.

Households	10	30	50	100
'Composite' - Median	49	89	115	135
'Individualised' - Median	57	87	72	65
'Composite' - Average	121	182	210	234
'Individualised' - Average	122	176	153	157
'Composite' - 95th Percentile	532	683	723	792
'Individualised' - 95th Percentile	478	699	599	649

The results show that while the 'composite' method shows an increase in all values with number of households, the 'individualised' method results are less consistent. This indicates that the more distinct output of the ‘individualised’ model requires more runs that the ‘composite’ model for small numbers of households as the overall variation in potential results is higher. As the number of households increases beyond 30, the higher output variation of the ‘individualised’ model reduces, but the benefits of the more distinct output in finding close matches is retained, driving an increasing performance improvement as the number of households increase.

The overall conclusion is that for the ‘individualised’ model between 100-200 runs are required to generate a representative range of results but that an excess of 500 is required to ensure the vast majority of potential demand patterns are captured.

7.6 Chapter Summary

This chapter detailed the development process for enhancements to the occupancy and demand sub-models that allow individual occupant and household behaviours to be better captured. The chapter highlights are as follows:

- Evidence of convergence in demand prediction as a result of composite group-calibration of both occupancy and cycle start time identification approaches identified in Chapter 5 was confirmed for household type groups and for a number of individual appliances and hot water uses.

- An improvement was proposed to reduce the occupancy model convergence by time-shifting the group-calibrated probability modules to account for the distribution in sleep, waking, leave and return times for individuals compared to the group average.

- The assessment of individual household appliance start times in comparison with the overall average determined that a significant number of households had distinct relative use behaviours for a number of electrical appliances and for high volume hot water events.

- To account for individual use behaviours, the cycle start time identification module is further manipulated for each household to account for different identified behaviours. This is achieved by restricting the random number generation for each household to simulate the different behaviours.

- A performance assessment of the developed ‘individualisation’ modules showed distinct improvements in comparison with the ‘composite’ group-calibrated model basis for both individual appliances behaviours and hot water overall demand. The results were less conclusive for the overall electricity demand model because of shorter measured data durations, conflicting behaviours between appliances, and a lower overall contribution from the individualised element.

- Further analysis determined that the overall demand model captured behaviours associated with 80-90% of individual households and could be used for comprehensive analysis for energy systems comprising a minimum of 15-20 households.

Chapter 8

Model Applications

8.1 Chapter Overview

The purpose of this chapter is to demonstrate that the stated aims of the demand model development in Chapter 1 have been achieved. These were to address shortcomings in current methods for analysis of small-scale energy systems. Three applications are used to illustrate the effectiveness of the developed probabilistic model; the calculation of average and peak demand, and overall time-dependent optimisation analysis.

For average electricity demand, it is shown that the developed and validated model, whilst computationally intensive, demonstrates that the BREDEM and SAP models ignore significant potential socio-economic demand drivers, and confirms existing analysis that they may overestimate demand in many cases. It is also shown that the average demand prediction uncertainty for single households exceeds a factor of two and remains potentially significant in excess of 100 households. Similar conclusions were also drawn for the hot water demand model, but less categorically as a result of lower household differentiation and the lack of independent validation data.

For peak electricity demand, the output of the developed model is first validated with measured demand data with good accuracy. In addition, the simplistic per-household design 'rules-of-thumb' used by electricity companies for distribution system sizing were shown to be potentially inaccurate, particularly for sub-100-household systems.

For peak hot water demand (diversity), the developed model demonstrates that the current preferred UK design basis of the Danish DS439 standard [63] potentially underestimates diversity for systems up to at least 50-75 households. The main conclusion from the presented analysis was that a more detailed study was required, accounting for the dynamics of typical district heating systems, to determine if the Danish basis is applicable for UK conditions. For both electricity and hot water peak demand, the impact of household characteristics was shown to be relatively small, with number of households the principal determinant.

For time-dependent optimisation analysis, two methods are introduced for selecting subsets of multi-run, high-resolution demand data for further analysis. One is a representative selection of the potential range of scenarios and the other is limited to the extreme cases. Using a nationally representative 100-household model, the subsets were used for two example design problems; generation and storage equipment sizing, and a full system analysis for generation utilisation and grid balancing requirements. The overall conclusion was that time-dependent demand uncertainty remains significant to at least 200 households and that further analysis of more complex systems is required.

8.2 Model Applications Background

Five key elements for the design of distributed generation-supplied grid sub-systems were identified in 1.6.1; sizing and balance, grid connection, storage, demand management, and seasonal matching assessments. Each element requires a detailed understanding of the dynamics of the energy system, which on the demand side range from prediction of peak demand at different timescales to high-resolution time series data for balancing and matching analysis. This chapter highlights the potential improvement in analysis for the five identified elements using the developed probabilistic, differentiated, high time resolution demand model, with particular focus on sizing and balance, storage, and grid connection assessments.

As addressed in 1.5, the degree and impact of prediction variation and uncertainty on system design analysis is currently poorly understood. The overarching conclusion of the work presented in the preceding chapters is that at the small-scale (i.e. <500 households) the range of household characteristics can significantly impact the average demand and relative timing, and, in addition, that the size of the population impacts the degree of uncertainty for any demand prediction as a result of the proportional impact of individual household behaviours. Furthermore, some or all of the household characteristics may not be known in advance of system design and this represents an additional prediction uncertainty.

In Chapter 3 it was determined that no existing model combines comprehensive and differentiated household calibration, a high time resolution, and probabilistic factoring to account for behavioural variations and associated prediction uncertainty. The developed electricity and hot water demand model addresses these shortcomings and allows a detailed assessment of the impact of this overall uncertainty on the five identified key design elements. In this chapter, the developed demand model applicability and capability for these assessments is reviewed, with specific reference to performance against existing modelling methods, where applicable, and residual areas for improvement.

Section 8.3 reviews the impact of overall population characteristics on average demand, and then determines the residual uncertainty from unknown household characteristics and from individual household behaviours. The degree of uncertainty is quantified for different sizes and types of energy system. Whilst average demand uncertainty prediction has limited applicability for detailed system design, it does provide a simple assessment of potential differences that can be expected from more detailed, time-dependent system analysis. Average demand prediction also allows the performance of low-resolution models, such as BREDEM [111] and SAP [48], to be analysed.

Section 8.4 assesses the performance of the developed model for the prediction of peak demand, an important assessment for network sizing: for electricity demand, maximum non-coincident demand (i.e. the sum of the peak demand for each individual household) and after-diversity maximum demand (i.e. the peak total system demand) are typically used; and for hot water, the diversity or simultaneity factor, which is equivalent to the after-diversity factor for electricity demand. The impact of different population types and sizes, and also the effectiveness of existing design standards, was reviewed.

Finally, 8.5 introduces a method for selecting representative and extreme behaviour datasets from multiple run analysis for time-dependent optimisation assessment using either equally probable or worst-case scenarios. Using the selected scenarios, a simple solar-matching model is used to indicate the potential impact of time-dependent demand uncertainty on prediction of system performance. This type of high-resolution analysis is critical for both sizing and balance, storage, and grid connection assessments, and has the potential to be extended to the other design elements (demand management and seasonal matching) with further model improvements or more detailed analysis, as discussed in 8.5.3. In addition, the impact of both demand variation and uncertainty are analysed for a solar panel plus storage system.

8.3 Average Demand - Variation and Uncertainty

This review of average demand variation and uncertainty uses the 'individualised' demand model basis introduced in Chapter 7, based on the core modelling methods detailed in Chapters 5 and 6. These chapters include validation of both the electricity and hot water models for average demand prediction performance in relation to characteristics-driven variations and prediction uncertainty. Therefore, no further validation is detailed in the following section.

8.3.1 Characteristics-Driven Variation

Assessment of the average predicted ('baseline') power use for a household or system can be used as a basic measure of relative demand or for low-resolution analysis. The primary existing sources of UK-calibrated average demand prediction are the BREDEM [111] and related SAP [48] models, that include average annual and monthly assessment of expected appliance, lighting, cooking, electric shower, and overall hot water demand; based primarily on number of occupants, and also floor area for appliance and lighting estimates.

For baseline analysis, the main difference between the developed model and the BREDEM and SAP methods is that it includes factors that vary depending on the socio-economic characteristics of a location. The age profile, employment probability, and income range being influenced by the location as outlined in 2.4 and 4.3. The predicted demand and impact of the additional factors within the developed model was therefore assessed against the BREDEM model output.

8.3.1.1 Electricity

The electricity demand model was run for four distinct, archetypal, small-household community types with 100 households each, where number of bedrooms (60 1-bed and 40 2-bed), number of occupants, and floor area per household were fixed. Modelled using BREDEM or SAP, these 'communities' would be predicted to have the same average demand, in this case 340.9W per-household. Two were modelled as 'retirement' with all over-65 years old occupants. The other two were modelled as 'urban' with no over-65 residents, as might be expected for certain city centre locations. Both types were modelled for the most and least deprived decile location characteristics, with the results shown in Table 8.1.

Table 8.1
Predicted average per-household electricity demand variance for four extreme small-household 'communities' by age and location with identical house types and sizes.

Community	'Urban'		'Retirement'	
Area IMD Decile	Least	Most	Least	Most
Average Power (W)	256.4	218.4	204.6	191.5

Similar analysis for the same extreme deprivation decile location cases for a 100-household nationally representative house type model (see Table 5.15), showed average demand varied from 396.8W per-household for the most deprived decile to 454.6W for the least. The BREDEM equivalent prediction in this case is 373.8W for both cases.

The maximum location-driven variance is c.±5-8% of the average prediction based on the extreme deprivation deciles. The results are consistent with the LSOA-level deprivation decile-based demand prediction validation exercise detailed in 5.14.6.4, which showed a consistent, and largely predictable, increase in demand with decreasing deprivation.

Palmer et al [180] compared the Household Electricity Survey (HES) data [89], used to calibrate the developed model, against both the BREDEM and SAP basis, and found that the HES demand data was significantly lower. Consistent with this, the developed model results highlight a tendency for the BREDEM model to significantly overestimate demand for smaller households and underestimate, but to a lesser degree, for larger households.

8.3.1.2 Hot Water

Similar to electricity demand, the BREDEM and SAP models estimate annual and monthly hot water use per-household based on number of occupants and shower type. Again, no allowance is made for different household types or income levels.

The income behaviour factoring within the developed model for hot water use differentiation (see 6.4.2.1) is a composite factor for all gas demand rather than hot water use specific. Therefore, any variation due to area socio-economic factors is only indicative that a difference exists. A hot water-specific income factor would be required for a reliable assessment.

As an indication of potential variation, for the 100-household nationally representative models based on deprivation decile extremes outlined in the preceding section, the average hot water use per household varied from 105.5 litres/day for the most deprived decile to 127.6 litres/day for the least. Again, this indicates that number of occupants alone is insufficient for detailed analysis. The BREDEM equivalent predicted demand is 117.5 litres/day, which is consistent with the model output average basis, and to be expected given that the BREDEM model is also calibrated from the EST hot water dataset [90].

8.3.1.3 Average Demand Assessment Summary

The primary aim of the developed model is high-resolution analysis and, therefore, not directly equivalent to the simpler, less computationally intensive approach and monthly resolution of BREDEM and SAP. The comparison does, however, highlight that simple baseline estimation requires an accurate assessment of more factors than occupant number and floor area. It is therefore recommended that consideration is given to updating the BREDEM/SAP approach to accommodate additional socio-economic factors associated with energy use behaviours and household location. The work presented also confirms the conclusion of analysis by Palmer et al that the BREDEM/SAP electricity demand calibration basis for number of occupants and floor area is potentially inaccurate.

8.3.2 Average Demand Uncertainty

As addressed in Chapter 3, there are no existing models that accurately address the demand prediction uncertainty resulting for variations in the following characteristics; appliance ownership and type, occupancy, income, and overall energy-use and appliance-specific use behaviours. The open-source Richardson et al [69] model, which is the current principal source of high-resolution UK domestic demand prediction data, incorporates a degree of differentiation for occupancy and appliance ownership, but has limited measured data derived, probabilistic calibration and no socio-economic factoring, that prevents an accurate assessment of uncertainty. For example, the appliance sub-model within the Richardson et al model, if set with identical appliance ownership per run, will generate the same total energy use with only time-dependent variation as a result of the occupancy sub-model. The developed model is therefore the first high time resolution model that incorporates the influence of all the identified characteristics to allow the demand uncertainty for small-scale energy systems to be analysed.

For any demand analysis, there are two levels of uncertainty to be determined: the potential uncertainty due to lack of household information; and the residual uncertainty from random behavioural variations. The following sections review both factors for electricity and hot water demand, for both a single household and a nationally representative 100-household 'community' (see Table 5.15).

8.3.2.1 Electricity Demand

The developed electricity demand model allows for different levels of input data and will probabilistically allocate unknown characteristics (see 4.3). Using the 100-household nationally representative set of households as a basis, the level of known information was varied to determine the predicted average demand range over 100 model runs. Five levels of known input data were analysed:

1. Location, house size, and deprivation decile ('Location')
2. As 1. plus household type ('Type')
3. As 2. plus household composition ('Composition')
4. As 3. plus employment and income status ('Income/Employment')
5. As 4. plus major appliances owned ('Major Appliances')

The results in Table 8.2 show the average per-household results, where the average household demand per run is assessed relative to the average for all equivalent model runs. For example, the minimum 'Location' result is the average of the lowest model-predicted relative demand per-household for all 100-households over the 100 'Location'

Table 8.2
Range of predicted average electricity demand for individual households relative to the mean prediction for 100 nationally representative households, over 100 model runs, by level of known household characteristics.

	Location	Type	Composition	Income/Employment	Major Appliances
Minimum	0.328	0.436	0.450	0.466	0.499
5th Perc.	0.430	0.544	0.550	0.560	0.593
95th Perc.	1.785	1.650	1.634	1.630	1.591
Maximum	2.651	2.557	2.429	2.363	2.285
Std. Dev.	0.456	0.391	0.373	0.363	0.340

model runs. The distribution of results is broadly normally distributed but left skewed with an extended tail at higher values, therefore the standard deviation shown is a comparative measure that is unsuitable for prediction.

The results indicate that the behavioural uncertainty for single households, as indicated by the residual uncertainty for a household where all characteristics including major appliances are known, is a factor of two on the low side (0.499) and slightly higher than a factor of 2 (2.285) on the high side. The additional uncertainty, where the household characteristics are unknown, is approximately one-third of the behavioural uncertainty. For individual households, therefore, behavioural uncertainty is three times more significant than the maximum associated with unknown household characteristics for average demand assessment.

The results presented above were the average for the 100 modelled households. For the 100 'Location' model runs, the lowest value for any household was 0.17 and the highest was 6.20, confirming the analysis of Haldi and Robinson [79] that behavioural variation was typically a factor of two but could significantly exceed this in extreme cases.

Table 8.3
Range of predicted average total electricity demand for 100 nationally representative households relative to the mean prediction, over 100 model runs, by level of known household characteristics.

	Location	Type	Composition	Income/Employment	Major Appliances
Minimum	0.900	0.920	0.910	0.919	0.940
5th Percentile	0.916	0.931	0.927	0.936	0.943
95th Percentile	1.081	1.068	1.054	1.052	1.070
Maximum	1.125	1.129	1.105	1.120	1.101
Standard Deviation	0.049	0.041	0.039	0.039	0.037

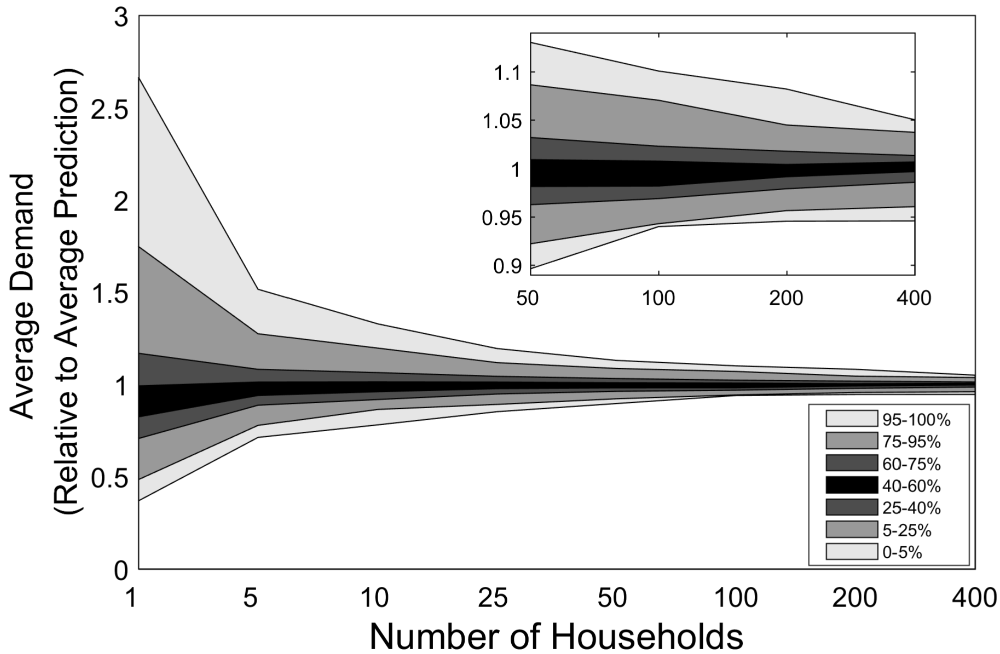


Figure 8.1. Percentile range of model predicted average electricity demand relative to the mean prediction over 200 model runs.

Increasing the scale of analysis to the total demand for the 100 households, and performing the same analysis, generates the results shown in Table 8.3. The relative contributions of behaviour- and household characteristics-driven uncertainty remains similar with approximately a three-to-one ratio. This ratio remains consistent up to 400-household systems, at which point the overall influence of uncertainty has diminished.

Figure 8.1 shows the impact of increasing scale on the distribution of results for the case where major appliance ownership is known. This analysis is based on 200 runs of the same 100 nationally representative households, with multiple runs randomly combined for the 200- and 400-household cases. For the less than 100-household cases, the sets of households are determined by selecting a fixed and representative sub-group from the 100-household set.

The results indicate that up to at least 100 households the uncertainty remains significant with a ±10% variation in average demand, which as a minimum would impact economic assessment of the system. As will be reviewed further in 8.5, this level of uncertainty translates to significant time-dependent variations in critical design parameters for distributed generation, such as solar matching and impact on the wider grid if connected to a constrained low voltage sub-system.

8.3.2.2 Hot Water Demand

The equivalent results to those in the above section for hot water demand are shown in Tables 8.4 and 8.5.

Table 8.4
Range of predicted average hot water demand for individual households relative to the mean prediction for 100 nationally representative households, over 200 model runs, by level of known household characteristics.

	Location	Type/Composition	Income/Employment
Minimum	0.175	0.211	0.216
5th Percentile	0.291	0.334	0.341
95th Percentile	2.241	2.058	2.015
Maximum	3.426	3.142	3.099
Standard Deviation	0.638	0.565	0.549

The results show that hot water demand uncertainty is higher than the electricity demand equivalent. This is to be expected as electricity demand uncertainty is reduced by the impact of constant-use appliances, and multiple appliances and demands that

Table 8.5
Range of predicted average total hot water demand for 100 nationally representative households relative to the mean prediction, over 100 model runs, by level of known household characteristics.

	Location	Type/Composition	Income/Employment
Minimum	0.891	0.868	0.868
Maximum	1.121	1.144	1.145
Standard Deviation	0.052	0.057	0.059

allow for a range of use behaviours. In the most extreme case for single household demand, with only 'Location' information known, the lowest relative value is 0.0721 and the highest is 5.451.

8.3.2.3 Average Demand Uncertainty Assessment Summary

Whilst average demand uncertainty alone does not provide information that can directly influence design decisions, as opposed to time-dependent demand uncertainty which is addressed in 8.5. It does provide a clear and quantifiable assessment of the potential impact at different scales and a justification for further detailed analysis for distinct system types up to at least 200-300 household systems.

The results also indicate that analysis at the individual household level, that does not account for at least a factor of two variation in overall energy demand, or the likely more significant time-dependent variations, risks poor performance resulting from the more extreme behaviours identified from the data and the model output.

8.4 Peak Demand Prediction

Peak demand assessment for an energy system is a critical element in system design, driving the sizing of the distribution network. For electricity demand, the primary sources of design guidelines are currently those used by the individual electricity companies (for example, [64] and [65]), based on experience and 'rules-of-thumb'. For hot water demand, as detailed in 1.6.1, several standards have been developed for hot water peak demand (diversity) prediction in district heating networks. However, the underlying analysis is typically not from UK demand data, and the source is either undefined or based on relatively small measurement campaigns. In contrast, the developed model allows a more detailed statistical assessment derived from UK occupancy and demand calibration data.

For the electricity demand model, no validation for peak demand replication performance was detailed in previous chapters, and is therefore included in this section. For the hot water demand model, the peak demand prediction performance was compared to the calibration dataset in 6.5.6, and close replication was shown. Without additional independent hot water data, no further validation is included in this chapter.

8.4.1 Electricity

Two factors can be used to assess the peak demand potential for an electricity network. The sum of the maximum peak demand for each individual household, typically known as the non-coincident maximum demand (NCMD), and the total instantaneous peak demand for all households, typically known as the 'After Diversity Maximum Demand' (ADMD). The former is a theoretical maximum total demand that is never approached in reality, and is used here primarily for further model validation, and the latter an assessment of the maximum total demand that can be expected for a multi-consumer network, which is a key measure for system sizing.

The HES dataset, with short measurement periods and limited measurement period overlap per household, cannot be used for reliable peak demand analysis. The Richardson et al (2008 data only) [69], REFIT [45], and Ashton Hayes (see 5.14.6) [173] datasets, with overlapping and longer measurement periods, have therefore been used for this purpose.

Of the existing models, the Richardson et al model is again the only open-source, UK data calibrated model that allows this type of analysis. However, as demonstrated in Chapter 3, the statistical basis of both the occupancy and appliance use models is insufficiently differentiated and probabilistically calibrated to provide reliable assessments over the extended period models required for peak demand assessment.

8.4.1.1 Non-Coincident Maximum Demand

Dataset Validation

For the 17 (out of 22) households in the Richardson et al dataset with consistent data and no evidence of secondary or water heating, the NCMD was 190kW. The

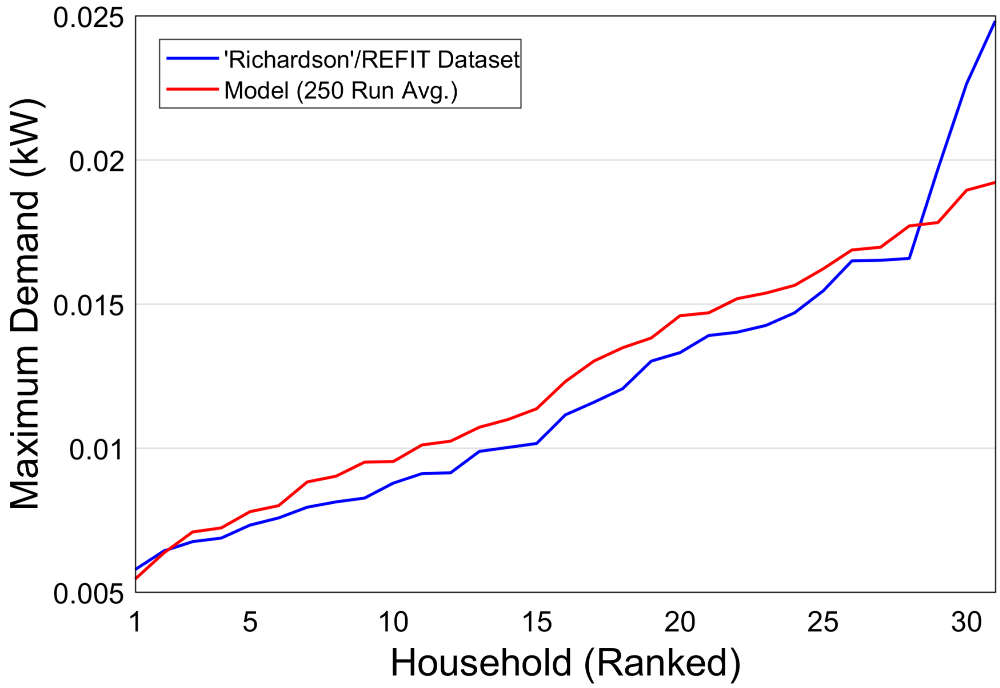


Figure 8.2. Ranked maximum electricity demand per household comparison for 'Richardson' and REFIT datasets ([69] and [45]), and dataset-equivalent models.

average equivalent for the model output over 250 runs was 196kW, with a range from 177kW to 234kW. Similar analysis for the 14 households (out of 20) in the REFIT dataset with no evidence of electric heating use, determined a NCMD of 188.4kW for the dataset and a model output average over 250 model runs of 181.6kW, with a range of 140.1kW to 217.3kW. There is no indication, therefore, that the model inaccurately predicts the average maximum demand per household for multi-household systems.

Analysis of the range of individual household maximum demand values from the same exercise, with the results interpolated into a 17-quantile ('Richardson') and 14-quantile (REFIT) distribution respectively, and compared to the actual maximum demand per household from the two datasets, generated the results shown in Figure 8.2. The results indicate a tendency for the model to slightly overestimate maximum demand in most households and to underestimate extreme values. The overestimation is potentially the result of the current model basis that does not restrict simultaneous use of different appliances and may, consequently, allow occasional periods of unrealistically high demand. However, in general, the model replicates the distribution with sufficient accuracy.

As the main use of this measure is for validation, with no direct influence on system design, no further NCMD analysis has been undertaken.

8.4.1.2 After Diversity Maximum Demand

Dataset Validation

After Diversity Maximum Demand (ADMD) is a measure of the maximum demand in an electrical network of multiple independent consumers, typically stated in kW per household. As the number of households increases the probability of coincident demand per household falls exponentially. For UK systems, ADMD design guidelines are typically electricity company specific, vary significantly, and are not scale-dependent. For example, for systems without electric heating: Eon: 2kW per household up to 4 bedrooms and 0.5kW for additional bedrooms [64]; and SP Energy: 1kW for non-detached 3-bedroom or smaller households, 1.5kW for detached 3-bedroom and 4-bedroom households, and 2kW for 5-bedroom or larger households, plus 8kW per system [65].

A more detailed study by Barteczko-Hibbert [181] reviewed the ADMD distributions for small populations (up to 100 households) based on Mosiac socio-economic indicators, and also for households with solar panels and electric vehicle charging. The analysis determined that at the 100-household scale, there was an average ADMD of 1.56kW, with lower values (c. 1.2kW) associated with older and lower income groups, and higher levels (c.2kW) associated with younger and suburban households. It is therefore clear that household characteristics potentially impact this measure for small-scale analysis. An additional conclusion from the same study was that as the population becomes smaller, in addition to the ADMD increasing exponentially, the degree of uncertainty in the predicted value increases. There is therefore potential design risk with using the typical or 'rule-of-thumb' type design rules identified above, particularly at the sub-100 household scale where the value is also highly scale-dependent.

For the 17 selected households in the Richardson et al dataset, the ADMD was 44.7kW or 2.63kW per-household. For 250 equivalent model runs, the ADMD was 45.7kW or 2.69kW per-household, with the measured data ADMD being in the 47th percentile of the model results.

For the 14 selected households in the REFIT dataset, there is overlapping data for approximately 9 months out of 11 months' total measurement duration from June to April, with the gaps spread randomly. The ADMD for the overlapping periods is 37.9kW. For 250 model runs for the equivalent households and period, the ADMD was 42.5kW, with the actual measurement being in the 18th percentile of the model results.

For the Ashton Hayes dataset (see 5.14.6.3), the maximum demand in the mid-May to August period (to discount the impact of secondary heating at other times) was 82.3kW on a 10-minute average basis. For the equivalent model over 100 runs the average result was 86.5kW, with a range from 72.4kW to 101.7kW. As stated in 5.14.2.2, the potential underestimate of holiday absences may account for a degree of overestimation.

Comparing the model output directly with the results from Barteczko-Hibbert [181] is difficult as it is not clear what proportion of households in that study have electric primary or secondary heating. However, the study includes analysis of homes with heat pumps, where the heat pump diversity is considered separately from the remainder of the electricity demand, and therefore primary electric heating influence can be discounted and it is assumed that secondary electric heating use is also low for these households. This group has a non-heat pump ADMD at 100 houses of 1.31kW with the model predicting an average of 1.27kW for groups of 100 nationally representative households.

In general, therefore, the model output for ADMD shows good correlation with multiple independent datasets, with no obvious consistent under- or overestimation and all measured data within the range of the model output.

Analysis

To determine the influence of area characteristics on peak demand, the nationally representative 100-household set model (see Table 5.15) based on the lowest and highest deprivation (IMD) decile area characteristics were run 250 times, with identical numbers of bedrooms and floor areas for each run, and with the household characteristics allowed to vary probabilistically. The lowest decile set is simulated as all private housing ('Lowest-PR') and the highest is all social housing ('Highest-SC') to represent extreme but realistic scenarios for this scale of system. The ADMD results are as shown in Table 8.6, with the smaller household set results generated by selecting fixed and representative subsets of the 100-household output.

Table 8.6
Model predicted After Diversity Maximum Demand (ADMD) average and maximum per-house values for different numbers of households for highest (IMD1) and lowest (IMD10) deprivation decile locations.

Households		10	25	50	100
Highest-SC	Avg. Demand (kW)	3.9	9.9	19.9	39.7
Lowest-PR	Avg. Demand (kW)	4.7	11.5	22.5	45.5
Highest-SC	Avg. ADMD (kW/house)	3.18	2.00	1.52	1.23
Lowest-PR	Avg. ADMD (kW/house)	3.35	2.09	1.65	1.32
Highest-SC	Max. ADMD (kW/house)	4.25	2.53	1.85	1.43
Lowest-PR	Max. ADMD (kW/house)	4.46	2.69	1.88	1.48

The results indicate a lower household characteristics influenced variation than the analysis of Barteczko-Hibbert, even if the lack of secondary heating use is considered. This suggests that the Mosiac-driven analysis is not representative of actual communities, which would have a mix of different household characteristics that are influenced, but not determined, by the location.

Based on this analysis, the 1kW assumption for 3-bed non-detached or smaller households from Eon [64] would potentially underestimate maximum diversity. Alternatively, the SP Energy basis [65] would potentially overestimate for systems of 50 households and above (allowing for an additional 0.25kW for secondary heating for 100-households based on comparison of [181] heat pump and standard heating system households, and 0.3kW for 50-households on a pro-rata basis).

Further analysis of the ten deprivation (IMD) decile 100-household sets analysed in 5.14.6.4, also shows a distinct but not overly significant influence on ADMD of the energy system socio-economic characteristics. This suggests that ADMD is primarily a function of the number of households and not overly influenced by area characteristics in realistic multi-household systems.

8.4.1.3 Peak Electricity Demand Assessment Summary

Without the inclusion of a secondary heating module, the developed model cannot currently be used for detailed assessment of maximum electricity demand. Further work to integrate a secondary heating model would therefore significantly enhance its capability in this area. The presented results, however, indicate that the model presents an alternative design method to the simple 'rules-of-thumb' currently used by electricity companies. As highlighted, there is significant scale-dependency and uncertainty in peak demand for sub-100 household systems that is not captured by these simplistic design guidelines.

8.4.2 Hot Water

The existing hot water diversity (peak demand) design standards make little or no differentiation based on the characteristics of the connected households. The literature on diversity is a combination of country standards with no explicit definition of how they were calculated and a number of independent sets of analysis, typically based on measured consumption on a relatively small scale. The significant variation in the existing standards and published distributions (see Figure 1.7) for smaller systems suggests either distinct localised behaviours or an uncertainty in predicted diversity that is not captured by a single, deterministic assessment.

The only open-source hot water model, developed by Vajen and Jordan [149], would require significant additional user calibration to be potentially useful as an assessment tool for system diversity and is principally used to generate typical single day use profiles. The developed model allows an assessment of the stochastic basis that underpins diversity analysis, and allows are more detailed assessment of the impact of household characteristics and outlier behaviours.

The following section first addresses the model predicted variation in diversity from differences in house type and household characteristics. In addition, a comparison between the model output and the current diversity standards identified in 6.5.6 has been undertaken, with additional focus on the timescale of analysis, and the predicted frequency and duration of periods where hot water use exceeds the design standard.

To provide 1-second diversity analysis, the 1-minute model output has been converted to a 1-second basis by starting hot water cycles randomly within the minute where a use is predicted, determining a flow randomly between 2.5 and 5 litres/min for <3 litre volume cycles, and 5 and 12 litres/min for other cycles, as detailed in 6.5.6, and determining the duration based on the cycle volume and flow.

8.4.2.1 House and Area Type Analysis

Average House Size

The average UK house has approximately 2.7 bedrooms [182], but with significant variation between different areas. Analysis of UK Census data at the LSOA level (typically 600-1000 households) shows that at this scale the variation is from 1.4 to 4.4 bedrooms, with a 5th and 95th percentile of 2.07 and 3.37 respectively [103], and this variation would be expected to increase as the scale of area is reduced. Varying the average number of bedrooms into 6 potential ranges for groups of 50 representative houses for each range in a 6th deprivation (IMD) decile area, the average 1-second diversity over 1000 model runs varies as shown in Table 8.7. (1000-run average is used for comparative analysis as the maximum predicted value varies inconsistently at this level of analysis)

Table 8.7
Variation in average modelled hot water diversity (1-second basis) for 50-household systems over 1000 runs by average number of bedrooms.

Average Bedrooms	1.4-1.6	2.2-2.4	2.4-2.6	2.6-2.8	2.8-3.0	3.0-3.2
Average	5.96	6.47	6.58	6.61	6.65	6.76

The impact on diversity is relatively small for mixed house size communities (i.e.

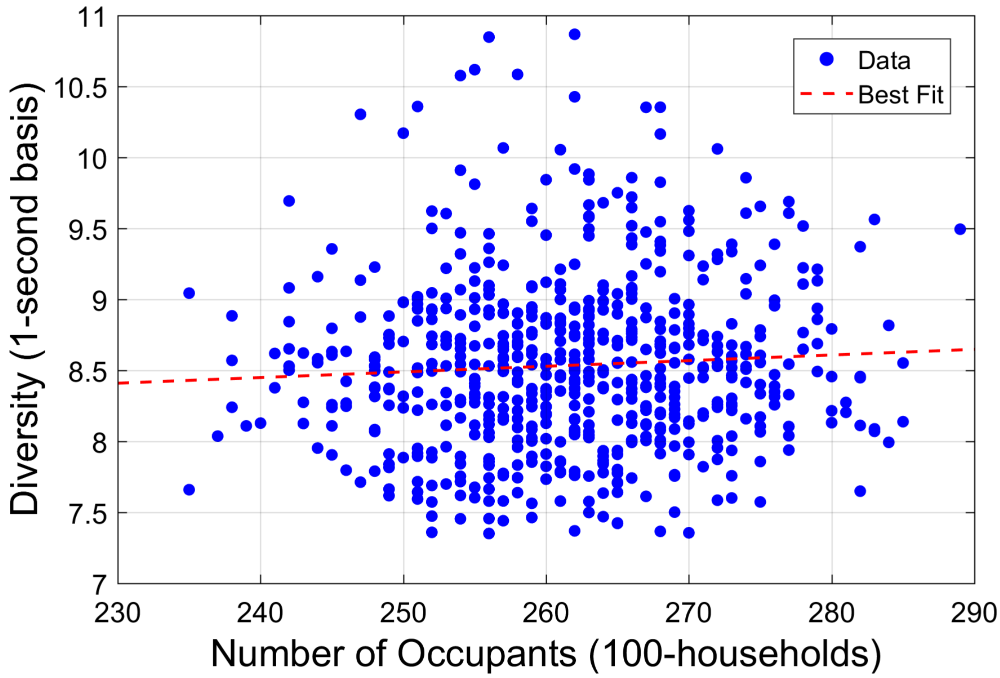


Figure 8.3. Variation in modelled hot water diversity (1-second basis) for a fixed nationally representative 100-household system over 1000 model runs by total number of occupants.

the 2.2 to 3.2 range) primarily as the average number of occupants per bedroom falls with an increasing number of bedrooms. The influence is more marked for less mixed areas, such as the predominantly small house type range example (1.4-1.6).

Number of Occupants

At the 600-1000 household LSOA level, the average occupancy per household varies from 1.4 to 4.8, with a 5th and 95th percentile of 1.96 and 2.88 respectively [183], therefore significant variation can be expected at smaller scales.

For a fixed set of house sizes, the number of occupants can vary significantly. By fixing the number of households and bedroom number, but allowing the total number of occupants to vary probabilistically, allows the influence of people number on diversity to be assessed. Figure 8.3 shows the distribution of total occupant numbers and predicted 1-second diversity values from 1000 runs of the hot water demand model based on the nationally representative 100-household set (see Table 5.15) in a 6th deprivation (IMD) decile area. The best-fit line indicates that while occupant number is predicted to influence diversity, the impact is small (c. 0.004 per additional person), particularly in comparison to the significant overall variation.

Area Type

The influence of the location socio-economics or community type on hot water diversity is not accounted for in diversity standards. However, it would be expected that higher or lower average use would also translate to differences in maximum instantaneous demand for a network, with further analysis required to determine if this variation is significant.

Comparing hot water model output over 1000 runs for extreme highest and lowest deprivation (IMD) decile area types, the results are shown in Table 8.8. The lowest decile set is all private housing ('Lowest-PR') and the highest is all social housing ('Highest-SC'), and both are based on the nationally representative set of households (see Table 5.15). In all cases the total number of occupants for each number of households was restricted to ±10% of the national average of 2.3 people per household to allow some variation but ensure occupant number was not a significant factor. Across all sizes of community, the difference between the two area types was approximately 10%.

Analysis Summary

The overall conclusion from the analysis based on number of bedrooms, number of occupants, and area type, was that the primary influence on diversity is number of households. The variation due to different household and area characteristics for a fixed number of households is relatively small, particularly in comparison with the probabilistic variation due to differences in individual household behaviours, and would not require significant adjustment of the design standards to account for this potential.

8.4.2.2 Existing Standards Analysis

DS439 Comparison

The Danish DS439 hot water diversity standard has been recommended for use in the UK by CIBSE, replacing the British Standard BS6700 which was deemed to significantly overestimate diversity. Table 8.8 shows a comparison between the DS439 and BS6700 predicted diversity values and results for two extreme location models based on a nationally representative mix of house types and sizes. The analysis is based on the maximum diversity on a 1-second basis.

Table 8.8
Variation in average and maximum modelled hot water diversity (1-second basis) for different numbers of nationally representative households for a highest deprivation decile area comprising all social housing ('Highest-SC') and lowest deprivation decile area comprising all private housing ('Lowest-PR'). 'DS439' data from [56] and 'BS6700' data from [59].

Households		10	25	50	75	100
DS439		2.37	3.76	5.59	7.17	8.64
BS6700		4.47	9.07	14.40	19.49	24.53
Highest-SC	Average	3.45	4.65	5.97	7.03	7.93
Lowest-PR	Average	3.73	5.06	6.57	7.80	8.91
	PR/SC Ratio	1.081	1.088	1.101	1.110	1.124
Highest-SC	Maximum	5.11	7.22	8.32	9.49	12.75
Lowest-PR	Maximum	5.38	7.07	8.65	10.49	12.65
	PR/SC Ratio	1.053	0.979	1.040	1.105	0.992

The results indicate that, up to 100 households, the model predicted average diversity over 1000 model runs exceeds the DS439 basis for all but the 75- and 100-household cases for the lower demand characteristics population. For all cases, the maximum predicted diversity significantly exceeds the DS439 basis. The 'error' becoming more pronounced as the number of households is reduced.

What is not clearly defined in any of the current standards is the basis for the predicted value in terms of time basis or an allowable tolerance for short periods where the demand exceeds the design basis. The following section address these two parameters.

Time Basis

The analysis presented above assumed a 1-second basis for the diversity analysis. Within the constraints of the model output and computational requirements, this represents instantaneous diversity. However, it is not clear if this is the basis used for the diversity standards, or whether this should be the basis for district heating analysis if the dynamics of the system are considered.

Table 8.9
Variation in maximum modelled hot water diversity by number of households and diversity time basis over 1000 model runs for lowest deprivation decile area comprising all private housing ('Lowest-PR'). 'DS439' data from [56].

Households		10	25	50	75	100
DS439		2.37	3.76	5.59	7.17	8.64
Lowest-PR	Max. (1s)	5.38	7.07	8.65	10.49	12.65
Lowest-PR	Max. (10s)	5.34	7.04	8.15	9.93	11.16
Lowest-PR	Max. (60s)	3.69	4.90	5.48	6.50	8.35

Table 8.9 shows the impact on maximum predicted diversity over 1000 runs of increasing the time basis to the average diversity over 10-second and 60-second rolling periods using the higher demand characteristics population ('Lowest-PR') (see above). Increasing the time basis reduces the predicted maximum diversity markedly. However, even with a 60-second basis, the potential underestimation if using the DS439 standard basis for sub-50 household systems remains.

Tolerance

Further analysis of the 1-second results highlights that the maximum diversity value is often a significant outlier result and therefore considering the absolute maximum value without also considering the duration of periods where the system demand may exceed the design standard is potentially misleading. Table 8.10 shows both the average and maximum per-run number of occasions ('events') and total duration when the model predicts that the DS439 basis is exceeded for nationally representative sets of households over an annual run. Again, the model predicts that the DS439 basis has the potential to be significantly exceeded on this basis for at least up to 25 households, although by 25-50 households the number of events and their duration have reduced to levels that may not be discernible to users.

Table 8.10
Average and maximum number of annual events and total duration where the predicted model output diversity on a 1-second basis exceed the DS439 basis for different numbers of nationally representative households.

Households		10	25	50	75	100
Events	Avg.	110	56.1	15.3	6.5	2.8
Events	Max.	307	178	52	22	16
Duration (mins)	Avg.	24.6	9.7	2.0	0.81	0.25
Duration (mins)	Max.	81.9	40.6	9.0	3.7	1.9

If we allow a small degree of tolerance for short periods of above-diversity demand, the predicted diversity values reduce significantly. Table 8.11 shows the results for the 'Lowest-PR' household set (see above), based on the maximum predicted instantaneous (1-second) value, and an allowance of 1 minute and 5 minutes annually where the demand exceeds the stated diversity value. The results again indicate that the DS439 basis potentially underestimates diversity at the lower end of the analysed household number range even when some flexibility for oversupply is allowed.

Table 8.11
Maximum model predicted diversity results with different levels of tolerance for the total annual duration for which the stated basis can be exceeded. 'DS439' data from [56].

Households	10	25	50	75	100
DS439	2.37	3.76	5.59	7.17	8.64
Diversity - Max.	5.38	7.07	8.65	10.49	12.65
Diversity - 1-minute exceeded	4.16	5.36	7.00	9.24	10.16
Diversity - 5-minute exceeded	3.61	4.63	5.79	6.98	8.00

8.4.2.3 Hot Water Diversity Assessment Summary

It is beyond the scope of the presented work to determine a suitable basis for diversity assessment, but clearly both the diversity time basis and tolerance of short periods of over supply need to be better defined. From the other perspective, system designers also need to better define the time basis that is critical for the system dynamics.

The results highlight that considering the dynamics of the system and user tolerances are important. The type of probabilistic, high-resolution assessment presented potentially allows for a better understanding of the system characteristics, particularly in comparison with existing deterministic assessments based on country-specific standards with limited or no calibration basis definition or measured results from small datasets that may not be representative or of sufficiently high resolution. The model output, for example, could be used to generate representative data for dynamic district heating system models for better assessment of the impact of short periods of demand that exceed the design diversity value.

Whilst the DS439 basis is clearly more accurate than the previous UK BS6700 basis. There is evidence, as a minimum, that its use in the UK should be carefully monitored for evidence of underestimation resulting in poor performance, and that its predicted basis for sub-50 household systems should be treated with caution.

8.5 Time-Dependent Optimisation Analysis

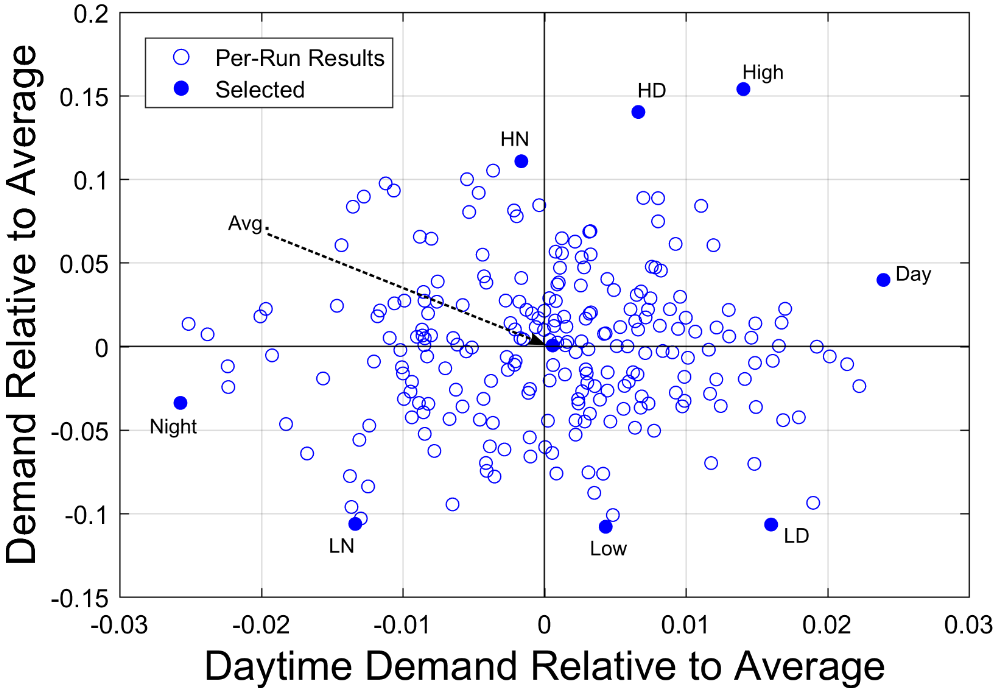
8.5.1 Representative and Extreme Case Modelling

The average demand uncertainty analysis detailed in 8.3.2 indicated that a significant degree of uncertainty, with respect to potential impact on energy system performance, remains for systems of at least 100 households. Furthermore, the time-dependency of the uncertainty is more directly linked to system performance, particularly for systems where supply and demand matching is critical, and is likely to significantly exceed the observed average demand uncertainty.

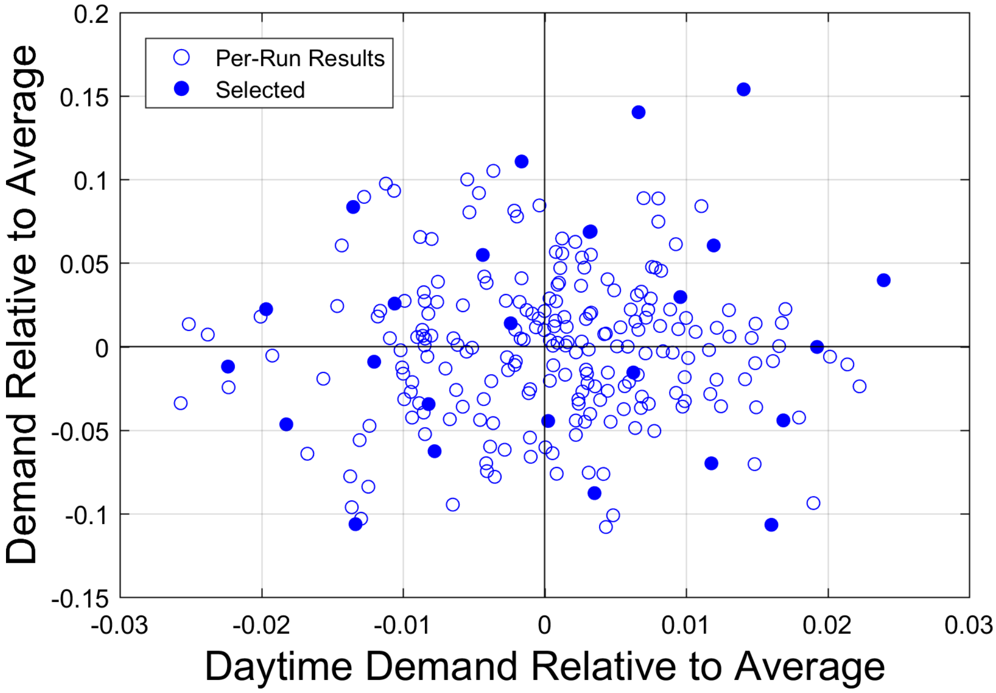
As shown in 7.5.5, highly probabilistic models must be run a sufficient number of times to obtain a representative distribution of results. However, in most cases further analysis based on the overall set of results is both impractical and unnecessary, and a filtered, representative subset can be used.

Two types of filtered assessment using the demand model output were therefore defined; a targeted 'stress-test' that determines the impact of the most extreme cases and a more refined assessment based on a representative, equally probable range of potential demand cases. These methods allow the predicted average demand uncertainty determined in 8.3 to be used as the filtering basis for more detailed analysis, assuming a close correlation between average demand and more refined time-dependent assessments can be shown.

To assess each model run, the primary variable is the average per-run demand



(a) Extreme subset



(b) Representative subset

Figure 8.4. Representative and extreme filtered subsets from 250 annual duration electricity demand model runs for the same 100-household nationally representative energy system.

relative to the average for all runs. Secondary relative assessments can also be added dependent on system type, such as daytime demand percentage for solar and wind powered systems, or overlap with heating and hot water use for CHP systems. The distribution of results based on a primary and secondary variable can be represented visually as shown in Figure 8.4.

For the extreme 'stress-test' case, nine results are selected; average, highest and lowest average ('High' and 'Low'), highest and lowest daytime proportion ('Day' and 'Night') plus the point furthest from the average in each quadrant that is not already selected ('HD', 'LN', etc.).

Representative samples can be selected using a variety of techniques, including the Kennard-Stone method and a method that uses a similar approach to the Kernel Density method used in this work (see Appendix A). In this case the Kernel Density method is used based on the *MATLAB* function, *kspxy*. As shown, if a sufficient number of points are selected, the majority of the 'extreme' cases are also included. Analysis of the minimum number of filtered cases required for a representative sample is reviewed below.

8.5.2 Analysis Examples

Annual demand at a 1-minute resolution was generated for three sets of 100 households; the nationally representative set (see Table 5.15) based on the 6th deprivation decile ('National Average'), and the 'Lowest-PR' and 'Highest-SC' sets identified in 8.4.2.1. 250 separate model runs were generated with only the house size, type, and tenure specified, and household characteristics determined probabilistically for each model run using the developed sub-model (see 4.3). This analysis therefore determines the maximum uncertainty based both on unknown characteristics and behaviour. For the equivalent 50-household system analysis, a fixed and representative selection of the 100-household data is used, and for the 200-household system analysis, two randomly selected 100-household sets are combined. The results have been filtered into the 'extreme' and 'representative' subsets as outlined.

Two types of analysis have been undertaken. The first is a solar generation and storage system sizing analysis based on utilisation targets for three community types using the average, minimum, and maximum cases from the 'extreme' filtered subset. The second is a full matching analysis for a simple solar energy system using the fully representative filtered datasets as outlined.

Both analyses utilised the energy system optimisation program, *Merit* [184], which allows a detailed assessment of the performance of selective case sampling and provides an indication of the potential for deviation from average system performance because of the demand variation and uncertainty predicted by the model. *Merit* allows the demand profiles at high resolution to be matched with generation from solar panels, wind turbines, and other sources, the impact of integrated storage to be assessed, and for both the need for a grid connection and the extent and timing of the import and export flows to be determined. The standard version of *Merit* has a 60-minute resolution but this has been modified for this work to allow comparison using the 1-minute resolution demand model output.

These type of design assessments for small-scale systems are complex undertakings, with optimisation criteria typically being project-specific. A review of published research in this area demonstrates the assertion in Chapter 1 that the occupancy and demand influence is the weakest element in optimisation analysis. Analysis is typically based on very limited monitoring data (e.g. [185], [186]) or generic demand examples integrated with the optimisation software [184], with probabilistic approaches being limited to simple statistical modelling of potential uncertainty (e.g. [187], [188]). The developed model offers a different approach, with a highly probabilistic, high-resolution demand prediction based on behaviour differentiated calibration.

Table 8.12
Assessment of solar and battery storage required for net supply and demand balancing and 50% solar utilisation for three demand cases and three location characteristic types.

Case	Type	Highest-SC	National Average	Lowest-PR
Lowest	Solar(kWp)	290.8	304.1	358.3
	Battery(MAh)	3.45	3.63	4.30
Average	Solar(kWp)	337.0	341.2	389.6
	Battery(MAh)	4.08	4.16	4.77
Highest	Solar(kWp)	370.2	393.5	426.8
	Battery(MAh)	4.48	4.78	5.21

8.5.2.1 Generation Equipment and Storage Sizing

To demonstrate the potential of the developed model for equipment sizing; the lowest, average, and highest average demand cases were assessed for generation and storage capacity required to achieve a balance between total supply and demand, and 50% utilisation of the solar panel output. This was done for the three identified sets of households ('Highest-SC', 'National Average', and 'Lowest-PR'). The results are shown in Table 8.12.

The results demonstrate a significant potential variation between the extreme uncertainty driven cases ('Highest' and 'Lowest') and lesser but still significant differences based on location characteristics. Design rules for smaller-scale systems that are based on limited data or experience do not allow the full variation in potential demand scenarios to be assessed, with the risk that the data sources are below-average examples or that the potential impact of outlier scenarios is underestimated.

8.5.2.2 System Balancing and Grid Connection

A simple solar system analysis (panels only with no storage) has been performed to indicate both the performance of the case filtering approach for balancing and grid connection analysis, and the capability of the model to perform probabilistic assessment of the potential operating scenarios. The solar array is fixed for all cases and sized such that the total annual generation matches the total annual demand for the model run closest to the average over the 250 runs. For the 250-run, 100-household case used, it is predicted that on average 43.8% of the generated electricity will be used by the households and the remainder balanced from the grid.

Table 8.13 shows the balancing and grid connection results for the 'extreme' analysis case. The grid export total varies from 91% to 107%, and import from 87% to 119%, compared to the average predicted values, which is a significant variance. The equivalent export values for equivalent 50- and 200-household sets are 88% and 108%, and 93% and 105% respectively. The equivalent values for grid import are 80% and 128%, and 90% and 117% respectively. The system scale therefore has an impact on the predicted range, as would be expected, but with the uncertainty remaining significant at 200 households.

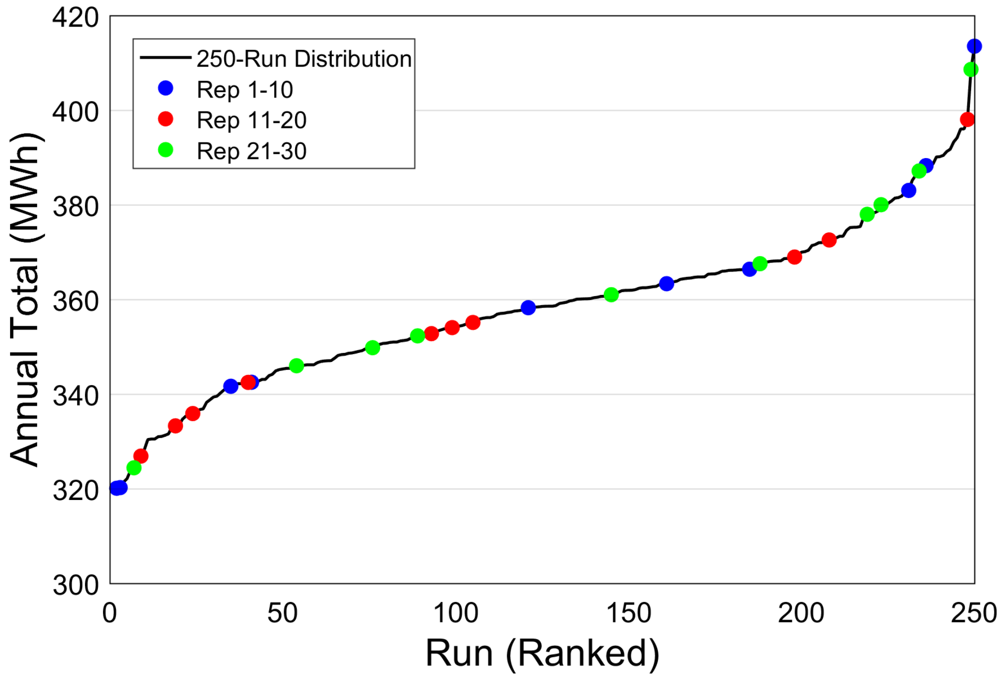


Figure 8.5. Representative cases by ranking range and position on average demand distribution.

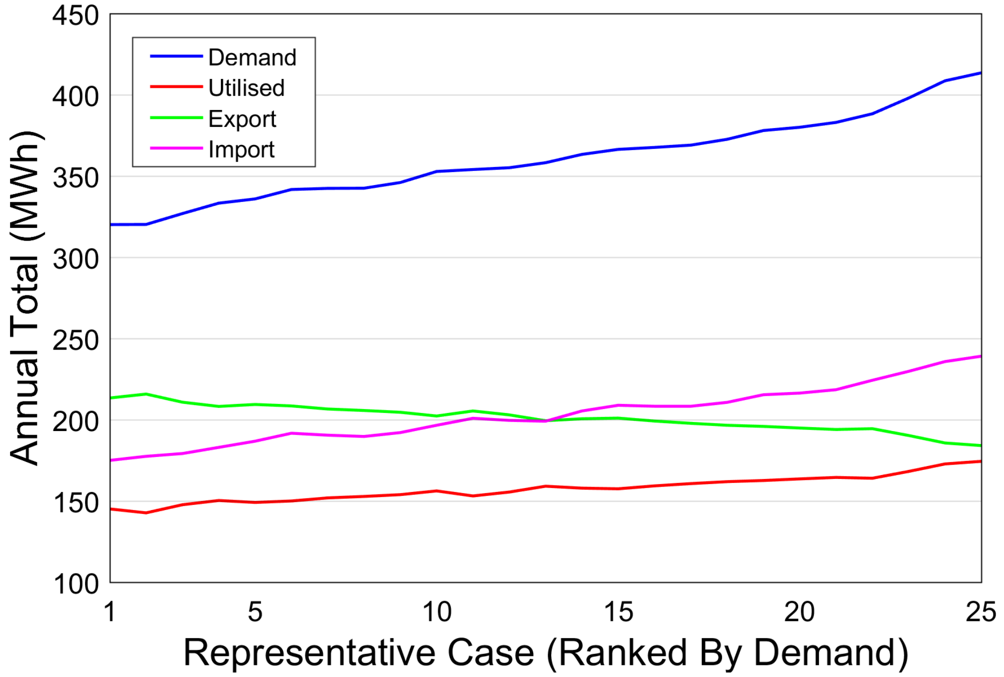


Figure 8.6. Range of total annual electricity demand plus total electricity utilised, imported and exported for 25 representative runs from 250 model runs for a 100-household solar panel supplied energy system.

For the more detailed representative analysis, first the optimum number of selected runs needs to be determined. Figure 8.5 shows the position of each representative case on the overall per-run demand distribution based on range position in the representative analysis ranking. This indicates that between 20 and 30 runs are required for an effective range of values, and therefore 25 runs (10%) was selected as the basis for a representative sample.

Table 8.13
Range of total annual electricity demand plus total electricity utilised, imported and exported for eight extreme cases and the average case over 250 model runs for a 100-household solar panel supplied energy system.

Case	Demand (MWh)	Utilised (MWh)	Export (MWh)	Import (MWh)
Average	358.5	157.1	201.4	201.4
High	413.5	174.4	184.1	239.1
Low	319.6	143.9	214.6	175.7
Day	372.6	161.9	196.6	210.7
Night	346.2	151.2	207.3	195.0
HD	408.6	172.8	185.7	235.8
HN	398.0	168.2	190.3	229.8
LN	320.2	142.7	215.8	177.5
LD	320.1	145.1	213.4	175.0

The results for the 25 selected representative cases are shown in Figure 8.6 in ascending order of predicted average demand. This gives an indication of the potential range of the predicted performance measures, allowing for a more detailed judgement of the system potential than if only the average or extreme cases are analysed. The results indicate that for a 100-household system, the variance in potential performance based on equally probable representative scenarios, is significant. Figure 8.6 also demonstrates that using average demand for initial screening is effective as each of the time-dependent assessments varies consistently with this simpler, easily determined measure.

8.5.2.3 Analysis Summary

Whilst the presented analysis is based on simple examples, and the degree to which the predicted variances are significant is situation-specific, the overall results indicate that for systems of at least 200 households, demand uncertainty would be an important design consideration, and would also significantly influence the system economics. Further analysis for specific types of community and different types of distributed generation system are required to fully understand and quantify the potential impact.

As outlined, no existing demand models accurately capture either the variation resulting from the location characteristics or the prediction uncertainty, that would allow the design risk to be assessed on a case-by-case basis or probabilistic design guidelines developed. Using the developed model for these types of assessment indicates a potential means to address this for future system development.

8.5.3 Other Design Elements

The analysis presented above is predominantly focused on the sizing and balance, storage, and grid connection design elements. The applicability of the developed model to the two remaining elements (demand management and seasonal matching) have not been reviewed in detail but are briefly discussed below.

8.5.3.1 Demand Management

The extent to which demand can be managed (i.e. shifted) is appliance-specific. For example, Zhu et al [189] identified three distinct classes of appliance with reference to this potential: non-shiftable (cooker, kettle, cold appliances, television, low-volume hot water use); time-shiftable (washing machine, dishwasher, dryer, high-volume hot water use); and power-shiftable (electric vehicles). Detailed analysis therefore needs to focus on individual household behaviours for the potentially shiftable demands.

As demonstrated in Chapter 7, the developed bottom-up approach with individual behaviour factoring allows a representative assessment of within and between household demand behaviours at the appliance-level. This allows both an assessment of maximum demand management potential and, critically, determination of the extent to which individual households would need to alter their behaviours to have a significant overall impact. Accurate assessment of realistic rather than maximum demand management potential also requires that the willingness of each household to make the necessary changes is also considered, which could be simulated with different factoring for each modelled household based on detailed analysis how this willingness varies. For different system scales, the developed model, as shown for other types of analysis, would also allow the degree to which demand could be shifted to be determined as a probabilistic range rather than an average assessment.

8.5.3.2 Seasonal Matching

Final validation of the developed model for seasonal variation is currently limited by the lack of a secondary heating module. For the other demands, a degree of seasonal factoring has been included for some appliances (kettles, microwaves, cookers) where a clear sinusoidal variation is discernible, and lighting use is directly influence by seasonal variations in external light levels.

However, there remains areas where additional work on identifying potential seasonal influences would be beneficial to model accuracy for this element. Analysis of the occupancy data did not identify any clear and consistent occupancy variations by month or season but further work in this area is required to determine if more complex seasonal influences on occupancy exist. This analysis should also extend to other appliances, particularly TVs and computers, where a degree of seasonal influence related to occupancy variations might be expected.

8.6 Model Applications Potential Summary

The stated aim of the model development was to provide a tool for comprehensive, differentiated, and probabilistic assessment of domestic demand at a high time resolution and over extended periods. This was to provide a means to assess five key design elements for small-scale energy systems with connected distributed generation (sizing and balancing, storage, grid connection, demand management and seasonal matching assessments). Existing models were determined to have limited or no calibrated differentiation for household characteristics, occupancy, appliance ownership and type, and energy use behaviour, and were therefore unsuitable for this type of analysis. The purpose of the analysis detailed in this chapter was to assess if the aim had been achieved.

The presented peak demand and overall system optimisation analysis demonstrated the potential for a probabilistic model that captures a comprehensive range of households by type and socio-economic factors, and also assesses the potential behaviour-driven variation. The developed model therefore has the capability to be used directly, or in parallel with limited measured data and existing deterministic guidelines, to assess design and operating risk that results from the overall prediction uncertainty. Using simple examples, the model was shown to provide detailed probabilistic analysis for sizing and balancing, storage sizing, and grid connection assessment elements. In addition, the individual behaviour enhancement detailed in Chapter 7 would allow the model to be used for appliance-level demand management analysis. Further model development and calibration would be required to extend the model capability to accurate seasonal analysis.

Whilst not directly comparable in terms of complexity and computational intensity, the presented analysis has demonstrated that existing simple, deterministic approaches, such as BREDEM/SAP for monthly demand assessments and electricity company guidelines for electricity network sizing, are overly simplistic and inaccurate when applied to specific types of small-scale system.

In addition to specific system analysis, the developed model also allows a higher-level assessment of the scale and impact of demand variation and uncertainty for different system types and sizes as also demonstrated in this chapter.

As detailed, three distinct elements related to potential variation and uncertainty must be considered to characterise demand for small-scale energy systems. The first is predictable variation as a result of the socio-economic characteristics of the connected households relative to the national average. This was determined to be approximately ±5-8% of the average predicted value and is not system scale-dependent.

The second is uncertainty resulting from unknown household characteristics, and the third is uncertainty because of potential individual household behavioural differences, both of which are scale-dependent. At the individual household level, the average behavioural uncertainty was a factor of two, with extreme cases well in excess of this. Unknown characteristics uncertainty adds an additional 33-50% to the total uncertainty if only location and house type are known.

For 100 households, the overall uncertainty reduces by approximately a factor of 10 to c.±10-15%, with the behavioural element falling by a more significant proportion. By 400-500 households the overall uncertainty, in terms of both average demand and more detailed system performance analysis, has diminished to less than ±5%, which is assumed to be generally within the scope of any correctly designed system.

It is outwith the scope of the presented work to determine for all types of system at which scale the uncertainty is within the typical design parameters of a system design or the system economics are not potentially compromised. What is clear, however, is that little work has been done previously to test small-scale system designs across a realistic range of design conditions and that the variations remain potentially significant, at least in terms of system economics, beyond 100 households. The main conclusion, therefore, is that the recommended optimisation and uncertainty analysis should be performed on systems up to c.500 households until a better understanding of the overall impact can be determined. The same conclusion would also apply to other types of demand-related analysis, such as the impact of casual gains on current and future low-carbon housing designs. As addressed, the developed model provides a means to perform this analysis either directly or in parallel with other existing methods.

8.7 Chapter Summary

This chapter detailed the analysis completed to determine the variation and uncertainty inherent in demand prediction at the small-scale and methods to assess potential small-scale energy system performance based on multiple results from a probabilistic model. The chapter highlights are as follows:

- Significant electricity and hot water demand variation for areas with identical housing can be expected based on the area socio-economic characteristics and age profile. The influence on peak demand is lower, with the number of households the primary determinant.

- At the household level, typical electricity demand behaviour-driven uncertainty exceeds a factor of two and for hot water demand exceeds a factor of three, with extreme examples significantly higher. At 100 households, average demand behaviour-driven uncertainty is c.±10-15% and at 400 households less than ±5%.

- In addition to the behaviour-driven uncertainty, uncertainty due to unknown household characteristics can add an additional 33% for single household analysis, and up to 50% for 100-household analysis.

- Hot water diversity is primarily driven by the number of households with less significant variations based on overall household characteristics. There is evidence that the Danish DS439 diversity standard basis, recommended for UK use, underestimates diversity for less than 50 household systems. Further discussion is also required on the timescale of the diversity assessment and the tolerance for short periods of over demand.

- Results from multiple probabilistic model runs can be filtered to representative or extreme selections by average demand for more detailed analysis. Small-scale system optimisation analysis using filtered data demonstrated that time-dependent demand uncertainty remains significant for at least 200 households and that further assessment for different types and scales of energy system is required to better understand when prediction uncertainty is no longer significant for design.

Chapter 9

Discussion

This chapter includes: a summary of the main contributions of the work; a discussion on the applications and limitations of the developed occupancy and demand models; a discussion on the implications of the work for distributed generation integration and small-scale energy system design, particularly in relation to system scale-, type-, and behaviour-driven demand uncertainty; suggested future work; and concluding remarks.

9.1 Contribution

The stated aim of this research project was the development of analysis methods and simulation tools to account for realistic occupant behaviour in domestic energy demand modelling. These should allow the influence of household characteristics and behaviours, and overall system scale to be accounted for within a probabilistic assessment of demand at high time resolution for small-scale (<1000 household) energy systems. The following outlines the main contributions of the presented work.

- ***Enhanced Domestic Occupancy Modelling Methods*** - Improvements to existing domestic occupancy modelling methods have been developed, including an effective higher-order Markov chain approach and a method to capture occupant interactions for couples and family households. In addition, the minimum calibration population size for effective modelling was assessed and significant occupant and day type differentiation based on this analysis used for occupancy model calibration. A performance comparison confirmed that the new methods, both individually and in combination, improved on existing approaches.

- ***Development of a Probabilistic, High-Resolution Electricity and Hot Water Demand Model*** - Existing bottom-up, high-resolution demand models were analysed and several potential areas for improvement identified, with the use frequency and timing prediction for intermittently used appliances and demands deemed the most significant. A new discrete-event method was therefore developed for intermittent demands, calibrated using high-resolution data. The developed approach was shown to be an improvement on the existing discrete-time models, particularly in the areas of identified weakness. In addition, enhanced versions of existing methods were used to model lighting, TV-use, always-on appliances, and miscellaneous low power appliances.

The developed model was shown to closely replicate the demand characteristics of both the calibration and independent datasets. The model also incorporated a significant number of additional probabilistic factors to account for variations in household demand resulting from differences in income, occupancy, and energy use behaviours. Further validation confirmed that the model could replicate the range of potential household demands for all but a small number with unusual patterns, and variations resulting from different area socio-economic characteristics.

- ***Critical Assessment of Group-Calibrated Occupancy and Demand Models*** - A critical assessment of the performance of Markov chain occupancy models calibrated using composite data from multiple individuals determined that significant convergence to the average group behaviour can be expected within the target model timescale (i.e. 1-year). Convergence was indicated by overly uniform timing of the main occupancy transitions (waking, sleep, etc.) and lower variance in average occupancy in comparison with equivalent real populations.

  A similar assessment of the demand model output, both overall and for each specific demand, confirmed convergence in demand timing associated with the occupancy model convergence and also as a result of the appliance cycle timing model calibration using multiple household data with only minor further differentiation. This convergence was highlighted by lower per-timestep demand variance between households for the model output in comparison with actual data.

- ***Development of Methods to Individualise Group-Calibrated Occupancy and Demand Models*** - Methods were developed to further refine both the occupancy and demand models to account for variations in individual behaviours relative to the group calibration average. For occupancy, the time basis of the Markov chain model was shifted and rebased for four key occupancy transitions. For demand, the cycle timing model was manipulated to account for different timing behaviours per household.

- ***Quantifying Demand Variation and Uncertainty in Small-Scale Energy Systems*** - Demand variation and uncertainty has been analysed using the developed probabilistic model for different types of small-scale energy systems. Demand variation because of different household and area characteristics has also been assessed. Uncertainty in predicted average, peak, and overall per-timestep demand has been analysed and quantified for different system sizes and for different levels of known household information. For distributed generation integration design risk, uncertainty in demand prediction was shown to remain potentially significant for systems in excess of 300-400 households.

- ***Creating Representative Subsets from Multiple Runs of a Highly Probabilistic Demand Model*** - A method has been developed to reduce the results from the significant number of individual model runs required for a representative output from the probabilistic demand model to useful filtered subsets for detailed analysis. Examples include representative and extreme subsets based on two key per-run demand variables.

9.2 Model Applications and Limitations

9.2.1 Occupancy Sub-Model

The Markov chain approach has been shown to remain an effective method for stochastic occupancy modelling from currently available data. An alternative discrete-event method was reviewed but shown to perform poorly in the key occupancy transition periods. The use of higher-order models was shown to be more effective for input dataset replication but will require significantly more input and validation data to determine if the improved performance is significant in relation to the overall prediction accuracy of this type of model.

The primary aim of the developed occupancy sub-model was to generate input occupancy data for a high-resolution, occupancy-driven energy demand model, with the objective to identify specific demand patterns for homogenous communities (e.g. retirement, social housing, commuter) and the specific influence of occupancy on any identified differences. Significant differentiation by occupant and day type improved the ability to capture individual behaviours. This was further enhanced by generating sequences of separately calibrated working and non-working day types per employed individual based on realistic variations in working patterns.

Further analysis determined an improvement in occupancy behaviour differentiation between occupant types but that behaviour convergence remained within each occupant type group. As a result, the Markov chain time basis has been manipulated to further account for individual occupancy behaviours relative to the group average.

The occupancy model output remains limited, both for calibration and validation, by the lack of large, multi-day occupancy datasets. The output is not suitable for comprehensive analysis at the single household level as occupants with unusual behaviours are not captured. However, the developed method is an improvement on existing models, and captures a broader range of distinct behaviours, particularly in relation to age and employment status, for the c.90% of individuals and households with typical diurnal patterns.

9.2.2 Demand Sub-Model

Calibration of existing demand models has typically focused on either a limited selection of household archetypes or large composite household groupings with minimal differentiation. Consequently, for real world assessment they do not adequately capture the breadth of household types and behaviours, and the output is not sufficiently representative to drive design decisions

The stated aim for the model development was to replicate the highly variable and individual nature of household demand, and to comprehensively capture different household types and specific demand behaviours associated with each type. In combination with the developed occupancy sub-model, the model allows distinct demand profiles for specific household types to be generated that can be used to provide improved demand prediction for specific types of communities in comparison with existing models.

Validation with both the calibration and independent datasets has shown that the developed model, particularly with the inclusion of additional individual household behaviour factoring for specific demands, is able to capture a wide breadth of the overall demand profiles from actual data. A small percentage of households (c.5-10%) exhibit unusual demand behaviours, which are beyond the capability of this type of model to capture. Therefore, the use of the model output for individual household analysis (i.e. heat gains, EV integration, house-scale renewables) requires a degree of caution. For comprehensive modelling, the developed model was shown to be broadly effective for systems with a minimum of 15-20 households in terms of capturing the overall range of expected demand patterns with a sufficient number of individual model runs performed to capture the variation and uncertainty predicted by the highly probabilistic model basis.

9.2.3 Model Applications

The development of a 1-minute resolution model with the aim to produce annual duration output is computationally intensive. However, current computing technology allows a 100-household electricity and hot water model to be completed in c.10 minutes on a standard 2013 Quad-core desktop, allowing multi-run analysis to be completed in a realistic timescale. Expected future improvements in computer technology will result in this type of high-resolution modelling becoming increasingly feasible for system analysis at all applicable scales.

In Chapter 8 the potential for the model to be used in different scenarios was assessed. It was shown to be applicable both as a research tool to quantify the degree of uncertainty highlighted by the highly probabilistic nature of the model and observed demand behaviours, and as a design tool for high-resolution modelling of small-scale energy system performance, particularly for highly time-dependent analysis such as system sizing, and matching of predicted demand with variable supply technologies.

The benefits of the probabilistic nature and high resolution are that they allow for different types of detailed analysis, particularly where the raw results are further filtered for representative subsets. In addition to average overall and per-timestep demand profiles, demand extremes and detailed distributed generation matching analysis can also be investigated. As an example, it was shown in Chapter 8 that the hot water sub-model results contradict the hot water diversity standard assessment for systems with less than 50 households, suggesting that this type of probabilistic model offers a complementary and potentially more robust method for assessing rare but significant demand extremes than analysis of demand in a small number of real buildings. This is discussed further in 'Further Work' below.

9.3 Implications for Small-Scale Energy System Design

Understanding and allowing for demand variation and uncertainty is small-scale energy systems is critical to the development of designs for distributed generation integration, to both low voltage networks and independent energy grids, that are sufficiently robust to account for all realistic scenarios. Analysis of the model output has shown that individual household average demand, even for households with identical characteristics, can vary by an order of magnitude for electricity demand and in excess of this for hot water use. The output also allows an assessment of the potential range of demands allowing a probabilistic rather than deterministic approach to design to be taken.

Further to this, a method has been developed to determine sets of either representative or extreme results to allow the probabilistic element of the output to be usefully incorporated in design decisions as it is recognised that accounting for every generated scenario is neither practical nor necessary. This allows systems to be 'stress-tested' to determine how sensitive to variations, and therefore how robust, the system design is.

There has been limited existing discussion or consensus on the implication of demand variation and uncertainty on system design. In the presented work, both the size and type of community linked to a centralised energy system has been shown to profoundly influence the demand characteristics. The reduction in uncertainty with scale is gradual and the point at which it is not significant for design likely to be system-type dependent, therefore additional work is required to assess the implications for actual systems.

As a minimum it can therefore be concluded that replacing generic design standards and 'rules-of thumb' with more specific analysis, such as the presented work, has the potential to improve the design of small-scale systems. In addition, significant further assessment is required of both the technical and economic performance of different types of distributed generation systems to determine if available designs are sufficiently flexible to account for the full range of potential operating scenarios.

An alternative view of the presented analysis is that, where practical, constrained or independent energy systems should be used for sufficiently large populations to reduce potential design and operating risk. The degree of demand uncertainty falls sharply for systems in excess of 200-300 households, indicating that systems at this scale would have a sufficiently diverse range of demand behaviours with a lower risk of demand-driven underperformance.

9.4 Further Work

Recommendations for further work in response to the presented work is split into three areas: further development of the model; reducing the calibration and validation data limitations; and further analysis of specific small-scale energy system designs.

9.4.1 Model Development

9.4.1.1 Lighting

The lighting module is currently the least developed of the various developed demand modules. This is partially the result of the low resolution of the lighting data in the Household Electricity Survey (HES) [89] dataset, with main lighting use only measured at central distribution boxes. The current migration from high to low energy bulbs also makes analysis of household data difficult as, even within households, a significant range of bulb types and wattages are used. The availability of room-level lighting data at current levels of low energy lighting use would allow the module to be significantly better calibrated.

The occupant-location driven model captures the broad tendency for occupants to change activity and therefore potentially location. The model is, however, currently hindered by the lack of understanding of how lighting is used in the daytime, multi-room lighting use when occupants move between locations, and the extent to which lighting is left on for security or safety purposes. The current model basis is also calibrated based on overall group behaviour and does not account of individual behaviour differences. More specific lighting use data would significantly improve the model calibration.

9.4.1.2 Hot Water

The assessment of hot water diversity would be improved with better definition of flowrates per specific use. Limited data is currently available and also the introduction of lower flow hot water appliances in relation to diversity is poorly understood.

9.4.1.3 Heating

Development of a heating model in parallel with electricity and hot water demand was hindered by the lack of data incorporating both heating use and sufficient socio-economic and household data to allow the influence of different parameters to be assessed.

The electricity model would be enhanced by a greater understanding of how secondary electric heating is used in households and for this to be integrated into the model. In some cases, current validation was restricted to the summer months, when significant heating use could be discounted.

9.4.1.4 Individualised Behaviours

The focus of the development to individualise occupancy behaviours has been on the timing of waking and sleep transitions. This has driven an improvement in the replication of typical variance in per-household demand levels in these periods. Similar developments targeted on the timing of the first absence from the dwelling and the final return to the dwelling have not resulted in similar improvement in replicating the variance in these periods (i.e. 7-10am and 4-8pm). Further work is therefore required to refine this approach or develop new methods to better capture demand variance at these times.

9.4.1.5 Higher-level Demand Correlations

The developed model captures variation is household demand characteristics driven by household type and income to a degree. However, it is expected that there are additional levels of behaviour correlation that are not captured. For example, the link between household type, income, and the maximum power of individual appliances has not been explored in detail. Neither has the link between floor area and number of owned appliances, and the potential link between number of owned appliances and maximum power of individual appliances. Further work and additional data collection is required to further explore the correlations between specific household characteristics and individual demands.

9.4.1.6 Seasonal Variation

Analysis of the occupancy data did not indicate any extreme seasonal variations in occupancy that were consistent across all populations and to an extent that could be easily modelled. There is some evidence of lower active occupancy in mid-summer and peaks in early spring and late autumn, and further analysis with the UK 2015 Time Use Survey (TUS) when released will allow seasonal variations to be further reviewed. The identified method for average occupancy variation in 7.3.2 could be used for seasonal or monthly manipulation if consistent variances can be identified.

The model incorporates several factors to account for seasonal variations in demand. The lighting model includes realistic solar cycles to ensure the lighting varies realistically through the year. Certain appliances (kettle, microwave, dryers) have clear sinusoidal variations in demand that have been incorporated within the demand sub-model. Seasonal variation in cooker and oven use was expected but a consistent variation was not observed in the HES dataset. Further work is also required to identify patterns in the miscellaneous appliance models. As outlined, accurate seasonal analysis would also require the potential for secondary electric heating use to be captured.

9.4.2 Data Limitations

All the work presented has to some degree been limited by data availability. The principal gap is the lack of data that links occupancy with energy demand. Assumptions have been required about the extent to which occupancy, both total and at specific times, impacts on the timing and frequency of specific demands, with the lack of linked data preventing the relationship being more accurately determined and modelled.

The other major data gap is long-term occupancy data for individuals. Without this it cannot be determined how consistent day-to-day occupancy patterns are and how closely the model captures the distribution of different day types and occupancy patterns at the individual level.

The impending release of the UK 2015 TUS dataset will allow the occupancy model calibration to be updated to reflect current behaviours. This will also potentially allow the IT appliance model to be converted to the same TUS-activity driven basis as TV use, as the UK 2000 TUS data was considered to be unrepresentative of current use. Comparing the two datasets will also allow it to be determined whether the datasets can be combined if sufficiently similar.

9.4.3 Small-Scale Energy System Design Basis

9.4.3.1 'Stress-Testing'

The main conclusions from the presented work are that demand uncertainty at different time-scales remains potentially significant for a range of potential small-scale energy systems and that the impact of the uncertainty is likely to be system-type specific. Therefore, further work is required to model different system designs and to determine if they are sufficiently robust, both technically and economically, to allow for the range of potential demands. It is possible that, at least for certain types of system, minimum system sizes would be recommended to reduce the uncertainty to tolerable levels. This would also allow it to be determined if, and at what scale, existing design standards and guidelines can be used, and to potentially generate more comprehensive standards or simplified analysis methods for smaller systems where the degree of variation and uncertainty exceeds the ability to use a single 'one-size-fits-all' design basis.

9.4.3.2 Hot Water Diversity

Hot water diversity analysis for district heating systems, particularly under UK conditions, is poorly understood. The conservative basis of BS6700 has been replaced by the

Danish equivalent, DS439, but without any distinction made except for house number and little definition of the timescales and tolerances considered. Much existing calibration work is based on actual measurements, which does not necessarily allow for the full range of user behaviours to be incorporated.

There is a strong suggestion from the model output that the worst-case instantaneous diversity for a system exceeds the DS439 basis up to at least 50 households. Further analysis, including a detailed dynamic analysis of systems under short durations where the design basis is exceeded, is required to determine a better diversity basis for UK conditions.

9.5 Concluding Remarks

The primary aim of the presented work was to generate a high-resolution, household-differentiated, probabilistic energy demand model to allow assessment of small-scale energy systems with homogeneous populations that deviate from the national average. Whilst data limitations prevented the inclusion of a heating module at this stage, new or enhanced methods were developed for domestic occupancy, electricity demand, and hot water demand modelling. It has been demonstrated that the new methods are an improvement over existing models in this field, particularly in terms of predicting time-dependent variations at the household-level that results from different characteristics and behaviours, and providing a comprehensive calibration basis that captures all household types that allow it to be used generally for UK systems.

The first stage of model development using group-calibrated modules in two key areas, was followed by a critical assessment of the output. This highlighted that this type of calibration has a tendency for rapid convergence to the data average basis, resulting in weak household differentiation performance. Further improved approaches were developed to account for this unrealistic convergence to improve the degree to which individual household behaviours were captured.

The inclusion of probabilistic factoring to account for the impact of both household characteristics and individual household behaviours on household demand has allowed the model to be used to assess both variation and uncertainty in demand for small-scale energy systems. Whilst variation in average behaviour based on the characteristics of the system consumers can be significant, the presented work has demonstrated that for small-scale systems the impact of less tangible behaviours can be several times more influential in determining system energy use. There is little evidence that this is both understood and incorporated in current energy system planning and design, which is a potential cause of system underperformance. It is hoped that the methods presented provide a pathway to improving small-scale energy system design.

Appendix A

Multivariate Kernel Density Probability Method

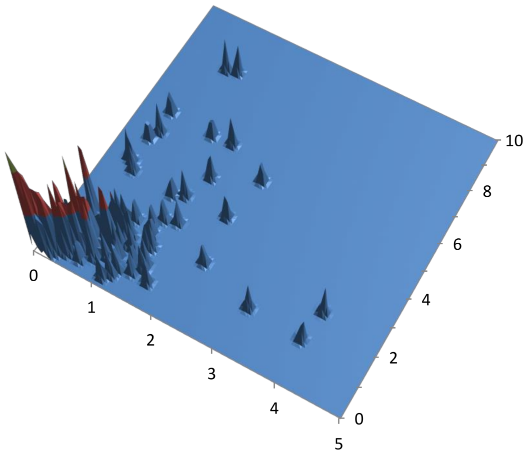
A.1 Appendix Overview

Kernel density estimation is a statistical method used to produce probability density functions from distributions of individual data points. Multivariate kernel density estimation is an extension of the basic technique to multiple variable relationships. This Appendix outlines how the multivariate method has been used to characterise a range of probabilistic relationships within the model that cannot be adequately captured by simple functions.

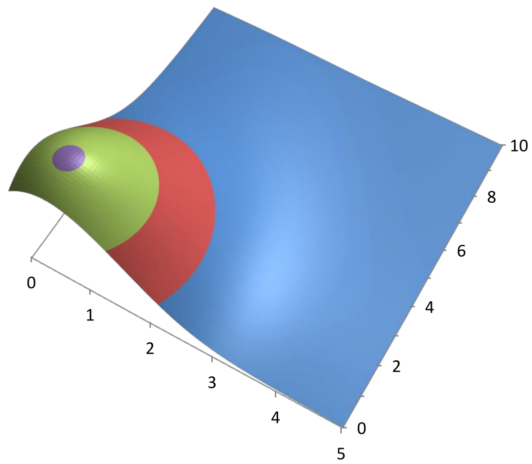
A.2 Kernel Density Method

For several relationships of interest there is a variable y that is strongly influenced by a continuous variable x and for which the complex probabilistic relationship cannot be easily simplified to a single mathematical function. The following method is therefore used to generate a 2-d probability surface from the available data. This is used within the model to generate values of variable y from a pre-determined value of x. The *MATLAB* function, *kde2d*, was used for the data analysis.

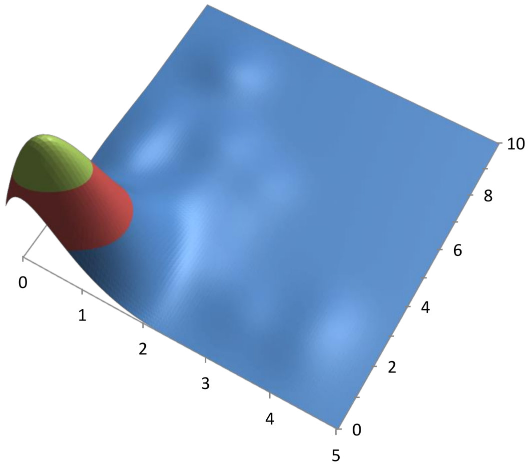
The method requires that each data point *(xi,yi)* is converted to probability density function (kernel) centred on the data point. Key parameters are the shape and width of the density function. The default Gaussian kernel has been used in this case with each data point therefore represented by a normally distributed function. The width of the kernel (known as the bandwidth) is user defined to produce optimally smoothed



(a) Undersmoothed



(b) Oversmoothed



(c) Optimised Bandwidth

Figure A.1. Generated overall probability 'surfaces' for different bandwidths to demonstrate the bandwidth influence on the surface characteristics.

final distributions. Too small and the resulting density function is 'undersmoothed' with excessive influence from individual data points. Too large and the function is 'oversmoothed' with much of the statistical detail lost (see Figure A.1 for examples).

The overall density function is determined by adding the contribution of each individual kernel at each point on the surface. To simplify the analysis the continuous functions x and y are reduced to a set of discrete values analogous to bins in a histogram analysis. For example, analysis of two variables, one with a range from 0 to 1 and the other from 0 to 10, would be converted to a 100x100 kernel density matrix with bin widths of 0.01 and 0.1 respectively. The output is a series of 100 cumulative probability distributions for y based on 100 equal-sized 'bin' ranges for x.

A.2.1 Example: Relationship between Cooker Average Daily Use Duration and Average Daily Cycle Number

The available data from 122 Household Electricity Survey (HES) [89] households shows a distinct relationship between daily cooker cycle number and total use duration, with cycle number increasing with duration. The relationship is however complex with significant deviations from the regression trendline (see Figure A.2), therefore the kernel density method is used to characterise the probabilistic range of potential relationships between the two variables.

Figure A.1(a) and (b) demonstrate undersmoothed and oversmoothed data (representing 0.2 and 5 times the optimum selected bandwidth). Visual analysis of each 1-d bin probability distribution determined that the result shown in Figure A.1(c) was the optimum between retaining the overall statistical relationship and smoothing data from individual data points.

The average daily cycle duration is determined for each household based on the household type average and a probabilistically selected factor from the overall distribution of ratios of household duration to the type average from the HES dataset. The average daily cycle number is then determined by selecting the appropriate bin range for the average daily duration within the kernel density model and then generating a random number between 0 and 1 to determine the average daily cycle number value from the cumulative probability distribution for that specific range.

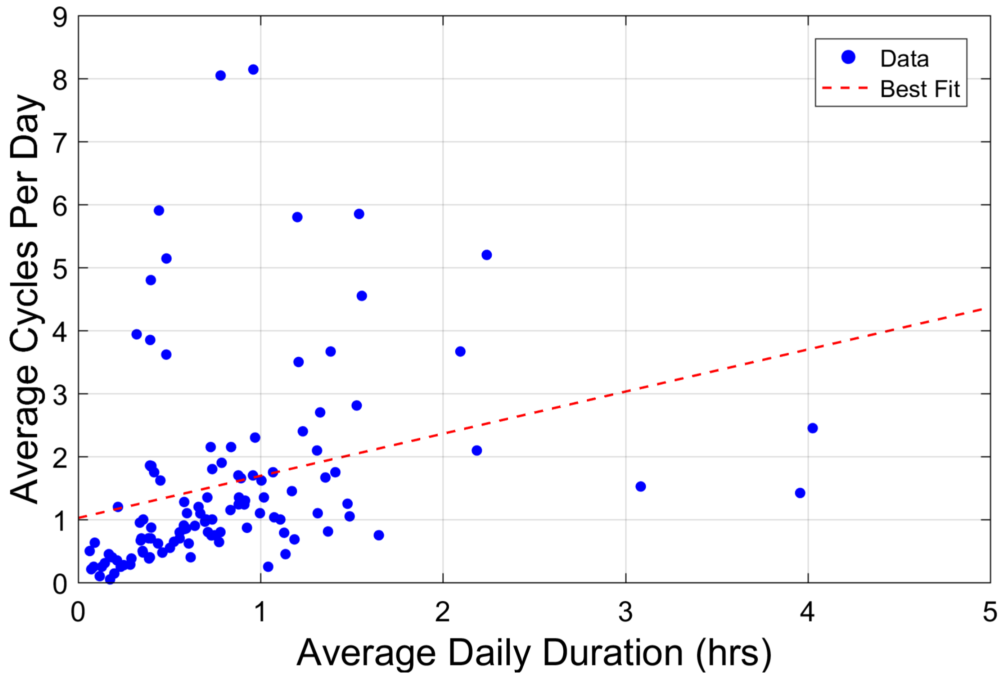


Figure A.2. Relationship between average daily cooker use duration and average number of cycles.

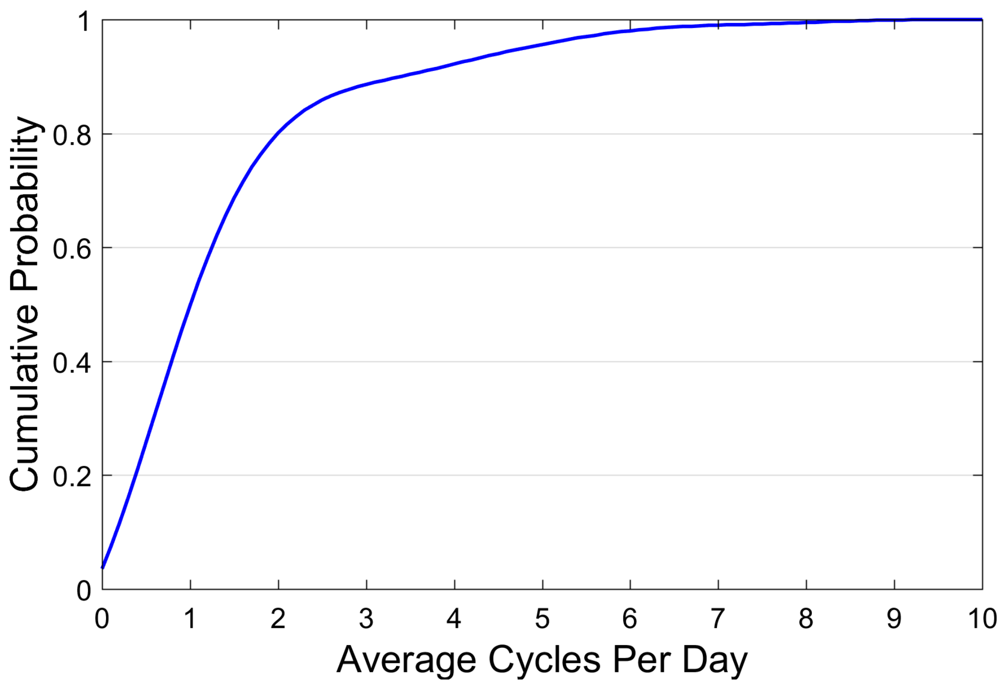


Figure A.3. Example average number of daily cycles cumulative probability distribution for an average daily total cooker use duration of between 0.252 and 0.294 hours.

For example, for an average daily cooker use duration of 0.28 hours, the 0.252-0.294 bin is selected (based on 100 'bins' of 0.042 hours width). The applicable cumulative probability distribution is shown in Figure A.3. A random number of 0.385, for example, determines that the average number of cooker cycles per day is 0.76 cycles.

Appendix B

Occupancy Model Module Populations

Single-Person Household Modules

No.	Day	Work	Age	TUS Age Range	TUS Diaries	Occ Avg.
1	Weekday	Full-time	18-33	18-34	236	18.6%
2	Weekday	Full-time	34-40	30-50	222	21.2%
3	Weekday	Full-time	41+	34+	234	23.1%
4	Weekday	Not-working	18-33	18-40	221	31.2%
5	Weekday	Not-working	34-46	28-56	214	39.7%
6	Weekday	Not-working	47+	40-67	349	43.7%
7	Weekday	Non-working	65-69	61-72	295	46.8%
8	Weekday	Non-working	70-79	68-79	353	48.3%
9	Weekday	Non-working	80+	78+	212	51.4%
10	Weekend	Full-time	18-33	18-50	199	21.1%
11	Weekend	Full-time	34+	33+	200	24.6%
12	Saturday	Not-working	18-33	18-40	225	27.5%
13	Saturday	Not-working	34-46	28-56	211	33.8%
14	Saturday	Not-working	47+	40-67	237	39.4%
15	Saturday	Retired	65-74	64-78	222	47.5%
16	Saturday	Retired	75+	75+	156	50.0%
17	Sunday	Not-working	18-33	18-40	242	30.3%
18	Sunday	Not-working	34-46	28-56	241	36.4%
19	Sunday	Not-working	47+	40-67	252	42.2%
20	Sunday	Non-working	65-74	64-78	207	47.2%
21	Sunday	Non-working	75+	75+	149	50.7%

Live-At-Home Adult Modules

No.	Day	Work	Age	TUS Age Range	TUS Diaries	Occ Avg.
1	Weekday	Full-time	16-18	16-19	261	16.4%
2	Weekday	Full-time	19-24	19-24	308	17.5%
3	Weekday	Not-working	16-18	16-19	246	28.0%
4	Weekday	Not-working	19-24	19-24	275	33.9%
5	Saturday	Not-working	16-18	16-19	202	29.2%
6	Saturday	Not-working	19-24	18-24	266	28.3%
7	Sunday	Not-working	16-18	16-19	228	26.0%
8	Sunday	Not-working	19-24	18-24	321	29.9%
9	Weekend	Full-time	16-24	16-24	270	16.7%

Nightworker Module

No.	Day	Work	Age	TUS Age Range	TUS Diaries	Occ Avg.
1	All	Full-time	All	All	330	22.1%

Single Parent Modules

No.	Day	Work	Age*	TUS Age Range*	TUS Diaries	Occ Avg.
1	Weekday	Full-time	All	All	176	27.4%
2	Weekday	Not-working	0-4	0-4	166	46.0%
3	Weekday	Not-working	5-15	5-15	246	41.8%
4	Weekend	Full-time	All	All	222	26.6%
5	Weekend	Not-working	0-7	0-9	257	40.6%
6	Weekend	Not-working	8-15	5-15	242	40.1%

*Note: Age = Youngest Child Age

Couple Modules

No.	Day	Work†	Age*	TUS Age Range*	TUS Diaries§	Occ Avg.
1	Weekday	FT-FT	18-42	18-43.5	159	25.5%
2	Weekday	FT-FT	43+	35+	151	31.1%
3	Weekend	FT-FT	18-42	18-43.5	163	25.7%
4	Weekend	FT-FT	43+	35+	157	31.1%
5	Weekday	FT-NW	18-54	18-55	170	46.3%
6	Weekday	FT-NW	55+	50+	171	52.5%
7	Saturday	FT-NW	18-54	18-55	228	46.0%
8	Saturday	FT-NW	55+	50+	207	52.2%
9	Sunday	FT-NW	18-54	18-55	209	46.0%
10	Sunday	FT-NW	55+	50+	188	52.5%
11	Weekday	NW-NW	18-54	18-58	148	48.3%
12	Weekday	NW-NW	55-64	50-64.5	231	52.9%
13	Weekday	NW-NW	65-69	60-73	334	55.6%
14	Weekday	NW-NW	70+	69+	253	58.6%
15	Saturday	NW-NW	18-48	18-52	160	39.9%
16	Saturday	NW-NW	49-54	38-62	229	48.1%
17	Saturday	NW-NW	55-69	55-70	237	52.5%
18	Saturday	NW-NW	70+	63+	219	56.7%
19	Sunday	NW-NW	18-48	18-52	177	44.7%
20	Sunday	NW-NW	49-54	38-62	209	50.4%
21	Sunday	NW-NW	55-69	55-70	204	54.1%
22	Sunday	NW-NW	70+	63+	205	58.1%

Note: *Age = Average Age †FT=Full-time, NW=Non-working §Number of combined diaries

Parent Modules

No.	Day	Work†	Age*	TUS Age Range*	TUS Diaries§	Occ Avg.
1	Weekday	FT-FT	0-7	0-9	162	31.7%
2	Weekday	FT-FT	8-15	5-15	207	32.6%
3	Weekend	FT-FT	0-7	0-9	172	31.9%
4	Weekend	FT-FT	8-15	5-15	219	32.7%
5	Weekday	FT-NW	0-2	0-5	236	51.8%
6	Weekday	FT-NW	3-7	3-9	226	49.3%
7	Weekday	FT-NW	8-15	5-15	278	48.4%
8	Weekend	FT-NW	0-7	0-9	146	50.9%
9	Weekend	FT-NW	8-15	5-15	155	49.5%
10	Weekday	NW-NW	All	All	214	52.6%
11	Saturday	NW-NW	0-2	0-5	162	49.7%
12	Saturday	NW-NW	3-7	3-9	165	46.7%
13	Saturday	NW-NW	8-15	5-15	235	47.0%
14	Sunday	NW-NW	0-2	0-5	169	48.3%
15	Sunday	NW-NW	3-7	3-9	178	50.3%
16	Sunday	NW-NW	8-15	5-15	252	50.5%

Note: *Age = Youngest Child Age †FT=Full-time, NW=Non-working §Number of combined diaries

Child Modules

No.	Day	School Status	Age	TUS Age Range	TUS Diaries	Occ Avg.
1	Weekday	Term	0-9	8-9	172	19.6%
2	Weekday	Term	10-11	10-11	187	20.0%
3	Weekday	Term	12-13	12-13	175	20.1%
4	Weekday	Term	14-15	14-15	153	21.2%
5	Weekday	Non-Term	0-10	8-11	99	21.1%
6	Weekday	Non-Term	11-12	10-13	115	27.4%
7	Weekday	Non-Term	13-15	12-15	112	30.4%
8	Saturday	Both	0-9	8-9	124	28.0%
9	Saturday	Both	10-11	10-11	152	32.8%
10	Saturday	Both	12-13	12-13	129	28.8%
11	Saturday	Both	14-15	14-15	125	28.7%
12	Sunday	Both	0-9	8-9	134	30.4%
13	Sunday	Both	10-11	10-11	136	30.4%
14	Sunday	Both	12-13	12-13	127	28.5%
15	Sunday	Both	14-15	14-15	105	30.5%

Appendix C

TV Shared Use Probabilities

This Appendix contains shared TV use probability data taken from the 2001 UK TUS survey [83] as detailed in 5.10.1. It should be noted that shared use with children watching is significantly lower than the equivalent for adults. There may be under-reporting of location sharing for children as their diaries are completed by others, and therefore child data should be used with caution.

Household	Watching	Units On	Probability
2-Person	2A	1	78.8%
2-Person	2A	2	21.2%
3-Person	2A	1	68.2%
3-Person	2A	2	31.8%
3-Person	2C	1	0%
3-Person	2C	2	100%
3-Person	1A1C	1	25.5%
3-Person	1A1C	2	74.5%
3-Person	3A	1	55.7%
3-Person	3A	2	34.2%
3-Person	3A	3	10.1%
3-Person	2A1C	1	20.3%
3-Person	2A1C	2	68.3%
3-Person	2A1C	3	11.4%
3-Person	1A2C	1	0%
3-Person	1A2C	2	65.5%
3-Person	1A2C	3	34.5%

A=Adult, C=Child

Household	Watching	Units On	Probability
4-Person	2A	1	59.6%
4-Person	2A	2	40.4%
4-Person	2C	1	0%
4-Person	2C	2	100%
4-Person	1A1C	1	19.4%
4-Person	1A1C	2	80.6%
4-Person	3A	1	48.8%
4-Person	3A	2	32.5%
4-Person	3A	3	18.7%
4-Person	2A1C	1	24.8%
4-Person	2A1C	2	57.9%
4-Person	2A1C	3	17.4%
4-Person	1A2C	1	0%
4-Person	1A2C	2	25.4%
4-Person	1A2C	3	74.5%
4-Person	4A	1	36.7%
4-Person	4A	2	30.0%
4-Person	4A	3	30.0%
4-Person	4A	4	3.3%
4-Person	3A1C	1	21.7%
4-Person	3A1C	2	68.6%
4-Person	3A1C	3	8.4%
4-Person	3A1C	4	1.2%
4-Person	2A2C	1	0.1%
4-Person	2A2C	2	20.4%
4-Person	2A2C	3	63.5%
4-Person	2A2C	4	16.0%
4-Person	1A3C	1	0%
4-Person	1A3C	2	3.2%
4-Person	1A3C	3	12.7%
4-Person	1A3C	4	84.1%

A=Adult, C=Child

Household	Watching	Units On	Probability
5-Person	2A	1	54.5%
5-Person	2A	2	45.5%
5-Person	2C	1	2.0%
5-Person	2C	2	98.0%
5-Person	1A1C	1	17.1%
5-Person	1A1C	2	82.9%
5-Person	3A	1	19.7%
5-Person	3A	2	37.3%
5-Person	3A	3	43.0%
5-Person	2A1C	1	16.6%
5-Person	2A1C	2	47.6%
5-Person	2A1C	3	35.8%
5-Person	1A2C	1	0.1%
5-Person	1A2C	2	29.7%
5-Person	1A2C	3	69.5%
5-Person	4A	1	13.9%
5-Person	4A	2	59.9%
5-Person	4A	3	22.2%
5-Person	4A	4	4.1%
5-Person	3A1C	1	17.3%
5-Person	3A1C	2	21.6%
5-Person	3A1C	3	22.5%
5-Person	3A1C	4	38.6%
5-Person	2A2C	1	0%
5-Person	2A2C	2	5.0%
5-Person	2A2C	3	25.3%
5-Person	2A2C	4	69.7%
5-Person	1A3C	1	0%
5-Person	1A3C	2	4.2%
5-Person	1A3C	3	29.8%
5-Person	1A3C	4	66.0%
5-Person	5A	1	30.2%
5-Person	5A	2	55.6%
5-Person	5A	3	0%
5-Person	5A	4	0%
5-Person	5A	5	14.3%
5-Person	4A1C	1	0%
5-Person	4A1C	2	0%
5-Person	4A1C	3	52.4%
5-Person	4A1C	4	47.6%
5-Person	4A1C	5	0%
5-Person	3A2C	1	0%
5-Person	3A2C	2	0%
5-Person	3A2C	3	2.4%
5-Person	3A2C	4	83.3%
5-Person	3A2C	5	14.3%
5-Person	2A3C	1	14.0%
5-Person	2A3C	2	0%
5-Person	2A3C	3	15.8%
5-Person	2A3C	4	23.6%
5-Person	2A3C	5	46.7%

A=Adult, C=Child

Bibliography

[1] National Grid, "2013 Electricity Ten Year Statement." Nov. 2013. [Online]. Available: http://www2.nationalgrid.com/UK/Industry-information/Future-of-Energy/Electricity-ten-year-statement/ETYS-Archive/

[2] E. J. Coster, J. M. Myrzik, B. Kruimer, and W. L. Kling, "Integration issues of distributed generation in distribution grids," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 28–39, 2011.

[3] E. A. Kremers, *Modelling and simulation of electrical energy systems through a complex systems approach using agent-based models.* KIT Scientific Publishing, 2013.

[4] The Scottish Government, "Low Carbon Scotland: Meeting the Emissions Reduction Targets 2013-2027 (RPP2)- The Second Report on Proposals and Policies." Jan. 2013. [Online]. Available: http://www.gov.scot/Resource/0041/00413150.pdf

[5] National Grid, "Future Energy Scenarios: UK gas and electricity transmission." July 2015. [Online]. Available: http://media.nationalgrid.com/media/1169/future_energy_scenarios_2015.pdf

[6] European Commission, "Energy Roadmap 2050," 2011. [Online]. Available: https://ec.europa.eu/energy/sites/ener/files/documents/2012_energy_roadmap_2050_en_0.pdf

[7] N. Kelly, A. Samuel, and J. Hand, "Testing integrated electric vehicle charging and domestic heating strategies for future UK housing." *Energy and Buildings*, vol. 105, pp. 377–392, 2015.

[8] S. Tipping, J. Chanfreau, J. Perry, and C. Tait, "The fourth work-life balance employee survey." July 2012, De-

partment for Business Innovation and Skills. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32153/12-p151-fourth-work-life-balance-employee-survey.pdf

[9] Office of National Statistics. A01: Labour market statistics summary data tables. 16 March 2016. Table 3. [Online]. Available: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/summaryoflabourmarketstatistics [Accessed Mar. 23, 2016].

[10] Office of National Statistics, "Characteristics of Home Workers, 2014." June 2014. [Online]. Available: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp171776_365592.pdf

[11] DECC. Energy Consumption in the UK (ECUK). Chapter 3: Domestic data tables. Table 3.12 - Number of appliances owned by households in the UK 1970 to 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/573270/Copy_of_ECUK_Tables_2016_Nov.xlsx [Accessed Aug. 04, 2016].

[12] K. Gram-Hanssen, "Efficient technologies or user behaviour, which is the more important when reducing households energy consumption?" *Energy Efficiency*, vol. 6, no. 3, pp. 447–457, 2013.

[13] DECC. (2015) Fuel used in electricity generation and electricity supplied. Energy Trends: Table 5.1. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/533218/ET_5.1.xls [Accessed Aug. 05, 2016].

[14] DECC. Energy Consumption in the UK (ECUK). Chapter 3: Domestic data tables. Table 3.02 - Domestic final energy consumption by end use and fuel 1970 to 2013. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/541168/ECUK_Tables_2016.xlsx [Accessed Aug. 04, 2016].

[15] Energy UK, "Pathways for the GB Electricity Sector to 2030: Full report." Feb. 2016. [Online]. Available: https://www.energy-uk.org.uk/publication.html?task=file.download&id=5722

[16] S. Abu-Sharkh, R. Arnold, J. Kohler, R. Li, T. Markvart, J. Ross, K. Steemers, P. Wilson, and R. Yao, "Can microgrids make a major contribution to UK energy supply?" *Renewable and Sustainable Energy Reviews*, vol. 10, no. 2, pp. 78–127, 2006.

[17] Energy Networks Association, "Distributed Generation Connection Guide: A Guide for Connecting Generation to the Distribution Network that Falls Under G59/3 and is 50kW or Less 3-Phase or 17kW or Less Single-Phase." June 2014. [Online]. Available: http://www.ukpowernetworks.co.uk/internet/asset/ab62c3e0-285b-40e6-b909-389ba22ee1eK/A+guide+for+connecting+generation+to+the+distribution+network+G59-3+50kW+or+less+.pdf

[18] National Grid, "System Operability Framework: Chapter 6 - Embedded Generation." Nov. 2015. [Online]. Available: www2.nationalgrid.com/WorkArea/DownloadAsset.aspx?id=44040

[19] Element Energy Ltd., "Research on district heating and local approaches to heat decarbonisation: A study for the Committee on Climate Change." Nov. 2015. [Online]. Available: https://www.theccc.org.uk/wp-content/uploads/2015/11/Element-Energy-for-CCC-Research-on-district-heating-and-local-approaches-to-heat-decarbonisation.pdf

[20] European Commission, "Energy Research: Key Advantages of Distributed Energy Resources." 2006, article No.1159. [Online]. Available: https://ec.europa.eu/research/energy/print.cfm?file=/comm/research/energy/nn/nn_rt/nn_rt_dg/article_1159_en.htm

[21] J. P. Lopes, N. Hatziargyriou, J. Mutale, P. Djapic, and N. Jenkins, "Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities." *Electric Power Systems Research*, vol. 77, no. 9, pp. 1189–1203, 2007.

[22] DTI/Ofgem, "Review of Distributed Generation." May 2007. [Online]. Available: https://www.ofgem.gov.uk/ofgem-publications/52326/review-distributed-generation.pdf

[23] Carbon Connect, "Distributed generation: from Cinderella to centre stage," *A Report by Carbon Connect*, 2012. [Online]. Available: http://www.eti.co.uk/wp-content/uploads/2014/03/CarbonConnect_DistributedGeneration_PDF.pdf

[24] K. L. Anaya and M. G. Pollitt, "Distributed Generation: Opportunities for Distribution Network Operators, Wider Society and Generators," in *Proceedings of the 23rd International Conference on Electricity Distribution - Paper 1222*. CIRED, 2015, pp. 1–5. [Online]. Available: http://cired.net/publications/cired2015/papers/CIRED2015_1222_final.pdf

[25] P. A. Strachan, R. Cowell, G. Ellis, F. Sherry-Brennan, and D. Toke, "Promoting community renewable energy in a corporate energy world." *Sustainable Development*, vol. 23, no. 2, pp. 96–109, 2015.

[26] IEEE, "Power systems of the future: The case for energy storage, distributed generation and microgrids." Nov. 2012. [Online]. Available: https://building-microgrid.lbl.gov/sites/all/files/THE_CASE_FOR_ENERGY_STORAGE_DISTRIBUTED.pdf

[27] E. Petrie, H. Willis, and M. Takahashi, "Distributed generation in developing countries." *Cogeneration and On-Site Power Production*, vol. 2, no. 5, pp. 41–42, 2001.

[28] D. Schnitzer, J. P. Carvallo, R. Deshmukh, J. Apt, and D. Kammen, "Microgrids for rural electrification." *Report for United Nations Foundation.*, Jan. 2014. [Online]. Available: https://rael.berkeley.edu/wp-content/uploads/2015/04/MicrogridsReportEDS.pdf

[29] DECC, "Energy Consumption in the UK (2015). Chapter 3: Domestic energy consumption in the UK between 1970 and 2014." [Online]. Available: http://webarchive.nationalarchives.gov.uk/20150423095958/https:

//www.gov.uk/government/uploads/system/uploads/attachment_data/file/338662/ecuk_chapter_3_domestic_factsheet.pdf

[30] DECC, "The Future of Heating: A Strategic Framework for Low Carbon Heat in the UK." Mar. 2012. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48574/4805-future-heating-strategic-framework.pdf

[31] A. L. Browne, M. Pullinger, W. Medd, and B. Anderson, "Patterns of practice: a reflection on the development of quantitative/mixed methodologies capturing everyday life related to water consumption in the UK." *International Journal of Social Research Methodology*, vol. 17, no. 1, pp. 27–43, 2014.

[32] DECC/Ofgem, "Smart Grid Vision and Roadmap." Feb. 2014. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/285417/Smart_Grid_Vision_and_RoutemapFINAL.pdf

[33] DECC, "Smart Metering Early Learning Project: Domestic Energy Consumption Analysis. Report and Technical Annex." Mar. 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/407542/2_ELP_Domestic_Energy_Consumption_Analysis_Report.pdf

[34] Society of Motor Manufacturers and Traders. Registration - EVs and AFVs. [Online]. Available: https://www.smmt.co.uk/category/news-registration-evs-afvs/ [Accessed Aug. 05, 2016].

[35] Brook Lyndhurst, "Uptake of Ultra Low Emission Vehicles in the UK: A Rapid Evidence Assessment for the Department of Transport." Aug. 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/464763/uptake-of-ulev-uk.pdf

[36] IEA, "Global EV Outlook: Understanding the Electric Vehicle Landscape to 2020." Apr. 2013. [Online]. Available: https://www.iea.org/publications/freepublications/publication/GlobalEVOutlook_2013.pdf

[37] G. Haines, A. McGordon, P. Jennings, and N. Butcher, "The simulation of vehicle-to-home systems–using electric vehicle battery storage to smooth domes-

tic electricity demand." in *Proceedings of the 2009 EVER Monaco International Exhibition and Conference on Ecologic Vehicles and Renewable Energies.*, 2009.

[38] P. Campbell, "Electric car drivers to sell power back to national grid." *Financial Times*, 10th May 2016. [Online]. Available: https://www.ft.com/content/7e75b7d2-169c-11e6-b197-a4af20d5575e

[39] IEA, "Technology Roadmap: Energy Storage," 2014. [Online]. Available: https://www.iea.org/publications/freepublications/publication/TechnologyRoadmapEnergystorage.pdf

[40] G. Flett, "Community Biomass Heating Networks: Modelling and Evaluation of Future Potential." 2013. [Online]. Available: http://www.esru.strath.ac.uk/Documents/MSc_2013/Flett.pdf

[41] The Scottish Government. (2016) Housing Statistics for Scotland - Key Information and Summary Tables. [Online]. Available: http://www.gov.scot/Topics/Statistics/Browse/Housing-Regeneration/HSfS/KeyInfoTables [Accessed Aug. 06, 2016].

[42] DCLG, "Net supply of housing: 2014-15, England." Nov. 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/475832/Net_Supply_of_Housing_England_2014-15.pdf

[43] DECC/BRE, "Energy Follow Up Survey, 2011." 2016, [data collection]. 3rd Edition. UK Data Service. SN:7471. [Online]. Available: http://dx.doi.org/10.5255/UKDA-SN-7471-3

[44] E. Störmer, C. Patscha, J. Prendergast, C. Daheim, M. Rhisiart, P. Glover, and H. Beck, "The future of work: jobs and skills in 2030." Feb. 2014. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/303334/er84-the-future-of-work-evidence-report.pdf

[45] D. Murray, J. Liao, L. Stankovic, V. Stankovic, R. Hauxwell-Baldwin, C. Wilson, M. Coleman, T. Kane, and S. Firth, "A data management platform for personalised real-time energy feedback." in *Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting*, 2015.

[46] Office of National Statistics. LSOA/MSOA 2011 historic electricity consumption. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/418828/2011.zip [Accessed Jun. 26, 2015].

[47] Office of National Statistics. (2011) LSOA/MSOA 2011 historic gas consumption. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/419273/2011.zip [Accessed Jun. 26, 2016].

[48] BRE, "The Government's Standard Assessment Procedure for Energy Rating of Dwellings: 2012 Edition. Version 9.92." Oct. 2013. [Online]. Available: https://www.bre.co.uk/filelibrary/SAP/2012/SAP-2012_9-92.pdf

[49] Datashine Scotland. [Online]. Available: http://http://scotland.datashine.org.uk [Accessed Apr. 14, 2016].

[50] R. Yao and K. Steemers, "A method of formulating energy load profile for domestic buildings in the UK." *Energy and Buildings*, vol. 37, no. 6, pp. 663–671, 2005.

[51] J. Torriti, "A review of time use models of residential electricity demand." *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 265–272, 2014.

[52] I. Staffell, P. Baker, J. P. Barton, N. Bergman, R. E. Blanchard, N. P. Brandon, D. J. Brett, A. Hawkes, D. Infield, C. N. Jardine *et al.*, *UK microgeneration. Part II: technology overviews.* ICE Publishing Ltd., 2010.

[53] H. L. Willis, *Distributed power generation: planning and evaluation.* CRC Press, 2000.

[54] M. Nehrir, C. Wang, K. Strunz, H. Aki, R. Ramakumar, J. Bing, Z. Miao, and Z. Salameh, "A review of hybrid renewable/alternative energy systems for electric power generation: Configurations, control, and applications," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 4, pp. 392–403, 2011.

[55] A. H. Fathima and K. Palanisamy, "Optimization in microgrids with hybrid energy systems–A review." *Renewable and Sustainable Energy Reviews*, vol. 45, pp. 431–446, 2015.

[56] Dansk Standard, *DS439: Code of Practice for domestic water supply.* Dansk Standard, 2009.

[57] D. H. McQueen, P. R. Hyland, and S. J. Watson, "Monte carlo simulation of residential electricity demand for forecasting maximum demand on distribution networks." *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1685–1689, 2004.

[58] B. Rigby, *Design of electrical services for buildings.* Taylor & Francis, 2005.

[59] British Standards Institution, *Design, installation, testing and maintenance of services supplying water for domestic use within buildings and their curtilages - Specification. .* London: BSI, 2016.

[60] S. Fjärrvärme, "District Heating Substations: Design and Installation." *Technical Regulations F*, vol. 101, Apr. 2008. [Online]. Available: http://www.svenskfjarrvarme.se/Global/Rapporter%20och%20dokument%20INTE%20Fj%C3%A4rrsyn/Tekniska_bestammelser/Kundanlaggningar/District_heating_substations_Design_and_installation_F-101_2008.pdf

[61] J. Thorsen and H. Kristjansson, "Cost considerations on storage tank versus heat exchanger for htw preparation." in *Proceedings of the 10th International Symposium on District Heating and Cooling, Hanover, Germany*, 2006.

[62] PEWO Energietechnik GmbH. Heating Substations: For Heat Distribution in Heating Networks and Buildings. [Online]. Available: http://www.simplybiomass.co.uk/wp-content/uploads/2015/02/pewoTherm_Prospekt_en_v10.9_lQ.pdf [Accessed Mar. 24, 2016].

[63] CIBSE/CHPA, "Heat Networks: Code of Practice for the UK. Draft for Consulation." 2014. [Online]. Available: http://www.cibse.org/getmedia/2b67857e-d29a-4d0c-a658-2fe7addf4f5d/Heat-Networks-Code-of-Practice-for-the-UK_draft_consultation.pdf.aspx

[64] T. Haggis, "Network Design Manual v7.7," *ON Central Networks: Coventry, UK*, Dec. 2006. [Online]. Available: https://www.scribd.com/document/33735606/Network-Design-Manual-v7-7

[65] N. Gourlay, "Framework for design and planning of LV housing developments, including u/g networks and associated HV/LV S/S. Issue 5." *SP Energy Networks Design Manual (SPD, SPM) Sec. 5d.*, May 2013. [Online]. Available: http://www.spenergynetworks.co.uk/userfiles/file/ESDD-02-012%20Issue%204%20-%20Greenfield%20Housing.pdf

[66] DECC, "Community Energy Strategy: Full Report. A report by the Department for Energy and Climate Change." Jan. 2014. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/275163/20140126Community_Energy_Strategy.pdf

[67] D. Yan and T. Hong, "EBC Annex 66 Text: Definition and Simulation of Occupant Behavior in Buildings." Nov. 2014, international Energy Agency. [Online]. Available: http://www.annex66.org/sites/default/files/pictures/EBC%20Annex%2066%20Text_0.pdf

[68] J. Tanimoto, A. Hagishima, and H. Sagara, "A methodology for peak energy requirement considering actual variation of occupants behavior schedules." *Building and Environment*, vol. 43, no. 4, pp. 610–619, 2008.

[69] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic electricity use: A high-resolution energy demand model." *Energy and Buildings*, vol. 42, no. 10, pp. 1878–1887, 2010.

[70] J. Widén and E. Wäckelgård, "A high-resolution stochastic model of domestic activity patterns and electricity demand." *Applied Energy*, vol. 87, no. 6, pp. 1880–1892, 2010.

[71] U. Wilke, "Probabilistic bottom-up modelling of occupancy and activities to predict electricity demand in residential buildings." Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2013.

[72] A. Hawkes and M. Leach, "Impacts of temporal precision in optimisation modelling of micro-combined heat and power." *Energy*, vol. 30, no. 10, pp. 1759–1779, 2005.

[73] A. Wright and S. Firth, "The nature of domestic electricity-loads and effects of time averaging on statistics and on-site generation calculations." *Applied Energy*, vol. 84, no. 4, pp. 389–403, 2007.

[74] M. Lave, J. Quiroz, M. J. Reno, and R. J. Broderick, "High temporal resolution load variability compared to PV variability." in *Proceeding of the 43rd Photovoltaic Specialists Conference (PVSC), Seattle.* IEEE, 2016, pp. 1831–1836.

[75] H. Bagge and D. Johansson, "Measurements of household electricity and domestic hot water use in dwellings and the effect of different monitoring time resolution." *Energy*, vol. 36, no. 5, pp. 2943–2951, 2011.

[76] N. Kreutzer and I. Knight, "Social housing electrical energy consumption profiles in the United Kingdom." in *Proceedings of the 2nd International Solar Cities Congress, Oxford*, 2006.

[77] Y. G. Yohanis, J. D. Mondol, A. Wright, and B. Norton, "Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use." *Energy and Buildings*, vol. 40, no. 6, pp. 1053–1059, 2008.

[78] F. McLoughlin, A. Duffy, and M. Conlon, "The generation of domestic electricity load profiles through Markov Chain modelling." *East-Asian Journal of Sustainable Energy Development Policy*, vol. 3, 2010.

[79] F. Haldi and D. Robinson, "The impact of occupants' behaviour on building energy demand." *Journal of Building Performance Simulation*, vol. 4, no. 4, pp. 323–338, 2011.

[80] S. Kelly, M. Shipworth, D. Shipworth, M. Gentry, A. Wright, M. Pollitt, D. Crawford-Brown, and K. Lomas, "A panel model for predicting the diversity of internal temperatures from English dwellings." July 2012, tyndall Centre for Climate Change Research. Working Paper 154. [Online]. Available: http://www.tyndall.ac.uk/publications/tyndall-working-paper/2012/panel-model-predicting-diversity-internal-temperatures

[81] Z. Gill, M. Tierney, I. Pegg, and N. Allan, "Low-energy dwellings: the contribution of behaviours to actual performance." *Building Research & Information*, vol. 38, no. 5, pp. 491–508, 2010.

[82] A. Grandjean, J. Adnot, and G. Binet, "A review and an analysis of the residential electric load curve models." *Renewable and Sustainable Energy Reviews*, vol. 16, no. 9, pp. 6539–6565, 2012.

[83] Office of National Statistics, "United Kingdom Time Use Survey, 2000." 2003, [data collection]. 3rd Edition. UK Data Service. SN:4504. [Online]. Available: http://dx.doi.org/10.5255/UKDA-SN-4504-1

[84] K. Fisher, J. Gershuny, and A. Gauthier, "Multinational time use study - Users guide and documentation. Version 5 - updated." Oct. 2012, Centre for Time Use Research. [Online]. Available: http://www.timeuse.org/sites/ctur/files/public/ctur_report/5715/mtus-user-guide-r5.pdf

[85] Eurostat, "Harmonised European time use surveys, 2008 guidelines." Dec. 2008, Cat. No.: KS-RA-08-014-EN-N. [Online]. Available: http://ec.europa.eu/eurostat/documents/3859598/5909673/KS-RA-08-014-EN.PDF/a745ca2e-7dc6-48a9-a36c-000ad120380e?version=1.0

[86] Office of National Statistics, "ONS Omnibus Survey, Time Use Module, February, June, September and November 2005." 2007, [data collection]. 2nd Edition. UK Data Service. SN:5592. [Online]. Available: http://dx.doi.org/10.5255/UKDA-SN-5592-1

[87] Sociaal en Cultureel Planbureau, "Tijdsbestedingsonderzoek 2005 - TBO 2005." 2005. [Online]. Available: http://dx.doi.org/10.17026/dans-znn-5xvz

[88] DECC, "Sub-national electricity and gas consumption statistics." 2015. [Online]. Available: https://www.gov.uk/government/collections/sub-national-electricity-consumption-data

[89] DECC. (2012) Household Electricity Survey. [Accessed Nov. 26, 2013].

[90] Energy Savings Trust, “Measurement of Domestic Hot Water Consumption in Dwellings.” 2008. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48188/3147-measure-domestic-hot-water-consump.pdf

[91] National Records of Scotland, “Household composition for specific groups of people in Scotland: Scotland’s Census 2011,” Aug. 2015. [Online]. Available: http://www.scotlandscensus.gov.uk/documents/analytical_reports/HH%20report.pdf

[92] Scottish Government, “Scottish Household Survey, 2007-2008.” 2010, [data collection]. 2nd Edition. UK Data Service. SN:6361. [Online]. Available: https://discover.ukdataservice.ac.uk/catalogue/?sn=6361&type=Data%20catalogue

[93] Scottish Government, “Scottish Household Survey, 2014.” 2016, [data collection]. 1st Edition. UK Data Service. SN:7964. [Online]. Available: https://discover.ukdataservice.ac.uk/catalogue/?sn=7964&type=Data%20catalogue

[94] DCLC, “Household Characteristics.” 2012, Statistical Dataset. [Online]. Available: https://www.gov.uk/government/statistical-data-sets/live-tables-on-household-characteristics

[95] The Scottish Government, “Scottish House Condition Survey, 2014 Key Findings.” 2015. [Online]. Available: http://www.gov.scot/Resource/0049/00490947.pdf

[96] DCLG, “English Housing Survey: Housing Stock Report, 2014-15.” 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/539600/Housing_Stock_report.pdf

[97] Scott-Wilson, “Housing standards: evidence and research - Dwelling size survey.” Apr. 2010, a report prepared by Scott-Wilson for CABE. [Online]. Available: http://webarchive.nationalarchives.gov.uk/20110118095356/http:/www.cabe.org.uk/files/dwelling-size-survey.pdf

[98] Office of National Statistics, "UK Perspectives 2016: An overview of the UK labour market." May 2016. [Online]. Available: http://visual.ons.gov.uk/uk-perspectives-2016-an-overview-of-the-uk-labour-market/

[99] Office of National Statistics. Table a49: Percentage of households by size, composition and age in each gross income decile group, 2014. [Online]. Available: https://www.ons.gov.uk/ons/rel/family-spending/family-spending/2015-edition/rft-a49.xls [Accessed Jul. 15, 2015].

[100] Office of National Statistics. Table 46: Percentage of households with durable goods by income group and household composition, UK, 2014. [Online]. Available: https://www.ons.gov.uk/ons/rel/family-spending/family-spending/2015-edition/rft-a46.xls [Accessed Jul. 15, 2015].

[101] J. Palmer, N. Terry, T. Kane, S. Firth, M. Hughes, P. Pope, J. Young, D. Knight, and D. Godoy-Shimizu, "Electrical appliances at home: tuning in to energy savings." *Further Analysis of the Household Electricity Survey Energy Use at Home*, Nov. 2013. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/275484/electricity_survey_2_tuning_in_to_energy_saving.pdf

[102] Office of National Statistics. 2011 Census: Dataset LC4104EW - Occupancy rating (rooms) by household composition. [Online]. Available: http://web.ons.gov.uk/ons/data/dataset-finder/-/q/dcDetails/Census/LC4104EW?p_auth=3ThVzaeB&p_p_lifecycle=1&_FOFlow1_WAR_FOFlow1portlet_dataset_navigation=datasetCollectionDetails [Accessed Jul. 17, 2015].

[103] Office of National Statistics. 2011 Census: Dataset LC1402EW - Household composition by number of bedrooms. [Online]. Available: http://web.ons.gov.uk/ons/data/dataset-finder/-/q/dcDetails/Census/LC1402EW?p_auth=3ThVzaeB&p_p_lifecycle=1&_FOFlow1_WAR_FOFlow1portlet_dataset_navigation=datasetCollectionDetails [Accessed Jul. 17, 2015].

[104] J. Pannell, H. Aldridge, and P. Kenway, "Market assessment of housing options for older people." Apr. 2012, A report for Shelter and the Joseph Rowntree Foundation. [Online]. Available: http://npi.org.uk/files/5213/7485/1289/Market_Assessment_of_Housing_Options_for_Older_People.pdf

[105] S. P. Jenkins, "The income distribution in the uk: A picture of advantage and disadvantage." Feb. 2015. [Online]. Available: https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2015-01.pdf

[106] Building Research Establishment, "Energy Use in Homes: A series of reports on domestic energy use in England: Fuel Consumption." 2005. [Online]. Available: https://www.bre.co.uk/filelibrary/pdf/rpts/FuelConsumption.pdf

[107] V. White, S. Roberts, and I. Preston, "Understanding High Use Low Income Energy Consumers." *Final report to Ofgem*, Nov. 2010. [Online]. Available: https://www.cse.org.uk/downloads/reports-and-publications/policy/energy-justice/understanding_high_use_low_income_energy_consumers.pdf

[108] T. Jamasb and H. Meier, "Household energy spending and income groups: Evidence from Great Britain." Cambridge Working Papers in Economics 1011, Faculty of Economics, University of Cambridge, Tech. Rep., Feb. 2010. [Online]. Available: http://www.eprg.group.cam.ac.uk/wp-content/uploads/2014/01/JamasbMeierCombined-EPRG10031.pdf

[109] Department of Work and Pensions, "Households Below Average Income (HBAI) Quality and Methodology Information Report 2012/13," July 2013. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/325492/households-below-average-income-quality-methodology-2012-2013.pdf

[110] K. Williams, "Space per person in the UK: A review of densities, trends, experiences and optimum levels." *Land Use Policy*, vol. 26, pp. S83–S92, 2009.

[111] J. Henderson and J. Hart, "BREDEM 2012; a technical description of the BRE domestic energy model." Jan. 2015, version 1.1, with corrections.

[Online]. Available: http://www.bre.co.uk/filelibrary/bredem/BREDEM-2012-specification.pdf

[112] DECC, "National Energy Efficiency Data - Framework: Summary of analysis using the National Energy Efficiency Data-Framework." June 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/437093/National_Energy_Efficiency_Data-Framework__NEED__Main_Report.pdf

[113] N. Terry, J. Palmer, D. Godoy, S. Firth, T. Kane, and A. Tillson, "Further Analysis of the Household Survey: Lighting Study (Final Report)," May 2013, Reference: 475/09/2012. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/326085/Lighting_Report.pdf

[114] R. Jones and K. Lomas, "Determinants of high electrical energy demand in UK homes: Appliance ownership and use." *Energy and Buildings*, vol. 117, pp. 71–82, 2016.

[115] R. Parris, "Still got a big box TV? High energy costs may make you think again." Which? article, dated 24/12/2011. [Online]. Available: https://blogs.which.co.uk/technology/tvs/still-using-a-big-box-tv-why-the-energy-costs-may-make-you-think-again/

[116] R. Hierzinger and J. Krivosik, "Comparison of energy efficiency requirements of the energy labels and ecodesign legislation." June 2012, Come On Labels. [Online]. Available: http://www.energyagency.at/fileadmin/dam/pdf/publikationen/berichteBroschueren/comeonlabels_vergleich.pdf

[117] C. Walker and J. Pokoski, "Residential load shape modelling based on customer behavior." *Power Apparatus and Systems, IEEE Transactions on*, no. 7, pp. 1703–1711, 1985.

[118] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, "A bottom-up approach to residential load modeling." *Power Systems, IEEE Transactions on*, vol. 9, no. 2, pp. 957–964, 1994.

[119] M. M. Armstrong, M. C. Swinton, H. Ribberink, I. Beausoleil-Morrison, and J. Millette, "Synthetically derived profiles for representing occupant-driven electric loads in Canadian housing." *Journal of Building Performance Simulation*, vol. 2, no. 1, pp. 15–30, 2009.

[120] J. Widén, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegård, and E. Wäckelgård, "Constructing load profiles for household electricity and hot water from time-use data - Modelling approach and validation." *Energy and Buildings*, vol. 41, no. 7, pp. 753–768, 2009.

[121] J. Torriti, "Price-based demand side management: Assessing the impacts of time-of-use tariffs on residential electricity demand and peak shifting in Northern Italy." *Energy*, vol. 44, no. 1, pp. 576–583, 2012.

[122] Office of National Statistics. MSOA 2014 historic electricity consumption. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/497487/Middle_Layer_Super_Output_Area__MSOA__domestic_electricity_estimates_2014_published_Jan_2016v3.xlsx [Accessed Mar. 26, 2016].

[123] A. Losi, P. Mancarella, and A. Vicino, *Integration of Demand Response Into the Electricity Chain: Challenges, Opportunities and Smart Grid Solutions.* John Wiley & Sons, 2015.

[124] I. A. Macdonald, "Quantifying the effects of uncertainty in building simulation." Ph.D. dissertation, University of Strathclyde, 2002.

[125] C. M. Clevenger and J. Haymaker, "The impact of the building occupant on energy modeling simulations." in *Proceedings of the 2006 Joint International Conference on Computing and Decision Making in Civil and Building Engineering, Montreal, Canada.* Citeseer, 2006, pp. 1–10.

[126] C. J. Hopfe and J. L. Hensen, "Uncertainty analysis in building performance simulation for design support." *Energy and Buildings*, vol. 43, no. 10, pp. 2798–2805, 2011.

[127] C. Spitz, L. Mora, E. Wurtz, and A. Jay, "Practical application of uncertainty analysis and sensitivity analysis on an experimental house." *Energy and Buildings*, vol. 55, pp. 459–470, 2012.

[128] A. Mahdavi and F. Tahmasebi, "The deployment-dependence of occupancy-related models in building performance simulation," *Energy and Buildings*, vol. 117, pp. 313–320, 2016.

[129] J. L. Leyten and S. R. Kurvers, "Robustness of buildings and HVAC systems as a hypothetical construct explaining differences in building related health and comfort symptoms and complaint rates." *Energy and Buildings*, vol. 38, no. 6, pp. 701–707, 2006.

[130] P. Hoes, J. Hensen, M. Loomans, B. De Vries, and D. Bourgeois, "User behavior in whole building simulation." *Energy and Buildings*, vol. 41, no. 3, pp. 295–302, 2009.

[131] S. Borgeson and G. Brager, "Occupant control of windows: accounting for human behavior in building simulation." Center for the Built Environment(CBE), University of California, Berkley., Tech. Rep., 2008. [Online]. Available: http://cbe.berkeley.edu/research/pdf_files/Borgeson2008-OperableWindowSimulation.pdf

[132] R. E. Nance, *A history of discrete event simulation programming languages.* ACM, 1996.

[133] M. Pidd, *Computer simulation in management science.* J. Wiley, 1998.

[134] I. Richardson, M. Thomson, and D. Infield, "A high-resolution domestic building occupancy model for energy demand simulations." *Energy and Buildings*, vol. 40, no. 8, pp. 1560–1566, 2008.

[135] J. Widén, A. M. Nilsson, and E. Wäckelgård, "A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand." *Energy and Buildings*, vol. 41, no. 10, pp. 1001–1012, 2009.

[136] M. Muratori, M. C. Roberts, R. Sioshansi, V. Marano, and G. Rizzoni, "A highly resolved modeling technique to simulate residential power demand." *Applied Energy*, vol. 107, pp. 465–473, 2013.

[137] H. Meidani and R. Ghanem, "Multiscale Markov models with random transitions for energy demand management." *Energy and Buildings*, vol. 61, pp. 267–274, 2013.

[138] M. L. Baptista, H. Prendinger, R. Prada, and Y. Yamaguchi, "A cooperative multi-agent system to accurately estimate residential energy demand." in *Proceedings of the 2014 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1405–1406.

[139] W. J. Stewart, *Introduction to the numerical solutions of Markov chains.* Princeton Univ. Press, 1994.

[140] A. J. Collin, G. Tsagarakis, A. E. Kiprakis, and S. McLaughlin, "Development of low-voltage load models for the residential load sector." *Power Systems, IEEE Transactions on*, vol. 29, no. 5, pp. 2180–2188, 2014.

[141] J. Conlisk, "Interactive Markov Chains." *Journal of Mathematical Sociology*, vol. 4, no. 2, pp. 157–185, 1976.

[142] M. Baptista, A. Fang, H. Prendinger, R. Prada, and Y. Yamaguchi, "Accurate household occupant behavior modeling based on data mining techniques." in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014, pp. 1164–1170.

[143] Y. Yamaguchi, M. Tanaka, and Y. Shimoda, "Comparison of occupant behavior models applied to a household." in *Proceedings of the 1st Asia Conference on International Building (ASIM12)*, 2012.

[144] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, and F. Descamps, "A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison." *Building and Environment*, vol. 75, pp. 67–78, 2014.

[145] Y. Yamaguchi and Y. Shimoda, "Evaluation of a behavior model of occupants in home based on Japanese national time use survey." *Proceedings of the 14th Conference of International Building Simulation (BS2015)*, pp. 617–624, 2015.

[146] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques." *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819–1835, 2009.

[147] M. Stokes, "Removing barriers to embedded generation: a fine-grained load model to support low voltage network performance analysis." Ph.D. dissertation, De Montfort University, 2005.

[148] U. Jordan and K. Vajen. DHW-calc Version 1.10. [Online]. Available: http://www.solar.uni-kassel.de [Accessed Nov. 23, 2013].

[149] U. Jordan and K. Vajen, "DHWcalc: Program to generate domestic hot water profiles with statistical means for user defined conditions." in *Proceedings of the ISES Solar World Congress, Orlando, FL, USA*, 2005, pp. 8–12.

[150] J.-P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, and C. Evans, "Household Electricity Survey: A study of domestic electrical product usage." May 2012, report issued by Intertek Testing & Certification Ltd for AEA Group. R66141. Final Report Issue 4. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/208097/10043_R66141HouseholdElectricitySurveyFinalReportissue4.pdf

[151] I. Richardson, M. Thomson, D. Infield, and A. Delahunty, "Domestic lighting: A high-resolution energy demand model." *Energy and Buildings*, vol. 41, no. 7, pp. 781–789, 2009.

[152] D. Robinson, U. Wilke, and F. Haldi, "Multi agent simulation of occupants presence and behaviour." in *Proceedings of the 12th Conference of International Building Simulation (BS2011)*, 2011, pp. 2110–2117.

[153] J. Palmer, N. Terry, and T. Kane, "Further Analysis of the Household Electricity Survey - Early Findings: Demand side management." *Report for DECC and DE-*

FRA, Nov. 2013. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/275483/early_findings_revised.pdf

[154] Y. Rubner, C. Tomasi, and L. J. Guibas, “The earth mover’s distance as a metric for image retrieval.” *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[155] E. McKenna and M. Thomson, “High-resolution stochastic integrated thermal–electrical domestic demand model.” *Applied Energy*, vol. 165, pp. 445–461, 2016.

[156] L. Weston, “Shift work,” 2014, HSE 2013: Vol 1 - Chapter 6. [Online]. Available: http://content.digital.nhs.uk/catalogue/PUB16076/HSE2013-Ch6-sft-wrk.pdf

[157] M. Steel, “Changes in Shift Work Patterns Over the Last Ten Years, 1999-2009.” 2010, office for National Statistics. RR887 Research Report. [Online]. Available: http://www.hse.gov.uk/research/rrpdf/rr887.pdf

[158] T. Mergoupis and M. Steuer, “Holiday taking and income,” *Applied Economics*, vol. 35, no. 3, pp. 269–284, 2003.

[159] J. Young and E. Hughes, “BDRC Continental Holiday Trends 2014 Report.” 2014. [Online]. Available: https://bdrc-continental.com/projects/holiday-trends-2014/

[160] Office of National Statistics. Annual Population Survey (Sep. 2014): Homeworking by employment type. [Online]. Available: http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/published-ad-hoc-data/labour/march-2015/homeworking-by-employment-type.xls [Accessed Sep. 27, 2015].

[161] Office for National Statistics., “Quarterly Labour Force Survey, April - June, 2015.” 2016, [data collection]. 3rd Edition. UK Data Service. SN:7781. [Online]. Available: http://dx.doi.org/10.5255/UKDA-SN-7781-3

[162] N. Bloom, J. Liang, J. Roberts, and Z. J. Ying, “Does working from home work? Evidence from a Chinese experiment.” *The Quarterly Journal of Economics*, vol. 130, no. 1, pp. 165–218, 2015.

[163] E. McKenna, M. Krawczynski, and M. Thomson, "Four-state domestic building occupancy model for energy demand simulations." *Energy and Buildings*, vol. 96, pp. 30–39, 2015.

[164] F. McLoughlin, A. Duffy, and M. Conlon, "Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study." *Energy and Buildings*, vol. 48, pp. 240–248, 2012.

[165] J. Buzek and D. Lopez Garrido, "Directive 2010/30/EU of the European Parliament and of the Council of 19 May 2010 on the indication by labelling and standard product information of the consumption of energy and other resources by energy-related products." May 2010. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0030&from=en

[166] DECC. Energy Consumption in the UK (2015). Chapter 3: Domestic data tables. Table 3.13 - Total stock of appliances by energy rating in the UK, 1996 to 2013. [Online]. Available: http://webarchive.nationalarchives.gov.uk/20150423095958/https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/375388/domestic.xls [Accessed Aug. 04, 2016].

[167] I. Mansouri, M. Newborough, and D. Probert, "Energy consumption in UK households: impact of domestic electrical appliances." *Applied Energy*, vol. 54, no. 3, pp. 211–285, 1996.

[168] CIBSE, "Code for Lighting: Part2," *The Chartered Institute of Building Services Engineers*, 2002.

[169] Market Transformation Programme, "BNDL KO01: Domestic Lighting - Government Standards Evidence Base 2009: Key Outputs," *A Briefing Note prepared for DEFRAs Market Transformation Programme, Version 1.1*, June 2010. [Online]. Available: http://efficient-products.ghkint.eu/spm/download/document/id/793.pdf

[170] J. Bright, C. Smith, P. Taylor, and R. Crook, "Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data." *Solar Energy*, vol. 115, pp. 229–242, 2015.

[171] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases." *Knowledge and Information Systems*, vol. 3, no. 3, pp. 263–286, 2001.

[172] Elexon. Average profiling data per Profile Class (regression data evaluated at 10-year average temperature. 2013-14. [Online]. Available: https://www.elexon.co.uk/wp-content/uploads/2012/01/Average_Profiling_data_201314_evaluated@10yearNET_v1.0.xlsx [Accessed Aug. 11, 2016].

[173] Scottish Power Energy Networks., "Ashton Hayes sub-station electricity monitoring dataset." 2012, Received from Scottish Power Energy Networks on 18th November 2013.

[174] DECC. LSOA estimates of households not connected to the gas network: 2014. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/497515/LSOA_estimates_of_households_not_connected_to_the_gas_network_2014_published_Jan_2016v2.csv/preview [Accessed Mar. 26, 2016].

[175] DECC/BRE, "Energy Follow Up Survey - Report 5: Secondary heating systems." Dec. 2013. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/274774/5_Secondary_Heating.pdf

[176] Energy Savings Trust, "At home with water." 2013. [Online]. Available: http://www.energysavingtrust.org.uk/sites/default/files/reports/AtHomewithWater(7).pdf

[177] L. Shorrock, "Analysis of the EST's domestic hot water trials and their implications for amendments to BREDEM and SAP." Mar. 2009. [Online]. Available: http://www.bre.co.uk/filelibrary/SAP/2012/STP09-DHW01_Analysis_of_EST_DHW_data.pdf

[178] Market Transformation Programme, "BNW AT08: Modelling projections of water using products," *A Briefing Note prepared for DEFRAs Market Transformation Programme, Version 1.1*, Mar. 2011. [Online]. Available: http://efficient-products.ghkint.eu/spm/download/document/id/961.pdf

[179] Energy Savings Trust, "At home with water 2." 2015. [Online]. Available: http://www.energysavingtrust.org.uk/sites/default/files/reports/AHHW2%20final.pdf

[180] J. Palmer, N. Terry, S. Firth, T. Kane, D. Godoy-Shimizu, and P. Pope, "Further Analysis of the Household Electricity Survey Energy Use at Home: Models, Labels and Unusual Appliances." *Report for DECC and DEFRA*, Feb. 2014. [Online]. Available: http://www.carltd.com/sites/carwebsite/files/Report%203_Models,%20labels%20and%20unusual%20appliances.pdf

[181] C. Bareczko-Hibbert, "After Diversity Maximum Demand (ADMD) Report." Feb. 2015, durham University for Customer-Led Network Revolution. Doc. No. CLNR-L217. [Online]. Available: http://www.networkrevolution.co.uk/wp-content/uploads/2015/02/After-Diversity-Maximum-Demand-Insight-Report.pdf

[182] Office of National Statistics., "2011 Census: Key Statistics for England and Wales, March 2011." Dec. 2012, Statistical Bulletinh. [Online]. Available: http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuskeystatisticsforenglandandwales/2012-12-11#rooms-bedrooms-and-central-heating

[183] Office of National Statistics. 2011 Census: Table KS105EW - Household Composition. [Online]. Available: https://www.nomisweb.co.uk/census/2011/ks105ew [Accessed Apr. 17, 2014].

[184] F. J. Born, J. Clarke, C. Johnstone, and N. Smith, "Merit - An evaluation tool for 100% renewable energy provision." in *Proceedings of the Renewable Energies for Islands Conference, 2001*, 2001.

[185] S. Chen, H. B. Gooi, and M. Wang, "Sizing of energy storage for microgrids," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 142–151, 2012.

[186] Q. Fu, L. F. Montoya, A. Solanki, A. Nasiri, V. Bhavaraju, T. Abdallah, and C. Y. David, “Microgrid generation capacity design with renewables and energy storage addressing power quality and surety,” *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 2019–2027, 2012.

[187] S. Dutta and R. Sharma, “Optimal storage sizing for integrating wind and load forecast uncertainties,” in *Proceedings of 2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*. IEEE, 2012, pp. 1–7.

[188] T. Niknam, F. Golestaneh, and A. Malekpour, “Probabilistic energy and operation management of a microgrid containing wind/photovoltaic/fuel cell generation and energy storage devices based on point estimate method and self-adaptive gravitational search algorithm,” *Energy*, vol. 43, no. 1, pp. 427–437, 2012.

[189] Z. Zhu, J. Tang, S. Lambotharan, W. H. Chin, and Z. Fan, “An integer linear programming based optimization for home demand-side management in smart grid,” in *Proceedings of 2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*. IEEE, 2012, pp. 1–5.