This version is available at https://strathprints.strath.ac.uk/59557/

A disaggregated, probabilistic, high resolution method for assessment of domestic occupancy and electrical demand

Graeme Flett*, Nick Kelly

Energy Systems Research Unit (ESRU), Department of Mechanical and Aerospace Engineering, University of Strathclyde, Glasgow, UK

Abstract

An integrated domestic occupancy and demand model with a 1-minute resolution has been developed which better captures the influence of different occupant behaviours than previous models. The occupancy model includes the fundamental link between occupancy and demand, and differentiates between different types and sizes of households. In particular, the likelihood of daytime occupancy is captured by age and employment differentiators. A novel method for identifying appliance use events and linking use to an occupancy profile has been developed that accounts for household specific appliance usage using an event-based approach calibrated directly from measured data. The method has been shown to perform better than both per-timestep probability models and models calibrated from time-use survey activity diaries. To further capture individual household behaviours, the demand model incorporates additional factoring to account for income and random behavioural influences. Whilst improving differentiation of individual household energy usage, due to limitations in the available data, the model incorporates some occupancy and use behaviour factors that are a composite of multiple households, leading to some behaviour averaging in the model output; consequently the model is best employed for energy demand assessment of multiple households.

Keywords: demand, occupancy modelling, disaggregated, building simulation, probabilistic model

1. Introduction

Demand prediction and its associated modelling methods have a variety of applications in energy system analysis.

At the regional and national-level, built environment demand prediction using stock models is used for centralised generation planning and infrastructure decisions. These models (e.g. SAP, BREDEM in the UK) make use of average energy usage characteristics, derived from large scale sources such as time-use and low-resolution demand surveys. However, the increasing focus on distributed generation and microgrids, involving the decentralisation of energy supplies ranging from town/district-scale energy systems to individual household microgeneration, requires that demand prediction is done at a very much smaller scale; a scale at which the diversity of demand is reduced, as distinct energy use behaviours associated with different types of end-user (e.g. retirement, suburban) and household types (e.g. retired, family etc.) become more influential. It is therefore essential to understand the variation in energy-use behaviours in order to better predict localised demand and thereby help develop better performing and more robust local energy systems in which energy supply and demand are well matched.

Focusing on electrical demand prediction, the work presented in this paper introduces several modelling techniques which aim to better account for household-type and specific household energy use behaviours. Facilitating the generation of more representative, high-resolution occupancy and demand profiles at the local scale, derived from combining specific, detailed, household demand profiles.

2. Demand Prediction Background

Analysis of the distribution of average power use per household shows a distinct pattern that is consistent across different types and scales (electricity, heat, appliance-specific demand etc.). Specifically, the majority of households are characterised by a broadly linear variation about the mean with a small percentage of extreme outliers, particularly at the high end of the distribution. This is illustrated in Figure 1 which shows the distribution of average electrical power demand from the 251-household Household Electricity Survey dataset [1] collected in the UK.
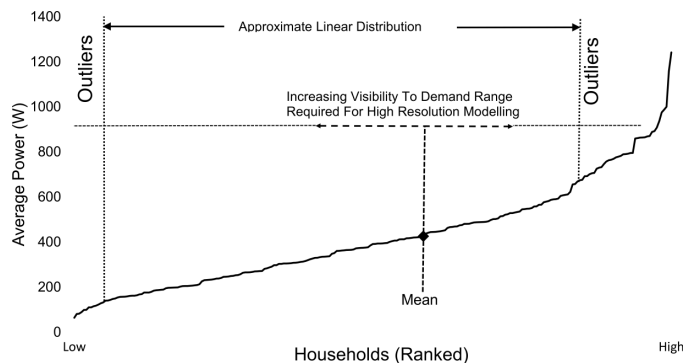

Fig. 1. Increasing importance of overall demand range and outliers for small-scale, high-resolution demand analysis.

*Corresponding author. E-mail address: graeme.flett@strath.ac.uk (G.Flett)

With small populations of energy users, such as the numbers of customers found in small energy networks, it is necessary to account for the full range of potential behaviours, including outliers, as these can have a significant impact on energy demand [2]. This would not be the case with larger networks where the scale ensures the overall energy demand converges to the mean.

2.1. Demand Factors

A number of factors have been shown to influence household energy demand characteristics that need to be accounted for in a small-scale demand model. Yohanis [3], and Haldi and Robinson [4] have shown that these include, but are not limited to; floor area, household size, bedroom number, occupant age, income, children, employment status, and tenure.

There remains a residual, seemingly random, element of overall and appliance-specific behaviour that cannot be directly attributed to identifiable household characteristics [5]. This could be due to different attitudes to energy use, prioritisation of energy spending, or individual needs driving specific appliance ownership and use. Haldi and Robinson [4] estimated that the behavioural impact on energy demand was at least a factor of two. Any predictive model for demand at the community scale therefore needs to account for both the influence of identifiable household characteristics and the stochastic element related to individual behaviours.

2.2. Demand Modelling Methods

Energy models can be broadly characterised by whether they are top-down or bottom-up, and deterministic or probabilistic.

- *Top-Down vs. Bottom-Up* - 'Top-Down' energy models typically use national-level data and linear regression to determine the influences of individual user variables for prediction. 'Bottom-up' models use high-resolution detailed usage data to build up models for individual elements that are combined to reflect the uniqueness of individual behaviours. [11]

- *Deterministic vs. Probabilistic* - 'Deterministic' models assume that behaviours and responses are predictable and repeatable, and can be modelled explicitly. 'Probabilistic' models attempt to reproduce the randomness and uniqueness of individual responses statistically.

Whilst at a high resolution human behaviour is inherently probabilistic, a lack of high quality behavioural data has led to the development of deterministic models and hybrid models combining both deterministic and probabilistic elements using either archetypes (example behaviours assumed to be representative) or composites of individual behaviours. For example, the occupancy model of Aerts et al [6] uses a probabilistic Markov Chain approach but is calibrated using seven archetypal occupancy profiles.

Capturing specific household demand variations is suited to bottom-up, probabilistic modelling approaches whilst recognising that data and computational limitations may require some simplification using macro-scale regression, composite behavioural models, etc. An example of the composite approach is the occupancy model used in this paper [12], which uses probabilistic calibration data from multiple individuals with similar characteristics, such as ages and relationships, to represent the behaviour of the group. This results in a model that exhibits variation between individual households but which converges to the overall group average when run for hundreds of cases. Similarly, the demand model of Richardson et al [7] predicts individual appliance uses using a per-timestep probability approach but with the probability model calibrated using the Time-Use Survey activity diaries from a large number of households with the same number of occupants. While the potential for convergence has been acknowledged [13], there has been no detailed evaluation of the potential impact on demand prediction accuracy.

2.3. Occupancy-to-Demand Relationship

A key determinant of the observed variation in total demand and demand timing between different household types and individual households is the extent to which the house is occupied [14]. In particular, large daytime occupancy variations driven by employment-related absences are clearly discernible.

Figure 2 shows that the daytime electrical demand profiles for different household types, if divided by the proportion of households with an active occupant at the same timestep, become broadly linear, indicating that a significant proportion of electricity use is occupant initiated. Whilst there is a degree of residual variation that is attributable to specific daytime activities, such as meal-times, and to constant demands (e.g. cold appliances), high-resolution, small-scale demand modelling clearly requires a robust linkage between occupancy and occupant-initiated energy use.
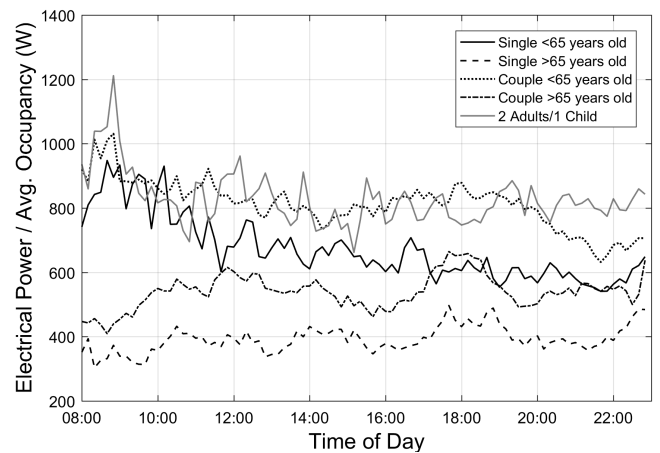
Fig. 2. Average electrical demand divided by average active occupancy for selected household types.

With a view to future changes in demand, it is also important to understand specific appliance use in more detail as initiatives such as demand shifting are highly appliance specific; with the greatest potential for washing and cooking activities [15]. Highly disaggregated, bottom-up demand models, which include realistic occupancy prediction and variation, allow the

Table 1
Existing high-resolution demand models.

Model	Occupancy Data	Occupancy Method	Occupancy Basis	Demand Basis
Aerts [6]	TUS	Survival Probability / Seven Typical Profiles	Profile-differentiated individual occupant model	Probabilistic power and duration linked to TUS Activity
Richardson [7]	TUS	First-order Markov / Occupancy States	Occupant number	Per-timestep probability / TUS activity-linked
Stokes [8]	BREDEM	BREDEM / Basic Occupancy	Occupant number with additional type factoring	Per-timestep probability / Demand data calibrated
Widen [9]	TUS	First-order Markov / TUS Activities	Undifferentiated individual occupant model	Directly TUS activity-linked
Wilke [10]	TUS	Survival Probability / TUS Activities	Household-type differentiated individual occupant model	Per-timestep probability / TUS activity-linked

occupancy influence on individual appliance use to be assessed as a baseline for the impact of additional demand shifting mechanisms.

2.4. Existing Occupancy-to-Demand Models

Existing domestic demand models can be split into two broad categories based on time resolution.

The first type are low resolution, top-down, deterministic models that seek to use generic rules to assess typical monthly or annual demand. An example of this type is the UK BREDEM model [16], and is used primarily to assess relative demand across the entire housing stock. These typically do not explicitly account for occupancy variations.

The second type are high resolution models that attempt to predict demand at sub-hour timesteps. These are typically bottom-up, probabilistic models that use occupancy as a key variable. Grandjean et al [17] compared the existing high resolution models in detail, many of which were calibrated with time-use survey (TUS) data.

Table 1 details several existing modelling methods that link occupancy to demand. A review of each has identified several areas for improvement that are common to some or all methods:

- *Undifferentiated Occupancy Profiles* – With the exception of Wilke [10], the models do not incorporate significant differentiation of the occupant and household types to capture specific occupancy patterns.

- *Activity and Use Time Sequencing* – The existing models typically determine whether an appliance demand has been initiated using a per-timestep analysis with either fixed or time-dependent probabilities. This approach does not consider either previous uses, previous TUS activity sequences, or realistic time gaps between uses, when predicting the timing of the next demand. While this type of 'memoryless' model can be used for low resolution modelling, it has the potential to generate an unrealistic distribution of uses when individual days are assessed.

- *Linking TUS Activities to Specific Appliances* – Widen and Wackelgard [9], and Wilke [10] incorporate TUS activities directly in the occupancy model, and Aerts [6] and Richardson et al [7] use TUS activity probability separately in the demand model. However, TUS activity def-

initions are very broad and, with few exceptions, do not necessarily infer use of a specific appliance. For example, the TUS 'Food Prep' activity category accounts for all cooking and food preparation activities and could involve the use of a number of different appliances (e.g. cooker, microwave, kettle, grill).

Figure 3 shows the average daily probability profile for the 'Food Prep' activity from the UK 2001 Time-Use Survey [18] compared to the demand profiles for three key cooking appliances and for all cooking appliances. The demand profile has been divided by the 'Food Prep' probability and the resulting plot shows no obvious correlation. This effect can also be shown for other TUS activity-appliance combinations, with the exception of TV use. Consequently, it can be concluded that activity data from time-use surveys alone is a weak proxy for individual appliance use.
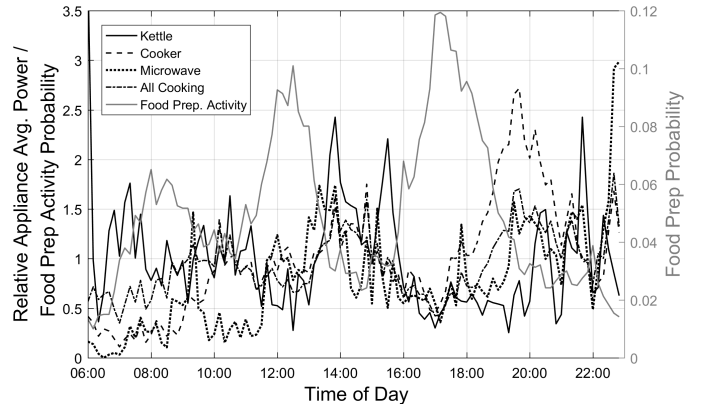
Fig. 3. Average relative cooking appliance power normalised for TUS 'Food Prep' activity probability.

3. Aim

The aim of the work presented here was to develop a high-resolution, electrical demand model that addressed the highlighted areas for improvement: 1) improving differentiation between households based on characteristics, 2) better capture of behavioural variations between households probabilistically, and 3) improving realism in the prediction of appliance use events.

The need to develop a model that captures, as far as possible, individual behaviours is very suited to a bottom-up, probabilistic approach with disaggregation of individual energy uses.

4. Demand Model Development

The new model builds on a previously developed household-differentiated occupancy model [12] (see 4.1) in order to simulate electrical demand for different household types. The contributions that this demand model makes are as follows:

- To capture the identified variation in demand between similar households, a set of overall household variance factors were developed that define appliance ownership and demand multipliers based on a) income, b) random energy use behaviours, and c) relative occupancy, that are applied probabilistically for each household modelled.

- Use patterns and typical power profiles for the most commonly owned appliances have been analysed and separate demand sub-models developed for three identified appliance groups that have similar characteristics. Each sub-model has a discrete-event based algorithm to determine the number of appliance uses per day which incorporate the identified behavioural and occupancy-driven factors.

- A discrete-event approach has also been developed to identify appliance use timing. Separation of the determination of the predicted number of uses and their timing is a different approach to existing models and is shown to be more effective at generating realistic appliance use timing sequences.

The following sections describe each of these aspects of the model in more detail. The developed model also includes modules for other types of demands (i.e. lighting, constant use appliances, low ownership appliances) which are based on previously published work ([19], [7], [20]) and are therefore not described in detail.

4.1. Household-Differentiated Occupancy Model

The underpinning occupancy model is described in a previously published paper by Flett and Kelly [12] and focuses solely on basic occupancy states (i.e. sleep, active, absent). Consequently only a summary is presented here. The occupancy model built on the Time Use Survey (TUS) calibrated, Markov-Chain probability approach developed by Richardson et al [7] with three key improvements.

- The model was calibrated using multiple occupant, household, and day types to better capture household-specific patterns. Working and non-working days were differentiated with separate profiles for weekdays, Saturdays and Sundays. TUS respondents submit one weekday and one weekend day occupancy diary, therefore each type module is calibrated from the occupancy data of multiple individuals.

- Interactions between related adults (i.e. co-habiting couples and parents) were captured by treating related pairs of occupants as a single entity. Child occupancy was directly linked to parent occupancy using a simplified Markov-Chain model. Other occupant types (e.g. single householders, adult children and households comprising unrelated adults etc.) were modelled as unlinked individuals. Each overall household occupancy profile are therefore generated from the relevant combination of couple, individual adult, and child models.

- The occupancy model also uses a higher-order Markov technique that takes the duration of an activity into account when predicting future activity. Separate Markov probability matrices are generated based on ranges of durations of typically between 30 minutes and 3 hours depending on the likelihood of the particular transition.

The performance of the model was compared with both the original Richardson et al method [7] and the event/duration based probability method developed by Wilke [10], and was shown to be an improvement [12].

4.2. Electrical Demand Data Analysis

The following sections detail the methods used to analyse the Household Electricity Survey (HES) dataset [1] to generate calibration data for the modelling of household-specific appliance use frequencies and timing. In the HES dataset individual appliances in 251 representative households were monitored for at least 1 month at a 2-minute resolution, with 26 households also monitored for a full year at a 10-minute resolution. Despite only comprising private households, the range of household types and social classes was nationally consistent. It was therefore assumed that the HES dataset was broadly representative of appliance use for UK households.

In order to develop a high-resolution demand model that predicts household-specific electrical demand behaviours and reflects variation between households, two main types of electrical dataset analysis were undertaken. The first focused on overall variation in energy demand between households of the same type. The second focused on appliance use behaviours, and variations between and within household type groups.

4.2.1. Overall Demand Analysis

Initial analysis was undertaken by splitting the HES households into eight basic types (mirroring the household differentiation in the previously reported occupancy model [9]). With the number of each household type in the dataset shown in brackets, the eight types were: 1-person (working age (36) and retired (34)), retired couple (54), family (2/3-person (24) / 4-person (42) / 5+ (12)), and multi-adult (2-person (32) / 3+ (17)).

Assessment of the average demand profiles for each of the defined household types (see Figure 13) showed distinct characteristics that would be expected from the different inherent lifestyles. For example; retired households show a more balanced demand across the day than their working-age equivalents; multi-adult households have less distinct morning and daytime demand patterns than family households. However,
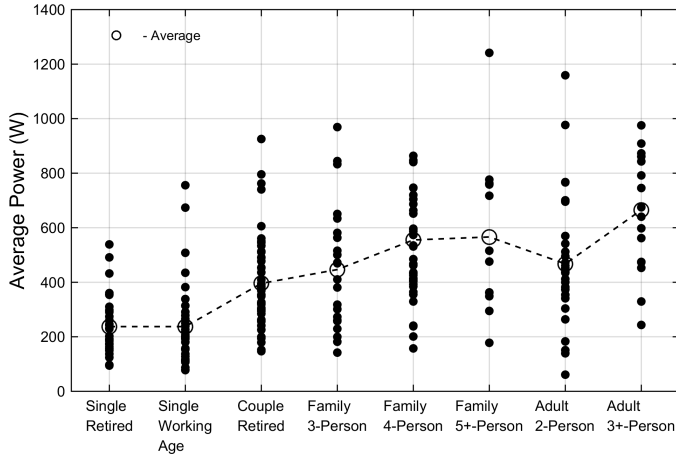
Fig. 4. Average electrical demand distribution per household type from the Household Electricity Survey (HES) dataset.

there are also characteristics that are common, such as the early evening demand increase driven primarily by higher occupancy, cooking activities, and increased lighting demand.

Further analysis of individual household occupancy and demand data shows a high degree of variability in behaviour, even between those households with broadly similar characteristics. Figure 4 shows the range of average power demand values for each of the eight defined household types from the HES dataset. It is worth noting that there is greater variation within household-type groups than between different groups. This suggests that factors such as appliance ownership, income, and energy use behaviours are at least as important as household type and that a significant proportion of the variation in electrical demand between households cannot be attributed solely to the household characteristics (i.e. type, size, age).

4.2.2. Specific Appliance Analysis

There are thirteen primary occupant-initiated electrical appliances which are commonly owned and for which a significant depth of use data was available; kettle, microwave, toaster, cooker, oven, washing machine/washer dryer, tumble dryer, dishwasher, laptop computer, desktop computer, iron, vacuum cleaner and hair dryer. Computer use is used to infer use of routers, monitors and printers. The remainder of this paper is focused on the model development for these key occupant-initiated appliances. Televisions, lighting, and low-ownership appliances are modelled using other methods that, for brevity, are not covered in this paper.

For each of the thirteen identified appliances, the HES dataset was analysed for use frequency, duration, and timing per household and then per defined household type as the basis for model calibration. A typical example of the generated calibration data is the average daily use data shown in Table 2 for kettles and microwaves.

4.2.3. Appliance Sub-Groups

Analysis of the HES dataset identified that the 13 primary appliances fall into three distinct groups based on typical patterns and durations of use:

Table 2
Average number of use events per day per household type for kettles and microwaves.

Household Type	Kettle	Microwave
Single Non-Pen	3.37	0.86
Single Pensioner	3.98	0.84
Two-Adult Non-Pen	4.59	1.34
Two-Adult Pensioners	6.10	1.34
Three-Adult+	6.50	2.14
3-Person Family	4.94	1.81
4-Person Family	4.94	2.45
5+-Person Family	6.22	3.09

- *'Simple'* appliances (kettle, toaster, microwave, and hair dryer) have a high daily usage probability, usage that is generally independent of use on previous days, and a limited range of use durations. The HES dataset was analysed for average and maximum uses per day per household, the ratio of average daily household use to the household-type average, and the range of total energy used per use.

- *'Flexible'* appliances (dryers, cookers/ovens, computers) have significant variation in daily use probability and duration, and are therefore defined by a sub-model that links both factors probabilistically. The HES dataset was analysed for daily usage probability, total daily use duration, and number of uses per day for each household type.

- *'Fixed'* appliances (washing machines, dishwashers, irons, vacuum cleaners) have a low multiple use per day potential, a use probability that is related to time since previous use, and a typical range of fixed or limited use durations. The HES dataset was analysed for daily usage probability, number of uses per day of use for each household type, and for typical power profiles and durations.

Three distinct demand sub-models were created in order to reflect the different usage patterns for each appliance sub-group (see 4.6).

4.3. Occupant-Initiated Appliance Use Model Overview

Based on the dataset analysis the required calculation sequence for a differentiated, probabilistic, bottom-up electrical demand model focused on the identified primary occupant-initiated appliances was determined. Five distinct sequential elements were identified as follows:

- *Household Behaviour Factor* - To account for the observed intra-household-type variation detailed in 4.2.1, in addition to appliance-ownership which is captured separately, the following were identified as potential causes; 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 model incorporates the influence of each as a combined multiplier as detailed in 4.4.

5

- *Appliance Ownership* - A 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. This element is not covered in this paper.

- *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 behaviour 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 timing is also captured. These elements are detailed in 4.5.

- *Daily Use Determination* - A discrete event-based approach [21] has been employed to separate the determination of the number of appliance uses per day and the timing. A use is defined as a separate demand event for appliances that are not in constant use. The model determines average daily use per household, and then use on individual days based on the average with factoring for occupancy and random variation. The average use is influenced by both the household- and appliance-level behaviour factors identified. The model basis for each of the three appliance types identified in 4.2.3 is described in 4.6.

- *Use Timing Determination* - The start time of each individual use is determined based on the modelled occupied periods and the probability distribution of start times for a specific use event (i.e. #x of total y per day). This differs from existing models which tend to incorporate a per-timestep probability calculation sequence calibrated to achieve an average use frequency. The new method was developed primarily to address the problems of unrealistic use sequencing and timing (see 2.4), where previously developed methods have not addressed the link between the number of uses in a day and their timing, and is described in 4.7.2. A performance comparison of the per-timestep and newly developed method is detailed in 5.1.1.

4.4. Household Behaviour Factor Development

As outlined, differences in overall 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 (HBF) is determined within the developed model by combining an income (IBF), overall relative occupancy (OROF), and random energy-use behaviour (RBF) factor as shown in Equation 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 4.5.1). This factor is applied either to daily use number, use probability, or total daily usage duration as a single multiplier depending on the appliance model type (see 4.5.2). The following sections outline how each factor was identified and is determined for each modelled household.

$$HBF = IBF \times RBF \times (OccUse\_H/OccUse\_T)^{OROF} \quad (1)$$

4.4.1. Effect of Income on Demand: Income Behaviour Factor

Analysis of the specific behavioural effect of income on energy use is rare. Data from both Jamasb and Meier [22] and White et al [23] allows the overall income effect on electrical demand to be determined. However, this indicated that the overall impact is not solely behavioural but also influenced by household type and size, occupancy, and appliance ownership; the influence of which are captured separately in the developed model. The plot of electricity demand against "equivalised" income (defined below) shown in Figure 5 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-based method was therefore developed that allowed these other income-related factors to be accounted for separately and the residual behavioural effect of income on energy use captured.

Published UK 2011 Census data has been separated into different sizes of area for comparative analysis. Area annual electricity demand data [24] is available down to the Lower Layer Super Output Area (LSOA) level for England and Wales, this corresponds to groups of households of between 600 and 1000. An LSOA is assumed to be sufficiently small that each area has distinct characteristics that can be used 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 (see below) 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.
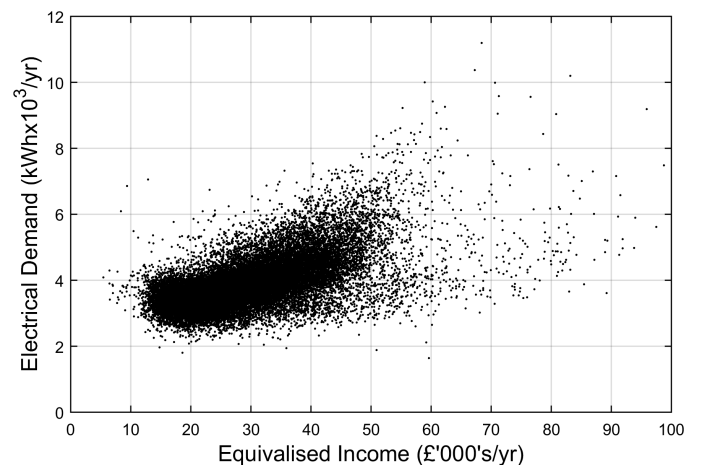

Fig. 5. Average household annual electricity consumption (kWh) by average equivalised income per English Lower-Layer Super Output Area (LSOA).

Using UK 2011 Census data various factors were determined for each LSOA to represent factors captured separately in the

model. The following were found to generate the most accurate regression model based on F-Stat and RMSE analysis:

- *SAP-Factored Occupant Number* – Number of people factored using the SAP N^0.4714 basis for demand relative to the household size (N) [16]. This value was shown to be better correlated with demand than the unfactored occupant number.

- *"Equivalised" Income* – Average gross income (2011 basis) factored by household size and reducing influence of number of people on expenditure based on the OECD method [25] (1st Person=0.58, 2nd Person=0.42, 3rd Person+=0.3). This value was shown to be better correlated with demand than the unfactored gross income.

- *Rooms per Person* – Average number of habitable rooms per person. Habitable rooms are living rooms, kitchens, bedrooms etc. but not toilets and hallways. This value was shown to be better correlated with demand than the unfactored 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. The demand model already accounts for this per-household factor.

- *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 for those factors). This demonstrates that the influence of increasing disposable income on electricity demand is significant.

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 1 to reflect the income-behaviour impact (based on 2011 "equivalised" income (IncE) values) as a relative Income Behaviour Factor (IBF).

$$IBF = 0.9852 + (0.440/3863.3) \times (IncE/10000)^{4.2} \quad (2)$$

This factor is used as a multiplier, first incorporated in the household behaviour factor (see 4.4), and then used to manipulate the appliance type-specific primary use defining parameter (daily number, probability, or total duration) used for each modelled appliance (see 4.6).

4.4.2. Random Energy-Use Behaviour Factor

Gill et al [5] determined that 37% of total household electricity use can be attributed to behaviour independent of identifiable household characteristics. An additional 'random behaviour' factor (RBF) is therefore applied; selected randomly between 0.77 and 1.23 (equivalent to a 37% variation). Assuming a linear distribution is likely to be an over-simplification, but further work is required to determine if there is a more complex link between this factor and other household characteristics.

As with the IBF, the Random Behaviour Factor (RBF) is first incorporated in the household behaviour factor (see 4.4), and then used to manipulate relative appliance demand per household (see 4.5.2).

4.4.3. Relative Occupancy Factors

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. Calibration of the model, 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 relative occupancy of the household to the household type average (OROF); and second that average use is determined for each household and adjusted for daily occupancy relative to the household average (DROF). Here, occupancy is defined as the proportion of time the house is actively occupied (i.e. not sleeping) 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 Time-Use Survey (TUS) data [18].

The relationship between relative occupancy and relative appliance use is assumed to be a power law (see Equation 1), with an exponent between 0 and 1. A value of 1 would represent an appliance with which use frequency was directly proportional to active occupancy. In reality, the value is much less 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 3 to estimate 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.

$$ROF = log(Occ_{ret}/Occ_{work})/log(Cyc_{ret}/Cyc_{work}) \quad (3)$$

Table 3
Overall and daily relative occupancy factors (OROF and DROF) for different appliances.

	Kettle	Microwave	Toaster	Computers
Overall/Daily	0.3	0.1	0.05	0.5
	Wash. Mach.	Dryer	Dishwasher	Cooker/Oven
Overall	0	0	0.5	0.1
Daily	0.3	0.5	0.67	0.2

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 within the household model (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 frequency 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 3 shows the overall (OROF) and daily (DROF) relative occupancy factors used in the model.

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

4.5. Appliance-Level Variance Factors

4.5.1. Occupancy-Timing-to-Appliance Use Relationship

The Relative Occupancy factors defined in 4.4.3 are not applied directly to occupancy but to a modified occupancy factor that also reflects appliance use probability during the occupied period(s). In order to do this, a probability density function (pdf) distribution for use time probability was generated for each appliance and household type from 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 value per timestep is the element $rp_{cyc}(t)$ in Equation 4. Figure 6 shows the modified distribution for kettle use in 1-person, retired households.
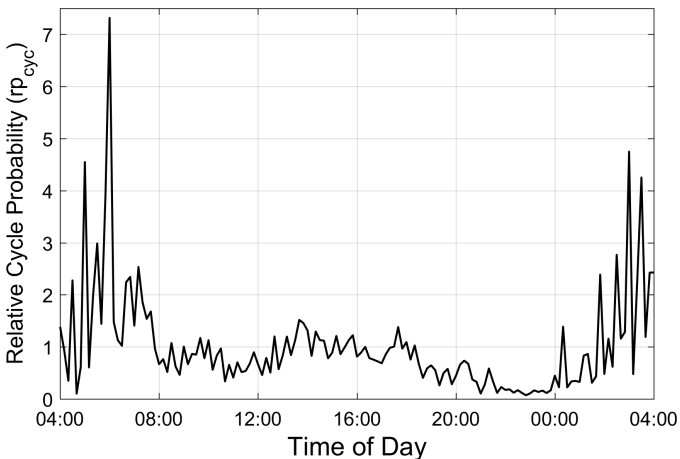

Fig. 6. Modified kettle use time probability with occupancy influence removed example for 1-person, retired households.

A combined factor ('OccUse') reflecting both occupancy and use probability summed for each of the 144 10-minute timesteps was then generated (see Equation 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.

$$OccUse = \sum_{t=1}^{144} p_{occ}(t) \times rp_{cyc}(t) \qquad (4)$$

The 'OccUse' factor is a unitless term that allows for a relative assessment of use potential rather than providing useful information directly. The average 'OccUse' value for a household (OccUse_H) compared to the population average (OccUse_T) gives a measure of 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 measure of the occupancy-driven variation in daily use to the household average. The determined ratios are factored by either the overall (OROF) or daily (DROF) relative occupancy factors as defined in 4.4.3.

4.5.2. Appliance Use Factor

For each appliance there is significant variation in the primary use defining parameter ('Simple' - number of daily uses, 'Fixed' - daily use probability, or 'Flexible' - daily use duration) for each household relative to the household type mean (see Table 2 for example). Analysis determined that for each of the appliances the ratio-to-(household type)mean distributions for each household type were similar and it was therefore acceptable to use a single combined distribution. The ratio-to-mean distributions for the 'Simple' appliances are shown in Figure 7. The distributions for all high-ownership appliances are similar. While the variation is partially attributable to household characteristics, a significant influence on the distribution is assumed to be random behavioural differences between households at the appliance level.
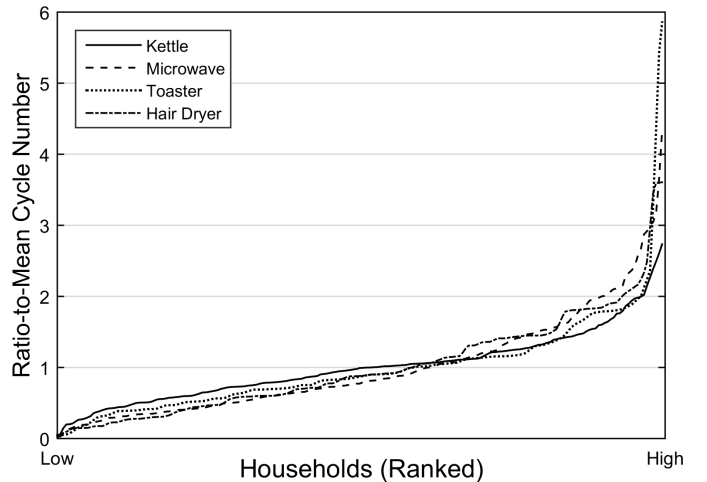

Fig. 7. Average household daily use ratio to household type mean distributions for all 'Simple' appliances.

For each modelled appliance, an 'Appliance Use Factor' (AUF) is randomly selected from the appliance-specific ratio-to-mean distribution. This is a multiplier that is combined with the average value of the appliance primary use defining statistic (see above) for the household type and the household behaviour factor (HBF) as per Equation 5 to determine the primary use

defining statistic for the 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 4.4. This suggests that any appliance-specific factoring could be allocated randomly and independently using the generated ratio-to-mean distributions, with additional modification using the household behaviour factor (HBF) (see Equation 5) to account for relative household-level behaviours.

4.6. Daily Use Determination Sub-Models

The following three sections outline the different methods used for the three distinct sub-groups of appliances defined in 4.2.3. The separate methods are used to predict number of daily uses and to define use energy and duration. A universal method is used to predict use time and is detailed in 4.7.2.

Prior to running the individual appliance sub-models, the 'Household Behaviour Factor' (HBF) is determined as outlined in 4.4 and the occupancy model is run to generate 1-minute resolution occupancy data for the household.

4.6.1. 'Simple' Sub-Model

Use Number Determination

The 'Simple' appliance use sub-model is applied to kettles, microwaves, toasters, and hair dryers and assumes the probability of use per day is independent of use on preceding and following days. The primary statistic that is used to define use per household is the average number of uses per day. The steps required to determine this value are as follows:

- *Step 1* - The 'Appliance Use Factor' (AUF) (see 4.5.2) is determined by randomly selecting a value from the appliance-specific ratio-to-mean distribution (see Figure 7).

- *Step 2* - The occupancy model output is converted to the combined factor that assesses both occupancy and appliance use likelihood ('OccUse') (see 4.5.1) 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 uses for the appliance (HhldUse) is determined from the following equation based on the average number of uses for the household type (TypeUse), 'AUF', and 'HBF' factors.

$$HhldUse = TypeUse \times HBF \times AUF \qquad (5)$$

- *Step 4* - For each modelled individual day, the household average number of daily uses is used to determine a baseline number of uses for the day 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 use number is the average predicted number based on the day-specific occupancy characteristics.

$$UseBase = HhldUse \times (OccUse\_D/OccUse\_H)^{DROF} \qquad (6)$$

- *Step 5* - The predicted number of uses for a specific day is determined using a binomial probability distribution (see below) with the baseline use number ('UseBase') as the average output and a random number generated between 0 and 1 to determine the specific value on the distribution to be used for each modelled day.

Binomial Use Probability Method

As outlined in Step 5 above, a binomial probability model is used to predict an actual number of uses in relation to the average predicted number of uses for a specific day. This further manipulation is based on the assumption that there is a natural variation in use 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 variations in use number in the HES dataset better than if not further manipulated and for other probability models, 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. For the use number model, N is the maximum use number for the household (probabilistically allocated from HES dataset analysis of the relationship between average and maximum use number values per appliance), and p the daily baseline use number (see Equation 6) divided by N. An example predicted use number distribution for a daily baseline of 3 uses and a household maximum of 6 uses is shown in Table 4.

Table 4
Binomial distribution probabilities for a daily use number baseline of 3 and household maximum daily number of uses of 6.

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

Power, Duration, and Energy

The 'Simple' model appliances are characterised by short, constant power uses. Each identified use in the HES dataset was analysed for total energy. Total energy is used rather than separate power and duration values to allow use data from multiple households to be combined for calibration purposes using a single value for simplicity.

The analysed data was converted to an average total energy per use for each household type and a ratio-to-mean distribution for the ratio to the household type mean of the total energy for each use ranked in ascending order. Each household is allocated a unit power from the observed distribution in the HES dataset, and the total energy allocated per use based on the average per household type and a randomly selected value from the ratio-to-mean distribution. The use duration is then determined from use energy divided by the unit power.

4.6.2. 'Flexible' Sub-Model

Use Number Determination

The 'Flexible' appliance use sub-model is used for cookers, ovens, dryers, and computer equipment. The primary statis-

tic used to define use per household is average total duration of use per day. The sub-model steps to calculate the number of appliance uses per day are as follows:

- *Steps 1-3* - Identical to Steps 1-3 for the 'Simple' model (see 4.6.1) except that average daily use duration (HhldDur) replaces average daily use number (HhldUse) as the output based on the household type average duration.

- *Step 4* - Using the Kernel Density method outlined below, the probabilistic relationship between average daily duration and use number is converted to a probability matrix that represents a series of use number cumulative probability distributions for the range of potential average daily durations increasing in equal increments. The average number of daily uses (HhldCyc) is determined probabilistically from the distribution appropriate for the determined average daily use duration (HhldDur).

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

$$CycDur = HhldDur/HhldCyc \qquad (7)$$

- *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.

$$DMR = (OccUse\_D/OccUse\_H)^{DROF} \qquad (8)$$

- *Step 7* - The HES data does not allow the relative impact of occupancy on number of uses and use duration to be determined. The model therefore assumes an equal impact with the daily baseline number of use (HhldUse) and use duration (CycDur) both being multiplied by the square root of DMR.

- *Step 8* - The predicted actual number of uses for a specific day is determined using a binomial probability distribution with the baseline use number as the average output in the same manner as the 'Simple' model (see 4.6.1).

Kernel Density Probability Method

The Kernel Density probability method used in Step 4 of the 'Flexible' sub-model is a standard method used to characterise a complex relationship between two variables while also allowing a degree of probabilistic variation to the specific behaviours used for calibration.

The method requires that each calibration 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 standard Gaussian kernel has been used in this case with each data point therefore represented by a normally distributed function.

An overall probability density function (pdf) distribution is determined by adding the contribution of each individual kernel at each point on the x-y surface. The continuous functions

x and y are reduced to a set of equal-sized discrete value ranges 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 range widths of 0.01 and 0.1 respectively. The width of the kernels is selected manually to achieve a distribution that retains the detail of the overall relationship but is not overly influenced by individual data points.

The pdf matrix is converted to a cumulative probability function (cdf) matrix along the axis related to the unknown variable. The unknown variable is then determined by generating a random number between 0 and 1 and finding the closest value in the applicable cdf distribution for the value of the known variable.
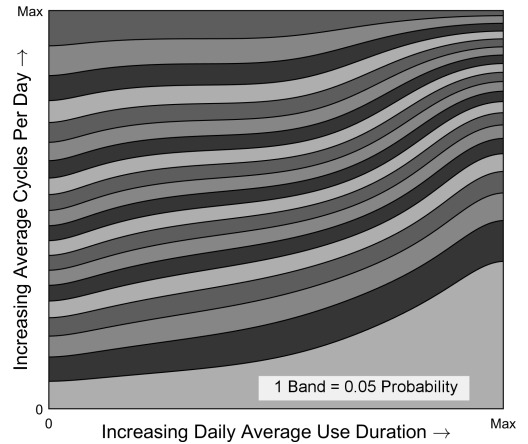

Fig. 8. Daily average use duration to average uses per day probability surface.

The resultant probability model is shown graphically in Figure 8. Each band represents a 0.05 probability, the x-axis daily use duration and the y-axis use number. A higher use duration increases the probability of a higher number of uses but the relationship is complex. This type of probability analysis captures the range of behaviours.

Power, Duration, and Energy

Cooker and oven demand profiles vary significantly. The majority of uses start with a period at a fixed power level as the element warms, followed by highly variable period of on/off power cycling at the same fixed power to maintain the chosen temperature or setting, which may also be regularly altered mid-use. To allow the overall impact on household power use to be adequately captured within a practical model, each identified use in the HES dataset was analysed for duration and a load factor, which is the average power per use divided by the observed unit peak power. Each modelled use was probabilistically allocated a duration and load factor based on this analysis.

Uses were modelled as an initial short (2-7 mins) period at peak power (peak power allocated probabilistically per household based on HES analysis) and then cycles of peak power use sequenced to achieve the target overall load factor. 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
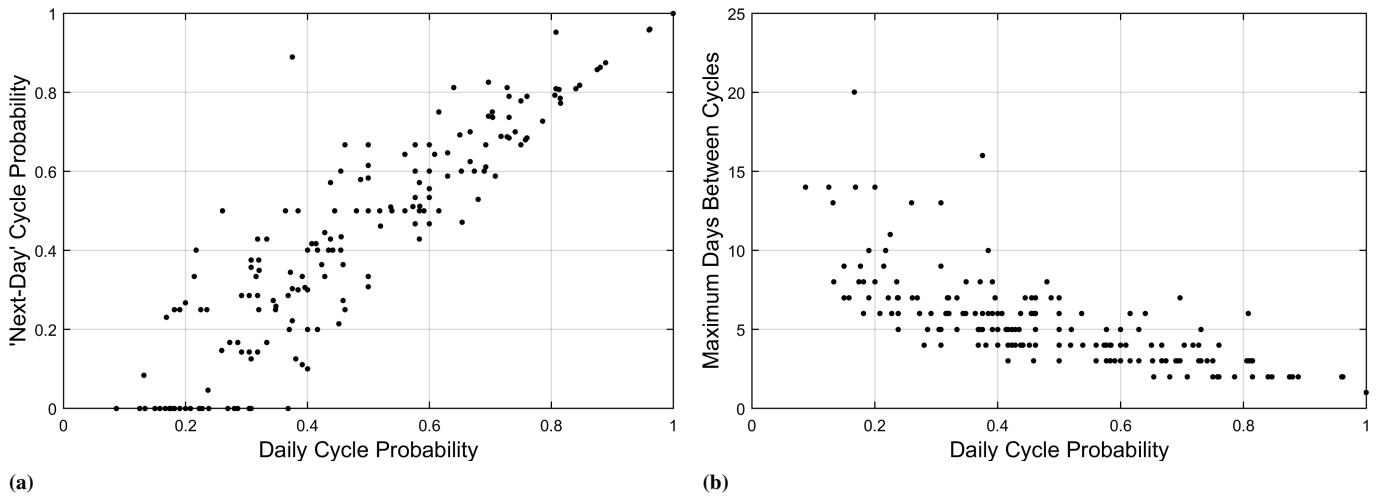
Fig. 9. Relationship between washing machine daily use probability and both 'next-day' use probability and maximum days between uses.

use, which is considered sufficiently accurate for an overall demand model.

For dryer and computer models, typical power levels and use durations were assessed directly from the HES dataset and applied probabilistically within the demand sub-model, as the power level was observed to be relatively constant during active use.

4.6.3. 'Fixed' Sub-Model

Use Number Determination

The 'Fixed' appliance use sub-model is used for washing machines, dishwashers, vacuum cleaners and irons. The primary statistic used to define household use of this type of appliance is daily use probability.

Analysis of the HES dataset also determined that there was a relationship between overall daily use probability and next-day (i.e. day following a use) probability (see Figure 9(a)), and an inverse relationship between daily use probability and the maximum time between uses (see Figure 9(b)). Beyond this the use sequences for households are highly variable. This analysis was used to model the observed variations in use probability with time since the previous use.

The Kernel Density method (see 4.6.2) has been used to allow the 'next-day' use probability and maximum time between uses to be probabilistically allocated to each modelled household based on the previously determined daily use probability. For each household the model then generates the probabilities for the number of days until the next use randomly but constrained by the 'next-day' probability, identified average, and maximum time between uses. These probabilities are further factored using the relative 'OccUse' factor for each day (see 4.5.1).

The following outlines the sequence of steps for the 'Fixed' appliance use probability analysis.

- *Steps 1-3* - Identical to Steps 1-3 for 'Simple' model except daily use probability (HhldPrb) replaces average daily

use number (HhldUse) as the output based on the household type average daily use probability.

- *Step 4* - The 'next-day' use probability and maximum number of days between uses is determined from the relationship to the daily use probability using the method defined above.

- *Step 5* - The cumulative probability function for number of days until the next use is determined using the method defined above.

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

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

Power, Duration, and Energy

For washing machines and dishwashers the HES dataset was analysed for typical power profiles differentiated by appliance energy rating. For each 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 for all modelled uses. Each household was probabilistically assigned an appliance energy rating and one of the four typical cycles per modelled use.

The cycle type allocation does not currently differentiate by household type due to difficulty in identifying appliance types and cycles for all households but analysis showed that the average power per use is similar for all households suggesting that use number is more significant than cycle type as a differentiator of relative power use.

11

4.7. Use Time Identification Sub-Model

4.7.1. Use Time Probability Distributions

Analysis of the distribution of use times per appliance showed that timing was dependent on household type, occupancy, number of uses per day, and specific time-specific drivers (e.g. meal times).

Separate start time probability distributions were generated for each primary appliance based on household type, total number of daily uses, and for each specific use number. Further, to allow these distributions to be used with the occupancy model output, the occupancy influence on each distribution needed to be removed. This is achieved by dividing the unmodified use 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 9) and generating cumulative probability models based on the modified pdf values (see Figure 10 for unmodified and occupancy-modified distribution examples).

$$pdf_{mod}(t) = pdf_{raw}(t)/(p_{occ}(t)/\bar{p}_{occ}) \qquad (9)$$

These modified distributions are used as the basis for linking occupancy with appliance use probability as described in the following section.
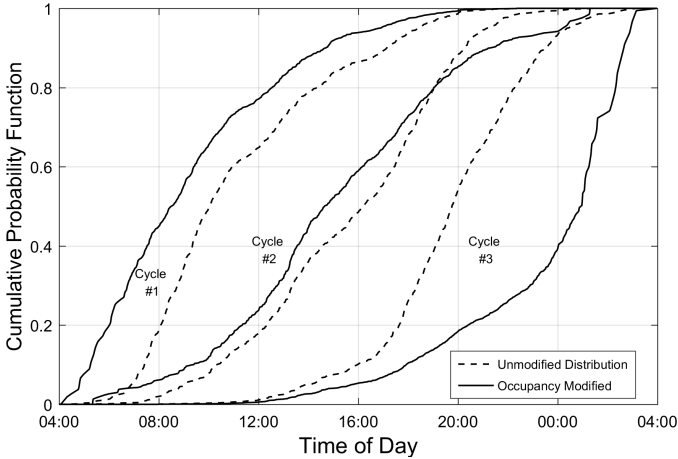

Fig. 10. Unmodified and occupancy-modified start time cumulative probability distributions example for a 3-use microwave use day.

4.7.2. Use Time Identification Method

For each household the appliance start times are determined based on the active occupancy periods generated by the occupancy model (see Section 4.1), the sub-model generated daily use number for each appliance (see 4.6), and the occupancy-modified, start time cumulative probability curves identified from the HES dataset for each appliance (see Figure 10 for example).

For each modelled day, the occupancy model output is converted into an event matrix that tracks occupancy transitions and appliance availability at the start and end of each period (see Table 5). Potential use 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.

Table 5
Daily event matrix example - microwave - Use #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
Cuml. Prob.	0.438	0.563	0.725	0.844	0.988	0.997	0.999	1.000

The appliance-use cumulative probability values at the start and end of each 'available' period are determined from the appropriate occupancy-modified curve (see Figure 10). A generated random number, limited to values within the 'available' periods, is then used to determine the start time (see Figure 11). Use events are therefore more likely during periods with higher use probability relative to occupancy to accurately reflect realistic appliance-specific behaviours.
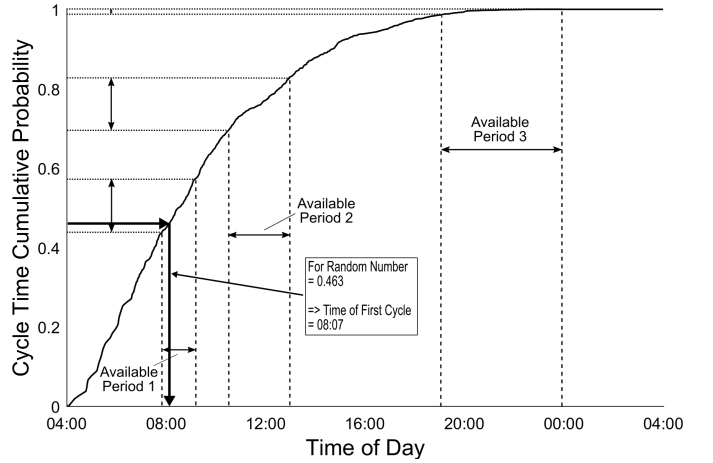

Fig. 11. Use time prediction model example.

For subsequent uses, previously identified use periods (including an arbitrary short dead period pre- and post-use to ensure adequate use separation) are made unavailable (see Table 6). The next use time is determined in the same manner using the next use time probability distribution.

Table 6
Daily event matrix example for Use #2 of 3 following a 4-minute microwave use 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
Cuml. Prob.	0.059	**0.062**	**0.066**	0.093	0.175	0.354	0.807	0.941
		↑	↑					

4.8. Other Demands

The occupant-initiated appliance demand sub-models only account for a proportion of household electricity demand. Other demands (i.e. TV-use, lighting, constant use (e.g. cold appliances), miscellaneous small appliances etc.) are modelled using a variety of techniques.

The lighting sub-model is based on the occupant location prediction approach used by Terry et al [19], and the miscellaneous appliance sub-model is similar to the Markov-Chain approach used by McLoughlin et al [20]. The constant appliance sub-model applies fixed power cycles and the TV-use sub-model is linked to specific Time-Use Survey TV activity data, both in the same manner as [7] among others. Consequently, none of these approaches is reviewed in detail.

5. Validation

5.1. Method Performance

The aim of the developed demand data calibrated, discrete-event method was to improve on appliance use sequence prediction compared to models that use per-timestep probability calculation methods and 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 [7].

5.1.1. Comparison with Per-Timestep Probability Method

The 'timestep' method analysed is similar to that used by [8], [7], and [10]. The appliance use 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 per-timestep pdfs is equal to the same baseline number of uses as the equivalent 'event' model. The methods were compared for both use number and timing prediction performance.

Use Number Prediction Performance

For the 'event' method the actual use number is identified using the binomial distribution method outlined in 4.6.1. For the 'timestep' method an equivalent degree of variance from the average number of uses to which it has been calibrated is inherent in that the method requires a sequence of independent probability calculations.

The kettle data for each of the single households in the Household Electricity Survey (HES) [1] dataset and the equivalent model results for both model types were analysed for the average difference per day between the actual and mean number of uses divided by the mean number of uses, and the average for each household determined. The model results used for the analysis were based on the average for 20 annual duration model runs to ensure a representative distribution of use behaviours were accounted for. For the HES dataset the average household value is 0.387. The equivalent for the 'timestep' models was 0.608 and for the 'event' model the value is 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 HES data and model results is a result of the lack of linked occupancy and demand data allowing 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 4.4.3) and binomial method (see 4.6.1) currently used.

Use Time Prediction Performance

Both the 'timestep' approach and the developed 'event' approach, if a single overall rather than specific use time distribution is used, are memoryless with regard to the sequence of uses 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 uses is compared to the measured data. This is determined by converting each use 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 use time standard deviation range for 6-use kettle days from all single HES dataset households is shown in Figure 12. Most households are in a range from 200 to 375 minutes with a mean of 285, which is indicative of a significant separation between uses (for example, uses 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 12 indicate that the 'event' method with multiple specific use distributions more closely approximates the distribution of the actual data, although there remains some discrepancy. Comparison of the two 'event' method basis (single and specific use timing distributions) highlights that the use-specific approach better captures the broad characteristics of the distribution and also the mean value (287 for the use-specific approach and 268 for the single distribution vs. 285 for the HES data). The per-timestep probability 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 use time varies excessively from the mean (c.14.00) as a result of unrealistic sequencing.
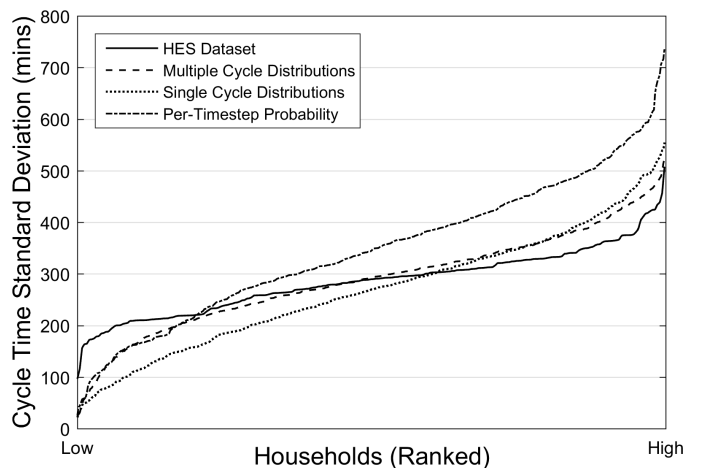

Fig. 12. Use time standard deviation range for 6-use kettle use days.

In both use number and timing comparisons, the multiple distribution 'event' approach shows a significantly better performance in capturing realistic use sequences. In this case, the residual error is the result of the timing sub-model calibra-

Table 7
Appliance model average use time cumulative probability equivalent results.

	Avg. Use CDF (Target=0.5)							
Model	Kettle	MW	Toaster	Cooker	Oven	Wash. Mach.	Dishwasher	Dryer
Time-Use Activity Calibrated Model	0.573	0.388	0.648	0.449	0.325	0.531	0.435	0.421
Demand Data Calibrated Model	0.521	0.494	0.504	0.499	0.529	0.529	0.515	0.540

tion which is based on the combined behaviours of multiple households. Further work is required to manipulate the calibration basis for individual households to account for different behaviours (e.g. typically early or later use).

5.1.2. Time-Use Activity Calibrated Model Comparison

To confirm the calibration of the use time identification method (see 4.7), the modelled use time results were converted to the associated cumulative density function (cdf) value from the relevant use time distribution. The distributions in this case are the unmodified versions prior to accounting for relative occupancy probability (see Figure 10).

The mean of the modelled cdf distribution should be close to 0.5 with a linear variation from 0 to 1. The results are also compared with the method utilised by Richardson et al [7] which was based on time-use survey activity probabilities. For the main cooking and washing appliances the mean cdf results are shown in Table 7.

The TUS linked model results show significant variation from the 0.5 target, with better performance shown for the demand data calibrated model. This confirms the conclusion from the initial analysis outlined in 2.4 that broad TUS activities are a weak predictor of specific appliance use. Residual deviation from the 0.5 target is a result of occupancy differences between the HES households and modelled equivalents, and significant variations in the number of cycles per HES household used for the calibration.

5.2. Demand Dataset Replication

Further validation undertaken has been to show that the model converges to the average electrical demand profiles and replicates the overall range of demand profiles from the HES dataset 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. Considering the level of probabilistic factoring used within the model, 500 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 HES sub-population

- To confirm that the range of model outputs is consistent with the HES dataset distribution.

- To compare individual household modelled and measured average demand profiles over a number of model runs to determine the similarity of the closest match.

5.2.1. Average Demand Profile Replication

The average demand results from the 500 runs for each of the eight identified household types were analysed to assess if the model converged to the average power value and the average time-specific demand profile for each household type from the HES dataset.

Figure 13 shows the results for all households and each of the eight defined household types for both the overall model output and for the combined results of the cyclic appliance sub-models only. In general, there is a good correlation between the model output and the HES data for both measures suggesting that the underpinning methods are effective. However, there are some discrepancies that need to be 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.

For some of the populations analysed, 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. The model 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. 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 use durations and power requirements, and likelihood of ownership of higher power appliances in larger households is, however, currently limited by data availability.

5.2.2. Demand Variance Replication

The HES-equivalent model results were also analysed to confirm if the range of household average demand values was consistent with the measured data. This is assessed by running 500 annual duration simulations and ranking the average demand results for each household. The ranking of the model run with an average demand closest to the value for the actual HES household on which it is based in then determined. The use of a ranking rather than a direct evaluation of demand allow the results for all households to be compared on the same basis. Consistency requires that the model output rankings for all households are linearly distributed between 1 and 500, which would indicate that the household type differentiation and probabilistic
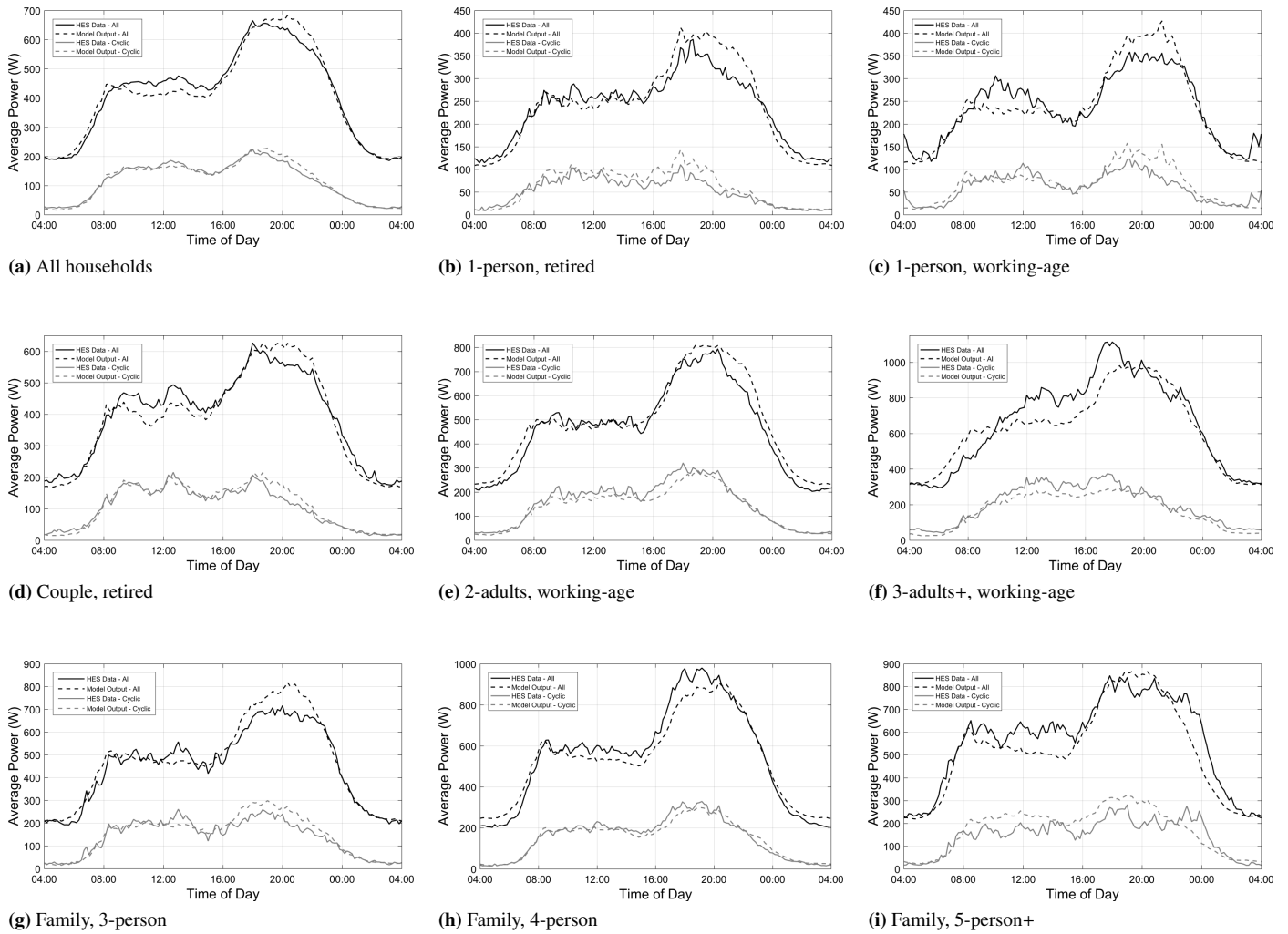
Fig. 13. Average overall and cyclic appliance electrical power per 10-minute timestep comparison between HES data and model output.

factoring used generates a realistic range of results, including outliers. The numbers of households whose measured data falls outside the model predicted results should also be small. To allow the performance of the household behaviour factor (see 4.4) to be determined, a further set of 500 runs was generated with the factor set to 1 and the same ranking distribution determined.

For overall average power, Figure 14 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 behaviours at the appliance level. 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.

However, the model with household-level factoring shows an improved performance, 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 (HBFs) generated for a single HES dataset equivalent model run is shown in Figure 15. The overall influence is to reduce the me-

dian demand value (the median HBF 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 14, where there is an improvement for the factored model in reducing both the demand 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 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 data, which do not fully capture the complex influence of income, occupancy, and randomly-driven variation. 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), appliance ownership, and individual appliance-level variation is insufficient. Further work is required to better calibrate this input.

The inclusion of the household behaviour factoring also reduces the number of HES household demands that are outwith the model output, suggesting that it is required to capture more
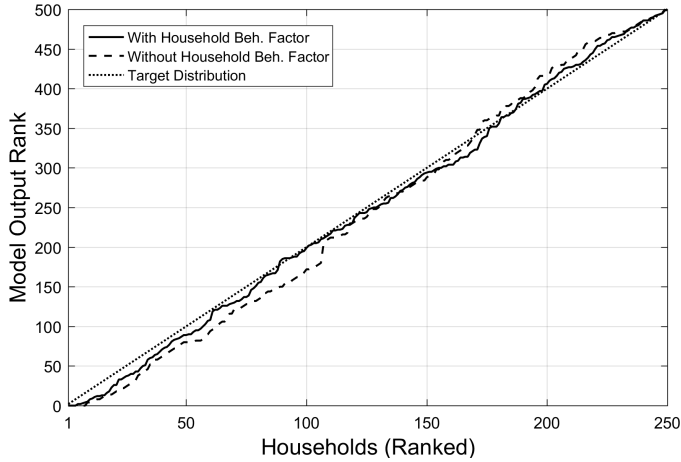
Fig. 14. Model output rank (from 500 runs) for HES household average electrical demand.
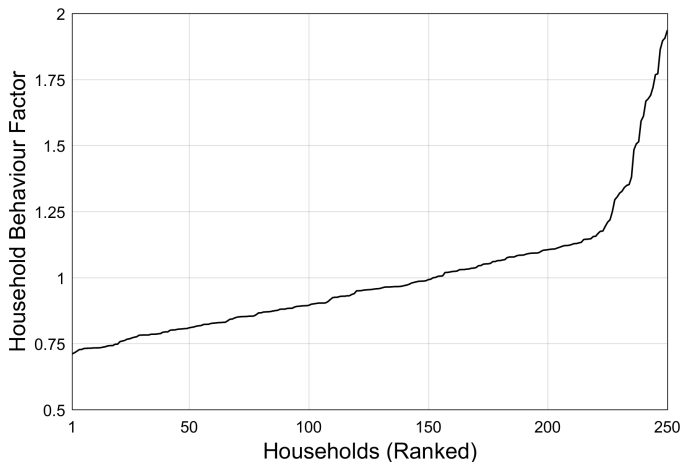

Fig. 15. Household behaviour factor (HBF) distribution for a single HES-equivalent model run.

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 also be other drivers of the weaker 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 developed model do not predict demand levels beyond those seen in the measured data and broadly capture the typical distribution driven by household characteristic and behavioural factors.

5.2.3. Individual Behaviour Replication

Having confirmed that the model captures average household type behaviours and a realistic distribution of average demands, further analysis is required to determine if the model is able to replicate household-specific per time step profiles. The following outlines a method to determine the similarity between the model output and measured data.

Similarity Analysis Method Development

The most commonly used numerical string similarity measure is the Euclidean distance, which is determined by the sum of the root-mean-squares of the difference 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 uses. At longer timescales the demand profiles tend to become smoother and more consistent. Euclidean distance analysis when used for erratic profiles generates results that do not properly reflect overall similarity and are too dependent on individual per time step differences.

A method is therefore required to reduce the time-series to a smoother profile that remains consistent to the overall profile. Piecewise Aggregate Approximation (PAA) [26] 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 was normalised and analysed based on variance to the time-series mean. For the PAA approach the time series was z-normalised (difference between actual and mean result divided by the standard deviation). The number of segments to be analysed was 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 profile while at 60-minutes too much of the detail was 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 16 shows an example conversion from a 10-minute time step daily 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 use timing

16
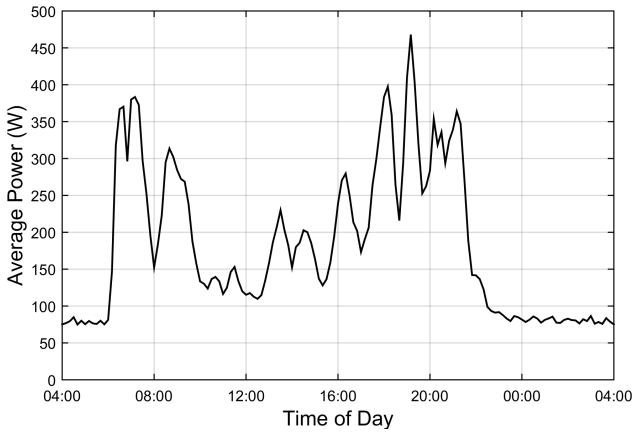
(a) Average electrical demand profile example (10-minute timestep)
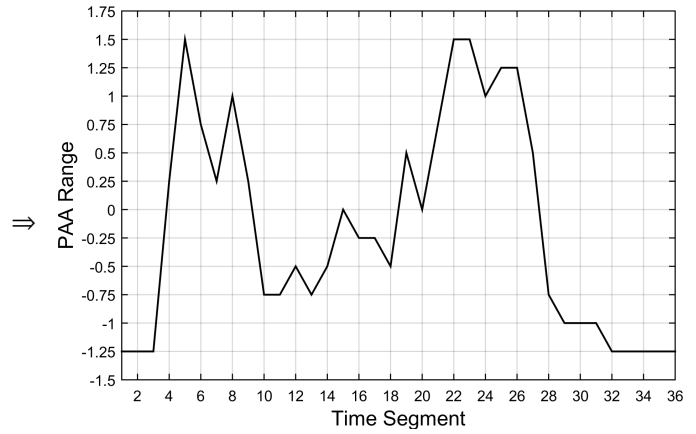
(b) PAA range 40-minute time segment conversion

Fig. 16. Monthly average profile (10-minute timesteps) to PAA range with 40-minute time segments conversion example.

model to be assessed, a separate 'timing' comparison was also undertaken. This is achieved by allowing each set of profiles to be 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.

Similarity Analysis Assessment Basis

For a 36 time-segment Euclidean Distance comparison of two time-series' 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 time-series 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 time series. A result between 2.5 and 3.5 retains broad similarity with evidence of either some time-shifted offset or failure to model 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', 'Good', 'Some', and 'Low' respectively. Example PAA value profiles for various PAA-ED scores are shown in Figure 17.

Similarity Analysis Results

Measured profiles from the HES dataset were compared with the model output for equivalent populations. The results have been grouped into the four defined similarity ranges. Detailed analysis of the data presented in Figure 18 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 more limited further improvement seen beyond 200-250 runs. The final results after 500 runs for both mean-normalised ('timing') and overall value analysis are shown in Table 8.

The 'timing' results show that 81% of the results are rated

Table 8
Model similarity results - 250 HES households - 500 dataset equivalent runs.

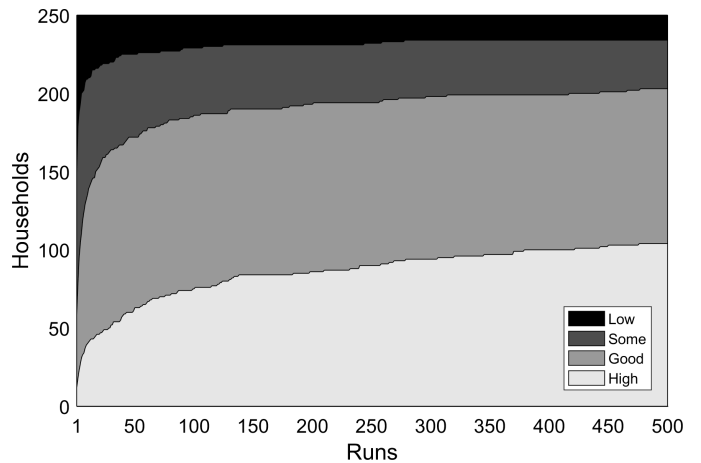| Similarity	'High'	'Good'	'Some'	'Low'
Timing	105 (42%)	98 (39%)	31 (12%)	17 (7%)
Overall	49 (20%)	118 (47%)	53 (21%)	31 (12%)

Fig. 18. Model similarity results - 250 HES households - mean-normalised analysis - 500 dataset equivalent runs.

'Good Similarity' or better after 500 runs. The results for the overall analysis are lower, as expected, but >65% are within the 'Good Similarity' or better range after 500 runs. The results suggests 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 able to be 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
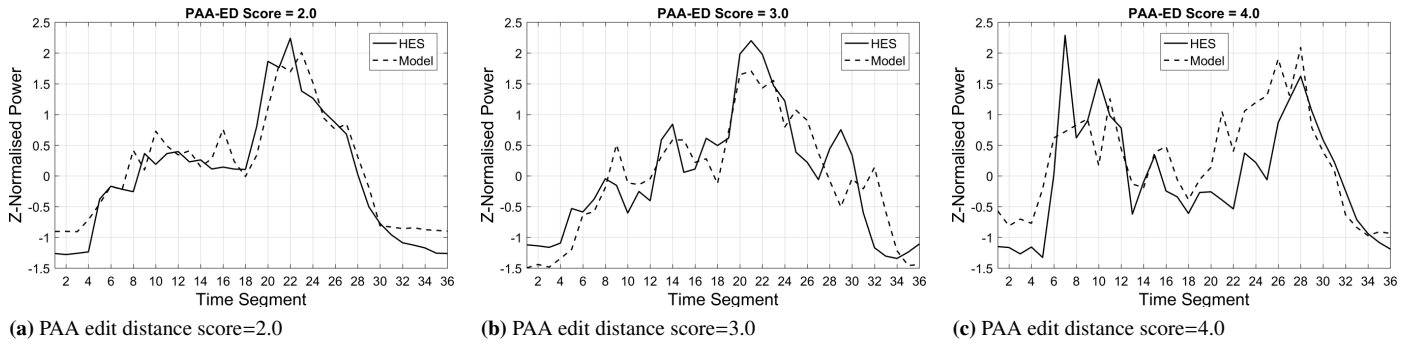
(a) PAA edit distance score=2.0 **(b)** PAA edit distance score=3.0 **(c)** PAA edit distance score=4.0

Fig. 17. Example PAA value profiles for different total PAA edit distance scores.

behaviours that are likely outwith the scope of probabilistic model calibrated from currently available data.

5.3. Residual Behaviour Averaging

The overall validation exercise shows that the model is able to capture general demand behaviours at the household-type level by incorporating a number of factors to differentiate occupancy and demand behaviours. Within each household-type group a range of individual household behaviours are captured, but some behaviours, and particularly those that deviate significantly from the average, are not.

The occupancy model basis for each defined occupant and day type and the appliance use time distributions are currently calibrated with data from a significant number of individuals and households, potentially forcing excessive behaviour averaging. To determine if this is the case, the standard deviation of the per-household mean-normalised demand per time step is compared for the HES measured and modelled data. The results are shown in Figure 19.
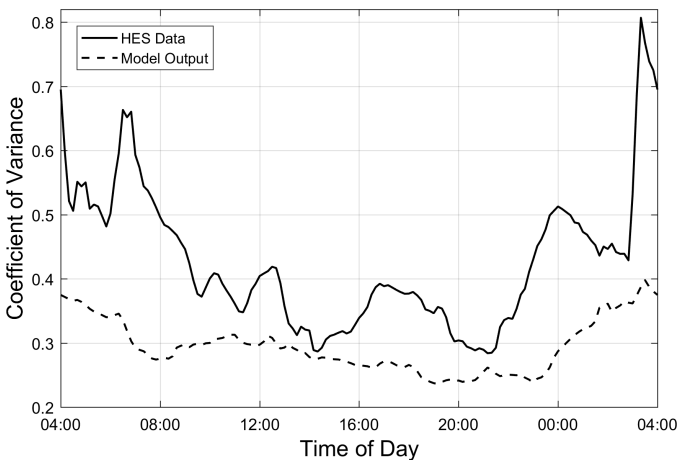

Fig. 19. Standard deviation of mean-normalised electrical demand per-timestep for measured data and modelled equivalent results for the 250-household HES population.

The results show that the model standard deviation is significantly lower than 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 individ-

ual behaviours in these time periods. This is addressed in the 'Further Work' section below.

6. Further Work

The initial verification that the developed model replicates the Household Electricity Survey (HES) [1] calibration dataset needs to be further validated with independent data. Initial analysis with two small UK datasets ([27] and [7]) shows similar performance to the HES dataset verification analysis. Further analysis with a larger dataset is required for confirmation.

The impact of composite behaviour averaging on prediction also requires further review. An improved method or calibration to capture more individual behaviour elements for both occupancy prediction and appliance use time identification would be beneficial to allow the model to be applied to individual households and smaller, distinct populations.

7. Discussion

A model has been developed that seeks to replicate the highly variable and individual nature of household electrical demand. By combining an occupancy model that incorporates occupant- and day-type differentiation and individual calendars with demand prediction that includes factors that capture individual appliance-use behaviours, the model allows distinct demand profiles for specific households types to be generated that can be used to provide improved demand estimation for homogeneous or heterogeneous communities than existing models. These have typically focused on either a limited selection of household archetypes or large composite household groupings to assess differentiated demand. Consequently, for real world assessment they do not adequately capture the breadth of household types and behaviours, and the results are not sufficiently representative to drive design decisions. The developed model has therefore focused on comprehensively capturing all potential household types and specific demand behaviours associated with each type.

The occupant-initiated appliance demand modelling method presented, directly calibrated using appliance-level demand data, aims to improve on the accuracy of existing models, particularly those calibrated using time-use survey activities. Using an event-based rather than timestep-based approach, and using separate start time probability data for each specific use,

18

reduces the likelihood of unrealistic use sequences, both intra-day and over extended periods.

Validation with both the input and independent datasets has shown that the model is able to capture a wide breadth of the individual behaviours seen. There remains a minority of households that the model does not yet capture accurately. While some households have highly unusual behaviours that are beyond probabilistic modelling, there is also evidence that the use of modelling factors that are composites of household-type behaviours results limits the degree of individual behaviours that can currently be captured. The current method is therefore primarily focused on modelling small communities, where a degree of behaviour averaging is acceptable, with further work required to determine the range of applicability and to enhance its applicability to individual households with further calibration and method development.

Acknowledgements

References

[1] J.-P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, C. Evans, Household Electricity Survey: A study of domestic electrical product usage. Final Report Issue 4., Intertek Testing & Certification Ltd.
URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/208097/10043_R66141HouseholdElectricitySurveyFinalReportissue4.pdf

[2] V. Cheng, K. Steemers, Modelling domestic energy consumption at district scale: A tool to support national and local energy policies., Environmental Modelling & Software 26 (10) (2011) 1186–1198, http://dx.doi.org/10.1016/j.envsoft.2011.04.005.

[3] Y. G. Yohanis, J. D. Mondol, A. Wright, B. Norton, Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use., Energy and Buildings 40 (6) (2008) 1053–1059, http://dx.doi.org/10.1016/j.enbuild.2007.09.001.

[4] F. Haldi, D. Robinson, The impact of occupants' behaviour on building energy demand., Journal of Building Performance Simulation 4 (4) (2011) 323–338, http://dx.doi.org/10.1080/19401493.2011.558213.

[5] Z. M. Gill, M. J. Tierney, I. M. Pegg, N. Allan, Low-energy dwellings: the contribution of behaviours to actual performance., Building Research & Information 38 (5) (2010) 491–508, http://dx.doi.org/10.1080/09613218.2010.505371.

[6] D. Aerts, Occupancy and activity modelling for building energy demand simulations, comparative feedback and residential electricity demand characterisation., Ph.D. thesis, Vrije Universiteit Brussel, http://dx.doi.org/10.13140/RG.2.1.2186.4484 (2015).

[7] I. Richardson, M. Thomson, D. Infield, C. Clifford, Domestic electricity use: A high-resolution energy demand model., Energy and Buildings 42 (10) (2010) 1878–1887, http://dx.doi.org/10.1016/j.enbuild.2010.05.023.

[8] M. Stokes, Removing barriers to embedded generation: a fine-grained load model to support low voltage network performance analysis., Ph.D. thesis, De Montfort University, http://hdl.handle.net/2086/4134 (2005).

[9] J. Widén, E. Wäckelgård, A high-resolution stochastic model of domestic activity patterns and electricity demand., Applied Energy 87 (6) (2010) 1880–1892, http://dx.doi.org/10.1016/j.apenergy.2009.11.006.

[10] U. Wilke, Probabilistic bottom-up modelling of occupancy and activities to predict electricity demand in residential buildings., Ph.D. thesis, École Polytechnique Fédérale de Lausanne, http://dx.doi.org/10.5075/epfl-thesis-5673 (2013).

[11] L. G. Swan, V. I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques., Renewable and sustainable energy reviews 13 (8) (2009) 1819–1835, http://dx.doi.org/10.1016/j.rser.2008.09.033.

[12] G. Flett, N. Kelly, An occupant-differentiated, higher-order Markov Chain method for prediction of domestic occupancy., Energy and Buildings 125 (2016) 219–230, http://dx.doi.org/10.1016/j.enbuild.2016.05.015.

[13] J. Torriti, A review of time use models of residential electricity demand., Renewable and Sustainable Energy Reviews 37 (2014) 265–272, http://dx.doi.org/10.1016/j.rser.2014.05.034.

[14] R. Yao, K. Steemers, A method of formulating energy load profile for domestic buildings in the UK., Energy and Buildings 37 (6) (2005) 663–671, http://dx.doi.org/10.1016/j.enbuild.2004.09.007.

[15] DECC, Demand Side Response in the Domestic sector - a literature review of major trials., reference: 12D/257 (2012).
URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48552/5756-demand-side-response-in-the-domestic-sector-a-lit.pdf

[16] J. Henderson, J. Hart, BREDEM 2012; a technical description of the BRE domestic energy model., version 1.1, with corrections (2015).
URL http://www.bre.co.uk/filelibrary/bredem/BREDEM-2012-specification.pdf

[17] A. Grandjean, J. Adnot, G. Binet, A review and an analysis of the residential electric load curve models., Renewable and Sustainable Energy Reviews 16 (9) (2012) 6539–6565, http://dx.doi.org/10.1016/j.rser.2012.08.013.

[18] Office of National Statistics, United Kingdom Time Use Survey, 2000., [data collection]. 3rd Edition. UK Data Service. SN:4504. (2003).

[19] N. Terry, J. Palmer, D. Godoy, S. Firth, T. Kane, A. Tillson, Further Analysis of the Household Survey: Lighting Study (Final Report), Reference: 475/09/2012 (2013).
URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/326085/Lighting_Report.pdf

[20] F. McLoughlin, A. Duffy, M. Conlon, The generation of domestic electricity load profiles through Markov Chain modelling., East-Asian Journal of Sustainable Energy Development Policy 3.

[21] R. E. Nance, A history of discrete event simulation programming languages., ACM, 1996.

[22] T. Jamasb, H. Meier, Household energy spending and income groups: Evidence from Great Britain., Tech. rep., Cambridge Working Papers in Economics 1011, Faculty of Economics, University of Cambridge (2010).
URL http://www.eprg.group.cam.ac.uk/wp-content/uploads/2014/01/JamasbMeierCombined-EPRG10031.pdf

[23] V. White, S. Roberts, I. Preston, Understanding High Use Low Income Energy Consumers., Final report to Ofgem.
URL https://www.cse.org.uk/downloads/reports-and-publications/policy/energy-justice/understanding_high_use_low_income_energy_consumers.pdf

[24] Office of National Statistics, LSOA/MSOA 2011 historic electricity consumption. (2011).
URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/418828/2011.zip

[25] Department of Work and Pensions, Households Below Average Income (HBAI) Quality and Methodology Information Report 2012/13 (2013).
URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/325492/households-below-average-income-quality-methodology-2012-2013.pdf

[26] E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, Dimensionality reduction for fast similarity search in large time series databases., Knowledge and information Systems 3 (3) (2001) 263–286, http://dx.doi.org/10.1007/PL00011669.

[27] D. Murray, J. Liao, L. Stankovic, V. Stankovic, R. Hauxwell-Baldwin, C. Wilson, M. Coleman, T. Kane, 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.