TOWARDS DETAILED OCCUPANCY AND DEMAND MODELLING OF LOW-CARBON COMMUNITIES

Graeme Flett, and Nicolas Kelly
Energy Systems Research Unit, University of Strathclyde, Glasgow, UK

ABSTRACT

The integration of low carbon energy systems with zero carbon housing and communities requires a detailed understanding of the scale and timing of energy demand. Household demand is closely related to occupant type and associated occupancy patterns. A high resolution occupancy modelling technique is proposed as a foundation for future demand modelling. Probability data is compiled for multiple occupant, household, and day types from Time-Use Survey data. A higher order Markov approach is then used to generate representative occupancy profiles over extended time periods. An improved method to model family interactions has also been developed.

INTRODUCTION

Background

Technical and commercial analysis of community scale zero-carbon energy projects requires a detailed understanding of the likely demand profile at both short and long time-scales. Intra-day demand diversity estimation is required for sizing of localised energy supply systems, assessing demand management potential, and the scope for grid import/export.

Demand prediction is of particular importance for low and zero carbon projects where generation may either be seasonal, intermittent or benefit from stable demand (Abu-Sharkh et al, 2006). Matching of supply and demand and adequate storage sizing are critical for ensuring that such projects perform as anticipated. The UK Government has identified a lack of energy demand data as a key barrier to growth in low-carbon community energy and demand management projects (DECC, 2014).

Previous research work (Yohanis (2008), Haldi and Robinson (2011), Kelly et al (2012)) has shown that a variety of factors can influence household energy demand. These include, but are not limited to, house size, household size, bedroom number, occupant age, income, number of children, and tenure; there is limited existing research that attempts to quantify the impact of each, or that considers how different combinations of households may influence the overall demand, and therefore the carbon emissions, of a wider community.

Most existing modelling methods (e.g. Richardson et al (2010)) assume average households or combine large, disparate populations to calibrate occupancy and demand models. They therefore generate demand profiles that are an unrepresentative composite of the combined behaviours. Where an energy scheme encompasses a large area, these may be appropriate. However, communities are often more homogeneous, with similar housing types, tenure and broad economic characteristics; examples include social housing or commuter estates. There is a need to evaluate how specific community characteristics impact demand, and to assess how the community size influences the potential for demand variation.

Critical Factors in Domestic Energy Demand

Detailed and differentiated energy modelling at the household level requires an understanding of a) how different household factors influence demand, b) how the factors are dispersed across the population of interest, and c) how this knowledge can then be combined to predict demand for all household types.

As identified by Capasso (1994), Yao and Steemers (2005), and Torriti (2012), amongst others, occupancy is a key determinant of overall energy demand, and particularly in its temporal characteristics. Limited work, however, is available that analyses occupancy data per occupant, household and over different day types (e.g. weekday, weekend) to allow the occupancy-driven impact on household energy demand to be determined.

For zero-carbon housing, electrical and hot water demand will predominate as heating demand falls as a result of improved building thermal performance. The residual heating load may also be more closely linked to actual occupancy as pre-heat times reduce and demand becomes more intermittent in highly insulated and passively heated houses. Understanding potential occupancy variations will therefore be required to predict zero-carbon housing performance under realistic demand conditions.

Yohanis et al (2008) determined that house type, floor area, bedrooms, income, age, location, and occupants all influenced electrical demand. In particular, income and employment status was important for the level and timing of demand. Kelly et al (2012) found similar results for heating demand.

There are also less well-defined behavioural factors to energy use. Gill et al (2010) identified that there was a 51%, 37% and 11% influence on heat, electrical and hot water use respectively based solely on surveyed energy efficiency behaviour. Fell and King (2012) found that household characteristics alone could not explain gas demand variations and that there must be other underlying behaviours.

At the household level, many conflicting factors are involved. There is a need, therefore, to determine and model the relative influence of occupancy, household characteristics, and individual behaviours.

Existing Occupancy Modelling Methods

Energy demand models are broadly characterised into two types, top-down and bottom-up. Top-down demand models, such as that developed by Kelly et al (2012), use regression analysis of large populations to determine the influence of individual factors. Bottom-up models use highly detailed models for each individual demand that are then combined to provide an aggregate total demand.

Grandjean et al (2012) provides a comprehensive review of existing demand models, primarily for electrical loads, concluding that bottom-up models based on Markov-chain occupancy probability techniques represent the best current method. Capasso (1994), Widen and Wackelgard (2009), and Richardson et al (2010), and have developed such models.

The models developed by Widen and Wackelgard, and Richardson provide an adaptable method, utilising widely available and globally consistent Time-Use Survey (TUS) data. Markov Chain probability profiles are generated for occupancy changes based on household size, time of day, and day type (i.e. weekday/weekend). These allow the occupancy status at a time *t* to be determined based on the probability of a status change at time *t-1*.

Both models utilise first-order Markov probabilities that have no ‘memory’, with only the status at the previous time-step considered regardless of the status duration or sequence. The result, particularly when using large populations with significant occupancy variations, is highly random models with poor duration prediction (Wilke, 2013).

An alternative non-Markov method has been developed by Wilke (2013), which uses the same TUS data to generate forward prediction of status duration and subsequent change of status. Aerts (2014) has extended this to incorporate different broad occupancy profiles identified by clustering analysis of the TUS data. This method improves duration prediction and reduces computation load, as there is recalculation per event rather than timestep.

Neither the Richardson nor Widen and Wackelgard models make any differentiation in household type beyond number of occupants. Further work using the same basic first-order Markov technique by Muratori (2013) has split the households into four archetypes (working/non-working, male/female). However, the influence of this split is not analysed in detail.

For the non-Markov approach, Wilke (2013) reviewed a variety of sub-population models with wider populations and found using sub-populations to be significantly more accurate. Aerts (2014) uses identified occupancy characteristics directly rather than populations to infer them. Extended diaries to determine how each characteristic day is distributed for each modelled occupant type would be required to develop this method further.

AIM

Existing occupancy prediction methods based on large, undifferentiated populations have been seen to result in unrepresentative individual occupancy profiles. To address this, the following sections describe – identification of the key household characteristics influencing occupancy; and – the development of a customised Markov model to generate representative occupancy data for each individual household in a community. The generated occupancy model will be used as an input for a future community-scale differentiated demand model.

METHOD

The following section outlines how the differentiated, customised Markov occupancy model was developed. This includes a description of, a) how the critical factors used to differentiate occupants and households with divergent occupancy patterns were identified; b) the structure of the higher-order Markov model; c) the method used to improve co-habiting couple and child models; d) secondary modelling techniques used to identify television use from the dataset; and e) how realistic working weeks were allocated within the model.

The UK Time-Use Survey (TUS) dataset compiled in 2000/2001 was used for the model development. The data set comprises around 10,000 single-day weekday and weekend diaries with a 10-minute resolution. The data includes basic location and more detailed activity information for each respondent.

While some of the specific activities within the survey may now be out-of-date, the basic occupancy data is assumed to remain broadly representative. Any new survey should be based on the same standard process and the model can be simply converted to any new dataset when available.

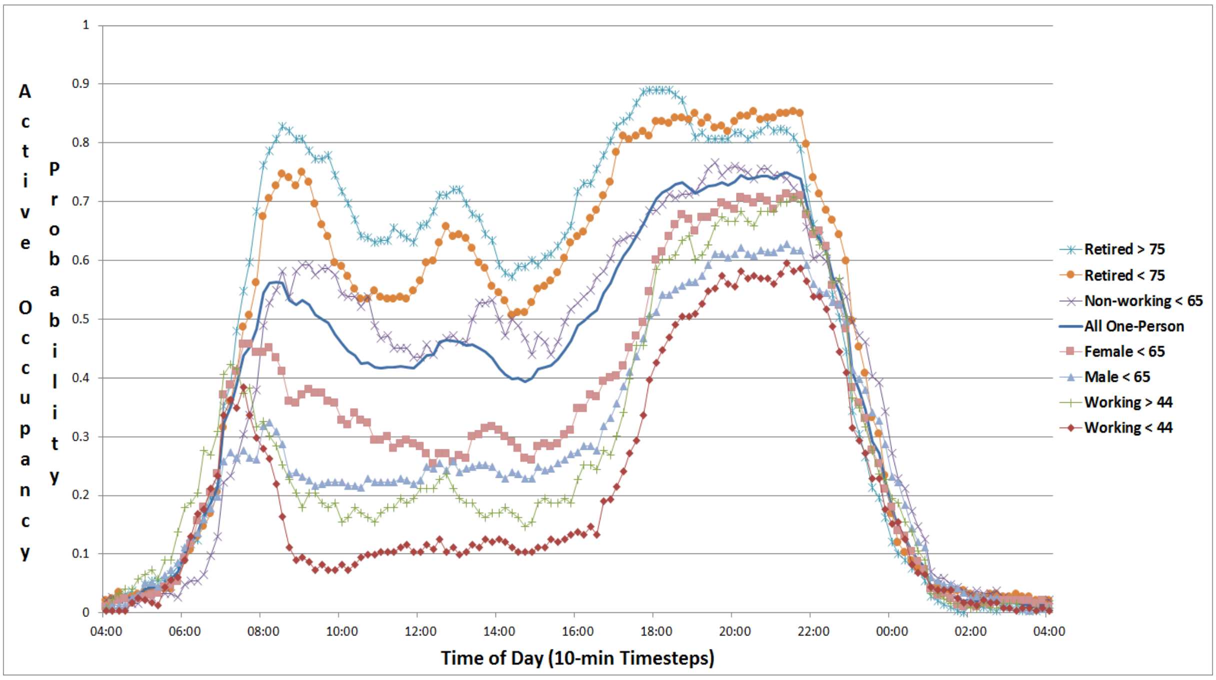

Figure 1: One-person Household Population Weekday Active Occupancy Probability

Identifying Critical Household Types and Characteristics

The UK TUS dataset was analysed in order to extract time-dependent occupancy probabilities based on key household characteristics; this involved splitting the dataset according perceived critical parameters to ascertain their influence on occupancy.

Analysis of the TUS dataset identified seven basic household types (single, retired single, couple, retired couple, family, single parent, mixed adult). These are used as the primary filter for generating household-differentiated occupancy models.

Figure 1 shows averaged weekday time-dependent occupancy probabilities for a range of one-person household types from the TUS dataset; their extreme variation highlights that other factors, in addition to household type or occupant number, are required to model occupancy accurately.

Age, gender, and employment status of occupants were also analysed to determine their influence on occupancy. Further, weekday, Saturday and Sunday datasets were considered separately.

For all day types, employment status had the greatest influence on occupancy. Occupant age also was an influence and so the TUS dataset was further split into overlapping age ranges for efficient use of the available data. For example, for working one-person households, the 18-37 TUS population was used for the 18-33 model, the 28-44 TUS population was used for the 34-40 model etc.

At the outset it was not known how many diaries would be required as a minimum to ensure a sufficient coverage of occupancy probability data to allow any model to remain stable. Splitting weekday diaries by household type, employment, and logical age ranges generated populations of typically between 100 and 200 diaries. This population size was used for initial analysis of model stability.

The overall conclusion from the TUS analysis was that the data should be split into household type, employment status, and age ranges. Each sub-dataset was then used separately to generate population-specific Markov occupancy probability matrices.

Higher Order Markov Occupancy Modelling

As stated previously, the existing first-order Markov models (e.g. Widen and Wackelgard (2009), Richardson et al (2010)) are 'memoryless'. The status at the next timestep is only dependent on the previous timestep status, duration is not considered.

This lack of 'memory' results in poor replication of the profile shape and status duration distribution of the input data. Generated occupancy profiles exhibit erratic behaviour patterns, particularly with regard to sleep and daytime absence periods where actual status changes are strongly linked to duration. This also prevents this technique from being used for extended period models that consistently replicate typical profiles over a number of days.

Two potential higher-order methods were identified that consider status duration. One uses 'event-based' modelling, similar to that developed by Wilke (2013) and Aerts et al (2014). This does not utilise a Markov process but assigns a next status and fixed duration at each change of status based on probabilities generated from the TUS data.

While this method was shown to improve the distribution of identified durations, to reduce computational time by a factor of 5, and to be useful

$$
\begin{array}{c} \\ S \end{array}
\begin{array}{ccc} S & A & O \\ (p_{SS} & p_{SA} & p_{SO}) \end{array}
\quad \Longrightarrow \quad
\begin{array}{c} \\ 0 < S < 2 \\ 2 < S < 4 \\ 4 < S < 6 \\ 6 < S < 8 \\ > 8 \end{array}
\begin{array}{c} \begin{array}{ccc} S & A & O \end{array} \\ \begin{pmatrix} p_{S2S} & p_{S2A} & p_{S2O} \\ p_{S4S} & p_{S4S} & p_{S4O} \\ p_{S6S} & p_{S6A} & p_{S6O} \\ p_{S8S} & p_{S8A} & p_{S8O} \\ p_{S8+S} & p_{S8+A} & p_{S8+O} \end{pmatrix} \end{array}
$$

Figure 2: Transition from First-Order to Higher-Order Markov Model for Current Sleep Status

for short duration modelling, it was not stable over extended periods. This method does not have the self-correcting nature of a Markov model, and several consecutive outlier durations can force the model into occupancy patterns that are unrepresentative, with limited associated probability data. It is also more difficult to assess durations that overlap different modelled day types.

An alternative method is developed here which maintains the Markov approach but uses the status duration to give a higher-order basis. This captures specific patterns (such as sleep and workday absence) more accurately. To differentiate between sleep patterns and daytime absence, separate ‘Sleep’ and ‘Out’ statuses are used for inactive periods.

Markov probabilities for each status, derived from the TUS dataset, were split into ranges of durations (e.g. < 2 hour, 2-4 hours etc.) (see Figure 2). The optimum ranges differ based on status, occupant type and day type (e.g. longer absent durations are more critical for assigned working days). Different ranges are used for each occupant model based on analysis of the relevant TUS dataset status durations.

In Figure 2, '*p*' is the probability of a particular transition, '*S*' refers to the sleeping, '*A*' is awake and active and '*O*' is out of the dwelling. The numbers refer to the range of durations included in hours.

Matlab has been used to develop models for both first and higher order Markov approaches to allow comparative analysis of both methods. The results of this comparison are presented later in the paper.

The model generates occupancy profiles based on the same 10-minute resolution as the input TUS data. The primary model differentiates between ‘sleep’, ‘active’, and ‘out’ conditions. A secondary model, described below, further splits 'active' periods into a ‘general’ and ‘television-viewing’ status.

Interaction of Couples and Children

Initial modelling work assumed that co-habiting couples could be treated as independent adults with a small error from underestimating combined behaviour. However, subsequent analysis indicated that the estimation error for single and double occupancy was greater than expected and would lead to an overestimation of total occupied time for a couple (See Figure 3).

A method was therefore developed to characterise the occupancy of couples. There are insufficient relevant diaries to allow each adult in a couple to be identified separately. However, if we assume that tracking each specific adult is not critical, the couple could be modelled as a single entity.

Using the same TUS analysis method as for individual occupants, each couple was analysed as having a single status based on the unassigned combination of individual states (i.e. Sleep/Sleep, Sleep/Active, Sleep/Out etc.). This reduces the number of potential status combinations from 9 to 6, reducing the data required for stable models.

Figure 3 shows that the joint model results are significantly closer to the TUS dataset than combining individual adult models. This method was also applied separately to parents in family households.

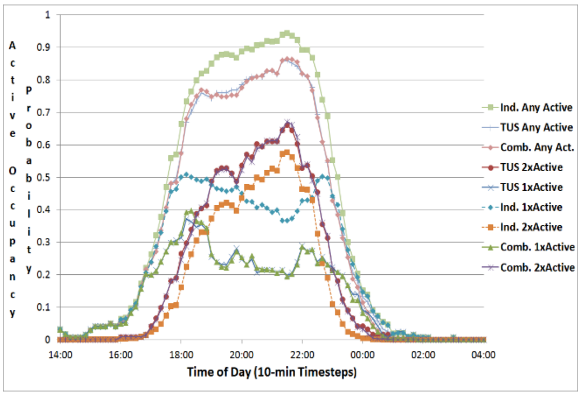

Figure 3: Impact of Combined Couple Model on Individual and Overall Occupancy

By distinguishing between family and non-family households for the co-habiting couple models, it was assumed that the influence of children was adequately captured. A simpler Markov probability method can therefore be utilised for children by linking child occupancy to adult occupancy (i.e. if adult is active/inactive is child active/inactive etc.) to limit the input data requirement.

Each child was modelled separately as there are insufficient diaries to determine occupancy interaction between siblings of different ages. Child occupancy was split by age range (e.g. 8-9, 10-11 etc.), and between school term and holiday periods.

There are insufficient school holiday TUS diaries to generate differentiated parent models; however there is no difference in average occupancy levels for parents in the TUS dataset (31.0% for term periods and 30.7% for non-term periods).

A separate independent model is used for 'adult' children (16-24) living in the parental home. It is assumed that their occupancy is independent of the other household members.

Secondary Activity Modelling

The majority of specific TUS diary activities (e.g. cooking, cleaning) do not explicitly identify an activity leading to energy demand. However, the TUS ‘Television’ activity can be directly linked to appliance use. Therefore, for each adult and child model a secondary Markov probability model is used to distinguish between 'General' occupancy and television use for each 'Active' period.

Economic Activity / Working Week Models

The TUS dataset also includes a one-week work/education duration diary. These show a wide variation in work profiles across the population, with fewer than 50% conforming to the 'typical' Monday to Friday daytime working profile.

The developed 'working day' models outlined in this paper only include people that were working over 5 hours on their diary day. For ‘non-working’ days, all people with less than 1.5 hours working time are combined, including those defined as ‘employed’. Active occupancy is slightly lower for the employed on ‘rest’ days but the difference is small and there is insufficient data to model separately.

5 hours working time was selected to define ‘full-time’ work as it was considered a sufficient minimum to establish a distinct pattern and allowed for distinct populations of sufficient size (>100 diaries) to be selected from the TUS dataset. Less than 1.5 hours working time was considered to be indistinguishable from other ‘out’ activities. TUS data for part-time (working hours between 1.5 and 5 hours) and night workers are currently not included but will be integrated at a later stage.

The occupancy model generates a working week based on the TUS work diaries and assigns an appropriate day type per day based on the generated calendar. This ensures that a realistic distribution of working and non-working days is identified for a community.

The different day types are as follows: working weekday, non-working weekday, working Saturday, non-working Saturday, working Sunday, non-working Sunday, and public holiday. The occupancy model can be used for both single-day, and extended time period analysis.

VALIDATION

Occupancy Model Validation Metrics

Time-Use Survey (TUS) data is limited to single day diaries. There are no datasets that track household occupancy over longer periods for a sufficient number of households. Final validation of the occupancy model will, therefore, first require integration with an energy demand model and then comparison with long term demand data. Consequently, the initial validation described here is restricted to confirmation that the model output is comparable with the TUS input data.

Three metrics are used to compare and validate the occupancy models.

Metric (1) - The first is a basic analysis that the model generates the same per-timestep average active occupancy probability as the TUS dataset over an annual profile when the results of multiple runs (typically 100) are aggregated. The difference between the TUS data average occupancy and modelled probabilities is added for each of the 144 10-minute timesteps and compared for each method used.

$$Err1 = \sum_{t=1}^{144} \left| \left(\overline{P}_{ACTIVE}^{MOD}(t) - \overline{P}_{ACTIVE}^{TUS}(t) \right) \right| \quad (1)$$

Metric (2) - An analysis of the cumulative proportion of status durations is used to determine if the generated model accurately replicates the range of durations from the input data up to a maximum of 24 hours (144 10-minute timesteps). The 'error' is the sum of the absolute difference between the model and TUS data cumulative proportion at each duration value. The analysis is typically an average of 100 separate model runs of an annual duration.

$$Err2 = \sum_{d=1}^{144} \left| \left(\sum_{d=1}^{d} P_{DUR}^{MOD}(d) - \sum_{d=1}^{d} P_{DUR}^{TUS}(d) \right) \right| \quad (2)$$

Metric (3) - The third metric uses the Levenshtein Edit Distance Method for character string analysis to compare generated single day status profiles (144 timesteps) against the closest match from the input TUS dataset. This is used to confirm that modelled profiles closely match actual profiles. The method is also used to confirm that the model remains stable over extended periods by analysing the metric for each modelled day of a multi-day run. The Levenshtein method used assigns a ‘cost’ of 1 for each edit (insertions, deletions, and replacements), the maximum error is therefore 144. For clarity the error is converted from a 10-minute/144 timestep basis to an error expressed in hours.

RESULTS

Firstly the impact of using smaller differentiated individual occupant first-order Markov models is examined. Secondly, the additional benefit when the higher-order Markov model is used with the differentiated models is considered. Finally, a separate analysis of the combined co-habiting couple model is also included.

The results presented are based on single day-type models (e.g. all working days). Validation using multiple day-type models based on assigned occupant calendars remains to be undertaken.

First-Order Occupant Type Model Analysis

To analyse the potential benefit of using smaller, differentiated individual occupant populations with similar occupancy characteristics, an average 1-person household occupant model (representative of the models developed by Widen and Wackelgard (2009) and Richardson et al (2010)) is compared with two smaller 1-person household groups from the TUS dataset (‘Working 18-37’, ‘Over 76’) using the first-order Markov model for a typical weekday.

<u>*Average Occupancy Probability Analysis*</u>

The results for the average occupancy probability analysis (metric (1)) are shown in Table 1. The first column identifies the first-order model populations used for the analysis and the second column identifies the TUS data population used for comparison with the model results. The results are the calculated 'Err1' values.

Table 1
Avg. Occupancy Probability Analysis (Err1) for 1st-Order Model and TUS Population combinations

MODEL	TUS	ERR1
All 1-person	Working 18-37	32.44
All 1-person	Over 76	19.56
Working 18-37	Working 18-37	3.33
Over 76	Over 76	3.40

As the results in Table 1 demonstrate, there is a significant improvement when smaller differentiated populations are used for the model.

<u>*Status Duration Analysis*</u>

The results for the status duration analysis (metric (2)) are shown in Table 2. The first column identifies the population used for the first-order model analysis and the second column identifies the TUS data population used for comparison with the model results. The results are the calculated 'Err2' values.

Table 2
Status Duration Analysis (Err2) for 1st-Order Model and TUS Population combinations

MODEL	TUS	SLEEP	ACTIVE	OUT
All 1-person	Working 18-37	8.20	9.72	22.43
Working 18-37	Working 18-37	2.26	0.79	2.87
All 1-person	Over 76	4.13	7.11	8.93
Over 76	Over 76	1.64	1.63	1.10

The results show that the 1-person model generated from the whole 1-person population fails to properly replicate the range of durations expected for the two specific 1-person populations identified. Significant improvements are shown where both the model and comparison data are extracted from smaller, more representative populations.

<u>*Single Occupancy Profile Analysis*</u>

Edit distance analysis (metric (3)) was used to compare the overall 1-person first-order model with the specific ‘Working 18-37’ and ‘Over 76’ first-order models. The ability to replicate the ‘Working 18-37’ and ‘Over 76’ TUS datasets was assessed.

The average lowest edit distance for the overall 1-person model to the closest ‘Working 18-37’ TUS dataset match for 260 modelled working weekdays was 5.35 hours. This compares to 1.88 hours for the first-order model based on the ‘Working 18-37’ population. The equivalent improvement for the ‘Over 76’ population was from 3.71 to 3.51 hours.

The overall conclusion is that there is an improvement in the first-order models ability to replicate observed behaviour using smaller, more representative populations. The degree of improvement would seem to depend on the observed behaviour of the differentiated population. Further analysis is required to determine why the improvement is significantly greater for the ‘Working 18-37’ population, and if this variations is seen for other populations.

Higher Order Model Validation

<u>*Average Occupancy Probability Analysis*</u>

In comparison with the first-order Markov model, the higher-order Markov model reduces metric (1) from 3.33 to 2.07 for the ‘Working 18-37’ population, and from 3.40 to 2.66 for the ‘Over 76’ population. The higher-order Markov method therefore replicates the TUS input data average occupancy probabilities more accurately.

<u>*Status Duration Analysis*</u>

Using metric (2); the distribution of status durations was compared for the first-order (FO) and higher-order (HO) models as shown in Table 3.

Table 3
Status Duration Analysis (Err2) for First and Higher Order Markov Models

MODEL	SLEEP	ACTIVE	OUT
Working 18-37 FO	2.26	0.79	2.87
Working 18-37 HO	1.42	0.47	1.43
Over 76 First-Order	1.64	1.63	1.10
Over 76 Higher-Order	1.26	1.15	1.03

This shows that the higher order model is significantly better at replicating durations from the input data, particularly for the working population.

Similar results were achieved for other populations. However, further work is required to confirm the impact of changing the population sizes and duration range splits on the accuracy.

Single-Day Occupancy Profile Analysis

For the ‘Working 18-37’ higher-order model, the average edit distance expressed in hours to the closest TUS dataset match for 260 modelled working weekdays was 1.53 hours, which compares to 1.88 hours for the first order model. For the ‘Over 76’ population the equivalent improvement was 3.51 to 2.22 hours. This represents a significant improvement in the higher-order model’s ability to mimic actual profiles from the input data, particularly for the ‘Over 76’ population that shows a relatively small improvement when the differentiated population is used with the first-order method.

Further analysis of the results showed that the number of generated profiles that have a lowest edit distance match of 2 hours or under increases from 66.5% to 79.2% for the ‘Working 18-37’ higher-order model, and from 7.3% to 50.4% for the ‘Over 76’ higher-order model.

Over the continuous model period of 260 days, the lowest edit distance profile remains consistent showing that the higher-order method remains stable over the extended run and is not impacted by smaller calibration datasets.

Couple Model Analysis

Average occupancy analysis (see metric (1)) is more difficult for 2-person models. However, Figure 3 highlights a significant improvement for the combined couple model compared to two individual models in tracking the TUS data occupancy patterns.

Table 4
Status Duration Analysis (Err2) for First and Higher Order Markov ‘Working Couple 28-50’ Models

MODEL	S-S	S-A	S-O	A-A	A-O	O-O
2xInd. FO	3.53	1.33	0.85	1.54	1.44	2.97
2xInd. HO	2.59	1.45	0.75	1.84	0.88	2.12
Comb. FO	0.99	0.37	0.84	0.65	0.30	1.67
Comb. HO	0.97	0.29	0.88	0.50	0.30	1.36

Status duration analysis (see metric (2)) for 100 260-day ‘Working Couple 28-50’ model runs (see Table 4) shows a significant improvement using the combined model approach, and a more limited additional benefit from using the higher-order Markov approach on this particular metric.

Table 5
Occupancy Profile Analysis for First and Higher Order ‘Working Couple 28-50’ Models

MODEL	AVG. LOWEST EDIT DIST. (HRS)
2xInd. FO	3.88
2xInd. HO	3.38
Comb. FO	3.28
Comb. HO	2.89

Results for single-day occupancy profile analysis of the ‘Working Couple 28-50’ model (see metric (3)) in Table 5 highlight a more significant benefit when using the higher-order approach.

Considering all metrics, the combined, higher-order model basis provides an improved approach for co-habiting households.

DISCUSSION

Several enhancements to existing high resolution, Markov-based occupancy models have been considered, with three primary improvements implemented and analysed. The first was to split the occupant models based on occupant and household characteristics to generate more representative occupancy profiles that reflect different lifestyles. The second improved the model consistency and status duration prediction by using a higher-order Markov process. The third was to differentiate between single and family households.

With significant variations in occupancy based on household type, age and employment shown in the TUS data, differentiating the occupancy models using these factors produces more accurate occupancy profiles. However, the degree of improvement varies for different sub-groups.

The developed higher-order modelling method shows a measureable additional improvement in output accuracy based on analysis for single person and cohabiting couple households. In particular, this method improves the generated profile accuracy for groups not significantly improved by using smaller populations with a first-order model. It is therefore necessary to combine both smaller populations with the higher-order method for maximum benefit.

The differentiation between related couples and individuals allows the occupancy interactions seen in the input data for co-habiting relationships to be captured. In particular, this reduces the overestimation of the total active occupancy period predicted when couples are modelled individually.

The occupancy model improvements outlined offer a means to improve occupancy prediction at the household level and for communities where individual household characteristics are significant and the composition deviates from the average.

Initial analysis suggests that the higher-order models remain stable with the current dataset size of between 100 and 200 diaries, perhaps due to the consistency within each sub-population. This method is therefore more stable for multi-day modelling in comparison with higher-order event-based approaches. Further work is required to confirm that no further consolidation is necessary, or whether increasing sub-population sizes but reducing higher-order model duration ranges is a better approach.

The ability to produce stable and representative occupancy profiles based on realistic distributions of day types over extended periods will allow this model to be integrated in the future with a demand cycle model that generates cycles that are consistent with occupancy over extended periods. This improves on existing models that typically use fixed per-timestep cycle probabilities.

CONCLUSION

In future zero carbon buildings, occupancy related appliance and hot water demand will constitute a significant portion of energy use. The work reported in this paper develops an improved method for household occupancy modelling. Existing Markov probability methods use status probabilities models derived from large, mixed populations. This approach does not adequately capture common occupant characteristics seen in reality.

Using various statistical techniques to analyse the results, it was determined that using smaller, differentiated populations produces more representative profiles. Further improvement was measureable when a higher-order Markov approach, based on ranges of current duration, was used.

The model was also shown to be stable over runs of 1 year in duration. This allowed individual occupant profiles to be generated based on realistic sequences of working and non-working days, which can then be combined to assess variations in overall occupancy at the household and the community level.

Further validation of the occupancy model, and analysis of the link between occupancy and demand, will be possible once this model is integrated with a demand model and compared with demand data.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support received for this work from the BRE Trust.

REFERENCES

Abu-Sharkh, S., Arnold, R., Kohler, J., Li, R., Markvart, T., Ross, J., Steemers, K., Wilson, P., and Yao, R. (2006). “Can microgrids make a major contribution to UK energy supply?” Renewable and Sustainable Energy Reviews **10**: 78-127.

Aerts, D., Minnen, J., Glorieux, I., Wouters, I., and Descamps, F. (2014). “A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison.” Building and Environment **75**: 67-78

DECC. (2014). Community Energy Strategy: Full Report. A report by the Department for Energy and Climate Change. January 2014.

Fell D., King G. (2012). “Domestic energy use study: to understand why comparable households use different amounts of energy.” A report by Brook Lyndhurst for the Department for Energy and Climate Change.

Gill, Z., Tierney, M., Pegg, I., and Allen, N. (2010). “Low-energy dwellings: the contribution of behaviours to actual performance.” Building Research & Information **38**(5): 491–508.

Grandjean, A., Adnot, G., and Binet, G. (2012). “A review and analysis of the residential electric load curve models.” Renewable and Sustainable Energy Reviews **16**: 6539-6565.

Haldi, F., Robinson, D. (2011). “The impact of occupants' behaviour on building energy demand.” Journal of Building Performance Simulation, **4**(4): 323-338.

Kelly, S., Shipworth, D., Gentry, M., Wright, A., Pollitt, M., Crawford-Brown, D., and Lomas, K. (2012). “A panel model for predicting the diversity of internal temperatures from English dwellings.” Tyndall Centre for Climate Change Research. Working Paper 154. July 2012.

Muratori, M., Roberts, M., Sioshansi, R., Marano,V., and Rizzoni, R. (2013). “A highly resolved modeling technique to simulate residential power demand.” Applied Energy **107**: 465–473

Richardson, I., Thomson, M., Infield, D., and Clifford, C. (2010). “Domestic electricity use: A high-resolution energy demand model.” Energy and Buildings **42:** 1878-1887.

Torriti, J. (2012). “Demand Side Management for the European Supergrid: Occupancy variances of European single-person households.” Energy Policy **44**: 199–206

Widen, J., Wackelgard, E. (2009). “A high-resolution stochastic model of domestic activity patterns and electricity demand.” Applied Energy **87:** 1880-1892.

Wilke, U. (2013). “Probabilistic Bottom-up Modelling of Occupancy and Activities to Predict Electricity Demand in Residential Buildings.” PhD thesis, École Polytechnique Federale De Lausanne. February 2013.

Yao. R., Steemers, K. (2005). “A method of formulating energy load profile for domestic buildings in the UK.” Energy and Buildings **37**: 663–671

Yohanis, Y., Mondol, J., Wright, A., and Norton, B. (2008). “Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use.” Energy and Buildings **40:** 1053-1059.