OCCUPANCY BASED THERMAL ENERGY MODELLING IN THE URBAN RESIDENTIAL SECTOR

TIM TRÖNDLE & RUCHI CHOUDHARY
University of Cambridge, UK

ABSTRACT

Energy use in the urban residential sector corresponds directly to operation of energy consuming devices. For instance, space-heating is a substantial element of residential energy consumption in the UK, but it is as much tied to occupant preferences and timing of their presence in a home, as it is to the physical characteristics of the dwelling. Timing, duration, and in-use efficiency of residential energy consumption are essential for increasing the utilization of district energy systems and demand-side management. The ongoing IEA-EBC Annex 66 and our own recent work attests that there is a need to develop new methods to represent occupant presence and energy consuming activities in energy demand modelling. This is as much a multi-scale simulation issue, as it is a computational challenge. This paper presents a proof-of-concept methodology of estimating thermal energy demand on the urban scale by introducing occupancy models to high resolution bottom-up energy models. A synthetic population is created from census data and the occupancy of every citizen is modelled using a time heterogeneous Markov chain which is calibrated using time use survey data. The methodology is applied to a case study where the thermal energy demand is found to be varying up to 50% in different locations at certain times of the week. Regions with less diverse energy demand and thermal power patterns can be identified and discriminated against those with more diversity in the demand.
Keywords: residential building occupancy, building energy model, agent-based model, population synthesis, micro-simulation.

1 INTRODUCTION

Accounting for 34% of global energy end-use, the building sector is the largest energy sink and a major contributor to global CO_2 emissions [1]. Three quarters of this amount are accountable for space heating and cooling purposes. When trying to reduce this energy impact, understanding the energy demand originating in the building sector and its drivers is crucial. With more than half of the global population living in cities and with on-going urbanisation [2], urban built environments are becoming more important in this regard. This is even more so true when looking at America or Europe where the urban population exceeds 70% of the total population already today [2]. Estimations of energy demand for space heating are valuable during the design phase of urban built environment and its energy supply infrastructure, but also for effective retrofitting actions and incremental performance evaluations. In particular, the task of designing efficient energy supply systems asks for a high temporal resolution of the estimation [3]. The built infrastructure and other factors impacting energy demand for space heating are diversely distributed in urban environments and hence analyses have recently been done on a high spatial resolution as well [4]–[6].

While thermodynamic models of buildings are well understood and assumed to achieve good results, simulated energy demand for space heating in urban environments deviate largely from measured data which is often accounted to micro-climate effects [7], [8] and to the impact of occupants [9]–[11]. "Buildings don't use energy: people do" [11] and hence the energy usage in a building can be accounted directly or indirectly to the way people use buildings. In residential buildings, people's impact on space heating demands stems mostly from the way people control the heating, ventilation, and air conditioning (HVAC) systems, the way people actively ventilate their buildings through window and door opening, the way

people shade their rooms through blinds and curtains, the way people functionally divide their residential homes, and people's occupancy and activities performed at home which attribute in the form of heat gains [12]. Given the complex nature of human behaviour a comprehensive understanding of the driving forces, relationships, and feedbacks is difficult. Several factors like awareness [11], socio-economic situations [13] and rebound effects [14], are discussed in the literature. Given the trend towards zero-energy buildings in which thermal losses from the building fabric are significantly reduced, the impact of occupants on energy demand is relatively increasing [11]. This trend emphasises the importance of quantifying the influence of occupants on energy consumption. On the city scale, former work includes high resolution energy models without explicit influence of occupancy [5], [15], [16], and models including people behaviour but only in aggregated form on lower spatial resolution [17].

The behaviour of occupants in energy simulation is typically represented through several sub-models. The foundation to most of them lies in determining time periods when the building is occupied [18]–[20]; sometimes even disaggregated into zones of the building [21]. On top of those, more detailed activity models [19], window opening models [22], and HVAC control models [23], are placed. Behaviour models are typically differentiated between deterministic and probabilistic types. Deterministic models assume a direct causal link between a driver of a certain behaviour and the actual derived action. While different methods exist, these types of models are based on rational decision making. Probabilistic models in contrast are based on likelihoods of different actions. The data source for occupancy and activity models in many cases is the so-called time use survey (TUS) data sets for which a standardising research centre exists [24].

This paper reports the progress of on-going work of introducing people occupancy models to high resolution bottom-up city models for space heating. Energy demand is driven by the occupancy of people in their households which itself is modelled through a time-heterogeneous Markov chain based on TUS data. Citizens are clustered through features of themselves and of the households they live in. A statistically viable urban population is formed through population synthesis. The model allows analyses of spatio-temporal patterns of energy use for space heating across a given district. It is implemented as an open-source agent-based model [25], which allows easy integration of other bottom-up effects impacting energy demand for space heating like people movement, people interactions, activity models, and urban microclimates, other energy uses like water heating and electricity, other environmental impacts e.g. on-air quality, or the supply side of building energy.

2 METHODOLOGY

2.1 Conceptual model

The general urban energy system as applied in this study consists of three distinct entities: citizens, HVAC controls, and dwellings whose models will be described in detail in the following sub sections. A dwelling forms a home for one to n citizens and incorporates exactly one HVAC control system. Fig. 1 shows a flow-chart of the model. The model is time-step based where in each time step k, each entity updates its state: first all citizens update their occupancy, i.e. determine whether they are at home or not. Second, the HVAC control system of each dwelling updates its heating set point, taking into account the occupancy of the dwelling. Lastly, each dwelling updates its indoor temperature and the thermal power needed for reaching it. An agent-based implementation of this conceptual model is available online and published under an open-source license [25].
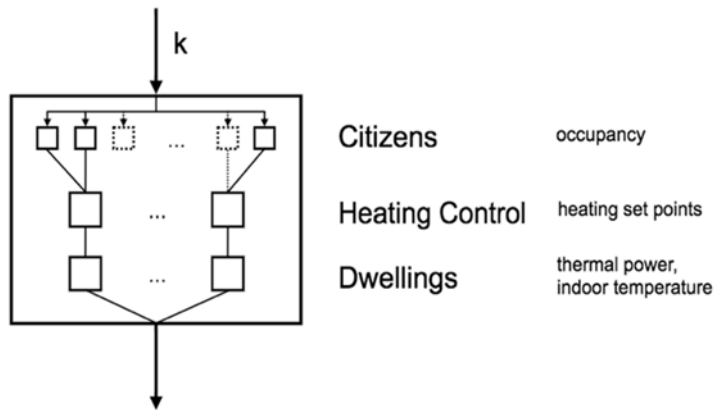
Figure 1: Flow chart of a single time step.

2.2 Heating system control model

The heating set point $\theta_{set,z}$ for a heating zone z can be understood as a function of P_z, the set of people inside the heating zone or related to it; L_{P_z} and A_{P_z}, the locations and activities of those people; and B_{P_z}, the heating behaviour which itself is influenced by many -- often unknown -- aspects like the comfort zone, awareness, socio-economic situation, and usage patterns of P_z.

For the simulation model of the heating system controls applied in this study, the following simplifications are made compared to the general model as defined above:

- zones are entire dwellings;
- time is discrete;
- location is equal to presence, i.e. we do not incorporate indoor positions;
- heating behaviour is based on three occupancy modes as described below.

The smallest unit considered is a dwelling d, by which we mean the fraction of a building that has a distinct energy meter and is occupied by a single household. Each dwelling d is of the set D, i.e. of the urban residential building stock. P_k^d is defined as a subset of the entire urban population P comprising of all people that occupy dwelling d at time k. Given these definitions, the heating set point for dwelling d at time $k \in K$ is defined as:

$$\theta_{set,k}^d = \begin{cases} \theta_{set,absent}^d, & \text{if } P_k^d = \varnothing \\ \theta_{set,active}^d, & \text{if } \{p \in P_k^d | \text{p is active}\} \neq \varnothing \\ \theta_{set,passive}^d, & \text{otherwise} \end{cases} \tag{1}$$

In this model, there are three distinct heating set points between which the control system toggles depending on occupancy. It shall be noted that assuming a desired comfort level defined by the set point temperatures $\theta_{set,absent}^d \leq \theta_{set,passive}^d \leq \theta_{set,active}^d$ this controller is close to optimal in terms of energy efficiency as the dwelling is minimally heated. As indoor

temperature lags behind occupancy it is not optimal in terms of comfort level. This effect is particularly strong when occupants enter a dwelling whose indoor temperature is far from $\theta_{set,active}^d$ or $\theta_{set,passive}^d$, but is ignored in this study.

2.2.1 Occupancy model

Citizens are modelled solely by the occupancy in their respective dwellings, using a probabilistic occupancy model that has been applied in several studies, albeit at the scale of single dwellings [18]–[20]. The occupancy model consists of a time-heterogeneous Markov chain with the following states: (1) not at home, (2) active at home, and (3) asleep at home. As the Markov chain is time heterogeneous, transition probabilities between the states of the Markov chain are time dependent, and hence the transition matrix Pr^p for person p at time k is defined by the collection of its entries $p_{i,j}^p(k) \forall i, j \in \{1, 2, 3\}$.

Every person has exactly one home, so we can define a time-invariant set of people P_d for every dwelling d such that the family of sets $P_D = \{P_d | \forall d \in D\}$ form a partition of population P and $\cup_{d \in D} P_d = P$ and $P_{d_1} \cap P_{d_2} = \varnothing \,\forall\, d_1 \neq d_2$ hold. The time dependent set of occupancy of dwelling d at time k as used in the heating system control model can then be given as $P_k^d = \{p \in P_d | p$ is active at home or p is asleep at home$\}$.

2.2.2 Thermal model of the dwellings

Dwellings are modelled following the simple hourly dynamic model of EN ISO 13790 [26]. Each dwelling is represented as a single thermal zone with squared floor area which is entirely heated. The dwelling consists of only one storey with equally sized windows facing north, south, east, and west. Ventilation and infiltration is limited to constant natural ventilation, and heat gains are limited to metabolic heat gains. Furthermore, heat transfer between dwellings is ignored.

The unknown and bounded heating power $\Phi_{HC,nd,k}^d$ is determined by the need to reach the set point temperature as defined by the heating system control. According to [26], it is assumed that the controller has a perfect dwelling model and can hence determine the necessary heating power in a precise manner.

2.3 Model calibration

The following subsection describes methods to calibrate the conceptual model as defined above. In particular two types of data sets are taken into account: (1) time use survey (TUS) data and (2) aggregated census data, which both are available for many regions of the world, and in the case of TUS data, even in a standardised manner [24]. The TUS data is used to calibrate the occupancy model, whereas the aggregated census data is used to generate a synthetic population. In addition to these, a micro sample of census data is necessary, i.e. fully detailed census data for a fraction of the population. For the approach described in this study, that data must be available in the TUS data set, i.e. for each participant whose time use is recorded in the study, we will demand features of the participant as well, e.g. their socio-economic situation.

2.3.1 Occupancy model

The transition matrix Pr^p for each citizen is derived from the TUS data set. The TUS data set contains location and activity data for each participant at a high temporal resolution in the form of a diary. Diaries span at least a day, though typically a weekday and a weekend day are recorded as time use varies between these days [24]. In the following, a mapping is

performed from each tuple of location and activity to one of the states of the Markov chain. As an example, the tuple (location = 'workplace, activity = 'lunch) is mapped to 1 = not at home, as is (location = 'second home', activity = 'sleep'). After the mapping, each diary can be understood as a concrete instance of a stochastic process that is described by the time heterogeneous Markov chain of the occupancy model. The set of participants is clustered by one or more household or people features F, e.g. age, which is available in both, the TUS data set and the aggregated census, and a time-heterogeneous Markov chain for each cluster following the approach used e.g. by [18,20] is created. The resulting set of transition matrices $\{Pr^f | \forall f \in F\}$ then allows us to deterministically allocate a transition matrix to a citizen based on the citizens feature value $f^p: Pr^p = Pr^{f^p}$.

The choice of household or people features is important for the quality of this approach. While theoretically all features could be used, in practise this might lead to sample sizes of that type of person which is too small. Hence, a subset of features must be chosen. Unfortunately, we are not aware of a deterministic way of choosing the *correct* subset of features. We instead acknowledge the inherent uncertainty and analyse features, their correlation among each other, and their correlation to the derived time series. We furthermore discuss the sensitivity of the results to the choice of features in the case study performed.

2.3.2 Synthetic population
The urban scale occupancy model as described above needs information on the composition of every household, and household and people feature vectors f^p for every individual p in the population of the study area. Such disaggregated data is typically not available and hence the population is synthesised from aggregated census data and typical population compositions given by micro samples of census data. Population synthesis as a way to initialise micro-simulations has been applied in the past mainly in land use and transportation models [27], [28], but more recently in energy models as well [29]. [30], provides an overview over the different approaches that are available.

In population synthesis, the aim is to estimate the joint probability mass function p_F for a set of n features $F = \{F_1, F_2, \ldots, F_n\}$ which describes the correct distribution of those features in the actual population, given a set of marginal probability mass functions $p_{\widetilde{F}}$ for a subset of features $\widetilde{F} \subset F$. As a simple example for the two-dimensional case of $F = \{\text{sex}, \text{age}\}$ consider that marginal probability mass functions $p_{sex}(f_1)$ and $p_{age}(f_2)$ are known and the joint probability mass function $p_{sex,age}(f_1, f_2)$ is to be estimated.

In the case of this study, the relationship between households and people must be retained when creating the synthetic population. Solutions for such hierarchical problems have been discussed in the past [31]–[33], and in the case of this study the approach of [32], is followed.

2.3.3 Thermal dwelling model and heating control system
To isolate the impact of occupants on the space heating energy demand in buildings in this proof-of-concept study, we are assuming the same physical conditions for all dwellings. We are hence defining a default dwelling and are allocating it to each household in the study area. This can be compared to normative building energy assessment where the object of study is the building and its impact on energy demand. Heating behaviour in these assessments is considered external to the object of study and equal among all buildings which allows to compare the physical structure of buildings only. Here, the object of study is the heating behaviour of occupants and its impact on energy demand. The physical structure of the building is considered external to the object of study and always equal among all households. Equal configuration is assumed for the heating control system as well.

3 CASE STUDY

The model is applied in a case study of the London Borough of Haringey which according to latest census data [34], is the home of 254,926 usual residents and comprises of 101,955 households. For this case study two data sets have been used: the UK Time Use Survey 2000 [35] and the UK 2011 Census [34]. The former contains time use diaries with a resolution of 10 min for a weekday and a weekend day for more than 11,500 individuals of nearly 6,500 households in the UK. The latter contains aggregated census data for so called output areas (OA) comprising of at least 40 resident households and 100 resident people to ensure confidentiality. Haringey is divided into 753 distinct output areas which are aggregated into 145 Lower Super Output Areas (LSOA) and furthermore into 19 wards. The source code for this case study is published under an open source license and is available online [36], [37]. Table 1 summarises the parameters of the thermal dwelling model applied for the dwelling of each household in Haringey.

3.1 Feature selection results

In the first step, features that are used to form the occupancy model and the synthetic population are examined. Associations between features and the occupancy time series in the TUS data set are determined and used as an indication of features that are more significant. Later, we will investigate the sensitivity of the simulation results to the choice of features.
To illustrate the necessity to cluster the sample population, Fig. 2 shows all occupancy time series in the data set (a) raw, and after clustering by (b) economic activity and (c) age. Each occupancy time series is depicted vertically using three colours: white for (1) not at home, dark grey for (2) active at home, and light grey for (3) asleep at home for both days starting from midnight weekday until the end of the weekend day. While some patterns like the stronger homogeneity at night times are clearly visible, the data appears noisy. After clustering by either economic activity or age, the in-cluster homogeneity becomes visually stronger and more patterns emerge.

Table 1: Model parameters applied in Haringey study.

| parameter name	value	parameter name	value
thermal mass capacity	9.9 MJ/K	initial indoor temp	20 ˚C
thermal mass area	150 m^2	absent set point	0 ˚C
floor area	60 m^2	passive set point	18 ˚C
room height	2.5 m	active set point	22 ˚C
window to wall ratio	0.19	max heating power	10 kW
U-value wall	0.26 W/(m^2 K)	metabolic heat active	140 W
U-value roof	0.12 W/(m^2 K)	metabolic heat passive	70 W
U-value floor	0.40 W/(m^2 K)	metabolic heat < 18	0.75
U-value window	1.95 W/(m^2 K)	natural ventilation rate	0.65 l/(s m^2)
transmission adjustment ground	0.91		
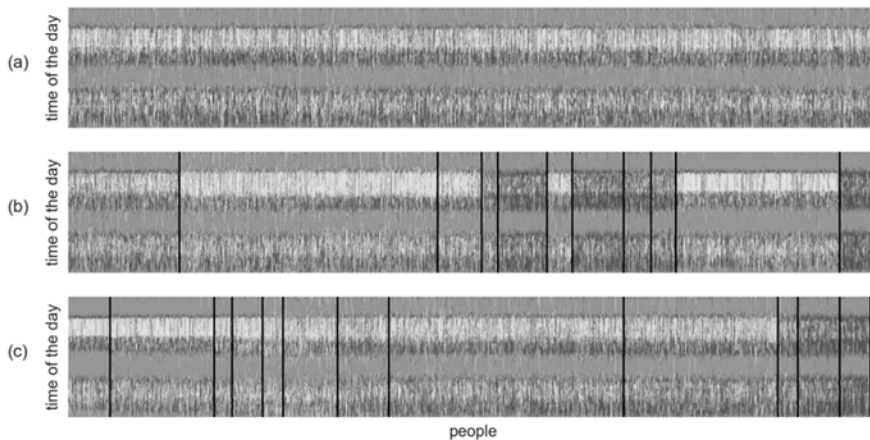
Figure 2:   Occupancy time series. (a) raw; (b) clustered by economic activity; (c) clustered by age groups. Occupancy is colour coded using white for (1) not at home, dark grey for (2) active at home, and light grey for (3) asleep at home.

In the following analysis, five people features and three household features are considered. People features are: the economic activity, highest qualification, age, and personal income of a person and whether that person is looking after another person. Household features include the composition of a household, the region it is located in, and the population density in the household's postcode sector. As all features and the occupancy time series are nominal, we will use Cramér's V method [38], to denote the strength of association between a feature and the occupancy, which ranges between 0 and 1, with Φ_C. The original method is shown to be a biased estimator [39] and we will hence use the suggested bias correction. We will measure association between a subset of features and occupancy for each of the 288 time steps of the time series. Fig. 3 shows the results for a selected number of subsets.
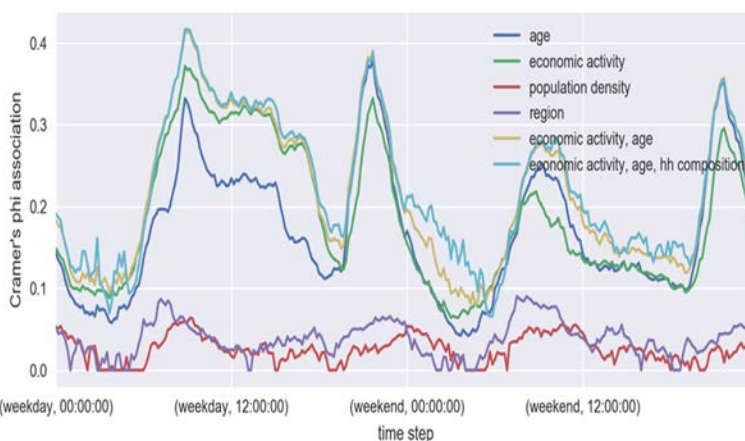

Figure 3: Time dependent association between people features and occupancy.

   A strong time dependency is visible, indicating that at certain times of the day occupancy is stronger associated to people or household features than on others. Weekday and weekend day show similar, but distinct patterns, and the weekend day reveals a weaker association at almost all time steps of the day. The region and the population density of the household's surrounding area don't show a significant association at any time of the day, indicating that there are no significant differences in occupancy patterns between, e.g. rural and urban environments. Among single features, economic activity and age show the highest association throughout both days; a finding that is in line with former research [19]. Combined features reveal an association that is higher than single features at most times of the day. Combination of three features, like the depicted combination of economic activity, age, and household type show the highest association but their curves are unsmooth, indicating overfitting and sample sizes that are too small.

   Table 2 shows the average association over the day for selected feature combinations and the average size of their clusters. The combination of economic activity, age, and household type reveals the highest average association, but also has a low average sample size. Among the tuple combinations of features, Table 2 shows the four with the highest average association, which have a significantly larger average cluster size than combinations of three features.

   Based on these results it is concluded that: (1) There exists no strong argument for assuming different occupancy patterns based on spatial location only, e.g. in rural or urban areas. Hence, although our case study is an urban environment, the TUS data set is not filtered by spatial location. (2) Combinations of 3 features reveal high association with the occupancy time series but seem overfitted. Their cluster sizes are too small to be used to fit the occupancy Markov model and hence combinations of 3 features are not considered. (3) Combinations of 2 features reveal higher associations with the occupancy time series than single features while their cluster sizes remain reasonably high. Several combinations show high association and the sensitivity of the model to those is investigated.

Table 2:   Average of time dependent association between people and household features $\overline{\Phi_C}$ and average cluster size for feature combinations with highest average association.

features	$\overline{\Phi_C}$	avg cluster size
economic, age, hh composition	0.22295	35
economic, age	0.21534	117
economic, qualification	0.19142	121
economic, hh composition	0.18996	202
economic, region	0.18953	66

3.2 Simulation results

Using economic activity and age as discriminating features, a model of Haringey can be formed by creating a synthetic population for every LSOA following a zone-by-zone approach [30] and the model can by simulated by the open-source simulation model [25]. The simulation is run for two winter days, a weekday and a weekend day, and the results for both days are scaled appropriately energy-wise to represent a full week. Due to the population synthesis, model results are aggregated on the LSOA level. Fig. 4 shows a choropleth map

of Haringey where the (a) average, and (b) standard deviation of the energy demand for one full week per dwelling is depicted for every LSOA.
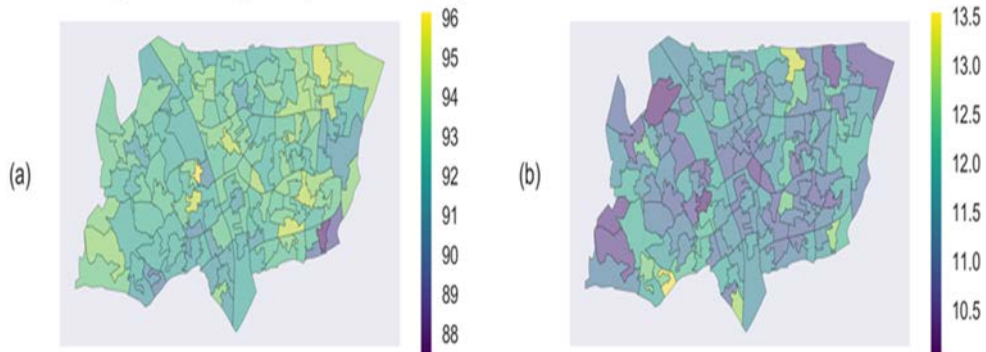

Figure 4: Simulated energy demand in each LSOA of Haringey for a week in January [kWh/week]. (a) Average across all households; (b) Standard deviation across all households.

The map reveals a few spatial patterns, most significantly a higher energy demand in the north-eastern corner of Haringey and a negative correlation between the two. The variation in average energy demand is though only marginal, with less than 6% difference between most LSOA.

Fig. 5 depicts the thermal power profiles for each LSOA, as average and standard deviation among all households in one area. The average thermal power varies between LSOA's with up to 400 W per household. Large variations can be seen during daytime, and in particular in the evening hours. The temporal pattern follow the ones that are visible in the occupancy time series depicted in Fig. 2. The standard deviation shows a more permanent spread across the LSOAs, with around 300 W per household peak and significant variations in particular in the evening hours of weekdays but in particular also during daytime on weekends.
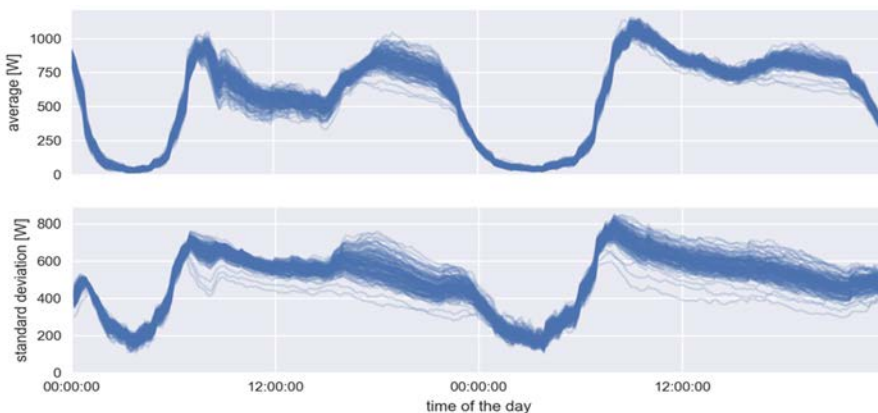

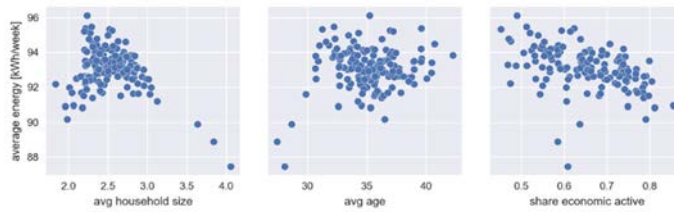Figure 5: Average thermal power per household across LSOAs.

Figure 6: Pairwise relationships of population features and simulated energy usage per LSOA.

To understand relationships between people and household features and the energy usage as seen above, Fig. 6 depicts the pairwise distributions between the average household size, the average age, and the share of economically active people in a LSOA and the energy usage as determined by the simulation model in terms of average energy usage per household. The only visible trend is a negative correlation between the share of economically active people and the average energy demand. In general, a direct link between single feature values of a LSOA and its energy demand does not seem possible. It shall be noted that the sensitivity of these results towards simulation parameters like season and length of the simulation must be further investigated.

3.3 Results for alternative feature selection

To explore the sensitivity of the methodology to the selected subset of features, further simulations are run with different choices of features for the clustering of the population sample and as a control variable for the population synthesis. In particular, the features age and highest qualification have been chosen to run. In addition, one simulation has been performed without clustering the population by not taking into account any feature. The results are shown in Fig. 7 in the form of the difference to the simulation above of aggregated thermal power per household on the ward level. The average per ward varies rather strongly during daytime on the weekday while all feature selections reveal similar results on the weekend. The difference in standard deviation is particularly strong in the evening hours and generally throughout the weekend.
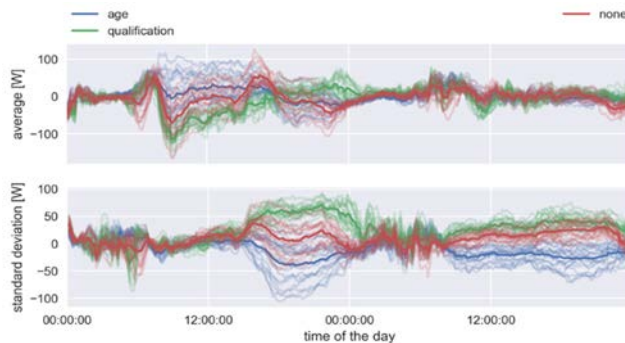

Figure 7: Difference to the original run in average thermal power per household in different wards.

4 DISCUSSION AND CONCLUSION

This paper evaluates the usefulness of microsimulations for urban energy systems by presenting the proof-of-concept of introducing occupancy models to high resolution bottom-up energy models. A synthetic population is created from census data and the occupancy of every citizen is modelled using a time heterogeneous Markov chain which is calibrated using time use survey data. The fraction of the variability of the thermal energy demand caused by building specifics is deliberately ignored and heating system control is considered to be based on occupancy only to highlight the impact of occupants. This study is hence hypothetical and aims at understanding underlying mechanics of energy demand rather than actual energy demand.

In the application of this methodology to the London Borough of Haringey, the average thermal energy demand for space heating varies only marginally between the 145 LSOAs of in average around 1700 citizens into which Haringey is divided and to which results are aggregated. The standard deviation of energy demand per dwelling varies much more among the LSOAs indicating a higher diversity of household compositions in these parts of the city. When looking at thermal power profiles similar temporal patterns to the ones from the input time use data can be seen: low variability in average thermal power during the night and on the weekends, and a larger variability during working hours. At noon on a workday for example, the LSOA with the lowest average power shows a thermal power of roughly 400 W per household, while the LSOA with the highest average power shows a thermal power above 600 W per household. The standard deviation of thermal power between dwellings varies strongly in the evening hours of workdays and during the weekend.

The results are sensitive to the selection of people and household features that are used to calibrate the model. Further work must be spent on understanding the uncertainty introduced through this approach. Probabilistic methods as proposed by [19], [40], should be compared to this deterministic approach.

The results support the suitability of models with low resolution when aggregated results are demanded. In cases where higher temporal or spatial resolution is needed, as it is for example for the design and optimised operation of decentralised energy supply systems, detailed micro simulations as performed in this study offer the potential to understand and predict urban energy demand better. Without correct representation of variability across population and household features, aspects like the timing, duration, and total amount of peak energy consumption are easily over or underestimated.

This study focusses on only one aspect of urban energy demand, namely occupancy, and generates energy profiles for a hypothetical scenario. The methodology and simulation framework allow to analyse more aspects like the spatial distribution of heating patterns, set point preferences [40], fraction of dwellings heated [40], and electricity demand [19] and incorporating those will allow to create a more comprehensive picture of urban energy demand.

REFERENCES

[1]   Global Energy Assessment. Toward a Sustainable Future. Cambridge University Press, Cambridge, UK; New York, NY, USA; the International Institute for Applied

Systems Analysis, Laxenburg, Austria. http://www.globalenergyassessment.org. Accessed on: 10 Apr. 2017.

[2]   United Nations. World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352). United Nations, Department of Economic; Social Affairs, Population Division. https://www.compassion.com/multimedia/world-urbanization-prospects.pdf. Accessed on: 20 Feb. 2017.

[3]   Omu, A., Choudhary, R. & Boies, A., Distributed energy resource system optimisation using mixed integer linear programming. *Energy Policy*, **61**, pp. 249–66, 2013.

[4]   Choudhary, R., Energy analysis of the non-domestic building stock of Greater London. *Building and Environment,* **51**, pp. 243–54, 2012.

[5]   Li, Q., Quan, S.J., Augenbroe, G., Yang, P. & Brown J., Building energy modelling at urban scale: integration of reduced order energy model with geographical information. *Building Simulation,* 2015.

[6]   Choudhary, R. & Tian, W., Influence of district features on energy consumption in non-domestic buildings. *Building Research & Information*, **42**(1), pp. 32–46, 2013.

[7]   Kolokotroni, M. & Giridharan, R., Urban heat island intensity in London: An investigation of the impact of physical characteristics on changes in outdoor air temperature during summer. *Solar Energy*, **82**(11), pp. 986–98, 2008.

[8]   Steemers, K., Energy and the city: density, buildings and transport. *Energy and Buildings*, **35**(1), pp. 3–14, 2003.

[9]   Seryak, J. & Kissock, K., Occupancy and behavioural affects on residential energy use. *Proceedings of the Solar Conference*, pp. 327–39, 2003.

[10] Fabi, V., Andersen, R.V., Corgnati, S & Olesen, B.W., Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models. *Building and Environment*, **58**, pp. 188–98, 2012.

[11] Janda, K.B., Buildings don't use energy: people do. *Architectural Science Review*. 2011.

[12] Van Raaij, W.F. & Verhallen, T.M.M., A behavioural model of residential energy use. *Journal of Economic Psychology*, **3**(1), pp. 39–63, 1983.

[13] Druckman, A. & Jackson, T., The carbon footprint of UK households 19902004: A socio-economically disaggregated, quasi-multi-regional inputoutput model. *Ecological Economics*, **68**(7), pp. 2066–77, 2009.

[14] Brännlund R., Ghalwash T. & Nordström J., Increased energy efficiency and the rebound effect: Effects on consumption and emissions. *Energy Economics*, **29**(1), pp. 1–17, 2007.

[15] Tian, W., Rysanek, A., Choudhary, R. & Heo, Y., High resolution energy simulations at city scale. *Proceedings of the 14th International Conference of Ibpsa-Building Simulation,* pp. 239–46, 2015.

[16] Zhang, Y., Soga, K., Choudhary, R. & Bains, S., GSHP application for heating and cooling at City Scalefor the city of westminster. *Proceedings world geothermal congress*, 2015.

[17] Bustos-Turu, G., Dam, K.H. van & Acha, S., Simulating residential electricity and heat demand in urban areas using an agent-based modelling approach. *Energy Conference (ENERGYCON), 2016 IEEE International*, pp. 1–6, 2016.

[18] Richardson, I., Thomson, M. & Infield, D., A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, **40**(8), pp. 1560–6, 2008.

[19] Aerts, D., Occupancy and activity modelling for building energy demand simulations, comparative feedback and residential electricity demand characterisation. *Architectural Engineering (ARCH)*. 2015.

[20] Widén, J., Nilsson, A.M. & Wäckelgård, E., A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand. *Energy and Buildings*, **41**(10), pp.1001–12, 2009.

[21] Liao, C., Lin, Y. & Barooah, P., Agent-based and graphical modelling of building occupancy, *Journal of Building Performance Simulation*. **5**(1), pp. 5–25, 2012.

[22] Andersen, R., Fabi, V., Toftum, J., Corgnati, S.P. & Olesen B.W., Window opening behaviour modelled from measurements in Danish dwellings. *Building and Environment*, **69**, pp. 101–13, 2013.

[23] Fabi, V., Andersen, R.V. & Corgnati, S.P., Influence of occupant's heating set-point preferences on indoor environmental quality and heating demand in residential buildings. *HVAC&R Research*. 2013.

[24] Centre for Time Use Research. https://www.timeuse.org. Accessed on: 1. May 2017.

[25] Tröndle, T. & Tang, D., timtroendle/energy-agents: v1.0.0. Zenodo. https://doi.org/10.5281/zenodo.579761. Accessed on: 6 May 2017.

[26] ISO 13790:2008: Energy Performance of Buildings: Calculation of Energy Use for Space Heating and Cooling. International Organization for Standardization, Geneva, Switzerland.

[27] Beckman, R.J., Baggerly, K.A. & McKay, M.D., Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, **30**(6), pp. 415–29, 1996.

[28] Arentze, T., Timmermans, H. & Hofman, F., Creating Synthetic Household Populations: Problems and Approach. *Transportation Research Record: Journal of the Transportation Research Board*, **2014** (1), pp. 85–91, 2007.

[29] Dam, K.H. van & Bustos-Turu, G., A methodology for simulating synthetic populations for the analysis of socio-technical infrastructures. Presented at *11th Conference of the European Social Simulation Association*, Groningen, Netherlands, 2015.

[30] Müller, K. & Axhausen, K.W., Population synthesis for microsimulation: State of the art. http://dx.doi.org/10.3929/ethz-a-006127782. Accessed on: 25 Oct. 2016.

[31] Ye, X., Konduri, K., Pendyala, R.M. & Sana, B., A methodology to match distributions of both household and person attributes in the generation of synthetic populations. Presented at *Transportation Research Board 88th Annual Meeting*, 2009.

[32] Müller, K. & Axhausen, K.W., Hierarchical IPF: Generating a synthetic population for Switzerland. http://www-sre.wu.ac.at/ersa/ersaconfs/ersa11/e110830aFinal00305.pdf. Accessed on: 6 May 2017.

[33] Pritchard, D.R. & Miller, E.J., Advances in population synthesis: fitting many attributes per agent and fitting to household and person margins simultaneously. *Transportation*, **39**(3), pp. 685–704, 2011.

[34] Office for National Statistics. United Kingdom 2011 Census. https://www.ons.gov.uk/census/2011census. Accessed on: 6 May 2017.

[35] Ipsos-RSL, Office for National Statistics. United Kingdom Time Use Survey, 2000. UK Data Service. http://doi.org/10.5255/UKDA-SN-4504-1. Accessed on: 6 May 2017.

[36] Tröndle, T., Choudhary, R. timtroendle/urban-occupants-paper: v1.1. Zenodo. https://doi.org/10.5281/zenodo.584113. Accessed on: 29 May 2017.

[37]   Tröndle, T., timtroendle/pytus2000: v0.5.1. Zenodo. https://doi.org/10.5281/zenodo.579661. Accessed on: 15 May 2017.

[38]   Cramér, H., *Mathematical Methods of Statistics*, Princeton: Princeton University Press, 1946.

[39]   Bergsma, W., A bias-correction for Cramérs and Tschuprows. *Journal of the Korean Statistical Society*, **42**(3), pp. 323–328, 2013.

[40]   Baetens, R. & Saelens, D., Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour. *Journal of Building Performance Simulation*, **9**(4), pp. 431–47, 2016.