COMMUNITY-SCALE HOUSEHOLD ACTIVITY MODELLING CONSIDERING HOUSEHOLD HETEROGENEITY USING JAPANESE TIME USE DATA

Y. Yamaguchi, Y. Kambayashi, T. Okada, Y. Shoda,
A.Taniguchi-Matsuoka, Y. Shimoda
Osaka University, Osak, Japan

ABSTRACT

Activity-based residential energy demand modelling, in which activity of household members is stochastically modelled and energy demand is quantified based on simulated activity, is useful for various applications, since the model enables to enhance analytical functions as it replicates the structure in which energy demand is determined in the real society. Instead of its usefulness, methods to apply activity-based modeling in community-scale energy demand modelling has not been well established. Based on this background, this paper proposes a method to link the composition of households living in communities with household activity modelling parameters as a component of an activity model as a component of a community scale energy demand model. The developed model first generates households living in simulated community based on population census. This household generation process enables to replicate the composition of households in assumed community. Then, parameters of activity modelling such as probability undertaking activtiies at each time day were given to each household member living in simulated community based on logistic regression models developed based on time use data. The logistic regression model was designed to consider demographics and household characteristics as predictor variables. As a result, the composition of households in simulated community is reflected in the activity modelling. A case study demonstrated how differences in household composition affect time use of people living in Japanese pin-code districts in a city.

INTRODUCTION

Community-scale energy demand modelling has attracted attention in energy management and community development. Researchers have developed methodologies to model energy demand at high temporal and spatial resolution (e.g. Baetens et al. 2016). One of the key modelling methods for energy management is so-called activity-based modelling (Sivakumar, 2013) in which activity of household members is stochastically simulated and energy demand is quantified based on simulated activity. Consideration of activity in energy demand modelling is useful to enhance models' analytical functions, since the model enables to replicate the structure in which energy demand is determined in the real society. There is also possibility to improve the spatiotemporal resolution of energy demand modelling.

There has been a number of methods developed to model people's activity. Widén et al. (2012) used Markov chain considering activity as transition states. Wilke et al. (2013) used logistic regression and survival analysis. In their model, activity is modelled as sequence of discrete event with start time and duration. Starting of activity was modelled as a selection of activity among all and selection probability was modeled by multinomial logit model. The duration of activity was modelled by survival analysis. Tanimoto et al. (2010) and Yamaguchi et al. (2017) proposed similar discrete event models. In the context of community-scale energy demand modelling, one of the most important criteria is the consideration of demographics and household characterisitcs in order to reflect characteristics of community in modeled energy demand. Wilke's model is the most advantageous because the method is capable of considering socio-demographic conditions of simulated people as predictor variables in the activity selection and duration determination processes. Instead of its usefulness, methods to apply these activity models in community-scale energy demand modelling has not been well established. This paper proposes a method to integrate an activity model with a community scale energy demand model as a core component for reflecting charactericitics of communities.

METHOD

The developed model first generates households living in simulated community based on population census. Households are modelled as a combination of household members with specific demographics such as age, gender, work/school status and occupation as listed in Table 1. By the combination of household members, household characteristics, such as household size and household composition can be defined. Housing condition is also given. For this process, a number of probability distributions are generated based on population census and the demographics and household characteristis are randomly sampled. This household generation process enables to replicate the composition of households in simulated community. It also enables to deal with individual household member as the unit of activity modelling and individual household as the unit of energy demand modelling.

Table 1. Demographics and household characteristics

Category	Item	Note
Demographics of household member	Age	Categorized with 5-year age range
	Gender	Male or female
	Occupation	Fulltime, part-time, and unemployed
	Children's school status	Preschool, elementary, junior and high school and university
Household characteristics	Household size	Number of household members
	Household composition	Couple's age, age difference between children and parents; number of children with their age, married couple's occupation status
	House ownership	Owner or renter
	Housing type	Single detached house or multi-family building

In the abovementioned existing activity modelling methods using discrete event modelling approach, models basically use 1) activity starting probability at which each activity (*n*) is started at each time step (*t*), $Pst_{n,t}$, 2) activity undertaking probability at which each activity is being undertaken at each time of step, $Pact_{n,t}$ and 3) activity duration modelled by the form of cummulative probability distribution, $CPdu_{n,t}$,. Based on Wilke et al. (2013), this study qunatifies these parameters based on time use data (TUD) and logistic regression analysis. $Pst_{n,t}$ and $Pact_{n,t}$ are modeled by the logit model shown in Equation (1):

$$\ln\frac{P_t}{1-P_t} = \beta_{t,0} + \sum_{m=1}^{M}\beta_{t,m}\cdot x_m \tag{1}$$

where P_t describes one of $Pst_{n,t}$ and $Pact_{n,t}$ at time *t*, x_m is the m^{th} predictor variable, $\beta_{t,0}$ and $\beta_{t,m}$ are the regression coefficients. $CPdu_{n,t}$ is modelled by ordered logit model shown by Equation (2):

$$\ln\frac{CPdu_{t,Du}}{1-CPdu_{t,Du}} = \beta du_{t,Du} - \sum_{m=1}^{M}\beta du_{t,m}x_m \tag{2}$$

where $CPdu_{t,Du}$ describes cummulative probability distribution of one of activity *n* starting from time *t* with activity duration, *Du*, $\beta du_{t,Du}$ and $\beta du_{t,m}$ are the regression coefficients. As shown in Equation (2), it was assumed that regression coefficients are constant for all activity duration, *Du*.

Japanese TUD (SBJ, 2006) was used as input data for the regression analyses. Japanese TUD has two types of TUD collected by using two diary formats, Diaries A and B. The survey using Diary A collected time use diaries from people living in approximately 80,000 households representing Japanese population with a classification of time use into 20 activities. The total number of samples in Diary B was 18 thousand people in approximately 8000 households. Diary B classified time use into 85 activities, which allowed more direct conversion from activities to energy use. For example, household maintenance activity, which was one of 20 activities in Diary A, was divided into 14 activities that included laundry and cooking activities that could be linked to the operation of the washing machine and the kitchen stove. Due to the sample size difference, $Pst_{n,t}$ and $Pact_{n,t}$ were quantified with 15-min interval (as same as Japanese TUD) when Diary A data was used, whereas data were quantified with 1-hour interval when Diary B data was used.

In order to take into account the influence of demographics and household characteristics on activity, twelve sets of regression models were developed for each $Pst_{n,t}$, $Pact_{n,t}$ and $CPdu_{n,t}$ by considering demographic segment made by gender, occupation, age and children's school status as listed in Table 2, as well as the distinction between weekdays and holidays. The twelve segments are considered because our analysis showed that the influence of predictors varies significantly among the segments, though these factors were considered as predictor variables in Wilke et al. (2013). In addition to this, only relevant demographics and household characteristics that can be given by household generation model (i.e. Japanese population census) were selected. For example, household income was not considered even though data is available in TUD, because household income is not available in population census. Table 3 lists the predictor variables considered in this study. All variables are modelled as a dummy variable. The predictor variables were selected in the regression models so that Akaike's Information Criterion (AIC) could be minimized. It should be noted that geographical/locational information is not considered when Diary B is used. Thus, geographical/locational difference is only generated by the composition of households in district regarding the items listed in the table.

Table 2. Demographic segments of household members

Segment	Description	Segment	Description
Student	Primary, secondary, high school, university/college and graduate school student	Adult male	Male aged 15 to 59 years old and finished education
Working female	Female aged 15 to 59 years old with a job and finished education	Unemployed female	Unemployed female aged 15 to 59 years old and finished education
Elderly male	Male aged 60 years old or older	Elderly female	Female aged 60 years old or older

Table 3. Predictor variables for modelling activity from TUD.

Variable	Item	Note
Age	Aged 10–29, 30–44, 45–59, 60–74, 75 or older	Not considered for students.
Children's school status	Primary, junior, and high school, and university student	Only considerd for student category.
Work status	fulltime, part-time, and unemployed	Not considered for students.
Occupation	Professional and technical workers, managers and officials, sales workers, protective service workers, agricultureal, forestry and fishery workers, workers in transport and communicational occupation, craftmem, mining workers, manufacturing and construction workers	Only considered when Diary A is used.
Housing	Owner-occupied house and rented house	
Income	Single and double income	We did not consider the income for female because there was a strong correlation between occupation and double income, because most adult male has a job.
Commuting time	Less than 30 minutues, 30 to 60 minutes, 60 to 120 minutes, longer than 120 minutes.	Only considered when Diary A is used.
City size	Four segment classified by population size of cities.	Only considered when Diary A is used.
Three-generation household	Living in the three generation household or not	Only considered for those aged 60 or older
Youngest children	Youngest child is in pre-school, in primary, junior-high or high school, and people without preschool- or school-aged children	Only considered for adult male and working and unemployed female
Household size	Number of household size is 1, 2, and 3 or larger	

CASE STUDY

This section presents a case study in which the model presented in the previous section was applied to 41 pin-code districts of Komae city, Tokyo, Japan, respectively to estimate $Pact_{n,t}$ of all household members based on Diary B. First, all households were generated based on population census data of each district. The regression models developed for $Pact_{n,t}$ were applied to all household members. The result was aggregated for each district to quantify mean values of $Pact_{n,t}$. This case study shows how differences in the composition of households influence people's activity conducted at home.

The city has various districts in terms of the composition of households. As an example, Figure 1 shows the proportion of households with different household size in the districts. The number of household in each district ranges from 320 to 2750. The minimum value of the mean household size was observed in District 4, Higashiizumi 4th, which was 1.57 persons/household. In the district, single households occupy 69% of the total. 51% of household heads were younger than 40 years old. 79% of households lived in a multi-family building. On the other hand, District 20, Iwadominami 4th, had the highest mean of household size, 2.55 persons/household. Single household only occupies 24% of the total, while households with 3 or more members occupied 42%. 44% of household heads were aged 40 to 59 and 32% were aged 60 or older.

Figure 2 shows the mean number of people who undertake activities at home in the two pin-code districts, Higashiizumi 4th and Iwadominami 4th. The figure summarizes activities into 9 activity categories listed in the figure legend. To quantify the values, $Pact_{n,t}$ on weekdays and holidays were first summed in each generated household. Then, the result was averaged among all households in each district. Due to the difference in household size, the two districts have considerably different profile of activity undertaken per household basis as shown in the figure. In order to clarify the difference in activity due to the composition of households, Figure 3 shows the difference in the probability undertaking each activity per person. To quantify the difference, probability shown in Figure 2 was divided by the mean household size. As shown in the figure, the proportion of sleeping is higher in Higashiizumi 4th from 4:00 to 11:00. On the contrary, more activity is conducted in Iwadominami 4th from 12:00 to 3:00, especially for watching TV or media use. However, the total amount of time spent at home was not so different, which was estimated to be 0.2 hours/person on both weekdays and holidays. Thus, the difference was attributed to the later waking time in the morning and later home arrival time from work and outing activity in Higashiizumi 4th compared to Iwadominami 4th. This result implies that the difference in the composition of households has a significant influence on the occupancy and time allocation of activities at home.

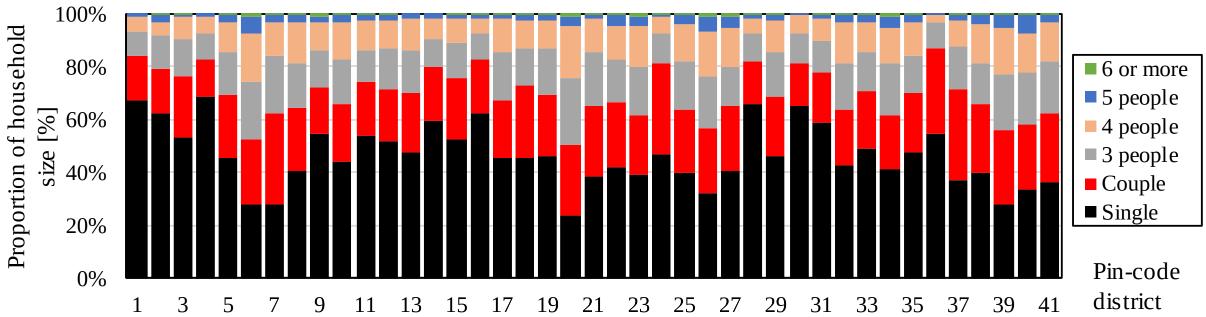

Figure 1. Proportion of household with different household size in the 41 pin-code districts in Komae city.

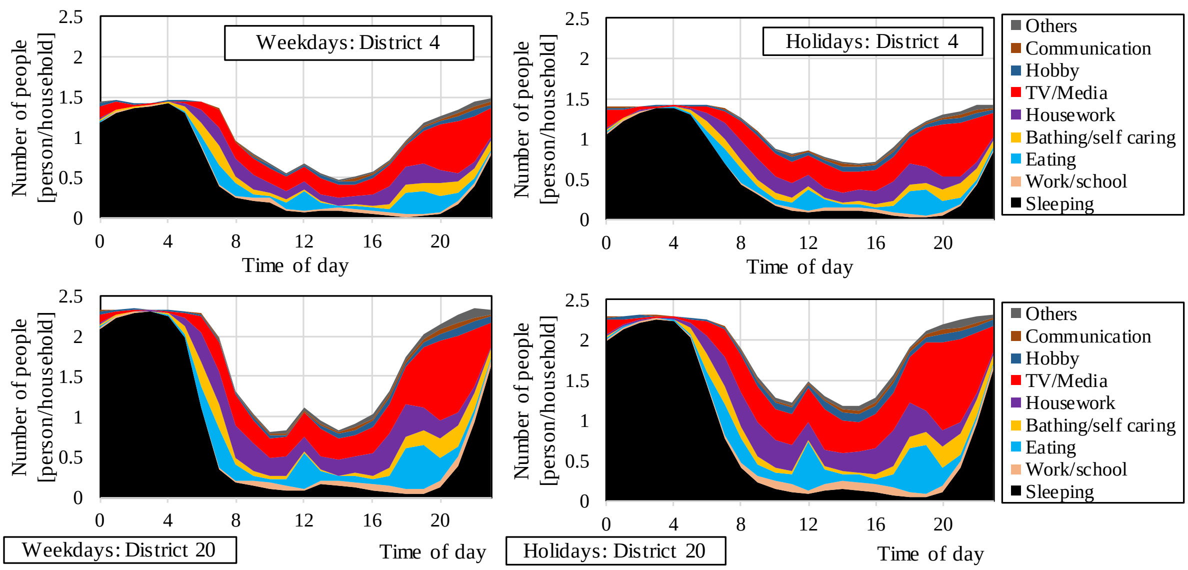

Figure 2. Mean number of people who undertake activities per household in two districts with the smallest (District 4: upper) and largest (District 20: bottom) household size.

Finally, in order to evaluate the influence of difference in the composition of households more comprehensively, we quantified mean number of people who undertake all activities at each time of day in each pin-code district. Figure 4 shows distribution of the mean and standard deviation among pin-code districts. Each dot shows one of the combinations of an activity and a time of day. This result includes the influence of household size in the districts, the influence was normalized by dividing the result by the mean household size of each district. The result is shown in Figure 5. The figure shows the relationship between the mean and standard deviation of undertaking probability of activities at each time of day. Standard deviation represents the variety in activity undertaking probability caused by the difference in demographics and household characteristics in pin-code districts. As shown in the figure, there are some combinations with relatively high standard deviation. Since it is difficult to observe the combinations with small mean probability, Figure 6 shows the distribution of the coefficient of variation given by dividing standard deviation by mean. The horizontal axis shows the mean activity undertaking probability, whereas the vertical axis shows its coefficient of variation. As shown in the figure, the coefficient of variation of most activities was smaller than 0.5. Although there are some combinations of activity and time that have a high coefficient larger than 0.5, their mean activity undertaking probability were smaller than 1%.

CONCLUSION

This paper proposed an approach to link household composition in communities with household activity modelling parameters as a core component of community-scale energy demand modelling. The key feature of the approach is its household generation process in which simulated households with demographics and household characteristics were randomly sampled from population census so that the composition of households is reflected. Then, activity of household members is stochastically simulated as sequence of activities modelled as discrete event based on model parameters, such as activity undertaking probability, given based on logistic regression model considering demographics and household characteristics of generated households as predictor variables. This approach would strengthen the capability of reflecting the composition of households in a community in estimated energy demand. The developed model was applied to 41 pin-code district in Komae city, Tokyo, to demonstrate how the composition of household affects time use of residents at home. Based on the case study, the following implications were derived:

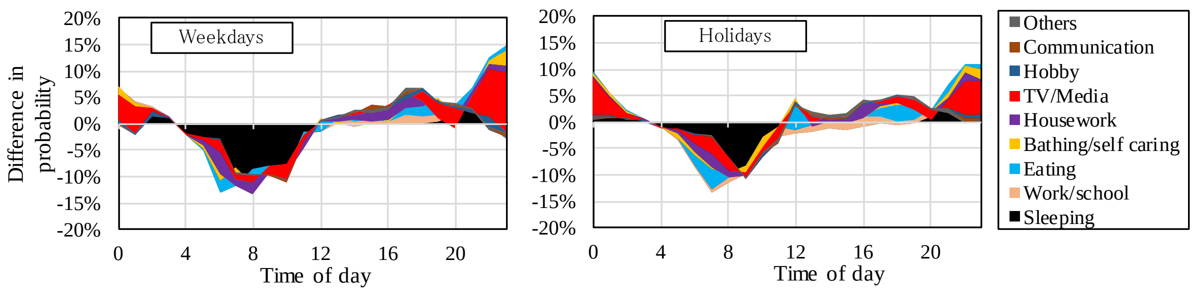

Figure 3. Difference in activity probability per households in the two pin-code district.

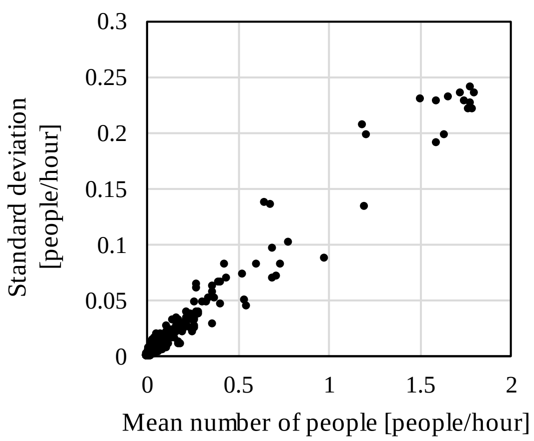

Figure 4. Mean and standard deviation of the mean number of people undertaking activities among districts

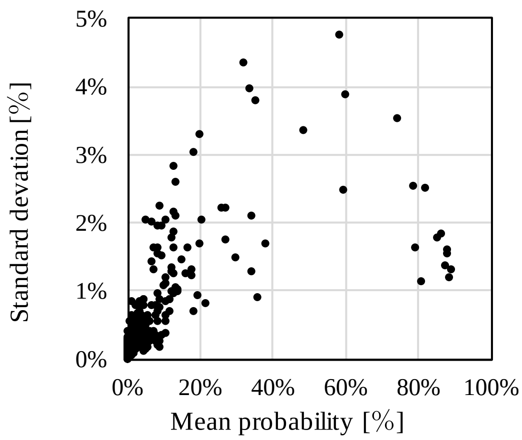

Figure 5. Mean and standard deviation of the activity undertaking probability

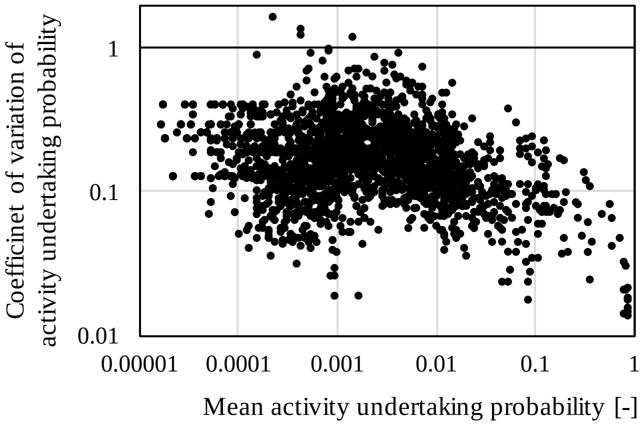

Figure 6. the relationship between mean activity probability and its coefficient of variation

- Activity undertaken in a community is considerably characterised by the composition of households;
- Household size is the most influential factor determining number of people spending time in community;
- Average time use per person is different among communities due to the difference in cumulated demographics and household characteristics in each pin-code. The influence is observed in activities with high activity undertaking probability, such as sleeping and watching TV. Although there are some other activities that undertaking probability vary widely among communities, their activity undertaking probability is relatively small (less than 5%);

ACKNOWLEDGEMENT

This work was supported by JST CREST under Grant JPMJCR15K4, Japan.

REFERENCES

Baetens, R., and Saelens, D. 2016. Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour. Journal of Building Performance Simulation 9(4): 431–447. doi:10.1080/19401493.2015.1070203

Sivakumar, A. 2013. Activity-based modelling for urban energy systems, in Urban energy systems: an integrated approach, edited by Keirstead, J. and Shah, N. Routledge, pp. 204-217.

Statistics Bureau of Japan (SBJ), Ministry of Internal Affairs and Communications. 2006 Survey on time use and leisure activities.

Tanimoto, J., Hagishima, A. 2010. Total utility demand prediction system for dwellings based on stochastic processes of actual inhabitants. Journal of Building Performance Simulation, 3(2), 155–167.

Yamaguchi, Y., Shimoda, Y. 2017. A stochastic model to predict occupants' activities at home for community-/urban-scale energy demand modelling. Journal of Building Performance Simulation, 10, 565–581.

Widén, J., Molin, A., Ellegård, K. 2012. Models of domestic occupancy, activities and energy use based on time-use data: deterministic and stochastic approaches with application to various building-related simulations. Journal of Building Performance Simulation, 5(1), 27–44. https://doi.org/10.1080/19401493.2010.532569

Wilke, U., Haldi, F., Scartezzini, J.-L., Robinson, D. 2013. A bottom-up stochastic model to predict building occupants' time-dependent activities. Building and Environment, 60, 254–264.