Spatial Analysis on Routine Occupant Behavior Patterns and Associated Factors in Residential Buildings

Yunjeong Mo, Ph.D., A.M.ASCE[1]; and Dong Zhao, Ph.D., M.ASCE[2]

[1]Dept. of Construction Management, College of Computing, Engineering, and Construction, Univ. of North Florida, Jacksonville, FL. Email: y.mo@unf.edu
[2]School of Planning, Design and Construction, Dept. of Civil and Environmental Engineering, Michigan State Univ., East Lasing, MI. Email: dzhao@msu.edu

ABSTRACT

Residents have full control over home systems and place a significant influence on energy consumption. Despite advanced building technologies and energy-efficient appliances, energy consumption in residential buildings remains high in the US and around the world. This can be explained by the interaction effects between building technology and occupant behavior. Given an increasing number of studies on occupant behavior, residents' routine daily energy use activities remain unclear in the literature. The goal of this study is to identify the components that influence the similarity and differences of energy usage-related activities. To achieve the goal, this study aims to (1) compare the components of routine occupant behaviors using the US national behavior data by region, and (2) identify if geographical location affects the characteristics of activities using GIS. The findings inform that duration, start time, and end time have more influences on the differences in energy usage-related activities, and watching TV, washing and grooming, and cooking and food preparation are more different by geographical location. The result can be used to provide more reliable information regarding energy and behavior to the occupants in residential buildings. Also, the result can be applied to the new energy and behavior strategies and policies about residential building energy plans.

INTRODUCTION AND BACKGROUND

Residential occupants have significant influences on and control over energy consumption compared to other types of building occupants, and it emphasizes the role of occupant behavior in residential energy savings. Despite advanced building technologies and energy-efficient appliances, energy consumption in residential buildings remains high in the U.S. and around the world. The energy consumption can be explained by the interaction effects between building technology and occupant behavior (Zhao et al. 2017). Given an increasing number of studies on occupant behavior, residents' daily energy use activities remain unclear in the literature.

The goal of this study is to identify the factors that influence the similarity and differences in energy usage-related activities. To achieve the goal, this study (1) compares the components of routine occupant behavior (ROB) using the U.S. national behavior data by region, and (2) identifies if and how geographical location affects the characteristics of activities using GIS.

American Time Use Survey (ATUS)

The ATUS provides a dataset conducted and maintained by the U.S. Bureau of Labor Statistics every year. The purpose of the survey is to record the respondents' activities, locations, and demographic information on a regular day from 4 AM to 4 AM of the next day (Diao et al. 2017).

The activities in the ATUS data are in a hierarchical tree structure with 3 tiers. The 1st tier consists of overall categories of activities, the 2nd tier consists of intermediate categories of activities, and the 3rd tier contains the most detailed activities.

The ATUS data have been used in many behavioral studies since the ATUS records detailed daily diaries for each respondent with activities, times, places, partners, etc. In addition, the ATUS provides the respondent's socioeconomic information, which supports behavior data analysis. Johnson et al. (2014) presented a statistical model for the behavior of residential occupants with the ATUS data. Diao et al. (2017) identified and classified occupant behavior with energy consumption outcomes. They used 8-17 activities from the 1st tier and 2nd tier ATUS activities. Aksanli et al. (2016) developed a residential energy modeling framework based on human activities to estimate the energy consumption in residential buildings. They used seven simplified activities: sleeping, personal grooming, cooking, cleaning, entertainment, working at home, and going to work, which are derived from the 1st tier activities in the ATUS. Unlike most current studies that use the 1st tier or 2nd tier activities, this study uses the 3rd tier activity list to provide a more detailed and realistic behavioral analysis.

Geographic Information System (GIS)

GIS has been used often in research. GIS is beneficial as a useful cognitive tool to analyze and gather spatial data with its visual interface, which can help experts from other areas understand the data easily (Fonseca and Schlueter 2015; Zhao et al. 2015). It allows the researchers and other stakeholders to quickly identify data patterns and outliers (Kolter and Ferreira Jr 2011). GIS is used not only as a tool to display data on the map, but also as a method to analyze data by geographical location.

Recently, building and construction fields have been actively employing GIS as a part of their research methods as well, since GIS can capture, store, analyze, manage, and present spatial or geographic data, including not only the energy or construction-related data but also the location (i.e., address, city, state) and physical properties of buildings (i.e., size, height) (Ma and Cheng 2016). Also, city-wide GIS databases have been available in many regions of the world and accessible to the general public (Reinhart and Davila 2016). GIS has a high potential to combine 3D models of buildings, energy simulations, and real-time databases at a large geographical scale. Most of the existing building and construction studies using GIS have focused more on building energy consumption, physical building properties, or demographic information of occupants. However, this study combines GIS with the Occupant Behavior Model, and uses spatial analysis (the grouping analysis with K-means clustering) to explain the similarities and differences in energy usage-related behaviors of residential occupants by state.

Occupant Behavior Prediction Model

The Occupant Behavior Prediction Model aims to predict occupant behaviors and identify routine activities (Mo 2018). The model incorporates the concept of habit with the components of activity frequency and context. Context includes time (start time, end time, duration), place, and situation (partner, weather, other circumstances). These components are derived from routine behavior studies to measure the strength of habit in occupant behavior (Mo et al. 2019). The predicted occupant activities, the identified routine and non-routine activities can be used for efficient building operation and control strategies, more effective interventions, or education on occupant energy usage-related behaviors (McCoy et al. 2018). In the previous study (Mo 2018), energy usage-related activities were selected from all of the 3rd tier activities in the ATUS. The selected activities were re-organized, and new codes were assigned to the energy usage-related activities. Table 1 explains the new codes for activities, energy types, and appliances associated with the activities.

Table 1. Activities and Associated Energy and Appliances

Code	Activity	Energy	Appliances (Electricity and Gas)
AA01	Washing, dressing, and grooming	E,W,G	Lighting, Shower, Hair dryer, Shaving
BB01	Interior cleaning	E	Lighting, Vacuum
BB02	Laundry	E,W,G	Lighting, Washer, Dryer
BB03	Food and drink preparation	E,W,G	Lighting, Oven, Stove, Toaster, Blender, Coffee machine, Cooker, etc.
BB04	Kitchen and food clean-up	E,W	Lighting, Dish washer
BB05	Heating and cooling	E,G	Lighting, HVAC
BB06	Gardening, ponds, pools, and hot tubs	W,G,E	Lighting
BB07	Care for animals and pets	E,W	Lighting
BB08	Vehicle repair and maintenance	E	Lighting, Repair tools
CD01	Physical care for children	E,W	Lighting
CD02	Physical care for/helping adults	E,W	Lighting
EF01	Work for job(s)/research/homework	E	Lighting, Computer
LL01	Television	E	Lighting, TV
LL02	Listening to/playing radio or music	E	Lighting, Computer, Music player, Radio
LL03	General computer use	E	Lighting, Computer

Note: E denotes electricity, W denotes water, G denotes gas

METHODS

Comparative Analysis for Energy Usage-Related Activities by Regions

Table 2. Main Routine Energy Usage-Related Activities

Code	Description
AA01	Washing, dressing, and grooming oneself
LL01	Watching TV
CD01	Physical care for children
BB03	Food and drink preparation
BB04	Kitchen and food clean-up

Table 3. Census Regions

Region (Code)	States
Northeast (R1)	CT, MA, ME, NH, NJ, NY, PA, RI, VT
Midwest (R2)	IA, IL, IN, KS, MI, MN, MO, ND, NE, OH, SD, WI
South (R3)	AL, AR, DC, DE, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV
West (R4)	AK, AZ, CA, CO, HI, ID, MT, NM, NV, OR, UT, WA, WY

To compare the energy usage-related activities in the ATUS data, a comparative analysis was performed. As defined by the Occupant Behavior Prediction Model (Mo 2018), the components of the activities were identified as *Frequency, Duration, Start Time, End Time*, and *Partner*. The previous studies (Mo 2018; Mo and Zhao 2021; Mo et al. 2020) evaluated the predictability of each energy usage-related activity and regarded the activities with higher predictability as more routine activities. Based on them, the top five most routine activities (Table 2) were compared by region (Table 3) to identify the geographical differences.

Mean and Mode. The mean values of the given conditions were compared for the numeric variables, including *Frequency, Duration, Start Time,* and *End Time*, and the mode values were compared for the categorical variable, *Partner*.

Analysis of Variance (ANOVA). The ANOVA provides a statistical test that generalizes the *t-test* to more than two groups, and it can be used to evaluate the statistical significance of differences among three or more group means. In this study, the ANOVA was used to evaluate the differences in energy usage-related behaviors among the four regions.

Spatial Analysis for Routine Energy Usage-Related Activities

Geographical visualization helps us understand the results of data analysis more easily and clearly, and geographical analysis considers the geographical distribution of the data. A more detailed geographical analysis was performed for the selected activities that showed significant differences by region in the comparative analysis. ArcGIS 10.1 by ESRI was used to identify if state-level geographical location affected the characteristics of the activity.

Comparison of Activities by States. The mean values of the numeric variables (*Frequency, Duration per act, Sum duration of an activity per day, Start Time,* and *End Time*) and the mode value of the categorical variable (*Partner*) of each activity were calculated for each state. Among the numeric variables, *Duration per act* is the duration of a single occurrence of an activity, and *Sum duration of an activity per day* is the total (sum) duration of multiple occurrences of activity by one person in a day.

The mean values and the mode value by the state were sorted and classified with the quantile method on the map with different colors. Quantile assigns the same number of data to each class, and the resulting map is suitable to explain the order or sequential comparison for linearly distributed data (ESRI 2018). In this study, five quantiles were used to display the data on the map. For example, to compare the total time spent on watching TV on a day across 50 states, the average values of the states were ordered, and each quantile had 20% of the data (10 states).

Grouping of Activities with K-means Clustering. The pattern of each activity was grouped by similar states using the Grouping Analysis of ArcGIS. The Grouping Analysis is a part of the Spatial Analysis in ArcGIS, and it uses K-means clustering. Since most of the features (*Frequency, Duration per act, Duration per day, Start Time, End Time*) were numeric data, and only *Partner* feature was categorical, K-means clustering was applicable for clustering activity data by states.

To determine a suitable number of K, the pseudo F-statistic was computed. The pseudo F-statistic is the ratio of between-cluster variance to within-cluster variance, which is explained as follows (Wilkinson et al. 2004), where K is the number of clusters at any step in the hierarchical clustering, and N is the number of instances, GSS is the between-group sum of squares, and WSS is the within-group sum of squares. Large values of pseudo F denote cohesive and separated clusters. Especially, peaks in the pseudo F statistic indicate greater cluster separation. For the Grouping Analysis in GIS, pseudo F static was run first, then the largest pseudo F value was used as the number of clusters in K-means clustering.

$$pseudo\ F = \frac{GSS\ /\ (K-1)}{WSS\ /\ (N-K)}$$

The result of the Grouping Analysis was displayed on a parallel box plot. The value ranges of the input features, *Frequency, Duration* (*Sum duration of an activity per day*), *Start Time, End Time,* and *Partner*, were standardized with z-transform to remove the unexpected weight effect from different variances of the features. Z-transform is explained as follows (Witten et al. 2016), where x is the actual value, μ is the mean of the feature, and σ is the standard deviation of the feature.

$$z = \frac{x-\mu}{\sigma}$$

RESULT

Comparative Analysis for Energy Usage-Related Activities

Table 4 summarizes the mean values (*Frequency, Duration per Act, Duration per Day, Start Time, End Time*) and mode values (*Partner*) of *Washing, dressing, and grooming* (AA01), *Food and drink preparation* (BB03), and *Watching TV* (LL01) by region.

Table 4. Average or Mode Values of Routine Behaviors by Region

Var.	Region	AA01	BB03	BB04	CD01	LL01
Frequency	R1	1.74	1.72	1.34	2.66	1.94
	R2	1.74	1.70	1.28	2.52	1.99
	R3	1.78	1.67	1.31	2.65	2.01
	R4	1.74	1.71	1.31	2.55	1.90
Duration/Act	R1	31.09	35.01	26.55	24.85	115.34
	R2	31.94	31.46	27.63	24.09	120.83
	R3	33.09	33.28	26.94	24.42	123.36
	R4	31.11	31.85	30.07	24.12	116.50
Duration/Day	R1	50.16	56.01	34.04	62.00	209.09
	R2	50.63	49.87	34.76	53.16	221.60
	R3	54.47	51.53	34.57	59.50	230.85
	R4	50.70	51.98	38.36	56.86	207.95
Start Time	R1	762.97	806.22	938.78	868.03	1026.86
	R2	732.07	803.18	905.89	894.20	1008.25
	R3	774.07	792.82	928.72	861.34	1005.78
	R4	761.52	784.08	914.14	922.73	1015.14
End Time	R1	779.15	839.76	963.83	892.88	1062.05
	R2	756.79	833.25	928.86	915.49	1054.79
	R3	796.02	824.52	950.03	883.79	1047.22
	R4	785.51	814.31	940.34	940.88	1071.83
Partner	R1	-1	1	1	2	1
	R2	-1	1	1	2	1
	R3	-1	1	1	2	1
	R4	-1	1	1	2	2

All are average values except for the values for Partner, which are mode values.

Table 5 summarizes the difference of activities by region, which was derived from the ANOVA. In general, *Duration per Act, Duration per Day, Start Time*, and *End Time* more affected the differences of the energy usage-related activities in different regions, and *Frequency* and *Partner* less affected them. Considering all the given six components, *Watching TV* (LL01), *Washing, dressing, and grooming* (AA01), and *Food and drink preparation* (BB03) were more different by region. The rest of the activities were not significantly different by region, which means that these activities were more similar in all regions.

Table 5. Difference in Activities by Region

Act.	Frequency		Duration/Act		Duration/Day		Start Time		End Time		Partner	
AA01	1.21	(0.30)	4.73	(0.00)*	8.10	(0.00)*	8.85	(0.00)*	7.72	(0.00)*	n/a	(n/a)
BB03	0.94	(0.42)	3.29	(0.02)*	3.30	(0.02)*	2.16	(0.09)	2.38	(0.07)	5.47	(0.00)*
BB04	1.05	(0.37)	2.24	(0.08)	1.95	(0.12)	1.78	(0.15)	1.70	(0.16)	1.30	(0.27)
CD01	0.53	(0.66)	0.11	(0.95)	1.86	(0.13)	4.47	(0.00)*	3.78	(0.01)*	0.63	(0.59)
LL01	4.25	(0.01)*	3.32	(0.02)*	8.54	(0.00)*	3.05	(0.03)*	2.68	(0.05)*	5.28	(0.00)*

*ANOVA: F-value (p-value), * denotes p-value < 0.05.*

Spatial Analysis for Routine Energy Usage-Related Activities

Among the routine activities in the previous step, *Watching TV* (LL01) was further compared by state using GIS since it showed the most difference by region. Also, occupants generally spent longer time for *Watching TV* than the other four activities (Table 4), which means that *Watching TV* more influences occupant behavior and energy consumption.

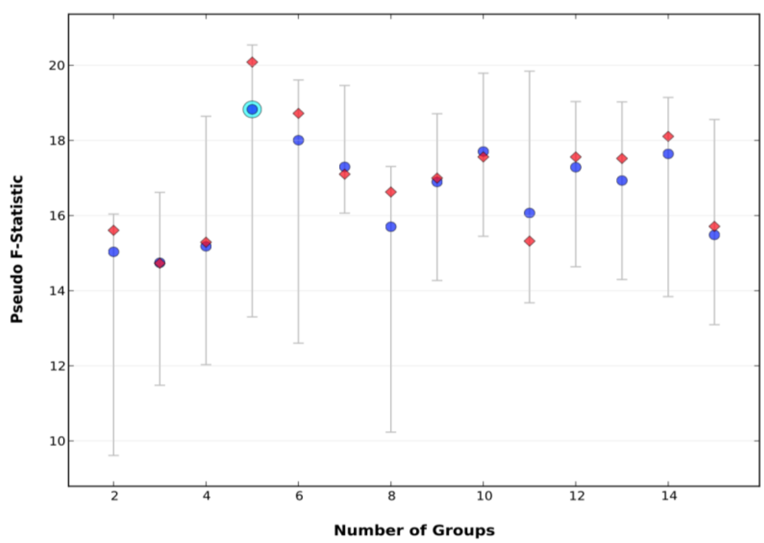

Figure 1. LL01 Number of K

Before running K-means clustering, pseudo F-statistic was calculated, and five was the most suitable number of clusters for *Watching TV* activity (Figure 1). Figure 2 explains the characteristics of the clusters. The boxplots represent the standardized values of the features, and dots stand for the mean values of the clusters. Figure 3 (1) displays the clusters and the states on the map.

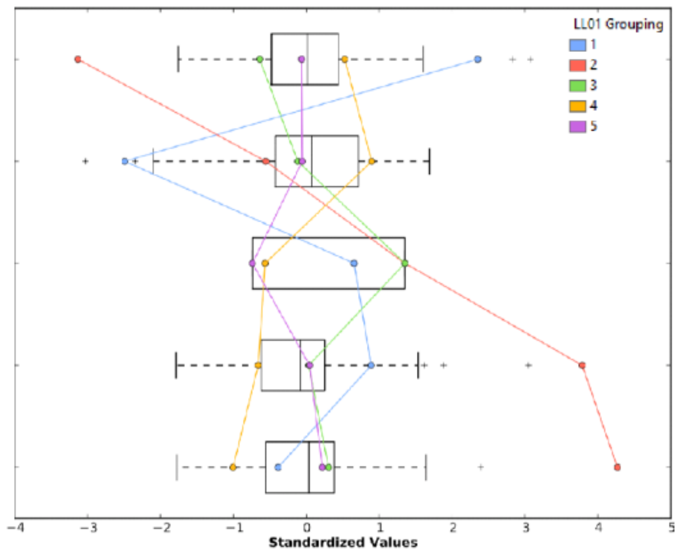

Figure 2. LL01 Group Analysis

- ***Cluster 1:*** The occupants in the states in Cluster 1 watch TV least frequently, but they spend the longest time a day compared to other clusters. They tend to watch TV more with family, and start to watch TV relatively early and end relatively late. Cluster 1 has Montana, Wyoming and Alaska, which are located in colder areas.
- ***Cluster 2:*** The occupants in the states in Cluster 2 watch TV relatively less frequently for the shortest time. They watch TV with family starting and ending latest. Cluster 2 has Hawaii.
- ***Cluster 3:*** The occupants in the states in Cluster 3 are in the middle for *Frequency* and *Duration* among the clusters. They watch TV with family and start relatively late and end relatively early. Cluster 3 has Oregon, Idaho, Utah, Arizona, Colorado, New Mexico, Kansas, Oklahoma, Missouri, Tennessee, North Dakota, South Dakota, Vermont, and Delaware.
- ***Cluster 4:*** The occupants in the states in Cluster 4 watch TV most frequently for a relatively long time. They tend to watch TV alone, starting and ending earliest among the clusters. Cluster 4 has Nevada, Wisconsin, Michigan, Ohio, Kentucky, West Virginia, Arkansas, Louisiana, Mississippi, Alabama, and Rhode Island.
- ***Cluster 5:*** The occupants in the states in Cluster 5 are in the middle for *Frequency, Duration, Start Time,* and *End Time*. They watch TV alone. Cluster 5 has Washington, California, Minnesota, Nebraska, Iowa, Illinois, Indiana, Texas, New Hampshire, Massachusetts, Maine, Connecticut, New York, New Jersey, Pennsylvania, Maryland, Virginia, North Carolina, South Carolina, Georgia, and Florida.

In Figure 3 (1), the geographical distribution of the clusters shows patterns with the cold area (Cluster 1), Hawaii (Cluster 2), central area (Cluster 3), central east area (Cluster 4), and coastal, central north and central south areas (Cluster 5). Cluster 2 had only Hawaii, and it explains that the geographical location of Hawaii strongly influences the different pattern of watching TV in this location compared to other states. Hawaii is remote from the other continental states, and has a unique climate, culture, economy, industry, lifestyle, etc. These differences might affect the difference in watching TV. Each state can consider the characteristic of the cluster where the state belongs for its energy control strategies and policy development.

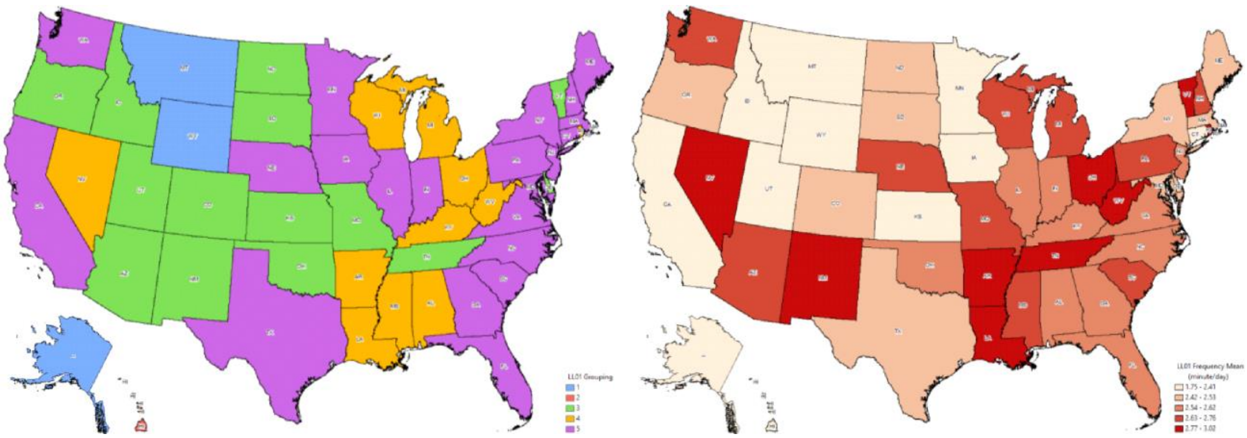

Figure 3. LL01 State Clusters (1) & Frequency by Quantiles (2)

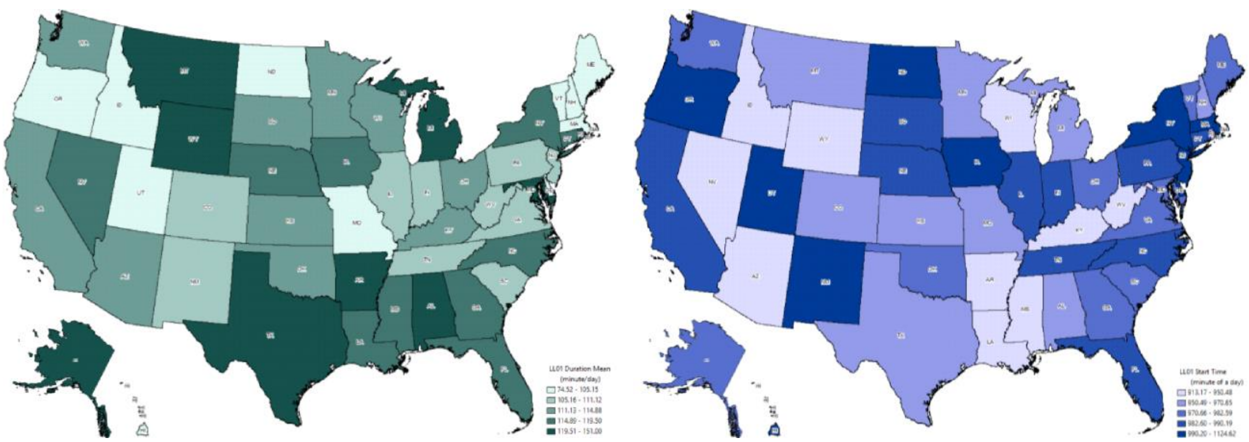

Figure 4. LL01 Duration (1) & Start Time (2) by Quantiles

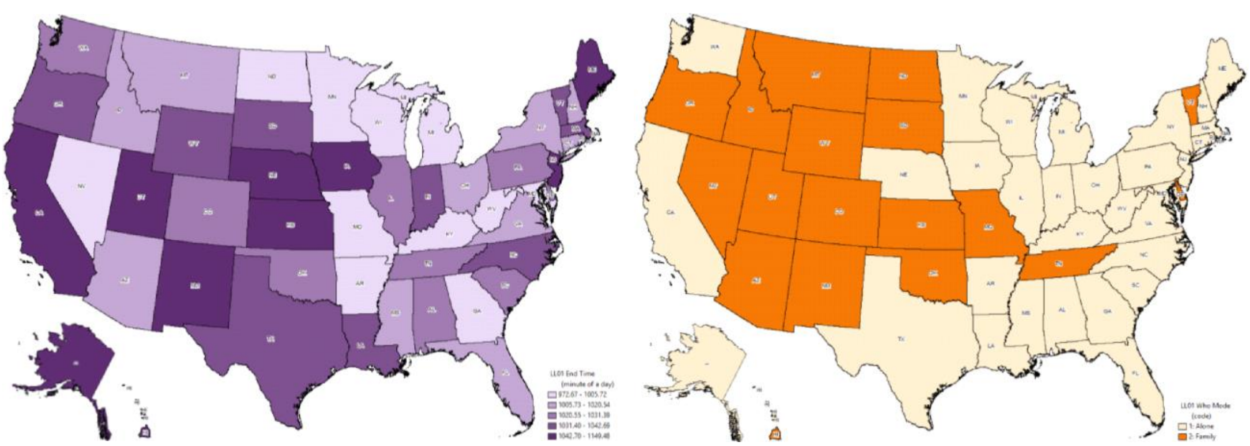

Figure 5. LL01 End Time (1) & Partner (2) by Quantiles

While Figure 3 (1) shows the result of the clustering analysis, which integrated the effects of *Frequency, Duration, Start Time, End Time,* and *Partner* together; Figures 3 (2), 4 (1), 4 (2), 5 (1) compare the mean values of individual components of watching TV by state with five quantiles. These maps more simply and directly explain the difference of each component by state. Darker colors indicate more *Frequency*, longer *Duration*, later *Start Time, and End Time*, and 20% of the states are indicated with the same color. Figure 5 (2) compares the mode value of *Partner*, and light color indicates watching TV more alone, and dark color indicates more with family. The map has only two-color levels since Partner has two values (1: Alone, 2: with Family). One of the possible explanations is that occupants in the agricultural states tend to watch TV with family, and they have more flexible time to watch TV compared to other states.

Frequency and *Duration* of *Watching TV* are more directly related to energy consumption, and *Start Time, End Time,* and *Partner* explain the lifestyle of occupants in the state. As shown in Figure 3 (2), regarding *Frequency*, occupants in Nevada, New Mexico, Arkansas, Louisiana, Tennessee, Ohio, West Virginia, and Vermont watch TV most frequently (2.77-3.02 times per day), while occupants in Alaska, Montana, Idaho, Wyoming, Utah, California, Kansas, Minnesota, Iowa, and Connecticut watch TV least frequently (1.75-2.41 times per day). In Figure 4 (1), regarding *Duration*, occupants in Alaska, Montana, Wyoming, Texas, Arkansas, Michigan, Alabama, and Maryland spend the longest time watching TV (119.51-151.00 minutes per day), while occupants in Oregon, Idaho, Utah, North Dakota, Missouri, Maine, Vermont, New Hampshire, Massachusetts spend the least time watching TV (74.52-105.15 minutes per day).

The spatial analysis efficiently explains the characteristic of energy usage-related activities, and helps to compare the activities considering the geographical locations of the occupants. It also enables researchers to effectively consider environmental, social, technological factors based on location.

CONCLUSION

In this study, the routine energy usage-related activities were analyzed by region and state. The findings include that duration, start time and end time have more influences on the differences in energy usage-related activities, and watching TV, washing and grooming, and cooking and food preparation are more different by geographical location. The differences in activities could be more efficiently explained using GIS analysis. The geographical locations enable to associated other diverse factors more effectively. The routine activities of occupants can be influenced not only by internal factors of the occupants, such as age, gender, job, income, education, number of family members, etc. but also by external factors including climate, economy, industry, policies, building technology and more of the location. The finding also shows the routine activities vary by different geographic locations, although they are persistent over time in the same location. GIS analysis effectively connects the geographical context of the information with these factors.

The result can be used to provide more reliable information regarding energy and behavior to the occupants in residential buildings. Also, the result can be applied to the new energy and behavior strategies and policies about residential building energy plans. In addition, the geographical comparison using GIS and the grouping analysis can be used to develop more efficient strategies for different locations of the residential buildings.

ACKNOWLEDGEMENT

This work is in part supported by National Science Foundation (NSF) No.2046374. Any opinions and conclusions in this work are of the authors and do not necessarily reflect the view of NSF.

REFERENCES

Aksanli, B., Akyurek, A. S., and Rosing, T. S. (2016). "User behavior modeling for estimating residential energy consumption." *Smart City 360°*, Springer, 348-361.

Diao, L., Sun, Y., Chen, Z., and Chen, J. (2017). "Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation." *Energy and Buildings*.

ESRI. (2018). "Data classification methods." ArcGIS Pro, <http://pro.arcgis.com/en/pro-app/help/mapping/layer-properties/data-classification-methods.htm>.

Fonseca, J. A., and Schlueter, A. (2015). "Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts." *Applied Energy*, 142, 247-265.

Johnson, B. J., Starke, M. R., Abdelaziz, O. A., Jackson, R. K., and Tolbert, L. M. (2014). "A method for modeling household occupant behavior to simulate residential energy consumption." *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES*, IEEE, 1-5.

Kolter, J. Z., and Ferreira, J., Jr. (2011). *A Large-Scale Study on Predicting and Contextualizing Building Energy Usage*. AAAI.

Ma, J., and Cheng, J. C. (2016). "Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology." *Applied energy*, 183, 182-192.

McCoy, A. P., Zhao, D., Ladipo, T., Agee, P., and Mo, Y. (2018). "Comparison of green home energy performance between simulation and observation: A case of Virginia, U.S." *Journal of Green Building*, 13(3), 70-88.

Mo, Y. (2018). *Occupant Behavior Prediction Model Based on Energy Consumption Using Machine Learning Approaches*. Doctoral Dissertation, Michigan State University, East Lansing, MI.

Mo, Y., and Zhao, D. (2021). "Effective Factors for Residential Building Energy Modeling using Feature Engineering." *Journal of Building Engineering*, 102891.

Mo, Y., Zhao, D., and Sorce, J. (2020). "Daily Occupant Behavior Pattern Affecting Energy Use in Residential Buildings." *Construction Research Congress 2020: Computer Applications*, American Society of Civil Engineers Reston, VA, 518-526.

Mo, Y., Zhao, D., and Syal, M. (2019). "Effective Features to Predict Residential Energy Consumption Using Machine Learning." *Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience*, American Society of Civil Engineers Reston, VA, 284-291.

Reinhart, C. F., and Davila, C. C. (2016). "Urban building energy modeling–A review of a nascent field." *Building and Environment*, 97, 196-202.

Wilkinson, L., Engelman, L., Corter, J., and Coward, M. (2004). "Cluster Analysis." *Systat* University of Illinois Urbana-Champaign, 65-124.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

Zhao, D., McCoy, A. P., Du, J., Agee, P., and Lu, Y. (2017). "Interaction effects of building technology and resident behavior on energy consumption in residential buildings." *Energy and Buildings*, 134, 223-233.

Zhao, D., McCoy, A. P., and Smoke, J. (2015). "Resilient Built Environment: New Framework for Assessing the Residential Construction Market." *Journal of Architectural Engineering*, B4015004.