Development of Occupants' Behavior Model in Urban Scale Using Dynamic Time Warping and Particle Swarm Optimization Algorithms Based on National Lifetime Survey

Hengxuan Wang
Graduates School of Human-Environment Studies
Kyushu University
Fukuoka, Japan
e-mail: whx7174615@gmail.com

Daisuke Sumiyoshi
Faculty of Human-Environment Studies
Kyushu University
Fukuoka, Japan
e-mail: sumiyoshi@arch.kyushu-u.ac.jp

***Abstract*—For the target of energy demand estimation of residential buildings in urban scale, occupants' behavior model has been paid much attention. In this paper, a new model for simulating occupants' behavior schedules in urban scale has been proposed using only public stochastic data (national lifetime survey) combined with Dynamic Time Warping and Particle Swarm Optimization algorithms. We use this proposed model to simulate the working-male's behavior schedules with 5-mintues interval in resting day as an example. The simulated results - percentages of occupants adopt the given behavior at specific moments are calculated and compared with public stochastic data to verify the accuracy. Compared with existing models, the proposed model is more efficient and accurate. We believe this model could be useful for building energy demand estimation in urban scale combined with appliance operation possibility based on occupants' behaviors.**

Keywords-occupants' behavior model; energy demand of residential buildings; public stochastic data; particle swarm optimization; dynamic time warping.

I. INTRODUCTION

A. Background

To get the target of decarbonized society in 2050 [1], the Japanese government is promoting the introduction of decentralized renewable energy devices in urban area to reduce carbon emissions. But without a suitable introduction plan, the surplus electricity generated from excessive devices would disturb the balance between the supply and demand of power system or failing to get decarbonization target because of insufficient devices. Therefore, in building sector, it is essential to develop the decentralized energy introduction plan based on the energy demand of buildings in urban area.

The renewable energy is limited to natural condition (e.g., solar energy) and outpower changes dramatically over time. Therefore, the energy demand of buildings should be estimated with high temporal resolution. Non-residential buildings (e.g., office) have temporal characteristics of energy demand because of fixed schedule of users. However, the energy demand of residential building is decided by appliances' operation, which is influenced by the behavior of the occupants with significantly personal characteristics. In previous studies about energy demand estimation for residential buildings, the behavior schedules of occupants had been set to several cases. This assumption would significantly affect the accuracy of results. The reason is that even the same type occupants in urban scale would have numerous kind behaviors at the same time, but there are only a few cases in these few schedules that would overlay the peak or trough energy demand amount. Thereby, the demand results and the amount of renewable energy devices need to be introduced would be a departure from reality. Thus, a method to simulate the occupants' behavior schedules in urban scale is very essential for the plan of introduction of decentralized renewable energy in urban scale.

B. Related Work

There is much previous research about the occupants' behavior model in urban scale [2]. The models could be divided into two types based on whether to use dataset called Time Use Data (*TUD*), which describe occupants' behavior by time.

For the first type without *TUD*, in [3], a model was developed using only public stochastic data of *TUD* include mean and standard deviation of behaviors' duration time in a day and percentages of occupants adopt the behavior at special moments by 15-mintues interval of a day. They firstly selected the behaviors according to probabilities and arranged their total duration time into 24 hours. Next, they placed the first behavior into the slot in timeline according to random number and placed the next behavior into the end of previous behavior one by one. As one merit, this method could generate the occupants' behavior schedules with only public stochastic data. But the accuracy of the simulation results was greatly influenced by the first behavior's inserted slot, which was decided randomly. Also, the results had not been validated.

For the type of models using *TUD*, in [4], they developed a model using the Markov Chain, which is a stochastic model to determine the transition of behavior from another only depend on the condition at the previous time step. They collected the *TUD* from a great number of households and analysis the transition probability between behaviors. But the behavior items were limited in at room or not.

In [5], they proposed a Markov Chain model and expanded analysis of the number of behavior items. They simulated the

 ISBN: 978-1-68558-054-4

household's members independently. These Markov Chain models considered and simulated the transitions probability between the behaviors precisely, but the accuracy of behavior duration time was dependent on the timing and number of behavior transitions. This could be a weakness for simulating the occupants' behavior schedules. In [6], they developed a occupants' behavior model dealing with above problems. They divided behaviors into routine and non-routine and considered them separately. The behaviors' duration time and transition probabilities between them were acquired by analyzing the *TUD* from the national time-use survey conducted by Statistics Japan in 2006 and been utilized for placing the behaviors into the timeline. They firstly placed the routine behaviors (including sleeping, commuting to work & school, dining and bathing) into timeline, then selected the non-routine behaviors according to the probabilities and placed them in the gap between the routine behaviors until all gaps had been filled. They improved the model in [7] considering the interaction among household members (e.g., household members always have dining together at one time and bathing one by one) and time-dependent characteristics of the specific behaviors (e.g., for a single person, the personal washing often happens immediately after waking up or breakfast, but it's not shown in *TUD* because it was originated from a wide range of people.). In [8], they explored several machine learning methods to pre-process the *TUD* to improve the accuracy of behavior model. Although the duration time and transition probabilities of behaviors were detailed considered in their model, predetermining the number of behavior occurrences with a subjective assumption was made. (e.g., three meals over a day, one sleeping in the evening with long period), but according to the public stochastic data, there are also sleeping at the daytime for many type people), this might be a weakness of their model for ignoring this specifical cases. On the other hand, raw *TUD* are required to make this kind of model while only public stochastic data are available in many countries.

As mentioned above, until now there are many developed occupants' behavior models in urban scale with own strengths and weaknesses. But there is still no precise occupants' behavior model that do not require prior analysis of large amount of raw *TUD*, pre-classification of behaviors according to routinely or not and predetermined number of occurrences with subjective assumption.

C. Purpose of the paper

In this paper, a new occupants' behavior model using only public stochastic data without raw *TUD* has been proposed. Compared with existing models in previous research, this model could ensure the accuracy and efficiency.

Section II introduces the detailed procedures of the proposed model. Section III corrects the simulation processes based on the simulation results. In Section IV, the final simulation result of working-male in resting day is shown. In Section V, the conclusions about model's features and weaknesses are introduced. Based on that, the directions of improving model in future are also introduced.

II. PROPOSED BEHAVIOR MODEL

A. Parameters of purposed occupants' behavior model

The proposed model simulates the occupants' behavior schedules with 5-min interval based on the public stochastic data called National Lifetime Survey in 2020 from Japan Broadcast Institution (*NHK*) [9]. It should be noted that target of simulating behavior schedules is to estimate energy demand of residential building, so the behaviors that have no relationship with energy demand in residential building (e.g., working outside, commuting to work, school) have not been simulated in this paper. This assumption, which is one of the differences between the previous studies, can greatly simplify the model. Table I shows the classification result of behaviors on public stochastic data. These behaviors have been simplified into 24 types and divided into interior and exterior. According to whether using appliances, the interior behaviors are further divided into two categories.

To make the model, the public stochastic data would be utilized include:

- ***PM***: probabilities of adopting given behavior by 15-minutes interval (the data has been processed into 5-minutes interval by liner interpolation).
- ***PA***: probabilities of adopting given behavior over a day.
- ***MTB***: average duration time of adopting given behavior.
- ***SDTB***: standard deviation of duration time of given behavior.

Some samples of public stochastic data are shown in Table II.

During the process of simulation, a blank timeline with 288 time slots (time of a day with 5-minutes interval) is generated firstly and prepares for filling up with behaviors

TABLE I. CLASSIFICATION OF BEHAVIORS

Interior Behavior (lighting & HVAC used)		**Exterior Behavior**
Appliance used	***Non-appliance used***	
eating	children care	shopping
washing	leisure	conversation personal relationships
sleeping	reading newspaper	work
hobbies, entertainment and culture (with Internet)	reading magazines comics	leisure and exercise
hobbies, entertainment and culture (without Internet)		class and lecture
cooking, cleaning, laundry		commuting
radio		sporting
household chores		

TABLE II. SAMPLE OF PUBILC STOCHASTIC DATA OF WORK-MALE IN SUNDAY AND TARGET BEHAVIOR -SLEEPING

Behavior	***PA***	***MTB***	***SDTB***
sleeping	99.20%	8:25	2:07
eating	97.60%	1:38	0:52
washing	96.00%	1:04	0:34

Time	***PM***
0:00	70.20%
0:05	71.27%
0:10	72.33%

separately. The given behavior's occurrences will span corresponding time slots in the timeline depend on its duration time length. The detail steps are explained as:

1) Iterates over all given behaviors in order of the ***PA***'s values and determines whether to adopt based on its ***PA***.
2) Once the given behavior has been adopted, the duration time (***TB***) is determined according to the Gaussian Distribution defined by ***MTB*** and ***SDTB***.
3) To insert these given behaviors into the timeline, it is critical to decide several parameters' solution of the given behavior including:
   a) ***n***: number of behavior occurrences. (***n*** =1~4 randomly)
   b) *sm*: start moment of each behavior occurrence. ***SMN***: $\left[sm_1, sm_2, \cdots sm_n\right]$
   c) *pn*: probability of each number of behavior occurrences. (e.g., pn_2: probability of behavior occurring twice) ***PNN***: $\left[pn_1, pn_2, \cdots pn_n\right]$
   d) *pt*: each occurrence's duration time as percentage of ***TB*** in a large number of schedules. ***PTN:*** $\left[pt_1, pt_2, \cdots, pt_n\right]$ (e.g., Figure 1 shows the difference between ***PTN₁***: [1/3,1/3,1/3] and ***PTN₂***: [1/6,1/3,1/2]

B. Start moments of behavior occurances

It is necessary to decide ***SMN*** 's solution to determine the positions of timeline where the behavior occurrences are going to be inserted.

Figure 2 shows the process of deciding the start moment of each behavior occurrence by the cumulative distribution of ***PM***. In detail, one day is divided into ***n*** time regions, which have the same sum of ***PM***. The start moment of splitting time region is generated randomly (e.g., 6:30). It is assumed that object behavior occurs once in each time region. Based on that assumption, the cumulative distribution function of ***PM*** in each time region has been calculated to determine the start moment of each occurrence.

C. Dynamic Time Warping

Different from ***SMN***, it is impossible to get ***PNN***&***PTN*** solution based on the existing public stochastic data merely.

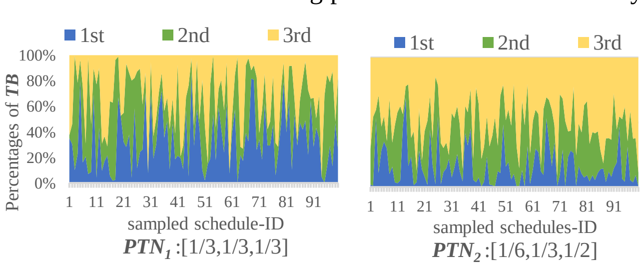

Figure 1. Percentages of each occurrence's duration time in sampled schedules

It is necessary to introduce parameter optimization method to obtain the optimal ***PNN***&***PTN*** solution.

To verify the fitness of ***PNN***&***PTN*** candidate solution, we introduce the objective function to compare ***PM*** and probability of adopting given behavior at 5-min interval, which is calculated by schedules generated using ***PNN***&***PTN*** candidate solution (***PM'***). For ***PM*** and ***PM'*** are both time series data, Dynamic Time Warping (*DTW*) introduced in [10] is used as objective function to measure their similarity. *DTW* of ***PM*** and ***PM'*** is calculated by (1):

$$DTW(\boldsymbol{PM}, \boldsymbol{PM'}) = \min\sqrt{\sum(x_i - y_j)^2} \quad (i,j) \in L \tag{1}$$

$$\boldsymbol{PM} = \left[x_1, x_2, \cdots, x_{287}\right],\ \boldsymbol{PM'} = \left[y_1, y_2, \cdots, y_{287}\right]$$

The list of index pairs $L = \left[l_0, l_1, \cdots, l_{287}\right]$ shows the matching pairs of the elements of ***PM*** and ***PM'*** (e.g., $l_k = (i_k, j_k)$ shows the x_{i_k} and y_{j_k} would be matched) that satisfies the following properties are shown in (2) (3) (4):

$$0 \le i_k, j_k \le 287 \tag{2}$$

$$l_0 = (0,0),\ l_{287} = (287,287) \tag{3}$$

$$l = (i_k - 1, j_k)\ or\ (i_k, j_k - 1)\ or\ (i_k - 1, j_k - 1) \tag{4}$$

Different from the traditional matching method, which would match ***PM*** and ***PM'*** at the same index pairs ((x_1, y_1), (x_2, y_2), …, (x_{287}, y_{287})). In *DTW*, based on the above properties, there is a large number (*T*) of possible matching solutions as candidates, which is shown in (5):

$$T\left[l_0, l_1, \cdots, l_{287}\right] = \begin{bmatrix} (x_0, y_0) & (x_1, y_0) & (x_1, y_1) & \cdots & (x_{287}, y_{287}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (x_0, y_0) & (x_0, y_1) & (x_0, y_2) & \cdots & (x_{287}, y_{287}) \end{bmatrix} \tag{5}$$

all matching solutions' distances between ***PM*** and ***PM'*** would be compared and the smallest one would be called *DTW*. By calculating the *DTW* obtained from different ***PNN***&***PTN*** candidate solutions, the most suitable

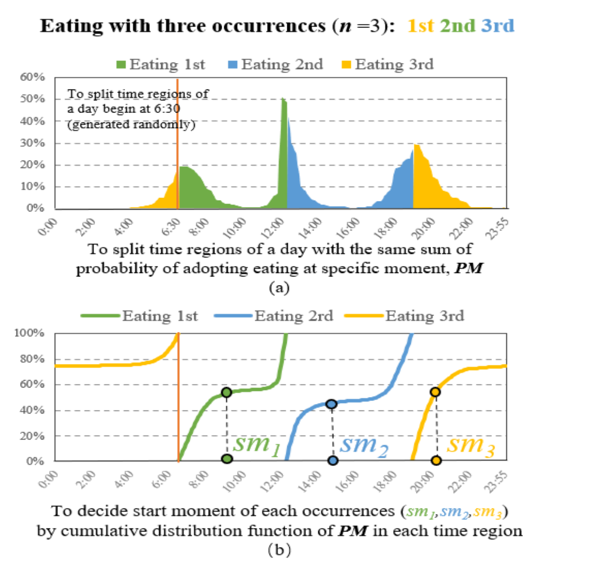

Figure 2. Processes of determining start moment of each occurrence of eating

PNN&PTN solution would be decided with the minimal *DTW*.

D. Patricle Swarm Optimization

As mentioned above, the best ***PNN&PTN*** solution can be found by finding the minimal *DTW*. In this paper, we use Particle Swarm Optimization (*PSO*) algorithm to find the minimal *DTW*. *PSO* is an evolutionary algorithm introduced in [11] that could optimize a problem by iteratively trying to improve a candidate parameters' solution to get the better position in a D-dimensional space (D is the number of parameters).

In the process of the *PSO* algorithm, firstly, a large number of particles have been generated and each particle is a candidate solution of ***PNN&PTN*** with different *DTW* result. At 1st iteration, particle' initial position (p_1) and velocity (v_1) are randomly generated. p_1 means ***PNN&PTN*** solution and v_1 means the distance between p_1 and p_2 (position at 2nd iteration) as showed in (6). These particles make up a cloud that covers the entire space, then the *DTW* of all particles are calculated to decide their fitness. Based on fitness values, the globally best particle position (pg_1) and locally best particle position (pl_1) are determined. As showed in (7), according to pg_1, pl_1 and p_1, v_1 would be updated to v_2, which would continue to update p_2 to p_3. With the iteration advancing, the cloud contracts gradually and performs the exploration for best ***PNN&PTN*** solution with minimal *DTW*.

$$p_{k+1} = p_k + v_k \tag{6}$$

$$v_{k+1} = wv_k + \varphi_1(pg_k - p_k) + \varphi_2(pl_k + p_k) \tag{7}$$

k :	k^{th} iteration	pl_k :	locally best particle's position at kth iteration
w:	inertia weight	φ_1, φ_2 :	$\varphi_1 = c_1 r_1$, $\varphi_2 = c_2 r_2$
v_k :	particle's velocity at kth iteration	r_1, r_2 :	random numbers in the range [0,1]
p_k :	particle's position at kth iteration	c_1, c_2 :	$c_1 = c_2 = 2$
pg_k :	globally best particle's position at kth iteration		

E. Process of proposed model

Figure 3 shows the proposed model's all processes for simulating the behavior schedules. Using this model, 1000 behavior schedules have been generated and evaluated.

III. Correction of Simulation Process

According to the schedule results, there are two significant errors include:

1) Figure 4 shows the results of sleeping are inaccurate thoroughly.
2) Figure 5 shows the delaying of start moments of ***PM'*** compared with ***PM***.

The above errors would be dealt with as follows:

A. Correction of sleep simulation process

The error 1) can be attributed to the inaccurate determination of start moments of sleeping. Different from other behaviors, people always have a long period sleeping in the evening and add a short period sleeping during the daytime. For this type of behavior with a clear temporal characteristic,

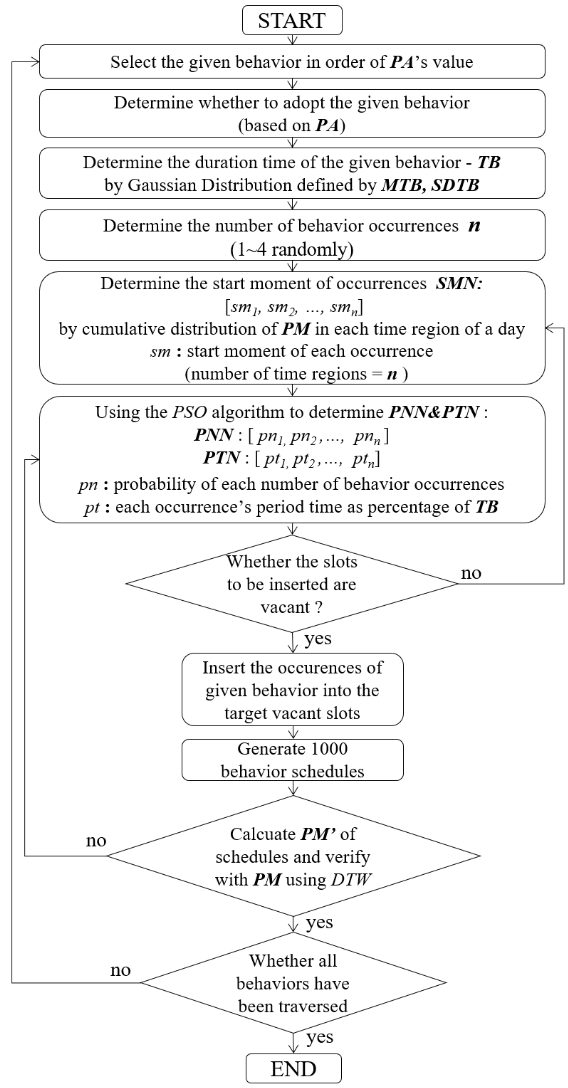

Figure 3. Simulation processes of proposed model

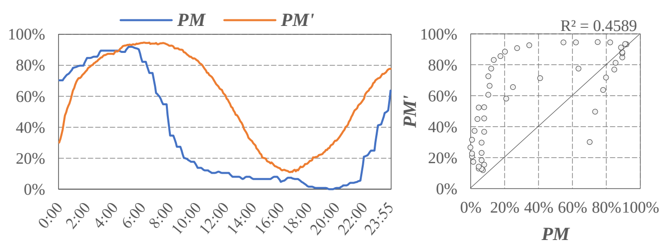

Figure 4. Comparation of simulated stochastic data, ***PM'*** and public stochastic data, ***PM*** (sleeping)

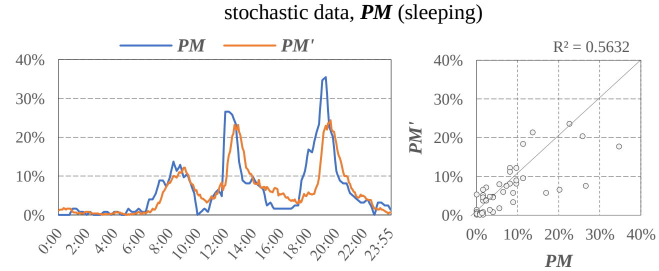

Figure 5. Comparation of simulated stochastic data, ***PM'*** and public stochastic data, ***PM*** (eating)

the method deciding behavior start moments based on sum of ***PM*** is not suitable any longer. To solve this problem, we revise the model to simulate sleeping behavior in the following process:

a) The number of sleeping occurrence (***n***) is set to 1~2. If sleeping occurs once, it occurs at night; if sleep occurs twice, the first and longer sleeping occurs in the evening and second one occurs at daytime.

b) For sleeping in the evening, people wake up at a more concentrated time than when they fall asleep. Therefore, we use the moments of waking up (ending of sleeping) to decide the position of sleeping in timeline where being inserted into.

c) The range of end moments of first sleeping in the evening is set as 0:00-12:00, the range of start moments of second sleep is set as 12:00-18:00. The specific moments in the range are searched by *PSO* method too.

To sum up, the parameters of sleeping for *PSO* method are reset showed as (8):

$$\boldsymbol{SMN} = [sm_1, sm_2] \quad \boldsymbol{PNN} = [pn_1, pn_2] \quad \boldsymbol{PTN} = [pt_1, pt_2] \quad (8)$$

After the calculation by *PSO*, Figure 6 shows the results of sleeping's ***PM*** and ***PM'*** by this revised process, which is better than original one.

B. Correction of start moment

For the error 2), the reason being considered is that the decision of start moment based on cumulative distribution function of ***PM*** always drop behind actual situation. To solve this issue, the new parameter *ad* is introduced to adjust the $\boldsymbol{SMN}: [sm_1 + ad, sm_2 + ad, \cdots, sm_n + ad]$. The decision of *ad* is also calculated by *PSO* method. Therefore, the solution of ***PNN***&***PTN***, *ad* would be decided together by minimal *DTW*. Figure 7 shows the simulation results of ***PM*** and ***PM'*** after the adjustment of behavior start moments and it demonstrates higher accuracy than before.

IV. SIMULATION RESULTS

Table III shows the calculation results of behavior washing's parameter solutions by *PSO* algorithm of working-male in resting day. In Figure 8, the probability distribution of all target behaviors is shown. The result shows that ***PM'*** agreed well with ***PM*** and it confirms our model's accuracy.

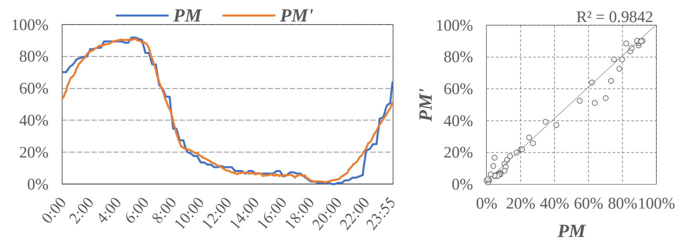

Figure 6. Comparation of simulated stochastic data, ***PM'*** after revising and public stochastic data, ***PM*** (sleeping)

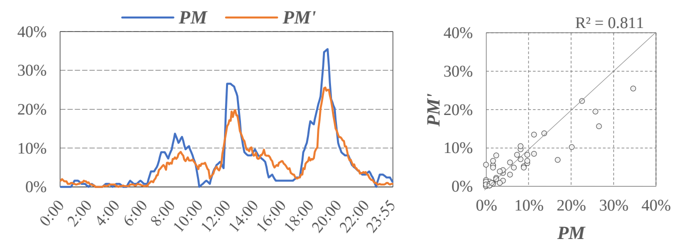

Figure 7. Comparation of simulated stochastic data, ***PM'*** after revising and public stochastic data, ***PM*** (eating)

TABLE III. PARAMETER SOLUTIONS OF BEHAVIOR WASHING CALCULATED BY *PSO* ALGORITHM

Behavior	***Parameter Solutions – PNN, PTN,*** *ad* $[pn_1, pn_2, pn_3, pn_4]$ $[pt_1, pt_2, pt_3, pt_4]$ $[ad]$
washing	***PNN*** = [0.9%, 18.7%, 31.8%, 48.6%] ***PTN*** = [21.7%, 20.6%, 18.2%, 39.5%] *ad* = [3]

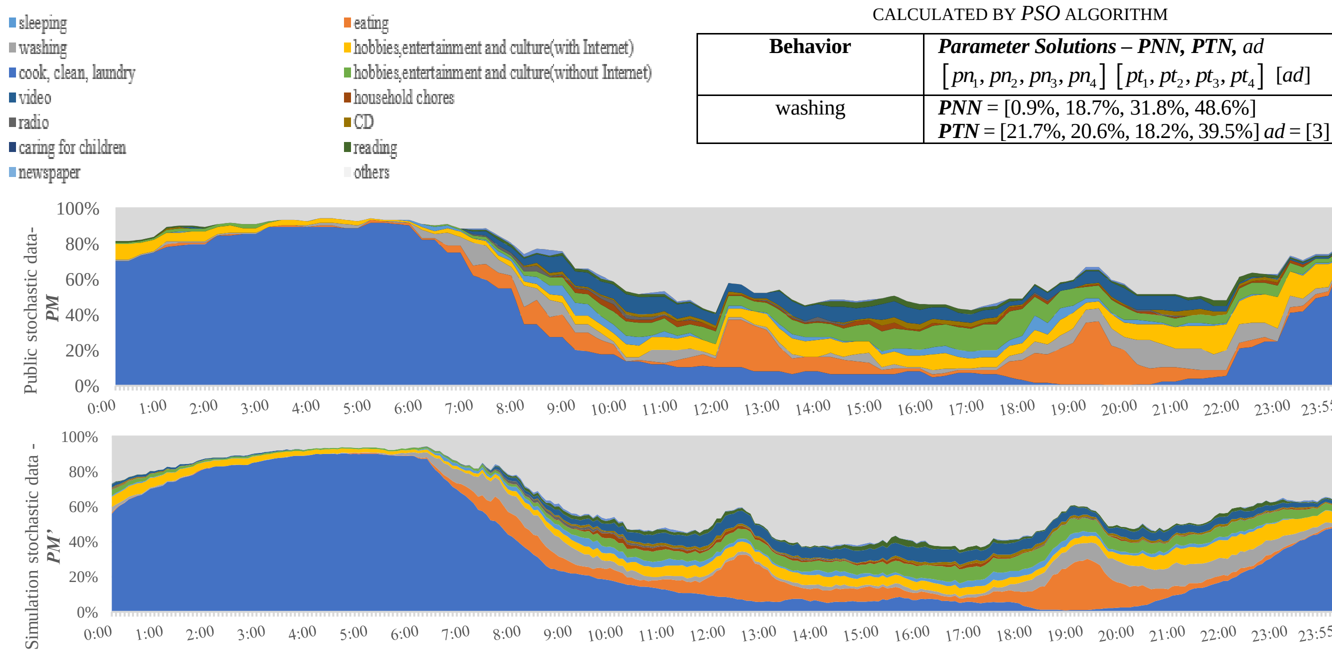

Figure 8. Comparation of simulated stochastic data, ***PM'*** and public stochastic data, ***PM*** of all behaviors by work-male in resting day. (Each area shows the probability that a behavior is undertaken at each time of day)

V. CONCLUSION AND FUTURE WORK

This paper proposes a model based on public stochastic data to generate occupants' behavior schedules at home. It could be used in energy demand estimation of residential buildings in urban scale. More than just simulation for behavior schedules based on past public stochastic data, it might be useful for assuming the future people behavior change by altering the inputting public stochastic data (e.g., the working time at home increased because of covid-19). Compared with existing behavior model's research, the proposed model has the following features:

- Generating occupants' behavior schedules based on public stochastic data only. Without statistical analysis of large amounts of raw *TUD*, which is not available in many countries, making the behavior model simpler and more efficient.
- No classifying the behaviors or setting the specific number and duration time of behavior occurrences. This feature could exclude the errors from subjective assumptions.
- Utilizing the *PSO* and *DTW* algorithm to search the suitable number of behavior occurrences and percentages of occurrences' duration time. It would make the simulation results match the public stochastic data as closely as possible.
- Deciding the start moments of behavior based on cumulative distribution of public stochastic data. As the simulation results do not agree with the public stochastic data, the start moments calculated by the above method have been corrected using *PSO* algorithm.

It should be noted that by *PSO* algorithm merely, the ***SMN*** could be determined without using cumulative distribution of ***PM***. But with the assistance of cumulative distribution, the *PSO* algorithm could narrow the search range and get solution quickly. Using this revised model, the working-male's behavior schedules in resting day have been generated. The result shows that our model has a good accuracy. But there are also several drawbacks:

a) During some time interval s (e.g., at 12:00~12:30 and 22:00~22:30, the public stochastic data - ***PM*** of eating, sleeping increase rapidly, but the simulation results - ***PM'*** fail to reflect such phenomenon.

b) No consideration of interaction between behaviors (e.g., people are likely to wash themselves when they wake up, but behavior transition between sleeping and washing can't be simulated in a single schedule).

c) Fail to consider the interaction between the family members (e.g., having a meal together). This interaction is import for residential building energy demand estimation.

In future, we are going to deal with these drawbacks to improve the behavior model. About drawback *a)*, more in depth analysis of public stochastic data especially in specific time interval will be done so that different weights will be given during these time intervals in simulation process. For *b)*, which had been raised in much previous research, analysis the raw *TUD* to get the results of behavior transition probabilities is a feasible option. And about last *c)* as mentioned in [7], when several kinds of schedules from different people in a family are required for residential building's electricity demand estimation, it could be a solution to choose schedules from generated schedule database, which have meals at the same time and bathing in sequence. Also, it should be noted that the behavior schedules could not be used directly in the energy demand estimation model without appliance operation possibility based on behaviors. More work about the relationship between behavior and appliance operation is also necessary in future.

REFERENCES

[1] Ministry of the Environment, Japan, Outlines of Japan's Long-term Strategy under the Paris Agreement, 2019.

[2] G. Happle, J. A. Fonseca, and A. Schlueter, "A review on occupant behavior in urban building energy models," Energy & Buildings, vol. 174, pp. 276-292, Sept. 2018, doi: 10.1016/j.enbuild.2018.06.030.

[3] J. Tanimoto, A. Hagishima, and H. Sagara, "Validation of methodology for utility demand prediction considering actual variations in inhabitant behavior schedules," Journal of Building Performance Simulation, vol. 1, no. 1, pp. 31-42, Mar. 2008, doi:10.1080/19401490701868471.

[4] I. Richardson, M. Thomson, and D. Infield, "A high-resolution domestic building occupancy model for energy demand simulations," Energy and Building, vol. 40, no. 8, pp. 1560-1566, Feb. 2018, doi:10.1016/j.enbuild.2008.02.006.

[5] J. Widén, et al., "Constructing Load Profiles for Household Electricity and Hot Water from Time-Use Data – Modelling Approach and Validation," Energy and Building, vol. 41, no. 7, pp. 753-768, Mar. 2009, doi:10.1016/j.enbuild.2009.02.013.

[6] Y. Yamaguchi and Y. Shimoda, "Evaluation of behavior model of occupants in home based on Japanese national time use survey," Proceedings of Building Simulation 2015: 14th Conference of IBPSA, Dec. 2015, pp. 2897-2904, doi: 10.26868/25222708.2015.2520.

[7] Y. Yamaguchi and Y. Shimoda, "A stochastic model to predict occupants' activities at home for community-/urban-scale energy demand modelling," Journal of Building Performance Simulation, vol. 10, no. 5-6, pp. 565-581, Jun. 2017, doi: 10.1080/19401493.2017.1336255.

[8] Y. Li, Y. Yamaguchi, and Y. Shimoda, "Impact of the pre-simulation process of occupant behavior modelling for residential energy demand simulations," Journal of Building Performance Simulation, vol. 15, no. 3, pp. 287-306, Mar. 2021, doi:10.1080/19401493.2021.2022759.

[9] Japan Broadcast Institution (*NHK*), National Lifetime Survey in 2020, Japan.

[10] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE transactions on acoustics, speech, and signal processing, vol. 26, no. 1, pp. 43-49, Feb. 1978, doi:10.1109/TASSP.1978.1163055.

[11] J. Kennedy and R. Eberhart, "Particle swarm optimization", Proceedings of ICNN'95 – International Conference on Neural Networks, Nov. 1995, pp. 1942-1948, doi: 10.1109/ICNN.1995.488968.