Application of Household Urban Micro-Simulation (HUMS) in Cities of Different Population Sizes and Comparison Between Model Parameter Characteristics

Munkhbat BATZAYA[1], Nao SUGIKI[2], Atsushi SUZUKI[3], Tomohiko SAKATA[4], Kojiro MATSUO[5]

[1]Student member of JSCE, Department of Architecture and Civil engineering, Toyohashi University of Technology
(Hibarigaoka 1-1, Tenpaku-cho, Toyohashi, Aichi 441-8150, Japan)
E-mail: batzaya.munkhbat.ws@tut.jp
[2]Member of JSCE, Department of Architecture and Civil engineering, Toyohashi University of Technology
(Hibarigaoka 1-1, Tenpaku-cho, Toyohashi, Aichi 441-8150, Japan)
E-mail: sugiki@ace.tut.ac.jp
[3]Member of JSCE, Department of Social Infrastructure Design Engineering, Meijo University
(Shiogamaguchi 1-501, Tenpaku-ku, Nagoya, Aichi 468-8502, Japan)
E-mail: atsuzuki@meijo-u.ac.jp
[4]Member of JSCE, Housing Urban Studies Group, Building Research Institute
(Tachihara 1, Tsukuba, Ibaraki 305-0802, Japan)
E-mail: sakata@kenken.go.jp
[5]Member of JSCE, Department of Architecture and Civil engineering, Toyohashi University of Technology
(Hibarigaoka 1-1, Tenpaku-cho, Toyohashi, Aichi 441-8150, Japan)
E-mail: k-matsuo@ace.tut.ac.jp

Understanding the nature of the urban population and household structure has become quite important due to declining birthrate and aging population in Japan. Household Urban Micro-Simulation (HUMS) model is considered to be a useful prediction method. We will apply the model to 3 cities, Toyohashi, Toyama and Sapporo which are different on population scale, to provide simulations of future scenarios and make comparison between model parameters and simulation results to verify model versatility. Furthermore, transferability of model parameter between different target areas can be checked.

Key Words : *Micro-Simulation, Household micro-data, Open-data, Model versatility*

1. INTRODUCTION

In Japan, owing to population decline, declining birthrates, and an aging population, understanding the family structure and nature of the urban population in detail has become quite important. As such, the household urban micro-simulation model is one of the useful prediction method for population demographics as it considers a variety of attributes. The cohort component method is often used to forecast future population, in which the future population is forecast based on past demographics with an aggregate value, such as the population or household number, in a spatial unit as the minimum unit. This makes it difficult to express nonlinear demographics and the detailed attributes of residences. In response, a household micro-simulation model has been proposed, which can analyze future population and household distribution by treating individuals and households as minimum units; consequently, this model can influence the formulation of urban policy. In a pioneering study, the individual life cycle and evolution of household location were modeled and calculated using sample data[1)]. Furthermore, recently, the relationship between land use and transportation has been widely recognized, and interest in the urban micro-simulation model has increased as an analysis method for describing likely changes in land use and transportation over time. Several research studies have been performed addressing this issue and applying it to real urban cases [2,3,4,5,6]. HUMS model includes the initial distribution generation of household micro-data using household sample data[7,8]. However, because these methods use household sample data obtained from a large-scale questionnaire, the versatility of the analysis method is limited. Therefore, this study uses only Open-data.

In this study, we will apply Household Urban Micro-Simulation (HUMS) model to cities of different population sizes for transferability of model parameter between different target areas and reproducibility of model when simulation is applied to such differing target areas.

2. HUMS MODEL STRUCTURE

The basic structure of the HUMS model developed in this study is shown in Fig. 1. This model is divided into first-generation household micro-data in the initial year of simulation required for analysis, followed by the application of the urban structure prediction model, which describes the changes happening each year after the first year. In the HUMS model, first, changes in individual attributes of household micro-data are represented by a life event model. At each time step of the simulation, life events (aging, mortality, marriage, divorce, fertility, enrollment, employment, acquisition/voluntary surrender of driver's license, and relocation) are generated probabilistically. The relocation of a household is then determined by considering these changes in individual and household attributes. Meanwhile, transferred households (from outside the target area) are generated separately from the household micro-data within the target area. Then, the location choice, such as house type or residence zone, is expressed for the relocated household. Finally, the land price model expresses the change in land prices due to changes in land use.

(1) Initial Household Micro-Data

In this study, we have used Simulated Annealing (SA) method to generate initial household micro-data which was developed by Sakata, et al.[9]. Current SA

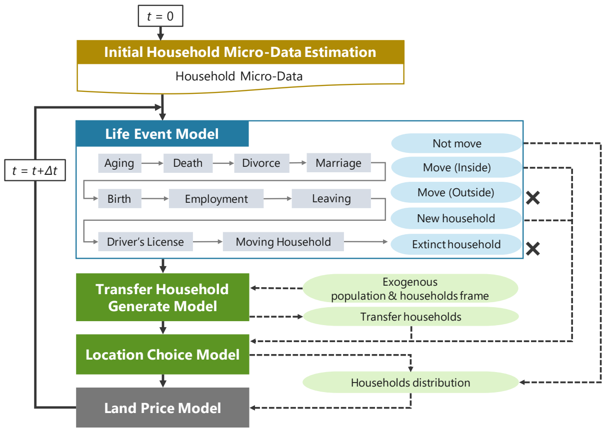

Fig. 1 The basic structure of HUMS model

method was used to minimize the objective error

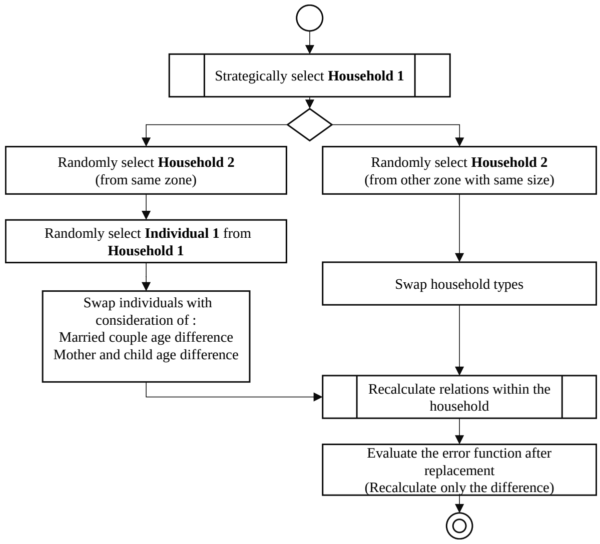

Fig. 2 Processing flow of SA method (for one iteration)

function that consists of age difference of mother and child, age difference of married couple, followed by other penalty functions to account the error for relation inside households. The processing flow of this method is shown on Fig. 2.

(2) Life Event Model

a) Aging

Aging event is assumed to occur in all individuals. If the simulation time step is given as Δt, the same Δt is added to the age of all individuals. The following life events are based on this concept of added age.

b) Mortality

Mortality event occurs probabilistically for all individuals. The mortality rate is estimated by gender and age based on survival analysis, assuming the Weibull distribution as the cumulative survival function

$$S(t) = \exp\left\{-\left(\frac{t}{\beta}\right)^{\alpha}\right\} \tag{1}$$

where t is age (survival time), and α and β are parameters.

c) Divorce

A divorce event occurs probabilistically for married individuals. The divorce of each couple is determined by the Monte Carlo approach, using the divorce rate based on the age of the husband obtained from statistical data. Furthermore, for those leaving households, three types of destinations were considered: home (household with mother), inside the target area, and outside the target area.

d) Marriage

Marriage events are assumed to occur probabilistically for males over 18 years old and females over 16 years old. Marriage is determined by the Monte Carlo approach using the "marriage rate by age and gender" data from the statistical data set. The matching of couples is determined by using the distribution of the age difference between married couples by age obtained from statistical data.

e) Fertility

Fertility events are assumed to occur probabilistically for married females aged 16 to 49 years. Fertility is determined by the Monte Carlo approach using the fertility rate by the mother's age and birth order taken from statistical data. The generalized log gamma distribution, which is used to estimate the birth probability, is

$$g_n(x) = \frac{C_n|\lambda|}{b_n\Gamma(\lambda_n^{-2})}(\lambda_n^{-2})^{\lambda_n^{-2}} exp\left[\lambda_n^{-1}\left(\frac{x-u_n}{b_n}\right) - \lambda_n^{-2} exp\left\{\lambda_n\left(\frac{x-u_n}{b_n}\right)\right\}\right] \tag{2}$$

where $g(x)$ is the fertility rate of the n-th child of an x year-old female, $\Gamma()$ is the gamma distribution, and C_n, λ_n, b_n, and u_n are parameters.

f) Employment and Enrollment

Employment and enrollment status is updated probabilistically. Status update from enrollment and graduation are determined by 'Enrollment and employment rate' from the statistical data.

g) Driver's License

The driver's license status is updated probabilistically for individuals over 18 years old. The acquisition and voluntary surrender rate of driver's license are calculated by shifting the "driver's license holder rate" by a 1 year; rise of shifted holder rate is taken as acquisition rate and fall of shifted holder rate is taken as voluntary surrender rate.

h) Independence

Leaving events are assumed to occur probabilistically for individuals. The leaving rate by gender and age is defined separately from leaving due to divorce, employment, or enrollment, and individual home departure is determined based on the leaving rate. The relocation of the departing individual is determined by the "house type choice" and "zone choice" models.

i) Moving

Moving events are assumed to occur probabilistically for households. Household relocation is determined using the census data of "current residence, place of residence five years ago, and gender and age" and by applying the Monte Carlo approach, using the choice probability based on the householder's age and the attributes of other household members.

(3) Location Choice Model

a) House Type Choice Model

For the relocation of households inside the target area, the house type is determined by the house type choice model. The choice of house type is represented by a multinomial logit model with household attributes as variables. The choice set H_n of house type h of household n is, $H_n = \{h = 1$ (own / detached), $h = 2$ (own / apartment), $h = 3$ (rent / detached), $h = 4$ (rent / apartment)$\}$. The choice probability and utility function of the multinomial logit model in the house type choice model are

$$P_{hn} = \frac{e^{V_{hn}}}{\sum_{h' \in H_n} e^{V_{h'n}}} \quad (h \in H_n) \tag{3}$$

$$V_{hn} = \sum_k \theta_k X_{khn} + c \tag{4}$$

where X_{khn} is a household attribute variable, such as householder age, household size, and number of children. These parameters are set based on the house type composition for each current household type.

b) Zone Choice Model

For the relocation of households inside the target area, the relocation zone is determined by the selection of a residence zone according to a multinomial logit model that chooses only one zone from each zone in the target area. In the zone choice model, the choice probability is considered for each of the four house types. Assuming the choice set of the residence zone of household n is Z_n, the multinomial logit model and utility function for the choice probability of zone i are as follows:

$$P_{ihn} = \frac{e^{V_{ihn}}}{\sum_{i' \in Z_n} e^{V_{i'hn}}} \quad (i \in Z_n) \tag{5}$$

$$V_{ihn} = \sum_k \alpha_k X_{kin} + \gamma LP_i + c \tag{6}$$

where X_{kin} is a zone attribute, such as transportation conditions and land conditions, and LP_i is the land price.

c) Land Price Model

The land price LP_i of each zone at the end of each simulation time step was calculated using the hedonic regression model. The land price model is given as

$$LP_i = \sum_k \gamma_k X_{ki} + \delta D_i + c \tag{7}$$

where X_{ki} is the zone condition, such as the distance to the station and land use zone, and D_i is the location density. With this model, the land price used in the next residential zone choice is updated.

3. MODEL APPLICATION TO REAL CITIES

(1) Target Areas and Data sources

In this study, our target area to apply the HUMS model will be Toyohashi, Toyama and Sapporo. The model is applied based on fourth-order grid (500m grid). Furthermore, census regional grid statistics are used as marginal distributions to generate individual and household attributes. In addition, the probability of life event is defined based on data from sources such as “national census” and “vital statistics”. Since we will set the initial year of the simulation as 2015, the data used in this study were all available open-data from 2015.

Table 1 Mortality event parameter estimation result

	Sapporo		Toyama		Toyohashi		Kosai	
par.	male	female	male	female	male	female	male	female
α	9.3	11.8	9.2	11.6	9.5	12.1	7.4	11.4
β	109.6	110.0	109.4	109.9	109.4	108.3	117.3	109.7
R^2	0.999	0.998	0.999	0.999	0.999	0.999	0.992	0.999

Table 2 Fertility event parameter estimation result

City	Child	Parameters			
		Cn	Un	bn	λn
Sapporo	1st	0.61	29.71	5.34	0.08
	2nd	0.41	32.24	4.88	0.26
	3rd	0.12	33.98	4.69	0.42
	4th	0.03	33.21	7.38	-0.09
	5th	0.02	37.39	8.49	6.19
Toyohashi	1st	0.72	29.31	4.74	0.08
	2nd	0.55	31.74	4.50	0.08
	3rd	0.17	33.37	4.28	0.41
	4th	0.03	36.43	2.20	3.85
	5th	0.02	37.75	8.37	6.62
Toyama	1st	0.72	28.86	4.77	0.08
	2nd	0.54	31.36	4.50	0.08
	3rd	0.18	33.27	4.38	0.12
	4th	0.04	41.13	1.74	8.38
	5th	0.02	38.21	8.52	6.81
Kosai	1st	0.71	28.92	5.06	0.08
	2nd	0.56	31.25	4.76	0.08
	3rd	0.18	32.60	4.34	0.12
	4th	0.04	40.21	1.55	9.84
	5th	0.02	37.22	8.15	6.51

Table 3 Land price parameter estimation result

Par.	Sapporo		Toyohashi		Toyama	
	par.	tval	par.	tval	par.	tval
log10(Dist. to city center (m))	-2.218	-1.067	-8.018	-7.88	-2.431	-1.31
log10(Dist. to closest station (m))	-4.296	-3.756	-1.378	-1.73	-2.828	-2.46
Loc. Density (Population) (person/ha)	-5.954	-1.274	3.903	2.719	0.014	0.61
Loc. Density (No. Employee) (person/ha)	7.275	35.333	-	-	0.034	6.45
Commercial area (Dummy)	-	-	-	-	4.148	2.96
Const.	26.4	3.737	38.767	12.06	20.021	2.6
R2	0.716		0.733		0.793	

Table 4 House type choice model parameter estimation

Par.	Toyohashi		Toyama		Sapporo	
	Own detached		Own detached		Own detached	
	par.	tval	par.	tval	par.	tval
Household size	0.351	6.990	0.539	8.876	-	-
Household head's age	0.011	6.100	0.007	3.387	0.007	4.945
No child (dum.)	-	-	-	-	-0.418	-5.496
Household head's gender (dum.)	-	-	-	-	-	-
Full time job (dum.)	-	-	-	-	-	-
Charactesric dummy variable	-0.675	-4.550	-0.188	-0.658	-0.247	-1.691
Sample size	5000		5000		10000	
R^2	0.443		0.508		0.226	

(2) Model Parameter Estimation Results

The model we considered in this study that estimates parameters were following; Mortality, Fertility, Location Choice model.

a) Mortality Event Parameters

Parameter estimation result of mortality event, where it assumes Weibull's distribution is shown in Table 1. The mortality rate is calculated by using statistical data from "Life Table" from 2015.

b) Fertility Event Parameters

Parameter estimation result of fertility event, where it assumes Generalized Log Gamma distribution is shown in Table 2. In current study, we have assumed maximum child birth per individual up to 5th order.

c) Land Price Model Parameters

In the land price model, explanatory variables of the land price were road distance to central station (city center), road distance to closest station and location density (population and employee). For this model, "public announcement of land price" from the National Land Information dataset in 2015 is used. Land prices were estimated for grid level by using spatial point data. The parameter estimation result of each city is shown in Table 3.

d) House Type Choice Model Parameters

The explanatory variables of the house type choice model are household size, age of head of household, and the dummy variable no children. Furthermore, since our study assumes four types of house type, we will show the result of own/detached type as an example in Table 4. Parameters were estimated by using 5000 randomly sampled household labels from the initial household micro-data set of the target area. The estimation results show valid sign conditions and significant t values for each parameter.

Table 5 Zone choice model parameter estimation result

Par.	Toyama		Toyohashi		Sapporo	
	Own detached		Own detached		Own detached	
	par.	t-val	par.	t-val	par.	t-val
log10(Land Price)	-4.358	-10.160	-0.533	-12.546	-4.006	-14.078
log10(Detached stock)	8.260	19.261	0.039	18.151	9.439	34.168
log10(Dist. to closest station)	-8.029	-18.025	-5.291	-13.130	-10.021	-38.616
Low Rise Building Area (Type 1&2)	1.371	6.344	-	-	2.074	18.552
Mid Rise Building Area (Type 1&2)	-	-	-	-	-	-
Residential Building Area (Type 1&2)	1.179	6.705	-	-	1.116	12.307
R^2	0.696		0.555		0.631	
Sample size	2000		2000		2000	

e) Zone Choice Model Parameters

The explanatory variables of the house type choice model are household size, age of head of household, and the dummy variable no children. Same as we previously did, zone choice Table 5 shows result of own/detached type as an example. The estimation results show valid sign conditions and significant t values for each parameter.

(3) Simulation result comparison

With the various model we have created so far, we can now use it to simulate the future prediction of the demographic changes of the cities. We have set the simulation timestep to between 2015 to 2025. Since the open-data between 2015 to 2019 is obtainable from each city's Basic Residence Register data, we can use it to compare with the HUMS model to verify how well it performs. In this case, Fig 3 shows total number of population and household for each cities compared with statistical data. Furthermore, Fig 4 shows population by 5 age group in 2019 for 3 target areas. The legend "HUMS" indicates simulation result whereas "BRR" indicates

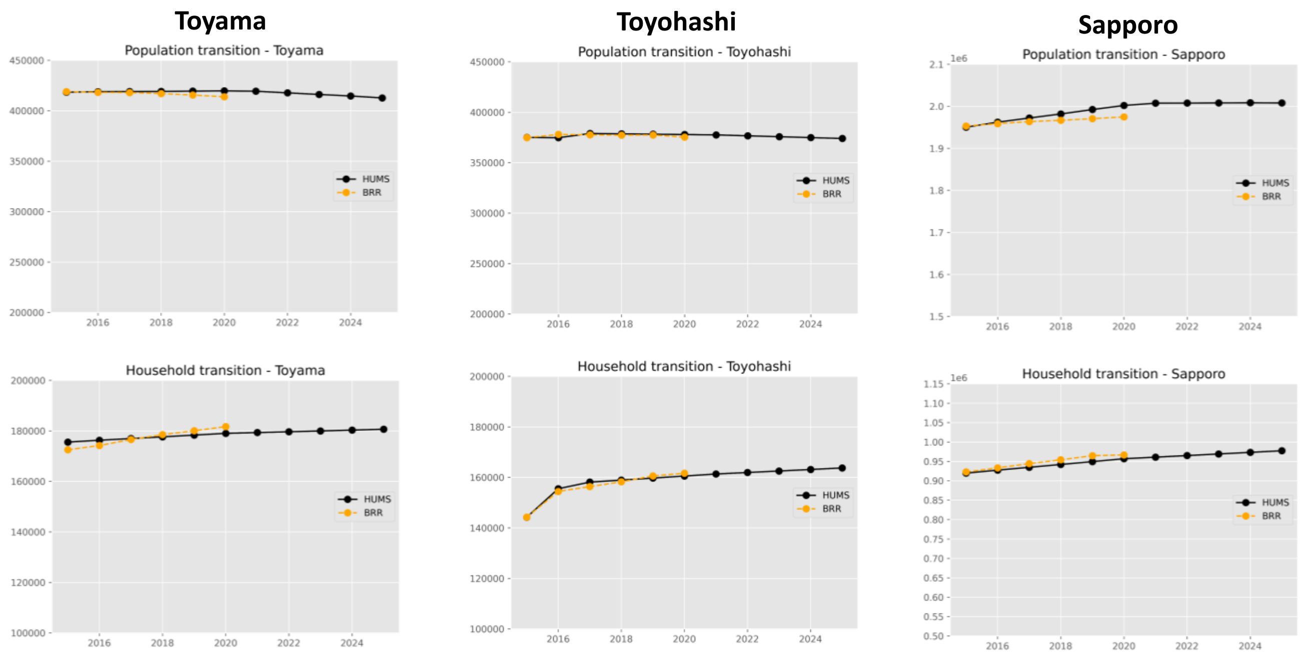

Fig. 3 Population and household transition for target cities compared with statistical data

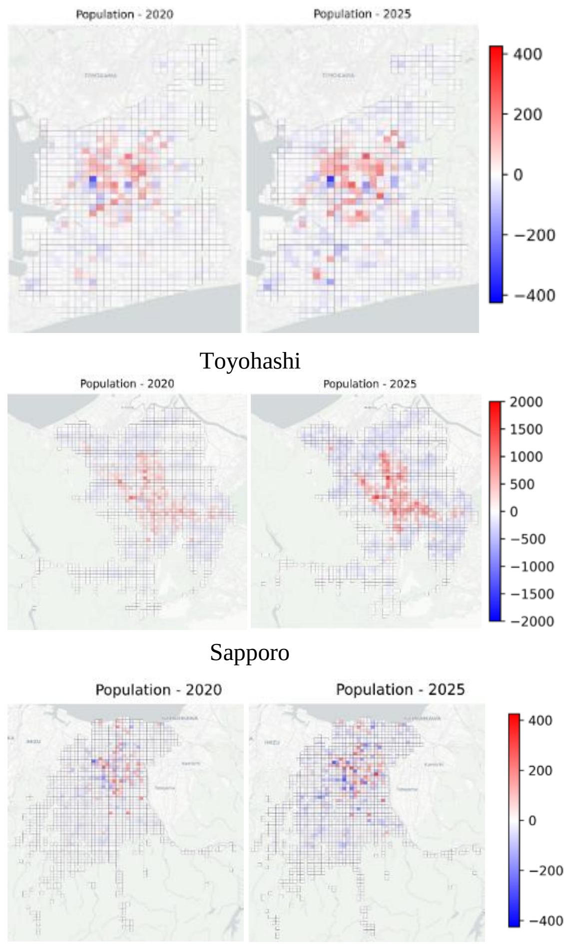

Fig. 5 Population by 5 age group from simulation compared with statistical data

actual statistical values from Basic Residence Register. Moreover, it is possible to measure the changes over time for each 500m grid areas by using location choice model. Demographic change due to relocating households for target cities in case of year 2020 and 2025 are shown in Fig 5 where it indicates population and household difference from year 2015.

4. PARAMETER TRANSFERABILITY

The HUMS model we used thus far involves many models to estimate parameters. If one were to apply this model to new target city, creating and estimating those from zero would be heavy burden. One way to handle such problem is to check whether if its possible to transfer the already estimated parameter from one target city to another target city, to check the reproducibility. In this section, we have set the Toyohashi city as a base target and used Sapporo and Toyama city's life event model parameters, namely fertility and mortality events as a

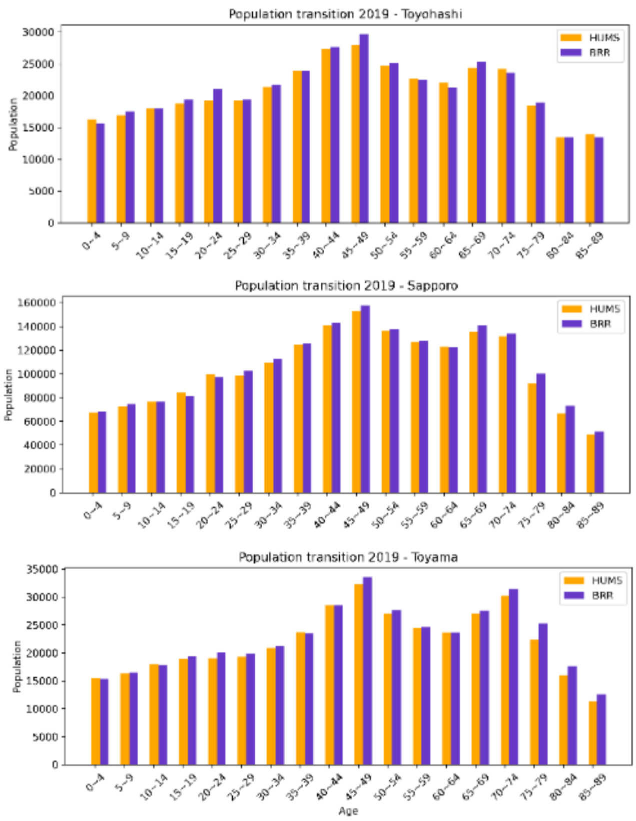

Fig. 4 Population by 5 age group from simulation compared with statistical data

transfer parameters. As we can see from the Fig 6, total population from simulation result when the target city is Toyohashi city was able to indicate negligible difference from actual statistical values. Furthermore, in Fig 7 we can also see the household type transition was also showing stable results when the other city's parameters were transferred.

5. CONCLUSION

In this study, we have applied HUMS model to real cities, Toyohashi, Toyama, Sapporo, where each city differ by their population sizes and other urban structures such as their main transportation mode, etc. The objective was to first simulate each target areas and check the result to verify the model versatility. As it was shown in the previous sections, the simulation result was compared with statistical data to verify. Second, we attempted to check if it is possible to transfer the model parameters between the cities. It was done by transferring mortality and fertility event parameters. However, it is also possible to transfer the

location choice model parameters between the target cities and verify it by grid differences for each simulation timestep. Furthermore, current study is yet to fully consider Kosai city as a target city.

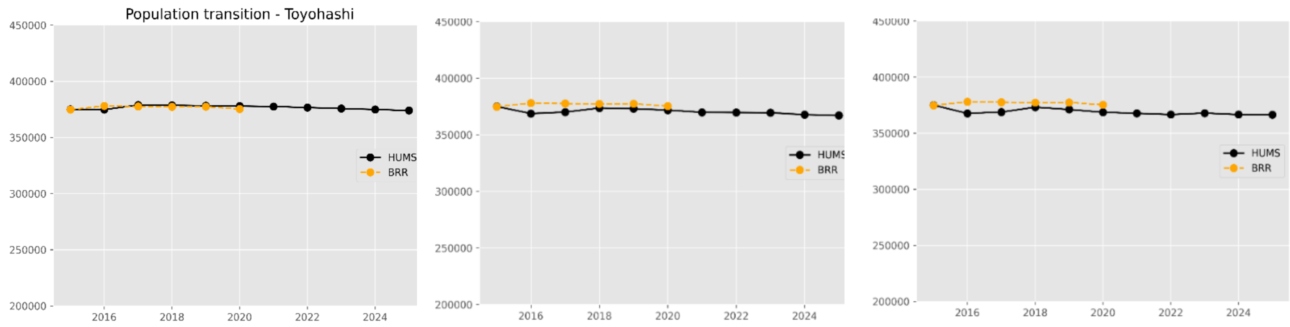

Fig. 6 Total population change of Toyohashi city (used parameters from left to right: Toyohashi, Toyama, Sapporo)

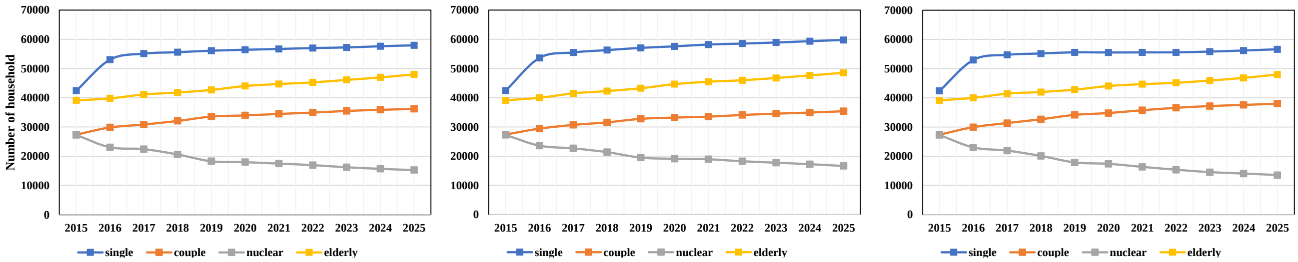

Fig. 7 Household type transition of Toyohashi city (used parameters from left to right: Toyohashi, Toyama, Sapporo)

However, once it is accomplished, we can verify the performance of HUMS model when it is applied to such a small scale city with small population.

6. ACKNOWLEDGEMENT

This research was supported by the Japan Society for the Promotion of Science (JSPS) under the Grant-in-Aid for Scientific Research (C) 20K04721. We would like to express our gratitude for this support.

REFERENCES

1) Hayashi Y, Tomita Y (1988) A Model for Zonal Forecast of Life Cycle Progress, Residential Location and Population Attributes Using Random Utility Models and A Micro-Simulation Technique. Journal of JSCE 395/Ⅳ-9: 85-94
2) Wegener M, Frurst F (1999) Land-Use Transport Interaction: State of the Art, TRANSLAND Integration of Transport and Land Use Planning. Work Package 2 Deliverable D2a, Institute of Spatial Planning, University of Dortmund
3) Wegener M (2003) Overview of Land-Use Transport Model. In: Proceedings of CUPUM'03
4) Waddell P (2002) UrbanSim: Modeling urban development for land use, transportation, and environmental planning. Journal of the American Planning Association 68(3): 297
5) Hunt JD, Donnelly R, Abraham JE, Batten C, Freedman J, Hicks J, Costinett PJ, Upton WJ (2001) Design of a Statewide Land Use Transport Interaction Model for Oregon. In: Proceedings of 9th World Conference on Transport Research
6) Chengxiang Z, Chunfu S, Jian G, Chunjiao D, Hui Z (2016) Agent-based joint model of residential location choice and real estate price for land use and transport model. Computers, Environment and Urban Systems 57:93–105
7) Sugiki N, Varameth V, Otani N, Kazuaki M (2012) Agent-Based Household Micro-Datasets: An Estimation Method Composed of Generalized Attributes with Probabilistic Distributions from Sample Data and Available Control Totals by Attribute. Asian Transport Studies 2(1): 3-18
8) Sugiki N, Muranaka T, Otani N, Kazuaki M (2015) Agent-based Estimation of Household Micro-data with Detailed Attributes for a Real City. In: Proceedings of CUPUM'15
9) Tomohiko Sakata, Atsushi Suzuki, Nao Sugiki, A Trial of Producing Initial Micro Data of Households Targeted at 2 Million City, Infrastructure planning conference JSCE, Vol.62, CD-ROM, 2020.

(Received October 1, 2021)