Development of Micropopulation Census through Disaggregation of National Population Census

Yuki Akiyama[1], Hiroyuki Takada[2], and Ryosuke Shibasaki[3],

[1]Earth Observation Integration & Fusion Research Institute, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan
[2] Graduate School of Frontier Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan
[3] Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa city, Chiba, Japan
aki@iis.u-tokyo.ac.jp; sturdy-runner@csis.u-tokyo.ac.jp; shiba@csis.u-tokyo.ac.jp

Abstract

The national census is widely used to monitor the distributions and dynamics of population. However, it is difficult for an open census to monitor populations in small areas such as city blocks or grids of arbitrary size. Recently, detailed and accurate population data are required for regional analysis of policies including welfare, medial, disaster prevention, and optimum services for social change. The limitations of the existing census data have a significant impact on these policies. Therefore, we estimate the distribution data of households and residents in order to disseminate the national population census stochastically to building locations. We use some demographic charts from the national population census about household and resident data. This data can be aggregated into an arbitrary spatial unit that is sufficiently reliable to overcome the foregoing restrictions by providing coverage throughout Japan.

1. Introduction

Recently, many kinds of statistical data can be acquired from national and municipality websites because many of them are being digitized. Many studies use the national census (undertaken by the Japanese Ministry of Internal Affairs and Communications) as detailed and reliable population data have been required to solve various urban problems, such as urban planning, disaster and crime prevention, as well as health and welfare policies.

However, not all open census data can provide the high-resolution spatial information required to solve these urban problems because much of the data has not been aggregated into administrative units. An increasing number of studies have sought to develop unrestricted population data by aggregating administrative units in response to the growing demand for spatially detailed population data to solve the aforementioned problems.

The first studies to properly use microdata appeared in the 1960s. Orcutt et al. (1961) used microdata for demography. Their study introduced the microsimulation method. Microdata have since been used in traffic, land use, and economic analyses. Some unique microsimulations have used microdata by developing data infrastructure and increasing demand after the 1990s. Notable examples include the “UrbanSim” for microsimulation about the locations of households and companies in the United States, the “ILUTE” for the monetization of the relationship between land use and traffic in Canada, and the “DELTA” for land use and economic analyses in the United Kingdom.

Japanese studies began to use microdata in the 1980s to monitor the population distributions of residents, particularly for urban and traffic planning. The development of the residential demand model by Miyamoto et al. (1986) and the population model by Hayashi et al. (1988) pioneered detailed population data. In the same period, Inagaki (1986) developed the “INAHSIM” microsimulation model for the future estimation of households to follow developments of microsimulation in foreign countries. Furthermore, Sugiki (2009) developed the estimating method of disaggregating population data, and Hanaoka (2009) developed the spatiotemporal method of dynamic microsimulation.

However, the results of these studies were aggregated data in administrative units or grid polygons. More recently though, there has been a demand for disaggregated population data with location information, which can be aggregated into arbitrary spatial units (Fig. 2).

1.1. The Necessity of Disaggregated Population Data

We used an example of the necessity of disaggregated population data in the case of damage estimation for a tsunami. The extant open population census and the results of previous studies were aggregated into administrative units or grid polygons (Fig. 1). Therefore, the estimation of population affected by the tsunami was calculated according to the area damaged. For example, the damaged populations in zone B, E, G, and H in Fig. 1 were calculated by multiplying the population of each zone by the damaged area. If buildings were not uniformly distributed in each zone, we could not use the above method to obtain a reliable result. However, we could acquire reliable results without the restriction of the existing aggregating units if the locations of all households and the number of residents could be estimated (Fig. 2). Such data is very meaningful not only for damage estimation by tsunami but also for the above fields.

Many fields require disaggregated population data. For example, the monitoring of the interzonal flows of people who have specific attributes, such as age group and gender (Esaki et al. (2005), and a study of the distribution and mobility of an aging household (Shimizu et al. (2009) showed the necessity of disaggregated population data. Studies outside Japan (Wegener (2003) have also demonstrated its value for the development of an urban model using microsimulation and for the assessment of the demand for hot water in the residential sector (Murshed 2009; Steinnocher et al., 2010).

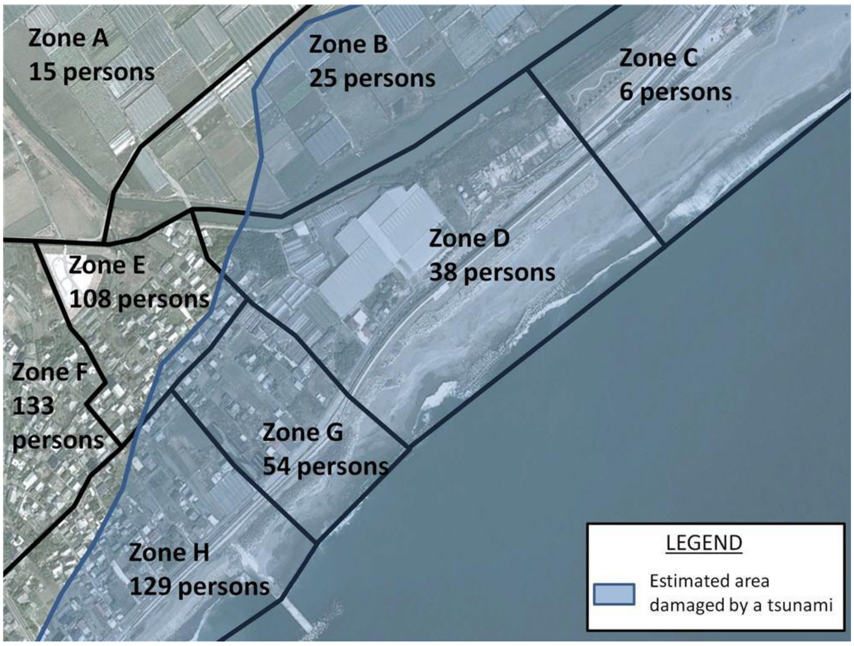

Fig. 1. Population data aggregated into administrative units and estimated area damaged by a tsunami

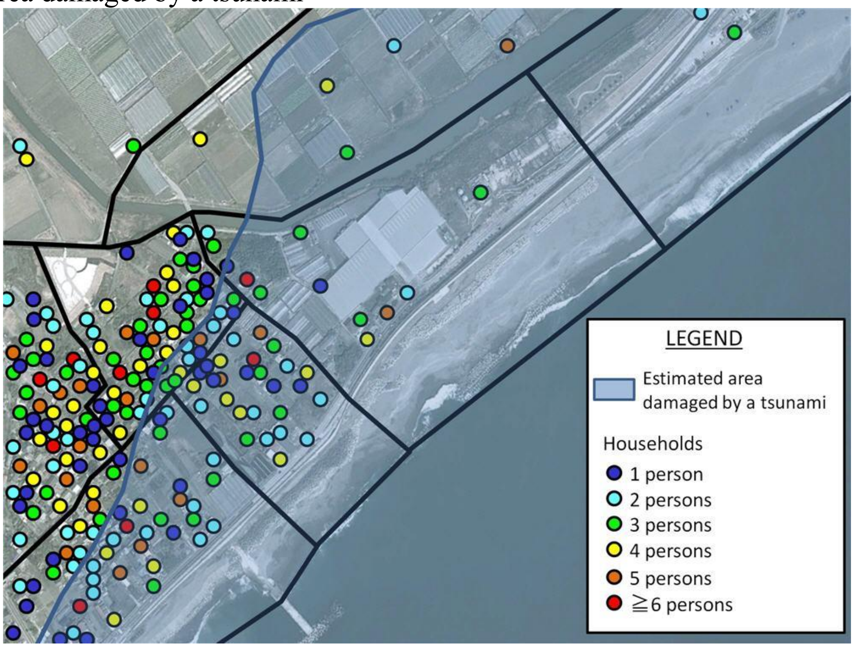

Fig. 2. Disaggregated population data and estimated area damaged by a tsunami

1.2. Objective

We developed estimates of the household and resident data of each building, called the “Micropopulation Census,” as per Fig. 2, to disaggregate Japan’s national population data into a detailed digital map called the “residential map” (“Z-map TOWN II” published by ZENRIN CO., LTD). This digital map can monitor the location and use of buildings, in addition to information about their residents, throughout Japan. This data can be aggregated into arbitrary spatial units because this is disaggregated estimating data. It can be expected to solve the above problems about spatial units. In addition, previous microsimulations and studies using microdata do not deal with building unit data.

2. Data Development

Fig. 3 shows the flow of data development in this study. First, the numbers of households in each aggregating unit were collected using the 2005 Japanese census. Households in the Japanese open population census were arranged in each administrative unit called “Cho-cho-moku” in a 1km square or 500m square grids. In this study, we used population data aggregated into 500m square grids. The buildings allocated household data were determined from the residential map in 2008 based on the number of households in each grid. Second, estimates of the distribution data of households were developed to allocate information about households from some statistical tables in the population census using the methods described in 2–2. Finally, estimates of the distribution data of residents, that is, the Micropopulation Census, were developed to allocate information about residents from some statistical tables in the population census using the methods described in 2–3 and the Japanese Vital Statistics (Ministry of Health, Labor, and Welfare of Japan) in 2005. The Vital Statistics are statistics on live births, deaths, fetal deaths, marriages, and divorces.

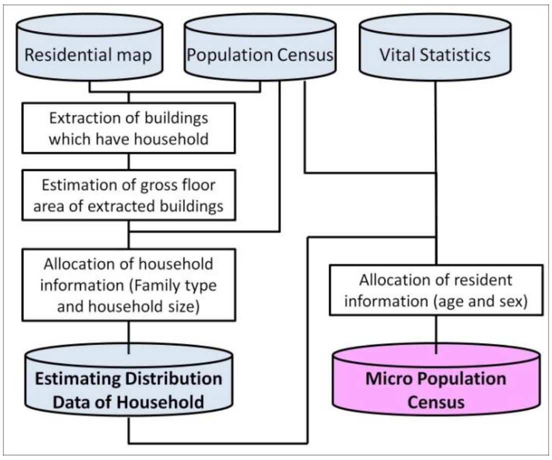

Fig. 3. Development flow of the Micro Population Census

2.1. Development of Estimated Distribution Data of Households

2.1.1. Deciding Which Buildings Have Household Information

It was necessary to decide which buildings had household information in order to allocate information in the population census to the residential map. All building data in the residential map have attributes of building use (detached houses, business establishments, multitenant building, landmarks, and others), building area, and the number of stories. In addition, the uses of all rooms in multitenant buildings can also be monitored (personal houses, business establishments, landmarks, and others). Other buildings and rooms featured these attributes but were not surveyed, including some households. The attribute information was used to decide the buildings and rooms likely to be allocated household information (Fig. 4). The number of buildings and rooms were counted in ascending order in each 500m square grid. If this total had the same value as the number of households in the population census, the buildings and rooms contained in the sum total were locations allocated household information from the census.

1) Detached houses (Pink buildings “B” in Fig. 4)
2) Personal houses in multiuse buildings (Pink rooms “D” in Fig. 4)
3) Others in multiuse buildings (Yellow rooms “E” in Fig. 4)
4) Other buildings (Yellow buildings “F” in Fig. 4)

5) Business facilities (Green buildings "G" in Fig. 4)

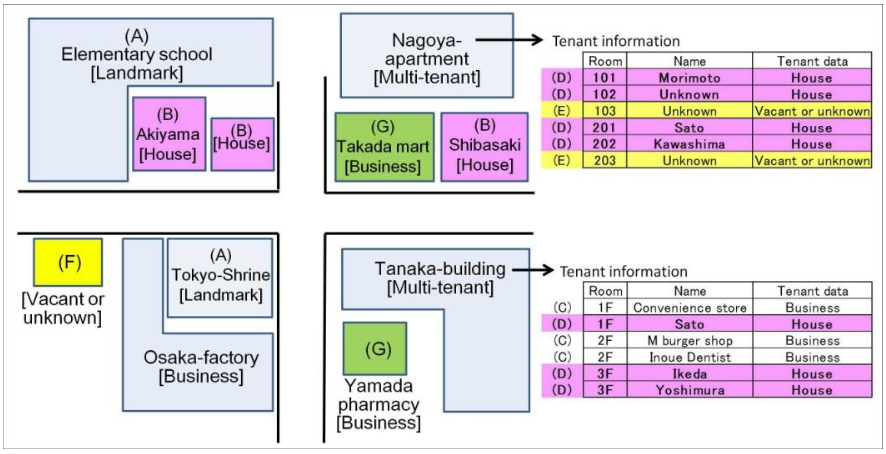

Fig. 4. Buildings and rooms in the residential map likely to be allocated household information from the population census are decided by this process

2.1.2. Estimation of Gross Floor Area of Each Building and Room

The gross floor area of extracted buildings and rooms were estimated by the following methods:

1) Detached buildings (building "B", "F", and "G" in Fig. 4)
Estimates of gross floor area in this case is given by
(Estimation of gross floor area) = (building area) × 1.5 (1)

This is because almost all detached buildings in Japan are two-storied and the floor area of the second floor is smaller than the first floor.

2) Rooms in multitenant buildings (room "D" and "E" in Fig. 4)
Estimation of the gross floor area in this case is given by

$$E_N = \frac{F_N S_N}{Sum_N} \tag{2}$$

Where Sum_N is the number of rooms of building N, F_N is the number of stories of building N, S_N is the area of building N, and E_N is the estimation of the gross floor area of building N. However, S_N contains common use spaces (i.e., common passageways or stairways). The area of common use spaces in typical Japanese multitenant buildings are approximately 20 percent of gross floor area. Therefore, S_N was multiplied by 0.80.

Consequently, all buildings and rooms allocated household information were given an estimated gross floor area.

2.1.3. Allocation of Family Type to Residential Map

There are many kinds of cross tables of gross floor area with household information in the Japanese census. Cross tables are available for all areas in Japan. In this study, information about family type was allocated to a residential map by estimating the gross floor area developed in 2–1-2 using cross tables of gross floor area (14 groups, Table 1) with family type (16 groups, Table 2) in each municipal unit. Table 3 shows the sample of the cross table.

Table 1. Groups of gross floor area in Japanese population census

Group number	Gross floor area[m^2]
1	0-19
2	20-29
3	30-39
4	40-49
5	50-59
6	60-69
7	70-79
8	80-89
9	90-99
10	100-119
11	120-149
12	150-199
13	200-249
14	250-

Table 2. Groups of family types in Japanese population census

Group number	Family type
1	Married couple
2	Married couple and some children
3	Husband and some children
4	Wife and some children
5	Married couple and their parents
6	Married couple and their parent
7	Married couple, some children, and parents
8	Married couple, some children, and a parent
9	Married couple and some relatives
10	Married couple, some children, and some relatives
11	Married couple, their parents, and some relatives
12	Married couple, some children, their parents, and some relatives
13	Siblings
14	Other family members (only relatives)
15	Others
16	Single

Note: Relatives do not contain householder's parents and children

Table 3. Sample of the cross table of gross floor area with family type

		Family type							
		1	2	3	4		14	15	16
		Husband and wife	Husband, wife and some children	Husband and some children	Wife and some children	…	Other relative	Not a relative	Single
Gross floor area [m^2]	0～19	124	39	8	27		5	26	2675
	20～29	592	342	38	196		19	74	7158
	30～39	1326	1165	80	611		41	123	5101
	⋮								
	150～199	1364	2515	87	385		124	7	531
	200～249	218	425	17	63		24	0	101
	250～	106	202	13	36		14	0	54

First, the rates of each family type were calculated in each group of gross floor area (14 groups). Second, the total households in each group of gross floor area were calculated to count the buildings and rooms in each group of gross floor area from building data developed in 2–1-2. Third, the numbers of households of each family type were calculated by multiplying these numbers of household by the rates of each family type. Finally, they were randomly allocated to buildings developed in 2–1-2. Consequently, new attribute information (i.e., family types) was allocated to the residential map.

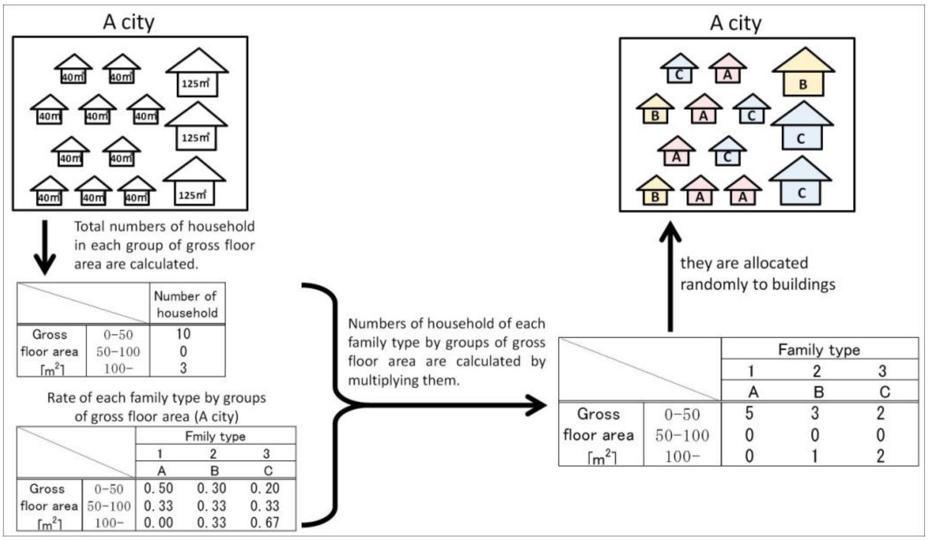

Fig. 5. Processing image to allocate information about family types to the residential map

2.1.4. Allocation of Household Size to Residential Map

There are cross tables of family types (16 groups) with household size—the number of residents in this household (7 groups)—in each municipal unit in the Japanese population census. This data was allocated to the residential map using the method of 2–1-3.

First, the rates of household size in each family type were calculated using cross tables. Buildings in the residential map already had family types after the processing in 2–1-3. Using this data, the total of households in each family type (1) could be calculated. Second, the numbers of household in each household size could be calculated by multiplying (1) by the rates of household size in each family type. Finally, they were randomly allocated to buildings developed in 2–1-3. Consequently, new attribute information (i.e., household size) was allocated to the residential map and the distribution data of households was estimated (Fig. 6.).

2.2. Development of the Estimates of Residents' Distribution Data

2.2.1. Allocation of Householder Information

There are cross tables of family types (16 groups) with information about the head of the household that includes age and gender in each municipal unit in the Japanese population census. This data was allocated to the data developed in 2–1-4. All households in the data developed by 2–1-4 have information about the head of the household, which was allocated to the residential map using the following method.

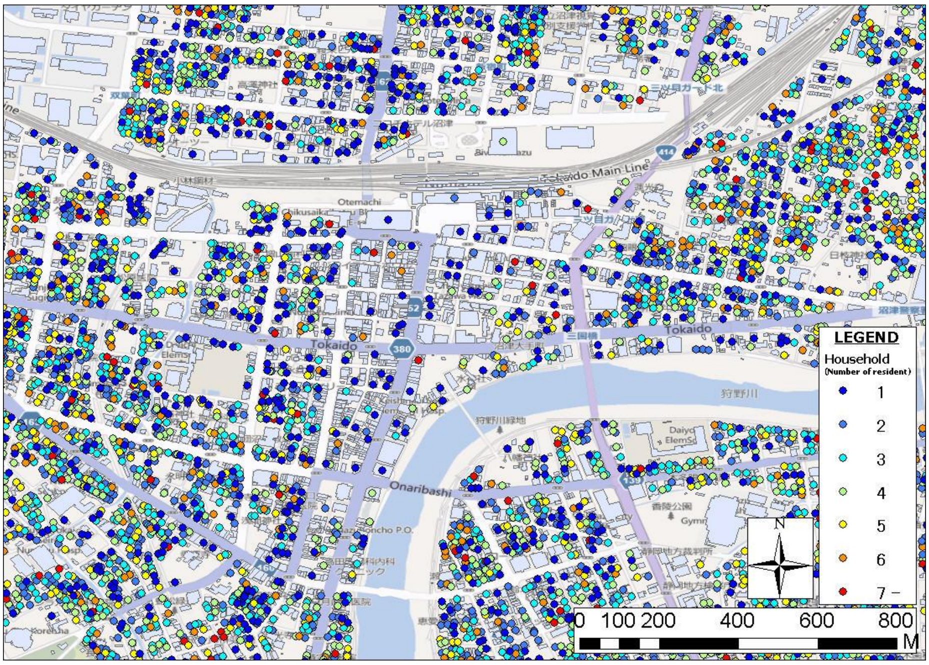

Fig. 6. Estimated distribution data of households (Numazu city, Shizuoka prefecture)

The population census has information about the age and gender of householders and family type. First, the ages of householders were re-aggregated into ages by every 5 years (21 groups) using the method in Table 4 and were then converted into rates (1). Second, the numbers of households of each family type were calculated using the household data estimated in 2–1-4 (2). Third, householders' population by age and gender in each family type were calculated by multiplying (1) by (2). Finally, they were randomly allocated to buildings developed in 2–1-4. Consequently, household information was allocated to all households.

2.2.2. Allocation of Spouse Information

This was necessary to confirm the presence or absence of spouses in each household and to allocate information about them.

In this study, we counted households in which couples have spouses. Table 5 shows the presence or absence of spouses in each family type. The following processes were applied to households with spouses.

There are cross tables of the gender and age of householders and their spouses in each municipal unit in the Japanese census.

First, the ages of householders and spouses by every year between 0 and more than 100 were re-aggregated into ages by every 5 years (21 groups), as per Table 4, using these cross tables.

Second, information about spouses was allocated. The ratio of men householders was much larger than the women ratio. Therefore, households with women householders were allocated information about spouses in preference to households with men householders to reduce processing errors. First, the population rates of husbands relative to wives in each age group were calculated. Second, the total populations of each age group in which the householder was women and had a spouse were calculated. Third, the population of men spouses in each age group of women householders was calculated to multiply them in each age group. Finally, they were randomly allocated to buildings with women householders. Consequently, age and gender (men) were allocated to households with women householders.

Finally, information on spouses was allocated to households with men householders. At first, the population rates of wives in each age group relative to husbands were calculated. The numbers of men already allocated as spouses were deducted to avoid double counts. Second, the total population of each age group in which the householder is men with a spouse were calculated. Third, the populations of women spouses in each age group of men householders were calculated to multiply them in each age group. Finally, they were randomly allocated to buildings in which the householders were men. Consequently, age and gender (women) were allocated to households with men householders.

Consequently, information about spouses (their ages and genders) was allocated to all households with spouses.

Table 4. Re-aggregation of age groups

Age Group	Age
1	0 - 4
2	5 - 9
3	10 - 14
4	15 - 19
5	20 - 24
6	25 - 29
7	30 - 34
8	35 - 39
9	40 - 44
10	45 - 49
11	50 - 54
12	55 - 59
13	60 - 64
14	65 - 69
15	70 - 74
16	75 - 79
17	80 - 84
18	85 - 89
19	90 - 94
20	95 - 99
21	100 -

Table 5. Family types and the presence or absence of spouses, householder's children, parents, and other relatives

Family group	Family type	Number of householder's			
		Spouse	Child	Parent	Other relative
1	Husband and wife	1	0	0	0
2	Husband, wife, and some children	1	1 -	0	0
3	Husband and some children	0	1 -	0	0
4	Wife and some children	0	1 -	0	0
5	Husband, wife, and parents	1	0	2	0
6	Husband, wife, and a parent	1	0	1	0
7	Husband, wife, some children, and parents	1	1 -	2	0
8	Husband, wife, some children, and a parent	1	1 -	1 -	0
9	Husband, wife, and a relative (not including some children and parents)	1	0	0	1 -
10	Husband, wife, some children a relative (not including some children and parents)	1	1 -	0	1
11	Husband, wife, and parents (not including some children and parents)	1	0	0	1 -
12	Husband, wife, some children, parents, and a relative	1	1 -	1	1
13	Siblings	0	0	0	1 -
14	Other relative	0	0	0	1 -
15	Not a relative	0	0	0	0
16	Single	0	0	0	0

2.2.3. Allocation of Child Information of Householders

First, it was necessary to confirm the number of children in each household in order to allocate child information to their householders. The number of children was calculated using family types and household size (i.e., the number of residents). The number of children was not clear by family types. However, it could be calculated by equation 3:

$$Nc_i = Nn_i - Nh_i - Ns_i - Np_i - No_i \qquad (3)$$

Where Nn_i is household size (the number of residents), Nh_i is the number of householders (= 1), Ns_i is the number of householders; spouses (=0 or 1), Np_i is the number of householders' parents (=0, 1 or 2), No_i is the number of other relatives, and Nc_i is the number of children of household *i*.

Table 5 shows the number of children in each family type calculated by this equation.

Second, the population rate of children in each age group and the gender rate in each age group of their parents were estimated using information on the relationship between the ages of the parents and the ages of their children from the Japanese Vital Statistics. The Japanese Vital Statistics contains the genders and mothers' ages when she gave birth to Nth children (N>0). Ages of Nth child were estimated by equation 4:

$$Ac(N)_i = Ah_i - A(N) \qquad (4)$$

Where Ah_i is the age of the householder or spouse, $A(N)$ is mothers' age when she gave birth to her Nth child, and $Ac(N)_i$ is the estimated age of Nth child of household *i*.

For example, if the current age of the first child is 15 and if the current age of his parent is 50, she gave birth to him when she was 35.

In addition, Ah_i is decided by the following method.

1) Ah_i is an age of spouse if spouse is women.
2) Ah_i is an age of householder if spouse is men.
3) Ah_i is an age of householder if householder is women and there is no spouse.
4) Ah_i is an age of householder if householder is men and there is no spouse.

Consequently, the genders and ages of all children in each age group of parents were estimated by the above method.

Finally, they were randomly allocated to households with children. Consequently, the number of children, along with their genders and ages, were allocated to the residential map.

2.2.4. Allocation of Parents Information of Householders

The method of allocating the parent information of householders was similar to the method of 2–2-3. At first, it was necessary to confirm the number of householders' parents in each household to allocate the parent information of their householders. The number could be determined by family types. Table 5 shows the number of parents that were householders.

Second, the rate of parents in each age group and the gender in each age group of their householders was estimated. This method is similar to the method in 2–2-3. They were also calculated using information on the relationship between parent ages and the ages of their children from the Vital Statistics. The ages of parents were estimated by equation 5:

$$Ap_i = Ah_i + A(N) \qquad (5)$$

Where Ah_i is the age of householder, $A(N)$ is the parent's age when she gave birth to her Nth child, and Ap_i is the estimated age of the parent of household i.

For example, the current age of the parent is 85 if the current age of householders (first child) is 50, and she gave birth when she was 35 years old. If there are two parents, the age of husband is calculated using $A(1)$, and the wife's age is calculated using $A(2)$. If there is one, the parent's age is calculated using $A(1)$.

Consequently, the ages and genders of all parents were allocated.

2.2.5. Allocation of Other Relatives Information of Householders

There is no information about the ages and genders of other relatives and the relationship between them and their families in the Vital Statistics. Therefore, the ages and genders of other relatives in each age and gender group were calculated using population data by gender and age group in each city, town, or village by the population census (Fig. 7).

First, the number of other relatives was calculated using equation 6:

$$No_i = Nn_i - Nh_i - Ns_i - Np_i - Nc_i \qquad (6)$$

Where Nn_i is household size (the number of residents), Nh_i is the number of householders (= 1), Ns_i is the number of householder's spouses (=0 or

1), Np_i is the number of householder's parents (=0, 1 or 2), Nc_i is the number of children of household *i,* and No_i is the number of other relatives.

Table 5 shows the number of other relatives calculated by equation 6. The total of other relatives in each city, town, and village could be calculated to totalize these values in all households. Population rates by gender and age group (2) could be calculated in each city, town, and village because the number of people by gender and age group could be monitored by the information shown in Fig. 7.

Second, population by gender and age group in each city, town, and village were calculated by multiplying (1) by (2).

Finally, they were randomly allocated to buildings with households that had other relatives.

Consequently, estimated population data that can monitor locations, family types, and the number of residents, including the age and gender of all households (i.e., the "Micropopulation Census"), becomes available. Fig. 8 shows the Micropopulation Census developed in this study.

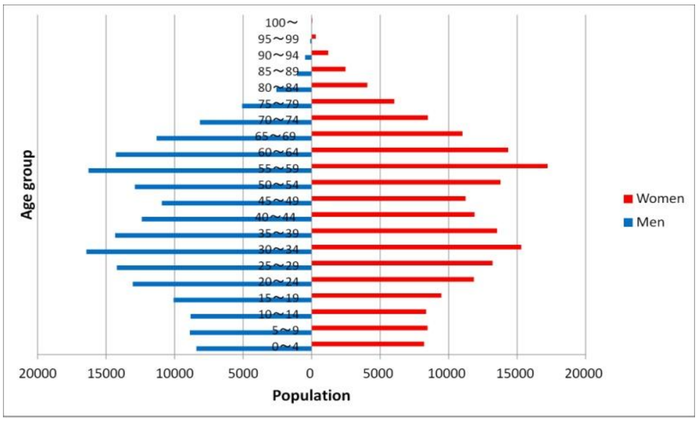

Fig. 7. Population by gender and age group in 2005 (Kashiwa city, Chiba)

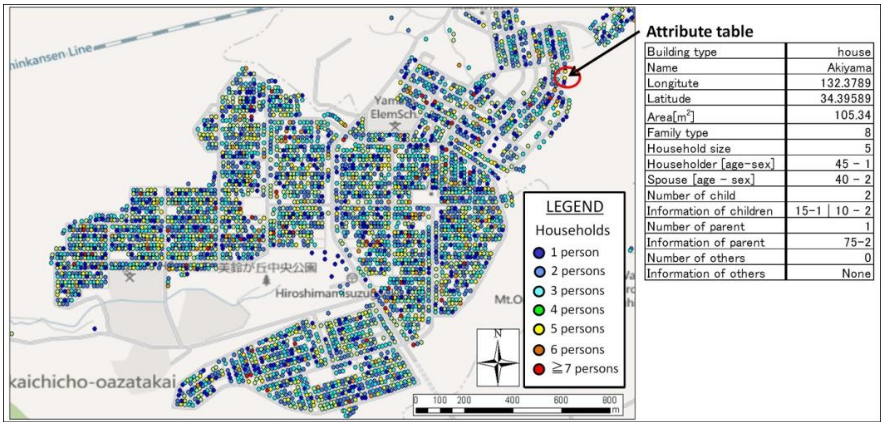

Fig. 8. This is an example of the Micropopulation Census. Information about all households and residents can be monitored using longitude and latitude.

3. Verification of the Reliability of the Micropopulation Census

In this study, we verify the reliability of the Micropopulation Census in Kashiwa city, Chiba prefecture, Japan (Fig. 9). Kashiwa city is located on the outskirts of Tokyo. This city developed in the 1970s as a bedroom suburb of Tokyo thanks to its advantageous location. In addition, a new railroad line recently opened and areas are developing along this railway line. However, there is evidence of depopulation and an aging population in the suburbs. We decided to use this city as a test field because of its various features.

The method of verifying the reliability was to compare the aggregated values of the Micropopulation Census developed to disaggregate the population census of Japan, which is aggregated by the 500m square grids with true values of the population census (in turn aggregated by other kinds of spatial units). Other aggregating units are an administrative unit called “Cho-cho-moku” (287 zones in Kashiwa city) and basic units (4821 zones in Kashiwa city). Fig. 10 shows administrative units and basic units. We compared the Micropopulation Census aggregated by these units with true values.

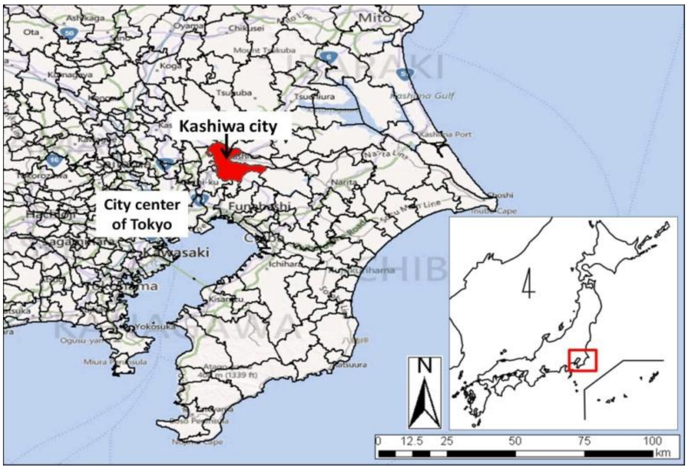

Fig. 9. Location of Kashiwa city

3.1. Verification of Reliability (in Cases of Aggregation by Administrative Units)

3.1.1. Verification of Reliability of Total Numbers of Households and Residents

Fig. 11 shows the results of a comparison between the numbers of households and residents by the aggregated results of the Micropopulation Census from the census of Japan. Both determination coefficients are over 0.97, which means the aggregated results of our data have high correlations with the population censuses. These results show that our data can reproduce the true values of the numbers of households and residents by successfully aggregating them into an administrative unit.

Fig. 10. This map shows some administrative units and basic units. An administrative unit is a zone based on an actual address (city districts or blocks) and a basic unit is a zone that includes 10 to 50 households.

In addition, the values of the population census tend to be larger than aggregated. This is because that room information of condominium buildings or apartment houses could not be collected adequately in zones where these kinds of buildings.

3.1.2. Reliability Verification of Total Numbers of Each Household Size

Fig. 12 shows results of comparison between numbers of household in each household size by aggregated results of the Micro Population Census with them by the population census of Japan. All results have high correlations especially in two-person household, three-person household, and four-person household. These results show that our data can reproduce true values of numbers of household in each household size successfully to aggregate into administrative unit.

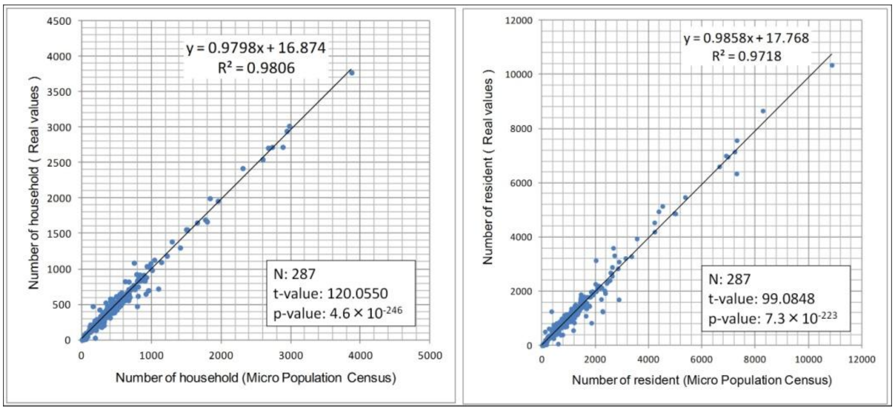

Fig. 11. Correlation of the number of households (left) and residents aggregated by administrative units (right)

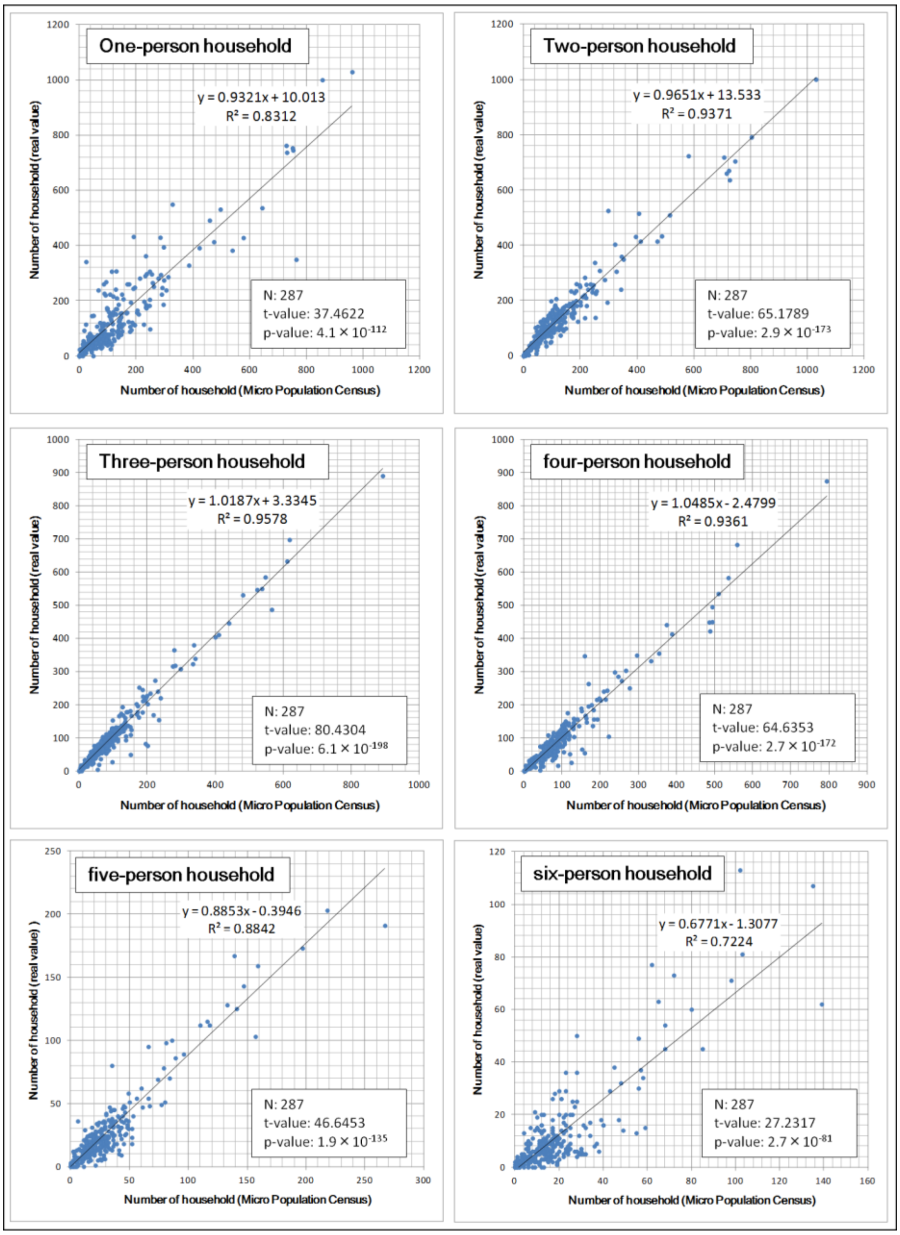

Fig. 12. Correlation of the number of each household size

3.2. Verification of Reliability (in Cases of Aggregation by Basic Units)

The aggregated results of our data by basic units were compared to real values. Fig. 13 shows the map of basic units in Kashiwa city. The population census aggregated by basic units are opened only numbers of households and residents in each unit. Therefore, we compared their aggregated result with the Micropopulation Census and the population census.

Fig. 14 shows the results of a comparison between the numbers of households and residents by the aggregated results of the Micropopulation Census with them by the population census. Both determination coefficients were over 0.8, which means that the aggregated results of our data correlated highly with the population censuses. These results show that our data can reproduce the true values of the numbers of household and residents and even aggregate them into a basic unit (i.e., the minimum spatial-aggregating unit).

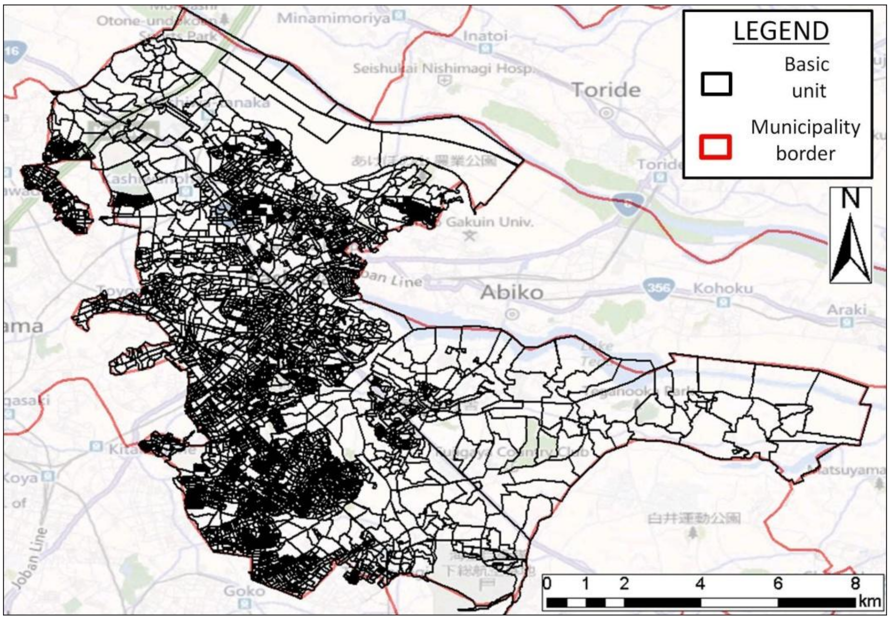

Fig. 13. Map of basic units in Kashiwa city (2005)

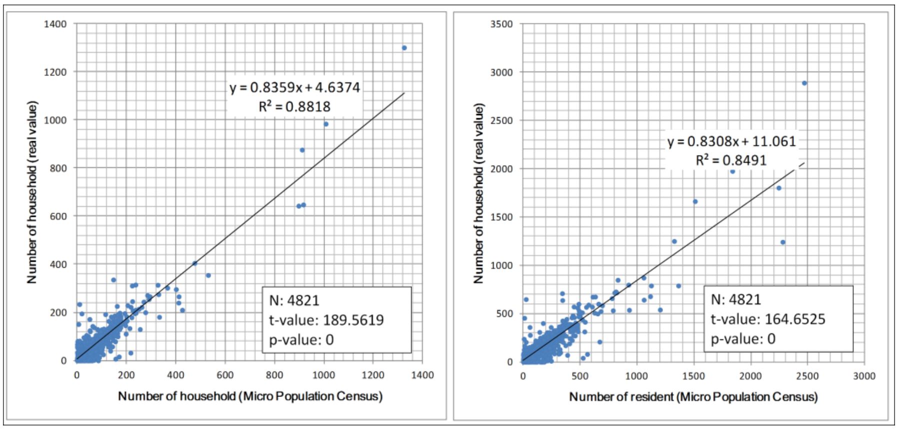

Fig. 14. Correlation of the number of households and residents aggregated by basic units

3.3. Error Verification between Multi Micropopulation Censuses

The development results of the Micropopulation Census changed with each processing because they were developed to randomly allocate various attributes to the residential map. Fig. 15 shows the development results of the first, 250th, 500th, and 750th attempts in cases where we develop 1000 kinds of data. Evidently, the development results are different. The numbers in each basic unit are the aggregated results, which are similar. If there are large errors between the aggregated results of our data and the real values of each data development, it is possible to develop the data with low reliability.

Therefore, we developed our data 1000 times and aggregated them into an administrative and basic unit using the method in 3–1 and 3–2. Finally, we compared the coefficient of determinations, t-values, and p-values between every 1000 times acquired by correlation analyses with real values.

First, Fig. 16 shows the lists of the determination coefficient among all 1000 data with real values. We analyzed the numbers of households of each size in cases of aggregation by administrative unit, and the number of residents in cases of aggregation by basic unit. Each result showed there were no large errors in the 1000 attempts.

Second, Fig. 17 shows the list of t-values. The t-value is a significance level of 0.1% is 3.29 when the degree of freedom is 1,000. The t-values in Fig. 19 and Fig. 20 are much larger than it. Consequently, all results were sufficiently significant.

Finally, Fig. 18 shows the list of p-values. These results show that all p-values in each aggregation are quite small. Consequently, all 1000 results were sufficiently significant.

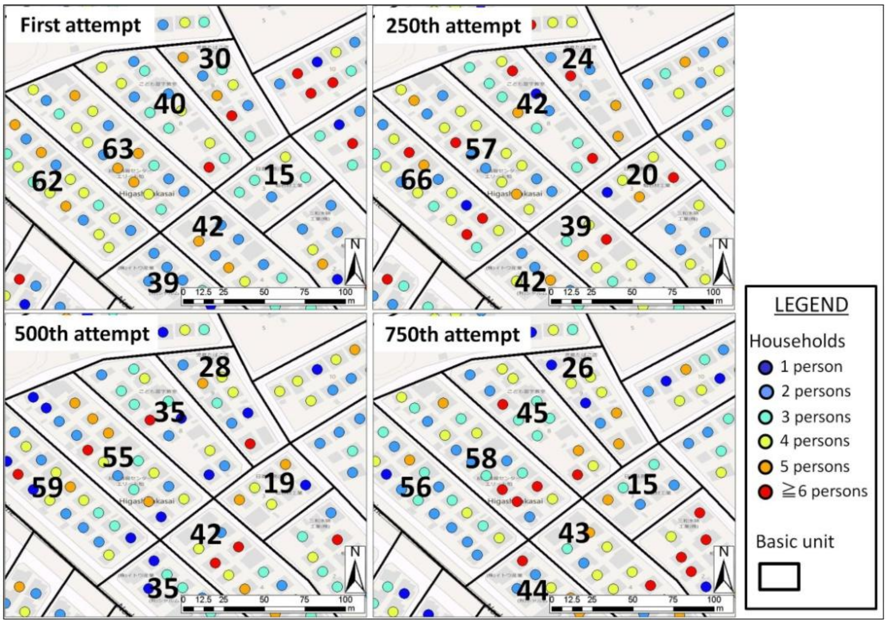

Fig. 15. Differences in the development results of our data in the same area

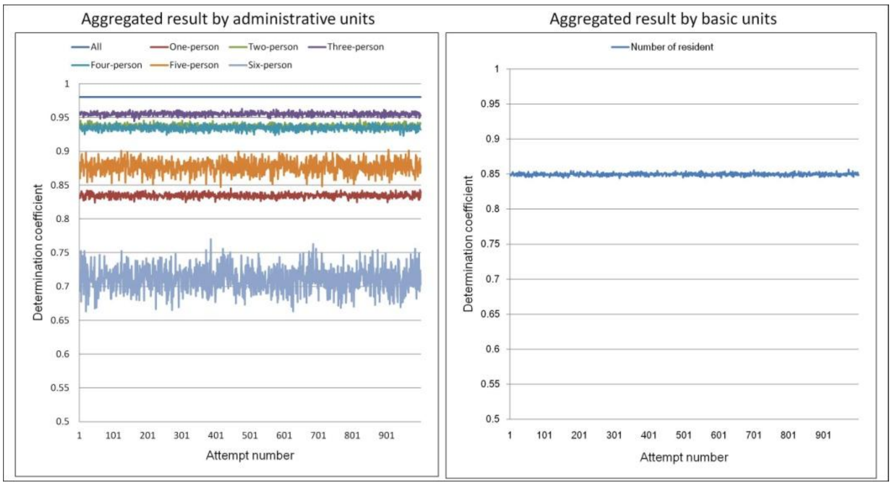

Fig. 16. List of determination coefficients aggregated by administrative units and basic units

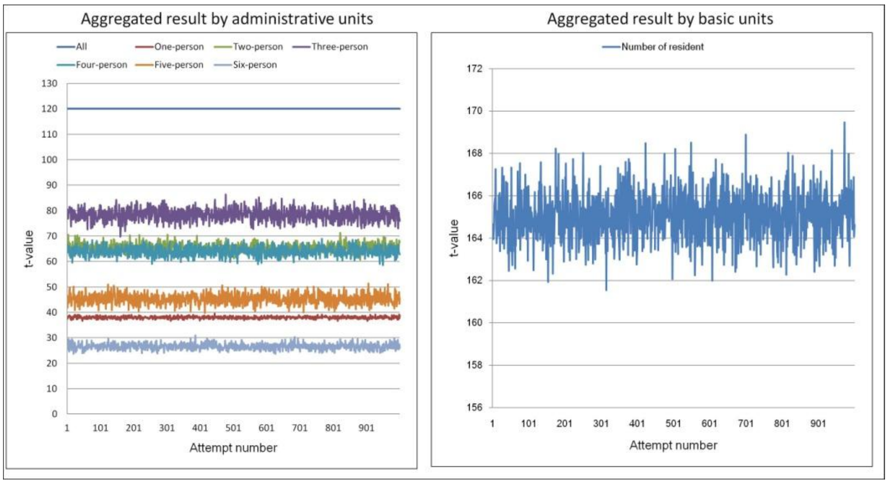

Fig. 17. List of t-values aggregated by administrative units and basic units

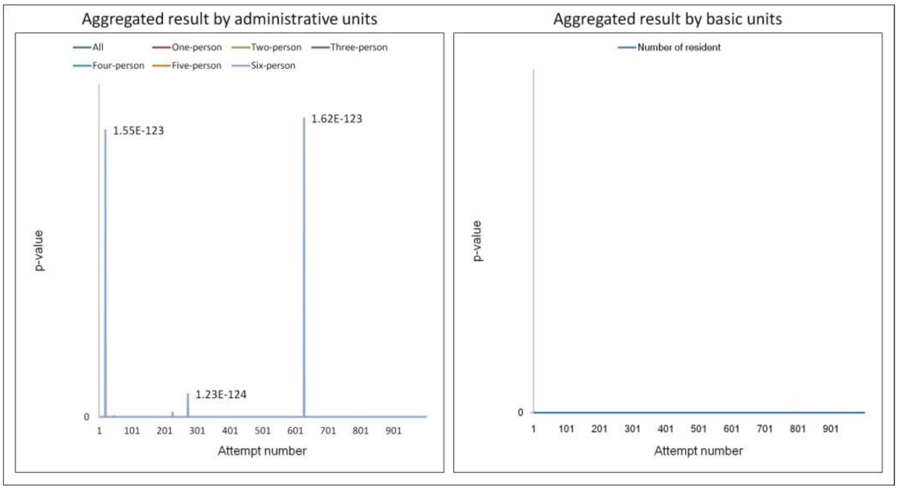

Fig. 18. List of p-values aggregated by administrative units and basic units

4. Result

We will demonstrate the possibilities of using the Micropopulation Census to address some concrete problems.

4.1. Detection of Aging Areas

Japan’s population is rapidly aging because of a declining birthrate. Detailed monitoring of aging areas in Japan is very significant in solving the problem. Fig. 19 and Fig. 20 show a 250 square-meter grid map of the population’s aging rate (more than 65-year-old population/total population) in the Pacific belt zone (i.e., from Tokyo to Osaka area). Fig. 19 shows many aging areas in rural and mountainous areas. In addition, Fig. 20 shows there are aging areas in the center of Tokyo and in its suburbs. In contrast, the aging rate in some cities, such as Kawasaki city and Urayasu city, is less rapid. Many young married couples and young couples with children are moving to these areas because of their proximity to the center of Tokyo and their development as new residential areas. We can monitor detailed regional structures of population by this grid map. This result is possible to contribute to monitoring of aging areas in Japan.

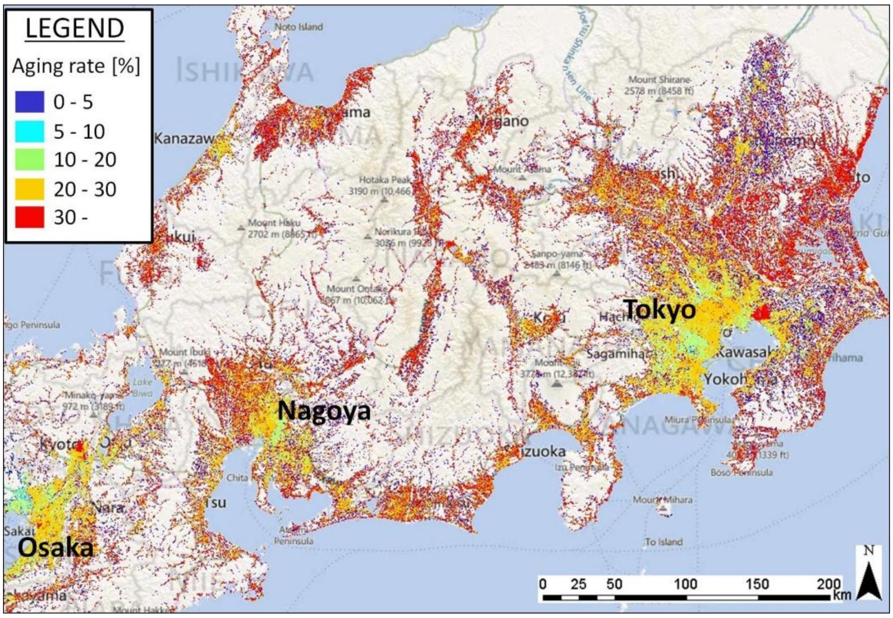

Fig. 19. 250 square-meter grid map of the population’s aging rate in the Pacific belt zone

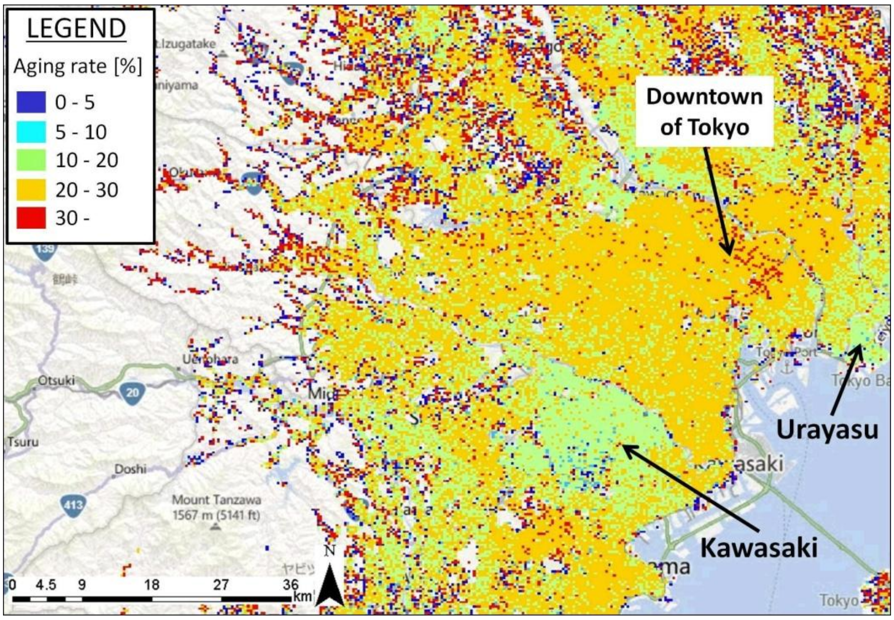

Fig. 20. 250 square-meter grid map of the population's aging rate in the Tokyo area

4.2. Analysis of the Shopping Weak

In recent years, the number of people who find it difficult to make a daily purchase, referred to as the "shopping weak," is increasing because of aging and the closure of grocery shops in their neighborhoods (Komaki (2010). It is difficult to monitor their distribution. Previous studies have estimated their number using statistical data or collecting information using field surveys. Therefore, it is difficult to monitor their distribution in a broad area.

However, the Micropopulation Census can monitor the spatial distributions of aging residents because our data includes the age and household location of all residents. The distance between residents and their nearest grocery shop can also be calculated because all residents have a quantitative location (i.e., longitude and latitude). Consequently, we can acquire reliable results about those who are shopping weak. Fig. 21 shows an example that detects aging singles that are shopping weak in the outskirts of one of Hiroshima city's residential areas. In addition, this method can be adapted for a broad area (Fig. 22).

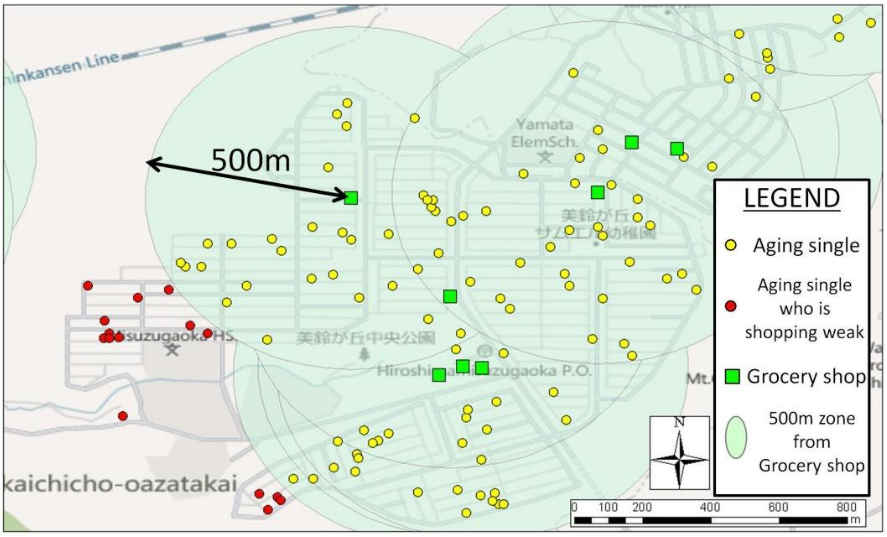

Fig. 21. Detection of aging singles that are shopping weak (in the outskirts of a residential area in Hiroshima)

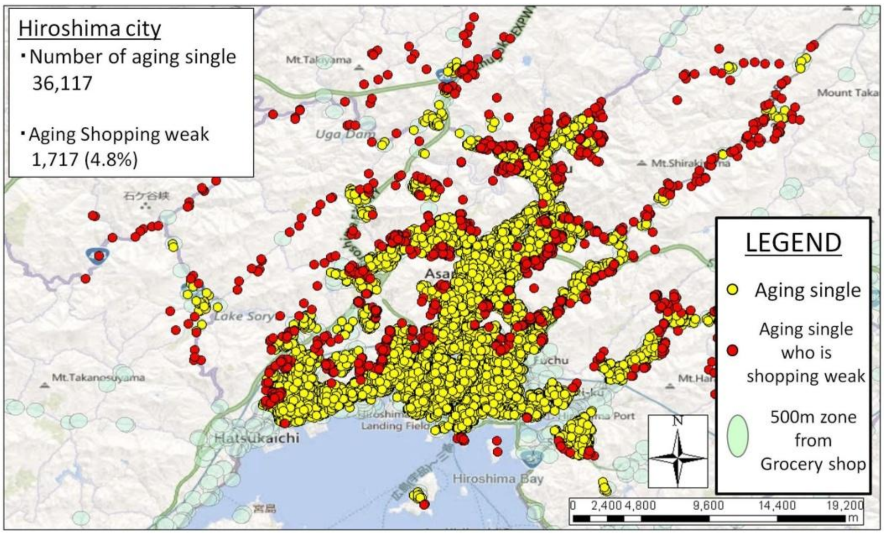

Fig. 22. Distribution map of aging singles that are shopping weak in Hiroshima city

4.3. Utilization Possibilities in Other Fields

It is expected that our data could be utilized to support welfare and health policies, marketing, or disaster and crime prevention because it can estimate the attributes of all residents by their locations.

At present, Japan's open censuses are aggregated into administrative areas or grids (1km or 500m). There are no higher-resolution open population censuses. It is possible to understand the spatial distribution of the aging shopping weak mentioned in 4–2, and the extent of the human suffering caused by a tsunami (Fig. 2) by monitoring the distributions of residents with specific attributes. In addition, our data can be used for the planning of facility placement and marketing support because it can estimate the age and gender structures of residents in arbitrary zones of specific locations.

There is a growing need for a detailed population census. This study is one of the foundations of a response to this need. We are working toward the practical use of our data by gradually improving its reliability.

5. Conclusion and Future Works

This study develops detailed population data of household locations in a "Micropopulation Census" to disaggregate the census to the residential map. This data can be aggregated into an arbitrary spatial unit because it is disaggregating point data. For example, the population data in Fig. 19 is the first grid population data aggregated by a 250 square-meter grid in Japan. In addition, we show the versatility of our method to develop this data in broad areas. Consequently, the Micropopulation Data indicates the possibility of a new population census.

There are some future objectives. First, we plan to improve the method, which currently integrates new census data based on common attributes in a step-by-step manner. It is expected that our data will more reliably increase the conjuncture among all allocating data. We plan to do this by adopting the Monte Carlo method based on the rates of each household type or the age and gender of residents.

We should verify the reliability of our data in other areas. It is not clear at this time that the reliability of our data is the same as the level of the results for Kashiwa city, even though we can develop it throughout Japan. We plan to verify its reliability in areas whose characteristics differ from Kashiwa city.

In addition, the most creative work is to use our data to resolve concrete problems. Our data have already been used to estimate disaster damage and to analyze the shopping weak in relation to the distribution of the elderly. We will improve the method for developing our data and attributing information by using our data to resolve concrete problems and conduct future research by using our data to resolve concrete problems and contribute to further research in future.

Acknowledgement

The authors were given a digital residential map during the joint research with Center for Spatial Information Science (CSIS), the University of Tokyo (research ID: 122), and we were supported by the Earth Observation Data Integration and Fusion Research Institute (EDITORIA). The authors would like to thank CSIS and EDITORIA for their contribution.

References

Asahara, M. and Matsumoto, Y.: Japanese Named Entity Extraction with Redundant Morphological Analysis. HLT-NAACL 2003, pp.8-15 (2003)

2. Asano, J.: A Study on the Actual Condition of Big Shopping Stores' Closing and the Tendency of Its Uses After the Closing in the Central Areas of Local Cities ~From Researches to Municipalities in Hokuriku-Koshin-Etsu Area~. Architectural Institute of Japan. Planning. Environment, No. 557, pp. 257—264 (2006) (in Japanese)
3. Establishment and Enterprise Census 2006, Ministry of Internal Affairs and Communications of Japan, http://www.stat.go.jp/english/data/jigyou/index.htm
4. Geocoding and Utilities, Center for Spatial Information Science, The University of Tokyo, http://newspat.csis.u-tokyo.ac.jp/geocode
5. Itai, Y., Yatabe, M. and Kamijima, S.: Research and Valuation on Space Use on Each Floor of Buildings at Mixed Area with Dweling and Factory Around Kawaguchi Station, Theory and Applications of GIS, Vol. 2, No. 1, 57-63 (1994) (in Japanese)
6. Ito, K. and Magaribuchi, H.: Method for Making Spatio-Temporal Data from Accumulated Information: Using the Identification by Resolving Geometric and Non-Geometric Ambiguity, Papers and Proceedings of the Geographic Information Systems Association, Tokyo, Japan, 23, September 2001. (in Japanese)
7. Jeffrey E. F. Friedl.: Mastering Regular Expressions, Third Edition, O'Reilly Media (2006)

8. Kondo, M.: An Analysis of Japanese Classical Literature Using Character-based N-gram model, The journal of humanities, Chiba Univ, vol. 29, 187-238 (2000)
9. Matsuura, T. and Motohashi, K.: Market Dynamics and Productivity in Japanese Retail Industry in the late 1990's, RIETI Discussion Paper Series, 05-E-001 (2005)
10. Osada, T., Akiyama, Y., Shibuki, T. and Shibasaki, R.: A study on Location Changes of Business Establishments by Extensive Urban Dataset and Its Accuracy Validation. The KAGIS 10th Anniversary International Conference, Daegu, Korea, 23, October 2007
11. Sagara, T., Makino, T., Kawaguchi, S., Ozawa, H. and Kituregawa, M.: A Method to Normalize Shop Names Using Their Address Descriptions for Accuracy Improvement of Shop Information Search form the Web, DEWS2006, 2C-o1, 392-398 (2006) (in Japanese)
12. Shibuki, T., Akiyama, Y. and Shibasaki, R.: A Method of Monitoring 3d Change of Shop and Office Tenants by Time Series Spatial Integration of Digital Map and Yellow Page, Journal of Architecture and Planning, Transactions of AIJ, Vol. 73, No. 626, 789-793 (2008)(in Japanese)