Projecting Households of Synthetic Population on Buildings Using Fundamental Geospatial Data

Takuya Harada * and Tadahiko Murata **

Abstract: In this paper, we propose a method to project households of synthetic population using fundamental geospatial data for real-world social simulations. That is, we assign each generated household on a building in a geographical map. When we try to conduct a real-scale social simulation, we need attributes of agents and their locations on a geographical map. We have already proposed a synthetic population method that generates attributes of agents or citizens from the statistics of the real world. To determine the locations of agents, we propose, in this paper, a threefold method to project generated households on buildings in a geographical map using the fundamental geospatial data. We apply the proposed method to project households generated from the statistics of Takatsuki City, Osaka, Japan and project them on buildings in the map. In order to cope with a problem of random assignment of households on buildings, we propose a modified method to consider types and area of buildings. Projection results show that households are assigned more reasonably to isolated houses and apartments.

Key Words: agent-based simulation, synthetic reconstruction method using statistics, fundamental geospatial data.

1. Introduction

Computational Science has been regarded as the third leg of science and accepted as a complement of traditional legs or pillars of theory and experimentation [1]. It is a growing multi- and inter-disciplinary field including science (biological, physical, and social), engineering, medicine, and humanities. Among them, social simulation has attracted attentions as one of the promising research fields. In this paper, we propose a method to project households of synthetic population using fundamental geospatial data for real-world social simulation.

When we develop a tool based on microsimulation or agent-based simulation [2],[3], we should prepare individual agents in the simulation tool. After defining properties and behaviors of agents, we can observe some consequences of a population through the simulation tool. Using several parameters in the simulation tool, we can find some extreme consequences that should be avoided or should be encouraged. Examining detail behaviors or connections among agents in such extreme cases allows us to find some important conditions or requirements to avoid or obtain such extreme cases. Recently social simulations try to show some concrete consequences in a real community with agents or citizens. Properties or characteristics of those agents should be prepared for such simulations.

First method to generate a synthetic population from statistical data was proposed in 1976 by Wilson and Pownall [4]. They proposed a method called a synthetic reconstruction method (SR method) that tries to reconstruct individual and household data from statistics and some real sample data using an iterative proportional fitting procedure (IPFP) [5]. Several succeeding algorithms are proposed based on the SR method. Barthelemy and Toint [6] indicated that the SR method has a difficulty to reconstruct populations fitting both the statistics of individuals and that of households simultaneously. In order to cope with this difficulty, Gargiulo et al. [7] and Barthelemy [6] proposed reconstruction methods not using sample data. Lenormand and Defuant [8] compare the reconstruction method without sample data and the SR method and show that the former can reconstruct a better population.

Essentially, the correctness or the accuracy of generated households cannot be confirmed since the real compositions of population in an area are not available for security reason. Besides, the number of available statistics is very few to generate a number of households. There must be multiple compositions that satisfy the available statistics. However, the need of real household compositions is growing as a technique of social simulations for the real world with households.

For example, Ichikawa et al. [9] tried to implement a real-scale social simulation on influenza spreads in Izu-Oshima Island, Tokyo, Japan. They employ an SR method without samples [10] to reconstruct a population with about 8,000 people and 5,000 houses in the island. They apply their infection model on the reconstructed population to compare two cases to see how long the infection continues in each case. In their model, they prepare a virtual city model for Izu-Oshima Island. Their virtual city model has four layers: the first layer corresponds to a city or Izu-Oshima Island itself, the second layer includes six main areas in the island, the third layer has 100 small areas in six main areas, and the fourth layer corresponds to 6,000 elements such as homes, schools, working places, and hospitals where each citizen has its own activity. They conduct their simulation using the synthesized population and the virtual structures of the island. Although they find interesting simulation results using their virtual model, they can conduct more realistic simulations by projecting their population on the map of the real island. Conducting more realistic simulation...
with the physical conditions such as capacities of roads and buildings on the real map enables them to find more concrete phenomena for influenza spreading.

Although the need of reconstructed populations is increasing, still the criticism against the accuracy of synthesized population remains. Any method that reconstructs or synthesizes populations from statistics cannot assure the accuracy of populations that are reconstructed. However, we can see two advantages of using reconstructed populations. First, a reconstructed population does not harm privacies of citizens those who live in a target area of simulations since the reconstructed population is not the same as the real population. Second, simulations using reconstructed populations can be tools that find potential danger or hidden success in a population that has the same statistical characteristics of the real population. Even if some reconstructed households do not exist in the real population, simulations with such households can show advantages and disadvantages with those households in the community. Governments can consider their measures to encourage or discourage to have such households in their policy making.

In this paper, we employ a reconstruction method without sample data proposed by Ikeda et al. [10]. Murata and Masui extended the method by modifying an objective function in the method and including a heuristic to improve directly some objective functions [11],[12]. They proposed their methods to minimize the errors of statistics of a generated population. However, they reconstructed a small-case population with the same statistical tendency. That is, they consider only 500 or 1,000 households. We have proposed a method [13] to reconstruct a real-scale population that includes 350 thousand households and about a million individuals. Now then, we try to project them on buildings in a real map.

Section 2 explains a proposed method to project a reconstructed population on a map. Results of projections are shown in Section 3. In Section 3, we find unreasonable assignments by projecting households randomly to buildings. We show the effectiveness of the proposed method that considers building area in projecting households. Finally Section 4 concludes this paper.

2. Proposed Method

In order to assign generated households on buildings in a map, we propose a threefold simulated annealing-based method as follows:

Step 1: Synthesize households in a city, town, or a village (a unit of a local government) using a simulated annealing method (SA1).
Step 2: Calculate the number of households by a district using another simulated annealing method (SA2).
Step 3: Assign households into a district using the other simulated annealing method (SA3).
Step 4: Assign each household on a building in the corresponding district.

As shown in the above procedure, we assign each household to a building in a map after applying simulated annealing methods three times. The following subsections explain each step.

2.1 Simulated Annealing

Simulated annealing [14] is known as one of the effective optimization algorithms for combinatorial optimization problems. To find global optimum solution, it starts with a randomly created initial solution; then it creates a neighborhood solution from the current solution. If the created neighborhood solution is better than the current one, the transition from the current one to the neighborhood is accepted. If the neighbor is worse than the current one, the transition probability is considered. That is, the transition is probabilistically accepted even if the neighbor is worse than the current one. The transition probability is controlled by the number of search iterated by the algorithm. In the early stage of the search, the transition probability is set relatively high, but the probability becomes low as the search runs. In order to design a simulated annealing algorithm to a specific problem, it is important how to design the expression of solutions, the neighborhood structure, the transition probability control, and so on.

2.2 Synthesizing Households in a Local Government

We synthesize households using the simulated annealing based synthetic reconstruction method (SA-based SR method) without samples [13]. Figure 1 shows an example of reconstructed households. We reconstruct nine family types shown in Fig. 2. These nine types are employed in the real statistics [15]. In Japan, these nine family types cover 95% of the entire population. Other family types are such as households with other relatives, households only with brothers and sisters, households with non-relatives, and others. Those households not in the nine family types are difficult to be reconstructed from the available statistics. For example, there are no statistics about other relatives in the same households. Although statistics about age difference among brothers and sisters are available, there are no statistics to estimate the age of a family member of such households only with brothers and sisters. Due to these statistical limitations, we consider only nine family types in Fig. 2.

As shown in Fig. 1, each household has its family members. Each family member has its attributes such as age, sex, household type, role in household, and kinship. The role in household shows a role of the member in its household. For example, a family member in a household of parents and children can
have a role of husband, wife or child. The kinship is an attribute that indicates a relation with other family members such as “husband–wife” or “parent–child” in a household of parents and children.

In the reconstruction method [13], the 21 statistics of age differences among some members in a household, male and female demographic pyramids in each family type are employed to reconstruct or synthesize a population with the same size of the real population in a local government such as city, town, or village. The following objective function for a real-scale population is used to minimize the errors in statistics employed for synthesizing a population:

\[ f_i(A) = \sum_{j=1}^{G_s} |c_{ij}(A) - R_{sj}|, \]  

where \( A \) is a reconstructed population, \( s \) is an ID of statistics among the above nine statistics, \( G_s \) is the number of terms in Statistics \( s \), \( c_{ij}(A) \) is the number of members for the \( j \)-th term of Statistics \( s \) in the reconstructed population \( A \), and \( R_{sj} \) is the real data of the \( j \)-th term of Statistics \( s \) in the real statistics.

The procedure of a simulated annealing (SA) method in [13] to minimize total error of the 21 statistics is shown as follows:

**[Procedure of SA1 (See [13] for details)]**

Step 1-1: Initialize reconstructed population probabilistically using statistics.

Step 1-2: Terminate the procedure at the prespecified termination condition.

Step 1-3: Change probabilistically an age of a randomly selected agent according to the ratio of people in demographic pyramids by age.

Step 1-4: Evaluate the generated solution using the 21 statistics and decide the transition.

Step 1-5: Update parameters.

Step 1-6: Go to Step 1-2.

### 2.3 Estimating the Number of Households in a District

Because the statistics for the number of households by family type and the number of family members are not available for district level in a local government, the number of households in each district should be calculated. In order to calculate it, we employ these three statistics by a district in each local government:

1: The number of households by the number of members in a household \(^1\).
2: The number of households by the family types \(^2\).
3: The number of population by the family types \(^2\).

These three statistics are prepared for each district in each local government in Japan. Tables 1 and 2 show an example of these three statistics in District \( p \). In order to find a combination that satisfies these three statistics simultaneously, we employ another SA method (SA2) using the following objective function \( g_s(x) \) in a local government:

\[ g_s(x) = \sum_{j=1}^{G_s} \sum_{p=1}^{P} |d_{sj}(x) - R_{sj}|, \]  

\(^1\) List 5 in [16].

\(^2\) List 6 in [16].

| Table 1 | Example of the number of households by the number of members in a household in District \( p \) of a local government. |
|---------|--------------------------------------------------|
| \# of members in a household | \# of households \( \left( R_{s1}^p \right) \) |
| 1 | 210 |
| 2 | 346 |
| 3 | 228 |
| 4 | 191 |
| 5 | 57 |
| 6 | 5 |
| 7 or more | 3 |

| Table 2 | Example of the number of households and population by the family types in District \( p \) of a local government. |
|---------|--------------------------------------------------|
| The family types | \# of households \( \left( R_{s1}^p \right) \) | \# of population \( \left( R_{s2}^p \right) \) |
| Single | 404 | 404 |
| Couple | 164 | 328 |
| Couple with children | 147 | 525 |
| Single parent with children | 39 | 90 |
| Households other than nuclear family | 18 | 73 |
| Households with non-relatives | 5 | 12 |
| Others | 1 | 2 |

where \( s \) is an ID of statistics, \( x \) is a control variable vector, \( G_s \) is the number of terms in Statistics \( s \), \( P \) is the number of districts in the local government, \( d_{sj}^p \) is the number of members of the \( j \)-th term of Statistics \( s \) for District \( p \) in the synthesized population, and \( R_{sj}^p \) is the real data of the \( j \)-th term of Statistics \( s \) in the real statistics.

Each element of the control variable vector \( x \) is \( x_{im}^p \), that denotes the number of households with \( m \) members in a household of the family type \( t \) in District \( p \). Therefore the number of elements of the control vector \( x \) is \( M \cdot T \cdot P \), where \( M \) is the number of members in the statistics (7 in Table 1), \( T \) is the number of family types (7 in Table 2), and \( P \) is the number of districts in a local government. We employ the following procedure of SA to minimize the sum of the objective function (2) for the three statistics.

**[Procedure of SA2]**

Step 2-1: Randomly initialize the control vector \( x \).

Step 2-2: Terminate the procedure at the prespecified termination condition.

Step 2-3: Randomly select \( x_{im}^p \), \( x_{im}^p \), \( x_{im}^p \) and \( x_{im}^p \) from districts \( p \) and \( q \).

Step 2-4: Operate \( x_{im}^p ++, x_{im}^p --, x_{im}^p --, x_{im}^p + + \), where ++ and -- are the increment and decrement operators, respectively.

Step 2-5: Evaluate the updated control vector \( x \) using the objective function (2), and decide the transition.

Step 2-6: Update parameters.

Step 2-7: Go to Step 2-2.

In Step 2-1, we employ the number of households by family type in Table 2 as initial value of the number of households in each district. In this procedure, we modify the values of four elements of \( x \) selected in Step 2-3 not to change the total number of households and population in a local government.
Step 3-1: Randomly assign each household into a district.

Step 3-2: Terminate the procedure at the prespecified termination condition.

Step 3-3: Randomly select \( h_{pq}^p \) and \( h_{pq}^q \) from Districts \( p \) and \( q \). \( h_{pq}^p \) and \( h_{pq}^q \) are households of the same family type with the same number of members. Then exchange them.

2.4 Assigning Household into a District

In order to assign each household generated by SA1 into a district according to the number of households determined by SA2 in Subsection 2.2, we employ the following statistics as shown in Table 3.*

1: The male population by age in District \( p $^3\).
2: The female population by age in District \( p $^3.

The following objective function is employed to minimize the errors between the synthesized population and the statistics:

\[
h_s(A) = \frac{1}{G_s} \sum_{j=1}^{G_s} \left[ e_{sj}(A) - \text{Round} \left( R_{pj}^s, m_{pj}(A) \right) \right],
\]

where \( G_s \) is the number of terms in Statistics \( s \) of District \( p \), \( e_{sj}(A) \) is the number of members for the \( j \)-th term of Statistics \( s \) in the reconstructed population \( A \) of District \( p \), \( R_{pj}^s \) is the ratio of the \( j \)-th term of Statistics \( s \) of District \( p \), and \( m_{sj}(A) \) is the number of members in the reconstructed population \( A \) of District \( p \). \( R_{pj}^s \) is calculated by the following equation using the value of \( R_{pj}^s \) in the statistics,

\[
R_{pj}^s = \frac{G_s}{\sum_{k=1}^{G_s} R_{pk}^s},
\]

where \( R_{pk}^s \) is a real value in the statistics (e.g., each value in Table 3). Since these values include males and females from all family types that include the other types than the nine family types in Fig. 2. In order to exclude people in households of the other types, we adjust \( R_{pj}^s \) by the following function.

\[
\tilde{R}_{pj}^s = R_{pj}^s - \sum_{t=1}^{T} \left( \text{pop}_{j, t} \right),
\]

where \( \text{pop}_{j, t} \) is the number of population of the \( j \)-th term in the family type \( t \) of the nine family types (\( T = 9 \) in the above equation), and \( \text{pop}_{j, t} \) is all the population of the term \( j \). The efficiency of this adjustment is shown in Section 3.

Note that we minimize the sum of the two statistics for each district using the objective function (3). We use the following procedure to assign each household into a district.

[Procedure of SA3]

Step 3-4: Evaluate the generated solution using the population by sex.
Step 3-5: Update parameters.
Step 3-6: Go to Step 3-2.

In Step 3-1, we randomly assign each household generated by SA1 according to the number of households by family type and the number of family members in each district obtained by SA2.

2.5 Assigning Household on a Building

After assigning each household into a district, we assign each household on a building in a map. We employ the fundamental geospatial data [17] released from the Geospatial Information Authority of Japan. The fundamental geospatial data includes the following information:

1: Geodetic control point,
2: Coastline,
3: Boundary of public facilities (river management boundary),
4: Boundary of public facilities (river management boundary),
5: Administrative boundary (town level; with a point in each polygon),
6: Road edge,
7: Riverside edge of levee crown,
8: Railroad track centerline,
9: Elevation (ground surface point where the elevation is known),
10: Shoreline,
11: Building outline,
12: Community boundary (with a point in each polygon),
13: Street block boundary (with a point in each polygon).

As shown in the above information, the fundamental geospatial data include the data of outline of each building (the 11th item) in the territory of Japan. These data include not only the latitude and longitude of circumference of each building, but also the information about the types of buildings such as Table 4.

In order to find buildings for residence, we select buildings of ordinary buildings, durable buildings, and high-rise buildings in Table 4. To one of those buildings we assign each household. We allow each building to have several households since some buildings are apartments or flats where several households live together in the same building.

---

* List 2 in [16].
* List 16-1 in [15].
* List 3-2 in [15].
When assigning each household to a building, we randomly select one building in the district where a household resides. As we see that the 13 items of information are included in the fundamental geospatial data, the administrative boundary is shown in the data; however, each building does not have information on the administrative district in a city or a town. Therefore, we assign such an attribute using the boundary data for districts in Population Census in 2010. We develop a tool that automatically adds a district information to each building by finding where the center of gravity of the building locates the inside or outside of the boundary of a district.

Random selection of building in a district may cause the following problems.

1: Assigning a household to too small building where any household cannot live,
2: Assigning a household averagely to each building with any area of space.

In order to assign households by types of buildings, we try to connect family types and types of buildings. Since there is no statistics that connects family types with types of building, we employ types of ownership of houses to connect these two attributes using four kinds of statistics:

1: The number of households by family types and types of ownership of houses.
2: The number of household members by family types and types of ownership of houses.
3: The number of household by types of building.
4: The number of household members by types of building.

Table 5 shows ownership of houses by households. There are five types of ownership of living places. Using these four kinds of statistics, we try to find the number of households by family types and types of buildings. We employ another SA method (SA-Adjustment) that minimizes the sum of errors of the above four kinds of statistics to add attributes of ownership type, and attributes of building type to reconstructed households by SA1.

When we assign each household to each district using SA3, we add other statistics to the statistics in Table 3. Since each household has the attribute of house ownership and the attribute of building type after the SA-Adjustment is applied. We employ the following four statistics:

1: The number of household by house ownership.
2: The number of household members by house ownership.
3: The number of household by types of building.
4: The number of household members by types of building.

The types of building have eight categories that are an isolated house, a row house, four types of apartments, business building and others. These statistics are available for each district in a local government in Japan.

Finally each household is assigned to each building in a district according to the construction area of each building. The construction area of each building can be calculated using the outline of each building shown in the fundamental geospatial data. We heuristically define the minimum and maximum areas of each type of building as shown in Table 6. The last row of Table 6 is explained later.

### 3. Results of Projection

We project households of Takatsuki City in Osaka, Japan using the fundamental geospatial data. Synthesized population of Takatsuki City is 139,749 households, and its population is 336,891. In the following computing processing, we employ a computer environment with Intel Core i7-3930K (3.2 GHz, 6 cores) and eight memories of DDR3-1600 8 GB.

#### 3.1 Synthesizing Households in Takatsuki City

As the termination condition in Step 1-2, we employ 100,000 evaluations of the objective function (1) per person. Table 7 shows errors between the synthesized population and the real statistics in the objective function (1), and computation time over ten trials using SA1 in Subsection 2.1. Since this is a total error of 21 statistics, the average error per statistics is only 1.33. That is only one or two difference in (1) for each statistics. The last row of Table 7 shows the average error and its standard deviation in the initial population. By comparing the initial error and the optimized error, we can see that SA1 can reduce the error by 99.99% from the initial error.

#### 3.2 Estimating the Number of Households in Each District

As the termination condition in Step 2-2, we employ 1,000 evaluations of the objective function (2) per household. Table 8 shows the average and standard deviations of errors of the objective function (2), and computation time over ten trials using

---

Table 5 Ownership of houses.

| Types of ownership | Description |
|--------------------|-------------|
| Owned house        | The household owns the house. |
| Rented house A     | The household rents a house managed by a government. |
| Rented house B     | The household rents a house managed by a public cooperation. |
| Rented house C     | The household rents a house managed by a private cooperation. |
| Other’s owned house| The household lives in a house owned by others. |

Table 6 Construction area of a building.

| Area             | Isolated house | Row house | 1–2 story apartment | 3–5 story apartment | 6–10 story apartment | 11–14 story apartment | 15 above story apartment | Business building | Others |
|------------------|----------------|-----------|---------------------|--------------------|----------------------|-----------------------|------------------------|-------------------|--------|
| Min (m²)         | 25             | 25        | 25                  | 200                | 300                  | 350                   | 350                    | 100               | 25     |
| Max (m²)         | 200            | 200       | 200                 | 500                | 1,000                | 10,000                | 10,000                 | 25                | 25     |
| # of households  | 114            | 39        | 1                   | 3,067              | 151                  | 0                     | 0                      | 1                 |        |

---

* List 25-2 in [15].
* List 16-2 in [15].

* List 7 in [16].

* List 8 in [16].
Table 7 Errors and computation time for reconstructing population of Takatsuki city, Osaka, Japan using SA1.

| Trial | Error | Computation time (s) |
|-------|-------|----------------------|
| 1     | 30    | 17,072.98            |
| 2     | 36    | 16,304.97            |
| 3     | 22    | 17,381.28            |
| 4     | 36    | 16,172.43            |
| 5     | 30    | 16,979.44            |
| 6     | 24    | 16,336.99            |
| 7     | 30    | 17,289.77            |
| 8     | 20    | 16,632.09            |
| 9     | 16    | 17,203.38            |
| 10    | 36    | 16,295.65            |
|       | Av.   | 28.0                 |
|       | St. Dev. | 7.2                |
| Ini. Av. | 322,154.4 |
| Ini. St. Dev. | – |

Table 8 Average and standard deviation of errors and computation time.

| SA2 | Error Computation time (s) |
|-----|-----------------------------|
|     | Average                     |
|     | 51.20                       |
|     | Standard Deviation          |
|     | 10.08                       |
| Ini. Av. | 19,210.8                   |
| Ini. St. Dev. | –                          |
| Ini. Av. | 322,154.4                  |
| Ini. St. Dev. | –                          |

Table 9 Average and standard deviation of errors and computation time.

| SA3 Method | Method A Error Time (s) | Method B Error Time (s) |
|------------|-------------------------|-------------------------|
| Av.        | 4,533.11                | 1,480.87                |
| SD         | 41.20                   | 33.64                   |
| Ini. Av.   | 89,847.97               | –                       |
| Ini. SD    | 556.15                  | –                       |

Table 10 Errors and standard deviations using SA3.

| Results by SA1 | Method A Error Time (s) | Method B Error Time (s) |
|----------------|-------------------------|-------------------------|
| Trial          | Av.                     | SD                      | Av.                     | SD                      |
| 1              | 30                      | 4,528.8                 | 41.0                    | 1,480.2                 | 32.2                    |
| 2              | 36                      | 4,549.7                 | 53.5                    | 1,483.2                 | 26.2                    |
| 3              | 22                      | 4,512.9                 | 50.3                    | 1,487.5                 | 30.1                    |
| 4              | 36                      | 4,530.5                 | 38.0                    | 1,473.3                 | 28.9                    |
| 5              | 30                      | 4,551.0                 | 33.6                    | 1,473.3                 | 35.2                    |
| 6              | 24                      | 4,525.7                 | 36.7                    | 1,496.3                 | 34.1                    |
| 7              | 30                      | 4,552.6                 | 41.0                    | 1,475.4                 | 38.5                    |
| 8              | 20                      | 4,531.1                 | 40.5                    | 1,465.7                 | 39.2                    |
| 9              | 16                      | 4,511.6                 | 37.6                    | 1,492.2                 | 36.4                    |
| 10             | 36                      | 4,537.2                 | 29.1                    | 1,481.6                 | 36.9                    |
| Average       | 4,533.11                | 41.20                   | 1,480.87                | 33.64                   |

SA2 in Subsection 2.2. Since there are 384 districts in Takatsuki City, the average error per a district is 0.133, that shows the number of households of almost all districts are estimated according to the employed statistics. The last row of Table 8 shows the average error and its standard deviation in the initial population. By comparing the initial error and the optimized error, we can see that SA2 can reduce the error by 99.73% from the initial error.

3.3 Assigning Households to Each District

As the termination condition in Step 3-2, we employ 1,000 evaluations of the objective function (3) per household. Table 9 shows the average and standard deviations of errors of the objective function (3), and computation time over ten trials using SA3 in Subsection 2.3. In the application of SA3, we employ the first result of SA2 for each of ten results by SA1. In Method A, raw data in the statistics are used in (3). On the other hand, adjusted data by (5) are used in Method B. By comparing Methods A and B, we can see that the proposed adjustment is effective to reduce the errors in assigning households to each district. The last row of Table 9 shows the average error and its standard deviation in the initial population for Method A and Method B. By comparing the initial error and the optimized error, we can see that SA3 with Method A and Method B can reduce the error by 94.95% and 98.34% from the initial error, respectively.

Table 10 shows errors and standard deviations over ten trials for each synthesized population reconstructed by SA1 in Table 7. Therefore we conduct 100 trials in total for Tables 9 and 10. Using Welch’s $t$ test, we can see 1% significant difference between Methods A and B. Using Method B, we can reduce errors in assigning household to each district.

3.4 Assigning Households to Each Building

Figure 3 shows the results of projection of synthesized household on the map of Makita-cho, Takatsuki City using the fundamental geospatial data. Polygons drawn by solid line show buildings with walls, and polygons with dotted line show buildings without walls. According to this map, we can see that the synthesized households are assigned only on buildings with walls of a map using the fundamental geospatial data. From this figure, we can see several households are assigned in the same building. Because we assigned synthesized households randomly to each buildings, each building averagely has the same number of households in a district. This seems strange since assigning households to each building is done regardless its construction space or area. We try to solve this problem using the modified assignment method in Subsection 2.5.

3.5 Modified Assignment of Households to Each Building

When we modify the assignment of households to each building, we employ another SA method to add attributes ownership type and attributes of building type to reconstructed households by SA1. The terminate condition of this SA method is 1,000 evaluations of total errors per household. Table 11 shows the average and standard deviation of errors and computation time.
over ten trials by this SA method. The last row of Table 11 shows the average error and its standard deviation in the initial population. By comparing the initial error and the optimized error, we can see that SA-Adjustment can reduce the error by 99.90% from the initial error.

Tables 12 and 13 show the results of reconstruction errors by SA3 using the modified statistics. The last row of Table 12 shows the average error and its standard deviation in the initial population. By comparing the initial error and the optimized error, we can see that Method A and Method B obtain 4.5 times and 12.3 times worse errors by modifying the statistics. This is because the four more statistics are considered in the objective function (3) as explained in Section 2. Besides, the errors are caused by mismatches in assigning households to each district without considering relations between ages and type of building. When we employ this projection considering building area, the effectiveness of the proposed method that considers the building area can be seen from the statistics.

As shown in Fig. 4, households in each district are assigned to a building in a district considering building area. Although we should consider how to reduce errors in assigning households to each district, we can assign each household to a building in a district considering building area. Figure 4 shows the results of projection of synthesized household on the map of Makita-cho, Takatsuki with considering building area. By comparing Figs. 3 and 4, we can see that more households are assigned to larger buildings by considering building area.

Table 14 shows the average number of households that assigned to each building when we consider the area of buildings in Makita-cho, Takatsuki-city, Osaka, Japan. First two columns in Table 14 show the area of buildings. The third columns in Table 14 show the number of households assigned to the corresponded building and the average number of households per building, respectively. The fourth and fifth columns of the random method show the number of buildings whose area is between the minimum area and the maximum area in the first two columns. The sixth and seventh show the number of households and the average number of households by the proposed method considering the building area.

The last row of Table 6 shows the number of households of the corresponding type of building based on the statistics. From this table, we can see that one household has the attribute of other building. According to the attributes of households shown in Table 6, we assign each household to the corresponding building area. According to Table 14, we can see the effectiveness of the proposed method that considers the building area.

As shown in Fig. 4, households in each district are assigned to each building considering their type of ownership of houses and the type of building. When we employ this projection result in traffic analysis of pedestrians or automobiles, we should take into account that there are possibility of different projection of households in that district. It should also be noted that types of buildings in the map are categorized only by the area of buildings shown in Table 6. Therefore, some households are possibly assigned to buildings of schools, hospitals, and business buildings. There is another map of buildings that has at-

| Table 11 | Average and standard deviation of errors and computation time. |
|----------|---------------------------------------------------------------|
|          | SA-Adjustment | Error | Computation time (s) |
| Average  | 60.90         | 92.02 |
| Standard Deviation | 14.43 | 1.24 |
| Ini. Av. | 63,711.60     | –     |
| Ini. St. Dev. | 259.43 | –     |

| Table 12 | Average and standard deviation of errors and computation time. |
|----------|---------------------------------------------------------------|
|          | SA3 | Method A | Method B |
|          | Error | Time (s) | Error | Time (s) |
| Av.      | 20,390.2 | 1,109.44 | 18,193.38 | 1,997.56 |
| SD       | 406.63  | 72.38    | 365.45  | 64.08    |
| Ini. Av. | 606,810.11 | –   | 606,253.07 | –     |
| Ini. SD  | 1,639.14 | –       | 1,635.55 | –       |

| Table 13 | Errors and standard deviations using SA3. |
|----------|------------------------------------------|
|          | Results by SA1 | Method A | Method B |
| Trial    | Error | $r_{ij}$ using $R_{ij}$ | $r_{ij}$ using $R_{ij}$ |
| 1        | 30    | 20,479.2 | 170.1 | 18,128.1 | 294.9 |
| 2        | 36    | 20,491.5 | 212.7 | 18,163.7 | 535.3 |
| 3        | 22    | 20,129.6 | 807.1 | 18,218.0 | 244.6 |
| 4        | 36    | 20,133.5 | 310.5 | 17,992.0 | 307.9 |
| 5        | 30    | 20,500.4 | 375.5 | 18,304.7 | 426.8 |
| 6        | 24    | 20,472.5 | 264.6 | 18,177.6 | 257.2 |
| 7        | 30    | 20,389.7 | 357.7 | 18,229.3 | 307.0 |
| 8        | 20    | 20,202.6 | 312.9 | 17,980.6 | 378.5 |
| 9        | 16    | 20,535.3 | 287.2 | 18,477.7 | 428.7 |
| 10       | 36    | 20,567.7 | 435.9 | 18,262.1 | 242.8 |
| Average  | 20,390.2 | 406.63 | 18,193.38 | 365.45 |

| Table 14 | The number of households assigned to each building. |
|----------|-----------------------------------------------------|
| Min      | Max       | # of B. | Random method | Proposed method |
| Building (m²) | # of H. | Av. | # of H. | Av. | # of H. | Av. | # of H. | Av. |
| 0        | 24       | 106   | 985         | 9.29 | 239 | 1.87 |
| 25       | 199      | 128   | 1,184       | 9.25 | 239 | 1.87 |
| 200      | 499      | 80    | 743         | 9.29 | 2,915 | 36.44 |
| 500      | 999      | 31    | 276         | 8.90 | 74  | 1.74 |
| 1,000    | 10,000   | 6     | 52          | 8.67 | 54  | 1.07 |

decrease these errors we need to ask local governments with smaller population to show more statistics.

Fig. 4 Projected households on a map of Makita-cho, Takatsuki-city. Gray scale and size of each circle denote the number of assigned households in a building. The map is generated using the fundamental geospatial data of the Geospatial Information Authority, Japan.
tributes of buildings such as residential house, schools and hospitals. However, such a map is available only for major cities in Tokyo, Osaka, and Nagoya in Japan. Usages of buildings should be considered in further research.

4. Conclusion

In this paper, we propose a method to project households of synthetic population using fundamental geospatial data for real-world social simulation. That is, we assign each generated household on a building in a geographical map. In order to assign households on a map, we propose a threefold method before projecting households on a map. That is, synthesize households in a local government using SA1. Then, estimate the number of households by a district in the local government using SA2. Finally, we assign synthesized household to meet the number of male and female population in each district using SA3.

When assigning households on a map, we firstly assign households randomly. However, some small house has several households by such random assignment. Therefore, we propose a method that considers the area of construction space of buildings using another SA method. By considering the area of buildings, we can assign households according to the area of buildings. Using the modified method larger buildings have more households. Using the proposed projection method, we can assign households that agree with the statistics of districts. With this household information on a map, we can develop real-world social simulation tools for traffic, marketing, or other real-scale simulations based on the statistics.

References

[1] Editorial Board of Journal of Computational Science: Aims and scope, Journal of Computational Science, Elsevier, https://www.journals.elsevier.com/journal-of-computational-science/ (visited in Mar, 2017).

[2] N. Gilbert: Agent-Based Models, SAGE Publications, 2008.

[3] D. Heard, G. Dent, T. Schifeling, and D. Banks: Agent-based models and microsimulation, Annual Review of Statistics and Its Application, Vol. 2, pp. 259–272, 2015.

[4] A.G. Wilson and C.E. Pownall: A new representation of the urban system for modeling and for the study of micro-level interdependence, Area, Vol. 8, No. 4, pp. 246–254, 1976.

[5] W.E. Deming and F.F. Stephan: A least squares adjustment of a sampled frequency table when the expected marginal totals are known, The Annals of Mathematical Statistics, Vol. 11, pp. 428–444, 1940.

[6] J. Barthelemy and P.L. Toint: Synthetic population generation without a sample, Transportation Science, Vol. 47, No. 2, pp. 266–279, 2012.

[7] F. Gargiulo, S. Ternes, S. Huet, and G. Deffuant: An iterative approach for generating statistically realistic populations of households, PLoS One, Vol. 5, No. 1, e8828. doi:10.1371/journal.pone.0008828, 2010.

[8] M. Lenormand and G. Deffuant: Generating a synthetic population of individuals in households: Sample free vs sample-based methods, Journal of Artificial Societies and Social Simulation, Vol. 16, No. 4, p. 9, 2013.

[9] M. Ichikawa, H. Tanuma, and H. Deguchi: Infectious disease simulation model for estimation of spreading, Developments in Business Simulation and Experiential Learning, Vol. 38, pp. 358–366, 2011.

[10] K. Ikeda, H. Kita, and M. Susukita: Estimation method of individual data for regional demographic simulations, Proceedings of SICE 43th Technical Committee on System Engineer, pp. 11–14, 2010 (in Japanese).

[11] T. Murata and D. Masui: Estimating agents’ attributes using simulated annealing from statistics to realize social awareness, Proc. IEEE International Conference on System, Man & Cybernetics, pp. 717–722, 2014.

[12] D. Masui and T. Murata: A two-fold simulated annealing to reconstruct household composition from statistics, Proc. IEEE International Conference on System, Man & Cybernetics, pp. 1133–1138, 2015.

[13] T. Murata, T. Harada, and D. Masui: Modified SA-based household reconstruction from statistics for agent-based social simulations, Proc. IEEE International Conference on System, Man & Cybernetics, pp. 3600–3605, 2016.

[14] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi: Optimization by Simulated Annealing, Science, New Series, Vol. 220, No. 4598, pp. 671–680, 1983.

[15] Ministry of Health, Labour and Welfare, Japanese Government, 2010 Population Census, Basic Complete Tabulation on Population and Households, Japan, 2011, http://www.e-stat.go.jp/SG1/estat/ListE.do?bid=000001034991 (visited in Mar, 2017).

[16] Ministry of Health, Labour and Welfare, Japanese Government, 2010 Population Census, Statistics for Each District, 27 Osaka-fu, 2012, http://www.e-stat.go.jp/SG1/estat/GL08020103.do?tttGL08020103&ttclassID=000001036634 (in Japanese. visited in Mar, 2017).

[17] The Geospatial Information Authority of Japan: http://www.gsi.go.jp/kiban/ (in Japanese. visited in Mar, 2017).

Takuya HARADA (Student Member)

He received his B.S. and M.S. degrees from Kansai University, Japan, in 2015 and 2017, respectively. Currently he is a Ph. D. candidate of Kansai University Graduate School. His research interests include large scale social simulations.

Tadahiko MURATA (Member)

He received his B.S., M.S. and Ph. D. degrees from Osaka Prefecture University, Japan, in 1994, 1996 and 1997, respectively. He is currently a Professor at Department of Informatics, Kansai University. He was a director of Policy Grid Computing Laboratory at Kansai University, Visiting Researcher at Computational Institute of Chicago University, Chair of IEEE SMCS Japan Chapter, Chair of IEEE SMCS Technical Committee on Awareness Computing, and Member of IEEE CIS Neural Networks Technical Committee. He is currently a board member of IEEE SMC Society. His research interests include social simulations, multi-objective optimization, and soft computing. He is a member of ISCIE, SOFT, IPS, JSEC, and IEEE.