Upgrading Spatiotemporal Demographic Data by the Integration of Detailed Population Attributes

Toshihiro Osaragi [a]*, Ryo Kudo [a]

[a] Tokyo Institute of Technology, osaragi.t.aa@m.titech.ac.jp, kudo.r.ac@m.titech.ac.jp

* Corresponding author

Abstract: In this study, a method was constructed for adding value to spatiotemporal data by integrating demographic information obtained from Mobile Spatial Statistics (MSS), Person-trip (PT) data, and the national census. We first constructed a model that provided spatiotemporal distribution of occupants in urban areas that vary according to clock time, location, and building use classification. The time, location, and building use classification were employed as keys to integrate demographic information. Weekday and weekend data for the central wards of Tokyo were employed to create estimates of the number of occupants with their detailed attributes. Using numerical examples, we demonstrated that the proposed model can provide demographic spatiotemporal distributions with far higher value than before; in which the buildings people occupy, their reasons for being there, their sex and age bracket, and their residential locations, can all be identified.

Keywords: census data, Person-trip data, Mobile Spatial Statistics, Building Point data, spatiotemporal distribution

1. Introduction

1.1 Background and Research Objective

Research for urban planning and regional disaster prevention must always consider the spatiotemporal distributions of people in urban areas. However, in large, economically thriving regions, the spatial distributions of such people change very rapidly over the course of a day, which means that it is extremely difficult to keep precise track of their shifting trends. Formerly, the results of national censuses, commercial statistical surveys, and other data sources were used to estimate daytime and night time demographics, but there are limitations to the utility of such static population distributions.

Person-trip data in the Tokyo Metropolitan Area (PT data), which consists of actual survey-based travel data, can provide some grasp of dynamic human population distributions. This data features detailed population classifications based not only on attributes, but also individual departure and destination locations, departure and arrival times, means of transportation, and purpose of travel, as well as other items. However, among the significant drawbacks of PT data are the facts that the sampling rate is only about 2% and that such data are taken only collected about once a decade. As a result, it cannot be considered "current" as of this writing.

In recent years, efforts have been made to estimate the spatiotemporal distribution of urban areas occupants from cellular telephone user population statistics based on user locations extracted from phone company tower data. In this study, the term "occupants" is used to mean people who are traveling to a destination as well as those who are already at their destination. These statistics, which provide regional data at very high sampling rates for any hour of the day everywhere in Japan, have the potential to be usefully employed in a variety of fields. In particular, NTT DOCOMO sells its Mobile Spatial Statistics (MSS: a trademark of NTT DOCOMO, INC.), for which the sampling rate in the Tokyo Metropolitan Area is as high as 50%, and various attempts have been made to confirm the accuracy of those statistical results (Seike et al., 2011). These demographic data represent the number of cell phones employed in the cellular network, combined with the cell phone market share enjoyed by NTT DOCOMO.

However, in a nod to privacy protections, few detailed attributes are provided with MSS and no distinctions are made between people in transit and stationary occupants. Thus, while the various population statistical datasets that are available offer high-quality information, each has its limitations and restrictions.

This study examines methods for improving the value of population statistics with the aim of preserving the advantages of each dataset while compensating for their shortcomings. The best features of each dataset, i.e., the immediacy and continuity of MSS (which are gathered and stored each hour of every day), the detailed personal attributes provided with the PT data (age, sex, occupation, etc.), as well as the purpose of individuals' stay in Tokyo, means of transportation, and the completeness of the population statistics in the national census (the precise number of residents, based on a complete enumeration), were integrated in an investigation of methods aimed at constructing a superior spatiotemporal demographic dataset. Next, this dataset was employed as part of an attempt to obtain some basic observations about the spatiotemporal distribution of urban area occupants.

1.2 Relevant previous research

Numerous previous studies have examined methods for estimating the attributes of urban area populations. For example, pioneering research was carried out by Tsuji (1981), who proposed a direct calculation method based on actual counts and an indirect calculation method using

employee numbers in local industries along with other measurements. In studies on exploiting PT data, some researchers have followed in Tsuji's footsteps by proposing to combine PT data and geographic information system (GIS) data to estimate the number of occupants at any given time and in any given locale (Osaragi and Hoshino, 2012; Osaragi, 2016).

In addition, increasing numbers of researchers are publishing studies employing demographic statistics of cell phone users. These include Seike et al. (2015), who published basic research attempting to predict the demographic attributes of cell phone users; Arimura et al. (2016), who estimated building populations based on their use classification, floor area, and number of wings; Yabe et al. (2016), who carried out crowd flow calculations of population movements during disasters by conforming a particle filter to a simulation of disaster-related activities. Furthermore, Calabrese et al. (2011) attempted to construct a real-time system for observations in Rome, Italy using the GPS function installed in cell phones, while Deville et al. (2014) compared current demographic statistics with national survey results in order to draw conclusions about the potential uses of cell phone user demographic statistics. Many other researchers are also attempting to monitor the activities of people in urban regions using their cell phone location data (Ahas et al., 2015; Gao, 2015; Järv et al., 2017).

Thus, there are numerous examples of research aimed at taking full advantage of the attributes offered by the abovementioned demographic data taken from mobile phones. However, few of those studies have attempted to raise the value of the data by combining datasets from multiple demographic statistical sources. Although some have combined national censuses with satellite observations (Chen, 1998, 2002; Sim, 2005, Steinnocher et al., 2006), those studies did not attempt to raise the value of the demographic statistical data itself.

2. Considerations in Methods for Combining Spatiotemporal Demographic Statistical Datasets

This paper examines a method of combining the detailed information obtained from PT data, including attributes and travel objectives, into the MSS base source, which offers superior levels of immediacy and consistency. The completeness of the demographic data obtained from the national census was also used to compensate for the anonymization of the MSS.

When combining multiple datasets, a variety of keys must be used. For example, when using GIS, each feature is designated with an ID that is common to every database. This feature provides a "key" that is used to tie the spatial information to the non-spatial information (attributes). However, since there are no explicit keys that directly tie MSS or PT data together, this study examines a method of using a single key to denote each combination of time, location, and building use in order to integrate the two databases (Fig. 1).
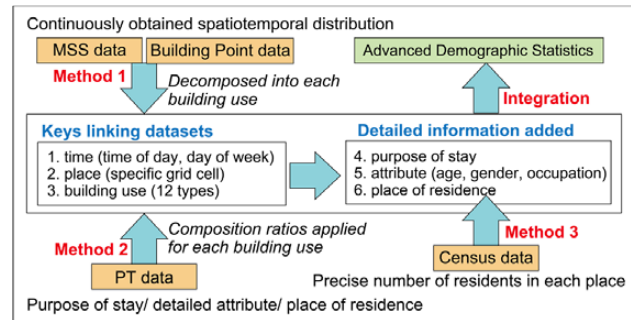

Figure 1. Framework of methods for integrating demographic statistics.

For example, when considering the attributes and travel objectives of occupants of an office building (building use) in the city centre (location) during a weekday (time), one would expect most of those occupants to be workers in the age range 30 – 59 years. Or, that most of the occupants of a school facility (building use) in a suburb (location) would be students with average ages around 20, or workers in their twenties through their sixties. On the other hand, a restaurant (building use) near a commercial zone (location) would be expected to have occupants with a variety of attributes, although these attributes would also be expected to vary with the day of the week and the hour (time).

As these examples illustrate, once the time, location, and building use have been identified, some of the attributes of the people occupying a building and their reasons for being there, are strongly implied. It is correspondingly expedient to have information regarding how many buildings in a given area have a certain use classification in order to anticipate the attributes and reasons the occupants have for being there at a given time of day. That is the rationale for using the time, location, and building use as a single key for classifying occupants in this study.

3. Analysis of Occupant Number by Building Use Classification (Method 1)

3.1 Model formulation

Although PT data contains information about the uses of buildings containing occupants, MSS does not. Therefore, we begin by considering a method for assigning use classifications to the locations indicated by MSS as containing occupants (Method 1 in Fig. 1).

The number of occupants y^t_i [occupants] in cell i at time of day t is found from the sum of occupants y^t_{ij} [occupants] inside or adjacent to all buildings of use classification j, as follows:

$$y^t_i = \sum_j y^t_{ij} \qquad (1)$$

Next, the number of occupants y^t_{ij} [occupants] is assumed to be proportional to the sum of the floor areas x_{ij} of all buildings of use classification j and is calculated using the occupant density D^t_{ij} [occupants/m²], which expresses the number of occupants per unit of building floor area at a given time:

$$y^t_{ij} = D^t_{ij} x_{ij} \qquad (2)$$

Occupant density D^t_{ij} can be separated into the temporal fluctuation component α^t_j, which depends on the time of day t and the building use classification j, and the location factor β_i, which depends on the cell (loc i), and can be formulated as the product of the two, as shown below:

$$y^t_{ij} = \alpha^t_j \beta_i x_{ij} \qquad (3)$$

The unknown parameters α^t_j, β_i can be estimated in a converging algorithm using multiple regression, as diagrammed in Fig. 2 (Osaragi and Hoshino, 2012).
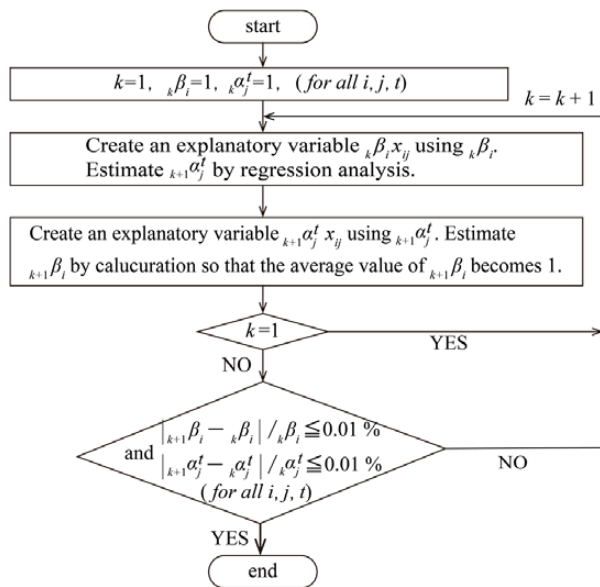

Figure 2. Converging algorithm using multiple regression for estimating parameters.

3.2 Study area and data employed

Figure 3 provides an overview of the data employed for this study. The study area was the central wards of Tokyo, which see vigorous spatiotemporal fluctuations in their occupant numbers throughout the day. The numbers of occupants (without attribute data) obtained from MSS at time t in cell i on weekdays and weekend days were used as the response variable y^t_t [occupants]. The Building Point (BP) data, which incorporates detailed attributes of each building, was employed to obtain the explanatory variables of cell i, building use classification j, and total floor area x_{ij}. Table 1 shows the twelve kinds of use into which the buildings were classified.
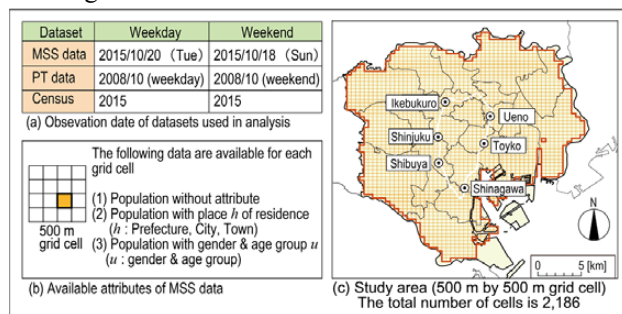

Figure 3. Study area and datasets used in analysis.

no.	Building use	Examples
1	public/religious facility	government office, museum, temple and shrine
2	educational facility	elementary/junior high/high school, university, vocational school, study cram school
3	medical/welfare facility	hospital, nursing care welfare facility, etc.
4	office	logistics, post office, finance, insurance, real estate, industry
5	restaurant/bar	home delivery, restaurant, bar
6	commercial facility	service, sales business, automobile industry (including automobile repair business)
7	mass retailer	supermarket, department store, large store, etc.
8	hotel	hotel, inn
9	apartment house	flat, apartment house, dormitory, company house
10	detached house	detached house
11	railway station	the number of railways
12	park	area of park

Table 2. Explanatory variables (Total floor area by building use taken from Building Point data)

With the respective use classification, the floor areas of combined facilities containing multiple buildings of various uses were divided in proportion to the numbers of agents operating the portions of each property. The residential areas of multiple-use residential facilities were estimated using the average area for apartment buildings in Tokyo that were quoted in the 2013 housing and land survey published by the Japan Statistics Bureau of the Ministry of Internal Affairs and Communications (MIC).

Calculations for railway stations were based on the 2015 National-Land Numerical Information (Railway). Here, it should be noted that even though parks are not buildings, it was deemed necessary to include them in the distribution of occupants for this model.

3.3 Model creation and validation

Our model was created using the datasets for both weekdays and weekend days. Figure 4 presents a portion of the results obtained when validating the compatibility of the estimated values used. Although both under- and overestimates are found in a small number of cells, description accuracy was good for both weekdays and weekend days. Of particular note is the fact that the estimation accuracy coefficients of determination were above 0.95 at 12:00 and 18:00. This result is quite significant because it indicates that the spatiotemporal distributions of occupants in the middle of the day and late afternoon, which are times of heavy population movements, were accurately reproduced. In comparison, the estimates at 24:00 and 06:00 on a weekday were rather poor. For the late-night weekday results, it is possible that this was due to factors other than building uses, which were incorporated here. Nighttime human movements, which are of a different character than daytime movements, require different explanatory variants and are more difficult to model. Although they are of great interest, they cannot be clearly described at this time, so they will be addressed in future studies.

The error rate of estimate E_i was defined as the statistical quantity in the following equation, and the spatial distribution of this equation is shown in Fig. 5.

$$E_i = \frac{\sum_t |y_i^t - \hat{y}_i^t|}{\sum_t y_i^t} \times 100 \ [\%] \tag{4}$$

Relatively high estimate errors can be seen in the vicinity of Tokyo Bay, which includes cells with large amounts of ocean water, canals, warehousing, and other atypical space types. Hence, there are few facilities whose explanatory variables are valid for the occupant estimates shown in Table 1. As a result, even though estimation accuracy levels suffered along such borders of the study area, the spatial distribution of those errors was fairly consistent and did not favor one portion over another.
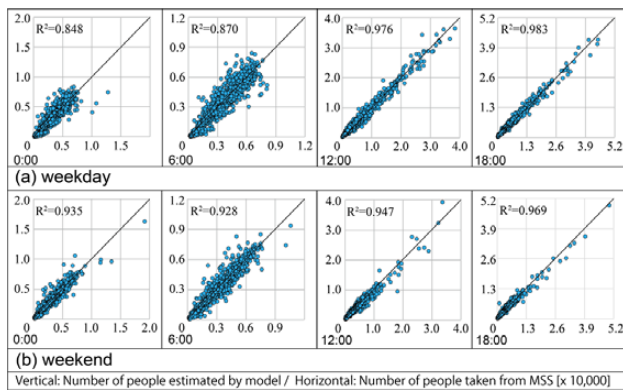

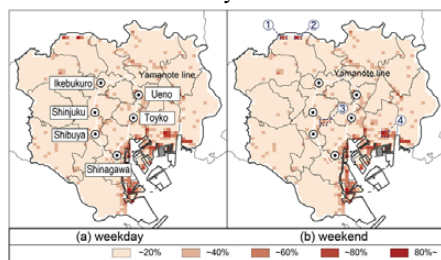Figure 4. Validation of models by the estimated number of people.

Figure 5. Spatial distribution of error rate (12:00).

3.4 Interpreting parameters employed for estimates

Part of the estimated values for the temporal fluctuation component α_j^t are shown in Fig. 6, where it is clear that the temporal fluctuation for occupants of a unit of area (100 m²) varied greatly with the building use classification. Variations were also significant when comparing weekdays to weekend days. For example, the occupant density increased abruptly after 08:00 in offices on weekdays and fell gradually after 18:00, while the opposite fluctuations were seen in apartment buildings. Meanwhile, restaurants/bars show a peak around 20:00.

On weekend days, the temporal fluctuation in office building occupant density is 1/4 to 1/5 that seen on weekdays, while the greatest reduction in apartment building occupant density occurred at 16:00. Meanwhile, the peak in commercial facilities was seen at 17:00 and was at levels about 2.5 times those seen on weekdays.
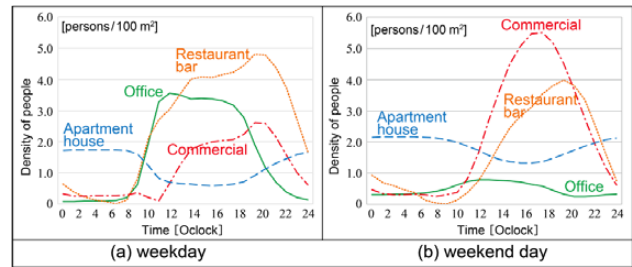

Figure 6. Temporal change in the density of number of people by building use (α_j^t).

Figure 7 shows the spatial distribution of the values estimated for parameter β_i, which is the location factor. The value of β_i on weekdays is high in the vicinity of Tokyo Station and other places with high concentrations of commercial facilities where the temporal fluctuations of occupant density are higher than in other districts. Comparing weekdays and weekend days, it can be seen that weekend days have smaller fluctuations in β_i, and show a fairly even distribution. Thus, the individual features of a given location have more to do with the occupant density on weekdays than on weekend days.
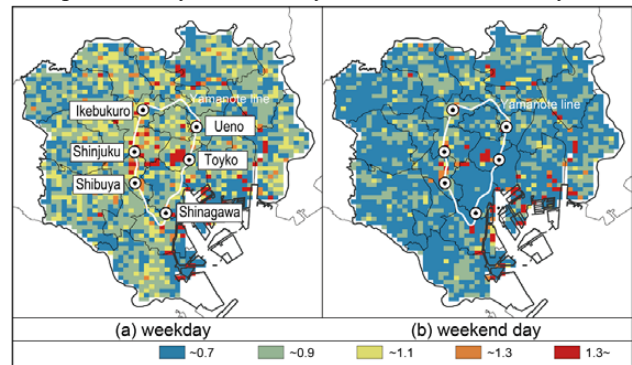

Figure 7. Spatial distribution of the estimated parameters representing location characteristic factor (β_i).

3.5 Estimation of number of occupants using building type

Using the above model, it is possible to group the total number of occupants estimated by MSS levels according to the building type classification. (This can be done by dividing the estimated total number in proportion to the buildings of each type.) Figure 8 presents part of the spatial distribution of the number of occupants categorized by the type of building they occupy at 14:00 on weekdays and weekend days. The reader can see how, on weekdays, (i) the total occupant population was distributed among, (ii) the office facilities in central Tokyo, (iii) detached houses on the edges of the study area, and (iv) commercial facilities. On weekend days, the results in (iii) show higher spatial occupant distributions (the number of those being at home) in the edges of the study area, thus predicting that more people will be at home. At the same time, although there are few occupants (ii) in the office facilities of central Tokyo, occupant levels are highly concentrated (iv) in central Tokyo cells containing large commercial facilities.
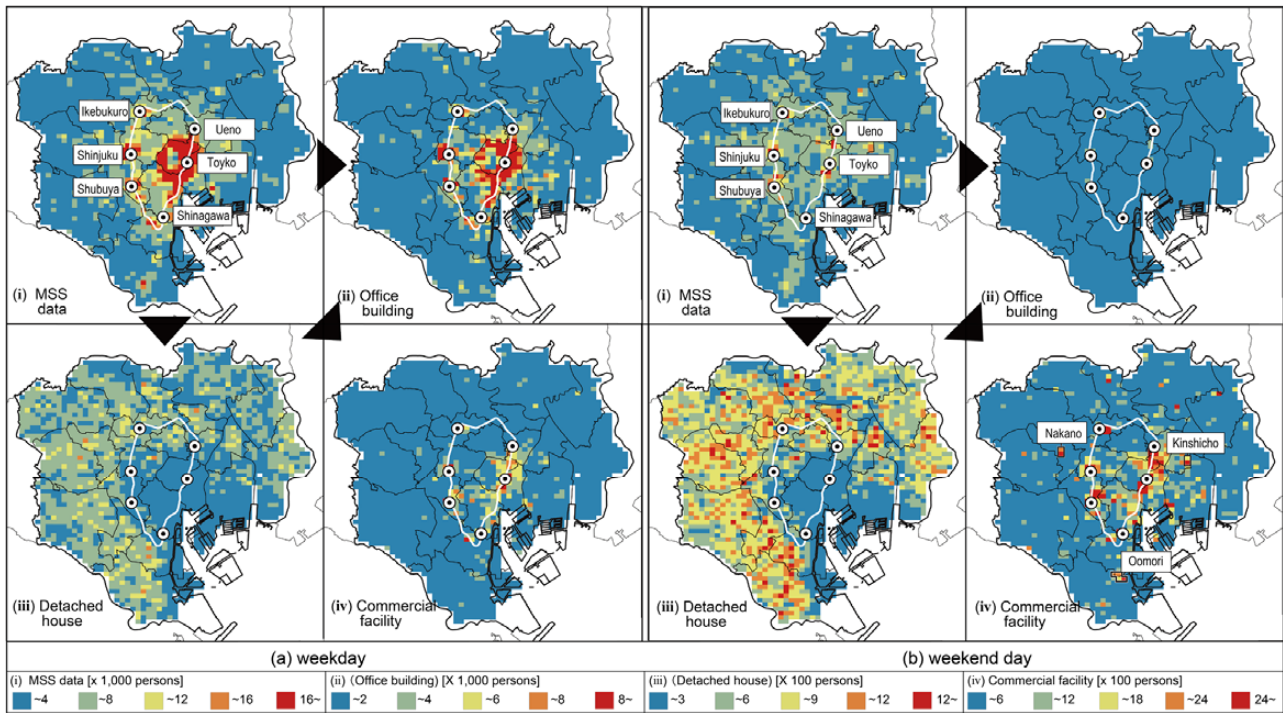
Figure 8. Spatial distribution of the number of people by building use at 14:00 on weekday and on weekend day.

4. Integration of Spatiotemporal Demographic Statistical Datasets

4.1 Pre-processing of PT data

Next, the information provided in the PT data is integrated with the occupant populations reclassified by building type y^t_{ij} that was calculated in Section 3. PT data provides detailed attribute information (sex, age bracket, residence location, etc.) along with information about the individual's travel (travel purpose, transportation means, use classification of destination, etc.). The basic spatial unit of area in the PT data was a "zone", i.e. an area typically comprised of between 10 and 20 city blocks with a nighttime population of about 15,000. It was assumed that occupants M^t_{sjk} with attribute k at time t in building of use classification j in zone s were uniformly distributed throughout the entire zone, and that the number of occupants y^t_{ijk} with attribute k at time t in the building of use classification j in cell i was calculated using the fraction of building floor area in cell i, which was counted in that use classification.
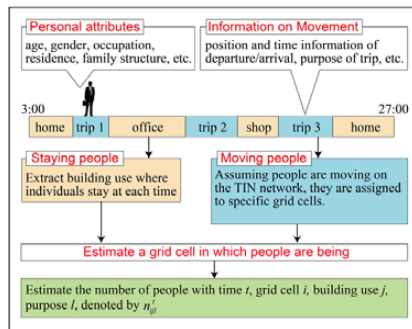

Figure 9. Method of extracting the spatiotemporal distribution of people.

The total number of occupants in each cell was calculated by assuming that the population y^t_{ijk} of traveling people in each cell associated with building use classification j traveled to the vicinities of their destination buildings of that classification in the way described above (Fig. 9). Travel on weekend days was simulated using the method previously reported (Osaragi, 2016) with the PT data collected after completing the same pre-processing.

4.2 Estimation of number of occupants on the basis of both building use and purpose of travel (Method 2-1)

It is possible that people will have differing purposes for going to the same building of a given use classification depending on the differing characteristics of the building location. The number of occupants n^t_{ijl} at time t in building of use classification j in cell i for purpose l was calculated from the PT data, and the fraction of the occupants based on the purpose of travel m^t_{ijl} was calculated using the time, location, and building use as keys, as follows:

$$m^t_{ijl} = \frac{n^t_{ijl}}{\sum_l n^t_{ijl}} \qquad (5)$$

The calculated m^t_{ijl} was multiplied by y^t_{ij}, which is the number of people in a building of a given use classification, in order to estimate Y^t_{ijl} [people], which is the number of occupants based on both building use and travel purpose.

$$Y^t_{ijl} = m^t_{ijl} y^t_{ij} \qquad (6)$$

Figure 10(a) shows an example of this calculation for the cell that includes *Shinjuku Station*. Here, the reader can see a dramatic difference between weekdays and weekend days in the numbers of occupants of buildings of various uses. On weekdays, a large number of people occupy buildings with a wide variety of use classifications, and

these levels vary significantly with the time of day. In contrast, the temporal fluctuations in occupant numbers were mild on weekend days, with the sole exception of the fluctuations seen at commercial facilities.

A more detailed view focusing on a use classification is provided for *Shinjuku Station* in Fig. 10(b), which shows the occupant numbers by purpose. Here, we find that most of the occupants at 08:00 on weekday mornings were commuting to work or school, while at 18:00 most were on their way home. On weekend days, there were a variety of purposes, including those related to eating, leisure, and shopping.
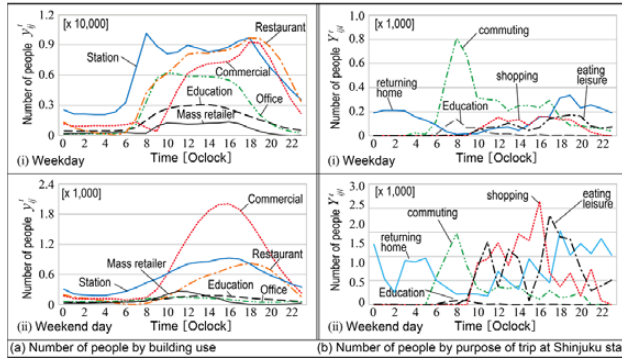

Figure 10. Temporal change of number of people by building use in cell where Shinjuku Station is located/ by purpose of stay at Shinjuku Station.

4.3 Estimation of number of occupants based on building use and sex-age classification (Method 2-2)

The number of occupants s^t_{iju} at time t in building of use classification j in cell i in the sex-age bracket classification u is calculated from the PT data, and the fraction of the occupants r^t_{iju} is calculated as follows:

$$r^t_{iju} = \frac{s^t_{iju}}{\sum_u s^t_{iju}} \qquad (7)$$

The calculated r^t_{iju} is multiplied by y^t_{ij}, which is the number of people in a building of a given use classification, in order to estimate Y^t_{iju} [people], which is the number of occupants on the basis of both building use and sex-age classification. However, MSS also provides a number of occupants y^t_{iu} of sex-age classification u in cell i at time t. In order to integrate these two data samples without contradiction, the number of occupants based on building use and sex-age classification Y^t_{iju} [people] is calculated thus:

$$Y^t_{iju} = \frac{y^t_{iu}}{\sum_j r^t_{iju} y^t_{ij}} r^t_{iju} y^t_{ij} \qquad (8)$$

Figure 11 shows the temporal fluctuations on a weekday in the number of occupants of the commercial facilities of a certain cell, grouped by sex-age classification. Most of the occupants of this cell, which includes *Shibuya 109* and other commercial facilities, were men and women in their twenties and thirties. Meanwhile, in the cell containing *Ginza Station*, a large majority of people in that age bracket were women. Interestingly, most of the daytime occupants of the commercial facilities in the cell

containing *Akihabara Station* were women in their twenties through forties, while they were outnumbered by men in their thirties through fifties in the evening. Thus, we can find variations in the sex-age classifications of the occupants of any given location that are in keeping with the characteristics of that location.
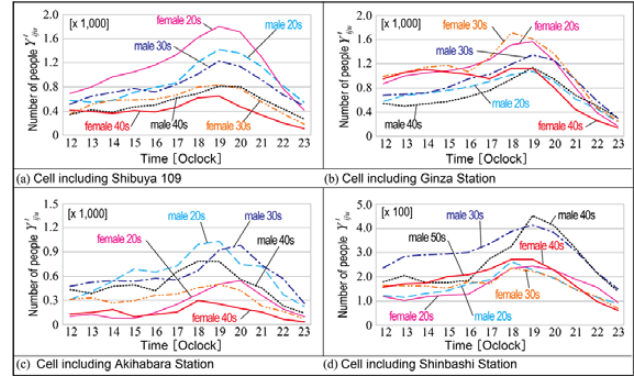

Figure 11. Temporal change in the number of people by personal attributes in commercial facilities in a specific cell.

4.4 Estimation of the number of occupants on the basis of building use and place of residence (Method 2-3)

The number of occupants v^t_{ijh} residing in location h and occupying building of use classification j in cell i at time t was calculated from the PT data, after which the fraction of the occupants on the basis of building use and place of residence w^t_{ijh} was calculated.

$$w^t_{ijh} = \frac{v^t_{ijh}}{\sum_h v^t_{ijh}} \qquad (9)$$

The extracted w^t_{ijh} was multiplied by y^t_{ij}, which is the number of people in a building of a given use classification, in order to estimate Y^t_{ijh} [people], which is the number of occupants based on both building use and place of residence. MSS also provides a number of occupants y^t_{ih} in residential area h in cell i at time t. The following equation was used after the PT data had been added to modify Y^t_{ijh}.

$$Y^t_{ijh} = \frac{y^t_{ih}}{\sum_j w^t_{ijh} y^t_{ij}} w^t_{ijh} y^t_{ij} \qquad (10)$$

In order to protect privacy, the MSS y^t_{ih}, which included time t, cell i, and place of residence h, is pre-processed to block access to the original data. The following section examines a method for modifying y^t_{ih} to compensate for the spatial information obtained from cell phones.

4.5 Compensation for anonymized populations (Method 3)

Because of the variation in the areas of the residential spaces h, the MSS for y^t_{ih} at time t, cell i, and housing h, y^t_{ih} is available by prefecture, by city, or by town. However, if the number of occupants y^t_{ih} falls below a lower limit N, data are anonymized (identifying data is scrubbed from the original data) in order to protect privacy. This can be verified by examining the total populations of the study areas at time t, as shown in Fig. 12.

For example, a person is less likely to need anonymity protection when he or she lives in a prefecture together

with a large number of other occupants. This is understandable because it is nearly the same as if one's residential information is hidden ("No attribute shown" in response to an inquiry). However, occupants who reside in a cell at the town level are anonymized (identifying information is scrubbed from the original data) because there are a limited number of occupants in towns.

Accordingly, the "total sums" of occupants of towns do not actually represent the total number of people in the cell. This is because there are significant amounts of movement among Tokyo residents within the Tokyo Metropolitan Area during the day, and these people return home at night. As a result, numerous individuals could be identified from the fact that they are now in their residential cells, which reduces the probability that the number of indicated occupants is underestimated due to anonymization (Fig. 12).
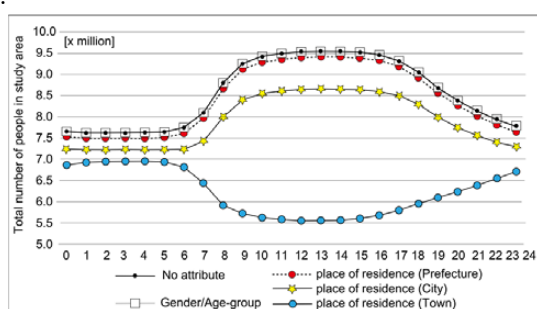

Figure 12. Total number of people counted according to attribute information in the study area.

Figure 13 shows a breakdown of how it is possible to compensate for anonymized information. Let us begin by assuming that numerous occupants live in district h, a town, A city, P prefecture, and are not anonymized. However, some occupants live in towns b and c, where the occupant number falls below N and must be anonymized. In this case, the occupants of a, b, and c (written $n(a)$, $n(b)$, and $n(c)$) are included within $n(A)$, which is the number of occupants of A, so the total exceeds N, thereby obviating anonymization. The relation $n(A) \geq n(a)$ also holds. If the occupants whose district h includes towns b and c are included within $n(A)$, the number of these occupants is represented by $n(A) - n(a)$. Although it is not possible to know exactly what the values of $n(b)$ and $n(c)$ are, if we assume that the number of anonymized individuals is proportional to the nighttime populations, we can use those proportions to estimate $n(b)$ and $n(c)$ by appropriately dividing the quantity $n(A) - n(a)$.

District h was used in the example above with towns, but the same principle of an h can be applied to municipalities of any size. From Fig. 12, we see that in spite of the lower likelihood of anonymization of occupants residing in cities compared to towns, some occupants are in fact anonymized. Hence, it is best to first apply the compensation method described above to the prefectural data (which has a low need for anonymization), then to city data, and lastly, to town data.

The number of occupants based on both building use and place of residence Y^{t}_{ijh} [people] was estimated using Eq.

(10) in order to combine the compensated MSS for occupant number y^{t}_{ih} in residential area h in cell i.
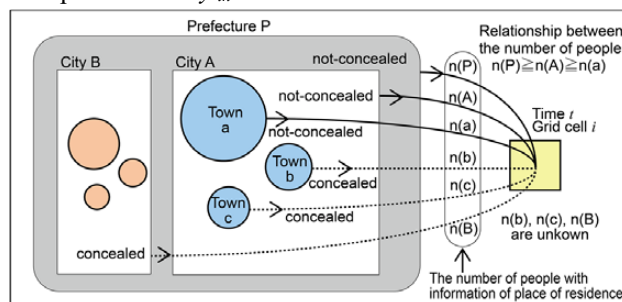

Figure 13. Concealing of the number of people by place of residence.

5. Summary and Conclusions

In this study, a method was constructed for adding value to spatiotemporal data by integrating demographic information obtained from MSS, PT data, and the national census. To accomplish this, we first constructed a model that provided spatiotemporal distribution estimates of the number of occupants in urban areas in terms of clock time, location, and building use classification. This data was obtained from the MSS demographic statistics that were estimated from cellular phone user location information and then combined with building point data, which contains information on building use, floor area, etc. The study area was the central wards of Tokyo. Model validation was carried out using actual data, and the results showed it provided good accuracy. Using this model with the time and location information obtained from MSS allowed the number of occupants to be broken down by building use classification.

Next, the time, location, and building use classification were employed as keys to integrating detailed PT data about the attributes of occupants and residents of Tokyo and their reasons for traveling within Tokyo, as classified by the times, locations, and types of buildings occupied. Weekday and weekend data for the central wards of Tokyo were employed to create estimates of the number of occupants that were: (1) classified by building use and their purpose for being there, (2) classified by building use and sex-age classification, and (3) classified by building use and residential location.

Until now, it has only been possible to establish the spatiotemporal distributions of populations by conducting large-scale surveys or examining highly specific locations. Now, however, it is possible to obtain demographic spatiotemporal distributions with far higher value, in which the buildings people occupy, their reasons for being there, their sex and age bracket, and their residential locations, can all be identified. In the future, we will be able to perform new analyses using the rich fund of information provided by these highly sophisticated demographic spatiotemporal datasets.

Although we attempted to address the statistical credibility of the estimates in the discussion in Section 3, this study employed multiple datasets and combined them in a complicated way. As a result, the statistical error in the datasets is non-uniform and complex. Essentially,

although the population estimates both in heavily populated locations and in locations with buildings tend to be relatively accurate, it is far from clear how accurate the estimates are in locations which fit both descriptions. Ultimately, it will be up to future researchers to find methods for statistically evaluating the reliability of the spatiotemporal demographic data developed in the ways described here, and to discuss this reliability.

Acknowledgements

References

Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M. and Zook, M. (2015) Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn, Issue 11: Geographies of Mobility: Applications of Location Based Data, International Journal of Geographical Information Science, Vol. 29, pp. 2017-2039.

Arimura, M., Kamada, A. and Asada, T. (2016) Estimating of Visitors Number in Mesh by Building Use by Integrated Micro Geo Data, Proceedings of JSCE, D3, Vol. 72, No. 5, pp. I_515-I_522.

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D. and Ratti, C. (2011) Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome, IEEE Transactions on Intelligent Transportation System, Vol. 12, No. 1, pp. 141-151.

Chen, K. (1998) Correlations between census dwelling data and remotely sensed data. In Proceedings: SIRC 98 – 10th annual colloquium of the spatial information research centre, Dunedin, New Zealand.

Chen, K. (2002) An approach to linking remotely sensed data and areal census data, International Journal of Remote Sensing, Vol. 23, No. 1, pp. 37–48.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, R.F., Gaughan, E.A., Blondel, D.V. and Tatem, J.A. (2014) Dynamic Population Mapping Using Mobile Phone Data, PNAS, Vol. 111, No. 45, pp. 15888-15893.

Gao, S. (2015) Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age, Spatial Cognition & Computation, Vol. 15, No. 2, pp. 86-114.

Järv, O., Tenkanen, H. and Toivonen, T. (2017) Enhancing spatial accuracy of mobile phone data using multi-temporal dasymetric interpolation, International Journal of Geographical Information Science, Vol. 31, pp. 1630-1651.

Osaragi, T. (2016) Estimation of Transient Occupants on Weekdays and Weekends for Risk Exposure Analysis, ISCRAM 2016: International Conference on Information Systems for Crisis Response and Management, Proceedings of the ISCRAM 2016 Conference, ISCRAM.

Osaragi, T. and Hoshino, T. (2012) Predicting Spatiotemporal Distribution of Transient Occupants in Urban Areas, AGILE 2012: Conference on Geographic Information Science - 15th, Lecture Notes in Geoinformation and Cartography, Bridging the Geographic Information Sciences, Springer, pp. 307-325.

Osaragi, T. and Kudo, R. (2019) Enhancing the Use of Population Statistics Derived from Mobile Phone Users by Considering Building-Use Dependent Purpose of Stay, 22nd Conference on Geo-Information Science (AGILE 2019), Geospatial Technologies for Local and Regional Development, Springer, Cham, pp. 185-203.

Seike, T., Mimaki, H. and Morita, S. (2015) Study on The Population Characteristics in a City Center District Utilizing Mobile Spatial Statistic: A New Statistical Method to Capturing Actual Population Distribution of the City through Day and Night, Journal of Architecture and Planning (Transactions of AIJ), Vol. 80, No. 713, pp. 1625-1633.

Seike, T., Mimaki, H., Hara, Y., Odawara, T., Nagata, T. and Terada, M. (2011) Research on the Applicability of "Mobile Spatial Statistics" for Enhanced Urbab Planning, Journal of the City Planning Institute of Japan, Vol. 46, No. 3, pp. 451-456.

Sim, S. (2005) A proposed method for disaggregating census data using object-oriented image classification and GIS, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI (Part 8/W27).

Steinnocher, K., Weichselbaum, J., Köstl, M. (2006) Linking remote sensing and demographic analysis in urbanised areas, In P. Hostert, A. Damm, & S. Schiefer (Eds.), Proceedings: First workshop of the EARSeL SIG on urban remote sensing, Berlin, Germany.

Tsuji, M. (1981) Survey of de Facto Population in Central Commercial Districts: Study on Methods to Estimate De Facto Population in Small Districts Part 1, Transactions of the Architectural Institute of Japan, No. 309, pp. 157-166.

Yabe, T., Sekimoto, Y., Kashiyama, T., Kanasugi, H. and Sudo, A. (2016) Real-time Prediction of Peoples Movement under Disaster Situations using Particle Filter, Journal of Traffic Engineering, Vol. 2, No. 2, pp. A_19-A_27.