

Spatiotemporal Interpolation Method for Population Flow Data in Urban Areas

Toshihiro Osaragi^{1,*}, Xianshu Nan¹, and Maki Kishimoto¹

¹ School of Environment and Society, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan -
 osaragi.t.aa@m.titech.ac.jp, nan.x.aa@m.titech.ac.jp, kishimoto.m.ac@m.titech.ac.jp

Commission IV, WG IV/9

KEY WORDS: Spatiotemporal Interpolation, Population Flow, Urban Demography, Tokyo, Denman-Beavers Iteration.

ABSTRACT:

Capturing the intricate dynamics of population movements in urban areas holds substantial implications for urban planning and management, particularly in the context of disaster mitigation. There is an attempt to introduce methods for estimating the spatiotemporal distribution of population flows, leveraging demographic data from various kind of sources. In earlier spatiotemporal interpolation methods, some key assumptions were made to obtain data at shorter intervals. In this study, we present an alternative spatiotemporal interpolation method by loosening assumptions and increase its versatility and facilitates flexible application across various contexts and objects. This is achieved by estimating the square root of the movement probability matrix for longer time intervals. The efficacy of our approach is demonstrated through its application to actual data from the Tokyo 23 wards, allowing for the estimation of the spatiotemporal distribution of population flows across various time intervals. Our results not only affirm the accuracy of the estimates but also provide insights into the intricate population flows within the densely populated regions of the Tokyo 23 wards. Moreover, by estimating population data at shorter time intervals, we explore the characteristics of these flows, offering an understanding of the dynamics that shape urban demography.

1. INTRODUCTION

The spatial distribution of individuals within metropolitan areas exhibits temporal variations, driven by human activities and facilitated by the mobility afforded by rapid urban transportation systems. Recently, there has been a burgeoning interest in the observation and analysis of macro-scale population flows across different timeframes. This heightened interest is motivated by the imperative to address various applications within the domain of urban planning. These applications encompass the alleviation of crowding at large-scale public events, the facilitation of evacuation guidance in the aftermath of major earthquakes, and the optimization of area marketing strategies within commercial industries. For these purposes, various spatiotemporal demographic data have been employed. Table 1 illustrates examples of data that are particularly valuable for macro-scale analyses, majorly employed within the context of Japan.

Notably, the Person Trip survey data (PT data) serves as a comprehensive source that is a questionnaire-based survey conducted by the government. It provides information on sex, age, purpose of trip, departure/arrival locations and times, among other details [Regions: Greater Tokyo Metropolitan Area/ Samples: Random sampling based on census data (about 1.4 million households from about 16 million households)/ Valid data: about 340 households (sampling ratio = 2.13 %)]. However, PT data is collected only every 10 years and has a low sampling rate.

The ubiquity of mobile phones has ushered in a novel opportunity to gather dynamically changing location information from users with higher sample rates (Deville et al., 2014; Ratti et al., 2006; Calabrese et al., 2011). Mobile Spatial Statistics (MSS) represents one such dataset derived from mobile phone location data, providing regional population information in grid-cell units at any desired time and day (Seike, 2011). From a macroscopic perspective, mobile data serves as a widely recognized proxy for comprehending the demographic dynamics of individuals residing at nighttime and engaging in activities at daytime within a specific area. Consequently, mobile data can be utilized to estimate population movements to a certain extent. From a microscopic standpoint, however, it does not offer insights into the directional vectors or distances associated with population flows.

Addressing the need for data on population flows and their trajectories, certain mobile phone datasets, such as Konzatsu-

Dataset	Person Trip survey (PT)	Mobile Spatial Statistics (MSS)	Konzatsu-tokei® (KT)
Content	travel behavior	spatial distribution	population flow
Sampling rate	■ about 2 – 3 %	○ about 40 %	■ about 0.5 %
Spatial unit	■ zone	○ 500m grid-cell	○ 250m grid-cell
Time resolution	○ every minute on weekdays	○ every 1 hour	○ every 5 minutes
Frequency	■ every 10 yrs.	○ everyday	○ everyday
Available information	○ sex, age, home address, dep./arr. time, purpose of trip.	■ sex, age, home address (prefecture or municipality zone)	○ sex, age

Table 1. Examples of population statistics (○: Advantage, ■: Disadvantage).

* Corresponding author

tokei® (KT), offer the population flows between grid-cells in 5-minute intervals. Despite this advantage, there are some limitation or shortfalls on the data's instability in grid-cells with data-exclusion due to privacy protection (Kamada, 2017). Despite the availability of these datasets, each presents its own set of limitations, encompassing factors like time interval, sampling rate, and spatial units. Consequently, there are some attempts have proposed methods that involve combining multiple datasets to generate new, more comprehensive datasets that leverage the strengths of each while compensating for their respective shortcomings. Osaragi and Kudo (2019), for instance, integrated multiple demographic datasets, including PT data and MSS, to obtain extensive data on static individuals in urban areas. Similarly, Osaragi and Hayasaka (2019) introduced a method for acquiring the spatiotemporal distribution of population flows, employing PT data, MSS, and KT through an entropy maximization approach.

Acquiring information at various time intervals is crucial for versatile usage of these datasets and analyzing population flow characteristics across different regions and time periods. Namely, given the potential to capture more detailed movement with shorter intervals in spatiotemporal demographic data, there is a compelling need to devise a spatiotemporal interpolation method. Numerous methods exist for interpolating time intervals in data. Tatsumi and Matsuba (2008) summarized several, including linear interpolation, neighborhood interpolation, and random interpolation. Kitakawa (2003) proposed a linear Gaussian state space model to address essential problems in time series analysis such as interpolation. Another study utilized population flow data at short intervals, combining the shortest route from departure to arrival areas and railway/road network data, revealing correlations between the number of stores and time slot population (Shimazaki et al., 2009). The fractal dimension was also employed in a method for generating interpolation values for economic time series (Chiba and Matsuba, 2005).

Building upon an integration method for multiplying spatiotemporal demographic data, this research introduces a novel spatiotemporal interpolation method designed to observe the spatiotemporal distribution of population flows at shorter time intervals. Assuming constancy in population flows within arbitrary time intervals, the population flows at shorter time intervals is computed through the square root calculation of the movement probability matrix. For regions exhibiting relatively large estimation errors, a selection process is proposed when utilizing actual data. Subsequently, utilizing the short time interval data derived from the spatiotemporal interpolation method, the spatial distribution of population flows in several regions is explored during various time periods. This exploration aims to capture more nuanced details regarding the moving ranges and directions of individuals within urban areas.

2. METHODOLOGY

2.1 Existing Spatiotemporal Interpolation Method

In existing interpolation methods, the linear interpolation method is commonly utilized for various spatiotemporal demographic data. For example, Osaragi and Hayasaka (2019) pre-processed MSS, PT data, and KT based on two assumptions to obtain shorter interval data: Assumption (1) is linear changes in population within a grid-cell between time t and $t+\Delta t$, while Assumption (2) is population flows follow a simple Markov

process based on a movement probability matrix (Figure 1). Linear interpolation was applied to derive population and static fraction at any desired time, and moving fraction in various time intervals was obtained by multiplying the movement probability matrix to standardize the data's time intervals. Subsequently, the population served as a constraint to differentiate between static individuals and population flows, with inter-grid-cell moving population estimated via the moving fraction using a maximum likelihood algorithm.

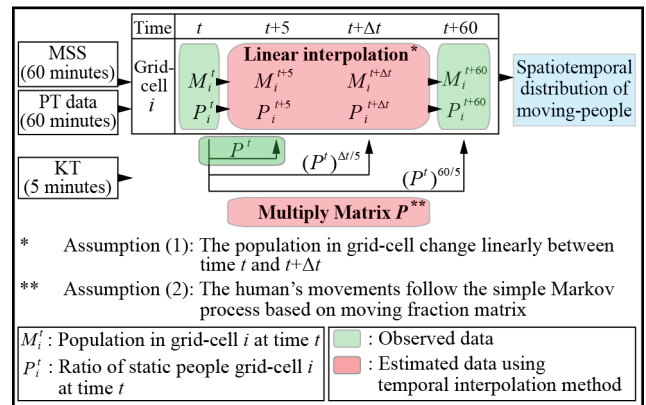


Figure 1. Standardization of the time intervals across datasets.

The preceding study obtained the spatiotemporal distribution of population flows at shorter time intervals based on the two assumptions. However, excessive reliance on assumptions in estimation procedures may lead to the generation of unrealistic results. Especially, Assumption (1) is considered hard to compatible for population flow estimation. By loosening Assumption (1) and relying exclusively on Assumption (2), we develop a method that increases its versatility and facilitates flexible application across various contexts and objects.

2.2 Proposing Spatiotemporal Interpolation Method

The spatiotemporal interpolation method is depicted in Figure 2. The variables employed in the calculation are as follows:

- M_i^t : The population in grid-cell i at time t .
- $s_{ij}^{t-t+\Delta t}$: The number of people move from grid-cell i to grid-cell j between time t and $t+\Delta t$ including $s_{ii}^{t-t+\Delta t}$ ($s_{ii}^{t-t+\Delta t}$ is the number of people stay in the grid-cell i).
- $f_{ij}^{t-t+\Delta t}$: The inter-grid-cell moving fractions from grid-cell i to grid-cell j between time t and $t+\Delta t$ including $f_{ii}^{t-t+\Delta t}$ ($f_{ii}^{t-t+\Delta t}$ is the fraction of people stay in the grid-cell i).
- $S^{\Delta t}$: The moving population matrix in time interval Δt whose element is $s_{ij}^{t-t+\Delta t}$.
- $F^{\Delta t}$: The movement probability matrix in time interval Δt whose element is $f_{ij}^{t-t+\Delta t}$.

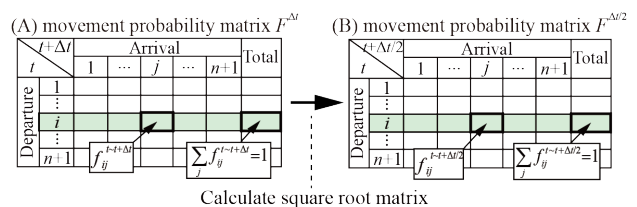


Figure 2. Comprehensive overview of the spatiotemporal interpolation method.

This research employed data from PT data, MSS and KT, which showed the spatiotemporal distribution of population flows

(Osaragi and Hayasaka, 2019). According to this data, we obtained the inter-grid-cell moving population $s_{ij}^{t-t+\Delta t}$ in time interval Δt ($\Delta t=60$ minutes). Using the method proposed in their study, the inter-grid-cell moving fraction $f_{ij}^{t-t+\Delta t}$ could be estimated using the population of grid-cell i (M_i^t and $s_{ij}^{t-t+\Delta t}$).

$$M_i^t = \sum_j s_{ij}^{t-t+\Delta t} \quad (1)$$

$$f_{ij}^{t-t+\Delta t} = s_{ij}^{t-t+\Delta t} / M_i^t \quad (2)$$

Next, we considered the population flows that composed of $s_{ij}^{t-t+\Delta t}$, and the movement probability matrix composed of $f_{ij}^{t-t+\Delta t}$. The same as their study, we assumed that inter-grid-cell moving fraction $f_{ij}^{t-t+\Delta t}$ is constant in any arbitrary time interval according to the simple Markov process based on movement probability matrix $F^{\Delta t}$. Under this assumption, the movement probability matrix in time interval $\Delta t/2$, could be obtained by calculating the square root of movement probability matrix $F^{\Delta t}$. The inter-grid-cell moving fraction $f_{ij}^{t-t+\Delta t/2}$ in time interval $\Delta t/2$ is the element of $F^{\Delta t/2}$.

$$(F^{\Delta t})^{1/2} = F^{\Delta t/2} \quad (3)$$

Finally, the number of people $s_{ij}^{t-t+\Delta t/2}$ who move from grid-cell i to grid-cell j between time t and $t+\Delta t/2$ can be obtained using M_i^t and movement probability matrix $F^{\Delta t/2}$ whose element is $f_{ij}^{t-t+\Delta t/2}$.

$$s_{ij}^{t-t+\Delta t/2} = f_{ij}^{t-t+\Delta t/2} \times M_i^t \quad (4)$$

2.3 Method for Calculating Square Root of Movement Probability Matrix

2.3.1 Denman-Beavers iteration: Various methods have been employed for calculating the square root of a matrix, with the Schur method and Newton's method standing out as widely utilized approaches (Björck and Hammarling, 1983). The Schur method involves simplifying the matrix to Schur triangular form, enabling the computation of the square root by evaluating the square root of the triangular matrix (Deadman et al., 2013). While this method is numerically stable, it has the potential to yield numerous imaginary numbers in the square root matrix, particularly when dealing with multidimensional matrices. Alternatively, the square root of a matrix can be computed using Newton's method, which is expressed as iterative equations. Despite its favorable mathematical convergence properties, Laasonen (1958) highlighted that Newton's method is numerically unstable in certain contexts.

To address this instability, Denman and Beavers (1976) introduced iteration equations as an extended variant of Newton's method, hereafter referred to as the DB method,

$$\left. \begin{aligned} Y_{k+1} &= \frac{Y_k + Z_k^{-1}}{2} \\ Z_{k+1} &= \frac{Z_k + Y_k^{-1}}{2} \end{aligned} \right\}, \quad k = 0, 1, 2, \dots \quad (5)$$

which considered matrix Y_0 is movement probability matrix $F^{\Delta t}$, matrix Z_0 is an identity matrix I whose dimension is the same as matrix Y_0 , Y_k^{-1} and Z_k^{-1} are the inverse matrices of Y_k and Z_k respectively. k is the iteration time.

Higham (1986) and Cheng et al. (2001) independently validated that the DB method ensures the numerical stability of computing the square root of a matrix. Furthermore, the DB method demonstrated its capability to compute the square root of a matrix, yielding solutions exclusively composed of real numbers, all within a concise calculation timeframe. In this research, the DB method was employed to calculate the

approximate square root matrix, relying on convergence calculations with matrices $F^{\Delta t}$ and I after undergoing k iterations.

2.3.2 Methodology for determining calculation iterations:

Round-off errors may occur during the calculation of inverse matrices in the division step of the DB method. These errors can lead to a floating result in the iteration calculations to some extent. Additionally, there is a potential for the accumulation of round-off errors with increasing iteration times, denoted as k . Consequently, it becomes necessary to establish the appropriate iteration times to achieve an accurate approximation of the square root of the matrix. Figure 3 illustrates the process of determining calculation iterations. Initially, employing matrix $F^{\Delta t}$ and matrix I with the DB method, we computed matrix Y^{k+1} and matrix Z^{k+1} for N iterations, where N represents the maximum number of iteration times. Subsequently, it was acknowledged that matrix F^{k+1*} was the square of matrix Y^{k+1} .

$$F_{k+1}^* = (Y_{k+1})^2 \quad (6)$$

Next, considering that F_{k+1}^* is a matrix with elements represented by $\hat{f}_{ij}^{t-t+\Delta t}$, we calculated the norm $\|F_{k+1}^*\|$ between matrix F_{k+1}^* and movement probability matrix $F^{\Delta t}$ over the time interval Δt .

$$\|F_{k+1}^*\| = \sqrt{\sum_i \sum_j (f_{ij}^{t-t+\Delta t} - \hat{f}_{ij}^{t-t+\Delta t})^2} \quad (7)$$

Comparing each norm from 1st time to N -th time, the Y_{k+1} with the smallest norm was considered to be the approximate square root of matrix $F^{\Delta t}$. By comparing the norms obtained from the 1st iteration to the N -th iteration, the Y^{k+1} corresponding to the smallest norm is deemed to be the approximate square root of the matrix $F^{\Delta t}$.

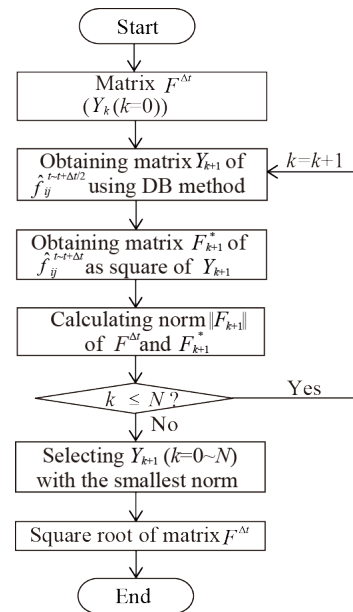


Figure 3. Process of determining the calculation intervals.

2.3.3 Adjustment for negative values in the square root matrix:

Due to the limitation that individuals can only move to a restricted number of grid-cells within the time interval Δt , the movement probability matrix $F^{\Delta t}$ inherently includes numerous 0 and small values. As a consequence of round-off errors stemming from the abundance of 0 or small values in matrix $F^{\Delta t}$, the matrix $F^{\Delta t/2}$ derived from $F^{\Delta t}$ often incorporates numerous negative values due to the effects of iterative calculations.

Figure 4 illustrates the process of adjusting negative values in the estimated approximate square root matrix of $F^{\Delta t/2}$. In instances where an element $\hat{f}_{ij}^{t-t+\Delta t/2}$ of matrix $F^{\Delta t/2}$ was a negative number, it was adjusted to 0. Nevertheless, following the adjustment of negative values, the summation of elements in each row of the matrix might exceed 1. Consequently, an additional adjustment was made to ensure that the sum of elements in each row equaled 1.

$$f_{ij}^{t-t+\Delta t/2} = \frac{\hat{f}_{ij}^{t-t+\Delta t/2}}{\sum_j \hat{f}_{ij}^{t-t+\Delta t/2}} \quad (8)$$

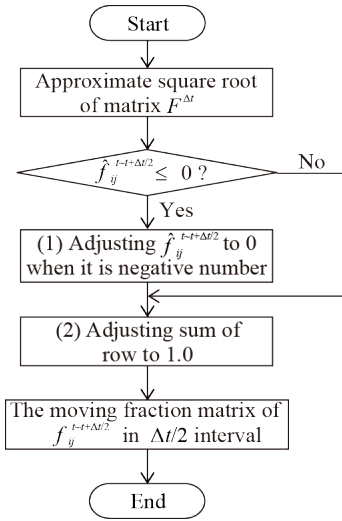


Figure 4. Method for adjusting negative values

Through the aforementioned procedures, we can derive the movement probability matrix $F^{\Delta t/2}$, wherein each element represents the inter-grid-cells moving fraction $f_{ij}^{t-t+\Delta t/2}$ during the time interval $\Delta t/2$. Subsequently, employing the methodology detailed in Section 2.2, we can determine the inter-grid-cell moving population $s_{ij}^{t-t+\Delta t/2}$.

Initially, based on the movement probability matrix $S^{\Delta t}$ during time interval t , two distinct movement probability matrices can be computed: one matrix in which the sum of elements in each row equals 1 (Figure 5A), and another movement probability matrix where the sum of elements in each column equals 1 (Figure 5B).

Subsequently, the square root of these two movement probability matrices (Figure 5 (C), (D)) was computed to derive movement probability matrices and corresponding moving population matrices for the time interval $\Delta t/2$. The summation of elements in each column in the matrix was denoted as $M_i^{*t+\Delta t/2}$, and the summation of elements in each row was denoted as $M_i^{**t+\Delta t/2}$. The population of grid i , at time $\Delta t/2$, denoted as $M_i^{t+\Delta t/2}$, was considered the average of the two estimated populations $M_i^{*t+\Delta t/2}$ and $M_i^{**t+\Delta t/2}$.

$$M_i^{*t+\Delta t/2} = \sum_j s_{ij}^{t-t+\Delta t/2} \quad (9)$$

$$M_i^{**t+\Delta t/2} = \sum_j h_{ij}^{t+\Delta t/2-t+\Delta t} \quad (10)$$

$$M_i^{t+\Delta t/2} = \frac{1}{2}(M_i^{*t+\Delta t/2} + M_i^{**t+\Delta t/2}) \quad (11)$$

We formulated Equations (12) and (13) to characterize the relationship between the population at time t , M_i^t , and at time

$t+\Delta t/2$, $M_i^{t+\Delta t/2}$, the inter-grid-cell movement $s_{ij}^{t-t+\Delta t/2}$, where n represents the number of grid-cells.

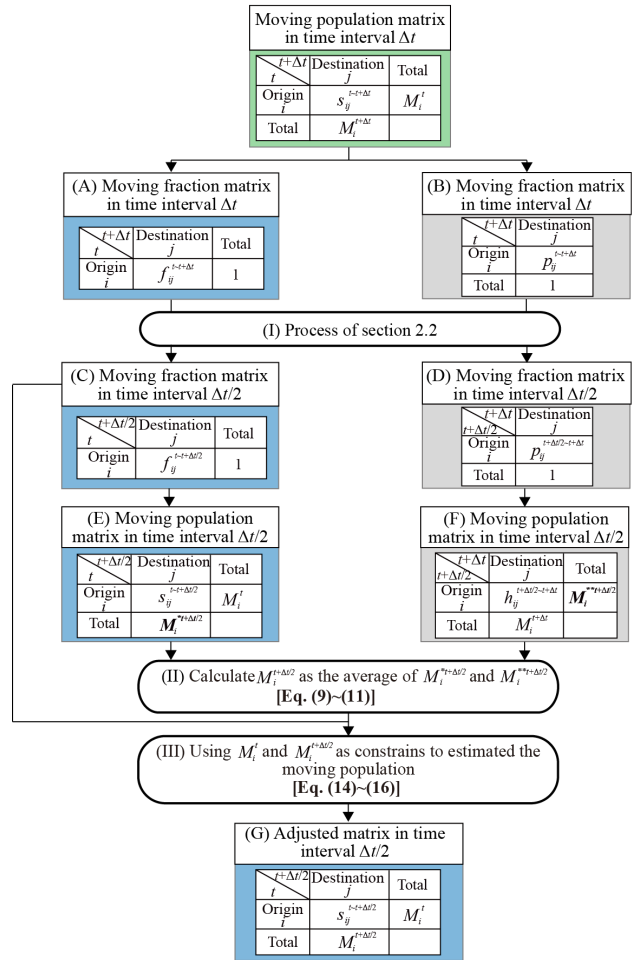


Figure 5. Procedure for refining the movement probability matrix within a shorter time interval.

$$M_i^t = \sum_j s_{ij}^{t-t+\Delta t} \quad (12)$$

$$M_i^{t+\Delta t/2} = \sum_j s_{ij}^{t-t+\Delta t/2} \quad (13)$$

The calculation of the number of people moving between grid-cells, $s_{ij}^{t-t+\Delta t/2}$, was carried out by utilizing the inter-grid-cell moving fractions $f_{ij}^{t-t+\Delta t/2}$ between time t and $t+\Delta t/2$ based on the maximum likelihood estimators, which yield the highest values for the occurrence probabilities.

$$s_{ij}^{t-t+\Delta t/2} = f_{ij}^{t-t+\Delta t/2} \times A_i^t \times B_j^t \quad (14)$$

$$A_i^t = \frac{M_i^t}{\sum_j f_{ij}^{t-t+\Delta t/2} \times B_j^t} \quad (15)$$

$$B_j^t = \frac{M_j^{t+\Delta t/2}}{\sum_i f_{ij}^{t-t+\Delta t/2} \times A_i^t} \quad (16)$$

The variables A_i^t and B_j^t exhibit mutual dependence; however, arbitrary initial values were selected for a converging calculation, yielding a unique value for the number of individuals moving between grid-cells $s_{ij}^{t-t+\Delta t/2}$.

3. CASE STUDY

3.1 Study Area and Data

The analysis focused on the Tokyo 23 wards, where human movements are notably active, and the study area was divided into grid-cells of spatial units measuring 500 m by 500 m. Demographic data from the Mobile Spatial Statistics (MSS) on a weekday in October 2008 were utilized (Figure 6). Individual locations were estimated from PT data at 60-minute intervals, with the PT data from weekdays being aligned with time use survey results. Given the low sampling rates for KT, the data from multiple days were amalgamated. Days with anomalies due to natural causes and those coinciding with significant public events were excluded, resulting in approximately half a year's worth of data from which one day's worth of weekday data was extracted.

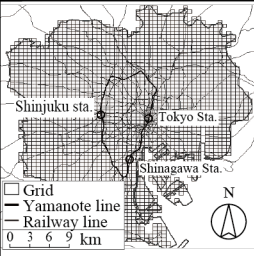
Time ranges	3:00–26:00	
Study area	Tokyo 23 wards divided into a 500m grid-cell	
Data for analysis	KT (223 normal weekdays from April. 1, 2015–Mar. 31, 2016), MSS (Oct. 20, 2015) and PT data (weekday in Oct. 2008)	

Figure 6. Study area and dataset employed for analysis.

Furthermore, an external region outside the study area, encompassing urban zones and water spaces, was considered. A considerable number of individuals were observed to move between this external region and the study area. Consequently, an external grid was established to depict the broader region beyond the study area. This external grid played a crucial role in the spatiotemporal distribution of population flows, ensuring the conservation of the total population across the entire region.

3.2 Excluding Boundary Effects from Estimated Data

Given the presence of an external grid representing a wide region outside the study area, a substantial movement of people occurred to/from this external grid, making its population significantly larger than that of any other grid-cells. Additionally, a considerable number of individuals engaged in inflow/outflow movements between the external grid and the grid-cells close to the boundary line of the study area, which we termed as boundary grids. To distinguish estimation results with potentially large errors, we introduced a method for defining boundary grids (Figure 7). Initially, we identified the boundary line of the study area by inputting the polyline of the Tokyo 23 wards boundary into ArcGIS, focusing on the boundary adjacent to other urban areas (Figure 7(a)). The coastline boundary was omitted from selection, as there were minimal movements of people to/from the external grid in the study area.

Next, to identify grid-cells in close proximity to the boundary line, we generated a buffer zone with a width d [m] from the boundary line (Figure 7(b)). For this study, d [m] was considered equivalent to the size of a grid-cell (500 m). Subsequently, grid-cells overlapping with the buffer zone were selected as boundary grids (Figure 7(c)). It is essential to acknowledge that the estimated results of population flows to/from these boundary grids may contain larger estimation errors than other grid-cells—a phenomenon termed as boundary

effect. Due to the reduced reliability associated with the boundary effect, we proactively excluded these boundary grids when assessing accuracy in the subsequent section.

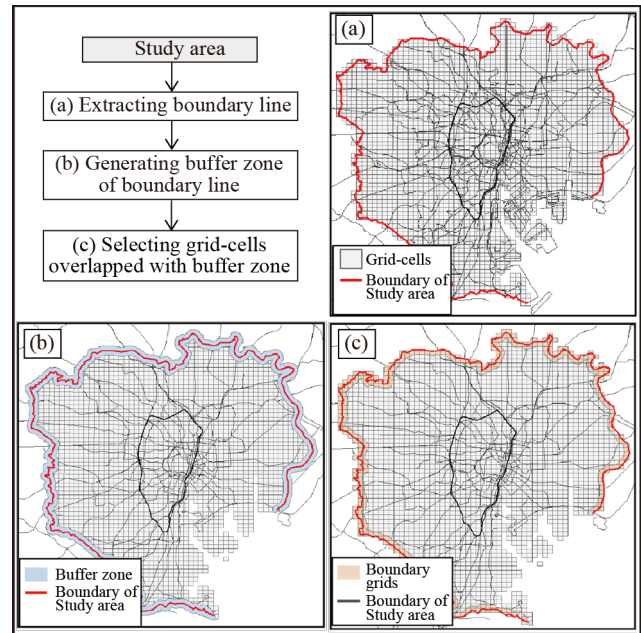


Figure 7. Method to eliminate the impact of boundary effects.

3.3 Validation of Estimate Accuracy

Given the absence of precise data depicting the number of people in fine spatiotemporal units, a method was devised to evaluate the accuracy of the spatiotemporal interpolation approach. Figure 8 illustrates the validation process of estimate accuracy. For this validation, we specifically choose data from the time frame between 14:00–16:00 to exclude morning and evening rush hours characterized by large fluctuations in population flow. This time window was selected to mitigate the impact of dynamic changes and enable a more focused evaluation of the accuracy of our estimates. The ensuing sections provide a detailed account of the validation process, its outcomes, and the implications for the reliability of our spatiotemporal interpolation method.

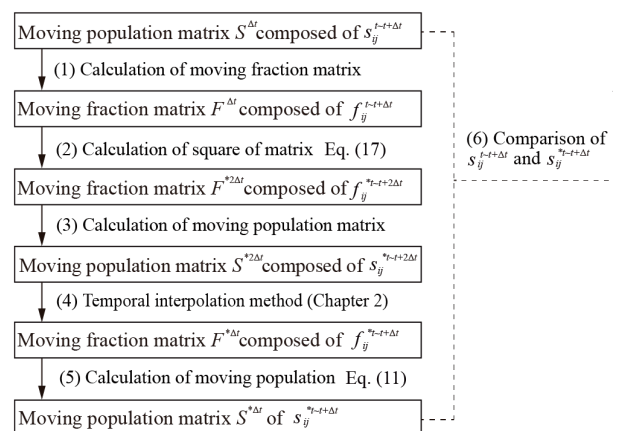


Figure 8. The procedure for evaluating accuracy.

Based on the data acquired during the 14:00–15:00 period ($\Delta t=60$ minutes), a movement probability matrix $S^{\Delta t}$ was derived.

We represented each element of this matrix $s_{ij}^{14:00-15:00}$, the population of grid i at 14:00 $M_i^{14:00}$, and the inter-grid moving fraction $f_{ij}^{14:00-15:00}$ (with the elements in every row summing to 1) (Figure 8(1)).

Initially, under the assumption that the moving fraction $f_{ij}^{15:00-16:00}$ between 15:00~16:00 is the same as $f_{ij}^{14:00-15:00}$ between 14:00~15:00, based on the simple Markov process, we estimated the movement probability matrix $F^{*2\Delta t}$. The elements of this matrix, denoted as $f_{ij}^{*14:00-16:00}$ between 14:00~16:00, were obtained by squaring the elements of the movement probability matrix $F^{\Delta t}$ representing $f_{ij}^{14:00-15:00}$ (Figure 8(2)).

$$F^{*2\Delta t} = (F^{\Delta t})^2 \quad (17)$$

Subsequently, the movement probability matrix $S^{*2\Delta t}$, representing $s_{ij}^{*14:00-16:00}$ between 14:00~16:00, could be derived from the matrix $F^{*2\Delta t}$ (Figure 8(3)). The estimated population of grid i at 16:00 ($M_i^{*16:00}$) could be obtained using the following equation:

$$M_i^{*16:00} = \sum_j S_{ij}^{*14:00-16:00} M_j^{14:00} \quad (18)$$

Utilizing the spatiotemporal interpolation method proposed in Section 2, we can ascertain the probability matrix of inter-grid-cell movement, denoted as $F^{*\Delta t}$, with elements representing $f_{ij}^{*14:00-15:00}$. Concurrently, the population matrix $S^{*\Delta t}$ can be estimated, featuring elements $s_{ij}^{*14:00-15:00}$, for the time interval between 14:00~15:00 (Figure 8(4)~(5)). Subsequently, the accuracy assessment involves a comparison between the elements $s_{ij}^{14:00-15:00}$ and $s_{ij}^{*14:00-15:00}$ of $S^{\Delta t}$ and $S^{*\Delta t}$ (Figure 8(6)).

Figure 9 presents a comparison between the moving and static populations during the 14:00~15:00 timeframe, derived from both observed and estimated data. In order to mitigate large estimation errors arising from boundary effects, the estimated values pertaining to the moving and static populations to/from boundary grids were precluded. The results indicate overall good accuracy in the estimated values. However, a detailed examination reveals instances where the estimated populations of inter-grid-cell movement and static populations within grid-cells were underestimated. This discrepancy can be attributed to the overestimation of populations moving to external grids through the employed comparison method, particularly in scenarios involving population flows between external grids and boundary grids.

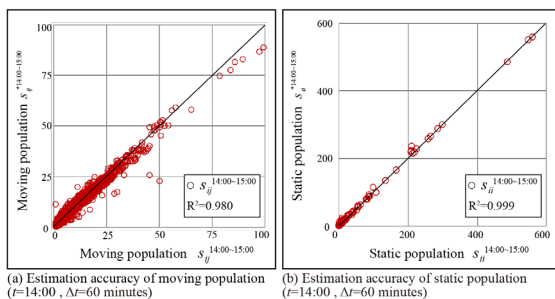


Figure 9. Precision of estimating both moving and static populations.

3.4 Spatiotemporal Distribution of Population Flows in Urban Space

Figure 10 illustrates the spatiotemporal distribution of population flows to/from the grid-cell containing Shinjuku station at intervals of 60 minutes, 30 minutes, 15 minutes, and 7.5 minutes during the 8:00~9:00 period. The visualization

reveals a diffusion of both inflow and outflow people over time, indicating an expanding moving distance and range.

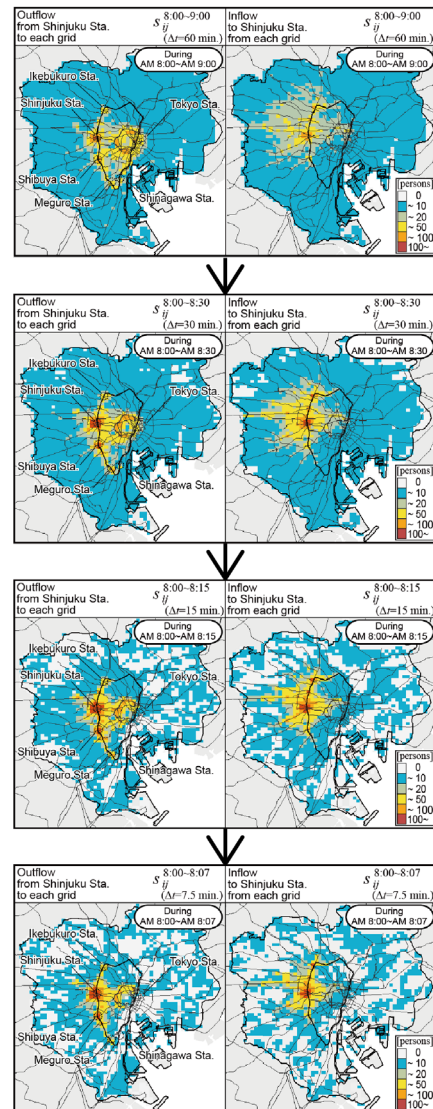


Figure 10. Spatial distribution of population flows to/from Shinjuku Station during different time intervals.

Notably, the population flow range is considerably broader with the 60-minute interval data compared to the 7.5-minute interval data. The 60-minute data demonstrates that individuals migrate to more distant grid-cells within Tokyo 23 wards. In contrast, the 7.5-minute data highlights a concentration of population flows in close proximity to railway lines connected to Shinjuku station. The shorter time interval data reveals a shorter moving distance for population flows.

Figure 11 depicts the spatial distribution of population flows from Shinjuku Station at intervals of 60 minutes and 7.5 minutes during various time periods. The 60-minute data readily highlights the impact of time periods on population flows. Specifically, during peak commuting hours (8:00~9:00 and 17:00~18:00), the movement range of individuals was notably broader compared to the daytime period (12:00~13:00). Conversely, utilizing shorter time interval data allowed for the observation of regional characteristics. For instance, the 7.5-minute interval data reveals that the movement range of individuals from Shinjuku Station predominantly aligns along

the railway lines. This observation indicates that individuals departing from Shinjuku Station primarily utilize the railway system. Given that Shinjuku Station serves as a major hub interconnected with various stations and regions through an extensive network of railway lines, it is more convenient for individuals departing from Shinjuku Station to rely on these railway connections.

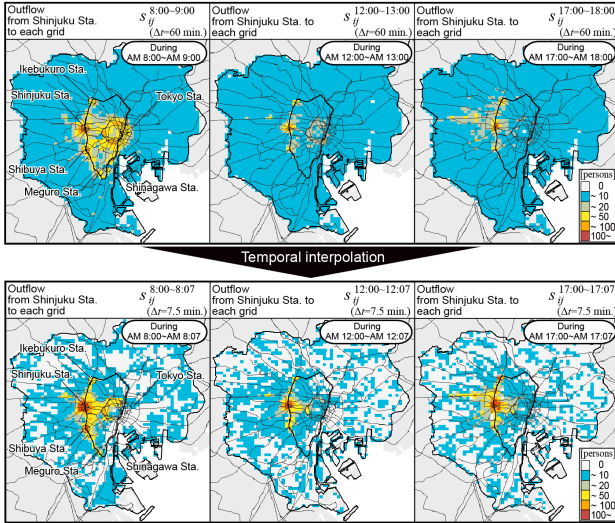


Figure 11. Spatial distribution of population flows from Shinjuku Station at intervals of 60 minutes and 7.5 minutes during different time periods.

4. COMPARISON OF EXISTING AND PROPOSED METHOD

Given the existence of a previously proposed spatiotemporal interpolation method in Osaragi and Hayasaka (2019), we attempted to compare the results obtained from the two distinct spatiotemporal interpolation methods, i.e., the existing method and our method proposed in this research.

This section specifically focuses on data selected from the time period between 14:00~16:00, characterized by relatively stable population flow. From the acquired data, the population of grid-cell i at 14:00 ($M_i^{14:00}$) and 16:00 ($M_i^{16:00}$) was derived from the Mobile Spatial Statistics (MSS), while the static occupant fraction at 14:00 ($P_i^{14:00}$) and 16:00 ($P_i^{16:00}$) was obtained from Person Trip (PT) data. The population and static occupant fraction at 15:00 were estimated through linear interpolation utilizing the following equations:

$$M_i^{15:00} = M_i^{14:00} + \frac{1}{2}(M_i^{16:00} - M_i^{14:00}) \quad (19)$$

$$P_i^{15:00} = P_i^{14:00} + \frac{1}{2}(P_i^{16:00} - P_i^{14:00}) \quad (20)$$

Subsequently, the inter-grid-cell moving population ($s_{ij}^{**14:00-15:00}$) between 14:00~15:00 was estimated by leveraging the population and static occupant fraction data for each grid-cell at 14:00 and 15:00, along with the moving fraction ($p_{ij}^{14:00-15:00}$) from KT. This estimation process facilitates a direct comparison between the moving fraction ($f_{ij}^{**14:00-15:00}$) and the previously estimated ($f_{ij}^{*14:00-15:00}$) from Section 3.3.

Figure 12 presents a comparative analysis of results obtained through two spatiotemporal interpolation methods. The x-axis represents the estimated outcomes using the method proposed in

Osaragi and Hayasaka (2019), while the y-axis corresponds to the estimated outcomes using the method introduced in this research. The comparison reveals a high degree of similarity in the distribution of population flows estimated by these two methods. However, it is noteworthy that the spatiotemporal interpolation method introduced in this research generally yields lower estimates for the population flows in comparison to the results obtained by the existing method.

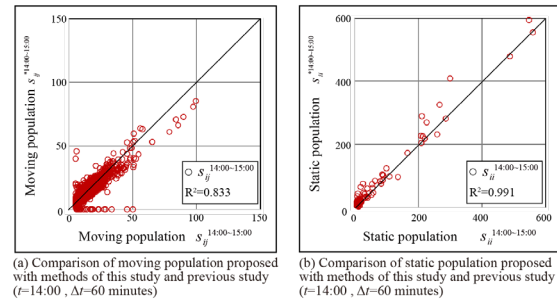


Figure 12. Comparison of the estimated results obtained through the existing method and our method.

Figure 13 illustrates the spatiotemporal distribution of population flows at 15-minute and 30-minute intervals using two distinct spatiotemporal interpolation methods. A noticeable trend emerges wherein the moving population estimated in this research is consistently lower than the moving population estimated using the existing method. Conversely, static population estimated in this research tends to be higher than the estimates for individuals staying in place as per the existing method.

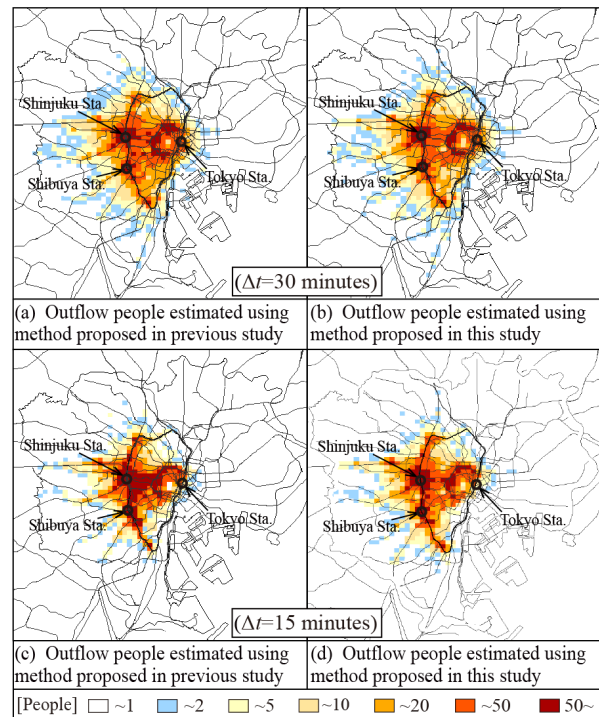


Figure 13. Comparison of outflow population estimates obtained through the existing method and our method proposed in this research.

5. DISCUSSION AND CONCLUSIONS

Based on the methodology developed for estimating the spatiotemporal distribution of population flows, this research introduces a novel spatiotemporal interpolation method aimed at estimating such distributions within shorter time intervals. Specifically, the population flow distribution in shorter time intervals is derived by estimating the square root of the movement probability matrix obtained from the moving population. The results of accuracy evaluation indicate that the estimation outcomes exhibit commendable accuracy.

Subsequently, utilizing the estimated data at shorter time intervals, this research elucidated the nuanced patterns of human movement within Tokyo 23 wards, delving into the characteristics of population flows in shorter time increments. Through an examination of the moving ranges and directions of individuals within shorter time intervals, this research revealed the capacity to discern differences in spatiotemporal distribution of population flows, influenced by transportation modes, geographical regions, and temporal variations.

The analysis underscored the critical importance of comprehending detailed human movement patterns in the development of simulation models for population flow distribution post-disasters or during special events. A notable limitation of this paper is the lack of real-world implementation of the proposed method in urban settings, especially in critical scenarios like emergency situations and traffic management or policy-making. Additionally, there is a need to explore the effectiveness of the methodology in different contexts or cities as part of future research efforts. For a prospective avenue for real-world implementation, the proposed spatiotemporal interpolation method is essential for enhancing estimation accuracy in consecutive steps of interpolation. With the availability of shorter interval data, there is potential to establish models for estimating human movements that dynamically respond to varying traffic conditions, thus playing a pivotal role in facilitating rapid responses to urban emergencies.

ACKNOWLEDGEMENTS

This paper constitutes part of the research outcomes supported by KAKENHI (Grant Number 23K26347). Data used in this research are not available because of commercial restrictions.

REFERENCES

- Björck, Å., Hammarling, S., 1983. A Schur method for the square root of a matrix, *Linear Algebra Appl.*, 52/53, 127-140.
- Calabrese, F., DiLorenzo, G., Liu, L., Ratti, C., 2011. Estimating Origin-Destination Flows Using Opportunistically Collected Mobile Phone Location Data from One Million Users in Boston Metropolitan Area, *IEEE Pervasive Computing*, 10(4), 36-44.
- Cheng, S.H., Higham, N.J., Kenney, C.S., Laub, A.J., 2001. Approximating the logarithm of a matrix to specified accuracy, *SIAM Journal on Matrix Analysis and Applications*, 22(4), 1112-1125.
- Chiba, R., Matsuba, I., 2005. Interpolation Financial Time Series using Fractal Dimension, *IEICE technical report*, 105(277), 19-24.
- Deadman, E., Higham, N. J., Ralha, R., 2013. Blocked Schur algorithms for computing the matrix square root, *Lecture Notes in Comput. Sci.*, 7782, Springer-Verlag, 171–182.
- Denman, E.D., Beavers, A.N., 1976. The matrix sign function and computations in systems, *Applied Mathematics and Computation*, 2, 63-94.
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, R.F., Gaughan, E.A., Blondel, D.V., Tatem, J.A., 2014. Dynamic Population Mapping Using Mobile Phone Data, *Proceedings of the National Academy of Sciences of the United States of America*, 111(45), 15888-15893.
- Higham, N.J., 1986. Newton's Method for the Matrix Square Root, *Mathematics of Computation*, 46(174), 537-549.
- Kamada, K., 2017. Toshikotsunnya ni okeru konzatutoukeideta no katsuyou ni tsuite, *Meeting of Ministry of Land, Infrastructure, Transport and Tourism Kinki Regional Development Bureau*, 19.
- Kitagawa, G., 2003. Numerical Methods for Time Series Analysis: A Computing Statistics Approach, *Bulletin of the computational statistics of Japan*, 15(2), 159-170.
- Laasonen, P., 1958. On the iterative solution of the matrix equation $AX^2 - I = 0$, *Mathematical Tables and Other Aids to Computation*, 12(62), 109-116.
- Osaragi, T., Hayasaka, R., 2020. Prediction of Spatiotemporal Distributions of Transient Urban Populations with Statistics Gathered by Cell Phones, *Proc. of GISTAM 2020, SCITEPRESS*, Science and Technology Publications, 1, 33-44.
- Osaragi, T., Kudo, R., 2019. Enhancing the Use of Population Statistics Derived from Mobile Phone Users by Considering Building-Use Dependent Purpose of Stay, *Geospatial Technologies for Local and Regional Development*, Springer, Cham, 185-203.
- Ratti, C., Pulselli, R. M., Williams, S., Frenchman, D., 2006. Mobile Landscapes: Using Location Data from Cell-Phones for Urban Analysis, *Environment and Planning B: Planning and Design*, 33(5), 727-748.
- Seike, T., Mimaki, H., Hara, Y., Odawara, T., Nagata, T., Mayayuki, T., 2011. Research on the Applicability of "Mobile Spatial Statistics" for Enhanced Urban Planning, *Journal of the City Planning Institute of Japan*, 46(3), 451-456.
- Shimazaki, Y., Sekimoto, Y., Shibasaki, R., Akiyama, Y., 2009. A study on the correlation between the number of stores and the time slot population based on person flow-The comparative analysis between the time and space interpolated person trip survey data and the census data, *Journal of the City Planning Institute of Japan*, 44(3).
- Tatsumi, K., Matsuba, I., 2008. Interpolation methods of time series data: survey and critique, *Gakushuin University Research Institute of Economics and Management bulletin*, 22, 35-43.