

Research Article

Travel Matrix Decomposition for Understanding Spatial Long-Distance Travel Structure

Hiromichi Yamaguchi ¹, Mashu Shibata,² and Shoichiro Nakayama³

¹*Institute of Science and Engineering, Kanazawa University, Kakuma-Machi, Kanazawa, Ishikawa 920-1192, Japan*

²*Graduate, School of Science and Engineering, Kanazawa University, Kanazawa, Japan*

³*Institute of Transdisciplinary Sciences, Kanazawa University, Kanazawa, Japan*

Correspondence should be addressed to Hiromichi Yamaguchi; hyamaguchi@se.kanazawa-u.ac.jp

Received 15 September 2022; Revised 19 December 2022; Accepted 9 January 2023; Published 8 February 2023

Academic Editor: Hiroki Sayama

Copyright © 2023 Hiromichi Yamaguchi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mobile phone location data enable us to obtain accurate and temporally detailed long-distance travel distribution. However, the traditional long-distance travel distribution model cannot normally handle this detailed temporal information. This study proposes an approach for handling temporally detailed information of long-distance travel distribution. Considering this approach, the origin-destination matrix decomposes into two variables (indicators): destination amenity and travel cost. They can be interpreted as composite indicators of several variables that are treated in the travel-destination choice multinomial logit model. Because they are calculated only from the origin destination, we can discuss their detailed temporal variations. In this study, time changes in destination amenities and travel costs of interprefectural travel in Japan are calculated to confirm the value of this approach. These indicators have succeeded in describing the pattern of domestic long-distance travel in Japan. These quantified indicators have facilitated the understanding of the national land structure. They are useful as outcome measures for policy-making. Moreover, these indicators explain the temporal applicability of the destination choice model. Specifically, the results of destination amenities have a large seasonal variation. This indicates that the parameters of the destination amenity model (i.e., the coefficients of the destination variables) are not seasonally stable. Therefore, this must be considered when dealing with destination choice for long-distance travel.

1. Introduction

Travel behavior outside metropolitan areas is more difficult to characterize than that within the metropolitan area, which is often routine and stable. In this study, we call such travel behavior outside the metropolitan area “long-distance travel” and address its difficulty caused by two features of long-distance travel: the small frequency of trips per person and the large variation in travel frequency among individuals. These characteristics make surveys with random sampling questionnaires inefficient because most respondents do not undertake long-distance travel during the year. However, some people travel long distances at such a high frequency that they cannot easily write a yearly record. Thus, few surveys can accurately capture country-wide long-distance travel.

To solve these data limitations, several scholars have developed techniques to complement the whole from small-sized data. For example, the gravity model [1], which is traditionally commonly used in transportation models, describes simple regularity. It shows the relationship between trip distribution and the zone size/travel level of service (LOS). This regularity was pointed out more than 70 years ago [2]. This model is commonly used in recent studies (e.g., Lenormand et al. [3], Zhang et al. [4], Cordera et al. [5], Chai et al. [6], and Grosche et al. [7]). Erlander and Stewart [1] showed that the parameters of the gravity model can be reasonably estimated even with samples as small as 1,000. Although their calculations are for inner-city traffic, the results show the ability of the gravity model to complement and predict the entire origin-destination (OD) matrix with a small sample of data. The disaggregated

discrete choice model approach (e.g., Kato et al. [8], Yao and Morikawa [9], and Fu et al. [10]), which deals with travel destination choice, is of similar values. These models allow us to understand complementary travel distribution and predict its future values. The data required to estimate the coefficients are the disaggregated response results of the questionnaire. Moreover, the required number of samples for this estimation is sufficiently smaller than that for all OD traffic volume estimations.

Passively collected location data (e.g., mobile phone data) can solve the sample size problem. There is the available information on a large sample to create an OD matrix for long-distance travel. Moreover, several studies have already used mobile phone location data to study long-distance travel [11–14]. Therefore, the requirement for complementing the travel distribution by simple regularities is smaller than it has been in the past.

However, attempts to apply mobile phone location data to the traditional model have the following problems. First, some information that is readily available in the questionnaire is lacking. For example, information about the “destination,” purpose of the trip, and the transportation mode is not directly available. However, this will be less of a problem in the future because several approaches have already been proposed to infer this information (such as purpose and destination identification) [13, 15, 16] and travel mode identification [17]. Second, the temporal detail of the model is constrained by the explanatory variables. We can now use mobility information with temporal and spatial details. On the contrary, the socioeconomic and travel LOS indicators often do not have sufficient temporal and spatial resolutions. Consequently, the temporal detail of mobility data may be sacrificed in the traditional approach. Pitombo et al. [18, 19] propose a more accurate distribution model by applying decision tree algorithms. Nevertheless, it has the same limitation for applying it to the temporal detail. Therefore, several approaches have been proposed to find new patterns (regularities) by applying the factorization approach to mobility data [14, 20, 21].

Here, the combination of mobile phone location data and traditional models is used to search for newly available information while avoiding the second problem. This study proposes an approach based on this idea. Regarding this approach, temporally detailed information is treated by applying a simple description by the traditional travel-destination choice model. A key feature of this approach is that the two indicators, namely, destination amenities and travel costs, are estimated directly from mobile phone data, rather than as functions of variables such as travel time or population. Destination amenities and travel costs have different spatial impacts. Therefore, these indicators can be estimated by a simple assumption and an OD matrix in which most of the elements are filled. If multiday OD matrices are used, these indicators can be estimated for each of those days. Thus, we can determine detailed temporal changes in travelers’ perceptions of destination amenities and travel costs. This information is difficult to obtain using the model that describes them as a population or travel time function. This model differs from other travel destination choice models that use large-scale data records [22] because these indicators are estimated directly. The “pairwise

constant” and region dummy in the models of Zhu and Ye [23] and Kristoffersson et al. [24] are conceptually partly similar to the two indicators used here. Nonetheless, it has a different meaning and spatio-temporal treatment.

Here, the proposed approach decomposes the wealth of information from the large sample by focusing on spatial patterns so that the travel costs and amenities due to unknown factors are also reflected. Thus, compared with the traditional approach, the proposed approach is less capable of understanding the mechanism, but it provides additional information to understand the national land structure without missing any information in the hypothesis setting phase.

To confirm the value of this approach, we calculate the time changes in destination amenities and travel costs of interprefectural travel in Japan. The results confirm the following two values: (1) the simple indicator describes the pattern of long-distance travel in Japan and (2) the time variability of these indicators informs the temporal applicability of the destination choice model. Considering the first value, the information obtained here is consistent with the expected (known) pattern. This indicates that we have successfully quantified the destination amenities and travel costs as perceived by the traveler. In addition, for the travel-cost indicator, we succeeded in revealing the national land structure without generating complex LOS data for long-distance travel.

The second value clarifies the additional information obtained from the proposed approach. The time-series features of destination amenities require attention because this indicator is constantly changing, and it has a large seasonal variation. This indicates that the parameters of the destination amenity model (the coefficients of the destination variables), which are estimated by information biased toward a particular season, are difficult to apply to other seasons. This point must be considered when dealing with destination choice for long-distance travel. To address this concern, we can use mobile phone location data as a powerful data source because of these data’s capability to provide a sufficient sample at any given time. In addition, we also succeeded in characterizing some seasonal variations in people’s sensitivity to LOS.

The rest of this study is structured as follows: Section 2 presents the approach of this study, which contains the OD travel matrix data and the methodology used to calculate and analyze the destination amenities and travel costs. Section 3 discusses the static results derived from the OD matrix of all-time points to clarify the spatial characteristics of the two indicators. Section 4 illustrates the estimation of time trends of the two indicators based on four years of daily data (1,461 days), and it summarizes the characteristics of their seasonal changes. Section 5 concludes the study. Finally, in section 6, we discuss the value of the results obtained in this study and the scope of application of the proposed approach.

2. Research Design

2.1. Data. We use the OD matrix data from “Mobile Spatial Statistics” [25], a statistic of population distribution that uses network-driven mobile phone network records. “Mobile

Spatial Statistics” is also used in the studies of Kubo et al. [26], Yamaguchi and Nakayama [14], and Hara and Yamaguchi [27]. These aggregate data are generated from the operational data of 85 million cellular phones provided by NTT DOCOMO, Inc. Moreover, it enables the collection Japan’s population distribution with an approximate hourly frequency. We will analyze 1,461-day interprefectural OD matrices that are created using the stay and residence information in the operational data. This is expressed in the equation as follows:

$$\mathbf{Q}_d = \begin{pmatrix} q_{1,1,d} & \cdots & q_{1,j,d} & \cdots \\ \vdots & \ddots & \vdots & \ddots \\ q_{i,1,d} & & q_{i,j,d} & \\ \vdots & & \vdots & \ddots \end{pmatrix}, \forall (d \in D), \quad (1)$$

where d indicates the date (i.e., $D = [2014.3.1, \dots, 2018.2.28]$). In addition, $i, j \in Z$ denote residence and travel destination, respectively, and are aggregated for each of the 47 prefectures. Therefore, $q_{i,j,d}$ is the estimated number of people staying in zone j . Furthermore, their place of residence from the mobile phone operation data is denoted by i . This study will be conducted using only prefecture-based aggregate data; i.e., no personal information will be handled. These aggregate data were generated by NTT DOCOMO, Inc. from mobile phone operational data after necessary processing to protect privacy (deidentification, estimation, and disclosure limitation). Terada et al. [25] explain the detailed procedures.

Here, we analyzed the aggregate data \mathbf{Q}_d , which were generated from information during the specific time period 13:00–14:00. This is when people most likely stay at their travel destinations. This indicates that we assume that the travel destination is the prefecture, where the traveler is staying at 13:00–14:00. This assumption may deviate from people’s perceptions of the destination. This study presents an analysis under this assumption. The proposed approach does not depend on this time assumption, and we can apply it directly if more accurate and detailed OD matrices such as those of Bachir et al. [28] are obtained.

Thus, Z is the set of all 47 prefectures in Japan. In this study, interprefecture travel is treated as an example of long-distance travel. The administrative boundaries of prefectures shown in Figure 1 are historically decided; in Japan, most boundaries are set in straits or mountain ranges. As a result, many of the locations coincide with the boundaries of the daily living area, so this matrix shows a large number of unusual trips. As can be seen from the scale in Figure 1, most of the nonadjacent OD pairs are more than 100 km in linear distance, which is also close to the definition (larger than 100 km) adopted in Axhausen et al. [29].

2.2. Basic Characteristics of the Target OD Matrix. Here, we summarize the basic characteristics of the data used in the analysis. The number of elements analyzed is $47 \times (47 - 1) \times 1,461$. In this set, only 19,138 (0.61%) of the combinations were either not observed or confidential information owing

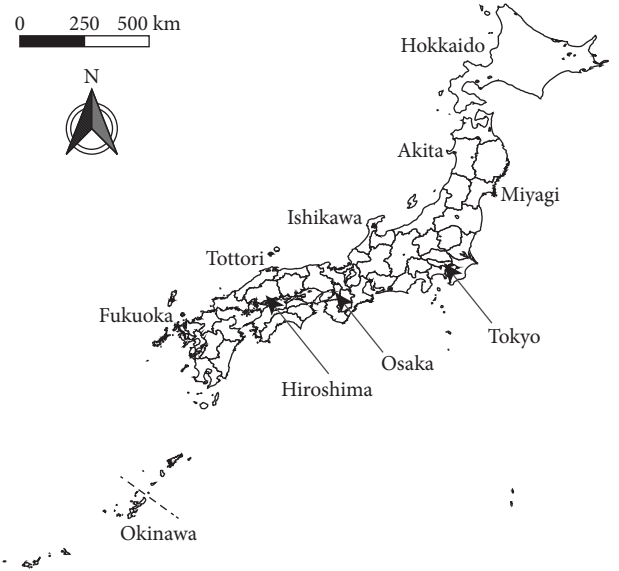


FIGURE 1: Target 47 prefectures.

to the small size of the sample. The data used in the analysis are mostly filled with numbers greater than zero.

Here, we consider the original day-to-day dynamics of the data. Figure 2 shows the day-to-day dynamics of the number of travelers for the two OD pairs over a one-year period. These figures show the following two characteristics of the time-series change for long-distance travel: (1) there are several peaks in the period. However, they differ among OD pairs. For example, the travel from Tokyo to Hokkaido (a) has large short-term peaks in May, August, and January. In contrast, the trip from Fukuoka to Tokyo (b) has relatively long and weak peaks in April and late January. (2) The volume of off-peak travel also varies with season, even on the same day of the week. For example, the travel from Hokkaido to Tokyo (a) in July and February differs significantly.

The time trends shown here illustrate the difficulty of the approach to classify the dates into several groups since it is difficult to establish a clear grouping. Special peaks are observed in more than five periods, as shown in Figure 2. The number of peaks may increase after checking other OD pairs. Furthermore, even if we focus on periods other than the peaks and consider each day of the week, seasonal differences still have their impact. In addition, seasonal variation is likely to be different for each OD pair. In contrast, our approach allows us to understand the characteristics of these complex temporal variations by extracting information on two types of spatial patterns at all-time points.

2.3. OD Matrix Decomposition Method. This section describes how to derive the two types of indicators by decomposing the OD matrices using the concept of the destination choice model.

We consider a multinomial logit model for long-distance travel destination choice as follows:

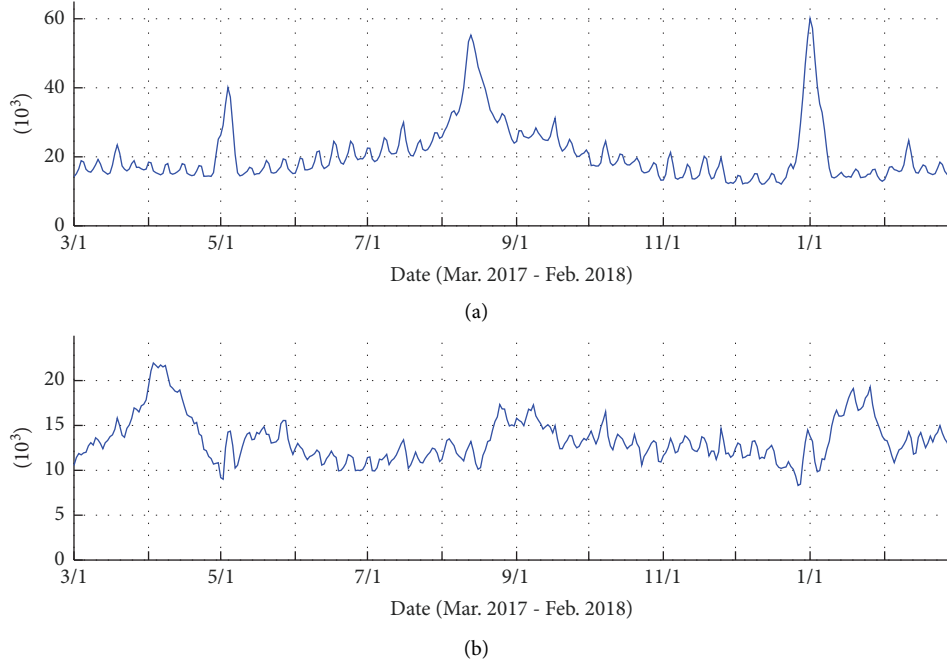


FIGURE 2: Day-to-day dynamics of the number of travelers for 2OD pairs. (a) (Residence zone and destination zone) = (Tokyo, Hokkaido). (b) (Residence zone and destination zone) = (Fukuoka, Tokyo).

$$P_{i,d}(j) = \frac{\exp(V_{i,j,d})}{\sum_{j \in Z} \exp(V_{i,j,d})}, \quad (2)$$

$$V_{i,j,d} = a_{i,j,d} + c_{i,j,d}, \quad (3)$$

where $P_{i,d}(j)$ is the probability that a resident of prefecture i will travel to prefecture j (stay at 13:00) on date d . This form is the most basic multinomial logit model [30]. If $i = j$, then it corresponds to the probability of “not traveling outside the residence prefecture.” These variables will be directly estimated in this study as unknown variables through maximum likelihood.

The systematic component of utility $V_{i,j,d}$ in most destination choice logit models is formulated as a linear sum of the relevant indicators multiplied by the coefficients. Considering the approach, socioeconomic indicators of destinations (destination variable) and the transportation level of service (LOS) are applied as the relevant indicators. For example, Marrocu and Paci [31] used the following variables as destination variables for destination choice models of long-distance travel: GDP, population density, natural environment (e.g., area of natural parks), and cultural and recreational attractions (e.g., number of museum visitors and restaurants). They also used the following variables as transportation LOS: geographical distance and prices. Similarly, Kato et al. [8] used the following variables as destination variables for destination selection in Japan: population, working population, GDP per capita, percentage of employees in the service sector, number of accommodations per area, and Hokkaido and Okinawa dummies. However, the temporal variation of these indices is smaller than that of the OD travel volume. Therefore, models using

the previously mentioned indices can only deal with temporally aggregated behavior, such as the total value for one year. They cannot deal with detailed temporal changes, such as equation (1).

In contrast, this study calculates two types of indices, namely, $(a_{i,j,d}$ and $c_{i,j,d})$, at each time point d . These two types of indicators are obtained by decomposing the OD matrix information into two types of spatial patterns. In this study, the same spatial pattern as in previous studies will be used. For example, the indicators used in the model of Marrocu and Paci [31] fall into the following two types of spatial patterns: indicators defined per destination location (GDP, population density, natural environment, and cultural and recreational attractions) and indicators defined per zone pair (transportation LOS). The approach proposed in this study estimates all indicator values themselves, applying only the constraints that indicate the two types of spatial patterns described previously, rather than population or other indicators. With this approach, two types of more accurate indicators can be calculated for each time point.

Thus, the time variation of interprefectural travel distribution can be explained by the time-series change of two calculated indices $(a_{i,j,d}$ and $c_{i,j,d})$. Here, the model’s form corresponds to the logit model of destination choice as in (2). Hence, the calculated indices $(a_{i,j,d}$ and $c_{i,j,d})$ can be interpreted as composite indices of multiple variables in several destination choice models.

The definitions of the two indicators $(a_{i,j,d}$ and $c_{i,j,d})$ are calculated in this study, and their meanings are explained as follows: first, $a_{i,j,d}$ is assumed to be a variable defined for each destination on each date, satisfying the following equations:

$$a_{i,j,d} = a_{k,j,d}, \forall (k, j, d) \in ([Z - \{j\}] \times Z \times D), \quad (4)$$

$$a_{i,i,d} = 0, \forall (i, d) \in (Z \times D). \quad (5)$$

Second, (4) indicates that the same value applies in the case of the same travel destination, excluding the $i = j$ component. Third, (5) shows that this indicator is zero when “not traveling outside the prefecture” (with $i = j$). Index $a_{i,j,d}$ is the composite index of the destination variables as described. This indicator is expected to account for cultural and recreational attractions, population, and economic size as pointed out by Marrocu and Paci [31]. Subsequently, we will refer to index $a_{i,j,d}$ as the “travel destination amenity.”

Let $c_{i,j,d}$ be a variable defined for each OD pair (i, j) and let it satisfy the following equations:

$$c_{i,j,d} = c_{j,i,d}, \forall (i, j, d) \in (Z \times [Z - \{j\}] \times D), \quad (6)$$

$$c_{i,i,d} = 0, \quad \forall (i, d) \in (Z \times D). \quad (7)$$

Furthermore, (6) shows that the variable remains constant though the OD direction is reversed. (7), together with (5), sets the deterministic utility of “not traveling outside the prefecture” to zero ($f_{i,i,d} = 0, \forall (i, d) \in (Z \times D)$). The variable set by Kato et al. [8] for each OD pair (i, j) , regardless of such direction, is a log-sum variable derived from the mode choice model. This variable is the expected utility of the mode choice result obtained by calculating and synthesizing transportation LOS, such as travel time, fare, frequency, and transit time for each mode. The $c_{i,j,d}$ calculated in this study is also an index calculated by synthesizing such multiple transportation LOSs. Subsequently, index $c_{i,j,d}$ will be referred to as “travel cost.”

In this study, these two types of variables are unknown, and the value that maximizes the log-likelihood is estimated as follows:

$$\begin{aligned} & \max_{(\mathbf{A}_d, \mathbf{C}_d)} \sum_{i \in Z} \sum_{j \in Z} (q_{i,j,d} \ln P_{i,d}(j)), \forall d \in D, \\ & \text{s.t. (2) - (7), (14),} \\ & \mathbf{A}_d = \begin{pmatrix} 0 & a_{1,2,d} & \cdots & a_{1,47,d} \\ a_{2,1,d} & 0 & \cdots & a_{1,47,d} \\ \vdots & \vdots & \ddots & \\ a_{2,1,d} & a_{1,2,d} & & 0 \end{pmatrix}, \quad (8) \\ & \mathbf{C}_d = \begin{pmatrix} 0 & c_{1,2,d} & \cdots & c_{1,47,d} \\ c_{1,2,d} & 0 & \cdots & c_{2,47,d} \\ \vdots & \vdots & \ddots & \\ c_{1,47,d} & c_{2,47,d} & & 0 \end{pmatrix}, \end{aligned}$$

where $(\mathbf{A}_d, \mathbf{C}_d)$ cannot be uniquely determined for all the variables owing to rank deficiency. This is because, in this model, the following relation holds for any k .

$$\begin{aligned} \mathbf{A}_d + \mathbf{C}_d &= (\mathbf{A}_d + k\mathbf{X}) + (\mathbf{C}_d - k\mathbf{X}). \\ \mathbf{X} = \mathbf{I} - \mathbf{E} &= \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \\ 1 & 1 & 1 & & 0 \end{pmatrix}. \quad (9) \end{aligned}$$

This issue can be resolved by fixing one variable, either \mathbf{A}_d or \mathbf{C}_d .

Therefore, we will estimate the unknown variables with one additional constraint:

$$a_{i,(j=\text{Tokyo}),d} = 0, \forall (i, d) \in (Z - \{j\} \times D). \quad (10)$$

This indicates that Tokyo’s destination amenity is fixed at zero, and all the other variables are calculated at their relative values. Notably, the destination amenity and travel cost presented as follows are relative values estimated under the constraints of (10). Thus, the absolute values are meaningless, and we need to focus only on spatial relationships of values. More care needs to be observed when comparing the same value to other periods since the change in value includes the change in destination amenity in Tokyo. In this study, we focus on indicators of spatial differences that are comparable even under this constraint.

The method described here is unique as it focuses on spatial patterns on the OD matrices. The two composite indices obtained here are identified only according to the spatial characteristics shown in the equations (4)–(7). The other basic approaches to analyze matrices (such as principal component analysis) cannot handle indices such as \mathbf{C}_d . This is because the basic matrix decomposition is focused on columns and rows. However, the travel OD matrix is characterized by the fact that its diagonal components have significantly different meanings from any other matrix (or table) data. Moreover, they are strongly related to elements whose directions are opposite. Our proposed approach is unique as it extracts components that are tailored to the nature of such travel information.

The proposed decomposition approach corresponds to a simple case of decomposing a matrix of odds ratios into two spatial patterns as well. The decomposed components can be interpreted as the components of the utility function of the destination choice behavior in the multinomial logit model. If destination amenities and travel costs are consistent with the assumptions of the spatial pattern, we obtain two complete sets of indicators that consider travelers’ sensitivities as well. We should note that the information obtained in the following sections resulted from calculations based on these assumptions.

2.4. Analytical Procedures for Japanese Interprefecture Travel Data. In this study, the proposed approach is applied to Japanese interprefecture data to clarify its characteristics.

At first, we apply the aggregated data in (11) to verify the static characteristics of the interprefectural OD matrix while

confirming the validity of the proposed approach. This is expressed as follows:

$$\mathbf{Q}_{\text{all}} = \sum_{d \in D} \mathbf{Q}_d. \quad (11)$$

where P_j is the population of prefecture j and W_j is the number of employees in the lodging industry in prefecture j . The source for the population is the 2015 census; the source for the number of employees is the 2014 economic census [32]. This number of employees is considered to indicate the

$$a_{j,\text{all}} = \alpha_{0,\text{all}} + \alpha_{1,\text{all}} \ln(P_j) + \alpha_{2,\text{all}} \ln(W_j) + e_{j,\text{all}}, \forall j \in Z, \quad (12)$$

By applying this matrix \mathbf{Q}_{all} to the proposed approach, the average destination amenity $a_{j,\text{all}}$ and travel cost $c_{i,j,\text{all}}$ can be calculated.

In this study, the characteristics of the calculated indicators are presented through several figures and regression analysis. For destination amenities, the following regression analysis is applied as follows:

size of the service industry for visitors, such as the tourism industry, in each prefecture.

The following regression analysis is applied to the estimated travel cost in order to clarify the relationship with transportation LOS.

$$c_{i,j,\text{all}} = \beta_{0,\text{all}} + \beta_{1,\text{all}} S_{i,j} + \beta_{2,\text{all}} H_{i,j} + e'_{i,j,\text{all}}, \forall (i, j) \in ([Z - \{j\}] \times Z), \quad (13)$$

where $S_{i,j}$ and $H_{i,j}$ are explanatory variables for the travel LOS between prefectural offices, respectively. $S_{i,j}$ is the straight-line distance between the prefectural offices, and $H_{i,j}$ is the shortest travel time using multiple transportation modes (air, rail, bus, and ships). For $H_{i,j}$, we applied the values published in the national government statistics [33].

Then, by applying the daily matrix to the proposed approach, destination amenities and travel costs are

calculated for each date. The variation of the two indicators over time is discussed, showing the correlation coefficients of the estimates and the coefficients of the regression analysis (shown as follows). The regression analysis is basically the same simple model as in the static case but applied to all days, respectively.

$$a_{j,d} = \alpha_{0,d} + \alpha_{1,d} \ln(P_j) + \alpha_{2,d} \ln(W_j) + e_{j,d}, \forall (j, d) \in (Z \times D). \quad (14)$$

Here, since the assumption in (10) affects only the constant term ($\alpha_{0,d}$), we can compare the other coefficients ($\alpha_{1,d}, \alpha_{2,d}$) among different dates d . A similar regression

analysis is also performed for the estimated travel cost C_d as follows:

$$c_{i,j,d} = \beta_{0,d} + \beta_{1,d} S_{i,j} + \beta_{2,d} H_{i,j} + e'_{i,j,d}, \forall (i, j, d) \in ([Z - \{j\}] \times Z \times D). \quad (15)$$

3. Decomposition Results of the Static OD Matrix

In this section, we describe the characteristics of the proposed decomposition method, showing the results of applying the proposed approach to OD matrix data aggregated at only one point in time, as shown in (11).

3.1. Difference between the Observed and Estimated Results. We compare the log-ratio matrix of the observed data \mathbf{B}_{all} to $\mathbf{A}_{\text{all}} + \mathbf{C}_{\text{all}}$ to verify the validity of the decomposition. More importantly, \mathbf{B}_{all} is the matrix defined by (16) and (17) as follows:

$$\mathbf{B}_{\text{all}} = \begin{pmatrix} b_{1,1,\text{all}} & \cdots & b_{1,j,\text{all}} & \cdots \\ \vdots & \ddots & & \\ b_{i,1,\text{all}} & & b_{i,j,\text{all}} & \\ \vdots & & & \ddots \end{pmatrix}, \quad (16)$$

$$b_{i,j,\text{all}} = \ln\left(\frac{q_{i,j,\text{all}}}{q_{i,i,\text{all}}}\right), \quad (17)$$

where $b_{i,j,\text{all}}$ has the following relationship from (2) to (3):

$$b_{i,j,\text{all}} \approx a_{i,j,\text{all}} + c_{i,j,\text{all}}. \quad (18)$$

Therefore, a comparison between the right and left sides of (18) presents the ability of two different indicators ($\mathbf{A}_d, \mathbf{C}_d$) for representing a long-distance travel distribution.

Figure 3 confirms the following three features observed in the OD table \mathbf{B}_{all} : (1) the diagonal components are all zero as defined in (17). (2) The closer the component to the diagonal component, the larger the value is (likely to be selected as a travel destination). Moreover, the farther the component from the diagonal component, the smaller the value is (less likely to be selected as a travel destination). The zones in this figure are arranged based on the prefecture codes assigned from northeast to southwest. The adjacent zones are spatially close to each other. Therefore, the aforementioned characteristics indicate that several travelers choose neighboring zones as their travel destinations. (3) Only the component (column) with Tokyo and Osaka as destinations has large values for most residential areas, regardless of location. This feature indicates that Tokyo and Osaka are the top travel destinations of several travelers despite their spatial distance. This may be because Japan's major economic functions are concentrated in these two cities.

A comparison between Figures 3 and 4 confirms that they are approximately in agreement. This finding can also be confirmed from the scatter plot of Figure 5 and the correlation coefficient ($r = 0.997$). These results indicate that the characteristics of the OD matrix \mathbf{B}_{all} between prefectures can be explained by the two components of the estimated travel cost to the migration \mathbf{C}_{all} and destination amenity matrices \mathbf{A}_{all} .

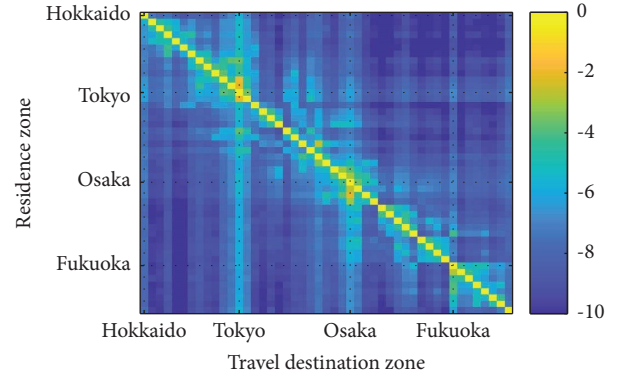


FIGURE 3: Observed matrix \mathbf{B}_{all} .

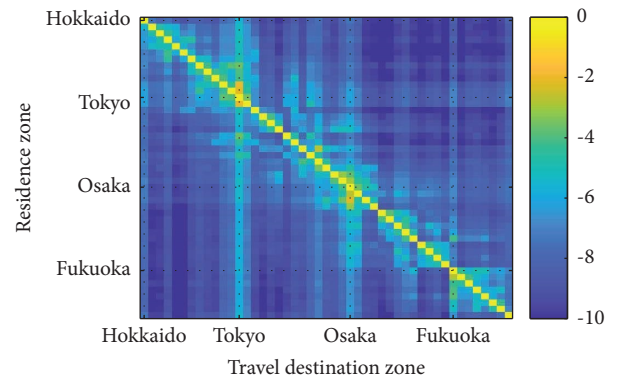


FIGURE 4: Estimated matrix ($\mathbf{A}_{\text{all}} + \mathbf{C}_{\text{all}}$).

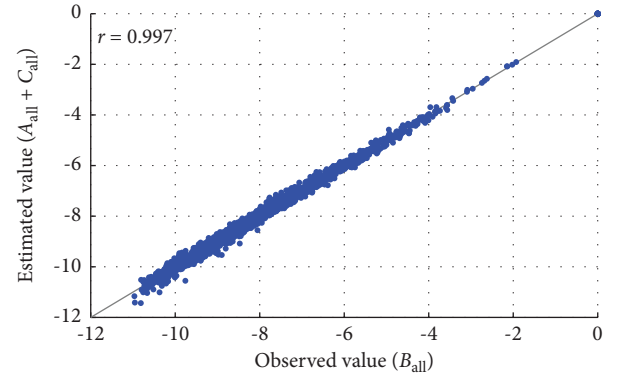


FIGURE 5: A scatter plot of the observed log-ratios (\mathbf{B}_{all}) and estimated ($\mathbf{A}_{\text{all}} + \mathbf{C}_{\text{all}}$).

In our decomposition method, the two types of spatial patterns represented by the two matrices are fully accounted for, as we estimate all elements of matrices with minimal spatial constraints. Under this assumption, the residuals, excluding the observation error, can only be caused by the following two reasons (cases where the model assumptions are not satisfied): (1) travel costs differ depending on the direction. For instance, the perceived LOS differs because of different time values. (2) The destination values vary by the residential zone. For example, Hokkaido's destination value is greater for people in metropolitan areas than for people in

rural areas. Figure 5 indicates that the effects of these differences are significantly small in the context of Japan’s OD matrix size. We disregard them and focus only on other information (destination amenities and travel costs) to clarify their characteristics.

3.2. Estimated Results of the Destination Amenity.

Figure 6 shows the estimation results of the destination amenity \mathbf{A}_{all} . The figure reconfirms the meaning of the three constraint equations for destination amenity, namely, (4), (5), and (10). First, constraint (4) implies that the destination amenity of each prefecture is the same for all the residential areas (excluding the diagonal component). Second, constraints (5) and (10) imply that the diagonal component and destination amenity of Tokyo are zero. The estimation results of Figure 6 indicate that the diagonal component and destination amenity of Tokyo are zero.

Subsequently, the differences in the destination amenities of each prefecture are examined. The destination amenity \mathbf{A}_{all} indicates the attractiveness of a destination toward travelers. For example, the probability of choosing a destination is expected to be proportional to the number of trips related to businesses, weddings, funerals, and other relationship-based trips. Figure 7 confirms this relationship. It shows the relationship between the estimated travel destination amenity and each prefecture population. The figure also confirms that the destination amenity of each zone is strongly correlated with the logarithm of the population. Therefore, the destination amenity is mainly determined by the population size. This result is consistent with a previous study that developed an inter-regional travel demand model for Japan based on a traditional approach Kato et al. [8].

In contrast, there are several exceptions. For example, Tokyo, Hokkaido, and Okinawa are located above the regression line. This indicates that they have high destination amenities. These prefectures have a higher value considering the urban functions and tourist attractions than the population size. This indicates that they can strongly attract tourists. On the contrary, regions such as Kanagawa, Saitama, and Nara prefectures are below the regression line (low destination amenity). Thus, they have a weak ability to attract travelers relative to population size.

Therefore, the ranking of the destination amenity by prefecture in Table 1 is examined. The top and bottom five prefectures in this table show two types of rankings. The ranking on the left is the estimated destination amenity, which is highly correlated with the population size as confirmed in Figure 7. The ranking on the right is the residuals from a simple regression by population as shown in Figure 7. Therefore, the ranking lists the prefectures with high- and low-estimated travel destination amenities per population size. Therefore, Okinawa, Hokkaido, and Nagano prefectures have high destination amenities per population size, in addition to Tokyo. These zones are recognized as resort areas, excluding Tokyo. These results indicate that the destination amenity is a quantitative index that integrates multiple attractions related to long-distance travel.

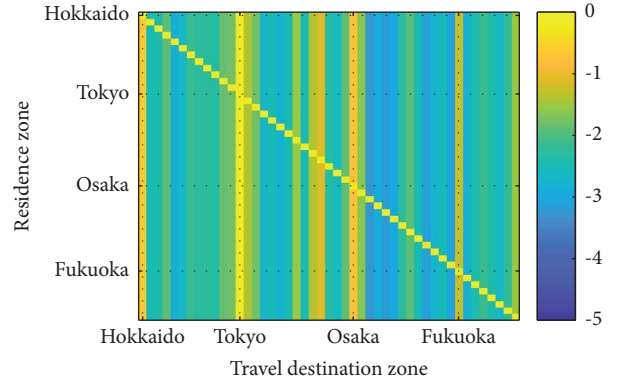


FIGURE 6: Estimated destination amenity matrix (\mathbf{A}_{all}).

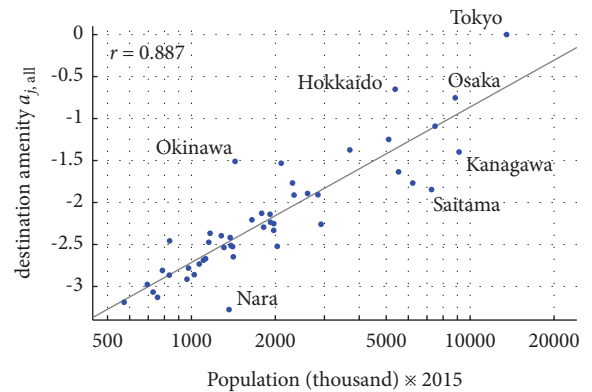


FIGURE 7: Relationship between the destination amenity $a_{j,\text{all}}$ and population.

TABLE 1: Rank of destination amenity.

Ranks	Destination amenities	Population size-corrected destination amenity (difference from the regression line of Figure 7)
1	Tokyo	+0.00
2	Hokkaido	-0.65
3	Osaka	-0.75
4	Aichi	-1.09
5	Fukuoka	-1.25
43	Shimane	-2.98
44	Kochi	-3.07
45	Tokushima	-3.13
46	Tottori	-3.19
47	Nara	-3.28

In contrast, Saitama and Nara prefectures have the smallest population-corrected destination amenity. These zones are adjacent to large cities similarly, such as Tokyo and Osaka. Therefore, several people in these prefectures commute to Tokyo and Osaka for work. When meeting those

from other prefectures, they may often prefer to meet at their workplace (Tokyo or Osaka) because of better accessibility and recreational attraction. Consequently, the destination amenity (ability to attract travelers) is smaller than the population size. Moreover, the destination amenities of Tokyo and Osaka are largely based on this mechanism.

In addition, Table 2 shows the results of the regression analysis defined in (12). Table 2 shows that these two variables are each statistically positively significant, as expected. The value of R^2 indicates that these two variables can explain most of the destination amenity variation. The results show that the results obtained by aggregating all periods are almost identical to the conclusions obtained with the traditional approaches. On the other hand, the value of our approach in terms of destination amenities will become clear in the next section, where a similar model is applied to several time points.

3.3. Estimated Results of Travel Cost. We examine the estimation results of the travel cost matrix C_{all} from Figure 8. The diagonal component is zero based on the condition in (7). In addition, the matrix is fully symmetric based on the condition in (6). The farther the distance from the diagonal component, the larger the absolute value of the travel cost tends to be. Each zone in Figure 8 is arranged in the same order as that in Figure 3 (i.e., from northeast to southwest). Therefore, the farther a pair of prefectures from the diagonal component, the farther the actual Euclidean distance is. The travel costs shown in Figure 8 are also close to the actual Euclidean distances.

An examination of the zone pairs of Tokyo departures and arrivals shows that the absolute values are relatively small, including the distant zones. This observation indicates that convenient transportation is available to most areas in Tokyo. Therefore, the air transportation network at Haneda Airport has a strong influence.

Therefore, a close examination of the diagonal components shows certain blocks with small absolute values of travel cost. The two yellow-colored squares in the lower right corner of Figure 8 represent the four and seven prefectures in Shikoku and Kyushu islands, respectively. This figure indicates that there are a lot of mutual travel activities inside a similar island. Other blocks from the southwest (from the lower right of Figure 8) can be identified as regional blocks.

The exponential value of this estimated travel cost C_{all} is used as the distance. Moreover, each prefecture is plotted on the two-dimensional coordinate to describe this distance matrix in detail. Here, the positions of Tokyo and Osaka are fixed, and the coordinates of each prefecture are calculated to match the ratio of the distances between the prefectures on the two-dimensional plane and that of the estimated travel cost $\exp(-C_{all})$. Figure 9 shows the location plot of each prefectural office using the original coordinates. Considering this figure, only the major rail links (high-speed rail (HSR) in purple) are shown as links connecting the nodes as shown in Figure 10.

Figure 10 shows the following three characteristics of the estimated travel cost: (1) Japan's transportation network is

centered on Tokyo, and the shape of the country derived from the estimated travel cost is a circle centered on Tokyo. This structure is formed because the travel cost is defined by the LOS of air routes over a certain distance. Moreover, direct flights are available from Haneda Airport to approximately all the prefectures. (2) Regarding the circular structure, the city axis on the Pacific Ocean and Sea of Japan sides are located on the inner and outer sides, respectively. The cities on the Sea of Japan side, such as Akita, Ishikawa, and Tottori, are located on the outer side of the circle. The distances between them are farther from those on the Pacific Ocean side. They are also distant from other cities in Japan. This pattern approximately reflects the shape of the current HSR network. (3) The cities along the Tokaido HSR (Tokyo-Osaka) and Tohoku HSR (Tokyo-Miyagi-Aomori) lines do not move away from Tokyo in a linear direction. They rather move in a distorted fashion. Here, the zones where express trains (Nozomi, Hayabusa) stop and have high-frequency service tend to be relatively close to the center although they are far from Tokyo. The travel cost derived in this study is an index that includes detailed LOS, such as those that have been mentioned.

Then, Table 3 shows the results of the regression analysis defined in (13). Table 3 shows that these two variables are each statistically negatively significant, as expected. On the other hand, the value of R^2 indicates that 25% of variances remain as a residual after being explained by these two variables. The residuals are expected to reflect other information such as service frequency, fares, and differences between transportation modes. However, it is not easy to generate data to create a model that can discuss these detailed issues. The proposed approach allows us to understand the state of transportation service levels in the country as a whole, shown in Figure 10, without detailed data preparation.

4. Time-Series Change in the Two Decomposed Indicators

The time variability of intercity travel in Japan can be described by time-series changes in the two indicators proposed and defined in this study. Therefore, this section presents a calculation of the two indices for each day and checks the time-series changes for 1,461 days.

First, we compare the estimated indices with those calculated from the all-time aggregate and OD matrix in Section 4. The correlation coefficients are calculated for each date. Here, the correlation coefficient is not affected by the assumption in (10) since the constant term is irrelevant. Table 4 shows the mean, one and five percentiles, and minimum values of the correlation coefficients for 1,461 days.

The correlation coefficient value of the destination amenity indicates that the average value is large (at 0.961), whereas the correlation coefficient is smaller than 0.877 at approximately 5% (approximately 70 days). Furthermore, the correlation coefficient is smaller than 0.584 at approximately 1% (approximately 18 days). This finding indicates that the magnitude of travel destination amenities among

TABLE 2: The regression analysis result of the destination amenity.

		Estimate	<i>t</i> -Stat
$\alpha_{0,\text{all}}$	Intercept	-13.508**	-22.593
$\alpha_{1,\text{all}}$	ln (Population)	0.312**	4.098
$\alpha_{2,\text{all}}$	ln (Number of employees in lodging industry)	0.725**	7.695
	Number of observations (prefecture)	47	
	R^2	0.909	

Level of significance: **1% and *5%.

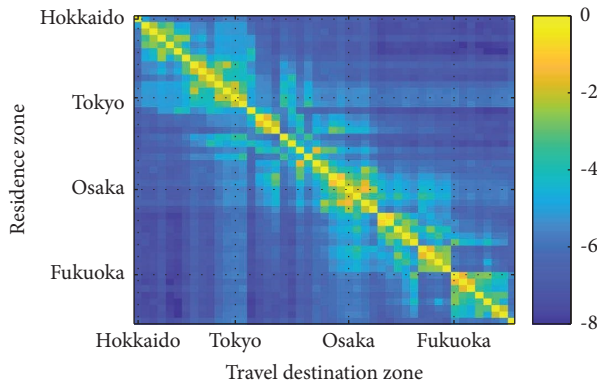
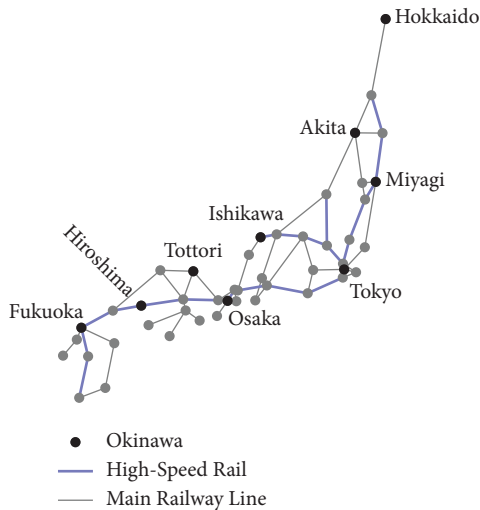
FIGURE 8: Estimated travel cost (C_{all}) matrix.

FIGURE 9: Spatial location of each prefecture (using latitude and longitude angles).

the prefectures is significantly different from the average level at a certain time point.

Figure 11 shows the time trend of this correlation coefficient value. This figure confirms that the correlation coefficient has been fluctuating approximately around the same value every year. Therefore, the seasonal variation is larger than the interannual variation in the travel destination amenity. The correlation coefficient values are small in three periods. These are grayed out in Figure 11. Considering the left to right, these periods are Golden Week (GW and consecutive national holidays), Obon (religious holidays in

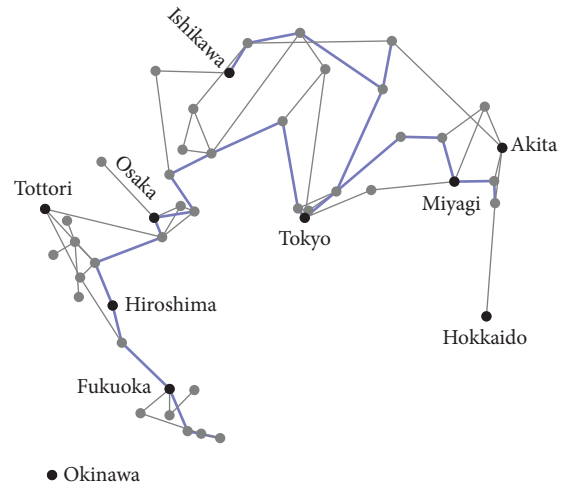


FIGURE 10: Spatial location of each prefecture by the estimated travel cost.

TABLE 3: The regression analysis result of travel cost.

		Estimate	<i>t</i> -Stat
$\beta_{0,\text{all}}$	Intercept	-2.560**	-35.893
$\beta_{1,\text{all}}$	Distance (100 km)	-0.638**	-32.943
$\beta_{2,\text{all}}$	Minimum travel time (hour)	-0.0940**	-13.204
	Number of observations (prefecture pairs)	1066	
	R^2	0.741	

Level of significance: **1% and *5%.

TABLE 4: Distribution of correlation coefficients between the destination value at one day and one of the all-time points summed.

	Destination amenities	Travel cost
Average value	0.961	0.983
5 percentiles	0.877	0.969
1 percentile	0.584	0.942
Minimum value	0.474	0.930

Japan), and New Year holidays. Among these periods, Obon and New Year holidays have smaller correlations than the other periods. Thus, the distribution of the destination amenity (attractiveness of travelers) in each prefecture significantly differs from the average.

Subsequently, the correlation coefficients of travel costs in Table 4 are examined. The correlation coefficients between

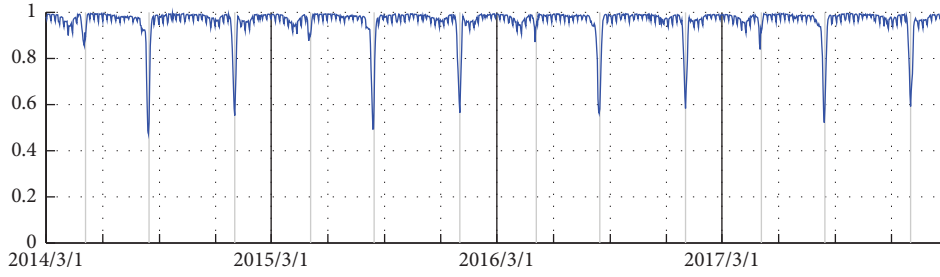


FIGURE 11: Time-series changes in correlation coefficients between the destination amenity at one day and one of the all-time points summed.

the travel costs at each date and the average values at the all-time points are 0.930, including those at the minimum points. This value is significantly large at the all-time points. Therefore, the relative distance between the prefectures indicated in Figure 10 is approximately the same at the all-time points.

Next, we applied the regression analysis (defined in (14)) to obtain more detailed time-series features of destination amenities. The value of R^2 for each day in Figure 12(c) indicates that the destination amenity for the three peak periods (GW, Obon, and New Year holidays) is hardly explained by the two variables. Thus, the results indicate the requirement of a different model for these periods. Furthermore, the estimated coefficients in Figures 12(a) and 12(b) show that the parameters of the model differ significantly depending on the time period. For example, the population coefficients are statistically insignificant in the three peak periods and relatively large in April and late January. The coefficient of the number of employees also shows large changes depending on the season as well as the day of the week. Thus, because of the large seasonal variation in destination amenities, the model with only a few groupings of periods in the traditional approach may be applicable only at certain limited points in time. The estimated $a_{j,d}$ provided us with these important caveats on the characteristics of destination amenities and the trend of time variation in Japan.

A similar regression analysis (defined in (15)) is also performed for the estimated travel cost C_d . The value of R^2 for each day in Figure 13(c) indicates that variances of approximately 25% remain in the residuals in all dates. The estimated coefficients in Figures 13(a) and 13(b) show some seasonal differences, although the correlation is high as shown in Table 4. For example, the coefficient of travel time is small in the New year and Obon periods. This indicates a lower resistance to travel to places that take more time in these periods. Thus, the trend is different during GW (early May) when the coefficient of distance is smaller, not that of time.

5. Discussion

Calculating the two indicators in the proposed approach has the following benefits.

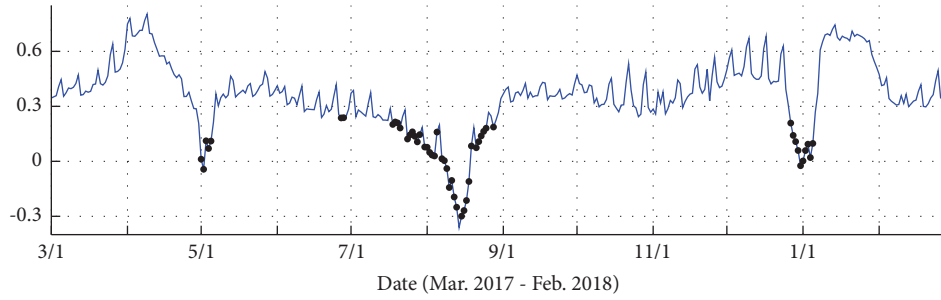
First, it evidences the temporal stability of the travel destination choice model in Japan. The spatial difference of the travel cost is constant over time (as shown in Table 4),

which is a favorable characteristic. This indicates that the coefficients estimated from the data at one point can be applied at other points. Moreover, seasonal differences in destination amenities were found. This finding raises an important concern for several destination choice models. This is because it indicates the possibility of the coefficients of the destination variables changing significantly from season to season. Thus, a model estimating from data that are biased toward a particular period cannot be applied to a broader point in time. Obtaining such data through questionnaires is difficult because of people's memory capacity and survey cost. To address this issue, we passively collected large sample data, such as mobile phone location data. These data are valuable because they easily provide a large sample over a long period.

Secondly, such a quantitative and simple description of the spatio-temporal characteristics of long-distance travel behavior is useful for transportation policy. The calculated travel costs demonstrate the structure of the country as perceived by the people. Figure 10 clearly shows the current status of long-distance transportation networks. This will be helpful in future discussions. In addition, the destination amenity identified here quantifies the ability of each prefecture to attract travelers. This may be a useful outcome measure for strategies to attract travelers because it can attract people from all over Japan. It can attract those who are less susceptible to the influence of few limited origins.

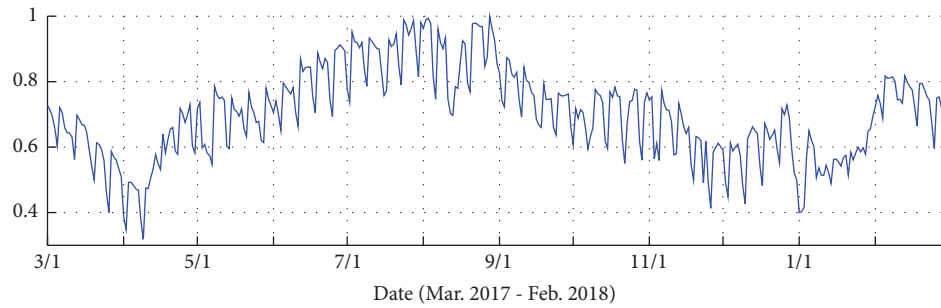
Furthermore, the decomposition approach proposed in this study is not limited to long-distance travel. It can be applied to any OD matrix data for which most of the elements are larger than zero. However, the analysis of OD matrix data for daily traffic, where many of the influencing factors have already been identified and where the phenomenon is stable over time, may not yield many new findings. Thus, we believe that the proposed approach is well suited for understanding highly variable patterns such as long-distance travel.

The approach and results described in this study have some limitations. To confirm the validity of the pattern obtained in this study, it is desirable to analyze the effect of assuming a time period set at $t = 13$. The critical limitation of this approach is the difficulty of applying the policy simulation. This is because the model does not contain variables that can be controlled directly. To discuss the effects of future policies, we determine that a model with traffic LOS and destination variables is more appropriate, which is similar to

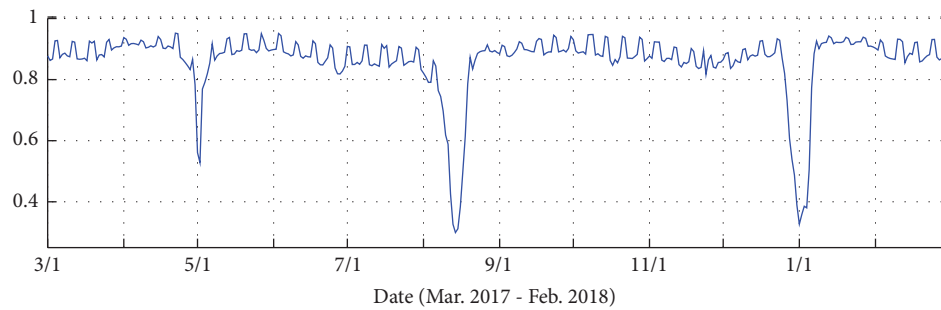


• : p-value > 0.05

(a)

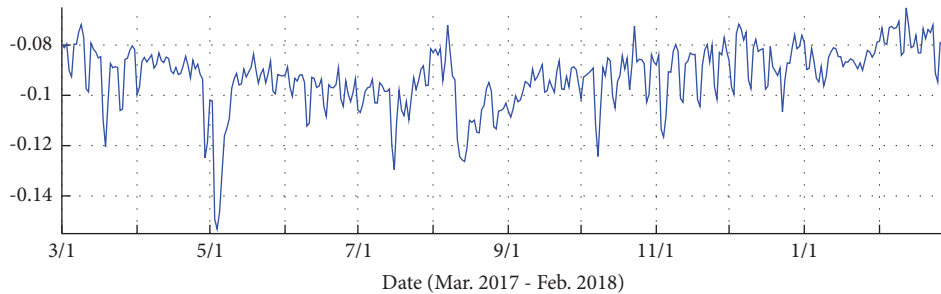


(b)



(c)

FIGURE 12: Regression analysis results for 365 days of destination amenities. (a) Estimated coefficients of $\ln(\text{Population})$. (b) Estimated coefficients of $\ln(\text{number of employees in the lodging industry})$. (c) R -squared value.



(a)

FIGURE 13: Continued.

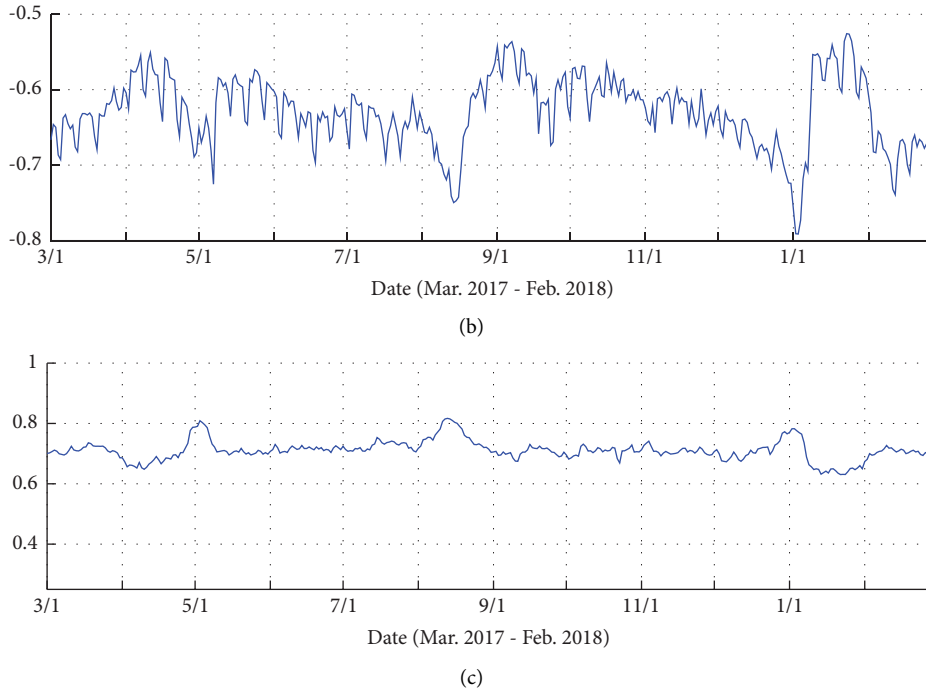


FIGURE 13: Regression analysis results for 365 days of travel cost values. (a) Estimated coefficients of distance (100 km). (b) Estimated coefficients of minimum travel time (hour). (c) R -squared value.

the case with traditional models. Applying the proposed approach has an advantage for calculating the outcome measures and confirming the prediction model's time stability as discussed. In addition, the results obtained in this study may be limited to long-distance travel in Japan. Particularly, more analyses are required to answer the question, "Do the warnings discussed here apply to other countries or other types of trips?" This question can be answered by applying the approach proposed in this study to other multipoint OD matrices.

6. Conclusion

This study proposes a method for decomposing the long-term and daily interprefectural OD matrix into two types of spatial patterns. The indices calculated in the decomposition of this study are referred to as the destination amenity and travel costs. These are identified by the differences in the influence space patterns on the OD matrix, and these patterns correspond to the variables treated in the traditional destination choice model. Therefore, the two indices (destination amenities and travel costs) are composites of several variables treated in any destination choice model.

Using mobile phone location data, we can derive detailed time-series changes in these indices because they can be calculated directly from the OD matrices. This study analyzes the changes in the two indices of interprefecture travel in Japan for over 1,461 days.

By calculating the two indicators in the proposed approach, the characteristics of interprefecture travel in Japan are briefly described as follows: (1) the destination amenity is strongly correlated with the population size. This value indicates zone

attractiveness toward travelers. (2) Hokkaido, Okinawa, and Tokyo have higher destination amenities than the population size. Hence, they are attractive travel destinations, in addition to their population size. (3) These destination amenities change significantly with the seasons. Therefore, seasonal variations should be taken into account when modeling destination amenities. (4) The calculation results of the travel costs show that Japan's national land structure is circular, with Tokyo at the center. Prefectures facing the Pacific Ocean are located inside the circle, whereas those facing the Sea of Japan are located at the outer edge of the circle. Long-distance transportation services are relatively unimproved in prefectures facing the Sea of Japan. (5) Additionally, it was found that this travel cost structure shows little variation over time, unlike destination amenities.

Data Availability

The MSS data used to support the findings of this study were supplied by DOCOMO InsightMarketing, INC under license and so cannot be made freely available. Requests for access to these data should be made to DOCOMO InsightMarketing, Inc. The same data are not freely available, but only the most recent data can be seen in <https://mobakumap.jp/>.

Ethical Approval

Not applicable.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Hiromichi Yamaguchi conducted conceptualization, data curation, formal analysis, methodology, validation, visualization, writing of the original draft, writing, reviewing, and editing of the study. Mashu Shibata conducted formal analysis, methodology, visualization, writing, reviewing, and editing of the study. Shoichiro Nakayama conceptualized, wrote, reviewed, and edited the study.

Acknowledgments

This study was supported by JSPS (Japan Society for the Promotion of Science) KAKENHI Grant-in-Aid for Scientific Research (B) 20H02267, 20H02270, and 21H01455 and by the Ministry of Education, Culture, Sports, and Science and Technology, Japan (MEXT) LEADER project.

References

- [1] S. Erlander and N. F. Stewart, *The Gravity Model in Transportation Analysis: Theory and Extensions*, Vsp, vol. 3, Utrecht, The Netherlands, 1990.
- [2] G. K. Zipf, "The P1 P2/D hypothesis :on the intercity movement of persons," *American Sociological Review*, vol. 11, no. 6, pp. 677–686, 1946.
- [3] M. Lenormand, A. Bassolas, and J. J. Ramasco, "Systematic comparison of trip distribution laws and models," *Journal of Transport Geography*, vol. 51, pp. 158–169, 2016.
- [4] Y. Zhang, F. Lin, and A. Zhang, "Gravity models in air transport research: a survey and an application," in *Handbook of International Trade and Transportation*, pp. 141–158, Edward Elgar Publishing, Cheltenham, UK, 2018.
- [5] R. Cordera, R. Sañudo, L. Dell'Olio, and Á. Ibeas, "Trip distribution model for regional railway services considering spatial effects between stations," *Transport Policy*, vol. 67, pp. 77–84, 2018.
- [6] D. Chai, D. Zhang, Y. Sun, and S. Yang, "Research on the city network structure in the yellow river basin in China based on two-way time distance gravity model and social network analysis method," *Complexity*, vol. 2020, Article ID 6680954, 19 pages, 2020.
- [7] T. Grosche, R. Klophaus, and A. Seredyński, "Market concentration in German air transport before and after the Air Berlin bankruptcy," *Transport Policy*, vol. 94, pp. 78–88, 2020.
- [8] H. Kato, Y. Kaneko, K. Kato, T. Shimizu, and K. Tanabe, "Inter-regional travel demand analysis using integrated model for practical travel demand forecast," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 9, pp. 239–254, 2011.
- [9] E. Yao and T. Morikawa, "A study of on integrated intercity travel demand model," *Transportation Research Part A: Policy and Practice*, vol. 39, no. 4, pp. 367–381, 2005.
- [10] X. Fu, T. H. Oum, and J. Yan, "An analysis of travel demand in Japan's intercity market: empirical estimation and policy simulation," *Journal of Transport Economics and Policy*, vol. 48, pp. 97–113, 2014.
- [11] R. Ahas, A. Aasa, Ü. Mark, T. Pae, and A. Kull, "Seasonal tourism spaces in Estonia: case study with mobile positioning data," *Tourism Management*, vol. 28, no. 3, pp. 898–910, 2007.
- [12] R. Ahas, A. Aasa, A. Roose, Ü. Mark, and S. Silm, "Evaluating passive mobile positioning data for tourism surveys: an Estonian case study," *Tourism Management*, vol. 29, no. 3, pp. 469–486, 2008.
- [13] M. Janzen, M. Vanhoof, K. W. Axhausen, and Z. Smoreda, "Estimating long-distance travel demand with Mobile phone billing data," in *Proceedings of the 16th Swiss Transport Research Conference (STRC 2016)*, Ascona, Switzerland, May 2016.
- [14] H. Yamaguchi and S. Nakayama, "Detection of base travel groups with different sensitivities to new high-speed rail services: non-negative tensor decomposition approach," *Transport Policy*, vol. 97, pp. 37–46, 2020.
- [15] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin-destination trips by purpose and time of day inferred from mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 240–250, 2015.
- [16] X. Li, M. Li, Y. J. Gong, X. L. Zhang, and J. Yin, "T-DesP: destination prediction based on big trajectory data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2344–2354, 2016.
- [17] T. Bantis and J. Haworth, "Who you are is how you travel: a framework for transportation mode detection using individual and environmental characteristics," *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 286–309, 2017.
- [18] C. S. Pitombo, E. Kawamoto, and A. J. Sousa, "An exploratory analysis of relationships between socioeconomic, land use, activity participation variables and travel patterns," *Transport Policy*, vol. 18, no. 2, pp. 347–357, 2011.
- [19] C. S. Pitombo, A. D. de Souza, and A. Lindner, "Comparing decision tree algorithms to estimate intercity trip distribution," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 16–32, 2017.
- [20] L. Sun and K. W. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transportation Research Part B: Methodological*, vol. 91, pp. 511–524, 2016.
- [21] D. Wang, Z. Cai, Y. Cui, and X. M. Chen, "Nonnegative tensor decomposition for urban mobility analysis and applications with mobile phone data," *Transportmetrica: Transportation Science*, vol. 18, no. 1, pp. 29–53, 2019.
- [22] A. Huang and D. Levinson, "Axis of travel: modeling non-work destination choice with GPS data," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 208–223, 2015.
- [23] J. Zhu and X. Ye, "Development of destination choice model with pairwise district-level constants using taxi GPS data," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 410–424, 2018.
- [24] I. Kristoffersson, A. Daly, and S. Algers, "Modelling the attraction of travel to shopping destinations in large-scale modelling," *Transport Policy*, vol. 68, pp. 52–62, 2018.
- [25] M. Terada, T. Nagata, and M. Kobayashi, "Population estimation Technology for mobile spatial statistics," *Docomo Technical Journal*, vol. 14, pp. 10–15, 2013.
- [26] T. Kubo, S. Uryu, H. Yamano, T. Tsuge, T. Yamakita, and Y. Shirayama, "Mobile phone network data reveal nationwide economic value of coastal tourism under climate change," *Tourism Management*, vol. 77, Article ID 104010, 2020.
- [27] Y. Hara and H. Yamaguchi, "Japanese travel behavior trends and change under COVID-19 state-of-emergency declaration: nationwide observation by mobile phone location data," *Transportation Research Interdisciplinary Perspectives*, vol. 9, Article ID 100288, 2021.

- [28] D. Bachir, G. Khodabandelou, V. Gauthier, M. El Yacoubi, and J. Puchinger, “Inferring dynamic origin-destination flows by transport mode using mobile phone data,” *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 254–275, 2019.
- [29] K. W. Axhausen, J. L. Madre, J. W. Polak, and P. L. Toint, *Capturing Long-Distance Travel*, Research Studies Press, Boston, Massachusetts, 2003.
- [30] K. E. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK, 2009.
- [31] E. Marrocu and R. Paci, “Different tourists to different destinations. Evidence from spatial interaction models,” *Tourism Management*, vol. 39, pp. 71–83, 2013.
- [32] Statistics Bureau of Japan, “Portal site of official statistics of Japan,” 2020, <https://www.e-stat.go.jp/en>.
- [33] MLITT, “Inter-regional travel survey in Japan,” 2015, https://www.mlitt.go.jp/sogoseisaku/soukou/sogoseisaku_soukou_fr_000016.html.