Shahriar AFANDIZADEH ZARGARI, Ph.D.[1]
(Corresponding author)
E-mail: zargari@iust.ac.ir
Amirmasoud MEMARNEJAD, Ph.D. Candidate[1]
E-mail: memarnejad@civileng.iust.ac.ir
Hamid MIRZAHOSSEIN, Ph.D.[2]
E-mail: mirzahossein@eng.ikiu.ac.ir
[1] Department of Transportation Engineering and Planning, School of Civil Engineering
Iran University of Science and Technology
16846-13114, Tehran, Iran
[2] Department of Civil Engineering – Transportation Planning, Imam Khomeini International University
34148-96818, Qazvin, Iran

Traffic Engineering
Original Scientific Paper
Submitted: 5 July 2021
Accepted: 16 Nov. 2021

NOVEL HYBRID METHOD FOR TRAVEL PATTERN RECOGNITION BASED ON COMPARISON OF ORIGIN-DESTINATION MATRICES IN TERMS OF STRUCTURAL SIMILARITY

ABSTRACT

Origin-destination (OD) matrices provide transportation experts with comprehensive information on the number and distribution of trips. For comparing two OD matrices, it is vital to consider not only the numerical but also the structural differences, including trip distribution priorities and travel patterns in the study region. The mean structural similarity (MSSIM) index, geographical window-based structural similarity index (GSSI), and socioeconomic, land-use, and population structural similarity index (SLPSSI) have been developed for the structural comparison of OD matrices. These measures have undeniable drawbacks that fail to correctly detect differences in travel patterns, therefore, a novel measure is developed in this paper in which geographical, socioeconomic, land-use, and population characteristics are simultaneously considered in a structural similarity index named GSLPSSI for comparison of OD matrices. The proposed measure was evaluated using OD matrices of smartphone Global Positioning System (GPS) data in Tehran metropolitan. Also, the robustness of the proposed measure was verified using sensitivity analysis. GSLPSSI was found to have up to 21%, 15%, and 9% higher accuracy than MSSIM, GSSI, and SLPSSI, respectively, regarding structural similarity calculation. Furthermore, the proposed measure showed 7% higher accuracy than SLPSSI in the structural similarity index of two sparse OD matrices.

KEYWORDS

structural similarity; travel patterns; OD matrix; traffic zones; Tehran metropolitan.

1. INTRODUCTION

An origin-destination (OD) matrix represents the travel demand distribution from different origins to different destinations. It provides transportation experts with comprehensive information on the number and distribution of trips. OD matrices are employed as primary inputs in many traffic studies and simulations. The values of OD matrix cells represent the number of trips (traffic flow) between OD pairs. Furthermore, a set of such cell values represents the travel demand distribution between traffic zones. Although previous studies mainly focused on developing OD matrix estimation methods [1–7], some studies investigated the similarities of OD matrices [8–11]. It is crucial to consider both the cell values and the trip distribution of traffic zones to compare OD matrices as it represents the matrix structure. The structural comparison of OD matrices have various applications, including estimation of constraint parameters in demand matrix [12], evaluation of the estimated OD matrix using ground-truth matrix [9], and travel pattern detection in different periods (on different days) [13].

Previous studies proposed different interpretations of an OD matrix structure. The structure of an OD matrix can be obtained by normalising the cells through dividing the amount of each cell to the total generated trips in each OD matrix row or by normalising the cells of each column through dividing its amount to the total attracted trips in each column [14]. Overall, the structure of an OD matrix is its configuration framework and represents the prioritisation and order of trips from an origin to different destinations [15]. In this regard, different travel destinations priority could be observed on different days. In the similarity comparison of OD matrices, the structural similarity is measured when the structures of the OD matrices are evaluated. Two matrices are completely structurally similar when they have similar structures and the same cell values simultaneously; this is the case when the two OD matrices are the same.

Many traditional statistical measures such as root mean square error (RMSE) [16–18], normalised root mean square error (NRMSE) [19, 20], and several other methods referred to in [21] have been employed to compare OD matrices in literature. These methods perform cell-to-cell comparisons of OD matrices using only mathematical equations and cannot compare a set of cells to detect structural similarities or dissimilarities between two OD matrices. Therefore, such methods cannot analyse and detect structural dissimilarities between two matrices due to different trip priorities from different origins [21]. Contrary to traditional methods, a small number of studies explored the structural similarity of OD matrices, such as:

1) The mean structural similarity index measure (MSSIM) [9]
2) Complementary methods on MSSIM, such as 4D-MSSIM [22], geographical structural similarity index (GSSI) [15], socio-economy, land-use, and population structural similarity index (SLPSSI) [21]
3) Optimisation-based methods, such as the Wasserstein distance [23], and normalised Levenshtein distance for OD matrices (NLOD) [12]

Although they can recognise the structural similarities/dissimilarities of two OD matrices, these methods have some drawbacks. For example, the results of MSSIM are significantly dependent on the local window size [12]. GSSI is based on MSSIM and considers geographical characteristics only in the selection of local windows. Although GSSI tackles the problems related to the local window size selection, it fails to utilise other important characteristics of zones in windowing [21]. Furthermore, SLPSSI is an augmented variant of MSSIM and GSSI. It considers the socioeconomic, land-use, and population characteristics of traffic zones and clusters these zones in a window. However, it does not use the geographical characteristics of the zones and cannot rationally describe the travel patterns of different zones. Moreover, the Wasserstein distance and NLOD methods are more time-consuming processes than structural similarity detection since they are optimisation-based approaches.

This paper focuses on the MSSIM and is developed to tackle the drawbacks mentioned above. In this study, the calculation structure of MSSIM is redesigned to select local windows based on geographical, socioeconomic, land-use, and traffic zone population characteristics at the same time. Therefore, in this study, geographical, socioeconomic, land-use, and population structural similarity index (GSLPSSI) is introduced.

2. LITERATURE REVIEW

Although various performance evaluation indices have been employed in transportation, only a few studies have been conducted on developing an index for the structural comparison of OD matrices. These OD matrix structural comparison methods are discussed below.

2.1 MSSIM method

MSSIM was first employed to compare two images, where one of them is the noised version of the other one [24]. It was demonstrated that two different images with the same mean square error (MSE) had different MSSIM values. Djukic et al. [9] used MSSIM to structurally compare OD matrices by considering each cell as a pixel of an image. They defined an m×m local window, which was necessarily smaller than the original matrix, and compared a set of extracted OD pairs from the local window. *Figure 1* shows the local windows for the comparison of matrices S_1 and S_2. *Figures 1a and 1b* depict the first local window for matrix comparison, while *Figures 1c and 1d* illustrate the next window.

Despite its ability to detect structural dissimilarities/similarities of OD matrices, MSSIM needs to be augmented since MSSIM values are sensitive to local window size, and no consensus has been reported on the correct selection of local window size. The sensitivity of MSSIM in structural similarity detection decreases as the local windows increase in size [15]. To eliminate this ambiguity, in some studies the entire matrix has been assumed as a window [25], however, by this assumption MSSIM will be insensitive to the structural dissimilarities

OD MATRIX - S_1

	D_1	D_2	D_3
O_1	70	40	20
O_2	30	90	80
O_3	10	60	110

a)

OD MATRIX - S_2

	D_1	D_2	D_3
O_1	120	50	70
O_2	110	80	60
O_3	30	50	160

b)

	D_1	D_2	D_3
O_1	70	40	20
O_2	30	90	80
O_3	10	60	110

c)

	D_1	D_2	D_3
O_1	120	50	70
O_2	110	80	60
O_3	30	50	160

d)

Figure 1 – Local windows in MSSIM

between the two matrices. Besides, the computational cost is higher when the local widows are smaller. For example, for two M×N matrices and m×n local windows, a total of (M-m+1)×(N-n+1) window pairs should be compared. In addition, the OD pairs that fall in the same local window do not necessarily correlate. Hence, structural similarity values would not have any interpretation for such OD pairs.

2.2 Complementary methods of MSSIM

Because of the sensitivity of MSSIM results to the local window size and the geographical adjacency of traffic zones, GSSI [15] and 4D-MSSIM [22] were developed. In addition, to enhance the accuracy of travel patterns analysis on different days, SLPSSI [21] was designed based on socioeconomic, land-use, and population characteristics of traffic zones.

4D-MSSIM uses the Euclidian distance to determine adjacent zones which are classified in the same group and forma local window. Since traffic zones are separated by natural features, e.g. rivers, or artificial ones, e.g. highways, the travel time between these zones is higher than two adjacent zones, therefore, Euclidian distance cannot accurately detect the vicinity of traffic zones.

GSSI arranges origins and destinations based on geographical positions and attributions. Then, local windows are determined for use in the MSSIM based on a higher-level geographical borders which is the result of dividing the whole area to a few main areas. In this method, the sizes of local windows vary, depending on the number of zones inside each geographical border. Although it provides a new windowing technique in the MSSIM, GSSI utilises only geographical distance (Euclidean distance between zones and the proximity of zones which puts them in the same group, like northern zones or southern zones) and predefined borders of geographical zones to determine the local window size as usual. Geographical borders do not necessarily determine zones with similar characteristics to trip production and attraction (trip generation).

SLPSSI was designed to cope with the drawbacks of the MSSIM and enhance the effectiveness and efficiency of the GSSI. It clusters traffic zones based on socioeconomic, land-use, and population characteristics. Zones are classified into several groups based on their trip production/attraction potential. These groups form the local windows of SLPSSI. Although it indicates new aspects of urban travel patterns on different days (either working or non-working days), SLPSSI does not enable geographical characteristic-based analyses since it does not consider geographical adjacency in selecting traffic zones in a local area window.

2.3 Optimisation-based methods

The Wasserstein distance is defined as the minimum total travel time to assign trips between OD pairs in matrix M using an assignment consistent with matrix N. This method is a linear programming problem that considers the distance based on travel time. It is based on optimisation techniques and has enormously higher computational cost than MSSIM-based techniques. Therefore, the Wasserstein distance cannot be employed for large-scale OD matrices such as large urban networks. Further details are provided by Ruiz de Villa et al. [23]. Likewise, NLOD compares OD flows of two matrices by using optimisation techniques. It defines the OD matrix structure as the priority of trip from an origin to a destination. This method is inspired by the traditional Levenshtein distance widely used in linguistics, such as comparing two strings of words.

It arranges each row of the OD matrix based on the number of trips. Then, the minimum variations are calculated for matrix rows based on the cell values and trip distribution in each row of the original matrix. Further details are provided by Behara et al. [12].

2.4 Summary of the literature review

Although MSSIM [9], GSSI [15], 4D-MSSIM [22], and SLPSSI [21] methods can detect the structural similarities and dissimilarities of two OD matrices, they still need to be further improved. For example, GSSI does not consider socioeconomic, land-use, and population characteristics in local window selection; 4D-MSSIM has errors in selecting adjacent zones; and SLPSSI does not consider geographical adjacency. Moreover, sophisticated methods, such as the Wasserstein distance and NLOD, have high computational costs since they employ optimisation techniques. Therefore, this paper seeks to provide a newly developed variant of MSSIM to detect the structural similarities and dissimilarities of OD matrices and provide new analyses of travel patterns more accurately.

3. METHODOLOGY

Following the investigated gaps in literature, this paper outlines a method procedure as shown in *Figure 2* to present the novelty in analysing traffic patterns based on the structural similarities and dissimilarities of OD matrices. The developed method divides the proposed traffic zones into main geographical areas, e.g. northern, southern, eastern, western, and central. Zones in each geographical area are correlated in terms of geographical adjacency (*Figure 2a*). Then, the zones in the same geographical areas are classified based on socioeconomic, land-use, and population characteristics. Car ownership per capita is the socioeconomic characteristic; resident population and employee population are the population characteristics; and the areas of commercial and administrative centres are the land-use characteristics of each traffic zone. These factors affect the potential of traffic zones on producing and attracting trips (trip generation). The car ownership per capita, resident population, and employee population are trip production factors in traffic zones, while commercial and administrative centre areas are trip attraction factors. All of these predefined factors are trip generation factors (*Figure 2b*). The normalised value of each feature is considered to be a score that presented the potential role of each feature in trip generation. For instance, for each zone and factor (e.g. zone number 110 and resident population factor with 5482 resident population size) a number between 0 and 1 that shows the normalised value of that factor is calculated, which is considered one of the pre-defined factors (*Figure 2c*). In order to detect the zones with the highest similarity in trip production and attraction (trip generation) potentials, the k-means clustering method was employed. Various methods have been used to cluster traffic zones and data, among which the k-means method has proved to have the highest performance [26, 27] (*Figure 2d*).

Consequently, simultaneous local windows are determined based on socioeconomic, land-use, population, and geographical characteristics. The zones within the same window have the highest similarity based on the aforementioned characteristics (*Figure 2e*). *Figure 3* illustrates the local windowing of the proposed method.

The MSSIM formulation consists of three equations to evaluate the similarities and dissimilarities of local windows by the mean, standard deviation, and covariance values. *Equation 1* is the luminance equation denoted as $l(x,y)$ and compares the mean entry of the selected window (μ_x,μ_y). *Equation 2* is the contrast and denoted as $c(x,y)$. It calculates the standard deviation (σ_x,σ_y). *Equation 3* is the structure

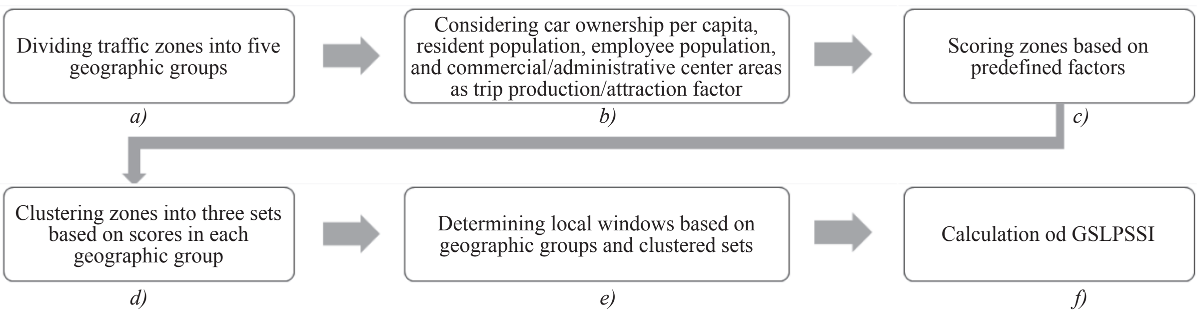


Figure 2 – Proposed procedure

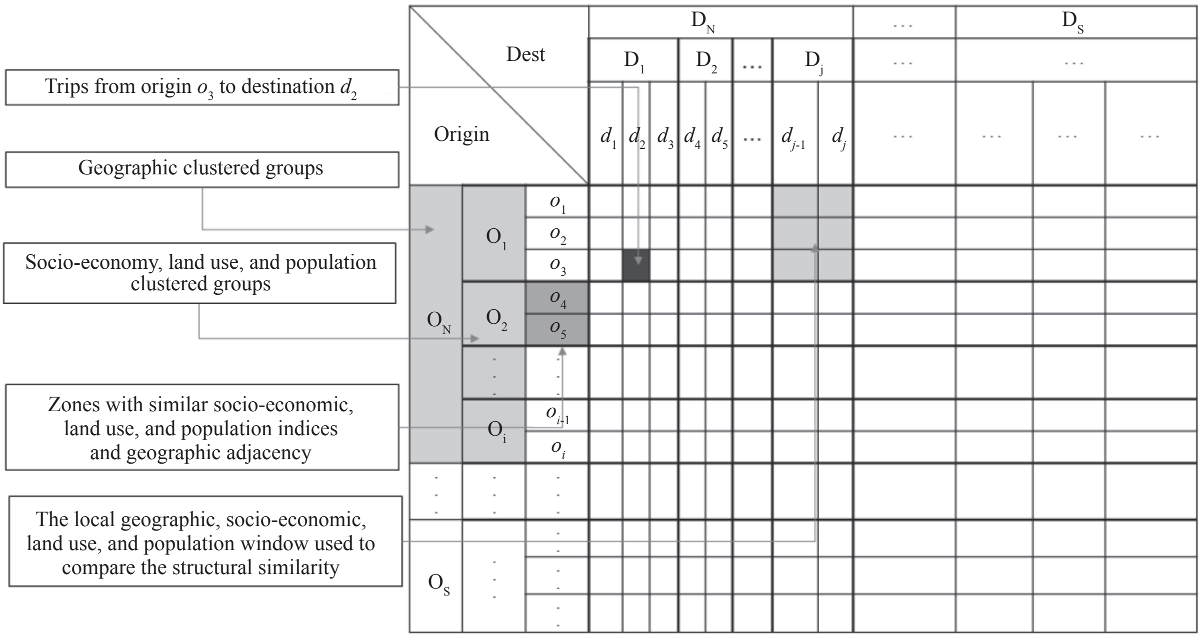


Figure 3 – Windowing in the proposed method

equation denoted as *str*(*x,y*) and evaluates the covariance σ_{xy} of the entries of the local window between the two OD matrices. *Equations 1 and 2* evaluate the numerical similarity of the two OD matrices, while *Equation 3* examines the structural dissimilarities. In these equations, *x* and *y* are sets of origins and destinations in the same window in matrices X and Y, respectively. *Equation 4* represents a combination of the aforementioned equations, where the coefficients c_1, c_2, and c_3 are utilised to stabilise the solutions when the mean or standard deviation is zero. In general, it is assumed that $c_3=c_2/2$. Previous studies set c_1 and c_2 to 10^{-10} and 10^{-2}, respectively [28]. These coefficients can be assumed to be zero when set to values that do not lead to zero mean and standard deviation. Furthermore, *α*, *β*, and *γ* represent the relative importance of the mean, standard deviation, and covariance. These parameters are typically assumed to be 1. As a result, structural similarity (SSIM) is formulated in *Equation 5*. The MSSIM of T local windows is calculated by *Equation 6*. The MSSIM varies from -1 to 1, in which 1 represents identical matrices, while -1 represents inverse matrices [9].

$$l(x,y)=\frac{2\mu_x\mu_y+c_1}{\mu_x^2+\mu_y^2+c_1} \tag{1}$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+c_2}{\sigma_x^2+\sigma_y^2+c_2} \tag{2}$$

$$str(x,y)=\frac{\sigma_{xy}+c_3}{\sigma_x\sigma_y+c_3} \tag{3}$$

$$\begin{gathered}SSIM(x,y)=[l(x,y)^{\alpha}][c(x,y)^{\beta}][str(x,y)^{\gamma}];\\ \alpha>0,\quad \beta>0 \text{ and } \gamma>0\\ \text{assuming } \alpha=\beta=\gamma=1 \text{ and } c_3=c_2/2\end{gathered} \tag{4}$$

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{5}$$

$$MSSIM(X,Y)=\frac{1}{T}\sum_1^T SSIM(x,y) \tag{6}$$

Therefore, the SSIM of each local window is calculated using *Equation 5*. By applying *Equation 5* to M local windows in the present paper, GSLPSSI is obtained as *Equation 7*. Furthermore, the structural component of the proposed method is calculated by *Equation 8* (*Figure 2f*). Actually, *Equation 8* is the structural part of the GSLPSSI method which is denoted as GSLPSTR and compares the covariance (σ_{xy}) of the local window entries between the two OD matrices.

$$GSLPSSI(X,Y)=\frac{1}{M}\sum_1^M SSIM(x_m,y_m) \tag{7}$$

$$GSLPSTR(X,Y)=\frac{1}{M}\sum_1^M STR(x_m,y_m) \tag{8}$$

Equation 7 and 8 compare the OD matrices. Also, x_m and y_m are a set of OD pairs in the local window *m*.

Evaluation of the proposed method in Tehran

Tehran is the capital of Iran, with a population of 8.7 million and an area of 1,200 square kilometres [29]. Tehran was zoned at two levels to perform traffic analysis. At the first level (the lowest level), the metropolitan area of Tehran is divided into 699 traffic analysis zones (TAZs). Comprehensive urban and suburban transportation studies were carried out at this zoning level, which is referred to as Area Level 1 (AL1). The next level, Area Level 2 (AL2), divides the Tehran metropolitan area into 122 zones based on natural and artificial features. The highest level (i.e., AL3) divides Tehran into 5 zones. *Figures 4 and 5* illustrate the zoning of Tehran.

Figure 4 depicts AL1 and AL3 zoning levels of Tehran, while AL2 and AL3 zoning levels can be seen in *Figure 5*. This study implemented investigations performed at AL2. In other words, the five

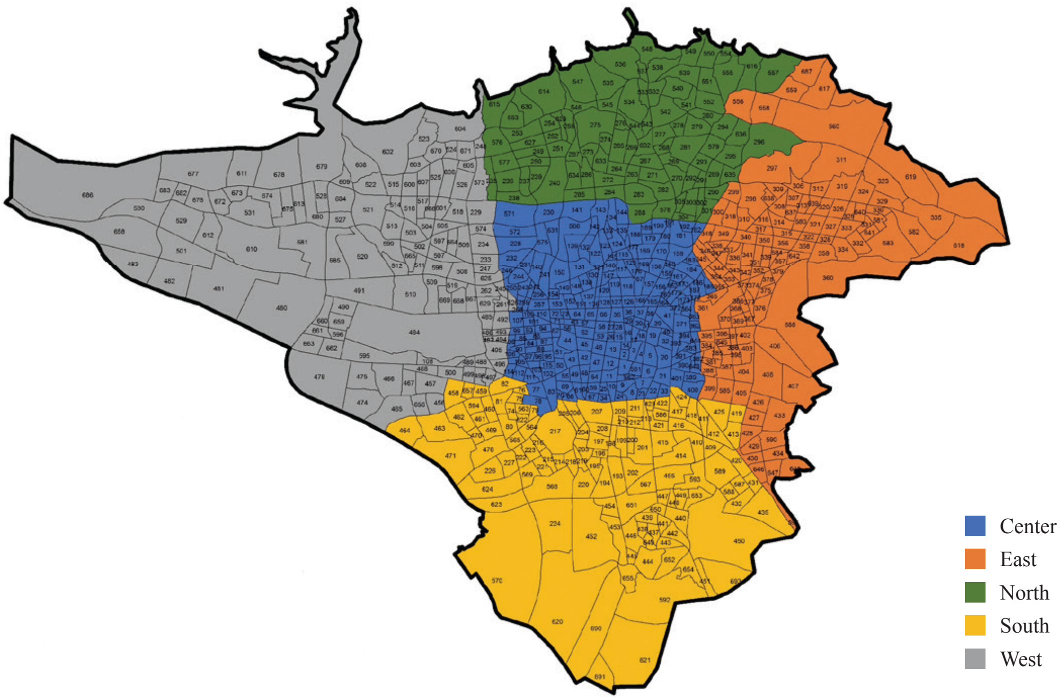


Figure 4 – AL1 and geographical zoning (AL3) of Tehran

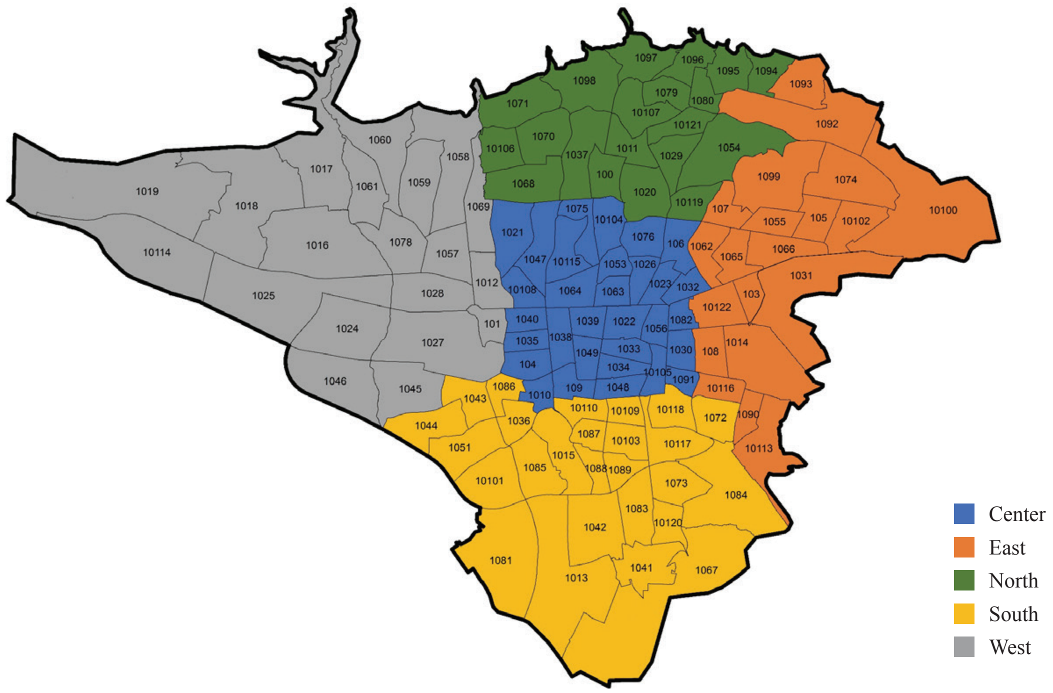


Figure 5 – AL2 and geographical zoning (AL3) of Tehran

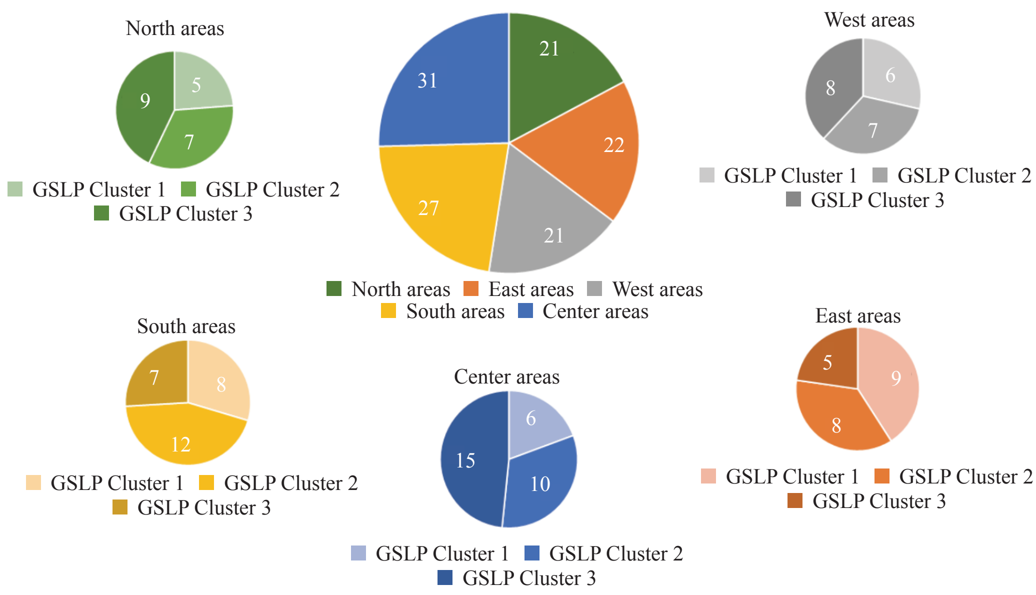


Figure 6 – Geographical zones and GSLP clusters

determinants were compared within AL2 classifying zones with trip generation similarities. To implement the proposed method, the 122 zones in AL2 were divided into five groups at AL3, as shown in *Figure 6*. Then, the zones of each group were divided into three clusters based on their potential of trip generation. GSLP cluster 1 had the lowest potential to generate trips, while GSLP cluster 3 had the highest trip generation potential. According to *Figure 6*, the main big pie chart shows the number of zones in different geographical areas (e.g. there are 31 zones in the centre area). The other five smaller pie charts in *Figure 6* illustrate the number of GSPL clusters in each geographical area. The darker colour in these pie charts shows GSLP cluster 3, the normal colour shows GSLP cluster 2, and the pale colour displays GSLP cluster 1. For instance, the small pie chart in blue colour, which is related to the centre area, has 15 zones in GSLP cluster 3, 10 zones in GSLP cluster 2, and 6 zones in GSLP cluster 1.

In order to evaluate the proposed method, OD matrices were extracted from the global positioning system (GPS) data of the Neshan navigation software in Tehran metropolitan. Raw GPS data, collected under privacy protection regulations, represent users' locations during a day. This data includes 326 million records of applicants' locations. The data were filtered by the rules which determine the origins and destinations. We can summarise the whole process in three main steps: (a) Defining origin and destination: a user ID gets stop-tag when it has 5 similar records in one zone or 15 minutes of continuous presence in that zone; (b) Move from one zone to another with the speed higher than 100 kilometre per hour is then eliminated (due to positional error); (c) Deleting the records which are observed in only one zone during the entire day. The origins and destinations were aggregated at AL2, forming a 122×122 daily OD matrix. It should be noted that extensive research has been conducted on the conversion of GPS data into OD matrices [4, 30, 31]. This study utilised the GPS-estimated OD matrix only to validate the proposed method. *Figure 7* illustrates the windowing of the GSLPSSI on a working day.

The proposed method was compared with the MSSIM, GSSI, and SLPSSI for validation. For this purpose, all the mentioned methods were implemented in Tehran metropolis as a test case. Following this goal, in the GSSI method AL2 zones were divided into northern, southern, eastern, western, and central groups, forming 25 local windows. Moreover, in order to implement the SLPSSI method, AL2 zones were divided into five clusters based on socioeconomic, land-use, and population characteristics, where SLP cluster 1 and SLP cluster 5 had the lowest and highest trip production and attraction potentials (TPAP), respectively (see *Table 1*). These clusters formed 25 local windows. The clustering process of SLPSSI is the same as that of the GSLPSSI mentioned in section 3.

Origin \ Destination			North area									...	South area								
			GSLP Cluster 1			GSLP Cluster 2			GSLP Cluster 3			...	GSLP Cluster 1			GSLP Cluster 2			GSLP Cluster 3		
			1011	...	1020	1071	...	1094	1037	...	10119	...	1051	...	1084	1089	...	1043	10109	...	10118
North area	GSLP Cluster 1	1011	679	...	10	76	...	4	16	...	8	...	158	...	18	34	...	78	54	...	9
		. . .	...	...	...	...	...	...	...	...	...	...		...			...			...	
		1020	140	...	345	49	...	54	7	...	3	...	45	...	24	65	...	13	56	...	78
	GSLP Cluster 2	1071	56	...	167	258	...	43	12	...	8	...	12	...	40	76	...	43	98	...	60
		. . .	...	...	...	...	...	...	...	...	...	...		...			...			...	
		1094	95	...	54	140	...	385	15	...	60	...	42	...	76	5	...	67	12	...	11
	GSLP Cluster 3	1037	5	...	25	88	...	170	980	...	17	...	7	...	32	41	...	58	84	...	14
		. . .	...	...	...	...	...	...	...	...	...	...		...			...			...	
		10119	8	...	45	54	...	6	230	...	470	...	54	...	59	15	...	32	90	...	79
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
South area	GSLP Cluster 1	1051	45	...	15	86	...	5	18	...	19	...	411	...	56	89	...	140	210	...	32
		. . .	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
		1084	120	...	29	66	...	58	19	...	41	...	170	...	392	270	...	168	173	...	51
	GSLP Cluster 2	1089	28	...	17	78	...	61	78	...	34	...	64	...	78	580	...	121	135	...	196
		. . .	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
		1043	86	...	43	57	...	29	19	...	75	...	16	...	32	45	...	432	76	...	139
	GSLP Cluster 3	10109	170	...	22	72	...	54	160	...	154	...	73	...	6	76	...	57	842	...	345
		. . .	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
		10118	67	...	32	29	...	17	189	...	215	...	61	...	19	116	...	175	297	...	956

Figure 7 – GSLPSSI windowing on a sample working day

Table 1 – Number and characteristics of similar zones in SLPSSI

Cluster name	Feature	Num. of zones
SLP Cluster 1	Very Low TPAP	8
SLP Cluster 2	Low TPAP	32
SLP Cluster 3	Medium TPAP	49
SLP Cluster 4	High TPAP	21
SLP Cluster 5	Very high TPAP	12

4. RESULTS AND DISCUSSION

In this section, GSLPSSI is compared with MSSIM, GSSI, and SLPSSI, and its advantages are demonstrated. The GPS OD from 5 October to 15 November 2019 was employed. It should be noted that Thursday and Friday are considered the weekend in the Iranian calendar (Thursday is a partial working day).

4.1 Structural comparison of week days

For the structural comparison of week days, the GPS OD data were gathered in one-month, averaged on every week day, and the seven OD matrices representing trips on each day of the week were obtained. The mid-week day, i.e., Monday in Iran, was then compared with other days of the week. As mentioned earlier, five clusters of zones were employed in the GSSI and SLPSSI. It should be noted that the entire matrix was considered a single-window in the MSSIM. Fifteen groups of zones were employed in the GSLPSSI based on geographical, socioeconomic, land-use, and population characteristics as compared in *Figure 8*. All methods indicate the structural similarity of Monday trips to other working days, specifically to Sunday and Tuesday which are middle working days of week in the Iranian solar calendar. The difference between Monday and Friday (the weekend) OD matrices can be seen in all methods. Contrary to the GSSI, SLPSSI, and GSLPSSI, MSSIM could not detect the structural difference between Monday and Thursday OD matrices. Moreover, the only method that indicates the structural difference in travel patterns of Wednesday (the last full working day) is the GSLPSSI which selects the local windows appropriately. On weekends, workers who work in Tehran metropolis and are residents of satellite cities return home, and the evening marks the recreational trips of Tehran residents, thus the major difference can be predicted precisely on Wednesday. The reverse process can be seen on Saturday and most of the indices detect it well, although it should be noted that the GSLPSSI could detect it more apparently.

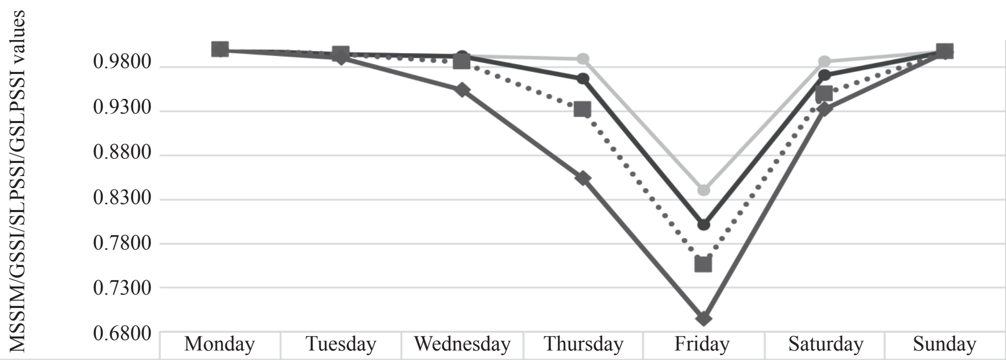
According to *Figure 8*, the GSLPSSI (0.8543) and SLPSSI (0.9319) values properly show the structural OD matrix dissimilarities of Thursday (the partial working day of the week) and Monday (the midweek working day). However, GSSI and MSSIM do not accurately show this difference. Based on the structural similarities comparison result of Friday to Monday, the proposed method seems to have approximately 9%, 15%, and 21% higher accuracy than SLPSSI, GSSI, and MSSIM, respectively, since the structural similarly of OD matrices on Friday is entirely different from those on other days of the week. The results of *Figure 8* show a monotonic decrease in the value of indices for MSSIM, GSSI, SLPSSI, and GSLPSSI respectively, which represent the strength of the prediction of each index in finding the OD matrix differences. It could be concluded that besides finding the OD matrix dissimilarities on Wednesday more clearly, the presented index could investigate the difference more accurately for other days of the week in comparison with the other indices.

4.2 Local windows travel patterns

The local windows of the GSLPSSI approach enable the comparison of travel patterns on two specific days of a week (for instance Monday to Friday) for a set of zones with similar socioeconomic, land-use, and population characteristics and geographical correlations which make a local window.

Each local window can have a different travel pattern (different structure) from other local windows and understanding these differences is helpful for transportation experts in allocation of resources. *Figure 9* compares Monday on 21 October 2019, with Friday (the weekend) on 25 October 2019, by the GSLPSSI method. In this comparison, trips from the origin with the lowest trip generation potential in western zones (W-GSLP cluster 1) are compared with other GSLP clusters.

According to *Figure 9*, the pattern of trips from origins in the W-GSLP cluster 1 to the E-GSLP cluster 2 was found to have the highest structural similarity (i.e., 0.9431). Overall, the travel pattern of Monday seems not to be significantly different from Friday for trips to eastern destinations (due to high similarity values). Therefore, travel demand solutions



	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
MSSIM	1.0000	0.9951	0.9929	0.9895	0.8400	0.9866	0.9980
GSSI	1.0000	0.9948	0.9922	0.9671	0.8012	0.9712	0.9975
SLPSSI	1.0000	0.9947	0.9859	0.9319	0.7558	0.9498	0.9972
GSLPSSI	1.0000	0.9904	0.9543	0.8543	0.6943	0.9324	0.9968

Figure 8 – Structural OD matrix comparison of Monday with other days of week

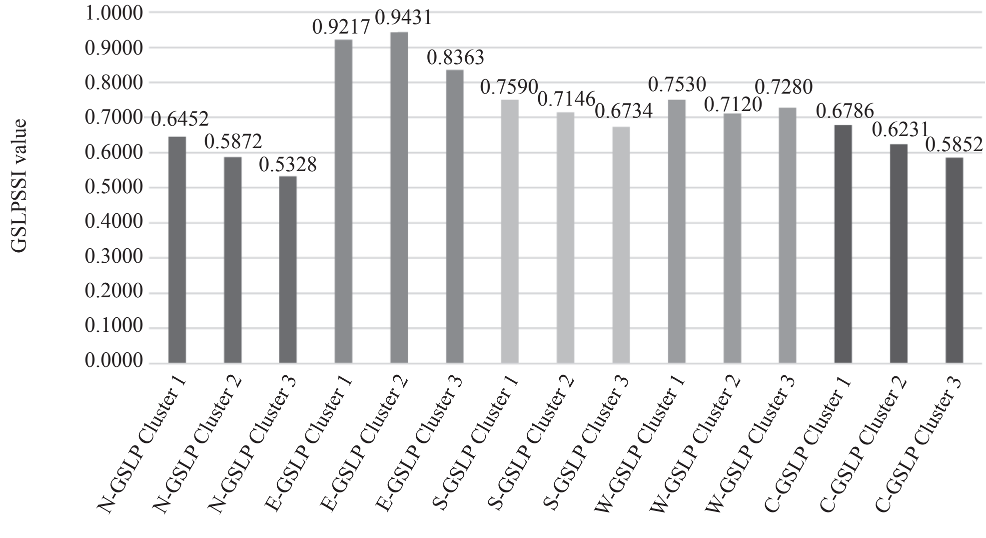


Figure 9 – Structural similarities of OD trips from origins in the W-GSLP cluster 1 on Monday vs. Friday

can be implemented for these group of zones, regardless of the day. Moreover, trips from origins in the W-GSLP cluster 1 to destinations in N-GSLP cluster 3 show the lowest structural similarity (i.e., 0.5328). According to the recreational centres placement, zones in this cluster have high potential for trip attraction at weekends, since such trips mostly do not occur on non-working days. Therefore, it is suggested that the public transportation between these zones is completely modified for weekends. C-GSLP cluster 3 shows the second-lowest structural similarity (i.e., 0.5852). Since the zones in this cluster are located in Central Business Districts (CBD) of Tehran, the zones in C-GSLP cluster 3 are the destination of work trips on working days, and the exclusion of such trips on weekends would completely change this pattern. Differences are observed in the structural similarities of each cluster in these five geographical groups despite their geographical location similarity; the difference in GSLPSSI values are due to their different trip generation potentials. Therefore, the proposed method seems to be able to identify and interpret the local travel patterns more appropriately.

4.3 Performance of the proposed method on sparse OD matrices

Since a considerable portion of transportation studies were conducted on the modelling and investigation of traffic flows at low zoning levels (e.g. AL1), it was required to evaluate the performance of GSLPSSI on AL1 OD matrices. GPS OD matrix has sparsity at AL1. For example, the OD matrix has a sparsity of 23% on Monday, 20 October at AL1. Therefore, the proposed method was evaluated on OD matrices with sparsity at AL1 in this section. For this purpose, the GPS OD data of Monday is compared to Friday, Thursday, and Tuesday of the second week of October 2019. *Table 2* shows the results of this comparison by different methods.

Table 2 – Comparison of the MSSIM, GSSI, SLPSSI, and GSLPSSI on sparse OD matrices

	Monday vs. Thursday	Monday vs. Friday	Monday vs. Tuesday
MSSIM	0.9472	0.8013	0.9523
GSSI	0.9365	0.7563	0.9483
SLPSSI	0.8767	0.7018	0.9413
GSLPSSI	0.8132	0.6542	0.9389

According to *Table 2*, all the methods could detect the structural difference of the sparse OD matrix of Monday from Friday. Considering the significant trip distribution difference of Friday from the working days (e.g. Monday), the GSLPSSI had a value of 0.6542, showing the highest accuracy in structural dissimilarity detection (7% higher accuracy than SLPSSI). MSSIM (0.9472) and GSSI (0.9365) were not able to correctly detect the structural dissimilarities of the travel pattern on Monday from that on Thursday, while SLPSSI (0.8767) and GSLPSSI (0.8132) showed much better performance. MSSIM had poor performance in calculating the structural similarity of Monday and Tuesday (0.9523) and Thursday (0.9472).

4.4 Sensitivity analysis

Sensitivity analysis measures the impacts of a change in input on output [32]. It is employed to measure the robustness and sensitivity of output to changes in input data. Although GSLPSSI and MS-SIM have the same mathematical formulation, they have different structural similarity outputs as they have different local windows. In this section the robustness of both GSLPSSI and GSLPSTR is evaluated. An efficient and effective statistical index for OD matrices comparison can simultaneously detect changes in the cell values of OD matrices and trip distribution variations. Study area is Tehran metropolitan and OD matrix X derived from the GPS OD data on Tuesday, 21 October 2019, was utilised as the reference matrix. Y matrices were produced by changing (and applying coefficients of) the reference matrix for comparison purposes. *Table 3* describes the sensitivity analyses of the GSLPSSI and GSLPSTR.

Table 3 – Sensitivity analysis conditions of GSLPSSI and GSLPSTR

Conditions	Tests	Structure	Cell values
Condition 1	Uniform scaling	Identical	Identical
Condition 2		Identical	Different
Condition 3	Random scaling	Different	Different

According to *Table 3*, the reference matrix is multiplied by constants in conditions 1 and 2. Constant is 1 in condition 1, and the product matrix is the same as the reference matrix. In condition 2, a constant number other than one is employed to build matrix Y. In this case, the cell values change, but the matrix structure remains unaffected. In condition 3, variable multipliers are employed to build matrix Y, in which cell values and matrix structure are different from the reference matrix.

Constant multipliers

In this part, the sensitivity of the GSLPSTR and GSLPSSI to constant multipliers are evaluated. Matrix Y_i was built by multiplying the reference matrix X by constant β. The β constant was selected from [0.2, 0.4, …,1.8, 2]. The proposed method is sensitive to input variations when (a) GSLPSSI is zero at β=1, and its value reduces as the difference of β from 1 increases, and (b) GSLPSTR is zero regardless of β. *Table 4* provides the GSLPSSI and GSLPSTR values for comparing matrices X and Y_i.

According to *Table 4*, both GSLPSSI and GSLPSTR are equal to one with β=1. Therefore, the proposed method is effective in condition 1. GSLPSTR remained 1 when multipliers other than one were applied. However, an increase or decrease in multipliers decreases GSLPSSI from 0.921 to 0.598 as the multiplier increases from 1.2 to 2.0. It also decreases from 0.954 to 0.307 as the multiplier decreases from 0.8 to 0.2. Therefore, both GSLPSSI and GSLPSTR are efficient and robust in condition 2.

Variable multiplier

In this part the sensitivity of the GSLPSTR and GSLPSSI to variable multipliers was discussed. Variable multipliers were applied using four random scaling percentages of α, i.e., 5%, 10%, 15%, and 20% with three demand scenarios. Demand scenarios are typically employed in travel demand modelling in simulation software [33, 34]. For each scenario, the reference matrix X was compared with 100 manipulated Y matrices, to calculate mean GSLPSSI and GSLPSTR. These scenarios are described below.

I. Low-demand scenario, derived from outdated studies

   In Scenario I, GSLPSSI and GSLPSTR were calculated for the X and $Y^l_{i,a}$ where $Y^l_{i,a}=X(0.6+\alpha\times rand[0,1])$ and $i\in[1,100]$. For example, $Y^l_{i,a}$ varies between 60 and 80% of X for α=20%. The same case is applied for other values of α.

II. Mid-demand scenario, derived from realistic estimations

   In Scenario II, GSLPSSI and GSLPSTR were calculated for the X and $Y^l_{i,a}$ where $Y^l_{i,a}=X(0.8+\alpha\times rand[0,1])$ and $i\in[1,100]$. For example, $Y^l_{i,a}$ varies between 80 and 100% of X for α=20%. The same case is applied for other values of α.

Table 4 – GSLPSSI and GSLPSTR under constant multipliers

	β=0.2	β=0.4	β=0.6	β=0.8	β=1	β=1.2	β=1.4	β=1.6	β=1.8	β=2
GSLPSSI	0.307	0.662	0.854	0.954	1	0.921	0.839	0.763	0.681	0.598
GSLPSTR	1	1	1	1	1	1	1	1	1	1

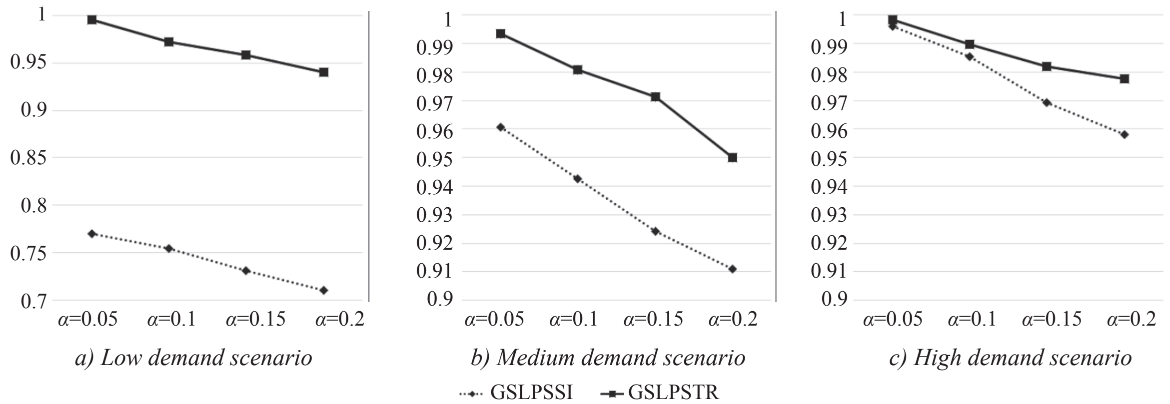


Figure 10 – Random multiplier results of the GSLPSTR and GSLPSSI

II. High-demand scenario in a traffic jam (peak hour)

In Scenario III, GSLPSSI and GSLPSTR were calculated for the X and $Y^l_{i,a}$ where $Y^l_{i,a}=X(1.05+\alpha\times rand[0,1])$ and $i\in[1,100]$. For example, $Y^l_{i,a}$ varies between 105 and 125% of X for α=20%. The same case is applied for other values of α.

The proposed method would be efficient and robust if the structural similarity for both GSLPSSI and GSLPSTR varies according to $Y^l_{i,a}$ in each scenario. It is expected that a rise/decline in the random multiplier reduces/raises the structural similarity of the two matrices. *Figure 10* shows the mean GSLPSSI and GSLPSTR based on random multipliers for a sample of 100 in each scenario. According to *Figure 10*, the mean GSLPSSI of the reference matrix and multiplied matrices reduce as the random multiplier α increases in all scenarios. For example, GSLPSSI was calculated to be 0.7698, 0.7543, 0.7312, and 0.7103 at α=0.05, α=0.10, α=0.15, and α=0.20, respectively, in Scenario I. This decreasing trend is also the case with the mean GSLPSTR, which only compares the structures of the two matrices. Therefore, it can be stated that the proposed method is efficient and robust in condition 3 for detecting the structural similarities of matrices produced by random multipliers.

5. CONCLUSION

In this study an OD matrix comparison method named geographical, socioeconomic, land-use, and population structural similarity index (GSLPSSI) is proposed and discussed. A comprehensive study is done to detect the structural similarities of two matrices in terms of both numbers (i.e., cell value differences) and structure (trip distribution difference). Generally, traditional methods only compare the numerical deviations of the two matrices and fail to consider their structural dissimilarities. As shown in literature, only a few studies investigated the structural similarity of matrices. For example, MSSIM selects local windows and moves them on two matrices to detect the structural similarity of the matrices. The efficiency and accuracy of MSSIM depends on local window size. Furthermore, complementary MSSIM methods, such as GSSI and SLPSSI, have been developed where GSSI defines local windows based only on geographical characteristics of zones. Indeed, the geographical positions cannot classify correlated zones in the same group in the best possible way. Furthermore, the SLPSSI only considers trip production and attraction (trip generation) factors to determine local windows and fails to consider the geographical adjacency of zones. Therefore, it cannot necessarily detect all differences in travel patterns. The proposed method classifies zones based on their geographical adjacency and their similarities in socioeconomic, land-use, and population characteristics at the same time. The results can be summarized as follows:

1) The proposed method detects the OD matrices structural similarities (i.e., cell values and trip distribution) of the days of the week more accurately than previous methods. GSLPSSI shows 9%, 15%, and 21% higher accuracy than SLPSSI, GSSI, and MSSIM, respectively, during the structural similarity/difference calculation of two OD matrices.
2) The proposed method can detect new travel patterns in local windows. It means that, based on the selection of local windows, the travel patterns of local window pairs can be analysed. These patterns cannot be observed in other methods. In addition, these travel patterns enable transportation experts to make managerial decisions.

3) The proposed method can not only compare the structures of sparse OD matrices but also calculate the structural similarities of two OD matrices with 7% higher accuracy.

As a result, the proposed method was found to be practical for the GPS OD matrix of Tehran. The weekend OD matrix is expected to differ from those of other days of the week since it has a different travel pattern. It was shown that the GSLPSSI can better demonstrate such differences of weekend from other working days.

The sensitivity analysis results demonstrate that the GSLPSSI is an efficient and robust method and could be employed to compare OD matrices. The results of constant multipliers show that the multipliers do not change the structural analysis of OD matrices in the GSLPSTR. Thus, the GSLPSTR part of our proposed method could be used when there is no significant difference between the trip distribution priorities of reference and estimated matrices (in OD matrix estimation processes) or between OD matrices with similar travel patterns (e.g. Monday and Thursday in the Iranian calendar). A competent OD matrix structural similarity index should be sensitive to random and variable multipliers; this was verified by the results obtained from applying random and variable multipliers to the proposed method. Moreover, this capability of the GSLPSSI could be used for optimisation algorithms to estimate OD matrices to evaluate the convergence of the reference and estimated OD matrices [35, 36].

ACKNOWLEDGEMENTS

The authors wish to thank Tehran Municipality and the Rajman Information Structures Company for providing the required data for this research.

روش ترکیبی نوین جهت تشخیص الگوهای سفر براساس مقایسه ساختاری ماتریس‌های مبدأ-مقصد

چکیده

ماتریس مبدأ-مقصد سفرها اطلاعات تحلیلی را در خصوص تعداد سفرها و نحوه توزیع آنها بین نواحی در اختیار متخصصان حمل‌ونقل قرار می‌دهد. به منظور مقایسه ماتریس‌های مبدأ-مقصد نه تنها بررسی اختلاف اعداد میان سلول‌ها بلکه مقایسه اختلاف ساختاری آنها که شامل اولویت سفرها از هرکدام از نواحی و الگوی سفرها در محدوده مورد مطالعه است، دارای اهمیت است. شاخص میانگین شباهت ساختاری، شاخص میانگین شباهت ساختاری براساس پنجره‌های جغرافیایی و شاخص میانگین شباهت ساختاری براساس پنجره‌های اجتماعی اقتصادی، کاربری زمین و جمعیت به منظور مقایسه ساختاری ماتریس‌های مبدأ-مقصد پیش از این توسعه پیدا کرده‌اند. این روش‌ها در زمینه تشخیص تفاوت در الگوی سفرها دارای اشکالات غیرقابل چشم‌پوشی می‌باشند، لذا در این مقاله شاخصی نوین که به صورت همزمان مشخصات جغرافیایی، اجتماعی-اقتصادی، کاربری زمین و جمعیت را در نظر می‌گیرد به منظور مقایسه ماتریس‌های مبدأ-مقصد توسعه پیدا کرده است. شاخص ارائه شده بوسیله ماتریس مبدأ- مقصد حاصل از داده‌های موقعیت مکانی سامانه موقعیت‌یاب جهانی تلفن‌های همراه در کلان‌شهر تهران مورد ارزیابی قرار گرفته است. همچنین میزان قدرتمندی روش پیشنهادی با استفاده از تحلیل حساسیت تائید گردیده است. روش پیشنهادی این مقاله به ترتیب ۲۱، ۱۵ و ۹ درصد دقت بیشتری نسبت به روش شاخص میانگین شباهت ساختاری، روش شاخص میانگین شباهت ساختاری براساس پنجره‌های جغرافیایی و روش شاخص میانگین شباهت ساختاری بر اساس پنجره‌های اجتماعی-اقتصادی، کاربری زمین

شهریار افندیزاده زرگری، استاد[۱]
(نویسنده مسئول)
ایمیل: zargari@iust.ac.ir
امیرمسعود معمارنژاد، دانشجوی دکتری تخصصی[۱]
ایمیل: memarnejad@civileng.iust.ac.ir
حمید میرزاحسین، دانشیار[۲]
ایمیل: mirzahossein@eng.ikiu.ac.ir
[۱] گروه حمل‌ونقل، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران، کد پستی: ۱۳۱۱۴-۱۶۸۴۶
[۲] گروه مهندسی عمران – برنامه‌ریزی حمل‌ونقل، دانشگاه بین‌المللی امام خمینی، قزوین، ایران، کد پستی: ۹۶۸۱۸-۳۴۱۴۸

و جمعیت داشته است. همچنین روش شاخص میانگین شباهت ساختاری براساس پنجره‌های جغرافیای، اجتماعی-اقتصادی، کاربری زمین و جمعیت (شاخص پیشنهادی این مقاله) ۷ درصد دقت بیشتری نسبت به روش شاخص میانگین شباهت ساختاری براساس پنجره‌های اجتماعی-اقتصادی، کاربری زمین و جمعیت در ماتریس‌های پراکنده داشته است.

کلمات کلیدی

شباهت ساختاری، الگوی سفرها، ماتریس مبدأ-مقصد، نواحی ترافیکی، کلان‌شهر تهران

REFERENCES

[1] Cipriani E, Florian M, Mahut M, Nigro M. A gradient approximation approach for adjusting temporal origin–destination matrices. *Transportation Research Part C: Emerging Technologies.* 2011;19(2): 270-82. doi: 10.1016/j.trc.2010.05.013.

[2] Djukic T, Flötteröd G, Van Lint H, Hoogendoorn S. Efficient real time OD matrix estimation based on Principal Component Analysis. *2012 15th International IEEE Conference on Intelligent Transportation Systems.* IEEE; 2012. p. 115-121. doi: 10.1109/ITSC.2012.6338720.

[3] Michau G, et al. A primal-dual algorithm for link dependent origin destination matrix estimation. *IEEE Transactions on Signal and Information Processing over Networks*. 2016;3(1): 104-13. doi: 10.1109/TSIPN.2016.2623094.

[4] Moreira-Matias L, et al. Time-evolving OD matrix estimation using high-speed GPS data streams. *Expert Systems with Applications.* 2016;44: 275-88. doi: 10.1016/j.eswa.2015.08.048.

[5] Osorio C. Dynamic origin-destination matrix calibration for large-scale network simulators. *Transportation Research Part C: Emerging Technologies.* 2019;98: 186-206. doi: 10.1016/j.trc.2018.09.023.

[6] Xiong X, Ozbay K, Jin L, Feng C. Dynamic origin–destination matrix prediction with line graph neural networks and kalman filter. *Transportation Research Record.* 2020;2674(8): 491-503. doi: 10.1177/0361198120919399.

[7] Novačko L, Šimunović L, Krasić D. Estimation of origin-destination trip matrices for small cities. *Promet – Traffic&Transportation.* 2014;26(5): 419-28. doi: 10.7307/ptt.v26i5.1501.

[8] Bierlaire M. The total demand scale: A new measure of quality for static and dynamic origin–destination trip tables. *Transportation Research Part B: Methodological.* 2002;36(9): 837-50. doi: 10.1016/S0191-2615(01)00036-4.

[9] Djukic T, Hoogendoorn S, Van Lint H. Reliability assessment of dynamic OD estimation methods based on structural similarity index. *Proceedings of the 92rd Annual Meeting of the Transportation Research Board*; 2013. doi: 10.13140/RG.2.1.4174.1929.

[10] Jin C, Nara A, Yang JA, Tsou MH. Similarity measurement on human mobility data with spatially weighted structural similarity index (SpSSIM). *Transactions in GIS*. 2020;24(1): 104-22. doi: 10.1016/S0191-2615(01)00036-4.

[11] Tavassoli A, Alsger A, Hickman M, Mesbah M. How close the models are to the reality? Comparison of transit origin-destination estimates with automatic fare collection data. In: *Australasian Transport Research Forum 2016 Proceedings, 1 January 2016, Melbourne, Australia*; 2016. p. 1-15.

[12] Behara KN, Bhaskar A, Chung E. A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies*. 2020;111: 513-30. doi: 10.1016/j.trc.2020.01.005.

[13] Andrienko G, Andrienko N, Fuchs G, Wood J. Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. *IEEE transactions on visualization and computer graphics.* 2016;23(9): 2120-36. doi: 10.1109/TVCG.2016.2616404.

[14] Bierlaire M, Toint PL. Meuse: An origin-destination matrix estimator that exploits structure. *Transportation Research Part B: Methodological.* 1995;29(1): 47-60. doi: 10.1016/0191-2615(94)00025-U.

[15] Behara KN, Bhaskar A, Chung E. Geographical window based structural similarity index for origin-destination matrices comparison. *Journal of Intelligent Transportation Systems*. 2020;26(1): 46–67. doi: 10.1080/15472450.2020.1795651.

[16] Barceló Bugeda J, Montero Mercadé L, Marqués L, Carmona Bautista C. A Kalman-filter approach for dynamic OD estimation in corridors based on bluetooth and Wi-Fi data collection. *12th World Conference on Transportation Research WCTR*; 2010.

[17] Frederix R, Viti F, Himpe WWE, Tampère CMJ. Dynamic Origin–Destination Matrix Estimation on Large-Scale Congested Networks Using a Hierarchical Decomposition Scheme. *Journal of Intelligent Transportation Systems.* 2014;18(1): 51-66. doi: 10.1080/15472450.2013.773249.

[18] Hellinga B, Aerde MV. A statistical analysis of the reliability of using RGS vehicle probes as estimators of dynamic O-D departures rates. *IVHS Journal.* 1994;2(1): 21-44. doi: 10.1080/10248079408903813.

[19] Antoniou C, Ben-Akiva M, Koutsopoulos HN. Incorporating automated vehicle identification data into origin-destination estimation. *Transportation Research Record.* 2004;1882(1): 37-44. doi: 10.3141/1882-05.

[20] Barceló J, et al. Robustness and computational efficiency of Kalman filter estimator of time-dependent origin–destination matrices: Exploiting traffic measurements from information and communications technologies. *Transportation Research Record.* 2013;2344(1): 31-9. doi: 10.3141/2344-04.

[21] Afandizadeh Zargari S, Memarnejad A, Mirzahossein

H. A structural comparison between the origin-destination matrices based on local windows with socioeconomic, land-use, and population characteristics. *Journal of Advanced Transportation.* 2021 Jun 14;2021. doi: 10.1155/2021/9968698.

[22] Van Vuren T, Day-Pollard T, editors. 256 shades of grey-comparing OD matrices using image quality assessment techniques. *2015 Scottish Transport Applications Research Conference Proceedings*; 2015.

[23] Ruiz de Villa A, Casas J, Breen M. OD matrix structural similarity: Wasserstein metric. 2014. *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*; 2014.

[24] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004;13(4): 600-12. DOI: 10.1109/TIP.2003.819861.

[25] Djukic T. *Dynamic OD demand estimation and prediction for dynamic traffic management.* PhD thesis. Delft University of Technology; 2014.

[26] Afandizadeh Zargari S, Safari F. Using clustering methods in multinomial logit model for departure time choice. *Journal of Advanced Transportation*. 2020 Jan 15;2020. doi: 10.1155/2020/7382569.

[27] Zheng Y, Zhao G, Liu J. A novel grid based k-means cluster method for traffic zone division. *Second International Conference on Cloud Computing and Big Data in Asia*; 2015. p. 165-178. doi: 10.1007/978-3-319-28430-9_13.

[28] Pollard T, Taylor N, van Vuren T, MacDonald M. Comparing the quality of OD matrices in time and between data sources. *Proceedings of the European Transport Conference*; 2013.

[29] United Nations PD. *The World's Cities in 2016*. Retrieved 2017-08-02.; 2017. https://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf [Accessed 14th June 2020].

[30] Ge Q, Fukuda D. Updating origin–destination matrices with aggregated data of GPS traces. *Transportation Research Part C: Emerging Technologies*. 2016;69: 291-312. doi: 10.1016/j.trc.2016.06.002.

[31] Dumbliauskas V, Grigonis V, Barauskas A. Application of Google-based data for travel time analysis: Kaunas city case study. *Promet – Traffic&Transportation.* 2017;29(6): 613-21. doi: 10.7307/ptt.v29i6.2369.

[32] Iooss B, Lemaître P. A review on global sensitivity analysis methods. In: Dellino G, Meloni C. (eds) *Uncertainty Management in Simulation-Optimization of Complex Systems. Operations Research/Computer Science Interfaces Series*. Vol 59. Boston, MA: Springer; 2015. p. 101-22. doi: 10.1007/978-1-4899-7547-8_5.

[33] Djukic T, et al. Advanced traffic data for dynamic OD demand estimation: The state of the art and benchmark study. *TRB 94th Annual Meeting Compendium of Papers*; 2015. p. 1-16.

[34] Hasnine MS, Nurul Habib K. Tour-based mode choice modelling as the core of an activity-based travel demand modelling framework: A review of state-of-the-art. *Transport Reviews*. 2021;41(1): 5-26. doi: 10.1080/01441647.2020.1780648.

[35] Behara KN, Bhaskar A, Chung E. A novel methodology to assimilate sub-path flows in bi-level OD matrix estimation process. *IEEE Transactions on Intelligent Transportation Systems*; 2020. doi: 10.1109/TITS.2020.2998475.

[36] Behara KN, Bhaskar A, Chung E. Single-level approach to estimate origin-destination matrix: Exploiting turning proportions and partial OD flows. *Transportation Letters*. 2021: 1-12. doi: 10.1080/19427867.2021.1932182.