Geographical window based structural similarity index for origin-destination matrices comparison

Krishna N. S. Behara, Ashish Bhaskar, and Edward Chung

School of Civil and Environmental Engineering, Queensland University of Technology, Brisbane, Australia; Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

ABSTRACT

Most traditional metrics compare origin-destination (OD) matrices based on the deviations of individual OD flows and often neglect OD matrix structural information within their formulations. Limited metrics exist in literature for the structural comparison of OD matrices. One such metric is mean structural similarity index (MSSIM) that computes statistics on groups of OD pairs defined by local sliding windows. However, MSSIM can result in different values based on the choice of the size of the window. In literature, no clear consensus has been reported on the level of acceptability of the window size and the resulting MSSIM values. Addressing this need, we propose the concept of geographical window, and develop geographical window based structural similarity index (GSSI) that exploits OD matrix structure by computing statistics on the group of OD pairs that are geographically correlated. Compared to traditional sliding window based MSSIM, the advantages of GSSI technique identified from real case study application are (a) it preserves geographical integrity; (b) it compares results with physical significance; (c) it captures local travel patterns; (d) it compares large-scale sparse OD matrices; and (e) it is computationally efficient. A thorough sensitivity analyses suggest that GSSI is a robust statistical metric and has potential for practical applications such as, benchmarking different OD estimation methods; improving the quality of solution by maintaining structural consistency in the OD estimation process; and identifying gaps in the transit service by comparing local (within a geographical window) travel patterns of car and public transit.

ARTICLE HISTORY
Received 14 June 2019
Revised 11 July 2020
Accepted 11 July 2020

KEYWORDS
Bluetooth; Brisbane City; geographical window; OD matrix structure; statistical area; structural similarity index

Introduction

An origin-destination (OD) demand matrix represents the distribution of travel demand between different OD pairs within a transport network. It is a key input into most traffic simulation models and it essentially provides two major aspects of travel information: first, the cell values of the OD matrix represent flows between individual OD pairs (we can also term it as the mass); and second, the arrangement of OD pairs reflects the underlying structure (or skeleton) of OD matrix.

In the OD estimation process, knowledge of OD structure helps to maintain structural consistency in the OD estimates (Behara et al., 2020b; Doblas & Benitez, 2005). Similarly, accounting for the OD structure in the comparison of the OD matrices from different solution algorithms (on a benchmark network) provides a holistic quality measurement (Djukic et al., 2013).

In literature, different methods (details below) are used to quantify the similarity between two OD matrices. If the structure of the OD matrices is also considered in the similarity estimation, then we term it as structural similarity. Two OD matrices have perfect structural similarity if their structures are similar with zero differences in the OD flows.

Some notable traditional measures widely used to quantify similarity of OD matrices are: root mean square error (RMSE) (Frederix et al., 2014; Hellinga & Aerde, 1994); normalized root mean square error (RMSN) (Antoniou et al., 2004; Jaume Barceló et al., 2013; Frederix et al., 2014); mean square error (MSE) (Cascetta, 1984); mean absolute error ratio (MAER) (Kim et al., 2005); mean absolute percent error (MAPE) (Nigro et al., 2018); goodness of Theil’s fit (GU) (J. Barceló et al., 2013; Jaume Barceló et al., 2013); R-squared (R²) (Tavassoli et al., 2016); and entropy measure (E) (Ros-Roca et al., 2018). Refer to...
Hollander and Liu (2008) for more information about the traditional metrics. Although the formulations of these metrics are mathematically simple they often fail to capture the structural differences between OD matrices (Djukic et al., 2013). Thus they are incapable of accounting the OD structural properties such as trip productions/attractions, destination choices, geographical correlations (Antoniou et al., 2016). Refer to example in the next section that demonstrates the inability of traditional metrics to capture OD matrix structural differences. Also, these standard measures are incapable of recognizing the spatial and temporal proximity of flows in the adjacent cells (Van Vuren & Day-Pollard, 2015). Total demand scale (TDS), proposed by Bierlaire (2002), is another measure to assess the quality of estimated OD matrices. It measures intrinsic under-determinacy of OD estimation problem by computing the difference between minimum and maximum levels of total demand of the corresponding OD estimates. Lesser the difference between the minimum and maximum OD combinations, better is the quality of estimated OD. However, TDS is not a suitable approach for applications other than OD estimation problem (e.g., comparison of OD matrices from different day types).

Limited metrics such as mean structural similarity index or simply MSSIM (Djukic et al., 2013), 4D-MSSIM (Van Vuren & Day-Pollard, 2015), Wasserstein metric (Ruiz de Villa et al., 2014), and normalized Levenshtein distance for OD matrices (NLOD) (Behara et al., 2020a) are available in the literature that could account for OD matrix structural information. These metrics are popular in other disciplines and are introduced in transport applications as discussed below.

The metric MSSIM computes statistics i.e., mean, standard deviation and covariance comparison on groups of OD pairs defined by local windows, and can identify structural differences better than traditional metrics such as R-squared and GEH (Day-Pollard & Van Vuren, 2015). Refer to set of Equations (2) to (3) in the next Section for mathematical formulations. Originally, MSSIM is applied for structural comparison of natural images considering that neighborhood pixels are correlated. However, the major drawback is that MSSIM is sensitive to window size and can result in different values based on the choice of local window size (i.e., size of OD pairs’ group). Additionally, application of MSSIM on OD matrices needs to account more attributes that are OD related. For instance, (a) the correlation between OD pairs can exist for a variety of reasons such as spatial proximity, and trip distance, and there is no guarantee that all OD pairs within the window are correlated; (b) the stability constants (refer to Equations (2) to (3)) used in its formulation could be network specific; and (c) the weightages to SSIM components need to be further explored especially while comparing sparse and dense matrices (Day-Pollard & Van Vuren, 2015). More details about the formulation are provided in the subsequent sections.

To define a group of OD pairs that are in spatial neighborhood, Van Vuren and Day-Pollard (2015) proposed 4D-MSSIM. It is an extension to MSSIM that additionally accounts for spatial proximity between OD pairs through Euclidean distance. The nearby OD pairs need to be computed for every OD pair in turn. The proximity estimation is only a one-time computation for a particular study area, and the idea of 4D-MSSIM is more in tune with OD matrices. However, it was still in an explorative phase as Van Vuren and Day-Pollard (2015) recommended further testing its implementation on a range of networks under wider set of circumstances.

Wasserstein distance is defined as the minimum total travel time required to assign the trips between OD pairs of matrix “X” using an assignment compatible with matrix “Y.” It is a linear programing problem that yields distance in terms of travel time or minutes per driver. This metric is based on optimization technique and is computationally more intensive than MSSIM or 4D-MSSIM. Refer to Ruiz de Villa et al. (2014) for more details. Similar to Wasserstein, normalized Levenshtein distance for OD matrices (or NLOD) also compares OD flows through an optimization approach. It defines OD structure as the preference of destinations from each origin. This approach is inspired from the traditional Levenshtein widely used in computational linguistics such as comparison of string sequence. Refer to Behara et al. (2020a) for more details.

Research on structural similarity of the OD matrix is limited. This paper aims to rationally extend the applicability of SSIM on OD matrix by considering the geographical boundaries. The motivation for which is twofold. First, there is lack of consensus on the size of the window for SSIM. Second, the transport analysis zone (TAZ) boundaries can be aggregated at different levels with similar socio-economic and land use pattern.

The increasing need to minimize the gap between land use and transport modeling motivates the modelers to standardize the statistical boundaries and TAZ boundaries. For instance, Australian Bureau of Statistics (ASGS, 2017) defines hierarchy of
geographical areas for the release of statistical information. This includes statistical areas (SA) for four levels: Statistical area level 1 (SA1) to statistical area level 4 (SA4). Generally, SA1 has a population of between 200 and 800 persons; SA2 level normally reflects the sub-urban level and is an aggregation of SA1s; SA3 is designed at the regional level and is an aggregation of SA2s; and SA4 reflects labor market within each State and Territory and is an aggregation of SA3s. For transport modeling the SA boundaries are being considered as TAZ with certain modifications where needed (Weston, 2019).

We propose to compute SSIM on the geographical windows that are defined by geographical boundaries and term this extension as geographical window based structural similarity index (GSSI). The sensitivity of proposed index is thoroughly tested with a case study application using Bluetooth based OD (B-OD) matrices from the Brisbane City Council (BCC) region, Australia.

The remainder of the paper is structured as follows: First, the study demonstrates the need for OD matrix structural similarity; and then presents the traditional MSSIM method. Next, it explains the development of the proposed measure – GSSI followed by discussion on the advantages of GSSI over MSSIM with a case study application on Brisbane city. The study then tests the robustness of proposed method through sensitivity analysis; discusses about its performance; and finally, concludes with future research directions. The Appendix section at the end of this paper provides a conceptual comparison of GSSI with other structural (dis)similarity metrics.

**Need for OD matrix structural similarity**

The OD flows between different OD pairs can have strong correlations due to sharing of similar structural properties such as trip production/attraction, trip purpose, spatial proximity, travel cost, and similar routes. While the temporal correlations exist among flows for the same OD pair (Nigro et al., 2018), the spatial correlations refer to the relationship between OD flows from different OD pairs during the same time period (Djukic, 2014).

Consider an example shown in Figure 1 for the demonstration of spatial correlation between OD pairs. In this example, we consider trips between OD pair/s represented at different zonal level hierarchy; that is, SA-3 and SA-2 for higher and lower levels as shown by thick and thin grey colored boundaries in Figure 1. The SA3 origin “O” and destination “D” are Nathan and Brisbane Inner (as highlighted by yellow colored boundaries), respectively connected through Pacific Motorway. Similarly, the trip boundaries at SA2 level are origins (o1, o2, o3, and o4 correspond to Tarragindi, Moorooka, Salisbury-Nathan, Coopers Plains, and Robertson, respectively); and destinations (d1, d2, d3, and d4 are Brisbane City, Spring Hill, Fortitude Valley, and New farm, respectively).

In Figure 1 we can see that SA2 OD pairs (i.e., at lower zonal level) could be correlated when they share similar structural properties. For instance, the SA2 zones (say, o1 to o3) are geographically correlated because they belong to the same SA3 zone (i.e., Nathan); they might have similar trip generation characteristics; and mostly share the same route (i.e., Pacific Motorway) to reach destination “D.”

The traditional metrics compare OD matrices based on deviations between individual OD flows and thus cannot recognize the spatial or temporal proximity of movements in adjacent cells and the complex OD structural information (Van Vuren & Day-Pollard, 2015). To demonstrate this, let’s consider an example of comparing OD matrices $M_i$ ($=M_1$ or $M_2$) with a reference OD matrix $M_R$ (Figure 2). Here, $M_1$ is simply 1.1 times $M_{R}$, and $M_2$ is chosen randomly. Traditional metrics namely MSE, RMSE, GU, and MAE are used for matrix comparison as shown in the Equations (1)–(1c) where $M_{R,w}$ and $M_{i,w}$ are the reference and query OD flows of $w^{th}$ OD pair from OD matrices $M_R$ and $M_i$, respectively (total number of OD pairs in both matrices is $W = 16$ each). The results of comparing matrices $M_1$ and $M_2$ with $M_R$ are presented in Table 1. The first column of Table 1 presents the metrics. The second and third column are the values from metrics for both cases, respectively. $M_1$ is structurally similar to the reference matrix than that of $M_2$. The traditional metrics fail to quantify the structural differences as the values of the
metrics are approximately same for the two comparisons.

\[
\text{MSE}(\mathbf{M}_R, \mathbf{M}_i) = \frac{1}{W} \sum_{w=1}^{W} (M_{R,w} - M_{i,w})^2
\]

\[
\text{RMSE}(\mathbf{M}_R, \mathbf{M}_i) = \sqrt{\frac{1}{W} \sum_{w=1}^{W} (M_{R,w} - M_{i,w})^2}
\]

\[
\text{GU}(\mathbf{M}_R, \mathbf{M}_i) = \frac{\sum_{w=1}^{W} M_{R,w} - M_{i,w}}{\sqrt{\sum_{w=1}^{W} M_{R,w}^2} + \sqrt{\sum_{w=1}^{W} M_{i,w}^2}}
\]

\[
\text{MAE}(\mathbf{M}_R, \mathbf{M}_i) = \frac{1}{W} \sum_{w=1}^{W} \left| \frac{M_{R,w} - M_{i,w}}{M_{R,w}} \right|
\]

From the above example, we could see that traditional metrics are unable to account for the OD matrix structural properties. This demands further need to focus research on structural (dis)similarity measures as additional metrics in transport applications that involve OD matrices comparison. One such metric is mean structural similarity index (MSSIM) which could potentially differentiate the above two matrix pairs. The application of MSSIM formulation (discussed in detail in the next Section) on the matrices illustrated in Figure 2 resulted in different structural similarity values equal to 0.9910 and 0.8213 for \( \mathbf{M}_1-M_R \) and \( \mathbf{M}_2-M_R \), respectively. The similarity values match quite well with the visual proximity, that is, the pair \( \mathbf{M}_1 \) and \( \mathbf{M}_R \) are more similar than the pair \( \mathbf{M}_2 \) and \( \mathbf{M}_R \).

**Mean structural similarity index (MSSIM)**

Wang et al. (2004) proposed mean structural similarity index (MSSIM) as a quantitative measure to compare the quality of two images. They highlighted the inability of traditional metrics in accounting the (visual) structural differences between the images. In their paper (refer to Figure 5 in Wang et al. (2004)) they showed that two images had the same MSE but different SSIM when compared with the same reference. Djukic et al. (2013) argued that SSIM can be applied to compare two OD matrices because OD pairs are analogous to pixels of an image (refer to Figure 6.2 in Djukic (2014)). The details of the SSIM formulation are presented in the following sub sections.

**Local sliding window and MSSIM formulation**

The traditional SSIM computes statistics on the local window (consisting group of pixels or OD pairs) that slides pixel by pixel (cell by cell) over the entire image (OD matrix). It is often referred as “local sliding window” and is generally of size \( m \times n \) (where, \( m \times n \ll M \times N \) i.e., size of OD matrix). The concept of sliding window is used to enhance the ability of SSIM to account image distortions by computing local statistics (Brooks et al., 2008). For ease of explanation consider an example as presented in Figure 3. Here, two \( 4 \times 4 \) OD matrices, \( \mathbf{X} \) and \( \mathbf{Y} \), are presented in column one and two, respectively. These two OD matrices are
compared using SSIM. The local sliding window of 2 × 2 sub-matrix (m = n = 2) is considered and is represented as colored cells. This window slides cell by cell over the entire OD matrix and in the current example we have 9 pairs of sub-matrices (refer Figure 3a–l). The local SSIM computes the structural similarity between the sub-matrices corresponding to the windows from both OD matrices. The MSSIM is computed by averaging the local SSIM values computed for all sliding windows.

The formulation for local SSIM is the product of three individual formulations (Equations (2)–(2b)) related to mean, standard deviations and coefficient of correlation between the groups of OD pairs.

$$L(x_s, y_s) = \frac{(2\mu_x\mu_y + C_1)}{\mu_x^2 + \mu_y^2 + C_1} \quad (2)$$

$$C(x_s, y_s) = \frac{(2\sigma_x\sigma_y + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2a)$$

$$\text{STR}(x_s, y_s) = \frac{(\sigma_{x,y} + C_3)}{\sigma_x\sigma_y + C_3} \quad (2b)$$

$$\text{SSIM}(x_s, y_s) = \frac{L(x_s, y_s)^2 C(x_s, y_s)^\beta \text{STR}(x_s, y_s)\gamma}{\alpha > 0, \ \beta > 0 \text{ and } Y > 0; \quad (2c)}$$

Assuming $\alpha = \beta = Y = 1$ and $C_3 = C2/2$

$$\text{SSIM}(x_s, y_s) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

$$; \quad [-1 \leq \text{SSIM}(x_s, y_s) \leq 1] \quad (3)$$

$$\text{MSSIM}(X, Y) = \frac{1}{S} \sum_{s=1}^{S} \text{SSIM}(x_s, y_s) \quad (3a)$$

Figure 3. An example of sliding window for SSIM calculation.
where $X$ and $Y$ represent two OD matrices to be compared; $x$ and $y$ represent the group of OD pairs within local windows in both matrices. The individual components of $L(x_s, y_s)$ and $C(x_s, y_s)$ compare the mean values ($\mu_x$ and $\mu_y$) and the standard deviations ($\sigma_x$ and $\sigma_y$) among OD pairs within the same ($s^{th}$) local windows from both matrices. The component $\text{STR}(x_s, y_s)$ compares the structure by computing coefficient of correlation between $s^{th}$ group of OD pairs in both matrices. The constants $C_1$, $C_2$ and $C_3$ are meant to stabilize the result when either mean or standard deviation is close to zero. Generally, $C_3$ is assumed to be $C_2/2$. Previous studies suggested values of $10^{-10}$ and $10^{-2}$ for $C_1$ and $C_2$, respectively (Pollard et al., 2013). In the analysis conducted for this study, the OD flows within the SSIM window are not all zero, hence we assume both $C_1$ and $C_2$ as zero. The parameters $x$, $y$ and $Y$ are used to adjust the relative importance of mean, standard deviation, and structural components, respectively, and are generally assumed to be equal to 1. The SSIM $(x_s, y_s)$ (Equation (3)) is the structural similarity of $s^{th}$ local window (of dimensions $m \times n$) from both matrices; and mean SSIM or MSSIM $(X, Y)$ (Equation (3a)) reports the overall structural similarity of the OD matrices, $X$ and $Y$ that is computed by taking average of SSIM values from $S$ sliding windows. The values of both SSIM and MSSIM lie between $-1$ and 1. The value of 1 implies that matrices are the same while the reverse is true when value is $-1$. In Figure 3a, the SSIM value for local window is 0.5963 and MSSIM computed over 9 local windows is 0.6777.

**SSIM extension in literature: 4D-MSSIM**

We know that consecutive zonal numbering (i.e., 101, 102, 103 and 104 in Figure 3) in OD matrices does not always guarantee geographical adjacency between the zones. If the zones are not adjacent, then spatial adjacency among OD pairs is further not guaranteed. To apply the SSIM’s rationale of “near-by pixels” in the OD matrix context, spatial proximity between OD pairs need to be considered. In this direction, Van Vuren and Day-Pollard (2015) developed 4D-MSSIM that accounts for spatial proximity among nearby OD pairs. The proximity is computed using Euclidean distance between the spatial coordinates of OD pairs. The Euclidean distance $(d(t^a_s, t^d_c))$ from OD pair $t^a_s$ (flow from zone $a$ to zone $b$) to OD pair $t^d_c$ (flow from zone $c$ to zone $d$) is expressed using Equation (4) where the coordinates of the centroid of zones $a$, $b$, $c$, and $d$ are $(x_a, y_a)$, $(x_b, y_b)$, $(x_c, y_c)$, and $(x_d, y_d)$, respectively.

$$d(t^a_s, t^d_c) = \sqrt{(x_a - x_c)^2 + (y_a - y_c)^2 + (x_b - x_d)^2 + (y_b - y_d)^2}$$

(4)

Contributions of nearby OD pairs are calculated using Gaussian weights ($\gamma$) expressed as shown in Equation (5). Here, “$x$” refers to Euclidian distance between OD pairs and “$\sigma$” is the network attribute that is generally average trip length or diameter of sample of zones in the study area. The shape of Gaussian plot with weight ($\gamma$) in $y$-axis and Euclidian distance in $x$-axis is shown in Figure 4.

$$\gamma = e^{(-x^2/\sigma^2)}$$

(5)

The final 4D-MSSIM is the MSSIM computed on the group of weighted OD flows. While the concept is interesting, it was still in an exploratory phase and there were few questions that needed to be answered. First, the cluster of “near-by” OD pairs is sensitive to the selection of initial OD pair; and second, Euclidian distance might not be an accurate measure of spatial proximity between OD pairs. For example, travel distance between two sets of OD pairs separated by a natural barrier (say, river) could be much higher than a mere Euclidian distance. In regards to Equation (5), Van Vuren and Day-Pollard (2015) mentioned that “The value in the denominator deserves some attention; it would make sense to link this to attributes of the network e.g., the average trip length or average zone diameter and would thus be mode and possibly purpose-specific. We have not yet analyzed the effect of different assumptions in this respect.”

**Need for improvements in the SSIM approach**

Although the formulation of SSIM has the ability to capture the structure of OD flows, the approach adopted needs further improvement due to the following reasons:
First, SSIM is sensitive to the size of local window due to which no clear consensus has been reported on the level of acceptability of sliding window size and the resulting MSSIM values. To demonstrate the sensitivity of SSIM toward window size, consider MSSIM values computed using different window sizes (from $3 \times 3$ to $20 \times 20$) for Monday and Sunday, and Monday and Tuesday Bluetooth based OD (B-OD) matrix pairs that are constructed from the BCC Bluetooth data (more details about B-OD matrix are presented in the next section). Figure 5 presents the results where blue line is for Monday and Sunday, and orange line is for Monday and Tuesday comparison. The x-axis represents the size of the local window and y-axis shows the MSSIM value. The order of OD pairs is same in the matrices for Sunday, Monday, and Tuesday. It is observed that as the size of sliding window increases, sensitivity of SSIM toward subtle structural differences between the OD matrices decreases. The MSSIM values increase as the sliding window size increases. Similar results are observed by Brooks et al. (2008) in comparing images using different window sizes. The rate of increment of MSSIM values is less for Monday and Tuesday pair as compared to Monday and Sunday pair. This is due to similar travel patterns between Monday and Tuesday (both working days) and less similar patterns between Monday and Sunday pair.

To circumvent this ambiguity about local window size literature suggests that SSIM should be computed over the entire OD matrix (Djukic, 2014). However, by doing so it will result in a statistical estimation that is less sensitive to structural changes within the OD matrix. According to the law of large numbers, the variance within a sample tends to decrease if the sample size increases. Larger window dimensions imply a greater number of OD pairs to be compared. Thus, the variance (distortion), and covariance (correlation distortion) parameters that capture structural changes within, and between OD matrices shall be reduced. Therefore, larger window size makes SSIM to be less sensitive to structural distortions, and thus, a smaller sized window is recommended.

Second, as the size of sliding window become smaller, the computational cost also increases. For instance, the number of pairs of local sliding windows that need to be compared from both OD matrices are $(M-m+1) \times (N-n+1)$.

Third, not all OD pairs within a local window are correlated, and the local SSIM value does not have any physical meaning attached to it. For example, the local SSIM value of 0.5963 in Figure 3a does not convey any physical significance if the OD pairs are not

![Figure 5. Sensitivity of MSSIM toward local window size.](image)
correlated within the local window. Djukic (2014) tried to address correlations among the OD pairs from their flow values (especially if volumes are high) by matrix-reordering (i.e., sorting each row of OD matrix in the order of OD pair volumes). Thus, correlated OD pairs can lie in the same neighborhood i.e., all high-volume OD pairs on one side and remaining on the other side (similar to the arrangement of pixels in any natural image). However, comparison of rearranged matrices might not be appropriate if the order of both matrices is different; for example, comparing Monday and Sunday OD matrices.

To this end, the study proposes an extension to MSSIM through the concept of geographical window as further discussed in the following section.

**Proposed geographical window based structural similarity index (GSSI)**

In the proposed approach, we first arrange the origins and destinations of OD matrix in order of the geographical identity, and subsequently define the windows for SSIM analysis that are consistent with the geographical boundaries. Here, the window size varies with the geographical boundaries considered in the rearranged OD matrix. This is different from the traditional SSIM application where the size of the window is fixed. The local window with a defined geographical boundary is termed as geographical window.

Let us explain the concept of geographical windows with the help of Figure 6 where each cell of the OD matrix represents a lower level OD pair (\(o_i\) to \(d_i\)). Here, the OD matrix is rearranged so that the lower level origins (rows) and destinations (columns) can be grouped into respective higher-level origin (\(O_k\)) and destination (\(D_l\)) zones. The number of higher zonal origins and destinations are \(K\) and \(L\), respectively. For instance, in Figure 6, \(o_1\) to \(o_i\) from \(O_1\) level is arranged together. The higher-level boundaries now define the geographical windows. The orange-shaded region represents a geographical window covering OD pairs from \(O_1\) to \(D_2\).

The STR and SSIM are computed over the geographical windows using Equations (2b) and (3), respectively. Similarly, the mean of STR and SSIM values computed over \(G\) geographical windows are referred as \(GSTR\) (Equation (6)) and \(GSSI\) (Equation (7)), respectively.

\[
GSTR(X, Y) = \frac{1}{G} \sum_{g=1}^{G} STR(x_g, y_g)
\]

\[
GSSI(X, Y) = \frac{1}{G} \sum_{g=1}^{G} SSIM(x_g, y_g)
\]

where \(X\) and \(Y\) represent two OD matrices to be compared; \(x_g\) and \(y_g\) represent the group of OD pairs within \(g^{th}\) local geographical window in both matrices. Other parameters are already explained in the earlier section. The values of both \(GSTR\) and \(GSSI\) lie between 0 and 1. Note that \(GSTR\) accounts only the structural component and \(GSSI\) compares overall structural similarity of OD matrices.

**Demonstration of geographical windows with an example from Brisbane city**

The study region that is considered for demonstration purpose is the BCC region (refer Figure 7). Several studies in the past have used Bluetooth in transport applications (Bhaskar & Chung, 2013; Chen & Bierlaire, 2015; Michau, Pustelnik, et al., 2017; Yang & Wu, 2018). The current study also uses Bluetooth travel information for the demonstration purpose. Figure 7 shows 845 Bluetooth MAC scanners (BMS) overlapped on SA3 zones of the BCC region. The trips observed as a sequence of BMSs are used to develop OD flows between scanner pairs (refer to Michau, Nantes, et al. (2017)) for more details about deriving Bluetooth-inferred trips). The OD flows at BMS level are further aggregated to SA3 level based on the concordance between BMS and SA3 zones. Thus, we have SA3 level B-OD matrices of 20 x 20 dimensions. Since they are aggregated trips, the B-OD matrices are dense (as shown for Monday in Figure 8a). It is to be noted that the Bluetooth trip ends do not represent actual trip ends of the Bluetooth equipped vehicles, and B-OD matrices are used only for the demonstration of proposed methodology.

Figure 8 demonstrates the application of SA4-based geographical windows for comparing SA3 (20 x 20) B-
OD matrices from March 7, 2016 (Figure 8a) and March 13, 2016 (Figure 8b). B-OD matrices are chosen randomly and belong to a typical week. The 7th March (Monday) was a typical working weekday and 13th March (Sunday) was not a part of any long weekend. These two days in the week are chosen because it is well known that weekend travel patterns are different from that of a weekday, and ideally, the comparison between the two should show dissimilarity. The SA4 zones used in designing geographical windows are Brisbane East, Brisbane North, Brisbane South, Brisbane West, and Brisbane Inner. For example, consider a geographical window of SA4 OD pair Brisbane East and Brisbane North. It consists of SA3 OD pairs e.g., 30101 to 30201, 30202, 30203, and 30204; 30103 to 30201, 30202, 30203, and 30204. These SA3 OD pairs are geographically correlated because they belong to the same SA4 origin (Brisbane East) and SA4 destination (Brisbane North). The size of local geographical window corresponding to this SA4 OD pair is $2 \times 4$ because Brisbane East and Brisbane North consist of 2 and 4 lower level (SA3) zones, respectively.

In the above example, the total number of geographical windows considered is equal to the number of higher order OD pairs, which is $5 \times 5 = 25$, and the GSSI for Monday-Sunday matrix comparison is 0.7231.

Advantages of GSSI over MSSIM: Brisbane case study examples

The advantages of the GSSI approach over traditional sliding window based MSSIM are discussed below in further detail with case study examples from the BCC region.

Structural comparison of local travel patterns

The concept of geographical window ensures geographical integrity and captures spatial correlation by computing statistics on all local (lower zonal level) OD pairs belonging to the same higher zonal level OD pair. Thus, it provides opportunities to compare the local travel demand distribution (travel patterns) between different suburbs of a region that a sliding local window is not capable of. From Figure 9, it can be

Figure 8. Splitting (a) Monday and (b) Sunday B-OD matrices into geographical (SA4) windows.
illustrated that Sunday travel patterns differ majorly for suburb pair—Brisbane South to Brisbane North. This is reflected from local SSIM value of 0.4653 (see Figure 9 (left) and bold value in Table 2). On the other hand, for another suburb pair—Brisbane South to Brisbane West—Sunday travel patterns are similar (if not exact) to that of Monday with a value of 0.8037 (see Figure 9 (right) and bold value in Table 2).

Analysis of local travel patterns during different day types can help policy makers to identify suburbs for better transport planning. For example, assuming that B-OD (car OD) matrices in Figure 8 are reliable, the travel demand distributed between Brisbane East and Brisbane West during Monday and Sunday are very similar with SSIM equal to 0.9517. Comparison of such local demand distributions with public transit patterns can help transport planners to identify gaps in the transit service. For instance, consider the suburb pair Brisbane East and Brisbane West. If SSIM (Monday, Sunday) for transit OD is far lower than 0.9517 (of car OD), then it implies that the transit service between the suburbs might need further improvement.

Comparison of window sizes

The size and shape of a geographical window is defined by the number of lower zonal OD pairs present within the respective higher zonal OD pair. Therefore, in this approach the local geographical window need not always be a square matrix and might not be of fixed dimensions. For instance, in Figure 8, the geographical window are of different sizes such as $2 \times 4$ for Brisbane East and Brisbane North, and $6 \times 4$ for Brisbane South and Brisbane West.

Although the dimensions of the window may vary, the GSSI values are found to capture structural distortions as effective as the small window size based SSIM does. To demonstrate this, we have compared a typical working Monday OD matrix (i.e., average of 45 working Monday OD matrices from 2016) with 40 B-OD matrices each from Saturdays (Figure 10(a)) and Sundays (Figure 10(b)), and with nearly 45 OD matrices from rest of the weekdays i.e., Tuesday, Wednesday, Thursday and Friday (refer Figure 10(c–f)). It is observed that GSSI plot closely align (same variation) with that of a $2 \times 2$ sliding window for weekends, and to that of a $3 \times 3$ sliding window for weekdays. Figure 10 demonstrate 12 different plots each (11 of which correspond to sliding windows of sizes ranging from $2 \times 2$ to $20 \times 20$ (dotted lines) and geographical window (continuous line)). For each plot, x-axis corresponds to different daily OD matrices and y-axis reflects MSSIM/GSSI values.

Ability to capture subtle structural distortions

Here, we compare GSSI and MSSIM with a real case study example. For the study area shown in Figure 7, we consider B-OD matrices from the days $7_{th}$–$13_{th}$ Mar (entire week), $10_{th}$ Aug (Queensland Ekka festival), $25_{th}$ Dec (Christmas Sunday), and $27_{th}$ Mar (Easter Sunday) in 2016, and compare each daily OD matrix with the reference Monday OD matrix (i.e.,

| MONDAY | Brisbane North |
|--------|---------------|
| Origin | 30201 | 30202 | 30203 | 30204 |
| Dest   | 30201 | 26  | 54   | 206  | 122  |
|        | 30202 | 74  | 178  | 312  | 93   |
|        | 30203 | 42  | 54   | 195  | 85   |
|        | 30204 | 55  | 104  | 238  | 76   |
|        | 30205 | 32  | 40   | 219  | 65   |
|        | 30206 | 11  | 25   | 100  | 36   |

| SUNDAY | Brisbane North |
|--------|---------------|
| Origin | 30201 | 15  | 32   | 50   | 63   |
| Dest   | 30202 | 46  | 163  | 163  | 79   |
|        | 30203 | 11  | 33   | 56   | 53   |
|        | 30204 | 6   | 12   | 76   | 35   |
|        | 30205 | 8   | 24   | 43   | 18   |
|        | 30206 | 6   | 14   | 36   | 24   |

| MONDAY | Brisbane West |
|--------|---------------|
| Origin | 30401 | 30402 | 30403 | 30404 |
| Dest   | 30401 | 23  | 371  | 117  | 48   |
|        | 30402 | 135 | 65   | 594  | 228  |
|        | 30403 | 51  | 37   | 231  | 106  |
|        | 30404 | 71  | 25   | 443  | 163  |
|        | 30405 | 184 | 9    | 505  | 60   |
|        | 30406 | 38  | 8    | 90   | 26   |

| SUNDAY | Brisbane West |
|--------|---------------|
| Origin | 30401 | 16  | 289  | 82   | 45   |
| Dest   | 30402 | 86  | 26   | 473  | 218  |
|        | 30403 | 44  | 25   | 156  | 75   |
|        | 30404 | 54  | 15   | 193  | 34   |
|        | 30405 | 102 | 7    | 263  | 24   |
|        | 30406 | 31  | 4    | 75   | 21   |

Figure 9. Insights into local travel patterns using geographical local window: (left) Brisbane South to Brisbane North and (right) Brisbane South to Brisbane West.
Mar 7, 2016) using GSSI/MSSIM and GSTR/STR as shown in Figure 11 and Figure 12, respectively. Note that in this application, 25 geographical windows (as shown in Figure 8) are used for GSSI/GSTR, and only a single window (of size equal to B-OD matrix) is used for SSIM/STR because the order of OD matrices are different for different day types.

The similarities in results of GSSI and SSIM comparisons are as follows:

1. The OD matrices during the weekdays are structurally more similar to the Monday OD matrix than rest of the day types. This is reflected in higher GSSI/SSIM (≥0.9899/≥0.9950) and GSTR/STR (≥0.9938/≥0.9995) values for the weekdays in Figure 11. A closer observation of GSSI/SSIM values in Figure 11 revealed that travel patterns during Tuesday and Wednesday (red rectangle) were slightly different from Thursday.

### Table 2. Local SSIM values over geographical windows: Monday vs Sunday.

|          | Brisbane East | Brisbane North | Brisbane South | Brisbane West | Brisbane Inner |
|----------|---------------|----------------|----------------|---------------|---------------|
| Brisbane East | 0.8319        | 0.2437         | 0.7650         | 0.9517        | 0.7755        |
| Brisbane North | 0.3311        | 0.7353         | 0.4034         | 0.7378        | 0.6299        |
| Brisbane South | 0.7771        | 0.4653         | 0.8062         | 0.8884        | 0.8165        |
| Brisbane West | 0.8340        | 0.7754         | 0.7562         | 0.8884        | 0.8165        |
| Brisbane Inner | 0.7711        | 0.6265         | 0.8257         | 0.8385        | 0.8750        |

![Figure 10. GSSI vs MSSIM for weekends (a,b) and weekdays (c–f).](image-url)
2. Both GSSI and SSIM identified that travel patterns during weekends and rest of public holidays are very different from that of weekdays. Among the non-weekdays, GSSI/SSIM values for Saturday-Monday showed higher proximity because the percentage of work-related trips occurring during Saturdays is higher as compared to Sundays and rest of public holidays (Bhat & Misra, 1999).

Despite the above similarities, we have found that SSIM is less sensitive to the structural distortions among the non-weekdays, especially Ekka, Saturday and Sunday, in Figure 13. We can observe that GSTR values (green oval) could differentiate more than STR does (orange oval) for these day types. To understand these differences, comparison of local travel patterns between different sections of the network is necessary which is only possible with the GSSI technique. Refer to Figure 13 that helps to visualize these local structural differences among these day types. The rows and column headers in Figure 13 are SA3 origins and destinations. The Ekka festival generally happens at Brisbane show grounds in Bowen Hills (Brisbane Inner). Due to this, the trips attracted to it and its neighboring suburbs such as Newstead-Bowen Hills, and Newmarket in the North; Ashgrove, and Milton in the West; and CBD, Valley, and South Bank in the Inner city have different demand distributions on the Ekka public holiday as compared to usual Saturday and Sunday.

Figure 11. Comparison among daily OD matrices using GSSI and MSSIM.

Figure 12. Comparison among daily OD matrices using GSTR and STR values.
Computational efficiency

Figure 14 present the computation time required to compare 415 B-OD matrices (from June 2015 to Dec 2016) with Monday OD matrix (Figure 8a) using different window size for MSSIM (hollow bars in Figure 14) and GSSI (solid bar in Figure 14). The testing is done on a Dell computer with Intel(R) Core(TM) i7-4770 CPU, 16GB RAM (3.40 GHz). It is observed that the computational time of GSSI (3.92 seconds) is much lower than that of the MSSIM with lower order slider window size (2/C2/2 sliding window based MSSIM required 39.5 seconds). The results are as expected, the comparison of 20/C2/20-dimension OD matrices via 2/C2/2 sliding window needs to be performed (20/C0/2 + 1)/C2/C2 = 361 times. On the other hand, GSSI is an average value of all local SSI values computed 25 times (due to 25 geographical windows). In general, if we have G geographical windows and m x n sliding windows to compare M x N OD matrices then, the GSSI approach is (M- m + 1)* (N-n + 1)/G times computationally more efficient than m x n sliding window based MSSIM. From this, it can be concluded that GSSI is computationally more effective than MSSIM (based on sliding window of smaller sizes).

Sparse OD matrices and geographical windows from different zonal levels

The previous section demonstrated GSSI’s application on dense OD matrices (SA2) and is based on geographical window from a particular higher zonal level (SA4). However, in most transport modeling applications we encounter OD matrices that are mostly sparse with zones much smaller than SA3. In such situations, it would be interesting to see how the GSSI approach performs for different types of zonal windows, and if it can compare sparse matrices better than the traditional MSSIM. Since BOD matrices constitute only partial observed flows, sufficient sparsity can be observed at SA2 level. Thus, the sparse OD matrices we considered for this analysis are B-OD matrices at SA-2 level (99 x 99). A typical Monday (sparsity around 20%), is compared with typical Saturday and Sunday (sparsity around 16% and 21%) matrices. The different geographical window levels used are SA3 and SA4 as shown in Figure 15. The MSSIM considers only single window because the order of re-arranged OD matrices is found to be different during the three days. Refer to Figure 16 for the visualization of sparsity in SA3 (30101 to 30401, 30402, 30403, and 30404) and SA4 (301 to 304) windows.

Table 3 demonstrates the GSSI comparisons applied using SA3 and SA4 windows as against the MSSIM values on SA2 OD matrices. We can see that MSSIM resulted in the highest similarity value but GSSI using SA4 and SA3 windows resulted in lower and lowest values of similarity, respectively. For example, Monday-Saturday comparison values are 0.9190, 0.8185, and 0.7202 by MSSIM, GSSI_SA4 and GSSI_SA3, respectively. There are hardly any structural differences between Sunday and Saturday travel patterns when compared to Monday. This is mainly due to the reason that the sensitivity of the SSIM formulation toward structural distortions reduces as the window size increases. Thus, we can see that GSSI is able to recognize the structural differences better than MSSIM despite sparsity in OD matrices.

Figure 13. Comparison of local travel patterns during Ekka, Saturday and Sunday.

|         | East | North | South | West | Inner city |
|---------|------|-------|-------|------|------------|
| East    | 1.0000 | 0.9973 | 0.9887 | 0.9991 | 0.9888 |
| North   | 0.7950 | 0.9560 | 0.8002 | 0.9741 | 0.9844 |
| South   | 0.9902 | 0.8039 | 0.9708 | 0.9944 | 0.9928 |
| West    | 0.7972 | 0.9826 | 0.9900 | 0.9983 | 0.9972 |
| Inner city | 0.9949 | 0.9841 | 0.9943 | 0.9977 | 0.9996 |

|         | East | North | South | West | Inner city |
|---------|------|-------|-------|------|------------|
| East    | 1.0000 | 0.9973 | 0.9887 | 0.9991 | 0.9888 |
| North   | 0.7950 | 0.9560 | 0.8002 | 0.9741 | 0.9844 |
| South   | 0.9902 | 0.8039 | 0.9708 | 0.9944 | 0.9928 |
| West    | 0.7972 | 0.9826 | 0.9900 | 0.9983 | 0.9972 |
| Inner city | 0.9949 | 0.9841 | 0.9943 | 0.9977 | 0.9996 |

Sunday (13th Mar, 2016) vs Mon (7th Mar 2016)
Regarding the selection of geographical windows (i.e., consideration of either SA4 or SA3 as shown in Table 3), the smaller the size of window, the better is the ability of SSIM to compute local structural differences. Thus, we recommend using geographical windows from immediately higher level in the zonal hierarchy; for instance, use SA4 windows for SA3 OD matrices, SA3 for SA2 or SA2 for SA1, etc.
Sensitivity analysis

Although the mathematical formulation of GSSI is same as that of MSSIM, they vary in their similarity values due to different choice of local windows. Also, we have demonstrated in the previous section that GSSI is able to capture better structural distortions (as compared to MSSIM) among daily OD matrices. This section further tests the robustness of both GSSI and GSTR through sensitivity analysis for different conditions of OD matrix combinations. A statistical measure is considered to be robust if it is sensitive to changes in both “OD flows” and the “OD structure.” The study site, data and the design of experimental set up are briefly discussed below.

1. Study site and data: The BCC region is the study area as shown in Figure 7. The reference OD matrix, \(X\) is the Bluetooth based OD matrix \((20 \times 20)\) observed on Monday, March 7, 2016 (refer to Figure 8(a)). The query OD matrices are developed specific to the experiments and are obtained by perturbing the reference OD matrix.

2. Experiments: The experimental set up for the sensitivity analysis of GSSI and GSTR is shown in Table 4.

We design two sets of experiments for the sensitivity analysis. Each experiment tests the sensitivity of both GSSI (\(X, Y\)) and GSTR (\(X, Y\)) to capture the structural similarity and the structure of OD matrices, respectively. The experiments are as follows:

a. Experiment based on uniform scaling effect: Here, the query matrices have the same structure as that of reference OD matrix while the magnitude of OD flows vary. If the uniform scale factor is one, then both OD matrices are exactly similar. Thus, abovementioned situation-1 and situation-2 are tested here (refer Table 4).

b. Experiment based on random scaling effect: Here, the structure and OD flows vary between query and reference OD matrices. Thus, above-mentioned situation-3 is tested here (refer Table 4).

In the following sections, we first discuss the design of experimental set up for both experiments, and then present the results of sensitivity analysis.

### Experimental set up for the sensitivity analysis of GSSI and GSTR

#### Criteria for uniform scaling effects

Here, sensitivity of GSSI and GSTR are tested for different uniform scaling percentages. The reference OD matrix, \(X\) is compared with \(Y_i\) where \(Y_i = \varphi * X\), and \(\varphi\) is chosen from \([0.1, 0.2, 0.3 \ldots 1.9, 2.0]\). GSSI is robust metric if:

1. The GSTR is zero for any value of \(\varphi\).
2. The GSSI value is zero for \(\varphi = 1\) and should decrease as \(\varphi\) deviated from unity.

#### Criteria for random scaling effects

Here, sensitivity of GSSI and GSTR are tested for four different cases of random scaling percentages i.e., \(\psi = [5\%, 10\%, 15\%, 20\%]\) over three types of demand scenarios. These demand scenarios are generally encountered in traffic demand modeling (refer Djukic et al. (2015)) and are as follows:

1. Outdated surveys (low demand),
2. The best historical estimates (medium demand), and
3. Congested traffic conditions (high demand).

Note that the scenarios are named as low (l), medium (m), and high (h) in reference to the total daily demand on the network, and do not refer to the demands of individual OD pairs. In each case of the demand scenario, reference OD (\(X\)) is compared with 100 replications of query ODs (\(Y\)). The details of the demand scenarios are as follows:

- **Low demand scenario:** Here, both GSSI and GSTR compare \(X\) and \(Y_i^\psi\) where \(Y_i^\psi = X^{(0.60 + \psi \cdot \text{rand}[0, 1])}\) and \(i \in [1, 100]\). For instance, if \(\psi = 20\%\), then \(Y_i^\psi\) ranges between 60% and 80% of \(X\), and similarly for other values of \(\psi\).
High demand scenario: Here, both GSSI and GSTR compare $X$ and $Y_{i,\psi}^{m}$ where $Y_{i,\psi}^{m} = X \ast (0.80 + \psi \ast \text{rand}[0,1])$ and $i \in [1,100]$. For instance, if $\psi = 20\%$, then $Y_{i,\psi}^{m}$ ranges between 80% and 100% of $X$, and similarly for other values of $\psi$.

Medium demand scenario: Here, both GSSI and GSTR compare $X$ and $Y_{i,\psi}^{h}$ where $Y_{i,\psi}^{h} = X \ast (1.05 + \psi \ast \text{rand}[0,1])$ and $i \in [1,100]$. The OD matrices for the high demand scenario represent demand during congested periods. Say, high daily demand can be witnessed during major events such as Commonwealth games. For instance, if $\psi = 20\%$, then $Y_{i,\psi}^{h}$ ranges between 105% and 125% of $X$ and, similarly for other values of $\psi$.

High demand scenario: Here, both GSSI and GSTR compare $X$ and $Y_{i,\psi}^{h}$ where $Y_{i,\psi}^{h} = X \ast (1.05 + \psi \ast \text{rand}[0,1])$ and $i \in [1,100]$. The OD matrices for the high demand scenario represent demand during congested periods. Say, high daily demand can be witnessed during major events such as Commonwealth games. For instance, if $\psi = 20\%$, then $Y_{i,\psi}^{h}$ ranges between 105% and 125% of $X$ and, similarly for other values of $\psi$.

The conditions for both GSSI and GSTR to be robust toward random effects are:

1. Both should reflect the random structural differences that exist between the OD matrices; that is, the similarity values should decrease/increase with increase/decrease in the magnitude of random scaling effects for all three demand scenarios.
2. The mean values of GSSI and GSTR for four different $\psi$ s in each demand scenario should be statistically different. To achieve this, we perform analysis of variance (ANOVA) test for both GSSI and GSTR for all three demand scenarios. The conclusion is drawn by comparing the F-statistic value with its critical value i.e., 3.54 at level of significance $\alpha = 0.05$ or 95% confidence. The two hypotheses are as follows:
   - Null hypothesis ($H_0$): For each demand scenario, there is no difference among the means of all four $\psi$ for both GSSI and GSTR.
   - Alternate hypothesis ($H_a$): For each demand scenario, means of GSSI and means of GSTR from four $\psi$ are different.

Results of sensitivity analysis

Results of uniform scaling effects

The results of uniform scaling for both GSSI and GSTR are shown in Figure 17. The plot illustrates that GSSI satisfies the conditions specified in the previous section i.e., GSSI values increased from 0.04 to 1 for $0.1 \leq \varphi < 1$ and decreased from 1 to 0.64 for $1 < \varphi \leq 2$. Similarly, GSTR remained unaffected (i.e., equal to 1) for both scaling-up and scaling-down cases (as described in the previous section). Thus, it is proved that both GSSI and GSTR are robust toward uniform scaling effects.

Results of random scaling effects

The plot shown in Figure 18(a) demonstrates decrease in the structural similarity as the magnitude of random fluctuations increase. For instance, GSSI for low demand scenario are 0.7759, 0.7675, 0.7489, and 0.7307 for $\psi = 5\%$, 10\%, 15\%, 20\%, respectively. A similar decreasing trend is observed for other demand scenarios. The structure component of GSSI, that is, GSTR also illustrates similar results in Figure 18(b).

Both GSSI and GSTR are further compared using F-statistic to check if there is any difference among the means. The results of ANOVA shown in Table 5 demonstrate that F values are far greater than 3.24 at 0.05 level of significance. The p value for each F value is close to zero as shown within the brackets in Table 5. Therefore, the null hypothesis is rejected implying that the means of all four cases in each demand scenario are statistically different among each other for both GSSI and GSTR. Thus, the results prove that both GSSI and GSTR are robust toward random scaling effects.

Therefore, sensitivity analysis conducted using different OD matrix combinations in different situations should give enough confidence on the robustness of the proposed metric.

Discussion

The results of sensitivity analysis proved that the proposed GSSI approach is sensitive to both uniform and random demand fluctuations. Sensitivity toward uniform scaling effects prove that GSSI can be used to compare OD matrices of similar travel patterns (structures) such as comparing OD matrices of typical Monday and typical Tuesday. The experiments related to random scaling effects suggest that GSSI is able to capture random fluctuations at different demand scales (low, medium and high). The query matrices in high demand scenario are designed such a way that...
they are closer to reference matrix as compared to other scenarios. The values of GSSI can reflect proximity with higher GSSI (and GSTR) values in high demand scenario as compared to any other scenario. Sensitivity of GSSI toward random demand fluctuations demonstrate its ability in various applications such as (a) comparing target and estimated demands in OD estimation problem, (b) benchmarking OD estimation methods, (c) checking the convergence/divergence of OD optimization algorithms, and (d) comparing the travel patterns based on type of the day (Behara et al., 2018) or modes (Hussain et al., 2019). The ANOVA test results further proved that the mean values of GSSI and GSTR are statistically different for all four cases in each demand scenario.

The study uses SA zones to demonstrate the concept. The SA zones are mainly defined considering the land use and population density. Travel behavior and demand is influenced by land use patterns and population density. Referring to the ATAP (2016) guidelines, “Transport zones should ideally contain homogeneous land use (for example, solely residential, industrial or commercial use or parking lots) and they should not cross significant barriers to travel (such as rivers, freeways and rail lines), and should have reasonably homogeneous access to the modeled transport systems. In this context, transport zones should match, as far as practically possible, Australian Bureau of Statistics (ABS) Statistical Area boundaries as defined by the Australian Statistical Geography Standard (ASGS). In general practice, transport zones

Table 5. Results of ANOVA for both GSSI and GSTR across all three demand scenarios.

|       | F (p) for Low | F (p) for Medium | F (p) for High |
|-------|---------------|------------------|---------------|
| GSSI  | 84.49(~0)     | 305.56(~0)       | 1536.8(~0)    |
| GSTR  | 734.24(~0)    | 1344.75(~0)      | 1448.7(~0)    |

Figure 18. Results of random scaling effects for (a) GSSI and (b) its structure component, GSTR.
will be aggregations of the Statistical Area Level 1 (SA1) or mesh block boundaries.”

The Transport analysis zones (TAZ) used by Queensland Department of Transport and Main Roads (TMR) are well aligned with majority of SA zones and the household travel survey is designed at SA1 level; thus SA1 zones are linked to the Brisbane TAZs in the transport models (Weston, 2019). Although we considered SA zones as a proxy for TAZ, the concept of geographical windows still holds good should we have different levels of TAZs; for instance, consider TAZs used by the Florida Department of Transportation Systems Planning Office (2007). It uses three types of TAZs for three different purposes: large-sized zone for state-wide planning, medium-sized zone for arterial planning, and small-sized zone for corridor analysis. To apply the GSSI’s rationale, medium sized TAZs (with approx. 4000 persons per zone) can be used to define geographical windows for small-sized TAZs (with approx. 1200 to 3000 persons per zone).

Conclusion

A holistic comparison of OD matrices should include both deviations of individual OD flows and the structural (dis)similarity between the OD matrices. The traditional statistical measures compare OD matrices using deviations of individual OD flows, and fail to account for OD matrix structural differences. Very limited studies in the past focused on developing such structural (dis)similarity measures (refer to Table A1 in the Appendix section for conceptual comparison of these measures). These metrics are popular in other disciplines and are introduced in transport applications. One such metric is MSSIM that computes statistics on groups of OD pairs defined by local windows. However, comparison of OD matrices using MSSIM can result in different values based on the choice of local window size. In literature, no clear consensus has been reported on the level of acceptability of sliding window size and the resulting MSSIM values. While a few researchers proposed variants of MSSIM, they needed further investigation. We propose an alternative way of addressing the limitations of MSSIM with respect to local window size and geographical adjacency through the concept of geographical window defined by geographical boundaries. We term this extension to MSSIM as geographical window based structural similarity index (GSSI). The findings of this study are (a) GSSI technique can capture local travel patterns better than a sliding window based MSSIM, (b) it has a physical meaning to local statistics, (c) it is computationally more effective, (d) it could compare large-scale sparse OD matrices, and (e) it can identify subtle structural differences among OD matrices. The study suggests that it is better to choose geographical windows at a level immediately higher to the zonal level of OD matrices that are to be compared. A thorough sensitivity analysis proves that GSSI is a robust statistical measure.

As a part of future research directions, we plan to study the optimality of stability constants in the SSIM formulation and weights specific to SSIM components (this is important while comparing sparse and dense OD matrices). Although, we have tried to address the issue related to selection of window size, the level of acceptability of GSSI still depends on different transport applications, and this requires further discussions with practitioners and scientific community in the transport domain.

The proposed formulation requires knowledge of the network in terms of its statistical boundaries and its hierarchy. Such information is generally prepared and maintained by the Bureau of Statistics and should be available to the transport analyst. In the current application, the hierarchical structure of the zones provided by the Australia Bureau of Statistics is utilized to define the dynamic window size. For future study, it is recommended to establish guidelines for the definition of the geographical boundaries. It should consider zonal socio-economic characteristics, land use pattern, and spatial proximity.

Acknowledgments

The authors are thankful to the Brisbane City Council (BCC), the Queensland Department of Transport and Main Roads (TMR), and the Queensland University of Technology for supporting the research. The Bluetooth data used in this study is from BCC. The conclusions of this paper reflect understandings of the authors, who are responsible for the accuracy of the research findings.

Disclosure statement

The authors declare no conflict of interest.

ORCID

Krishna N. S. Behara http://orcid.org/0000-0002-9064-367X
Ashish Bhaskar http://orcid.org/0000-0001-9679-5706
Edward Chung http://orcid.org/0000-0001-6969-7764
References

Antoniou, C., Barceló, J., Breen, M., Bullejos, M., Casas, J., Cipriani, E., Ciuffo, B., Djukic, T., Hoogendoorn, S., Marzano, V., Montero, L., Nigo, M., Perarnau, J., Punzo, V., Toledo, T., & van Lint, H. (2016). Towards a generic benchmarking platform for origin–destination flows estimation/updating algorithms: Design, demonstration and validation. Transportation Research Part C: Emerging Technologies, 66, 79–98. https://doi.org/10.1016/j.trc.2015.08.009

Antoniou, C., Ben-Akiva, M., & Koutsopoulos, H. (2004). Incorporating automated vehicle identification data into origin–destination estimation. Transportation Research Record: Journal of the Transportation Research Board, 1882(1), 37–44. https://doi.org/10.3141/1882-05

ASGS. (2017). Australian Statistical Geography Standard (ASGS). Australian Bureau of Statistics (ABS).

ATAP (2016). Australian Transport Assessment and Planning Guidelines. Commonwealth Department of Infrastructure and Regional Development.

Barceló, J., Montero, L., Bullejos, M., Linares, M., & Serch, O. (2013). 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: Journal of the Transportation Research Board, 2344(1), 31–39. https://doi.org/10.3141/2344-04

Barceló, J., Montero, L., Bullejos, M., Serch, O., & Carmona, C. (2013). A kalman filter approach for exploiting bluetooth traffic data when estimating time-dependent OD matrices. Journal of Intelligent Transportation Systems, 17(2), 123–141. https://doi.org/10.1080/15472450.2013.764793

Behara, K. N., Bhaskar, A., & Chung, E. (2018). Classification of typical Bluetooth OD matrices based on structural similarity of travel patterns: Case study on Brisbane city [Paper presentation]. Transportation Research Board (TRB) 97th Annual Meeting, Washington, DC.

Behara, K. N., Bhaskar, A., & Chung, E. (2020a). A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. Transportation Research Part C: Emerging Technologies, 111, 513–530. https://doi.org/10.1016/j.trc.2020.01.005

Behara, K. N., Bhaskar, A., & Chung, E. (2020b). A novel methodology to assimilate sub-path flows in bi-level OD matrix estimation process. IEEE Transactions on Intelligent Transportation Systems. 1–11. https://doi.org/10.1109/TITS.2020.2998475

Bhaskar, A., & Chung, E. (2013). Fundamental understanding on the use of Bluetooth scanner as a complementary transport data. Transportation Research Part C: Emerging Technologies, 37, 42–72. https://doi.org/10.1016/j.trc.2013.09.013

Bhat, C. R., & Misra, R. (1999). Discreetionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends. Transportation, 26(2), 193–229. https://doi.org/10.1023/A:1005192230485

Bierlaire, M. (2002). The total demand scale: A new measure of quality for static and dynamic origin–destination trip tables. Transportation Research Part B: Methodological, 36(9), 837–850. https://doi.org/10.1016/S0191-2615(01)00036-4

Brooks, A. C., Zhao, X., & Pappas, T. N. (2008). Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions. IEEE Transactions on Image Processing, 17(8), 1261–1273. https://doi.org/10.1109/TIP.2008.926161

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. Transportation Research Part B: Methodological, 18(4-5), 289–299. https://doi.org/10.1016/0191-2615(84)90012-2

Chen, J., & Bierlaire, M. (2015). Probabilistic multimodal map matching with rich smartphone data. Journal of Intelligent Transportation Systems, 19(2), 134–148. https://doi.org/10.1080/15472450.2013.764796

Day-Pollard, T., & Van Vuren, T. (2015). When are origin-destination matrices similar enough? [Paper presentation]. 94th Annual TRB Meeting, Washington, DC.

Djukic, T. (2014). Dynamic OD demand estimation and prediction for dynamic traffic management [PhD thesis]. Delft University of Technology, TU Delft.

Djukic, T., Barceló Bugeda, J., Bullejos, M., Montero Mercadé, L., Cipriani, E., van Lint, H., & Hoogendoorn, S. (2015). Advanced traffic data for dynamic OD demand estimation: The state of the art and benchmark study [Paper presentation]. TRB 94th Annual Meeting Compendium of Papers, Washington, DC.

Djukic, T., Hoogendoorn, S., & Van Lint, H. (2013). Reliability assessment of dynamic OD estimation methods based on structural similarity index [Paper presentation]. Transportation Research Board 92nd Annual Meeting, Washington, DC.

Doblas, J., & Benitez, F. G. (2005). An approach to estimating and updating origin–destination matrices based upon traffic counts preserving the prior structure of a survey matrix. Transportation Research Part B: Methodological, 39(7), 565–591. https://doi.org/10.1016/j.trb.2004.06.006

Florida Department of Transportation Systems Planning Office. (2007). A recommended approach to delineating traffic analysis zones in Florida [Paper presentation]. Florida Department of Transportation Systems Planning Office, Tallahassee, FL.

Frederix, R., Viti, F., Himpe, W. W., & Tampere, C. M. (2014). Dynamic origin–destination matrix estimation on large-scale congested networks using a hierarchical decomposition scheme. Journal of Intelligent Transportation Systems, 18(1), 51–66. https://doi.org/10.1080/15472450.2013.773249

Hellinga, B., & Aerde, M. V. (1994). A statistical analysis of the reliability of using RGS vehicle probes as estimators of dynamic OD departure rates. I V H S Journal, 2(1), 21–44. https://doi.org/10.1080/10248079408903813

Hollander, Y., & Liu, R. (2008). The principles of calibrating traffic microsimulation models. Transportation, 35(3), 347–362. https://doi.org/10.1007/s11116-007-9156-2

Hussain, E., Behara, K., Bhaskar, A., & Chung, E. (2019). A framework to quantify the differences between multi-modal travel demand: A case study on Brisbane network using
Bluetooth and smartcard data. Australasian Transport Research Forum (ATRF), 41st, 2019, Canberra, ACT, Australia.

Kim, S.-J., Kim, W., & Rilett, L. (2005). Calibration of microsimulation models using nonparametric statistical techniques. Transportation Research Record: Journal of the Transportation Research Board, 1935(1), 111–119. https://doi.org/10.1177/0361198105193500113

Michau, G., Nantes, A., Bhaskar, A., Chung, E., Abry, P., & Borgnat, P. (2017). Bluetooth data in an urban context: Retrieving vehicle trajectories. IEEE Transactions on Intelligent Transportation Systems, 18(9), 2377–2386. https://doi.org/10.1109/TITS.2016.2642304

Michau, G., Pustelnik, N., Borgnat, P., Abry, P., Nantes, A., Bhaskar, A., & Chung, E. (2017). A primal-dual algorithm for link dependent origin destination matrix estimation. IEEE Transactions on Signal and Information Processing over Networks, 3(1), 104–113. https://doi.org/10.1109/TSIPN.2016.2623094

Nigro, M., Cipriani, E., & del Giudice, A. (2018). Exploiting floating car data for time-dependent origin–destination matrices estimation. Journal of Intelligent Transportation Systems, 22(2), 159–174. https://doi.org/10.1080/15472450.2017.1421462

Panaretos, V. M., & Zemel, Y. (2019). Statistical aspects of Wasserstein distances. Annual Review of Statistics and Its Application, 6(1), 405–431. https://doi.org/10.1146/annurev-statistics-030718-104938

Pollard, T., Taylor, N., van Vuren, T., & MacDonald, M. (2013). Comparing the quality of OD matrices in time and between data sources [Paper presentation]. Proceedings of the European Transport Conference, Frankfurt, Germany.

Ros-Roca, X., Montero, L., Schneck, A., & Barceló, J. (2018). Investigating the performance of SPSA in simulation-optimization approaches to transportation problems. Transportation Research Procedia, 34, 83–90. https://doi.org/10.1016/j.trpro.2018.11.017

Ruiz de Villa, A., Casas, J., & Breen, M. (2014). OD matrix structural similarity: Wasserstein metric [Paper presentation]. Transportation Research Board 93rd Annual Meeting, Washington, DC.

Tavassoli, A., Alser, A., Hickman, M., & Mesbah, M. (2016). How close the models are to the reality? Comparison of transit origin-destination estimates with automatic fare collection data. Australasian Transport Research Forum (ATRF), 38th, 2016, Melbourne, Victoria, Australia.

Van Vuren, T., & Day-Pollard, T. (2015, May 20). 256 shades of grey – Comparing OD matrices using image quality assessment techniques. STAR Conference.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Weston, M. (2019). What goes on inside a zone? The secrets of intrazonal modelling [Paper presentation]. AITPM National Conference, Brisbane.

Yang, S., & Wu, Y.-J. (2018). Travel mode identification using Bluetooth technology. Journal of Intelligent Transportation Systems, 22(5), 407–421. https://doi.org/10.1080/15472450.2017.1384698
## Appendix

**Table A1. Comparison among different OD structural proximity measures.**

| Motivation behind | MSSIM | 4D-MSSIM | Wasserstein | Levenshtein | GSSI |
|-------------------|-------|----------|-------------|-------------|------|
| **Mathematical formulation** | MSSIM is popularly used in the comparison of images represented by pixels. Considering OD pairs analogous to these pixels, Djukic et al. (2013) applied MSSIM for comparison of OD matrices. Same as MSSIM but customized to OD matrices. Wasserstein is extensively used to solve mass transport or optimal transport problems. It is the minimal cost required to reconfigure the probability mass of one distribution to the other. Wasserstein in the OD matrix context attempts to compute minimal cost to reconfigure the demand flows of one OD to match the distribution of flows in another. Levenshtein is widely used in computational linguistics. It quantifies structural deviations between any pair of strings (words). Motivated by this approach Behara et al. (2020a) proposed sequence or preference of destinations from each origin analogous to the sequence of strings. Same as MSSIM but customized to OD matrices. |
| **Mathematical properties** | MSSIM is expressed in a single mathematical equation. It includes three components related to comparison of mean, standard deviation and covariance. No optimization is involved. The main formulation is same as MSSIM. Additionally, it uses geographical co-ordinates of zones to calculate spatial proximity between an OD pair in terms of Euclidian distance. No optimization is involved. Same as MSSIM. The spatial structure is accounted by exploiting network topology i.e., location of each centroid in the network. No spatial or network related attributes are considered. Same as MSSIM. |
| **Additional trip/network related attributes** | Satisfies non-negative, symmetric, and triangular inequality properties. Satisfies non-negative, symmetric, and triangular inequality properties. Satisfies non-negative, symmetric, and triangular inequality properties. Same as MSSIM. No additional trip/network related attributes are accounted. The spatial proximity between OD pairs are accounted through Euclidian distance. The spatial structure is accounted by exploiting network topology i.e., location of each centroid in the network. The geographical integrity and spatial proximity among the OD pairs are accounted through geographical windows. |
| **Comparison of OD flows and OD structure** | The mass components; that is, OD flows are compared using mean and standard deviation. The OD structure is compared using covariance. All the three components are defined explicitly. The mass components; that is, OD flows are compared using mean and standard deviation. The OD structure is compared using covariance. All the three components are defined explicitly. Differences in both OD flows and OD structures are implicit in the formulation. No additional trip/network related attributes are accounted. The spatial proximity between OD pairs are accounted through Euclidian distance. Differences in both OD flows and OD structures are implicit in the formulation. |
| **Computational efficiency** | Computational efficiency is expressed in a single mathematical equation. It includes three components related to comparison of mean, standard deviation and covariance. No optimization is involved. Computational efficiency is expressed in a single mathematical equation. It includes three components related to comparison of mean, standard deviation and covariance. No optimization is involved. Computational efficiency is expressed in a single mathematical equation. It includes three components related to comparison of mean, standard deviation and covariance. No optimization is involved. Computational efficiency is expressed in a single mathematical equation. It includes three components related to comparison of mean, standard deviation and covariance. No optimization is involved. |

(Continued)
| Feature                                      | MSSIM                                  | 4D-MSSIM                             | Wasserstein                          | Levenshtein                           | GSSI                                      |
|----------------------------------------------|----------------------------------------|--------------------------------------|--------------------------------------|----------------------------------------|-------------------------------------------|
| **Unit of measurement**                      | The similarity comparison ranges      | Od pairs, then it involves additional | The distance comparison ranges      | sliding windows used in MSSIM.          |                                           |
|                                              | between −1 and 1.                      | computational cost.                  | between −1 and 1.                    |                                       |                                           |
|                                              | Thus no measurement units.             |                                     | Thus no measurement units.           |                                       |                                           |
| **Level of detail**                          | SSIM facilitates local demand         | 4D-MSSIM can compare groups of OD    | The Wasserstein metric yields overall| Same as MSSIM.                         |                                           |
|                                              | comparison using sliding windows.      | pairs that are spatially correlated. | distance and does not facilitate     |                                       |                                           |
|                                              | However, the comparison has no        | Thus, it enables comparison of local  | local comparison.                    |                                       |                                           |
|                                              | physical interpretation. On the other  | structural differences.              |                                     |                                       |                                           |
|                                              | hand, use of one window does not help |                                     |                                     |                                       |                                           |
|                                              | in explaining any local structural     |                                     |                                     |                                       |                                           |
|                                              | deviations.                            |                                     |                                     |                                       |                                           |
| **Limitations**                              | First, MSSIM values change as the size | First, 4D-MSSIM is sensitive to the  | First, the metric is computationally  | First, the metric is computationally   | The stability constants are network      |
|                                              | of local sliding window changes.      | threshold value used to define the   | expensive. Second, the use of        | intensive as compared to                | specific and need further exploration.   |
|                                              | Second, the method is sensitive to the| contributions of OD pairs.           | average travel time might be          | MSSIM/GSSI/4D-MSSIM. However, it is    | GSSI does not consider other network     |
|                                              | arrangement of OD matrix (unless a    | Second, the use of Euclidian distance| inappropriate in all situations.     | better than Wasserstein. It does not   | properties such as travel cost between    |
|                                              | single window is used). Third, the     | might not be a better measure for    | For instance, travel time could      | consider other network attributes       | OD pairs. The window is only based on    |
|                                              | order of OD matrix might not always    | clustering nearby OD pairs.          | differ during a Monday and a         | such as travel time, and geographical    |
|                                              | result in spatial adjacent zones/OD    | For instance, two OD pairs separated | Sunday between an OD pair. Use of    | coordinates.                           | boundaries but does not consider other    |
|                                              | pairs. Fourth, the stability constants | by a natural barrier (say river) can  | average travel time might not be      |                                       | information such as socio-economic       |
|                                              | used in its formulation are specific  | be located quite distant away as      | justified while comparing matrices    |                                       | characteristics, and land use.           |
|                                              | to the network conditions and need     | compared to what Euclidian distance   | from different day types.             |                                       |                                           |
|                                              | further exploration. Fifth, the local  | measures. Travel time could be a      |                                     |                                       |                                           |
|                                              | window comparison does not have any    | better distance measure. Third, as it|                                     |                                       |                                           |
|                                              | physical interpretation and does not   | uses the same MSSIM formulation,      |                                     |                                       |                                           |
|                                              | consider any network properties.       | stability constants                  |                                     |                                       |                                           |

**Table A1.**