Insights into geographical window based SSIM for comparison of OD matrices

Krishna N. S. Behara¹, Ashish Bhaskar¹ Edward Chung¹
¹Smart Transport Research Centre, School of Civil Engineering and Built Environment, Queensland University of Technology, Brisbane, QLD 4000, Australia
Email for correspondence: krishnanikhilsumanth.behara@hdr.qut.edu.au

Abstract

The rich structural information of travel patterns in terms of number of trips distributed to different destinations within a region is an essential element that cannot be ignored during any statistical assessments such as comparison of Origin-Destination (OD) matrices. While most of the traditional indicators fail to compare the structural differences between OD matrices, Structural Similarity (SSIM) index outperforms them by comparing the structure, mean and standard deviations within one single formulation. In the existing literature with respect to OD matrices comparison, the application of SSIM is still theoretical in nature and there is no clarity on the level of acceptability of SSIM values due to its sensitivity towards the local window size. Thus SSIM in the context of OD matrices has to be further refined especially by emphasising on the size and physical significance of local window and local SSIM values, so that it can be applied best in practice. In this light, the paper proposes a practical approach for computing SSIM based on geographical windows that has a physical significance in terms of size and shape of windows, geographical correlation and its potential to analyse local travel patterns due to different travel distributions in different sections of the network. Also, using real Bluetooth zonal OD matrices from Brisbane network, the study also demonstrates the potential of SSIM over traditional indicators.

Keywords: structural similarity (SSIM), OD matrices comparison, geographical window, Bluetooth OD, local travel patterns

1. Introduction and literature review

Estimation of Origin-Destination (OD) matrices has long been a topic of research and researchers have focused on both the static and dynamic estimation of OD matrix (Antoniou et al. 2016). Traditionally, matrix estimation techniques have used loop detector data only for optimisation. With technology advancements, other data sources such as Bluetooth (Barceló Bugeda et al. 2010) have also been explored. Most of the research has focused on the estimation of OD matrix, and there is a limited research on the use of indicators for OD matrices comparison.

The ‘structure’ of an OD matrix is defined as distribution of the demand between different OD pairs within the matrix. Most of the traditional indicators namely, Root Mean Square Error (RMSE), Mean Square Error (MSE), are widely used as goodness of fit measures in OD matrix estimation because of simplicity in their formulations. However, these traditional indicators compare individual cells of the OD matrices and fail to capture the overall distribution of the demand (structure) in the OD matrices. This is further illustrated using real data in Section 5. This structural property cannot be ignored during any statistical
assessment such as comparison of OD matrices. Thus there is need for a potential indicator in addition to existing traditional metrics for a better comparison of OD matrices.

Limited indicators such as Structural Similarity Index (SSIM) (Djukic, Hoogendoorn and Van Lint 2013) and Wasserstein metric (Ruiz de Villa, Casas and Breen 2014) focus on assessing structural similarity of OD matrices. Wasserstein metric is based on optimisation formulation and is computationally intensive for large scale networks. The concept of SSIM was originally developed in the context of images comparison (Wang et al. 2004). However, the state-of-the-art application of SSIM in transportation perspective is theoretical in nature and needs further exploration of its potential in more realistic settings by emphasising on the physical meaning of it, so that it can applied best in practice (Pollard et al. 2013).

This research considers SSIM and investigates its appropriateness and demonstrates it as a potential tool to compute the structural similarity between OD matrices.

The rest of the paper is structured as follows. Description of study site and data is presented in section 2. Sensitivity of SSIM to local window size is illustrated in section 3; Geographical window based SSIM is proposed in section 4; SSIM potential over traditional indicators is investigated using real Brisbane data in section 5 and finally the paper is concluded in section 6.

2. Study site and Data

Brisbane City Council (BCC) region is chosen as the study site. Raw Bluetooth data, representing temporal detections of MAC IDs (Bhaskar and Chung 2013), is collected by BCC from over 845 Bluetooth MAC Scanners installed along many key corridors and intersections within the BCC region (see Figure 1(a)). Based on population distribution, BCC region is divided into four Statistical Areas (SA) namely SA4, SA3, SA2 and SA1 (order from higher to lower) respectively (see Figure 2(b) for SA3 and SA4 zones for BCC region).

Trips identified from Bluetooth detections (Michau et al. 2014) are critical construct for the bOD matrices (of size 845 x 845) between scanner pairs. The dimensions of bOD matrices are then reduced to 20 x 20 by geographically integrating the trips from Bluetooth detections with Statistical Area-3 (SA3) obtained from Australian Bureau of Statistics (ABS). This provides a 20 x 20 OD matrix where each cell represents an OD pair at SA3 level. In this study, Brisbane East refers to a portion of entire Brisbane East that is equipped with Bluetooth scanners.

Figure 1: (a) Location of Bluetooth Scanners within BCC region; (b) SA4 and SA3 levels in BCC region

3. Sensitivity of SSIM to local window size

To demonstrate the sensitivity of SSIM with window size, consider mean SSIM (MSSIM) values computed using different window sizes (3x3 to 20x20) for Monday-Sunday (Blue line)
and Monday- Tuesday (Orange line) OD matrix pairs as shown in Figure 2. Note: Here the order of OD pairs is same in both the matrices. It is observed in this study that, larger the size of sliding window, lesser is the sensitivity of SSIM towards fine correlation distortions within the OD matrix. In Figure 2, x-axis represents the size of the local window and y-axis is the MSSIM value. The MSSIM values increase as the sliding window size increases. The rate of increment of MSSIM values is less for Monday- Tuesday pair as compared to Monday- Sunday pair. This attributes to similar travel patterns between Monday-Tuesday (both of them being working days) and less similar patterns between Monday-Sunday pair. Thus, if a sliding window is used, then there is no clear consensus on the level of acceptability of the window size and its corresponding SSIM values. To overcome this limitation, SSIM based on geographical window is proposed in the next section.

Figure 2: MSSIM value increases with increase in the size of the window

![MSSIM values increase with increase in the size of the window](image)

<table>
<thead>
<tr>
<th>Size of the sliding window</th>
<th>Mon-Sun</th>
<th>Mon-Tue</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3X3)</td>
<td>0.7337</td>
<td>0.9939</td>
</tr>
<tr>
<td>(6X6)</td>
<td>0.7892</td>
<td>0.9975</td>
</tr>
<tr>
<td>(8X8)</td>
<td>0.807</td>
<td>0.9985</td>
</tr>
<tr>
<td>(15X15)</td>
<td>0.8164</td>
<td>0.9986</td>
</tr>
<tr>
<td>(20X20)</td>
<td>0.8292</td>
<td>0.9986</td>
</tr>
</tbody>
</table>

4. Proposed approach-Geographical window based SSIM

Here, we propose to apply SSIM by first arranging the origin and destinations of the OD matrix in order of the geographical similarity. Thereafter, define the window for SSIM analysis as per the geographical boundaries. The origin and destinations are first grouped into respective SA4 level and the matrix is rearranged accordingly so that its cells can be grouped to form SSIM windows, where the window is having a geographical representation at SA4 level. By doing so the SSIM window has a physical significance. Note: For this the size of the window need not to be a square matrix. Since the number of SA4 OD pairs are 25, the number of geographical windows for the OD matrix is also 25 (see Figure 3).

The SA4 zones for BCC region are Brisbane East, Brisbane North, Brisbane South, Brisbane West and Brisbane Inner (see Figure 1 (b)). Figure 3 demonstrates the application of SA4 based geographical windows for comparing SA3 (20 x 20) OD matrices of Monday (Figure 3(a)) and Sunday (Figure 3(b)), respectively. For example consider a geographical window of SA4 OD pair “Brisbane East” and “Brisbane North”. It consists of SA3 OD pairs i.e. 30101-30201, 30101-30202, 30101-30203, 30101-30204, 30103-30201, 30103-30202, 30103-30203, and 30103-30204. These SA3 OD pairs are geographically correlated because they have same SA4 origin i.e. “Brisbane East” and SA4 destination i.e. “Brisbane North”. Since “Brisbane East” and “Brisbane North” consist of 2 and 4 SA3 zones respectively, the size of the local geographical window is 2 x 4. The local SSIM values are then calculated for all geographical windows exclusively and the overall MSSIM is the average of all local SSIM values. For example, MSSIM for Sunday-Monday matrices pair, computed based on geographical window is 0.7231 (see Table 1).
By averaging, it implies that, the overall SSIM value is obtained by smoothing over all local values. Although mean SSIM values are used in this study, the local SSIM values based on geographical windows have practical significance in their own respects. For example, local SSIM computed for any local window provides valuable insights towards local travel patterns between different suburbs of the region. If the purpose is to compute the similarity of Sunday and Monday travel patterns between major suburbs, then the concept of sliding window of any size will not work. From Figure 3, it can be observed for suburb pair- "Brisbane South to Brisbane North", Sunday travel patterns are less similar to that of Monday with a local SSIM value of 0.4653 (bold in Table 1). On the other hand, for another major suburb pair- "Brisbane South to Brisbane West", Sunday travel patterns have a better similarity value of 0.8037 as compared to that of 0.4653 (of suburb pair- "Brisbane South to Brisbane North"). This can also be justified from the fact that, local trips are more dominant during weekends. Since the South and West suburbs are closer to each other as compared to other pair, it is obvious that the total number of trips are higher, resulting in higher SSIM value as compared to their corresponding Monday trips. Thus this approach provides some valuable insights towards local travel patterns by identifying the sections of network that experience different demand distributions.

**Figure 3:** Splitting Monday (a) and Sunday (b) OD matrices into geographical (SA4) windows

<table>
<thead>
<tr>
<th>Brisbane East</th>
<th>Brisbane North</th>
<th>Brisbane South</th>
<th>Brisbane West</th>
<th>Brisbane Inner</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brisbane East</td>
<td>0.8319</td>
<td>0.2437</td>
<td>0.7650</td>
<td>0.9517</td>
</tr>
<tr>
<td>Brisbane North</td>
<td>0.3311</td>
<td>0.7353</td>
<td>0.4034</td>
<td>0.7378</td>
</tr>
<tr>
<td>Brisbane South</td>
<td>0.7771</td>
<td>0.4653</td>
<td>0.8037</td>
<td>0.8037</td>
</tr>
<tr>
<td>Brisbane West</td>
<td>0.8340</td>
<td>0.7754</td>
<td>0.7562</td>
<td>0.8884</td>
</tr>
<tr>
<td>Brisbane Inner</td>
<td>0.7716</td>
<td>0.6265</td>
<td>0.8257</td>
<td>0.8385</td>
</tr>
<tr>
<td>Mean SSIM (MSSIM)</td>
<td>0.7231</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. MSSIM vs Traditional indicators using real OD matrices

To demonstrate the potential of MSSIM over traditional indicators, real Bluetooth OD matrices for 88 days (consisting of Sundays, Saturdays, Public holidays and School holiday weekdays) are compared with a typical working weekday using all five indicators (see Table 2). A typical working weekday OD matrix is computed by taking average of 75 regular working weekdays from the year 2016. Before comparing, with other distance metrics, MSSIM is converted into a dissimilarity measure (i.e. 1-MSSIM). SSIM can be an effective tool when traditional indicators are not capable of capturing structural differences. For example, Table 2 presents the OD matrices comparison results for 27th March 2016 (Easter Sunday) and 26th December 2015 (Saturday following Christmas day) as compared to aforementioned typical working weekday.

Table 2: MSSIM overcomes limitation of traditional indicators- real OD matrices comparisons

<table>
<thead>
<tr>
<th>Date (Day)</th>
<th>1-MSSIM</th>
<th>MSE</th>
<th>RMSE</th>
<th>GU</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>27th March 2016 (Easter Sunday)</td>
<td>0.3352</td>
<td>1.31\times10^9</td>
<td>3.63\times10^4</td>
<td>0.35</td>
<td>0.55</td>
</tr>
<tr>
<td>26th December 2015 (Saturday following Christmas day)</td>
<td>0.4296</td>
<td>1.31\times10^9</td>
<td>3.63\times10^4</td>
<td>0.35</td>
<td>0.53</td>
</tr>
</tbody>
</table>

It is expected that both these days (Easter Sunday and Christmas break) should have higher non-work related trips such as private travel towards different destinations (parks, relatives or friends' homes etc.) as compared to a regular working weekday. However, since, 26th December 2015 is amidst of a long 15-day Christmas and New Year Holidays, it can have different travel pattern as compared to that of Easter Sunday (from a short (4-day) holiday span). This difference is not captured by MSE, RMSE, GU and MAE but it is predominant in the MSSIM values.

6. Conclusion

The paper proposes geographical window based SSIM to address the sensitivity issues of SSIM to local window size. This is ensured by adjusting the local window boundaries such a way that it captures only those lower level zonal OD pairs (SA3) belonging to a particular higher level zonal (SA4) OD pair. This approach allows SSIM to adapt to local sub-matrix statistics rather than computing on the entire size of OD matrix. SSIM computed based on geographical window has physical significance especially when local SSIM values throw some light on the local travel patterns comparison. Also it seems to be more appropriate than choosing a sliding window of random size without any physical understanding of what the size represents.

Although many statistical indicators exist for OD matrices comparison, we choose, to select SSIM as it is robust in accounting the structural properties of OD matrices i.e. travel patterns. Using real OD matrices, SSIM values are compared with that of traditional indicators’ and it is observed that, when traditional indicators fail to distinguish the structural differences between OD matrices, SSIM can be used as a potential metric.

As a part of future scope, travel patterns shall also be analysed from the route choices perspective and shall be compared with current study SSIM interpretations.
7. References


