Comparing job accessibility in diverse spatial patterns in the Greater Golden Horseshoe area, Ontario, Canada

Bruno Dias dos Santos*
* School of Earth, Environment and Society (SEES)
Georgia Institute of Technology, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4K1
dossanb@mcmaster.ca

***Abstract*—This work aims to analyze job accessibility across diverse spatial patterns in the Greater Golden Horseshoe area, Ontario, Canada. Using variables extracted from satellite images, we identified six distinct spatial patterns with variations in building density, land use, and street layout. Afterwards, we compared accessibility to employment considering different modes of transportation in the identified spatial patterns. Results show that higher building density spatial patterns exhibit elevated accessibility for all transportation modes. The job accessibility is concentrated in approximately 10% of the GGH, posing challenges for spatial planners to devise a transportation system that accommodates these territorial differences. Statistical analyses confirm significant differences in job accessibility among spatial patterns.**

***Index Terms*—Job Accessibility; Spatial Patterns; Transportation Modes; Greater Golden Horseshoe; Urban Planning**

Introduction

Accessibility is defined as the potential to reach spatially distributed places and opportunities (Páez et al., 2012), such as jobs, parks, cultural activities, health, and education services. The accessibility to these diverse opportunities is directly dependent on the transportation network and the geographical distribution of activities, making it a key output of spatial planning.

A survey conducted by Palm et al. (2023) revealed that labor force participation ranks among the top five most compelling topics for expanding the applications of accessibility studies . The survey included over 50 participants from five Canadian provinces, with half representing the government, the majority of whom were employed in local government. Job accessibility serves as a crucial tool for comprehending urban form, the spatial mismatch between jobs and housing, and the balance between job and housing locations.

With the general context presented, this work is guided by the following question: How does job accessibility levels vary considering diverse spatial patterns? We posit the following hypothesis: job accessibility differs among various spatial patterns, even when different transportation modes are considered. The primary objective of this study is to analyze and compare job accessibility levels across different spatial patterns, focusing on the Greater Golden Horseshoe area (GGH) in Ontario, Canada. We compared the transportation modes: walking, cycling, public transportation off peak and on peak.

Fig. 1. Greater Golden Horseshoe Study Area. Source: Ontario, 2020.

Study area

The Greater Golden Horseshoe (GGH) constitutes the urban region centered around the City of Toronto, positioned at the western terminus of Lake Ontario (Figure 1). It extends northward to Georgian Bay, southward to Lake Erie, westward to Wellington County and Waterloo Region, and eastward to the counties of Peterborough and Northumberland. With a population of 10 million people and accommodating 4.9 million jobs in an area of 26,804 km², the GGH serves as the economic hub of Ontario and stands out as one of the rapidly advancing regions in North America (Ontario, 2020).

METHODS

Our methodology comprises four primary steps. The initial step involves feature extraction and dataset creation, where urban form attributes are extracted from satellite imagery and road network data, subsequently mapped onto a hexagonal grid. The second step employs Principal Component Analysis (PCA) to reduce dataset dimensionality. Utilizing only the principal components, we conduct a non-supervised classification to identify clusters representing the distinct spatial patterns within the GGH. In the final step, we compare job accessibility levels among the identified spatial patterns. Two statistical tests, the Kruskal-Wallis test and the pairwise Wilcoxon test, are employed to evaluate variations in job accessibility levels. For this work, we used already processed and classified satellite images provided by the Global Human Settlement Layer (GHSL) (European Commission. Joint Research Centre., 2023) and the Copernicus Global Land Service (CGLS) (Buchhorn et al., n.d.); Spatial Access Measures (SAM) dataset (Statistics Canada, n.d.), which measures accessibility to different opportunities, considering different modes of transportation; and a road network (OpenStreetMap, 2023) - more information about the data can be seen in the appendix.

Feature extraction

We integrate the selected datasets onto a hexagonal grid with an area of 500 m², comprising 53,609 regular cells across the study area. These hexagons served as spatial units of analysis, providing a standardized reference for merging data from disparate sources, including the road network and remotely sensed imagery.

For each hexagon, we extracted statistical measures — mean, maximum, minimum, standard deviation, median, sum, and majority values — for the GHSL and CGLS images. For the road network, we computed road sinuosity and extracted road network attributes, such as the total number of roads, intersections, intersection density (intersections per road), and statistical values related to road sinuosity.

Additionally, spatial variables were generated through neighborhood connections between hexagons. For each variable "i", we computed the neighbor mean value (NM_i) and the normalized difference (DNM_i) between each hexagon and its neighborhood mean. The methodology employed for deriving these neighborhood variables was devised by Dos Santos et al. (2022) , utilizing a distance matrix to identify neighboring cells within the grid. In total, 191 attributes were extracted for each hexagonal cell.

Reducing data dimension

Understanding intraurban differences is crucial for comprehending variations in accessibility levels. With the dataset at hand, we selected cells based on a threshold value for the presence of built-up areas. Hexagons with less than 10% of the sum of residential and commercial buildings were excluded from the analysis. This criterion was implemented to enhance the clustering process. Without this restriction, the clustering algorithm tended to create numerous clusters outside the urban area, without showing the intraurban differences.

We employed Principal Component Analysis (PCA) to reduce the dimensionality of the data. PCA is a technique used to select features and decrease dimensions in the presence of multiple variables. It achieves this by generating a reduced number of linear combinations of the original variables while preserving a substantial amount of the information contained in them (Ringnér, 2008) . We retained all components where the cumulative variance explained accounted for more than 90% of the total variance in the data. Subsequently, we standardized the components into Z-scores for the application of PCA.

Identifying spatial patterns

Using only the selected components, we applied an unsupervised classification with the K-mean algorithm to identify the spatial patterns of GGH. K-Means is a clustering algorithm, which takes as input a number k of clusters and the algorithm divides the instances into k clusters. The algorithm optimizes the separation between the groups by minimizing the internal variation of the clusters - the inertia. The clustering process is accelerated by randomly assigning the instances to one of the k clusters and then redistributing the instances in order to reduce the distance (in our case, the Euclidian distance) between each observation and the central point (the centroid) of its cluster.

Evaluating job accessibility

We transferred the SAM data from the dissemination blocks to the cell grid using a weighted area interpolation. This technique enables us to disaggregate accessibility values to another spatial unit, smoothing variations between polygons. It achieves this by using the area of overlapping geometries to apportion variables.

Following this, we conducted an initial statistical test to assess job accessibility levels across spatial patterns. Given that the job accessibility variables did not exhibit a normal distribution, we employed the Kruskal-Wallis test to determine whether the accessibility medians differed among the clusters. If a median difference was identified between the groups, we then applied the pairwise Wilcoxon test to discern which specific group differed from the others. We adopted a significance level of 0.05 for both tests.

RESULTS

Principal Components Analysis

For the clustering analysis, thirty-five principal components were selected, collectively explaining over 90% of the total variance. Table I highlights the three most significant variable contributions to the first four principal components. Across all components, variables related to neighbor relations demonstrated the most substantial loadings. Component 1 exhibits higher absolute loadings in the neighbor mean of built-up type variables, predominantly associated with more densely built-up areas in the study area; Component 2 displays higher absolute loadings in variables related to the fraction of water on neighboring cells, indicating a correlation with waterfront areas; Component 3 focuses on the presence of built-up areas and residential use in neighboring cells; and Component 4 correlates with the presence of trees in neighboring cells.

TABLE I
HIGHER VARIABLES CONTRIBUTIONS TO THE FIRST FOUR PRINCIPAL COMPONENTES (PC)

Variable	**PC1**	**PC2**	**PC3**	**PC4**
Built-up sum (NM)	**0.12**	-0.03	-0.03	0.02
Built-up mean (NM)	**0.12**	-0.03	-0.03	0.02
Comercial area (NM)	**-0.11**	0.02	0.01	-0.03
Water mean (NM)	-0.01	**0.19**	-0.02	-0.11
Water sum (NM)	-0.01	**0.19**	-0.02	-0.11
Water median (NM)	-0.01	**0.19**	-0.03	-0.11
Built-up mean (DNM)	-0.06	0.03	**-0.23**	0.04
Built-up sum (DNM)	-0.06	0.03	**-0.23**	0.04
Residential area (DNM)	-0.06	0.02	**-0.23**	0.03
Tree mean (DNM)	0.07	-0.04	0.05	**-0.16**
Tree sum (DNM)	0.07	-0.04	0.05	**-0.16**
Tree median (DNM)	0.06	-0.03	0.02	**-0.16**
Proportion of Variance	35%	9%	6%	5%
Cumulative Proportion	35%	45%	51%	56%

Clustering result

Only one-quarter of the study area was analyzed to perform the clustering classification, as only these cells had the sum of residential and commercial areas exceeding ten percent of the cell area. A series of experiments and visual analyses were conducted to determine the optimal number of clusters. After several experiments and visual analyses, we settled on six final clusters, considering the best trade-off between identifying meaningful clusters and their quantity (Figure 2).

Each final cluster can be interpreted as a distinct spatial pattern with unique levels of built-up area density, land use, and street layout. We named the spatial patterns according to their features. The first cluster, "Cluster 1 - High Building Density," corresponds to 17% of the analyzed clusters and exhibits the highest building density (in area, height, and volume) among all clusters. It also has the highest number of intersections per road, suggesting more connections between its streets. In turn, "Cluster 2 - Tree-lined Neighborhoods" is mainly characterized by the highest presence of trees, with about 20% of its area dedicated to residential use, no commercial areas, and a low inclination for farms. Eighteen percent of the clusters belong to this class, located at the periphery of urban areas. The third cluster, "Cluster 3 - Medium Building Density," is the largest, accounting for 28% of the analyzed area. Situated at the border of the first clusters, it features the second most residential and commercial areas. "Cluster 4 - Mixed Rural Use" represents 10% of the analyzed area and, as the name suggests, has a mosaic of rural use, including forests,

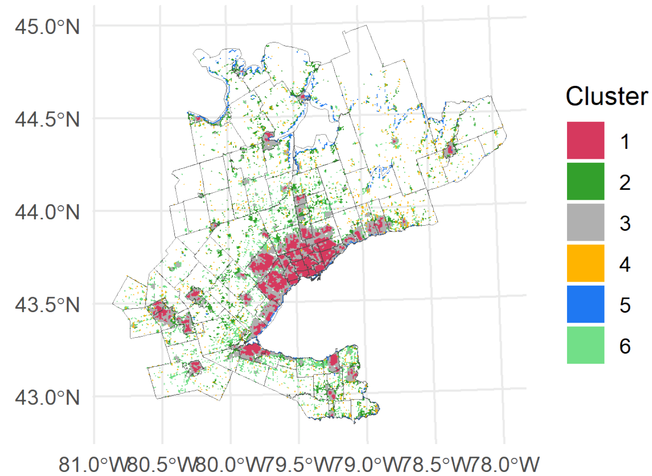

Fig. 2. Study area spatial patterns.

grasslands, shrubs, and crop areas. The least represented cluster, with only 6% of the clusters classified in this category, "Cluster 5 - Waterfront Housing," is located in front of the Ontario, Sinkoe, Erie, and Huron lakes, in neighborhoods with very low building density. Finally, "Cluster 6 - Farms and Rural Neighborhoods" represents 21% of the study area, scattered throughout the study area, sometimes close to major roads, and sometimes in the interior. It exhibits the highest presence of crop land.

DIFFERENCES IN JOB ACCESSIBILITY

Figure 3 illustrates job accessibility by mode in the Greater Golden Horseshoe study area. Notably, the region around downtown Toronto exhibits the highest levels of accessibility, with a decrease in accessibility as you move away from Toronto, a trend particularly noticeable on the walking map. Table II presents job accessibility by spatial patterns and modes. Notably, "Cluster 1 - High Building Density" and "Cluster 3 - Medium Building Density," characterized by the presence of commercial areas, demonstrate the highest mean job accessibility values. In contrast, the remaining clusters display considerably lower job accessibility indices, posing challenges in distinguishing them based solely on average values.

TABLE II
MEAN VALUES FOR JOB ACCESSIBILITY BY SPATIAL PATTERN AND MODE OF TRANSPORT

Cluster	**Walking**	**Cycling**	**Public transit**	**Public transit (peak)**
1	0.0478	0.1694	0.2097	0.2244
2	0.0023	0.0098	0.0044	0.0046
3	0.0176	0.0711	0.0797	0.0862
4	0.0016	0.0076	0.0029	0.0031
5	0.0046	0.0174	0.0157	0.0167
6	0.0021	0.0107	0.0042	0.0044

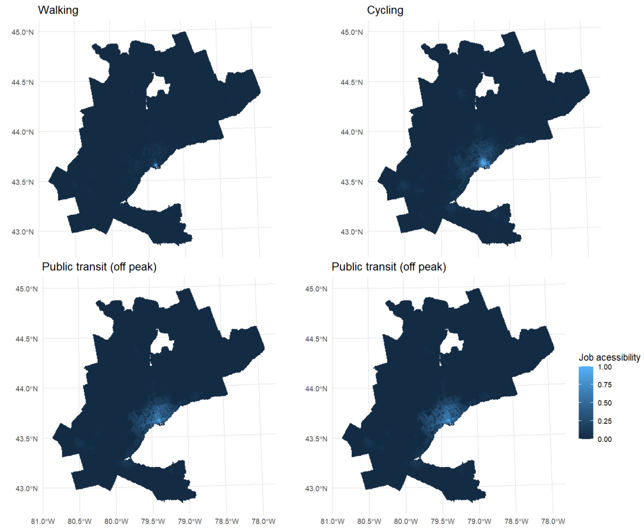

Fig. 3. Job accessibility by mode in the Greater Golden Horseshoe Study Area.

The spatial patterns exhibit differences in their median values of job accessibility for all transportation modes: walking, cycling, public transportation (off-peak), and public transportation (peak) (see Table III). The Kruskal-Wallis test results for all transportation modes yield p-values $<$ 2.2e-16, suggesting that there is significant statistical evidence against the null hypothesis of no difference in job accessibility between spatial patterns in the GGH.

TABLE III
COMPARISIONS OF JOB ACCESSIBILITY BETWEEN SPATIAL PATTERNS.

Mode	**Statistically different (Kruskal-Wallis test)**	**No statistical difference (Pairwise Wilcoxon)**
Walking	1, 2, 3	4 and 5; 5 and 6
Cycling	1, 2, 3, 6	4 and 5
Public transit (off peak) Public transit (peak)	1, 3, 4, 5	2 and 6

Additionally, we observed that the spatial patterns "Cluster 1 - High Building Density" and "Cluster 3 - Medium Building Density" exhibit statistical differences in terms of job accessibility compared to the other clusters. "Cluster 4 - Mixed Rural Use" and "Cluster 6 - Farms and Rural Neighborhoods" show no statistical difference for walking when compared to "Cluster 5 - Waterfront Housing". Cluster 4 also shows no statistical difference for cycling when compared to Cluster 5. On the other hand, "Cluster 2 - Tree-lined Neighborhoods" and Cluster 6 show no statistical difference across both public transit modes (off-peak and on-peak).

Conclusion

The main objective of this study was to analyze and compare the levels of accessibility to work in various spatial patterns in the Greater Golden Horseshoe (GGH) region in Ontario, Canada. The comparison involved four modes of transportation: walking, cycling and public transit during peak and off-peak hours.

Initially, we identified six distinct spatial patterns in the study area, characterized by differences in building density, land use, location and street layout. Our analysis revealed that accessibility to work varies across the study area, with spatial patterns that exhibit higher building density also demonstrating high values of accessibility to work, regardless of the mode of transport considered.

Notably, the study area has a low overall density, with less than 25% of cells containing more than 10% residential and/or commercial area. In addition, accessibility to work is concentrated in approximately 45% of the area analyzed, which corresponds to around 10% of the GGH. This concentration represents a complex challenge for spatial planners, as employment opportunities are confined to a small part of the territory. This scenario can lead to housing shortages in urban areas and require complex strategies to develop a transportation system capable of dealing with these territorial differences.

References

Buchhorn, M., Smets, B., Bertels, L., Roo, B.D., Lesiv, M., Tsendbazar, N.-E., Herold, M., Fritz, S., n.d. Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2018: Globe. doi:10.5281/ZENODO.3518038

Dos Santos, B.D., De Pinho, C.M.D., Oliveira, G.E.T., Korting, T.S., Escada, M.I.S., Amaral, S., 2022. Identifying Precarious Settlements and Urban Fabric Typologies Based on GEOBIA and Data Mining in Brazilian Amazon Cities. Remote Sensing 14, 704. doi:10.3390/rs14030704

European Commission. Joint Research Centre., 2023. GHSL data package 2023. Publications Office, LU.

Ontario, 2020. Planning transportation for the Greater Golden Horseshoe | ontario.ca.

OpenStreetMap, 2023. OpenStreetMap Dataset.

Páez, A., Scott, D.M., Morency, C., 2012. Measuring accessibility: Positive and normative implementations of various accessibility indicators. Journal of Transport Geography, Special section on accessibility and socio-economic activities: Methodological and empirical aspects 25, 141–153. doi:10.1016/j.jtrangeo.2012.03.016

Palm, M., Teel, S., Tiznado-Aitken, I., Soukhov, A., Paez, A., Farber, S., Hain, M., 2023. Developing data driven equity standards: Stakeholder perspectives.

Ringnér, M., 2008. What is principal component analysis? Nature Biotechnology 26, 303–304. doi:10.1038/nbt0308-303

Statistics Canada, n.d. Spatial Access Measures.