Modifiable Areal Unit Problem (MAUP) effects on accessibility to supermarkets in Montreal, Canada

Jasso Chávez José Arturo[1], Manaugh Kevin [2]

[1]Department of Geography, McGill University, jose.jassochavez@mail.mcgill.ca
[2]Department of Geography and Bieler School of Environment, McGill University, kevin.manaugh@mcgill.ca

ABSTRACT

Most place-based accessibility analyses work at aggregate spatial scales, such as census tracts or dissemination areas. It is well known that this introduces errors as it does not capture the variance in travel distances and times within any aggregate zone. Only a few studies have examined these issues related to the Modifiable Areal Unit Problem (MAUP), specifically in the context of accessibility studies. This paper quantifies the misclassification of accessibility when comparing point-to-point accessibility from every residential building to supermarkets in Montreal versus the aggregation at census tracts and dissemination areas. Using linear regression models, we found that the size of the census geography, population density, and income correlate with the overestimated accessibility. Also, this effect is more significant in census tracts than in dissemination areas, where 61% and 50% of the residential buildings are misrepresented, respectively.

1. Introduction

Studying accessibility allows researchers to understand a city's relationship between land use and transportation. Carrying out accessibility studies focuses efforts on increasing residents' quality of life. Progress in computational capabilities and detailed data on transportation systems, such as General Transit Feed Specification (GTFS) and detailed information on the city's structure (buildings, facilities, roads, etc.) allow new opportunities for accessibility analysis to achieve these goals (Stępniak & Goliszek, 2017).

These technical advances establish new opportunities and challenges that allow cities to be studied in detail and quickly (Stępniak & Goliszek, 2017). Among these new opportunities, one addresses past problems with greater precision, such as the Modifiable Areal Unit Problem (MAUP). MAUP, or the Openshaw effect, is a statistical bias associated with aggregating data at different spatial scales or zoning configurations (Goodchild, 2022; Kwan & Weber, 2008). The choice of a particular spatial scale or zoning is often arbitrary, and it is chosen because it may be more convenient in computational terms or is the only spatial resolution available. These resolutions may follow political-administrative limits or be based on the territory's characteristics (population, land use, transportation); some others are based on regular geometry figures such as grid cells in the shape of hexagons or squares of different sizes (Jelinski & Wu, 1996; Lowell, 2008).

 Volume DOI: 10.17605/OSF.IO/N6DU5

Many studies have measured and compared spatial and statistical results when studying phenomena in two or more different spatial scales. Most conclude that there are significant differences and recommend using the most disaggregated scale (Bell et al., 2013; Horner & Murray, 2004; Wan, n.d.; Zhang & Kukadia, 2005). However, those comparisons are merely relative as they compare two aggregated scales (e.g., census tracts with dissemination areas) and do not compare with the most disaggregated or the basic unit of analysis: residential buildings. This would be the most accurate spatial scale as it is a single unit of analysis and cannot be divided into smaller and are directly connected to the roads, so the starting and ending point of the trip is more accurate than larger areas such as census tracts or dissemination areas.

There is not enough literature to study differences in accessibility levels between spatial scales versus residential lots. Those who have approached the problem concluded that the bias is significant and that residential buildings are misclassified, but it is unclear the direction of the misestimation (overestimation and underestimation) and the factors explaining this (Bryant & Delamater, 2019). This research aims to quantify the misclassification (overestimation and underestimation) of accessibility between residential lots versus census tracts and dissemination areas using the cumulative opportunities method. Further, we examine the variance across different socioeconomic groups. Specifically, accessibility to supermarkets will be examined in the city of Montreal. Accessibility to supermarkets has gained attention from decision-makers who have aimed to understand spatial accessibility to places that sell nutritious and accessible food, mainly supermarkets or grocery stores (Farber et al., 2014).

2. Methods & Data

The methodology is divided into two sections:

1. Generation of several travel time matrices
2. Measuring accessibility and statistical análisis

2.1 Time travel matrices

Several travel time matrices between multiple origins and destinations were generated using Rapid Realistic Routing with R5 in R (r5r), an R-based package based on Java. This is done by building a multimodal transport network using the OpenStreetMap street network and Montreal's current public transit information in a General Feed Transit Specification (GTFS) data format and a Digital Elevation Model (DEM).

The time travel matrices were generated at different spatial scales: census tracts (CT), dissemination areas (DA), and residential buildings to supermarkets. For the CT's and DA's, the travel time matrix was calculated using the centroid of the polygon of origin to the centroid of the destination polygon; and another one from every residential building to every supermarket.

The travel matrix was calculated by public transit (using all the transportation agencies in Greater Montreal). The parameters used were a maximum walking distance to access and egress of public transport was 1500 meters with a speed computed at 3.6 km/h. The chosen day was on Wednesday, a "normal day," and the travel time matrices were calculated every minute from 10 and 11 am and then averaged.

The number of units and a general description for each scale is shown in Table 1.

Table 1. Description of different spatial scales

Spatial scale	Description	No. of units	Avg. area (sq. km)	Source
Residential buildings	buildings classified as apartments or housing or located in residential land use were used for the analysis	780,000	0.00016	OpenStreetMap (2021)
Dissemination areas	are small, relatively stable geographic areas with a population between 2,500 and 8,000 people	6,469	0.72	Statistics Canada (2016)
Census tracts	an area composed of one or more neighboring dissemination blocks and is the smallest standard geographic area for which all census data are released	970	4.7	Statistics Canada (2016)

The location of supermarkets in Montreal were obtained from DMTI (DMTI Spatial, Markham, ON)), which contains the location and characteristics of economic activities in Montreal and its metropolitan area. The database was cleaned by searching keywords with the word "supermarket," "supermarche" and other combinations of those words both in English and French, as well as the name of well-known supermarket chains. 480 supermarkets were identified.

Figure 1 shows an example of the structure of these geographies in Montreal.

Source: DMTI (2020), OSM (2022) and Statistics Canada (2016).

Figure 1: Comparison of different spatial scales and boundaries.

2.2 Accessibility measures and statistical analysis

The second part of the analysis measured accessibility using the cumulative opportunity measure using a 30-minute threshold and counts the number of opportunities reached within a travel time or distance. We used a 30-minute threshold as is suggested as a feasible time to satisfy life's necessities by transit and which has been adopted by governments around the world (Levinson, 2019). The formula is:

$$Ai = \sum_{j} OiBj$$

Where Ai is the accessibility between two places, Bj is a binary value equal to 1 if zone j is within the predetermined time threshold and 0 otherwise, and Oi is the number of opportunities at zone j.

Then, we compared differences in accessibility at different spatial scales to quantify the number of residential buildings misclassified and then by different income groups: high-income groups (median of > $45,000) and low-income groups (median of < $45,000).

3. Results

3.1 Differences in accessibility

Figures 2 and 3 represent the % of difference between the accessibility at every residential building and the centroid of census tracts and dissemination areas. The warm colors mean that census tracts or dissemination areas overestimate accessibility, while cold colors mean underestimation; yellow colors mean slight or no difference.

Figure 2 shows that when comparing residential buildings and census tracts, buildings outside of the island of Montreal face extreme changes, having overestimations or underestimations of 50% or more; in the island of Montreal, this misrepresentation is smaller or null. There is still a difference between residential buildings versus dissemination areas, but it is not as extreme, and the differences tend to be smaller. Figure 3 shows those effects in a higher resolution. We identify that the difference in the levels of accessibility difference depends on several factors. The first is the network walking distance between a given building and the centroid of census polygon. The further away, the greater the misestimation. Likewise, the closer the building is to a public transport station than the centroid of a geographic area, the building will have greater accessibility than the centroid of the polygon, causing the aggregate area to underestimate the values.

 Volume DOI: 10.17605/OSF.IO/N6DU5

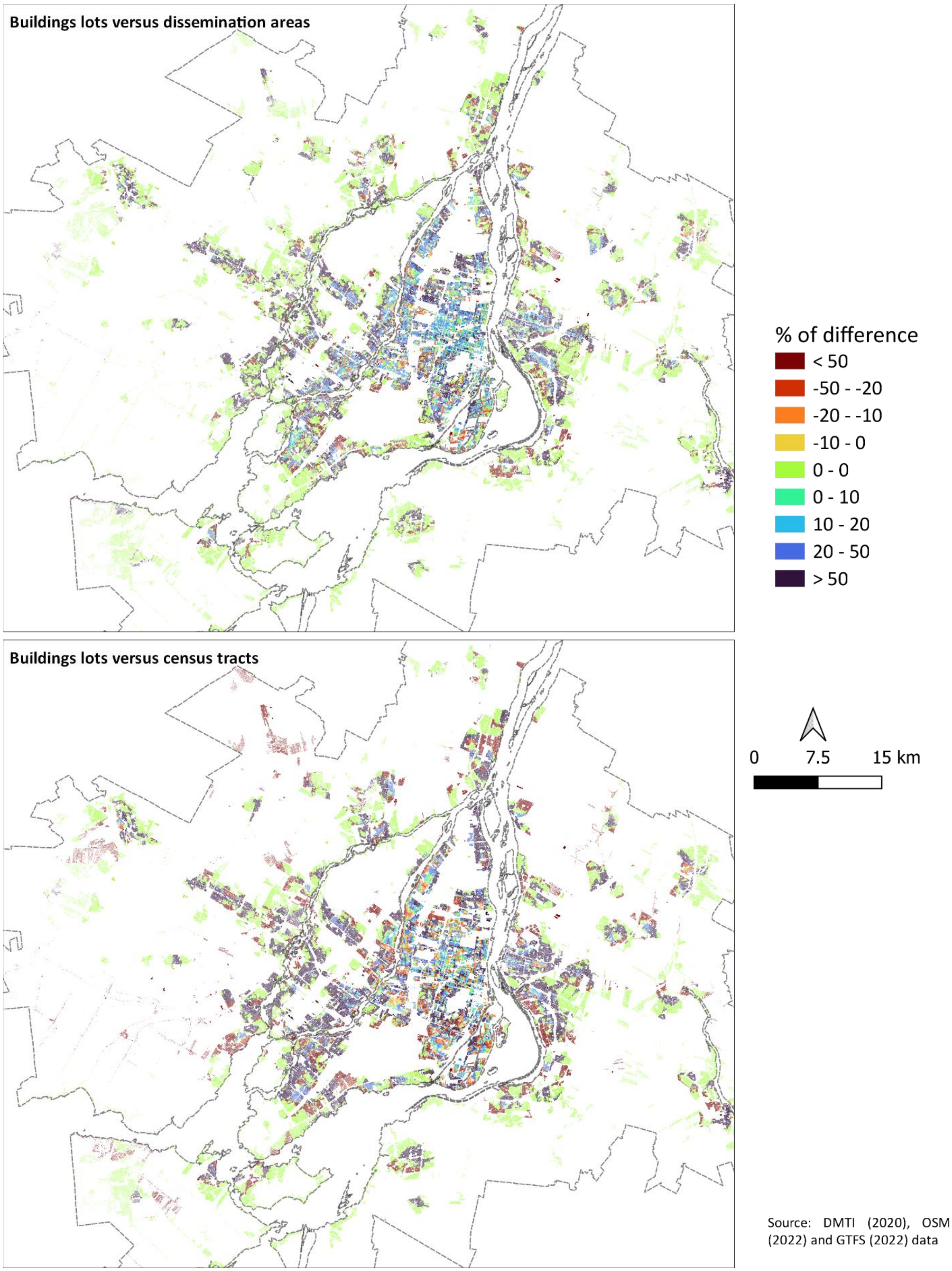

Figure 2: Comparison of different spatial scales and boundaries.

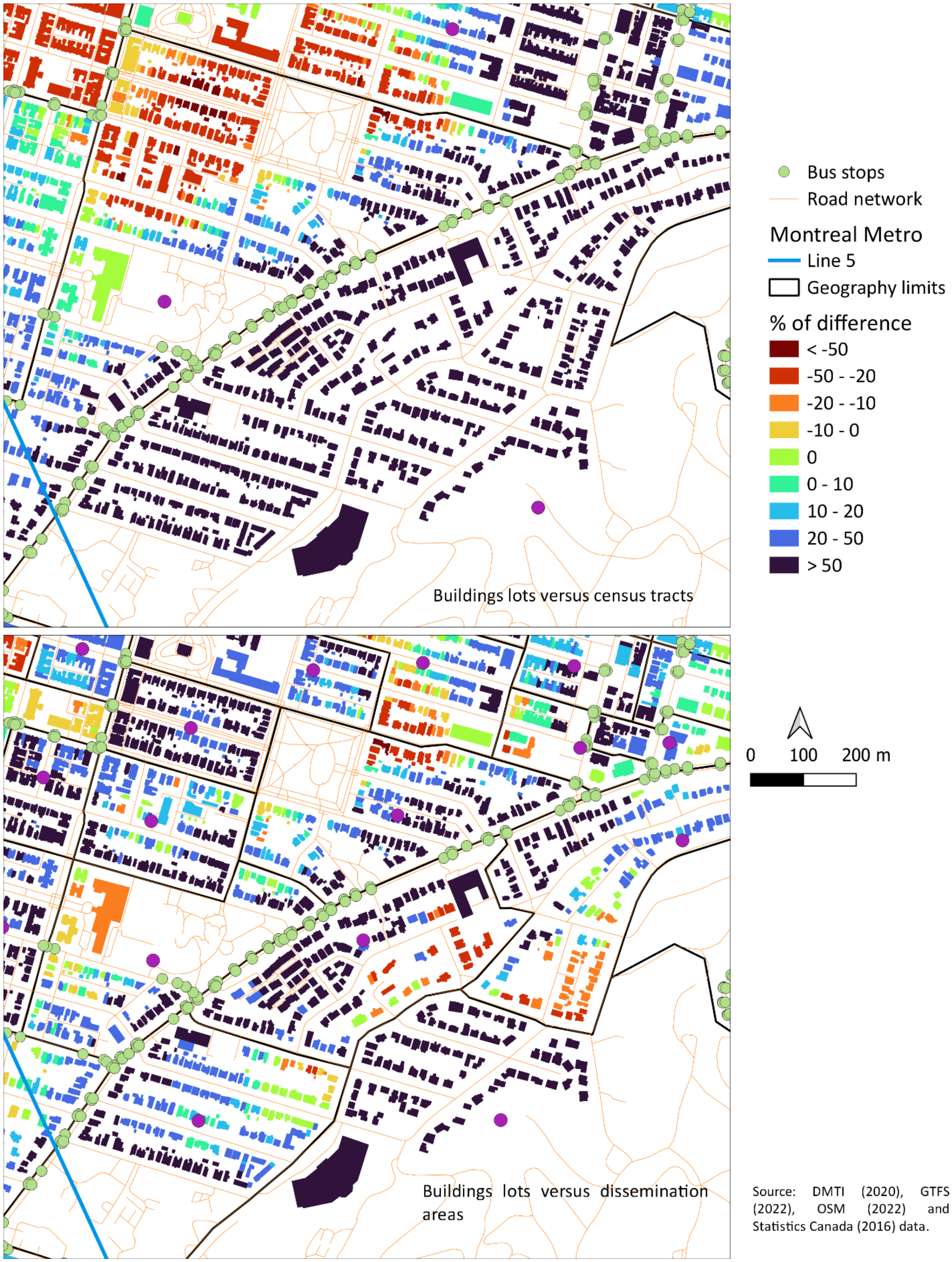

Figure 3. Difference in accessibility between spatial scales.

 Volume DOI: 10.17605/OSF.IO/N6DU5

3.2 Misclassification of buildings

Table 2 summarizes the misestimation of Montreal residential buildings when choosing dissemination areas and census tracts. We found that census tracts misestimate about 61% (~480,000) of the residential buildings in Montreal, while the dissemination areas misestimate 50% (390,000). The direction of the misclassification varies depending on the scale. The underestimation is more present in CTs and DAs, with around 40%. However, the overestimation levels differ, 22% for census tracts and 12% for dissemination areas.

Regarding the percentage of misestimation, we found that the overestimation groups of "more than 50%" for both census tracts and dissemination areas are the second in levels of misestimation while the third place is those that are "less than 50%" being constant for both scales. This is mainly because the places that have a difference of "more or less than 50%" are places that are in the periphery where there are is a low access to supermarkets and are more sensitive to any changes in spatial scale.

Table 2. Misestimation of Montreal residential buildings when measuring accessibility at different spatial scales. The smallest resolution is the reference.

	Census tracts		**Dissemination areas**	
% of misestimation	**#of buildings**	**% respect to the total**	**$ of buildings**	**% respect to the total**
More than -50	125,073	16	57,985	7.4
-20 to -50	26,832	3.4	20,472	2.6
-10 to -20	13,001	1.7	9,074	1.2
-5 to -10	5,374	0.7	4,072	0.5
-1 to -5	1,278	0.2	1,295	0.2
0	301,242	38.7	388,497	49.8
1 to 5	1,753	0.2	2,170	0.3
5 to 10	5,640	0.7	8,804	1.1
10 to 20	16,958	2.2	27,270	3.5
20 to 50	54,372	7	74,879	9.6
More than 50	227,827	29.2	184,832	23.7
No misestimation	301,242	39	388,497	50
Overestimation	171,558	22	92,898	12
Underestimation	306,550	39	297,955	38
Total misestimated	478,108	61	390,853	50

The effects of the misclassification of residential buildings vary depending on socioeconomic groups (Table 3). We found that residential buildings in higher-income geographies are more misclassified than those in low-income ones. It is essential to mention that dissemination areas and census tracts with low income are not overestimated; they are underestimated.

Table 3: Percentage of residential buildings misclassified per median income.

Income	Misestimation	Dissemination areas	Census tracts
		%	%
High income	No misestimation	62	51
	Overestimation	8	19
	Underestimation	30	30
Low Income	No misestimation	48	37
	Overestimation	10	22
	Underestimation	42	40

3.3 Linear regressions

Table 4 shows the results of the different linear regressions when trying to identify the factors that explain the differences in the percentage of change when comparing accessibility at a building scale versus other spatial scales. Two models were run for each spatial scale: one to predict over-estimation and another to predict underestimation.

The dependent variable is the percentage difference (%) between the accessibility to supermarkets in each residential building and the accessibility at the centroid of each polygon.

Table 4. Linear regression results.

	Buildings versus DA		Buildings versus CT	
	Overstimation	Underestimation	Overstimation	Underestimation
Intercept	-56.1244	150.598	-52.2027	78.9613
Shape area	-0.501986	2.21367	-0.101243	0.230642
Population density	2.49262	-6.451	3.65323	-9.72267
Island of Montreal	27.0839	-16.3798	18.7641	10.4076
Income	-0.0759	0.026	-0.108	0.377
R Squared	0.4199	0.098	0.529475	0.0640023
Degrees of freedom	92886	297810	171489	306358

All variables significant at $p<0.001$'
Note: Shape area = sq. km; population density = 1000 persons per sq. km;
Located in the main Island of Montreal: 1 = yes, 0 = no; Income in hundred thousand dollars.

Volume DOI: 10.17605/OSF.IO/N6DU5

The results indicate that the area of the census polygon, population density, income and whether a location is on the island of Montreal explain the overestimation of accessibility and these models have a relatively high R squared. However, in the underestimation models, these same variables are significant but the models have a low R squared.

For overestimated values, smaller polygons are associated with smaller overestimations. If the population density increases, the overestimation decreases, and if the area/unit is on the Island of Montreal, the overestimation decreases (central areas). Those results are consistent in all the overestimation models. Those variables explain 50% of the variation for the residential buildings versus CT and 42% for the residential buildings versus DA.

For the underestimation models (all values are positive), the results indicate that if the shape area increases, the underestimation increases; if the population density decreases, the underestimation decreases; if the residential building is located in Montreal, the underestimation decreases. However, these models have low levels of explanation R (squared lower than .10).

4. Discussion & Conclusion

The results indicate that aggregating the data misestimate accessibility to supermarkets in Montreal when comparing disaggregate residential buildings to aggregate with CTs and DAs, with 62% and 50% of the buildings being misrepresented, respectively. The area of geography, population density, and income are correlated to the misrepresentation. Those variables are highly correlated (50% and 42% for CTs and DAs) with the overestimation of accessibility but lowly correlated with the underestimation of accessibility (<10% for both).

A larger polygon means that residential buildings are more likely to be further from the centroid, making the misrepresentation bigger. For this reason, buildings close to the centroid have less distortion than those further away. To understand why underestimation is less correlated than overestimation, more research is needed by adding traffic-related variables. Also, we found that low-income groups should be considered when we aggregate data. In some sense, this is something positive as this means that they have higher levels of accessibility.

Despite the residential buildings being closer to reality, a trade-off versus the computation times needs to be considered as it could increase the total computation time when calculating the travel time matrices. In this context, measuring accessibility at a building level implies generating a maximum of 23,040,000,000 observations (origin-destination pairs). Measuring accessibility at dissemination areas is around ten times smaller, and at a census tract scale, around 400 times smaller in terms of observations. Likewise, another disadvantage of using residential buildings is that it does not have detailed information on variables such as income and population density which can limit analysis.

Studying accessibility to supermarkets using residential buildings is an alternative to studies using big data which have limited availability of information on residential addresses. Combining this approach with GPS that has real-time data on the location of people and how much time people spend on certain locations and could reduce other problems as the Uncertain Geographic Context Problem (UGCP).

The results will mark a precedent in accessibility studies for the elaboration of models using residential buildings compared to aggregation scales to understand how much those geographies overestimate and/or underestimate accessibility to urban amenities and how it affects differently the population groups. Understanding the bias levels of spatial aggregation can inform decision makers about the importance of using residential buildings for accessibility studies in order to correctly allocate resources.

We conclude that it is essential to use residential buildings when calculating accessibility to supermarkets. If not possible, it is essential to consider misclassification levels when aggregating the data. Also, those results only explain the case of supermarkets in Montreal. Further research is needed to understand the effects of the MAUP on accessibility to other facilities and cities worldwide.

References

Bell, S., Wilson, K., Bissonnette, L., & Shah, T. (2013). Access to primary health care: does neighborhood of residence matter?. Annals of the Association of American Geographers, 103(1), 85-105.

Bryant, J., & Delamater, P. L. (2019). Examination of spatial accessibility at micro- and macro-levels using the enhanced two-step floating catchment area (E2SFCA) method. *Annals of GIS*, *25*(3), 219–229. https://doi.org/10.1080/19475683.2019.1641553

Farber, S., Morang, M. Z., & Widener, M. J. (2014). Temporal variability in transit-based accessibility to supermarkets. *Applied Geography*, *53*, 149–159. https://doi.org/10.1016/j.apgeog.2014.06.012

Goodchild, M. F. (2022). The Openshaw effect. *International Journal of Geographical Information Science*, *36*(9), 1697–1698. https://doi.org/10.1080/13658816.2022.2102637

Horner, M. W., & Murray, A. T. (2004). Spatial Representation and Scale Impacts in Transit Service Assessment. *Environment and Planning B: Planning and Design*, *31*(5), 785–797. https://doi.org/10.1068/b3046

Jelinski, D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, *11*(3), 129–140. https://doi.org/10.1007/BF02447512

Kwan, M.-P., & Weber, J. (2008). Scale and accessibility: Implications for the analysis of land use–travel interaction. *Applied Geography*, *28*(2), 110–123. https://doi.org/10.1016/j.apgeog.2007.07.002

Levinson, D. M. (2019). The 30-Minute City: Designing for Access. *Network Design Lab*.

Lowell, K. (2008). Fiat boundaries: Some implications for interpretation, decision-support, and multi-temporal analysis. *Environmental and Ecological Statistics*, *15*(4), 369–383. https://doi.org/10.1007/s10651-007-0060-x

Stępniak, M., & Goliszek, S. (2017). Spatio-Temporal Variation of Accessibility by Public Transport—The Equity Perspective. In I. Ivan, A. Singleton, J. Horák, & T. Inspektor (Eds.), *The Rise of Big Spatial Data* (pp. 241–261). Springer International Publishing. https://doi.org/10.1007/978-3-319-45123-7_18

Wan, M. (n.d.). *Modifiable Areal Unit Problem (MAUP) Effects on Assessment of Accessibility via Public Transit*. 95.

Zhang, M., & Kukadia, N. (2005). Metrics of Urban Form and the Modifiable Areal Unit Problem. *Transportation Research Record: Journal of the Transportation Research Board*, *1902*(1), 71–79. https://doi.org/10.1177/0361198105190200109