Spatial Knowledge and Information Canada, 2019, 7(5), 2

Correlation of Public Transit Accessibility Measures with Actual Ridership

SARAH BREE
Geography and Planning
University of Saskatchewan
sarah.bree@usask.ca

EHAB DIAB
Geography and Planning
University of Saskatchewan
ehab.diab@usask.ca

SCOTT BELL
Geography and Planning
University of Saskatchewan
scott.bell@usask.ca

ABSTRACT

Transit accessibility measures are important tools used by planners to understand the effects of changes to the public transit system. However, it is not clear how existing accessibility measures (models) described in the literature correlate with actual public transit ridership data. Public transit systems vary dramatically according to the regions they serve, and no single model has been identified that accurately measures accessibility across the spectrum. This paper evaluates several transit system accessibility models by correlating the accessibility metric they produce with actual ridership data, using the City of Saskatoon as a case study. The results show that frequency based models result in higher correlation than coverage based models and a distance decay function based on the distance from demand location to service location further increases the correlation. This paper provides transportation planners a better understanding of the correlation between different transit accessibility measures and actual transit ridership.

1. Introduction

Saskatoon is continuously improving its transit system, including the planned introduction of a Bus Rapid Transit (BRT) system with an estimated cost of between 90 and 150 million dollars (City of Saskatoon, 2018). Successful implementation of such an infrastructure project requires that costs and impacts be accurately estimated and reported to planners, decision makers, and citizens (Ding 2018; Kim 2018). One method to assess planned changes to transit systems is to develop service metric models. These models estimate how the service metrics change when model inputs change (i.e. transit system configuration).

Accessibility is a key measure of public transit system performance. It refers to the ease with which locations can be accessed from other locations (Morris, Dumble, & Wigan, 1979). Several researchers examined the concept of accessibility. For example, Thill and Kim (2005) and Lei (2010) proposed several options to calculate accessibility based on distance to service using gravity functions.

Luo and Wang (2003) proposed the Two Step Floating Catchment Area model (2SFCA) to estimate geographical accessibility of medical services. Their model considered supply of surrounding services at a particular demand location, and the total demand on the services by surrounding locations. Subsequently, McGrail and Humphreys (2009) examined the use of the 2SFCA model in rural Victoria, Australia, and Dai (2010) examined the use of the 2SFCA model for estimating access to health care in Detroit, Michigan.

Luo and Qi (2009) proposed an enhanced 2SFCA (E2SFCA) model that applies a distance-decay to both steps of the original 2SFCA model. They proposed discrete weightings that change in a stepwise fashion at defined distances. Langford et. al. (2012) proposed a transit-enhanced E2SFCA model for estimating geographical access into transit systems, which is described in more detail later in this paper. Recently, Walk Score (Seattle WA) introduced *Transit Score* to quantify local accessibility to transit (Walk Score, 2018).

This paper evaluates several transit system accessibility models by correlating the accessibility metric they produce with actual public transit ridership data for Saskatoon.

2. Data and Methods

Saskatoon Transit's *General Transit Feed Specification* (GTFS) data was accessed on June 1, 2018. This dataset includes the location of every stop, route, departure direction, and the time of every weekly departure. At that time, Saskatoon Transit operated 41 bus routes serving 1465 stops (Figure 1), with 261,868 weekly departures.

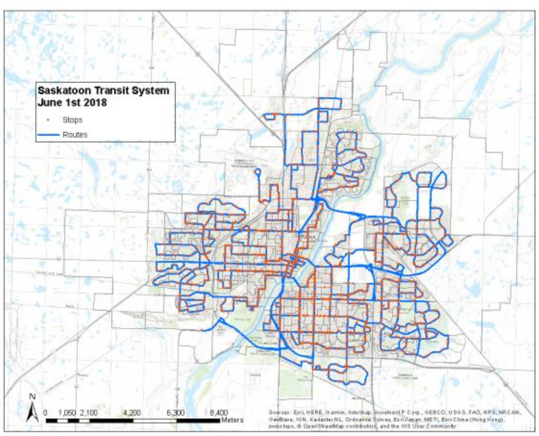


Figure 1. The June 1, 2018 Saskatoon Transit System Configuration.

Population data and transit ridership data per Dissemination Area (DA) was obtained from the latest Statistics Canada Data Census Population CSV per Dissemination Area report (2016). The percentage of riders in each DA is illustrated in Figure 2.

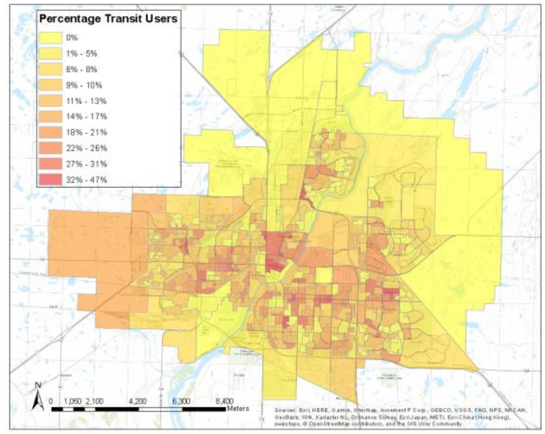


Figure 2. Percentage of transit users by DA from Statistics Canada (2016).

2.1 Service Area Partitioning

A transit system serves a geographical area. The smallest geographical sub-areas for which statistics such as population and transit ridership are available are DAs. The most recent data from Statistics Canada (2016) defines 362 DAs for Saskatoon. However, the DAs are not uniform: they range in size from 0.022 km² to 40.121 km². Some DAs are convex with externally located centroids. This is a typical Modifiable Areal Unit Problem (MAUP) issue, in which results can be skewed depending on the boundaries that are drawn to aggregate the data (Openshaw, 1983).

To overcome the MAUP issue, a grid of 100m by 100m cells was overlayed on the bounding box containing all Saskatoon DAs. and intersections computed for each DA. In many cases the grid cells were bisected by DA boundaries. That is, while there are many grid cells within DAs, the grid cells along the DA boundaries are usually clipped into smaller, non-square shapes. This operation resulted in 21,807 grid cells, each with an internal centroid.

Next, 400m network-constrained buffers were calculated around each bus stop. Buffers that intersected a grid cell were considered within the grid cell's catchment area. Accessibility measures were then computed for each grid cell. The average value of for the grid cells within a DA was then used to compute an accessibility measure for the DA as a whole.

2.2 Service Models

Transit access depends on the level of service. In this paper, a unit of service is defined as a departure, on a route, in a specific direction. In the analysis, the term *route* implies a *route-direction* combination. Service parameters can be evaluated in different ways by different models. In this paper we consider five model types:

1. *Stop Model*. This model counts the total number of stops within a demand location.
2. *Coverage Model*. This model counts the total number of routes serving all stops within a demand location. The service provided by each stop is determined by the number of different routes served.
3. *Frequency Model*. This model counts the departures from all stops within a demand location over some time interval (e.g., 1 hour).
4. *Filtered Coverage Model*. In this model, departures on the same route from stops farther away from the demand location are filtered (i.e., not considered).
5. *Filtered Frequency Model***.** In this model departures on the same route from stops farther away from the demand location are filtered (i.e., not considered).

Each one of the model types described above involves service locations at different distances from the demand location. In each case, a weighting W_{jk} based on the distance between service location j (bus stop) and demand location k (grid cell centroid) can be applied to the score contributed by each stop. Langford (2012) suggested a Butterworth filter given by:

$$W_{jk} = 1/\sqrt{1 + x(d_{jk}/d_{pass})^n}$$

with x=1, n=6 and d_{pass}=250m. For this paper, each of the model types described above was run with and without distance decay weighting. Distance decay was calculated with d_{pass}=250m and the network distance d_{jk} from the service location j to the demand location k.

2.3 Transit Enhanced Two Step Floating Catchment Area Model (E2SFCA)

The E2SFCA model, proposed by Langford et. al. (2012), is similar to the *Filtered Frequency Model* with distance decay. The service provided at service location j depends upon the demand location k. Therefore the service provided by service location j to demand location k is denoted S_{jk}. Using the *Filtered Frequency Model* with distance decay weighting, the accessibility measure at location k is given by:

$$A_k = \sum_{j \in \{B_{jk}\}} W_{jk} S_{jk}$$

where B_{jk} denotes the set of filtered service locations j that fall within demand location k's catchment buffer. The E2SFCA model differs from the *Filtered Frequency Model* by using a *service-to-demand* ratio R_{jk} in place of the service S_{jk} such that:

$$A_k = \sum_{j \in \{B_{jk}\}} W_{jk} R_{jk}$$

where:

$$R_{jk} = S_{jk}/D_j$$

D_j is the demand at service location j and is the sum of the weighed populations P of locations k that fall within location j's catchment buffer:

$$D_j = \sum_{k \in \{B_{kj}\}} W_{kj} P_k$$

In addition to the five model types described in Section 2.2, this paper also considered Langford's Transit Enhanced 2SFCA model described above. However, as shown in the results (Section 3), Langford's model performed poorly. Based on that poor performance, another model, dubbed the E2SFCA-2 model, was also considered.

In the E2SFCA-2 model, the demand (i.e., the population surrounding the service location) was treated as a potential supply of transit riders and was used to increase the service as shown in equation below.

$$R_{jk} = S_{jk}\sqrt{D_j}$$

2.4 Walk Score's Transit Score

Walk Score's Transit Score is a filtered frequency model that uses departures per week as its service metric (Walk Score, 2018). All departures on a route are ignored except for those from the stop located closest to the demand location. The score is computed for demand locations on a 500 foot grid. A distance decay, as shown in Figure 3, is then applied to the service scores. Distances are computed using the road network. Once computed, a log of the score is taken (Figure 4).

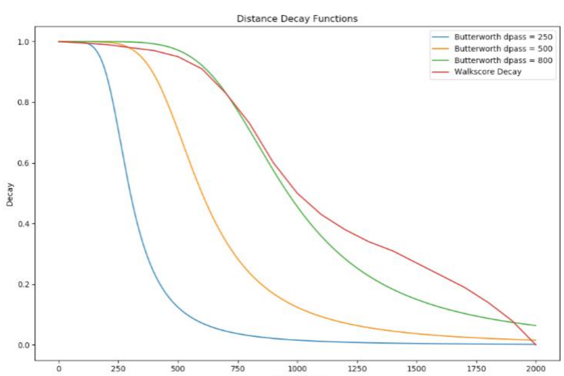


Figure 3. Butterworth distance decay functions and the Walk Score distance decay function.

2.5 Correlation with Transit Ridership

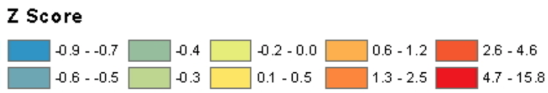
The DA data from Statistics Canada includes transit ridership estimates. Because of widely varying DA populations, transit ridership as a percentage of DA population was computed and used for correlation with the accessibility measures. The percentage of transit users by DA ranges from 0% in many smaller sized and less populated DAs to a maximum of 46.8%. To protect individual's privacy, the transit ridership estimates are intentionally coarse. The data reports a population of 246,376 with 24,980 transit users. A Pearson's correlation coefficient was computed to quantify the relationship between ridership percentage and each accessibility measure for Saskatoon at the DA level.

3. Results

Each of the five models identified in Section 2.2 was run twice: once with d_{pass}=250 m and once with no distance decay. Z-scores were calculated to allow visual comparison between figures. Figures 4 to 10 show the results with distance decay applied. Table 1 shows the correlation results for all models. As seen in the table, the best performance was obtained with the E2SFCA-2 (figure 10).

Table 1: Pearson's r correlation results

Model	Fig.	Distance Decay (d_{pass})	Pearson's
E2SFCA		250	0.036
Filtered Coverage			0.327
Transit Score	4	Walkscore	0.336
Filtered Coverage	5	250	0.342
Stop Count			0.345
Stop Count	6	250	0.349
Coverage			0.361
Coverage	7	250	0.364
Frequency			0.395
Filtered Frequency			0.398
Frequency	8	250	0.401
Filtered Frequency	9	250	0.418
E2SFCA-2	10	250	0.431



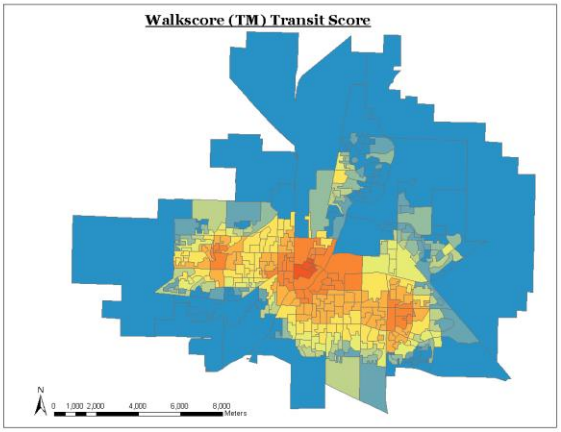


Figure 4. Walk Score's Transit Score

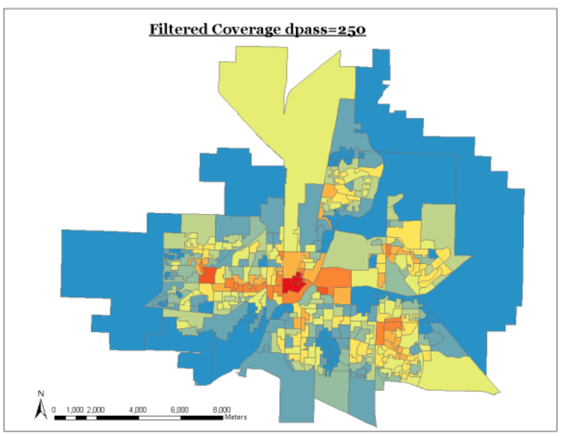


Figure 5. Filtered Coverage

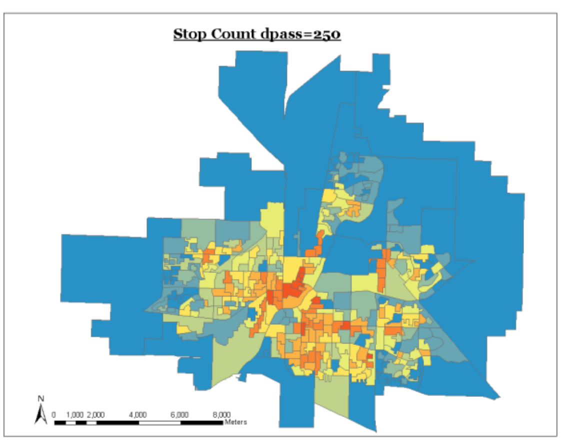


Figure 6. Stop Count

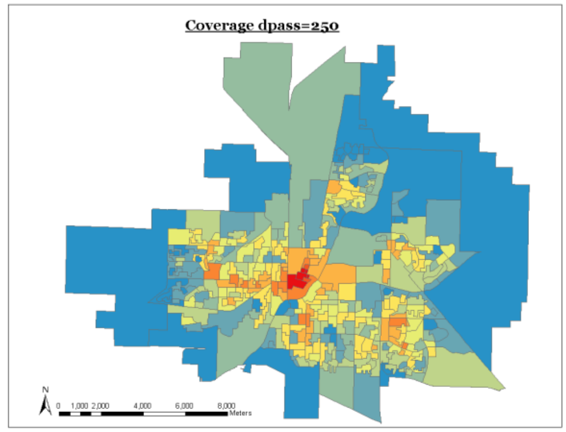


Figure 7. Coverage

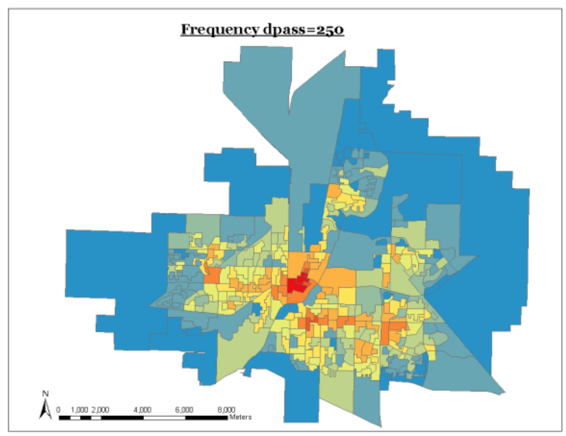


Figure 8. Frequency

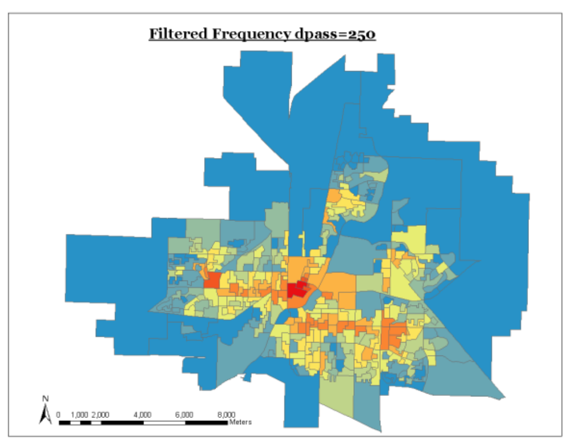


Figure 9. Filtered Frequency

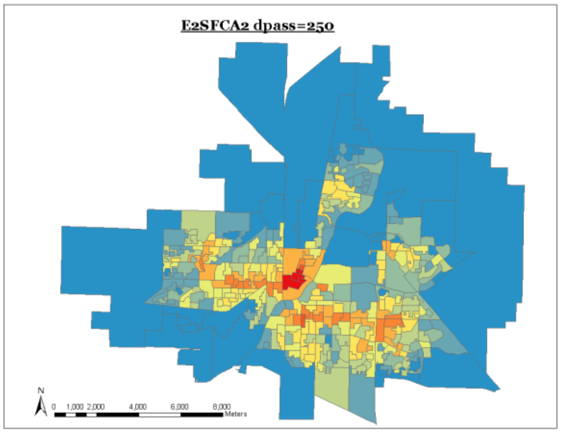


Figure 10 .E2SFC-2

A Pearson's r-value of 0.431 (the best obtained result) is not an especially strong correlation. However, it should be noted that there is a relationship between transit service supply and demand. While the demand for transit is influenced by its supply, transit supply itself is adjusted by transit agencies in response to demand changes over time (to provide an efficient service). In other words, demand and transit accessibility measures should be correlated.

In every case, a model with distance decay resulted in a better correlation. This confirms that distance to a bus stop is an important factor in accessibility. However, filtering the service when computing service levels has a mixed result. It improved the performance of the Frequency Model but decreased the performance of the Coverage Model.

As expected, the E2SFCA model performed poorly (Table 1), indicating no correlation between the model and actual transit ridership. Therefore, the E2SFCA was rerun with a modification that is labeled E2SFCA-2 in Table 1 and it resulted in the best performance of all the models considered. The choice of $\sqrt{Dj}$ was arbitrary and future work is required to determine how to best consider population demand in scenarios such as urban transit systems.

4. Conclusion

This paper examined accessibility models by correlating their scores with actual ridership data. Accessibility scores were computed for 362 DAs in Saskatoon and correlated with ridership as a percentage of the DA population using Pearson's r. In all cases, incorporating distance decay resulted in improved model performance. Service frequency, as a service metric, performed better than coverage, and a filtered frequency model, in which all but the closest departure locations were discarded, improved performance even more. The E2SFCA algorithm resulted in the worst performance when treating population as a demand that reduced accessibility. Conversely, that model performed best when modified to treat population as a supply.

It should be noted that many factors beyond physical accessibility have an impact on transit ridership. Therefore, generating statistical models to isolate such impacts is the next step of this research. Although transit system accessibility may be good in low population locations (e.g., centers of work, study, and employment), the transit ridership data collected by Statistics Canada links the users to their home DAs. Therefore, perhaps using origin and destination data could be recommended for future research.

Based on preliminary results, investing in higher frequency service rather than expanded coverage might result in greater transit ridership gains per dollar spent. Saskatoon's proposed BRT system prioritizes frequency over coverage.

Future work based on the results of this paper include treating population as a supply rather than a demand. Factors such as the catchment buffer size and distance decay parameters could also be varied. A model tuned to maximize performance for a particular transit system could be used to predict ridership changes in that system when the system is reconfigured, such as Saskatoon's proposed BRT configuration.

Acknowledgements

The authors would like to thank the Interact Team (https://teaminteract.ca) for inspiration and funding; and Michael Bree for help with Python.

References

City of Saskatoon, and Robert Dudiak. 2018. *"Transit Plan"* Technical Report. City of Saskatoon. https://www.saskatoon.ca/engage/transit-plan

Dai, Dajun. 2010. *"Black Residential Segregation, Disparities in Spatial Access to Health Care Facilities, and Late-Stage Breast Cancer Diagnosis in Metropolitan Detroit"*. Health and Place 16 (5). ISSN: 1353-8292.

Ding, Jishiyu, Yi Zhang, and Li Li. 2018. *"A New Accessibility Measure of Bus Transit Networks"*. IET Intelligent Transport Systems 12 (7): 682-688. ISSN: 1751-956X. DOI: 10.1049/iet-its.2017.0286

Kim, Hyun, and Yena Song. 2018. *"An Integrated Measure of Accessibility and Reliability of Mass Transit Systems"*. Transportation 45, no. 4 (July): 1075-1100. ISSN: 1572-9435. DOI: 10.1007/s11116-018-9866-7 https://doi.org/10.1007/s11116-018-9866-7

Langford, M., R. Fry, and G. Higgs. 2012. *"Measuring Transit System Accessibility using a Modified Two-Step Floating Catchment Technique"*. International Journal of Geographical Information Science 26 (2): 193-214. DOI: 10.1080/13658816.2011.574140 https://doi.org/10.1080/13658816.2011.574140

Lei, T. L., and R. L. Church. 2010. *"Mapping Transit-Based Access: Integrating GIS, Routes and Schedules"*. International Journal of Geographical Information Science 24 (2): 283-304. ISSN: 1365-8816. DOI: 10.1080/13658810902835404

Litman, Todd. 2012. *"Evaluating Public Transit Benefits and Costs: Best Practices Guidebook"*. http://library.usask.ca/scripts/remote?URL=http://www.deslibris.ca/ID/230688

Luo, W., and F. Wang. 2003. *"Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a case study in the Chicago region"*. Environment and Planning B: Planning and Design 30 (6): 865-884. ISSN: 02658135

Luo, Wei, and Yi Qi. 2009. *"An Enhanced Two-Step Floating Catchment Area (E2SFCA) Method for Measuring Spatial Accessibility to Primary Care Physicians"*. Health and Place 15 (4): 1100-1107. issn: 1353-8292

McGrail, Matthew R., and John S. Humphreys. 2009. *"Measuring Spatial Accessibility to Primary Care in Rural Areas: Improving the Effectiveness of the Two-Step Floating Catchment Area Method"*. Applied Geography 29 (4): 533-541. ISSN: 0143-6228

Morris, J., Dumble, P., and Wigan, M. 1979. "Accessibility indicators for transport planning". Transportation Research Part A: General, 13(2), 91-109

Openshaw, S., 1983. *The Modifiable Areal Unit Problem.* Norwick: Geo Books. *ISBN 0860941345. OCLC 12052482*

Thill, Jean-Claude, and Marim Kim. 2005. *"Trip Making, Induced Travel Demand, and Accessibility"*. Journal of Geographical Systems 7 (2): 229-248. ISSN: 1435-5930.DOI: 10.1007/s10109-005-0158-3

Walkscore. 2018. *"Walkscore's Transit Score Methodology"*. https://www.walkscore.com/transit-score-methodology.shtml.