# *Analysis of demand–supply gaps in public transit systems based on census and GTFS data: a case study of Calgary, Canada*



**As Published:** https://doi.org/10.1007/s12469-020-00252-y

**Publisher:** Springer Berlin Heidelberg

**Persistent URL:** https://hdl.handle.net/1721.1/131390

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

# Analysis of demand–supply gaps in public transit systems based on census and GTFS data: a case study of Calgary, Canada

**Cite this article as:** Koragot Kaeoruean, Santi Phithakkitnukoon, Merkebe Getachew Demissie, Lina Kattan and Carlo Ratti, Analysis of demand–supply gaps in public transit systems based on census and GTFS data: a case study of Calgary, Canada, Public Transport https://doi.org/10.1007/s12469-020-00252-y

# Analysis of demand-supply gaps in public transit systems based on census and GTFS data: A case study of Calgary, Canada

**Abstract:** Bridging the gap between demand and supply in transit service is crucial for public transportation management, as planning actions can be implemented to generate supply in high demand areas or to improve upon inefficient deployment of transit service in low transit demand areas. This study aims to introduce feasible approaches for measuring gap types 1 and 2. Gap type 1 measures the gap between public transit capacity and the number of public transit riders per area, while gap type 2 measures the gap between demand and supply as a normalized index. Gap type 1 provides a value that is more realistic than the gap type 2, but it requires detailed passenger data that is not always readily available. Gap type 2 is a practical alternative when the detailed passenger data is unavailable because it uses a weighting scheme to estimate demand values. It also uses a newly proposed normalization method called M-score, which allows for a longitudinal gap analysis where yearly gap patterns and trends can be observed and compared. A five-year gap analysis of Calgary transit is used as a case study. This work presents a new perspective of hourly gaps and proposes a gap measurement approach that contributes to public transit system planning and service improvement.

Key words: public transit system; supply-demand gap; GTFS; supply index; demand index

**Koragot Kaeoruean**
Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Thailand
koragot_kaeoruean@cmu.ac.th

**Santi Phithakkitnukoon* (ORCID: 0000-0002-5716-9363)**
Excellence Center in Infrastructure Technology and Transportation Engineering, Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Thailand
santi@eng.cmu.ac.th
*corresponding author

**Merkebe Getachew Demissie**
Department of Civil Engineering, University of Calgary, Canada
merkebe.demissie@ucalgary.ca

**Lina Kattan**
Department of Civil Engineering, University of Calgary, Canada
lkattan@ucalgary.ca

**Carlo Ratti**
Senseable City Lab, Massachusetts Institute of Technology, USA

ratti@mit.edu

## 1 Introduction

Public transportation plays an important role in modern society. Public transportation systems provide many benefits to individuals, communities, and the local economy. According to the American Public Transportation Association, or APTA, every dollar invested in public transportation generates $4 in economic returns, and every $1 billion invested in the sector creates and supports 50,000 jobs. An investment of $10 million in public transportation generates about $32 million in increased business sales (APTA, 2018). Homes located near public transit with high frequency service have residential property values 42% higher than the average. Public transportation produces less air pollution per passenger mile than a standard private car carrying a single driver (DART First State, 2018). Since public transportation can convey people in much less space than would be required for individual automobiles, it helps to lower traffic congestion. This, in turn, reduces the stress that comes from daily driving in highly congested areas. Using public transportation can free up a driver's time and attention. On average, daily commuters spend 35 minutes driving to work in their cars, which adds up to 152 hours or 19 full work days a year (Knoblauch, 2019). This time can be freed up by having someone else doing the driving, allowing riders to spend time reading, working, or studying instead of paying attention on the road.

While public transportation brings several benefits to its users, its operation and service quality can produce serious negative impacts. Service disruptions and unexpected transit delays can lead to lost wages or other consequences for missed time at work. Adding time and delay to peoples' commutes can put the brakes on the economic output of public transit.

Accessibility of transit services within a reasonable walking distance of an individual's origin and destination is one of the primary factors in peoples' decisions to use transit. Another factor is service temporal coverage. Over the past few years, many cities have experienced a decline in public transit ridership (Bohte & Maat, 2009)(O'Toole, 2018a)(O'Toole, 2018b)(Advancing Public Transport, 2017). Failing to meet growing public transit needs is one of the reasons for declined ridership. An increasing world population due to rapid urbanization has had a dramatic effect on the increased demand for jobs, housing, energy, and transportation services. The growing number of workers generates a larger travel demand and creates an emerging public transit need in new areas. It also results in changes in travel behavior and demand due to the built environment (Parady, Takami, & Harata, 2017).

It is critically important to balance the public transit supply and demand to minimize the gap between the level of transit service (supply) and the transit service need (demand). The balanced result is an efficient public transit service that meets demand to improve ridership. Areas that are undersupplied (i.e., places that lack necessary public transit service) and oversupplied (i.e., places with excessive public transit service) must be identified to create effective planning responses that will allow public transportation to meet the required demand while optimizing operational costs.

Limited availability of transit data has been one of the major impediments to empirical analyses and theoretical advances for the characterization and modeling of transit service demand and supply (Prommaharaj, Phithakkitnukoon, Demissie, Kattan, & Ratti, 2020). In recent years, however, transit agencies have begun to produce large volumes of data as part of their daily operations. These data are obtained from multiple collection streams: on-board sensors introduced by Automated Vehicle Location (AVL) systems to track the positions of transit vehicles; Automated Passenger Counting (APC) systems to record the number of boarding and alighting passengers; and Automated Fare Collection (AFC) systems to record trips taken by users and facilitate an integrated ticketing system across transport modes and operators (Dickens & Hughes-Cromwick, 2019). Transit agencies also utilize opportunistic sensing datasets produced from other sources, including mobile phone networks (Demissie et al., 2016), social media (Zannat & Choudhury, 2019), and transit apps ("Transit app," 2020). Although the AFC, APC, and AVL datasets provide detailed information that allow transit agencies to track transit ridership trends and improve transit service operation and planning, these datasets are usually shared in proprietary formats specific to each vendor or agency (Barbeau, 2017). In the case of Calgary Transit, neither the bus nor the LRT systems have smartcard ticketing systems (e.g., AFC).

Recently, Google has introduced GTFS realtime, a real-time counterpart to GTFS static. This feed specification allows public transportation agencies to provide realtime trip information like updates, service alerts, and vehicle positions. GTFS realtime is relatively new and faces various challenges such as a lack of clear documentation and openly available validation tools (Dawoodjee, 2020); large data sets in need of frequent updates, which makes manual inspection time-prohibitive (Barbeau, 2017); and data that is difficult to access and that lacks available time series (Altman, 2011). This study uses GTFS static data, which is distributed in a common format for public transportation schedules with the associated geographic information. Compared to GTFS realtime, GTFS static data has the advantage of characterizing the transit supply as it is intended to operate. It also contains long term time series information, which allows for a time series analysis of the data.

A gap is the difference between a transit system's supply and demand values, so measuring the gap requires data sources for both values. Supply can be estimated from the transit service accessibility and frequency (Currie, 2010) or from the number of transit stations and routes, as well as length of routes in an area (Jiao & Dillivan, 2013) (Jiao, 2017). Transit demand can be approximated from the number of people in the area who are socially disadvantaged (Currie, 2010) or the population of people who are transit dependent, i.e., people that cannot drive because they are too young, too old, too poor, or physically unable (Jiao & Dillivan, 2013). Since both supply and demand values are often represented in different units, a simple value subtraction cannot be used to directly calculate the gap between them. We must first project the values onto the same scale for comparison by performing a value normalization using a method such as min-max (Currie, 2010) (Toms & Song, 2016) or z-score (Jiao & Dillivan, 2013) (Jiao, 2017).

The work presented in this paper builds on the state-of-the-art and extends beyond the concepts of transit deserts (Jiao & Dillivan, 2013) and transport/social needs systems (Currie, 2010) that encompass undersupplied areas without adequate public transit service. It is equally important to consider and understand the other end of the service supply level, oversupplied areas. This research explores both under- and oversupplied areas to facilitate a discussion of effective transportation systems that provide an adequate level of service to all areas.

We present two approaches for measuring transit demand-supply gap. The first approach (type 1) does not require a normalization process since it yields a comparative unit across the demand, supply, and gap values: the number of transit users per area. This unit is a much more realistic and applicable quantity than a traditional index value that requires further interpretation. When detailed information about the number of transit users unavailable, the first approach cannot be applied. The second approach (type 2) uses a weighting scheme and census data to estimate the proportion of the transit user population. It also leverages a newly proposed normalization method to measure the gap value across different areas. The advantage of the second approach is that its population weighting scheme enables a longitudinal gap analysis where yearly gap patterns and trends can be observed and compared. We demonstrate this by creating a comparative study of gaps over a five-year period.

The gap between demand and supply varies not only across areas of transit service, but also over all time scales, including yearly, monthly, and even hourly. It is important to capture the hourly gap variation for strategic planning of transit service. Therefore, both proposed approaches introduce an hourly gap that provides a new temporal dimension for gap analysis that has not yet been explored before in the literature. We believe that our work contributes to the literature with its new perspectives on the gap analysis, which is essential for improving the efficacy of public transportation systems.

Our methodology is described in the next section, which also provides an overview of the analysis, details of datasets used in the study, and an explanation of the gap measurement. Section 3 presents the results of our gap analysis. The paper concludes with a summary, discussed limitations, and an outlook on possible future research.



# 2 Methodology

## 2.1 Overview

Our gap analysis study used the Calgary Transit System in Calgary, Canada as a case study. Our approaches leverage three main datasets: Canada Census (2016), Calgary Census (2013-2017), and GTFS static feeds. These datasets are shown in a flow diagram in Figure 1. The Canada Census supplements information from the Calgary census because it contains essential attributes like the number of public transit users in the country. This information is aggregated at the level of dissemination areas (DAs). To measure the gap between a public transit system's supply and demand, we must first estimate those values.

We use two different approaches: *gap type 1* and *gap type 2*. The demand value for each approach is based on the census datasets (Canada and Calgary). Since both censuses are reported in different geographic units, we must align the data sources by mapping between the two censuses. The Canada Census divides land into DAs, while the Calgary census is separated into community boundaries. We mapped the DAs into community boundaries so that the gap types 1 and 2 were comparable. This also allowed us to conduct a gap analysis per community class and sector.

Once the data have been remapped, the Canada Census data can be used to calculate demand type 1 for each Calgary community based on the passenger data (labor force, the number of public transit users overall, and the number of transit users per hour). The unit for this data is the number of public transit riders. Supply type 1 is calculated based on public transit trips extracted from GTFS data. All transit stops are mapped into a community, which allows us to calculate the number of trips per community. Supply type 1 is derived from the public transit vehicles' capacities (i.e., the number of passengers the vehicles can hold). Gap type 1 can be measured as the number of public transit riders.

In the case where detailed passenger data is unavailable and demand type 1 therefore cannot be calculated, demand type 2 is an alternative option. Typically, passenger data is collected every five years. Demand type 2 uses passenger data from a different time period to project the demand for the time period where the passenger data is unavailable. The result is a weighted calculation of public transit users as a ratio of public transit users to labor force. Hourly demand can be estimated in a similar fashion.

There are also two approaches for estimating public transit supply, types 1 and 2, which can be compared to demand types 1 and 2, respectively. Public transit trips are extracted from GTFS feeds and are mapped into communities based on transit stop locations. The number of trips per community is then calculated based on the estimate for supply type 1 according to the capacities of the public transit vehicles. The unit of this calculation's result is the number of riders, which aligns supply type 1 with its counterpart, demand type 1. Supply type 2 is an estimate based on the number of public transit vehicle visits per stop, per community. Since both demand type 2 and supply type 2 are in different units, their values must be normalized before gap type 2 can be calculated, which results in the unit for gap type 2 being a normalized index value.

When detailed passenger data is available, gap type 1 has the advantage of providing a more realistic and applicable value than gap type 2. Nonetheless, gap type 2 offers a practical alternative approach for a longitudinal gap analysis since it can be derived from projected transit demand over a time span. For the purposes of our case study, gap type 2 is calculated for a five-year period (2013-2017), while gap type 1 is calculated only for 2016, which is when the detailed passenger data is available.

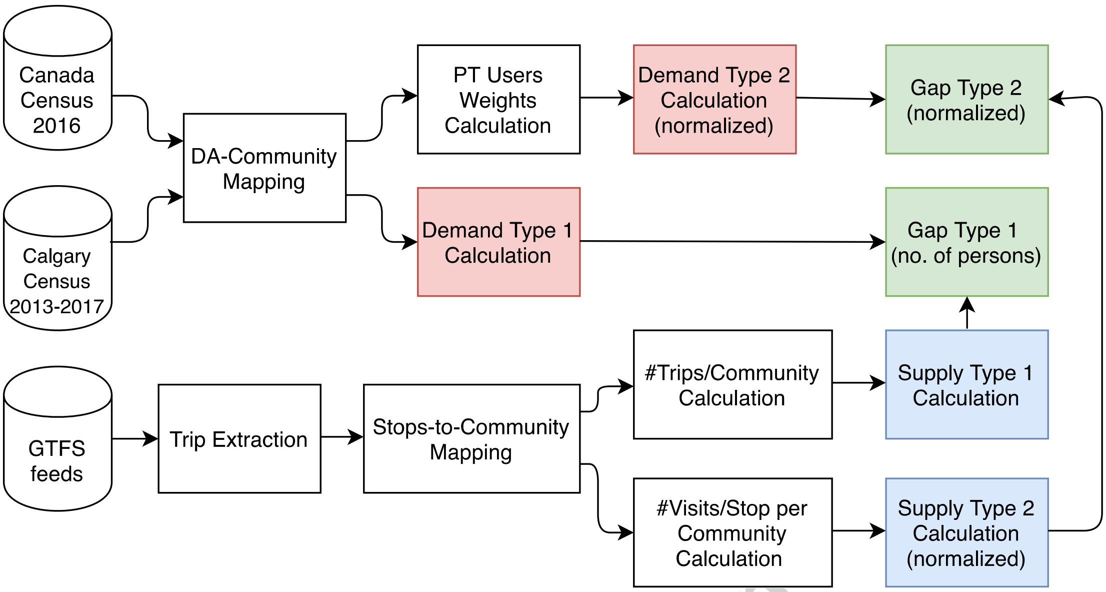



**Figure 1. Gap measurement flow diagram (although demand type 1 can be calculated directly with Canada Census data, we require the DA-Community mapping to compare both gap types)**

## 2.2 Datasets

Three datasets are used in this study: GTFS static feeds, Canada Census, and Calgary Census.

### 2.2.1 GTFS Feeds

The General Transit Feed Specifications (Google, 2012), or GTFS, defines a common format for public transportation schedules and their associated geographic information. There are seven *.txt* file categories, each of which contains different metadata about the Calgary Transit System:

1. *agency.txt* contains information about the transit agency, including the agency's name, URL, and the time zone of the transit agency's location.
2. *stops.txt* contains information about transit stops or stations, including the stop's ID, name, latitude, and longitude.
3. *routes.txt* contains information about transit routes, including the route's ID, name, and type (e.g., rail or bus).
4. *trips.txt* contains information about the journey made from an origin point to each transit vehicle's final stop, including the trip ID, the associated route ID, and a set of dates when service is available for those routes.
5. *stop_times.txt* contains information about trip arrival and departure times, including stop ID, associated trip ID, arrival time, departure time, and the order of stops for each trip.
6. *calendar.txt* contains information about dates of service based on a weekly schedule, including the day of the week, the start date, and the end date.
7. *shapes.txt* contains the path that a vehicle travels along a route alignment. Shapes are associated with trips and consist of a sequence of points through which the vehicle passes in the order in which they are reached. Its attributes include the shape point's latitude and longitude, the sequence in

which the shape points connect to form the shape, and the distance traveled along each segment of the shape.

All seven file categories are clearly related. When they are explored together, they describe Calgary's public transit system in detail. The relational mapping of all seven file categories of Calgary Transit's GTFS feeds is shown in Figure 2.

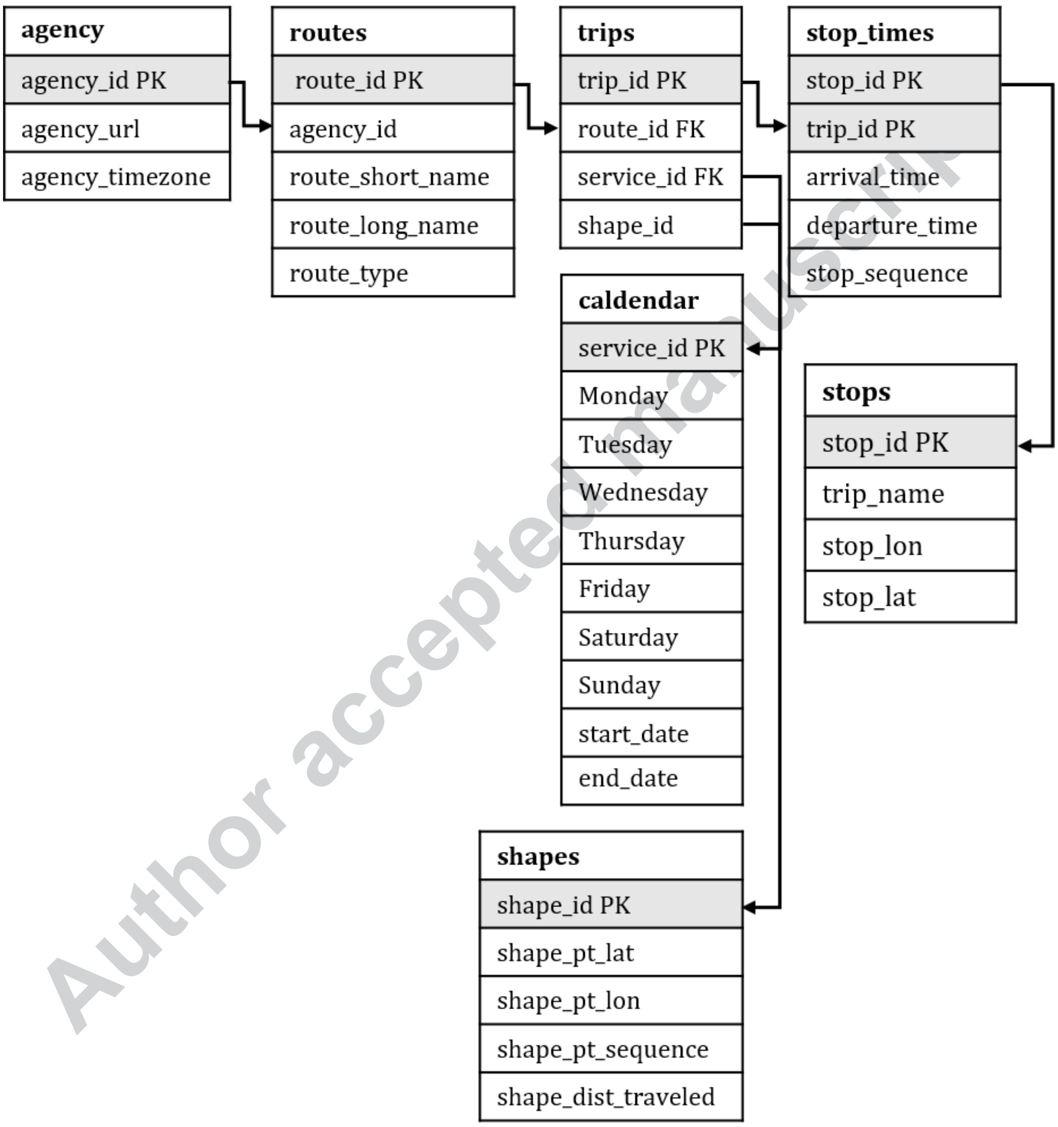



**Figure 2. Relational table of GTFS data**

From time to time, new developments and changes in infrastructure and scheduling require updates to the public transit system. These changes are reflected in the GTFS feeds. For this study, we obtained the GTFS data and all included updates from Calgary Transit for the 41 time periods that fell between June 2013 and July 2017.

For the purposes of this study, we only considered weekday transit service. A representative service date is selected for each weekday for each year. Transit schedules on holidays differ from regular weekday schedules, so we selected only GTFS feeds from regular weekdays for our analysis. Table 1 shows the service operation dates of each set of GTFS feeds that were selected to represent the transit service of that day of the week. We carried out our analysis using the data from the selected dates shown in Table 1.

**Table 1: Dates of service operation of the representative sets of GTFS feeds**

| Year | Day of the week | | | | |
|---|---|---|---|---|---|
| | Monday | Tuesday | Wednesday | Thursday | Friday |
| 2013 | 18/11/2013 | 19/11/2013 | 13/11/2013 | 14/11/2013 | 15/11/2013 |
| 2014 | 17/11/2014 | 18/11/2014 | 19/11/2014 | 13/11/2014 | 14/11/2014 |
| 2015 | 16/11/2015 | 17/11/2015 | 18/11/2015 | 19/11/2015 | 13/11/2015 |
| 2016 | 14/11/2016 | 15/11/2016 | 16/11/2016 | 17/11/2016 | 18/11/2016 |
| 2017 | 13/11/2017 | 14/11/2017 | 15/11/2017 | 16/11/2017 | 17/11/2017 |



Figure 3 shows the locations of both types of transit stops that are active during the morning hours (5 a.m.-9 a.m.) for each date listed in Table 1 over the five-year span from 2013 to 2017. The number of active stops gradually increases over time. Figure 4a shows the average number of hourly stop visits by transit vehicles during the morning hours to show the increase in service frequency. The average number of stop visits is calculated by taking the number of visits by transit vehicles to each stop over a one-hour period and then averaging that number over all stops such that the unit of this variable is visits/hour. Figure 4b shows the average number of trips by transit vehicles per morning hour. This is calculated by taking the number of trips made by all transit vehicles for an hour and finding the average such that the unit is trips/hour. Figure 5 shows the total number of unique routes (based on unique route ids) of each transit type during the morning hours of the five-year study period. The average number of hourly stop visits shown in Figure 4 are similar over time, but a smaller number of unique transit routes (Figure 5) and hourly trips (Figure 4b) were observed in 2013 than in 2017. The existing LRT infrastructure was built and opened for service prior to this study's start point of 2013. The LRT system consists of two routes, with each route operating in two lines. These routes remained similar throughout the duration of the study period (Figure 5).

Calgary Transit has experienced unprecedented ridership growth. The need for additional LRT lines to accommodate ridership growth and to serve newly developing areas has been identified. The West LRT, which consists of six new LRT stations and over 8 km of double track, is the most recent large infrastructure project that has been undertaken in Calgary. It was opened for service in December 2012 (Calgary Transit, 2014). In addition to the new LRT infrastructure, a network of 32 bus routes was modified into 25 new and revised bus routes in December 2012 (Calgary Transit, 2014).

Historical GTFS static data from Calgary Transit is available from November 2013 to present ("OpenMobilityData," 2020). Our analysis is based on GTFS data from 2013 to 2017. Data from November 2013 is used to capture the transit supply-demand gap at this early date to enable a comparison between the earliest and latest transit service years.

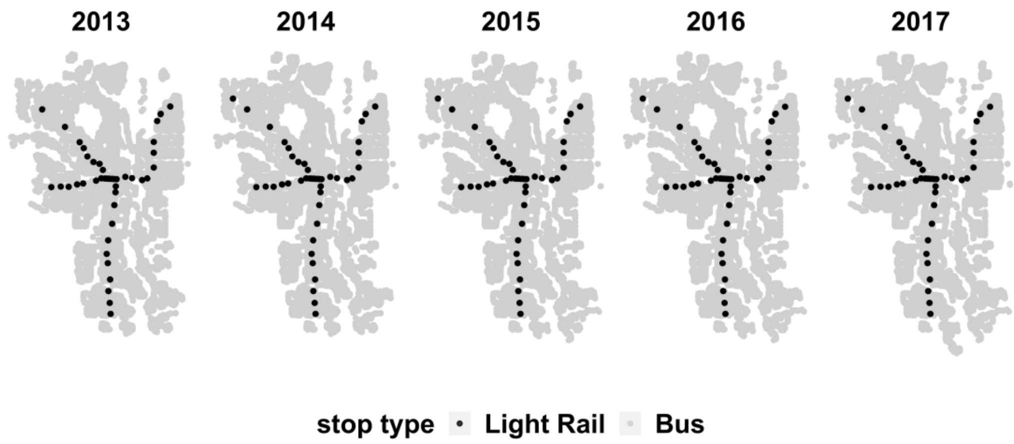



Figure 3. Locations of Calgary public transit stops from 2013 to 2017

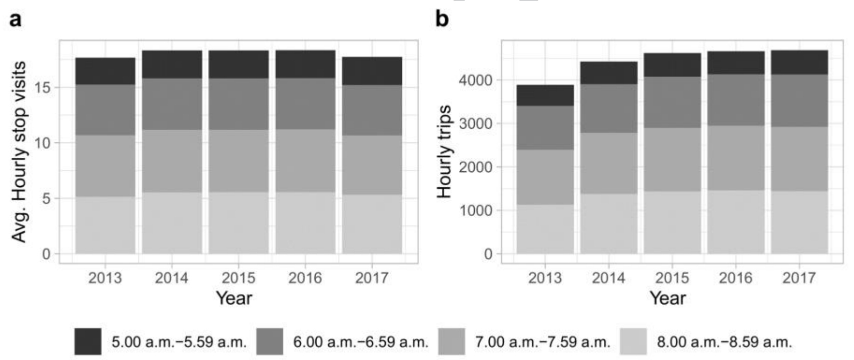



Figure 4. (a) Average number of hourly stop visits and (b) average number of hourly trips

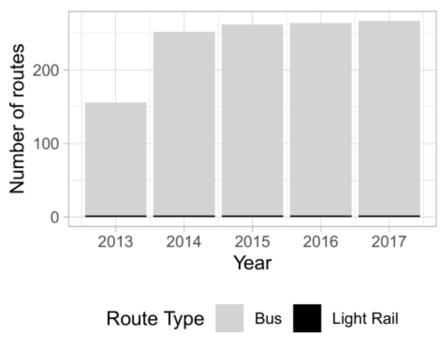



Figure 5. Total number of unique routes by type from 2013 to 2017

### 2.2.2 Canada Census

The 2016 Canada Census provides statistical information about Canadian residents. It covers 14 topics, including age, sex, income, commuting information, type of labor, etc. The census encompasses 10 provinces and three Canadian territories. It is conducted by the Statistics Canada every five years (Statistics Canada, 2012).

The dissemination area (DA) is the smallest standard geographic area for the Canada Census. Designed by population size rather than by area, DAs cover the entirety of Canada. Each DA contains an average of 400-700 residents. DA boundaries follow census subdivisions and census tracts boundaries, as well as roads and railways (Statistics Canada, 2016a).

This study uses labor force information from the Canada Census. This information encompasses the population aged 15 years and over, the number of public transit commuters, and the Journey to Work statistic, which reports on the total number of people leaving for work each hour from 5 a.m. and 9 a.m.

### 2.2.3 Calgary Census

Calgary is a city in southern Alberta with an area of 825.56 $km^2$ (Statistics Canada, 2016b; Statistics Canada, 2001). The largest city in Alberta and the third-largest municipality in Canada (Statistics Canada, 2016b), Calgary has a 74.6% participation rate in labor force and 10.3% unemployment rate. Its GDP is US$ 97.9 billion (Statistics Canada, 2001).

The Calgary Census is conducted by the community (Calgary Open Data, 2016) to create an official count of dwelling units and the population within those units. The population data includes the number of residents classified by age and gender. The boundary for the data is community districts within Calgary. There are a total of 298 communities and 100 attributes.

The Calgary Census also provides geographic information about the community. Figure 6 shows Calgary's community districts divided into different land-use classes and area sectors. There are four major land-use classes: Industrial, Major Park, Residential, and Residential Sub Area. The eight community sectors are: Centre, East, North, Northeast, Northwest, South, Southeast, and West. In this study, we assume that the total number of residents aged 15 and over is the available labor force in each community.

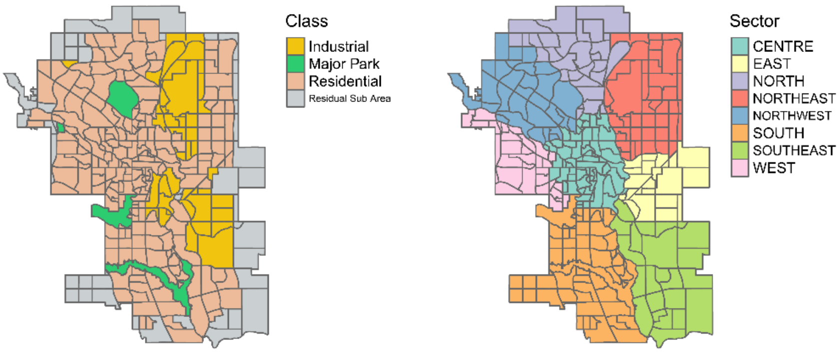



Figure 6. Calgary's community districts divided into different classes and sectors

### 2.2.4 DA-Community Data Mapping

To compare the gaps that are determined by each of the two approaches (types 1 and 2), we must remap the data so that they use the same geographic boundaries. The Canada Census uses DAs, while the Calgary Census uses community areas. Mapping also enables us to conduct a gap analysis per community class and sector, so we need to remap the 1,594 DA sub areas into the 298 communities, as shown in Figure 7.

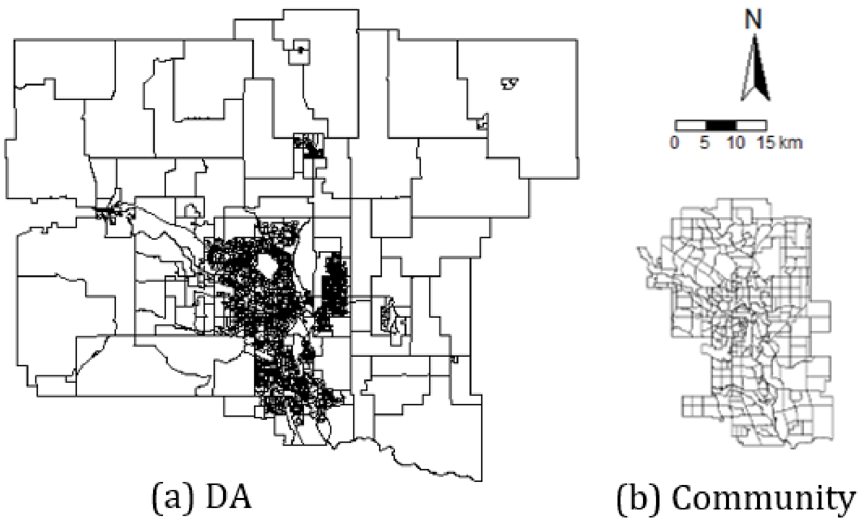



Figure 7. Boundaries of the Canada Census DAs and the Calgary Census community areas

Our approach is to use the centroid of the DA sub area as a reference point. The DA sub area is mapped into the community area within which its centroid is located. There are three scenarios in this approach. In the first scenario, the DA centroid is outside of the community area boundaries. In this case, those DA centroids that do not overlap with the community area are simply taken out. In the second scenario, the DA sub area overlaps with multiple community areas. This is addressed by mapping the overlapped DA into those communities. The third scenario occurs when multiple DA centroids overlap with a single community. This scenario is resolved by mapping all overlapped DA sub areas that community area. All three mapping scenarios are demonstrated in Figure 8.

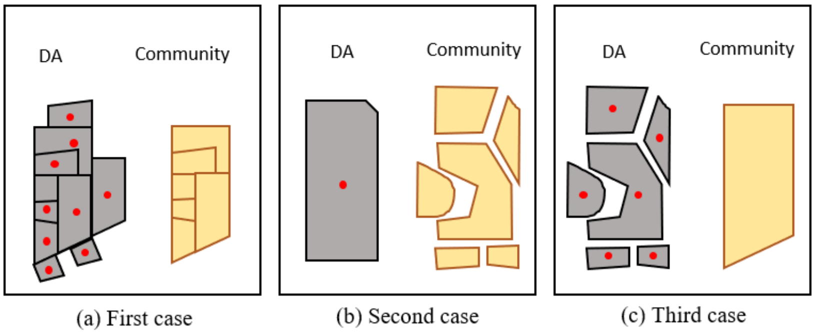



Figure 8. DA-community area mapping scenarios

Once the mapping is completed, census data like public transit passenger information per community area can then be gathered for our gap analysis.

## 2.3 Gap Measurement

A gap occurs when there is a difference between the values of transit demand and transit supply. The gap is the difference between the two values. If the demand and supply values are in the same unit, the gap can be determined through simple subtraction. If the units are different, then both values must be normalized before the gap calculation can be completed.

This research presents two approaches to the gap calculation, called gap type 1 and gap type 2. Both approaches yield an hourly gap that provides a temporal dynamism for the gap analysis beyond the state-of-the-art (i.e., a static gap). The calculation for gap type 1 requires detailed passenger information to determine the number of riders (demand) compared to the available space (supply). Gap type 2 is an alternative to gap type 1 that can be used when detailed passenger data is unavailable. It describes the gap in the form of a normalized index. Both gap types are classified as *undersupplied* when supply is less than demand, and as *oversupplied* when supply is greater than demand.

### 2.3.1 Gap Type 1

Gap type 1 uses the 2016 Canada Census data, which includes detailed public transit passenger information. Demand type 1 is calculated as a ratio between the number of public transit passengers who are leaving hourly for work and the area of community in which the transit service is operating (Currie, 2010). This is shown in Eq. (1).

$$D1_t = U_t \cdot W_c^{PC} / A_c \quad (1)$$

where $D1_t$ is demand type 1 during hour $t$ in the unit of persons/km$^2$, $U_t$ is the number of commuters who are leaving for work during hour $t$, $A_c$ is the area of community $c$

in $km^2$ where c = 1, 2, 3, …, 298 in our case, and $W_c^{PC}$ is a weight for public transit users in community $c$. $W_c^{PC}$ can be calculated as the ratio of the number of public transit users ($PT_c$) to the labor force ($L_c$) of community $c$:

$$W_c^{PC} = PT_c / L_c \tag{2}$$

To calculate supply, the Calgary Transit's capacity by vehicle information is used to estimate the transit service capacity per hour (Calgary Transit, 2002). The light rail's capacity is 195 riders per trip, while a bus's capacity is 80 riders per trip. Supply type 1 is thus defined by Eq. (3).

$$S1_t = P_{c,t} \cdot C_v / A_c \tag{3}$$

where $S1_t$ is the supply type 1 during hour $t$ (in the unit of persons/$km^2$), $P_{c,t}$ is the number of public transit trips within community $c$ during hour $t$, $C_v$ is the capacity of vehicle type $v$ (light rail or bus in our case), and $A_c$ is the area of community $c$ in $km^2$.

Gap type 1 ($G1_t$) at time $t$ can therefore be calculated in the unit of persons/$km^2$ as the difference between $S1$ and $D1$, as defined in Eq. (4). If $G1_t$ is positive, then the area is means oversupplied. A negative gap indicates that an area is undersupplied.

$$G1_t = S1_t - D1_t \tag{4}$$

### 2.3.2 Gap Type 2

Gap type 2 is an alternative calculation that can be used when the detailed passenger information, which is typically collected for every five years, is unavailable for the study's period of interest. In our case, commuter data for people that leave for work during hour $t$ (or $U_t$) is only available for 2016. If we wish to perform the same gap analysis for the other years of our study period, then we must make use of a combination of that year's labor force data and extrapolate the passenger information from the available year (i.e., 2016 in our case).

Demand type 2 can be defined as a ratio between the number of people in the labor force who are using public transit to leave for work at time $t$ and the area of community (Currie, 2010). This yields a result in the unit of persons/$km^2$, as shown in Eq. (5).

$$D2_t = L_c \cdot W'^{PC}_c \cdot W'^{PT}_t / A_c \tag{5}$$

where $D2_t$ is demand type 2 during hour $t$, $L_c$ is the current year's labor force (obtained for this research from the 2013-2017 Calgary Census), $W'^{PC}_c$ is a weight for the public transit users in community $c$ as defined in Eq. (2) from the 2016 Canada Census data, $A_c$ is the area of community $c$ in $km^2$, and $W'^{PT}_t$ is a weight for public transit users who are leaving for work at time $t$. The $W'^{PT}_t$ variable is based on the 2016 Canada Census data where the passenger data is available, and is the ratio between the number of commuters who are leaving for work at time $t$ and the total numbers of commuters who are leaving for work, as shown in Eq. (6).

$$W'^{PT}_t = U'_t / \sum_t U'_t \tag{6}$$

Note that an apostrophe on the variable indicates that its value is derived from census data that makes detailed passenger available, which in the case of this research is the 2016 Canada Census. This data is an estimate, so it may be different from the year in which the gap is analyzed.

Supply type 2 estimates the hourly level of service provided in each community area. It is calculated based on Currie's method (Currie, 2010), which considers the access distance and service frequency of public transit. The access distance for each public transit station or stop is defined as 0.4 km for a bus stop and 0.8 km for a rail station. A circle with the appropriate radius for the type of transit service is drawn around each public transit station. The areas of these circles are calculated in the unit of $km^2$ and then aggregated for the entire community area. Thus, supply type 2 is the sum of a community's access area multiplied by the service frequency within the community area, as shown in Eq. (7).

$$S2_t = \sum_{i=1}^{N} \left( \frac{\alpha_{c,t,i}}{A_c} \cdot F_{c,t,i} \right) \tag{7}$$

where $S2_t$ is supply type 2 at time $t$, $A_c$ is the area of community $c$ in $km^2$, $\alpha_{c,t,i}$ is an access area of transit station $i$ within community $c$ during time $t$ where $i$ = 1, 2, 3, ..., $N$, and $N$ is the total number of stops (or stations) within the community area, and $F_{c,t,i}$ is the service frequency of the station $i$ within the community $c$ during time $t$. Thus, the unit for $S2_t$ is the number of stop visits.

Since supply type 2 and demand type 2 are calculated in different units (i.e., persons/$km^2$ versus stop visits), we cannot use simple subtraction to calculate gap type 2. We must therefore normalize the units before we can use Eq. (8) where accent mark '$\widehat{\ }$' denotes a normalized value. A positive $G2_t$ implies an oversupplied gap value and a negative $G2_t$ implies an undersupplied gap value.

$$G2_t = \widehat{S2}_t - \widehat{D2}_t \tag{8}$$

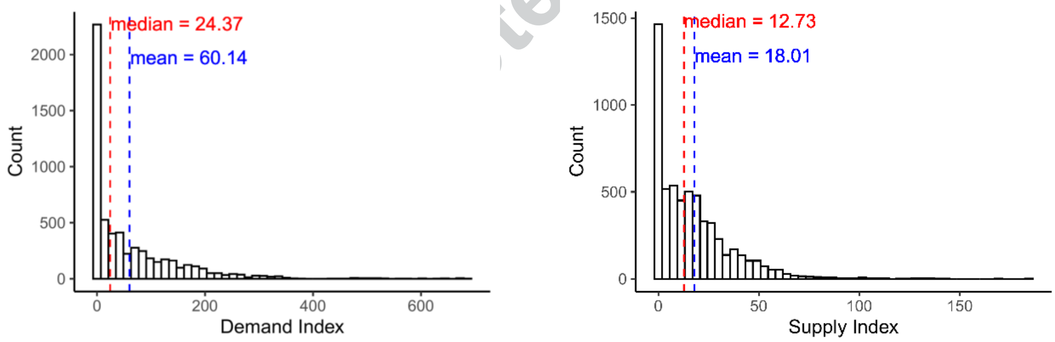


### 2.3.3 M-Score Normalization

The calculation for gap type 2 must be normalized since the supply and demand values are not measured in the same unit. Previous research has explored a few normalization methods, including min-max (Currie, 2010); percentage rescaling, which is similar to min-max except that it rescales from 0 to 100 instead of from 0 to 1 (Toms & Song, 2016); and Z-score (Jiao, 2017). The drawback to min-max is that the values are normalized such that the maximum and minimum value are 1 and 0, respectively, and the values in between are normalized accordingly. In our case, an area with maximum supply and maximum demand will have a normalized value of 0, which is not intuitive. The Z-score approach addresses this shortcoming, but Z-score normalization relies on the assumption that the values are normalized, which is not the case for the public transit demand and supply values, as shown in Figures 29 and 30 in the Appendix.

The statistical mean or average used by the Z-score normalization may not be the best indicator of the central tendency of the demand and supply values. Several studies have suggested that, compared to the mean, the median is a better measure of the center of skewed distributed values (Smith, 2016)(Moore & McCabe, 2003, p. 43) (Agresti & Finlay, 1997, p. 50)(Thorne & Giessen, 2000, pp. 81-82). Based on our data exploration for the transit demand and supply values, the median is indeed a better measure of the center than the mean.

The overall distributions of the type 2 transit demand and supply indices over the five years are shown in Figure 9. The mean demand is 60.14 and the median demand is 24.37. Of the demand values, 65.22% (3,707) fall below the mean and 34.77% (2,009) are above. The demand values are more evenly distributed around the median, with 50% (2,858) falling on either side. The supply's mean is 18.01 and its median is 12.73. Of the supply values, 62.50% (3,630) fall below the mean and 37.50% (2,178) are above. Like the demand values, the supply values are more evenly distributed around the median, where 50% (2,904) of the values fall on either side. These observations confirm that, as suggested by the literature, the median is a more suitable measure of center than the mean for the skewed distributions of the transit demand and supply indices.




(a) Demand value distribution (b) Supply value distribution

**Figure 9. Overall distributions of public transit demand and supply indices with their mean and median values**

For this research, we take a similar approach to the Z-score while addressing its shortcomings. The result is the *M-score* method, a normalization based on the statistical *median*. Eq. (9) defines the method used to standardize the value around the median.

$$\widehat{X_c} = (X_c - \tilde{X})/CD \tag{9}$$

where $\widehat{X_c}$ is the normalized value of data $c$ (or community $c$ in our case), $X_c$ is the original (non-normalized) value of data $c$, $\tilde{X}$ is the median value, and $CD$ is the coefficient of dispersion. $CD$ can be calculated as

$$CD = \frac{1}{n}\left(\frac{\sum_{c=1}^{n}|\tilde{X} - X_c|}{\tilde{X}}\right) \qquad (10)$$

where $n$ is the total number of data points, which in our case is the number of community areas. It is important to note that the stability of $CD$ can be affected by extreme values and an open-ended distribution because $CD$ measures the data's variability around the median. A single extreme data point can potentially have a disproportionately large effect since $CD$ is based on the difference between each data value and the median. An open-ended distribution occurs when the distribution does not have a specific boundary for the highest or lower data values.

## 3 Results

This section presents the results of our gap analysis based on the approaches presented in the previous section. The results discuss gap types 1 and 2, and analyze both gap types based on Calgary's geographical units.

### 3.1 Gap Type 1

Gap type 1 is calculated using data from 2016. It is measured for each community area based on the calculated demand and supply values. The public transit demand and supply values are ranked and grouped into quartiles: Very Low, Low, High, Very High. The values showing when public transit users are leaving for work in the morning are considered hourly between 5 a.m. and 8 a.m. Hourly demand and supply values for Calgary's communities are shown in Figures 10 and 11, respectively.

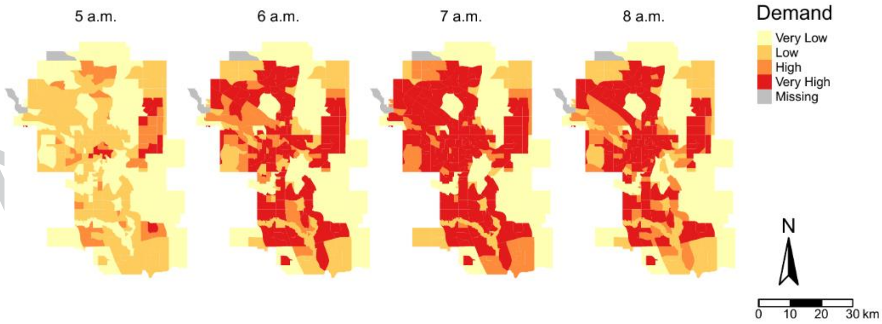



Figure 10. Public transit demand levels (type 1) per community in hourly intervals from 5 a.m. to 8 a.m.

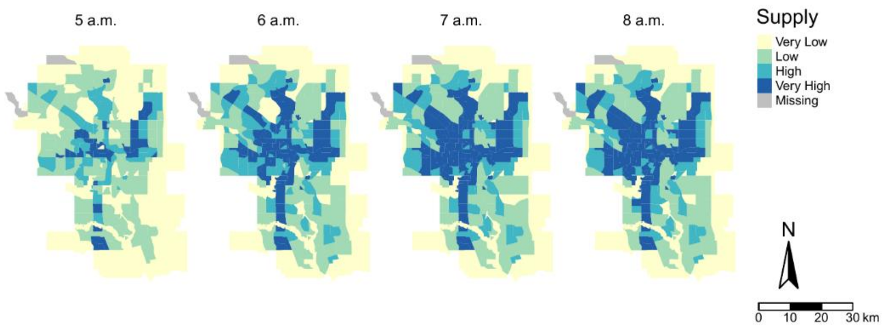



Figure 11. Public transit supply levels (type 1) per community in hourly intervals from 5 a.m. to 8 a.m.

The overall demand begins to rise at 5 a.m. and peaks during the 7 a.m. hour before dropping slightly during the 8 a.m. hour. Demand is consistently High along the LRT lines. The supply level reflects the amount of public transit service provided in different communities during different morning hours.

The values of gap type 1 for different communities during the morning hours are shown in Figure 12. A community where demand is greater than supply is considered undersupplied (US), while a community where supply is greater than demand is considered oversupplied (OS). For display purposes, gap values are grouped into six levels based on standard deviation: High US, Medium US, Low US, Low OS, Medium OS, and High OS. Communities with no data are labelled as 'missing'.

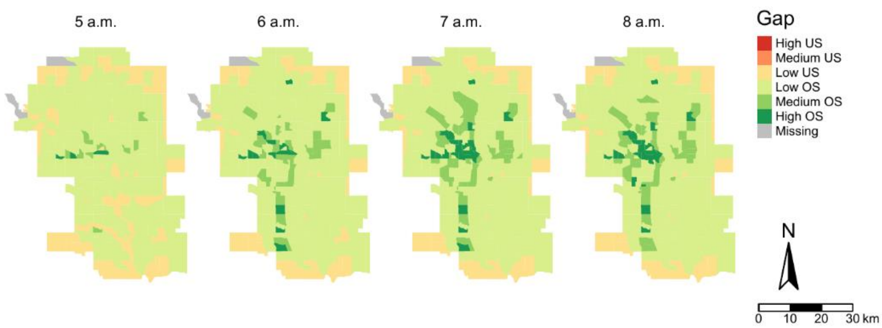



Figure 12. Gap type 1 per community in hourly intervals from 5 a.m. to 8 a.m.

Figures 13, 14, and 15 show demand type 1, supply type 1, and gap type 1, respectively, per community sector. Each sector's value is an average of all its community-based values. Demand is Low in the early morning (5 a.m.) and grows on the western side of the city by the 6 a.m. hour. By the 7 a.m. hour, demand rises to Very High in the Centre sector. Demand drops slightly during the 8 a.m. hour. During this time, the Western sectors maintain a High level demand, while the sectors on the Eastern side have a Low level of demand.

At 5 a.m., transit supply is High in the Centre sector and Low for most other sectors, with the exception of the Southeast, where it Very Low. By the 6 a.m. hour, the supply level has risen to Very High for the Centre sector. Supply is High for the Northwest, West, and Northeast sectors, and Low for the rest of the city. These levels remain steady through the 8 a.m. hour. Gap type 1 indicates that most parts of the city are at the Low OS level with the exception of the Centre sector, which is considered High OS at the 7 and 8 a.m. hours.

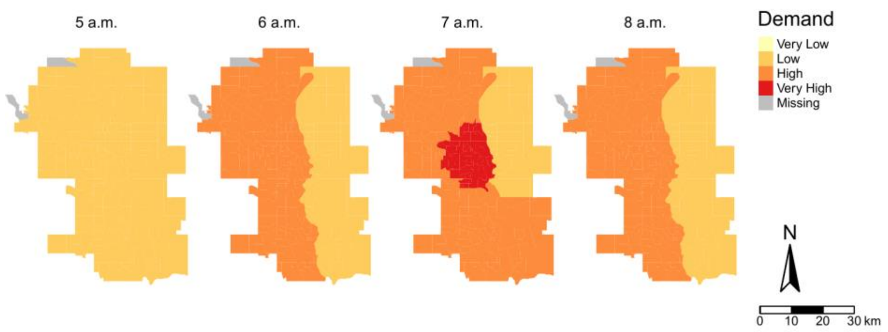



Figure 13. Public transit demand type 1 levels per community sector in hourly intervals from 5 a.m. to 8 a.m.

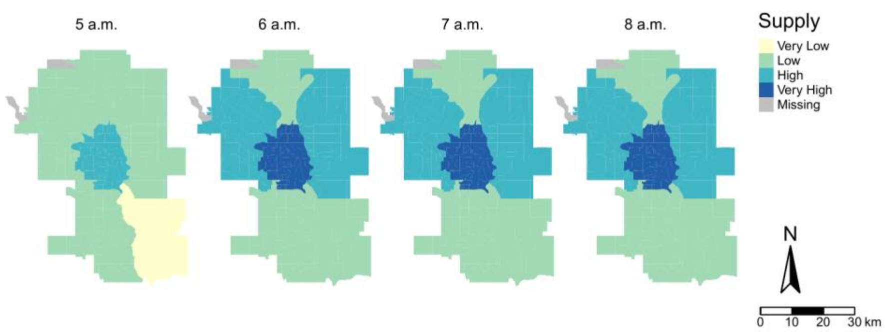



Figure 14. Public transit supply type 1 levels per community sector in hourly intervals from 5 a.m. to 8 a.m.

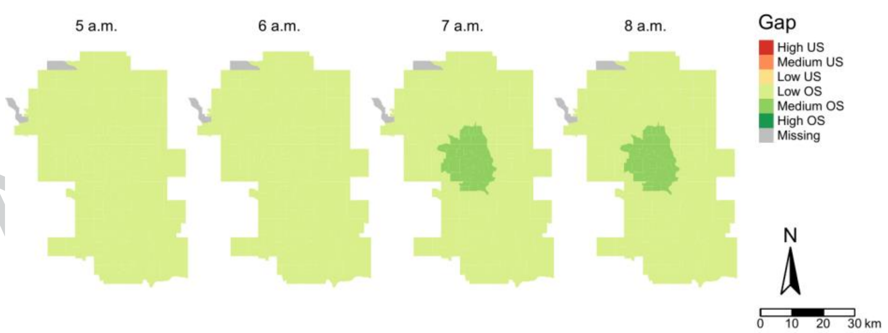



Figure 15. Gap type 1 per results community sector in hourly intervals from 5 a.m. to 8 a.m.

Figures 16, 17, and 18 show the demand type 1, supply type 1, and gap type 1, respectively, separated into four community classes. Throughout the morning hours, demand from the Residential areas is the highest compared to the other classes. This demand rises to Very High during the 7 a.m. hour. The Major Park and Residual Sub Areas have the lowest demand throughout the morning, starting at Very Low demand

and increasing to Low demand from 6 a.m. onwards. Industrial areas have a steady Very Low demand level throughout the study's time period.

Residential areas have the highest transit supply of all classes, starting at the low level and moving to the high level from 6 a.m. onwards. Major Park and Industrial areas have a low supply level from 6 a.m. onwards. Residual Sub Areas have very low supply throughout the study's time frame. Interestingly, gap type 1 remains unchanged, with the Residual Sub Area considered low US and the rest of the areas considered low OS.

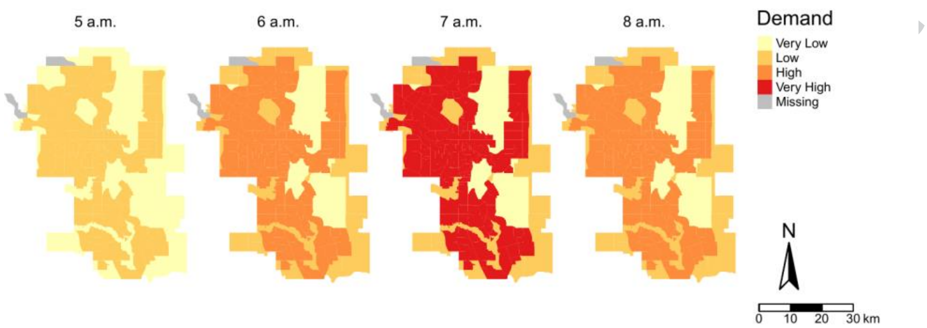



Figure 16. Public transit demand type 1 levels per community class in hourly intervals from 5 a.m. to 8 a.m.

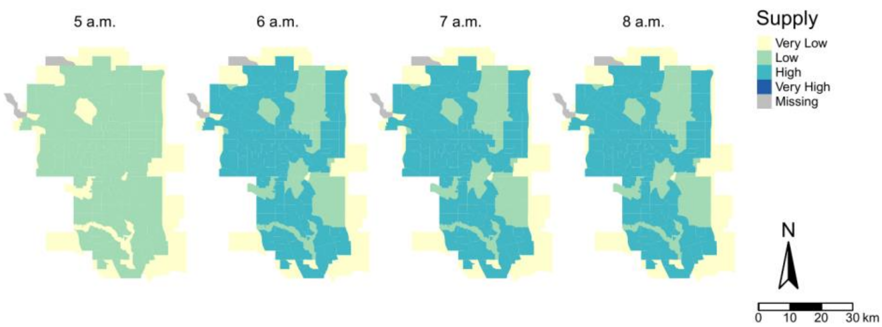



Figure 17. Public transit supply type 1 levels per community class in hourly intervals from 5 a.m. to 8 a.m.

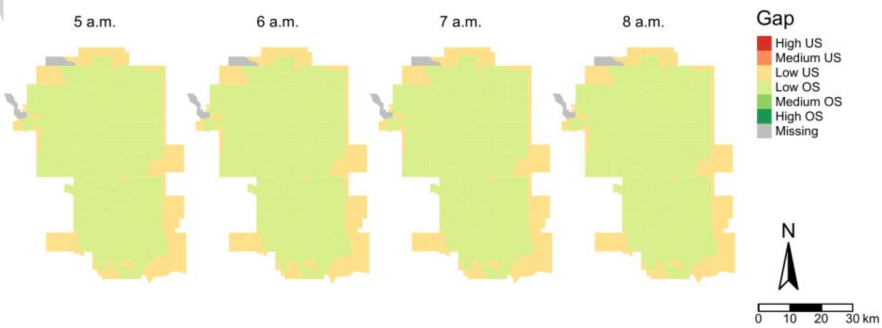



Figure 18. Gap type 1 per community class in hourly intervals from 5 a.m. to 8 a.m.

### 3.2 Gap Type 2

Gap type 2 is analyzed for each of the five consecutive years from 2013 to 2017. Like gap type 1, gap type 2 is measured for each community area based on calculated demand type 2 and supply type 2 indices. The demand and supply indices for public transit are normalized using the M-score method and then ranked and grouped into quartiles: Very Low, Low, High, Very High. The normalized values are evaluated hourly from 5 a.m. to 8 a.m. to give us snapshots of the times people are leaving for work for each of the five years. The resulting gap measurements can then be compared across the five-year span.

Figures 19 and 20 show the normalized hourly demand type 2 and supply type 2 values across different communities, respectively. Similar demand patterns can be seen across the five-year study period as time progresses from 5 a.m. to 6 a.m., with demand peaking during the hour of 7 a.m. and dropping slightly during the hour of 8 a.m. The supply level shows a similar hourly trend for each year, but its overall level increases sharply between 2013 and 2014, drops slightly in 2015, and then drops sharply in 2016. The supply level remains steady between 2016 and 2017.

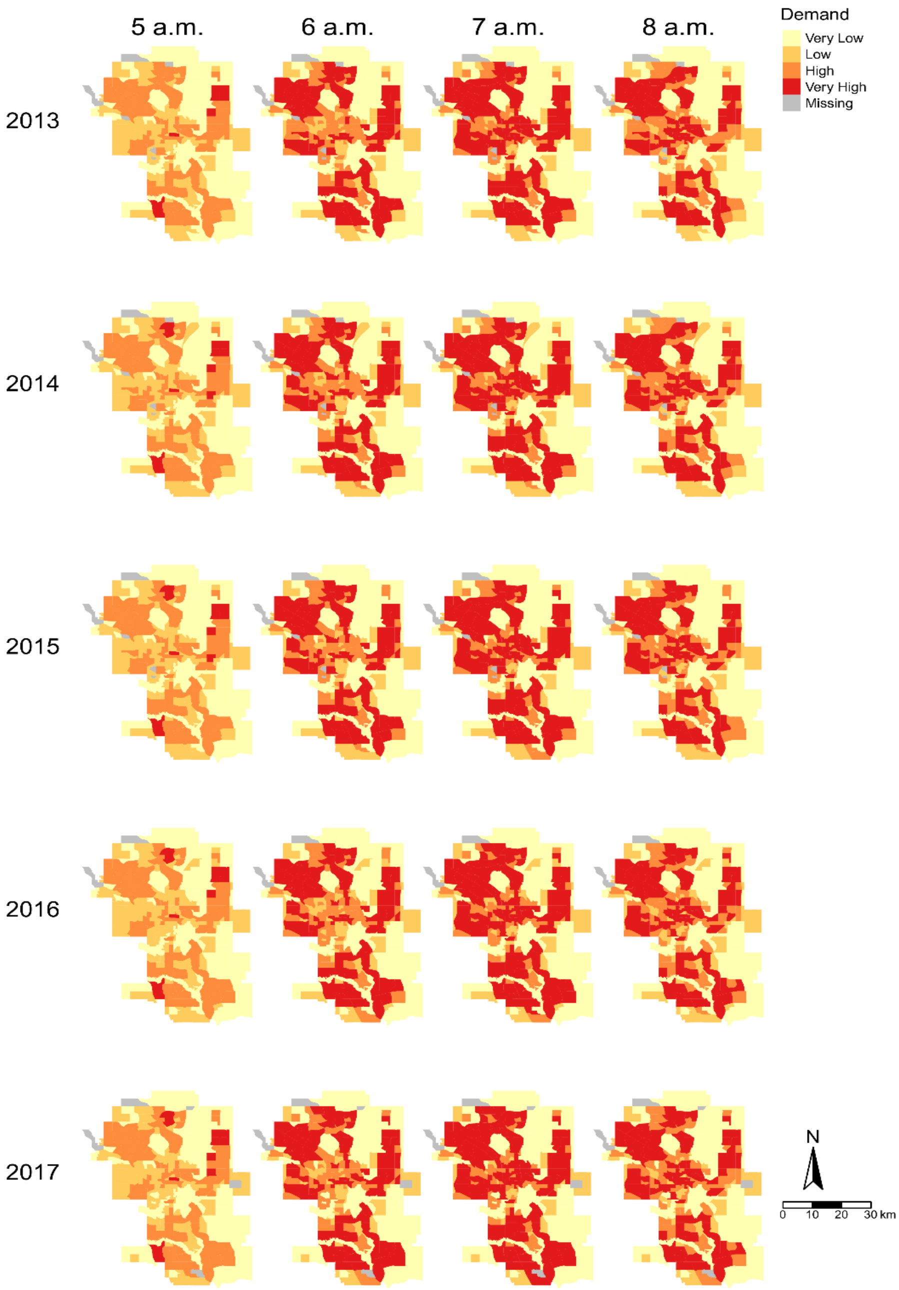



**Figure 19. Hourly public transit demand type 2 levels per community in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017**

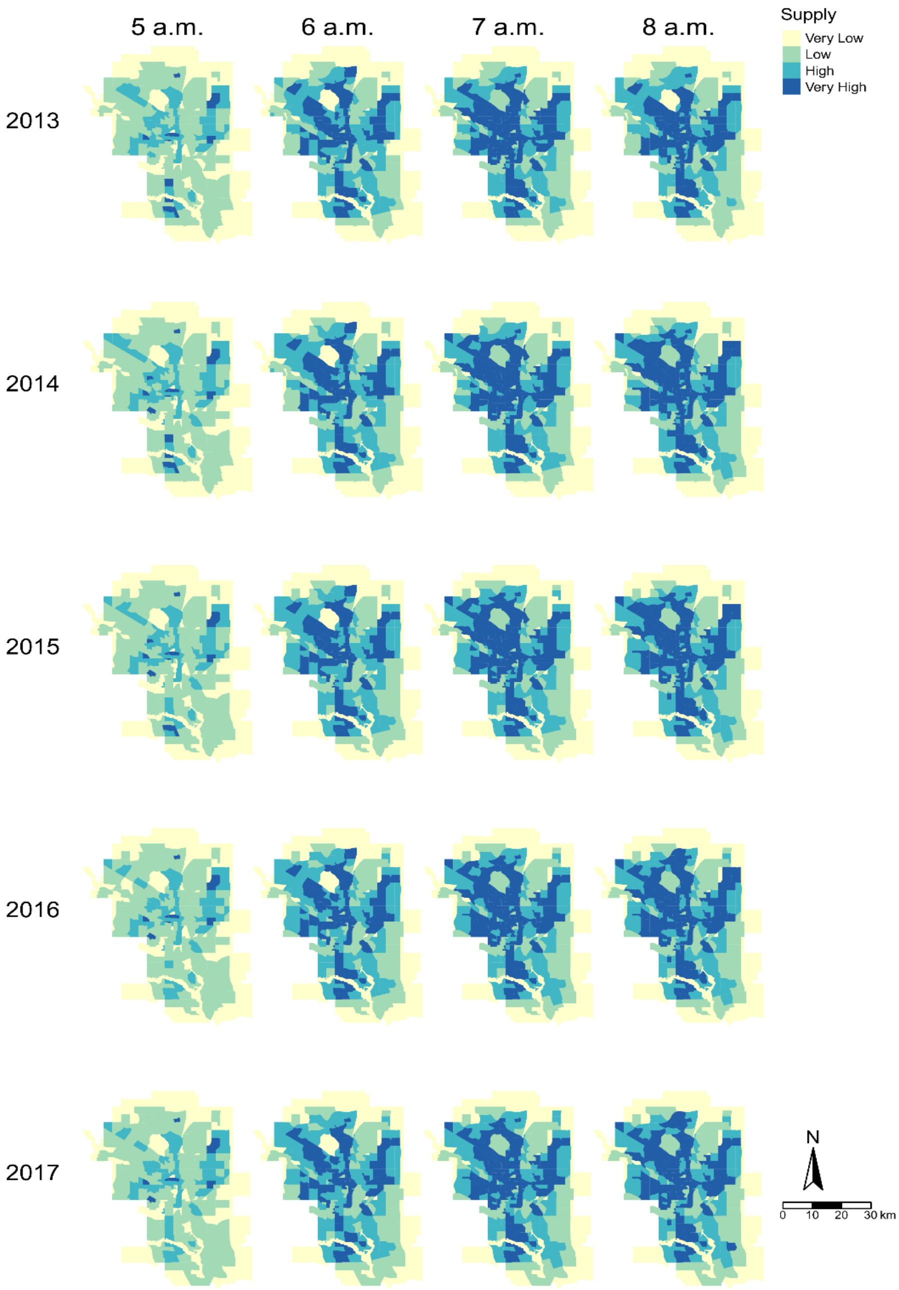



**Figure 20. Hourly public transit supply type 2 levels per community in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017**

Gap type 2 is calculated based on the normalized demand and supply values. The resulting values are shown in Figure 21, where they are grouped into six categories based on their relative deviations from their respective mean values for each community (Toms & Song, 2016). A trend emerges between 2013 and 2015, where the overall gap gradually increases from 5 a.m. to 7 a.m. and then drops slightly during the 8 a.m. period. Undersupplied gaps are seen mostly in the central part of the city, while oversupplied gaps are mostly scattered around the outer part of the city.

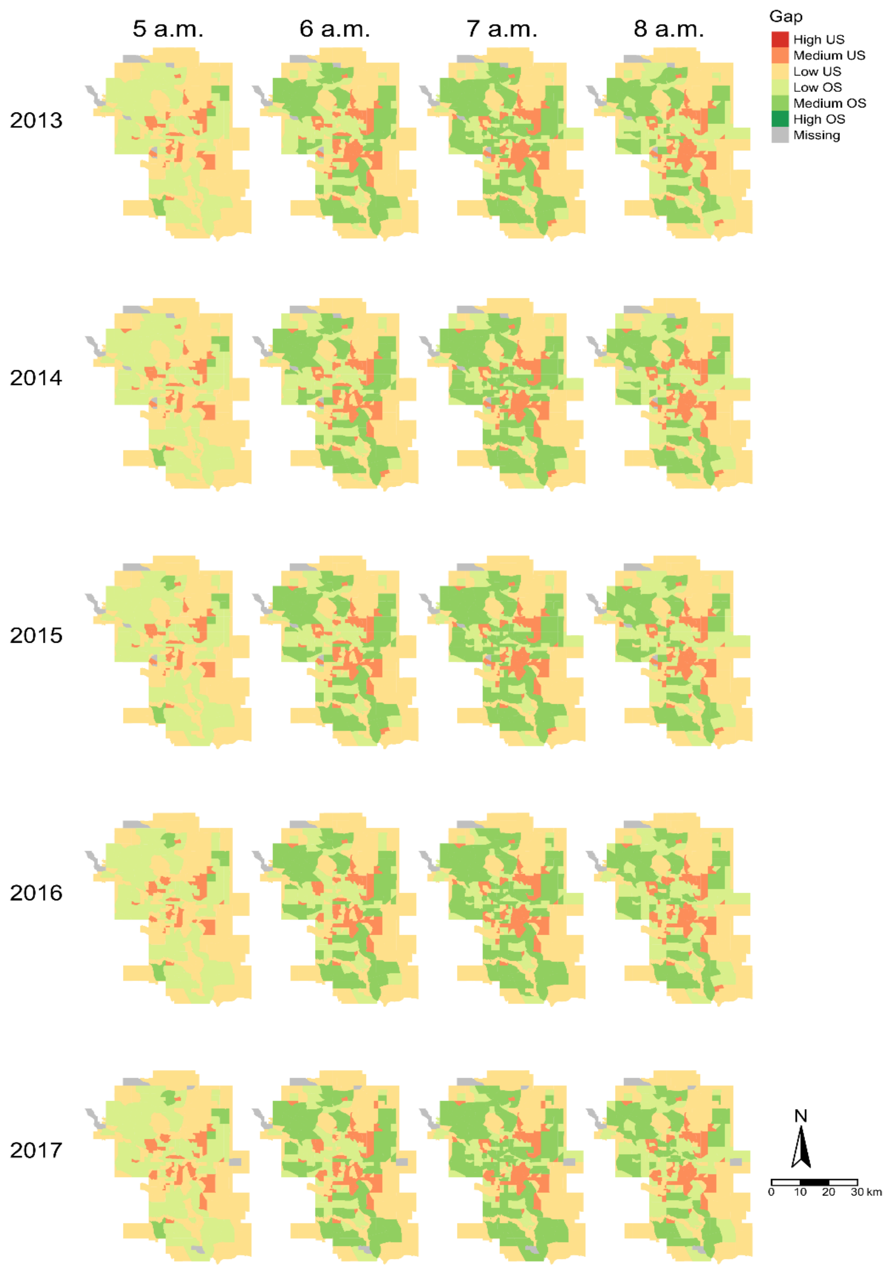



**Figure 21. Hourly gap type 2 per community in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017**

Figures 22, 23, and 24 show the demand, supply, and gap type 2 per community sector, respectively. A similar demand pattern can be seen each year: demand is Low in all sectors during the 5 a.m. hour before rising to High in all but the Eastern sectors during 6 a.m. and 7 a.m. hours. The North sector's demand level drops back to Low during the 8 a.m. hour, but the other sectors remain at the same level.

When the supply level is grouped by sector (i.e., taking an average across enclosed communities), it trends differently across the years of the case study. The supply level during the 5 a.m. hour is Low for all sectors for the duration of the study period. In 2013, the supply levels for the Central, Northwest, West, and Northeast sectors rose to High during the 6 a.m. period. In the later four years, the East sector's supply level also rose to High during the 6 a.m. period. The Central sector's supply level is Very High for the 7 a.m. time period between 2013 and 2016. From 2013 to 2015, the East sector's supply level remains High from 6 a.m. onwards. In 2017, no sector ever reaches the Very High level.

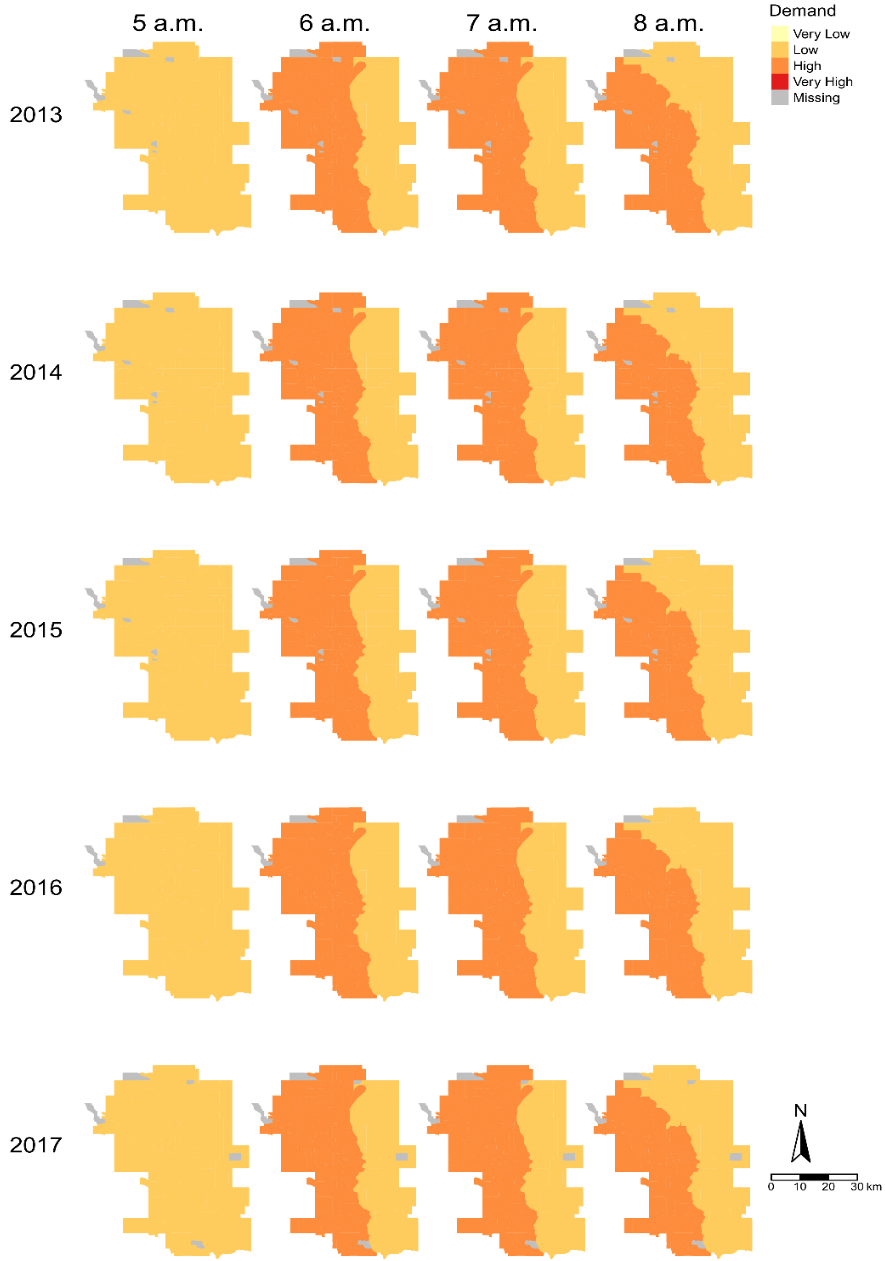



**Figure 22. Hourly public transit demand type 2 levels per community sector in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017**

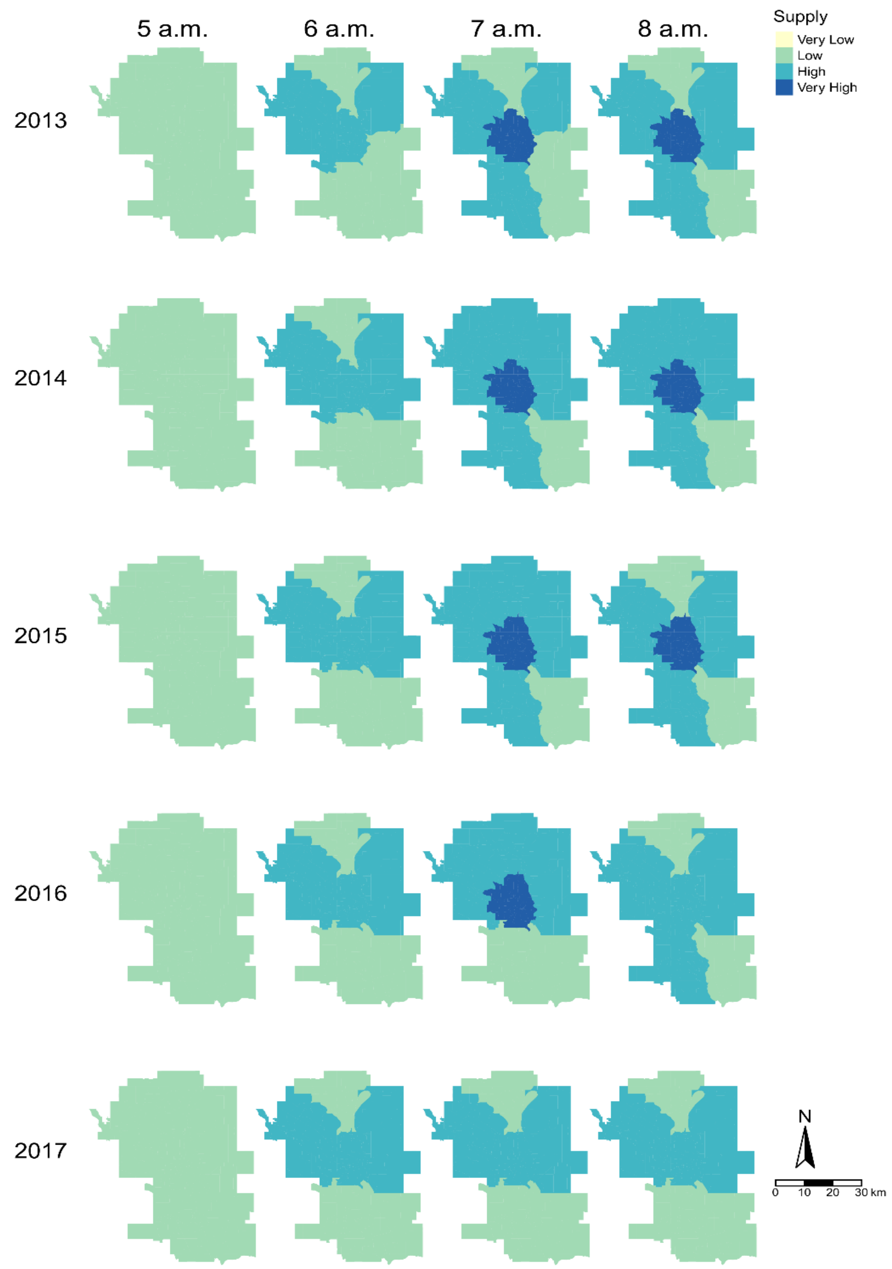



**Figure 23. Hourly public transit supply type 2 levels per community sector in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017**

When we look at gap type 2 per community sector, most sectors throughout the study period are at the Low OS level. The Central sector consistently starts the day at the Low US level. In 2014 and 2017, the East sector also begins at the Low US level. All other sectors begin the day at the Low OS level. This suggests a small undersupplied gap in the central part of the city during the earliest morning hour. During the 6 a.m. hour in 2013, the Central sector is still at the Low US level. All other sectors are at the Low OS level during this time. In 2013 and 2014, only the East sector is at the Low US level, while other sectors remain at the Low OS level during the 8 a.m. hour. The gap level in the other study years remains at the Low OS across all sectors for the 6 a.m., 7 a.m., and 8 a.m. hours. This may suggest that there is a small oversupply of service in general for the morning service operation hours.

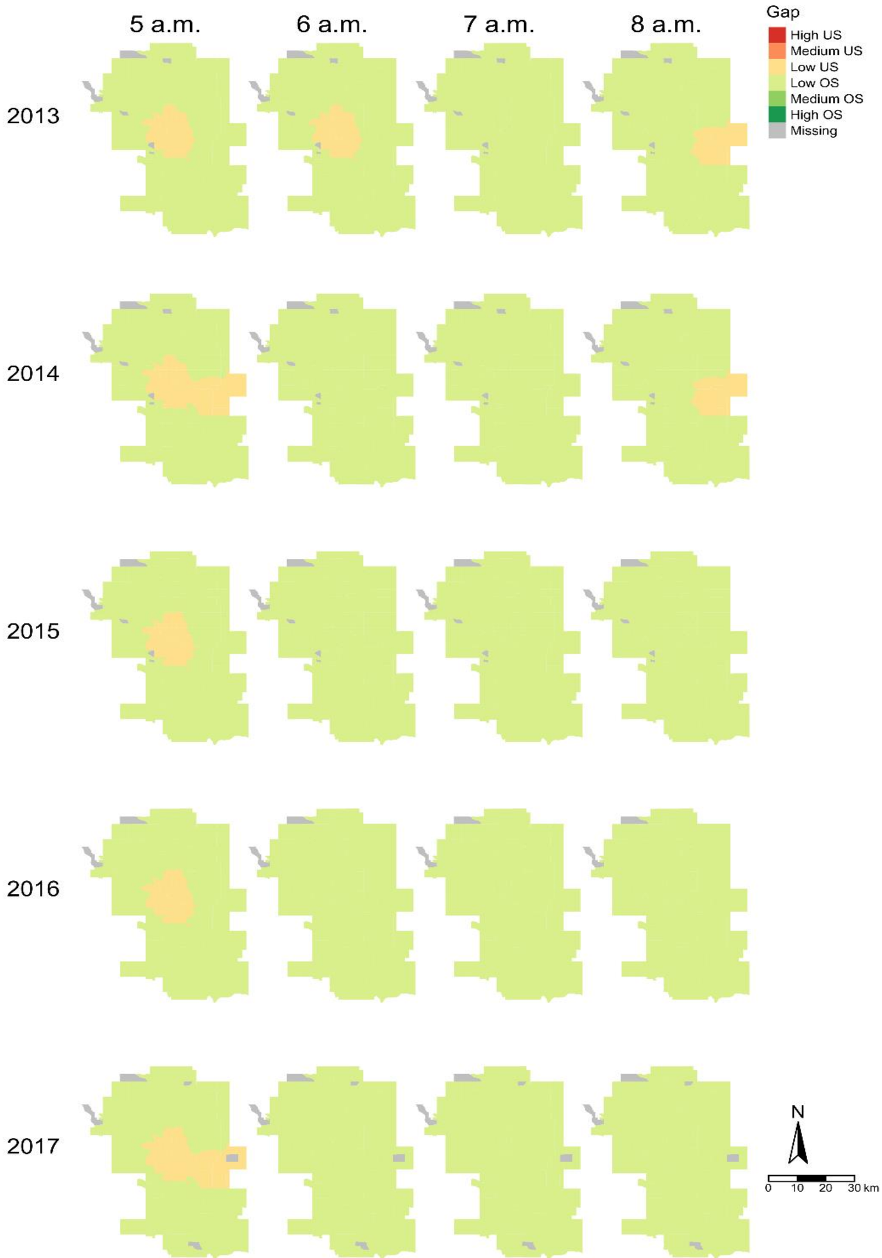



Figure 24. Gap type 2 per community sector in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017

Figures 25, 26, and 27 show the demand, supply, and gap type 2, respectively, by community class. Demand levels across different community classes for each of the observed years display a similar pattern, where the Residential areas have a higher level of demand than other classes. Residential areas begin with Low demand during the 5 a.m. hour, with that demand rising to High at 6 a.m. and remaining at that level for the duration of the morning. This observation is the same across the five-year study period.

The supply level exhibits a similar pattern to the demand level across the study period. At 5 a.m., the Residential and Industrial areas begin the morning with a Low supply level, while the Major Park and Residual Sub Area begin the morning with a Very Low supply level. The Residual Sub Area remains at the Very Low level and the Industrial area stays at the Low level throughout the morning hours. The Residential area rises to a High supply level at 6 a.m. and remains at that level for the duration. The Major Park class increases to and remains at the Low supply level from 6 a.m. onwards.

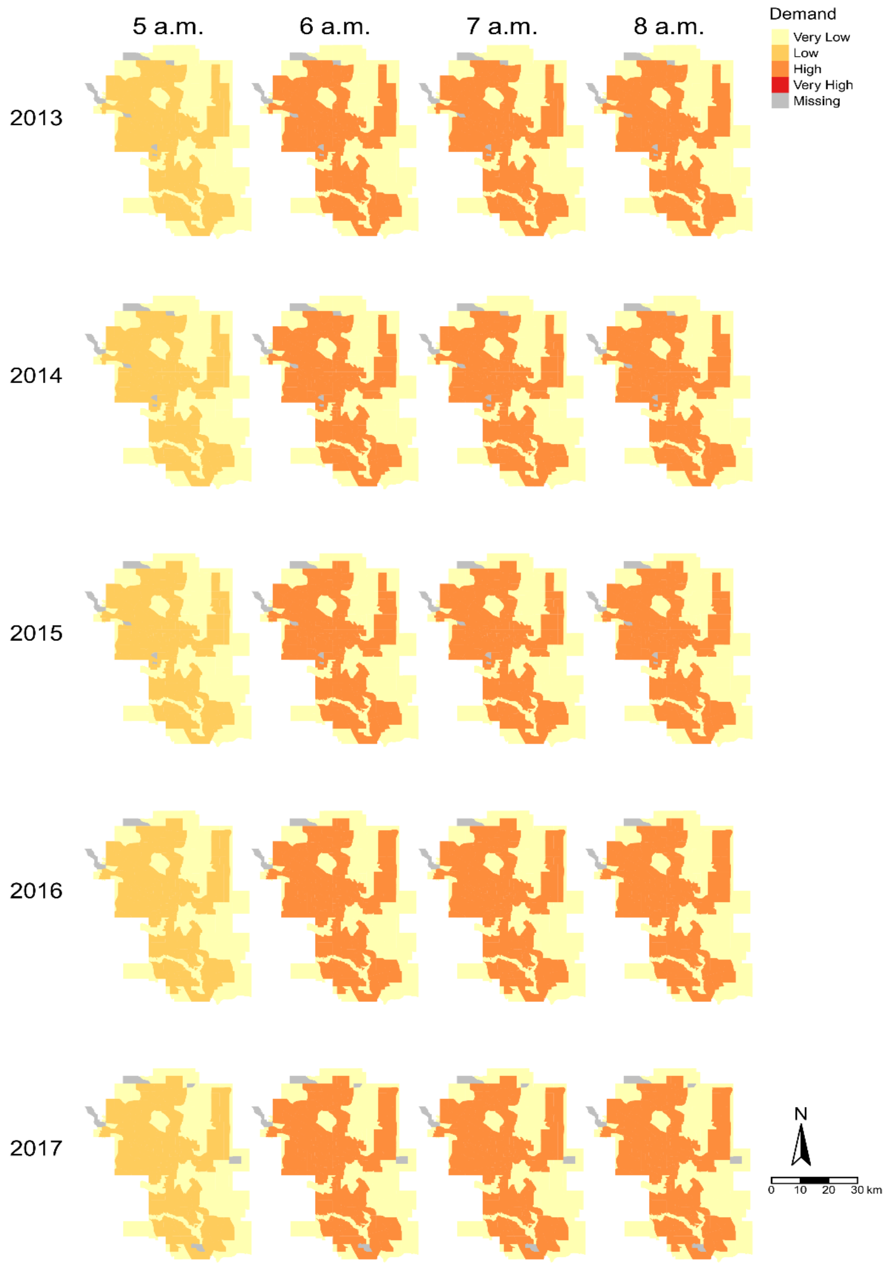



Figure 25. Hourly public transit demand type 2 levels per community class in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017

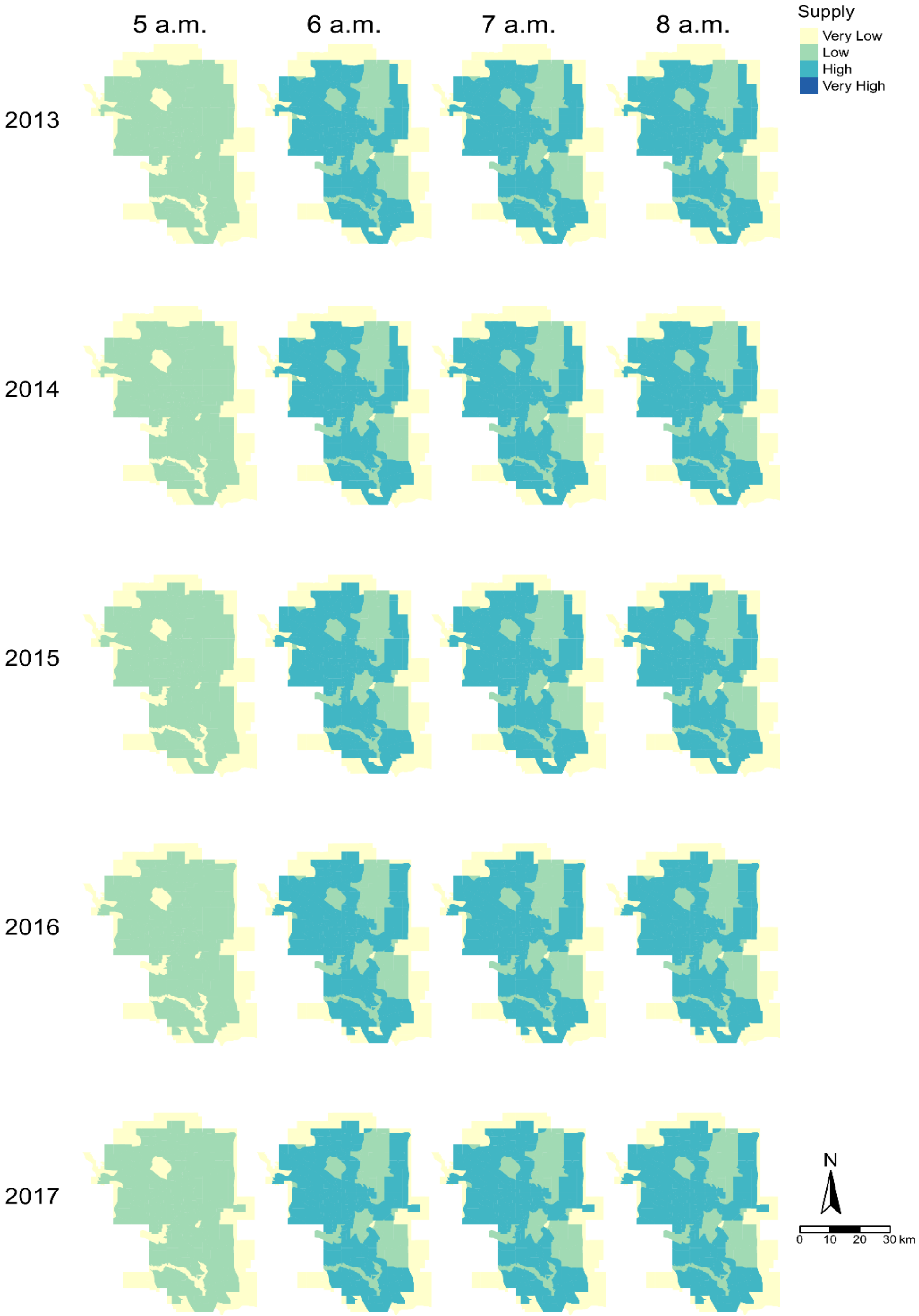



Figure 26. Hourly public transit supply type 2 levels per community class in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017

When gap type 2 is grouped by the four community classes, a similar pattern to the patterns exhibited by the supply and demand can be seen during the morning service operation hours for each of the five years. At 5 a.m., a Low OS gap is observed for the Residential area. During this time, all other classes have a Low US gap. Only the Industrial class's gap level changes after the 5 a.m. hour, rising to Medium US at 6 a.m. and remaining there for the duration of the morning. This may suggest that there is a need for additional transit service in Industrial areas during 6 a.m. to 9 a.m. timespan.

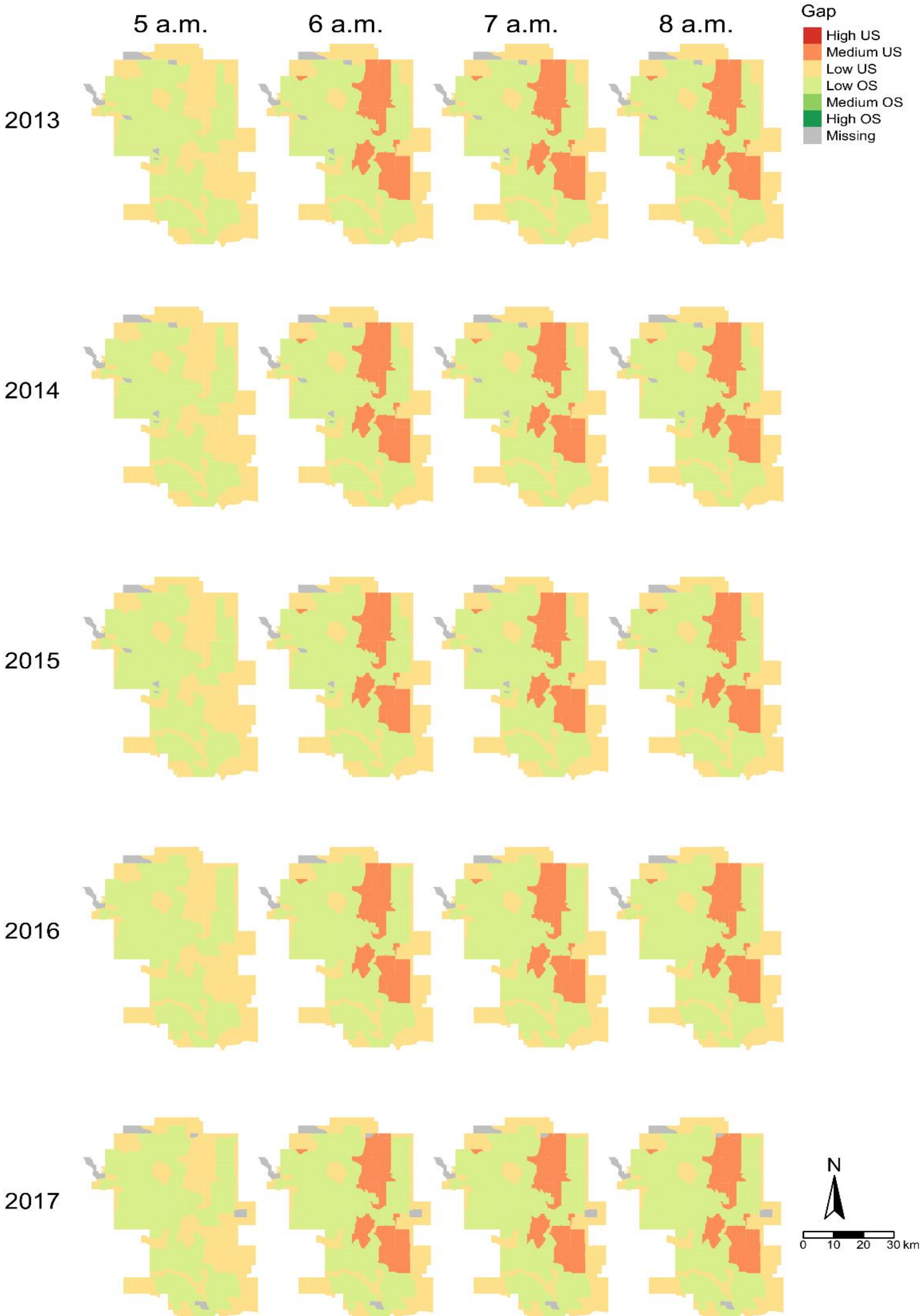


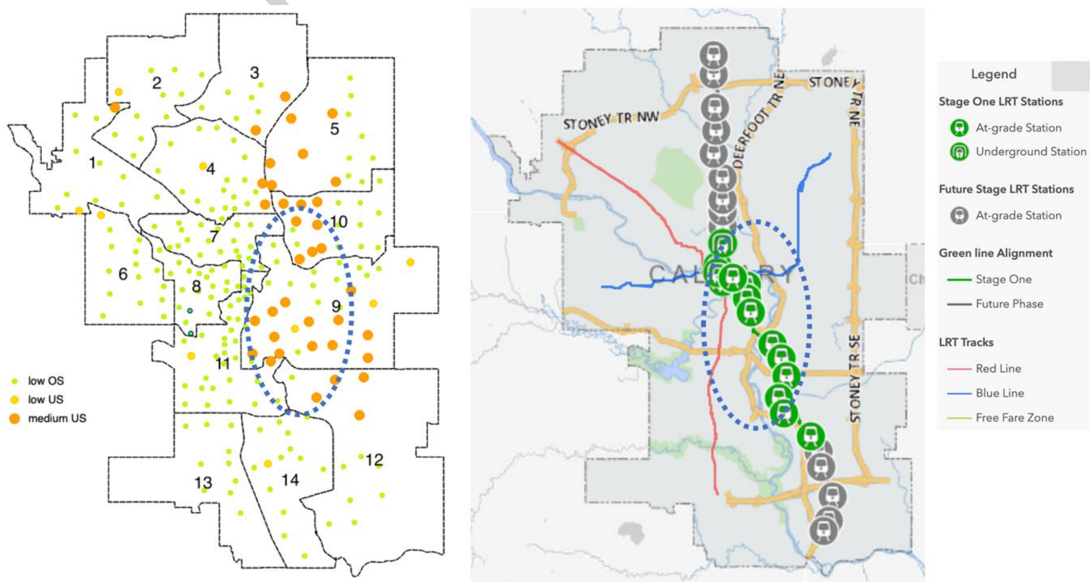

**Figure 27. Gap type 2 per community class in hourly intervals from 5 a.m. to 8 a.m. between 2013 and 2017**

The values of gap types 1 and 2 that were observed per sector and community class seem to be lower than the values that were calculated per community. This is because the sector and class values are calculated as an average of enclosed communities. A larger number of low-level communities may have dominated the average, which could explain these discrepancies. Nonetheless, the results are still valuable for trend observation and analysis.

Some previous research has addressed the gaps between transit supply and demand across different parts of a city, but there has been very little research to address gaps between transit supply and demand that vary over time (i.e., yearly and hourly). Characteristics of the built environment such as the transit trip generation and attraction roles of each dissemination area vary over time. Transit supply must be improved to adapt to the volume of transit demand moving in and out of each dissemination area based on the time of day. It is important to capture hourly gap variations for strategic planning of transit service.

Transit agencies can use the information gleaned from this research to identify locations and time periods with a high gap between transit demand and supply across the city. For example, Figure 28a shows Gap type 2 values per community class during the 8 a.m. hour in 2017. These values are shown using different colors and are associated with the centroids of each community class boundary for better presentation clarity. The Gap type 2 values highlight areas of the city that might need more attention in future transit planning efforts. This is especially interesting as it aligns with another city policy, the Green Line light rail transit infrastructure project. The Green Line is a planned LRT line that will run between north-central and southeastern Calgary (Figure 28b). Ward 9, which is shown to be underserviced, will receive eight of the 28 Green Line stations. Our analysis also shows a high concentration of medium US Gap type 2 values within Wards 5, 9, and 10 (Figure 28a). Transport planners can use this information to improve transit planning and operations in these specific areas.






(a) (b)

Figure 28. (a) Gap type 2 per community class centroids at 8 a.m. in 2017; (b) A planned Green Line LRT line in the city of Calgary (https://www.calgary.ca/Transportation/TI/Pages/Transit-projects/Green-line/home.aspx)

## 4 Conclusion

This paper presents two new approaches for measuring public transit demand-supply gaps, which are representations of the difference between the demand and supply values for transit. An area is considered undersupplied if the demand is greater than supply and oversupplied if the supply is greater than the demand. The first approach used in this paper, called gap type 1, measures the gap in the unit of the number of public transit users per area. The supply and demand values for this gap type are calculated based on detailed census passenger data (demand type 1) and GTFS feeds (supply type 1). Gap type 1 does not require value normalization because its demand and supply values are in the same unit, which makes this type of gap more applicable than a conventional gap index, which is normalized and therefore requires further interpretation by public transit system planners.

In a scenario where detailed passenger data is unavailable, gap type 2, which was introduced in this research, is a practical alternative to gap type 1. Gap type 2 uses a weighting scheme to estimate a portion of the transit user population from census data. It also uses a newly proposed normalization method called the M-score, which is based on the statistical median. Although gap type 2 yields a normalized index value, which is less desirable, there are advantages to this approach since gap type 2 enables a longitudinal gap analysis where yearly gap patterns and trends can be observed and compared.

The public transit system in Calgary, Canada was used as a case study for this research, which culminated in a comparative study of gaps over five years. Since gaps vary across not only the space in which transit service is provided but also over time, it is important to capture both types of gap variations to promote better strategic planning for transit service. Both proposed approaches therefore introduced the concept of hourly gaps to provide a new temporal dimension to the gap analysis. This has yet to be explored in the literature.

This research has explored gap measurements from three different perspectives: from a context of regional boundary (i.e., the community level), from the perspective of land use (i.e., the class type), and from a geographic perspective (i.e., the sector). Each perspective provides a different spatial context of the gap, which can be useful for transit service assessment and planning. We believe that our work contributes to the literature with its new perspectives on gap analysis, which is essential to improving the efficiency of public transit systems.

The results of this study can be used by policy makers and transit agencies to support data-driven recommendations for better planning of transit service, and to help them learn about emerging land use patterns. Transit agencies frequently make major service changes over the course of a year. For instance, Calgary Transit makes four major service changes annually. The changes are typically made based on feedback from passengers and drivers, political decisions, changing ridership levels, and new development areas. Our study creates a framework that can provide a transit demand-supply gap index over time. This index can identify locations and time periods with large transit demand-supply gaps to enable more efficient and effective transit planning. It can also highlight areas that are in need attention for future transit planning sessions.

There are several limitations to our study. We were only able to analyze weekdays, which were selected as a proof of concept demonstration for our approaches since they are periods of heavy use for public transit systems. Future studies should consider weekends, which would presumably exhibit different gap patterns and trends. We were limited in our analysis to only gaps for morning hours (5 a.m. to 9 a.m.) because the census data available on the Journey to Work's passenger information page only provides commuter data during the morning hours. The GTFS feeds that we used were carefully chosen as representative samples of each day of the week, but different choice may slightly alter the results. Our approach considers only unidirectional demand and supply, so the derived gap only reveals one aspect of the under- or over-served public transit service in the study area. The directional gap is thus worth future investigation. Our final notable limitation is that both demand types are based on reported public transit passenger data. These data cannot account for potential passengers like those who want to use public transit services but are unable to because of its inaccessibility or vulnerable groups who should be public transit-dependent. There may also be other hidden demands that are dependent on the local socio-demographics and land use. This look into potential transit demand paves the way for future studies that can better address these limitations. Future research that focuses on the development of graphics that synthesize the gap findings into a simpler visualization that allow the reader to understand the point at a glance is also essential, especially for a longitudinal gap analysis.



## 5 Acknowledgments

This research was supported by the Eyes High Postdoctoral Fellowship Program, Alberta Transportation, and the Alberta Motor Association (AMA). This research work was partially supported by Chiang Mai University.

## References

Advancing Public Transport. (2017). *Statistics Brief: Urban Public Transport in the 21st Century*. Retrieved from https://www.uitp.org/sites/default/files/cck-focus-papers-files/UITP_Statistic Brief_national PT stats.pdf

Agresti, A., & Finlay, B. (1997). *Statistical Methods for the Social Sciences* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Altman, S. (2011). Overview of GTFS-Realtime and Comparison to TSIP-RTDP. New York state department of transportation. Retrieved May 17, 2020, from http://ngtsip.pbworks.com/w/file/fetch/49201665/GTFS-RT_RTDP_Overview_v1.1.pdf

American Public Transportation Association (APTA). (2018). Public Transportation Benefits. Retrieved August 3, 2019, from https://www.apta.com/news-publications/public-transportation-benefits/

Barbeau, S. J. (2017). Quality control - lessons learned from the deployment and evaluation of gtfs-realtime feeds. Retrieved May 19, 2020, from https://ppms.trec.pdx.edu/media/project_files/GTFS-realtime_lessons_learned-v11.pdf

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2008.11.004

Calgary Open Data. (2016). Community Boundaries. Retrieved May 9, 2018, from https://data.calgary.ca/d/ab7m-fwn6?category=Base-Maps&view_name=Community-BoundariesOpens in new window

Calgary Transit. (2002). *A Review of Bus Rapid Transit*. *Transit Planning*.

Calgary Transit. (2014). West LRT One Year Review. Retrieved December 3, 2019, from https://www.calgarytransit.com/sites/default/files/reports/west_lrt_one_year_review.pdf

Calgary Transit. (2018). Statistics | Calgary Transit. Retrieved October 26, 2018, from http://www.calgarytransit.com/about-us/facts-and-figures/statistics

Canadian Geographic. (2012). Clean commute. Retrieved December 21, 2018, from https://www.canadiangeographic.ca/article/clean-commute

Currie, G. (2010). Quantifying spatial gaps in public transport supply based on social needs. *Journal of Transport Geography*, *18*(1), 31–41. https://doi.org/10.1016/j.jtrangeo.2008.12.002

DART First State. (2018). The Environmental Benefits of Riding Public Transit. Retrieved July 21, 2019, from http://www.dnrec.delaware.gov/dwhs/info/Pages/OzonePublicTrans.aspx

Dawoodjee, L. (2020). GTFS Validators. Center for Urban Transportation Research at the University of South Florida. Retrieved May 17, 2020, from https://github.com/luqmaan/awesome-transit#gtfs-validators

Demissie, M. G., Phithakkitnukoon, S., Sukhvibul, T., Antunes, F., Gomes, R., & Bento, C. (2016). Inferring Passenger Travel Demand to Improve Urban Mobility in Developing Countries Using Cell Phone Data: A Case Study of Senegal. *IEEE Transactions on Intelligent Transportation Systems*, *17*(9), 2466–2478. https://doi.org/10.1109/TITS.2016.2521830

Dickens, M., & Hughes-Cromwick, M. (2019). Leveraging Big Data in the Public Transportation Industry. American public transportation association. Retrieved May 21, 2020, from https://www.apta.com/wp-content/uploads/Big-Data-Policy-Brief.pdf

Google. (2012). Google Transit Feed Specification (GTFS) Overview.

Jiao, J. (2017). Identifying transit deserts in major Texas cities where the supplies missed the demands. *The Journal of Transport and Land Use*, *10*(1), 529–540. https://doi.org/10.5198/jtlu.2017.899

Jiao, J., & Dillivan, M. (2013). Transit Deserts: The Gap between Demand and Supply.

*Journal of Public Transportation*. https://doi.org/10.5038/2375-0901.16.3.2

Knoblauch, M. (2019). Americans spend 19 full work days a year stuck in traffic on their commute. Retrieved August 1, 2019, from https://nypost.com/2019/04/19/americans-spend-19-full-work-days-a-year-stuck-in-traffic-on-their-commute/

Moore, D. S. ., & McCabe, G. P. (2003). *Introduction to the Practice of Statistics* (4th ed.). New York: Freeman.

O'Toole, R. (2018a). Charting Public Transit's Decline. *Policy Analysis*, *853*, 1–18.

O'Toole, R. (2018b). Urban Transit Decline. Retrieved August 4, 2019, from https://www.downsizinggovernment.org/urban-transit-decline

OpenMobilityData. (2020). Retrieved May 18, 2020, from https://transitfeeds.com/p/calgary-transit/238?p=8

Parady, G. T., Takami, K., & Harata, N. (2017). Built environment and travel behavior: Validation and application of a continuous-treatment propensity score stratification method. *Journal of Transport and Land Use*, *10*(1), 763–788. https://doi.org/http://dx.doi.org/10.5198/jtlu.2017.890

Prommaharaj, P., Phithakkitnukoon, S., Demissie, M. G., Kattan, L., & Ratti, C. (2020). Visualizing public transit system operation with GTFS data: A case study of Calgary, Canada. *Heliyon*, *6*(4), e03729. https://doi.org/10.1016/j.heliyon.2020.e03729

Smith, M. K. (2016). Summary Statistics for Skewed Distributions. Retrieved October 10, 2019, from https://web.ma.utexas.edu/users/mks/statmistakes/skeweddistributions.html

Statistics Canada. (2001). Calgary-Edmonton Corridor. Retrieved October 26, 2018, from http://geodepot.statcan.ca/Diss/Highlights/Page9/Page9d_e.cfm

Statistics Canada. (2012). *Overview of the Census: Census year 2011*.

Statistics Canada. (2016a). Dissemination area (DA). Retrieved August 1, 2018, from https://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo021-eng.cfm

Statistics Canada. (2016b). Population and Dwelling Count Highlight Tables, 2016 Census. Retrieved October 26, 2018, from https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=302&SR=1&S=86&O=A&RPP=9999&PR=48

Thorne, B. M., & Giessen, J. M. (2000). *Statistics for the Behavioral Sciences* (3rd ed.). Mountain View, CA: Mayfield.

Toms, K., & Song, W. (2016). Spatial Analysis of the Relationship Between Levels of Service Provided by Public Transit and Areas of High Demand in Jefferson County, Kentucky. *Papers in Applied Geography*, *2*(2), 147–159. https://doi.org/10.1080/23754931.2015.1115365

Transit app. (2020). Retrieved May 21, 2020, from https://transitapp.com

Zannat, K. E., & Choudhury, C. F. (2019). Emerging Big Data Sources for Public Transport Planning: A Systematic Review on Current State of Art and Future Research Directions. *Journal of the Indian Institute of Science*, *9*, 601–619. https://doi.org/10.1007/s41745-019-00125-9