JTLU

Development of a temporal and spatial linkage between transit demand and land-use patterns

Sang Gu Lee [a]
East-West Gateway Council of Governments

Mark Hickman [b]
University of Queensland

Daoqin Tong [c]
University of Arizona

Abstract: We are exploring ways to capture the temporal and spatial dimensions of the use of public transit. Specifically, we are investigating how different land uses affect the spatial and temporal demand for public transit services. Spatially, the availability of new data collection technology in public transit allows us to examine transit demand at the individual stop level. Our hypothesis, however, is that transit users' activity may not be originated from or destined to an individual stop per se; rather, the activity is associated with a specific location in the vicinity of the stop, and this location may be "covered" by several adjacent transit stops. More importantly, understanding the transit demand at this aggregate level (an aggregate "catchment" area) can enhance the ability to define a specific land-use type and the temporal characteristics related to passengers' activities. Temporally, we seek to understand the relationship between the demand for public transit service at specific times of the day and the associated land uses that may strongly influence the timing of that demand. To explore these dimensions, this study: 1) proposes a method of stop aggregation; 2) generates transit service areas based on these aggregated stops; 3) develops a set of metrics to better represent land-use types within these service areas; and 4) examines the spatial and temporal characteristics of transit demand for these service areas. These methods are applied to a case study using land-use and transit demand data from the Minneapolis–St. Paul metropolitan area.

Keywords: Transit stop aggregation, smart card data, transit service areas, transit and land use

1   Introduction

We are exploring ways to capture the temporal and spatial dimensions of the use of public transit. Specifically, we are investigating how different land uses affect the spatial and temporal demand for public transit services. Although data collection technology in public transit has been developed at the individual stop level, there may be limitations to this approach. A single stop is often associated with a single direction and perhaps a single route. From the point of view of access to land and activities, other stops in the immediate vicinity, serving other directions and other routes, may also provide access to the same land uses and activities. Transit users' activity itself is not originated from or destined to an individual stop (except transfers); the actual location of the activity is assumed to be near the stop. For this reason, adjacent stops (in either the same or the opposite direction) within a small area are more likely to cover the same or similar land uses with respect to transit users' activity. Aggregating stops within a small area may better capture the land uses served in that vicinity. In addition, stop aggregation is able to reduce the complexity of the transit network. The advantage of stop aggregation is that stop groups that serve common or similar land uses and activities can be represented by a single node.

Based on the research conducted previously (Lee and Hickman 2011), there is considerable potential for intersection-level identification of boarding stops through the analysis of smart card transaction data using GIS and database management tools. Because of the nature of GPS accuracy, as well as the fact that transfers usually occur within the same intersection or transit center, the approach at an aggregate level can provide a stronger foundation to investigate the interaction between transit demand and land-use types. We may also argue from the realm of transit that the use of aggregation better represents the actual spatial reference that transit passengers may use in describing their transit trips.

To elaborate, the need for a stop aggregation model is as follows.

- One critical reason for stop aggregation is because of the inherent problem of the geographic specificity of the given fare collection system. The smart card transaction data provides the current location of the transaction (in the presumed vicinity of a boarding stop), instead of the actual boarding stop itself. Even if the route information (without directional information) is available, the actual stops used by a passenger are not designated. As shown in Figure 1, in order to correctly

[a] sanglee@gmail.com
[b] m.hickman1@uq.edu.au
[c] daoqin@email.arizona.edu

---

identify or infer the bus stop location, it is necessary to search in the area around the location of the transaction.

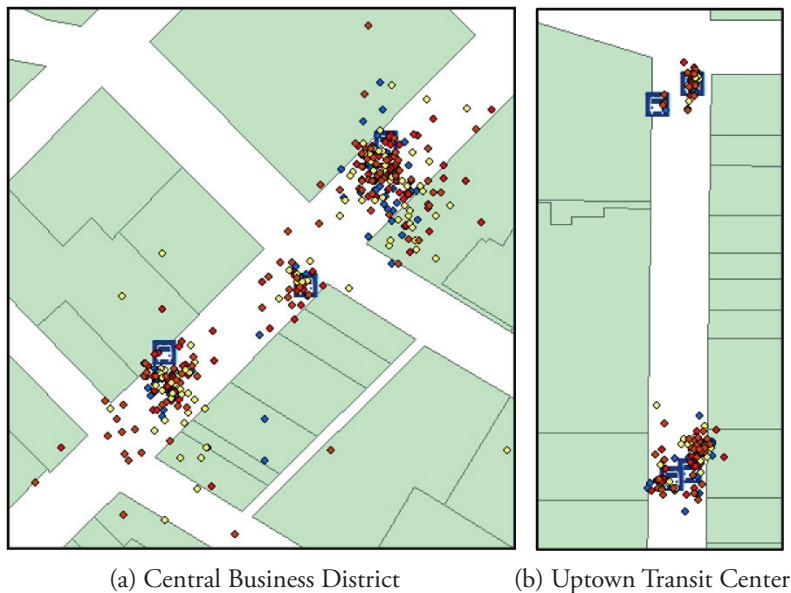(a) Central Business District (b) Uptown Transit Center

Figure 1: Current location of smart card transaction data. This figure shows the difficulty in identifying and inferring actual boarding and alighting stops from the transaction data (different colors indicate different days).

- Each bus stop itself serves only one direction. The choice of boarding and alighting locations, however, covers stops within the same general area. It is expected that all transit users within this same area should have an opportunity to take a bus in any desired direction from that area. For this reason, the stop aggregation model can provide specific nodes that connect to transit services in all directions, coming and going.
- Typically, on-board surveys ask each passenger's origin/destination and boarding/alighting stops in a more aggregate manner (e.g., a specific street intersection). This also suggests that an aggregate representation of stops can be a reasonable way to deal with major trip generators and attractors.

More importantly, understanding the transit demand at this aggregate level would greatly enhance the ability to define a specific land-use type and the temporal characteristics related to passengers' activities at that location. For example, commuters who live in a residential area may be more likely to take a bus between the hours of 6:00 a.m. and 8:00 a.m. during the week. Conversely, people who work in commercial or industrial areas are more likely to board a bus between the hours of 4:00 p.m. and 6:00 p.m. as they conclude their work activities. These aspects of travel, usually derived from survey information, give us strong motivation to integrate transit demand and land use, to explore the spatial and temporal distributions of transit demand.

For this study, transit demand is analyzed first using smart card data from Metro Transit in the Minneapolis–St. Paul metropolitan area. This case study illustrates one method for stop aggregation, using the Twin Cities case study, and then shows how to investigate transit service characteristics using data based on Google's General Transit Feed Specification (GTFS).

To investigate the interaction between transit demand and land-use types within an aggregate stop catchment area, this study uses a database of passenger fare card transactions recorded by an electronic fare collection system. Previous studies focus mainly on passenger origin-destination survey information to measure accessibility (e.g., distance to or from a transit stop) and to forecast stop-level transit ridership (e.g., based on demand from each parcel). In contrast to previous studies using the relatively small samples from passenger surveys, we use a much larger data set that has been collected from various electronic technologies: Automated Fare Collection (AFC) systems and Automated Passenger Counter (APC) systems.

This study contributes to our understanding of land use and public transit in several respects. First, this study provides a method of stop aggregation, which is employed to define transit service areas (i.e., catchment areas) based on transit users' activities. Second, this study develops a set of metrics to better represent land-use types in a specified transit service area, taking advantage of transit demand data from AFC and APC systems. Finally, utilization of various data sources (e.g., smart card data, GTFS data, and parcel-level land-use data) supports more quantitative and direct measures to estimate the temporal distribution of transit demand within a specific service area.

Following a review of the literature in the next section, we describe the methodology used to analyze the data. The fourth section provides a data description and preparation for analysis. An exploratory study on several cases using the Twin Cities is illustrated in the fifth section. The paper concludes with some final remarks and proposed future work.

2 Literature review

2.1 Stop aggregation

The aggregation of stops is proposed for generating a transit route origin-destination (O-D) matrix (Furth and Navick 1992). On-off count data have been aggregated to the segment level when there is no reliable small-sample survey or old O-D matrix at the stop level to serve as a seed. To determine the location of bus boardings, AFC and APC information, combined with the scheduled time of service at stops, allow transactions to be located by bus stop. However, it is useful to consider aggregating nearby stops into a stop group (several adjacent bus stops). The need for relevant stop aggregation to combine the scheduled time between bus stops from the AFC transaction

data is discussed in Barry et al. (2002). Based on their idea, one approach uses stop groups in the context of estimating aggregate-level origin-destination travel (Lee et al. 2011). That work emphasizes the usefulness of a more aggregate treatment to identify boarding and alighting stop locations from smart card data, as some transit demand is generated from other nearby stops on cross-streets or other locations. The methodology in this study expands this work to group stops that may have identical or similar transit catchment areas.

2.2 Service areas and walking

Service (catchment) areas and bus stop locations are typically planned based on some trade-off in supply (e.g., efficient operation, less travel time) and demand (e.g., more coverage). In the view of demand, a transit passenger's choice of bus stop is often associated with the nearest stop. Recent studies about this behavior have favored typical assumptions: Transit users generally tend to use the nearest stop/station, and they generally end their last trip of the day at the stop/station where they began their first trip of the day (Barry et al. 2002). However, what if the nearest stop does not serve the express or limited stop routes, or there are other stops within walking distance that are served by other routes? The transit service available at a stop is one important factor for defining the catchment area. One issue investigated here is the difficulty in determining mutually exclusive service areas for each stop. Generating mutually exclusive network-based service areas, especially for light-rail stations, is proposed by Upchurch et al. (2004). This is, however, complicated in the real world due to possible overlapping service areas and different service areas for each direction of travel. The effects of overlapping bus stop service areas were explored by Kimpel et al. (2007), who examined the decay of demand as a function of distance from the stop. Walking paths and the stop's catchment areas are commonly analyzed using shortest paths and Voronoi diagram methods applied to the street network (Furth et al. 2007a). Another study indicates that stop service areas are different for boarding and alighting stops and for different directions of travel (Furth et al. 2007b). Furthermore, compared to traditional neighborhoods with grid street patterns, suburban neighborhoods in which streets often are curvilinear with cul-de-sacs make it almost impossible to have separate (non-overlapping) service areas. For this reason, our approach to define service areas allows overlapping bus stop service areas.

Historically, the availability of transit service has been measured with regard to the number of persons that reside within some reasonable walking distance of transit routes or stops (Polzin et al. 2002). To determine a transit service area, the primary access mode to and from the transit stop (station) is generally assumed to be walking. The majority of previous studies (O'Sullivan et al. 1996; Hsiao et al. 1997; Loutzen-

heiser 1997; Beimborn et al. 2003; Zhao et al. 2003; Schlossberg et al. 2004; Kimpel et al. 2007; Weinstein et al. 2007; Crowley et al. 2009; Cheng et al. 2010; El-Geneidy et al. 2010) have focused heavily on walking as the primary access mode, including factors such as:

> 1) The walking distance to or from the transit stop: 400 meters (0.25 miles) for a bus stop and 800 meters (0.5 miles) for a rail station are generally accepted;
> 2) A decay in access by distance: a negative impact on transit use due to increasing walking distance; and,
> 3) Buffer analysis: a commonly used GIS method using the street network to define access areas (buffers) for transit service area analysis.

Based on previous studies, our approach to define a transit catchment area uses a street-network-based buffer with distance limits. Also, different values of these distance limits are explored through a sensitivity analysis, to measure the land-use type as well as its impact on transit demand.

2.3 Transit accessibility

Although transit accessibility has been recognized as one of the key factors that affect transit use, there are many varied definitions and analytic methods associated with "accessibility." Access (the ease with which people can reach the transit stop) and accessibility (the suitability of the transit system in helping people get to their destinations in a reasonable amount of time) are two very important issues in public transit planning (Murray et al. 1998). In one study, a public transit accessibility index was developed with a number of key variables including the distance to a bus stop, the number of services within walking distance, the frequency of service, the walking speed, and the time of day and day of week (Kerrigan and Bull 1992). Another study developed a transit accessibility and availability analysis tool that explicitly incorporates the time dimension, both from a supply and demand perspective. The result measures transit accessibility with regard to the number of daily trips per capita that are exposed to transit service (Polzin et al. 2002). A more recent study introduced a time-based method to analyze the transit service area, which incorporates total trip travel time into the transit service area (Cheng and Agrawal 2010). All of these studies enhance or incorporate the time dimension in measuring transit accessibility. For our study, transit accessibility is defined as the ability of residents and workers to reach transit facilities, including bus stops (Zhao 2003). As such, this study focuses on spatial accessibility to define service areas.

2.4   Integrated transportation and land use

One of the important considerations in transportation is the linkage between transportation and land use. This linkage is crucial to planning and implementing a transportation system, and has been studied by many professionals. The main stream of investigation into this linkage focuses on transit-oriented development (TOD), which aims to improve land use and transit operational efficiency (Cervero et al. 2002). According to Cervero and Kockelman (1997), three main land-use factors affecting transit use are the three D's: density, diversity, and design. Most prior models of this linkage have focused on how the built environment of TOD generates an increase in transit ridership. Measuring the impacts of land-use patterns on the transportation system, on the other hand, is another advantage of TOD. The disaggregate approach with individual bus stops gives a better understanding of the unique land use that surrounds each bus stop (Hsiao et al. 1997). Additionally, a specific land-use type will affect individual trip frequencies as well as the times of day and the days of the week when trips will occur (Hess et al. 2001). By looking at the number of dwelling units per parcel of land, the measure of potential transit demand at the most disaggregate level can be addressed (Kimpel et al. 2007). Our approach is very similar, but looks more generally at the variety of land-use classifications within a transit catchment area. Then, the temporal distribution of transit demand can be linked to a quantitative representation of the land-use type(s).

3   Methodologies

3.1   Two-level stop aggregation

In the first part of our analysis, we seek to aggregate transit stops based on some common characteristics and common service areas for the stops. A two-level (lower and upper level) stop aggregation method is introduced to better capture transit catchment areas. Figure 2 illustrates this method conceptually, showing the consideration of the distance between stops and a textual comparison of stop names. At the higher level, service characteristics at these consolidated stops are also considered.
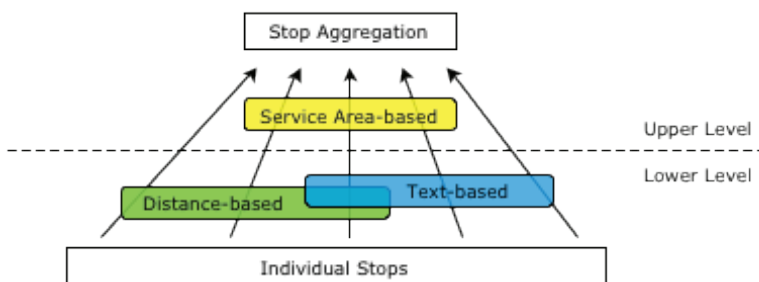

Figure 2: A hierarchical structure of stop aggregation. The hierarchical structure allows us to conduct the analysis at an individual stop or higher level.

Lower-level stop aggregation

In order to decide which individual stops should be combined geographically, two parameters (distance and text) were determined to be applicable. To infer the alighting stop and time from the transaction data, a method of aggregating stops is proposed (Lee et al. 2011), and Figure 3 illustrates how individual stops are aggregated.
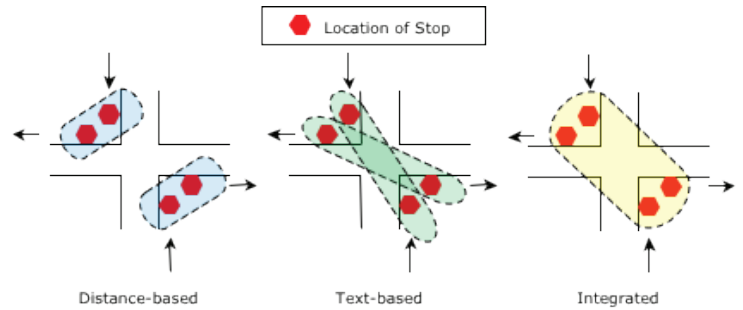

Figure 3: Conceptual approaches of lower-level stop aggregation.

- Distance-based method: This method groups stops based on a measure of proximity between individual stops. The Euclidean distance is the most commonly used similarity measure in cluster analysis techniques. This method relies on measuring the geographical proximity between stops. To determine an applicable distance parameter, bus stop spacing is also investigated, which varies depending on the agency and location (e.g., CBD, low-density residential, rural). For this study, as bus stop spacing in the CBD commonly ranges from 300 to 1000 feet (300 meters) (Fitzpatrick et al. 1996), 50 meters is selected as the distance parameter, considering mutually exclusive service areas between neighboring stops. With this, we combine stops into a single group if they are within 50 meters of each other.
- Text-based method: This method groups stops based on an identical or similar stop name. This method is based on an identical or similar street name around the same location (e.g., an intersection). So, if two stops share a common stop name, they are combined into a single group.
- Integration of distance- and text-based method: The distance- and text-based methods complement each other. So, in order to make a final list of aggregate stops, the groups from the distance-based method and the groups from the text-based method are combined to produce a set of aggregate stops.

Table 1: Example of lower-level stop aggregation. These four stops at the same intersection can be aggregated based on the integration of the distance- and text-based methods.

Stop ID	Stop_Name	Stop_Desc	Stop_City	Stop Aggregation	
16465	36 Av S & Lake St E	Near side S (SB)	Minneapolis	Text-based	Integration of distance- and text-based method
17590	36 Av S & Lake St E	Near side N (NB)	Minneapolis		
16920	Lake St E & 36 Av S	Near side E (EB)	Minneapolis	Text-based	
16921	Lake St E & 36 Av S	Near side W (WB)	Minneapolis		

Upper-level stop aggregation (Service area-based method)

The lower-level stop aggregation forms larger stop groups using the spatial and textual similarity between stops. Even though the lower-level stop aggregation is an effective treatment for capturing stops that ultimately influence the transit catchment area, this method falls short if the aggregation results in a stop group consisting only of one stop or of multiple stops in a single direction. In the first case, a single stop has not been "grouped" with any other stops. Such individual stops are easily observed when they are either located at a unique point along a one-way route or at a location where inbound and outbound stops are not co-located. Nonetheless, the transit catchment area may be affected by the same land uses. Based on this fact, service area-based stop aggregation (upper level) is introduced.

In this upper-level aggregation, we identify lower-level stop groups that have only one stop on a route, in a single direction. We wish to match this single stop with at least one stop for service traveling in the opposite direction. In this way, the aggregate stop allows a passenger to board and to alight along the same route. Practically, we match these lower-level stop groups so that each upper-level stop group has at least one stop in each direction, on each route serving the stop group. This matching is performed based on the most proximate stop groups containing stops in both directions on each route.

For example, let us assume that the travel of transit users on a given day starts and ends in the same location. A resident who lives in the parcel shown by the yellow polygon (in Figure 4) may use different stop groups at the lower level (red circle) depending on his or her direction of travel. This resident may use one stop group (the second red circle from the bottom) for boarding a bus headed northbound and a different stop group (third red circle from the bottom) for alighting during southbound travel on a return trip. Thus, these two stop groups may be aggregated at the upper level (blue circle). Figure 4 illustrates the procedure and results of the upper-level stop aggregation (blue circles) for this example.

3.2 Transit demand of smart card users

Each transaction in the Metro Transit AFC system has associated operational information, including the date, time of day, route number, use type (whether it is an initial boarding or a transfer), fare type, etc. Since passengers can take advantage of free bus-to-bus transfers within 2.5 hours of the first transaction, a passenger's transactions with "use type = 9" indicate an originating trip on that day, whereas any transactions with "use type = 1" indicate transfer trips within the passenger's tour. For the purposes of describing transit "trips," transfer trips are not counted as generating transit demand at a given stop.

Because the AFC data have the GPS location of transactions (not the actual bus stop used), identification of the boarding stop is the first step. A method of inferring boarding stops using a geographic information system (GIS) and database management tools was developed, showing that intersection-level boarding stops are successfully inferred (Lee and Hickman 2011). Using this methodology, the boarding stop of each transaction can be identified at the upper level. Similarly, by looking at both the duration between consecutive transactions and whether the transaction is an initial boarding or a transfer, a passenger's activity can be observed temporally and spatially. From our data set, we also can observe such passenger behavior over several days.

Lower level   Upper level
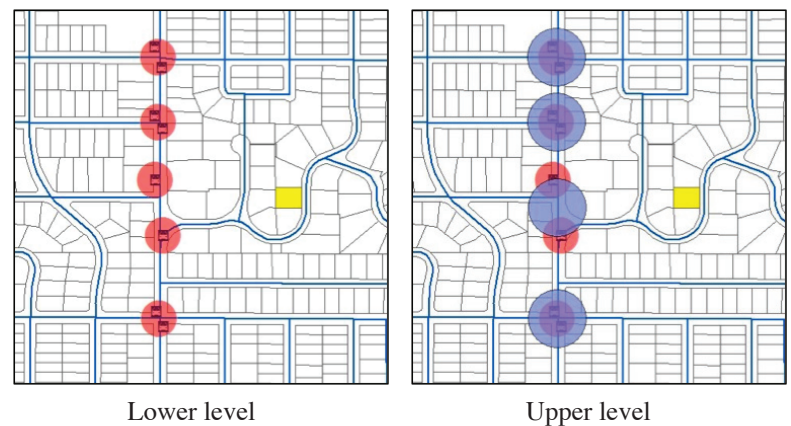
Figure 4: Developing upper-level stop aggregation.

3.3   Transit service areas

To delineate the service area of a stop, a feature-based proximity analysis is considered. This analysis can be defined by either a straight-line (e.g., Euclidean) or a network distance (e.g., walking along roads). O'Neill et al. (1992) describe a procedure for determining transit service areas, comparing results between the straight-line and network distance measurements. They also suggest developing a buffer area by identifying each street segment within a specified distance of the stop and buffering it. This procedure is applied to determine transit catchment areas (Hsiao et al. 1997). However, a circular buffer with a predefined radius seems unrealistic because it may ignore any indirect paths and obstructions, or it may define an area inaccessible to pedestrians. Upchurch et al. (2004) propose a combination of Thiessen polygons and maximum-distance Euclidean buffers for service areas that are both mutually exclusive and bounded by a fixed radius. The network-based method, which delineates transit service areas that are equally distant from a stop along all available network paths, is popular in GIS applications (Kimpel et al. 2007; Cheng and Agrawal 2010). Gutiérrez and García-Palomares (2008) describe the overestimation of catchment areas caused by the straight-line distance method when compared with the network-based method. As a result, in our approach, the network-based method is used for determining transit service areas.

We use the Network Analyst extension for ArcGIS (ESRI 1996) to develop these service areas. First, individual stops are turned into candidates for facilities using the "Service Area" function in ArcGIS. Second, we create polygons going from the facility to a parcel, using an accessible walking path. Third, we join and merge polygons of multiple facilities using the upper level stop aggregation. Fourth, the parcel-level land-use data is selected by using the catchment area polygon. Figure 5 illustrates the result.
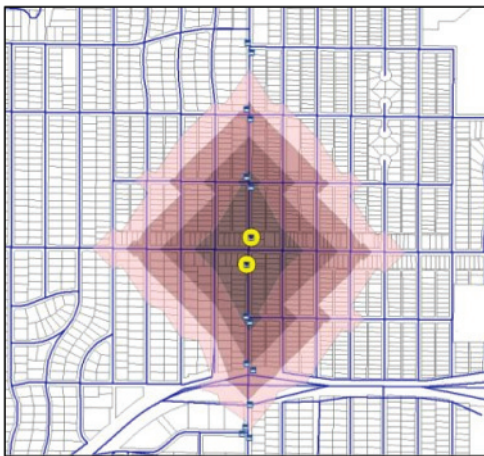

Figure 5: Example of defining transit service area at the upper level. This figure illustrates stop catchment area based on two stops (yellow, at the upper level) with respect to different distances (200, 300, 400, and 500 meters).

3.4   Measure of land-use types

Once the transit service area at the upper level is captured, it is important to identify trip generators/attractors (e.g., land-use types) within this area. For this, several inherent attributes, such as the land-use description and parcel acreage, can be used to measure land-use types that are likely to influence transit use. For this study, as shown in Table 2, seven conventional land-use types are chosen for reclassification. This classification is simply one possible method of aggregation of land-use types; practitioners also can customize this reclassification to suit their own interests.

A parcel, however, may have multiple descriptions, if the property has mixed uses (e.g., residential and commercial). In this case, specific heuristics should be applied consistently in the data set. For example, when a parcel indicates residential as its primary use, but commercial as its secondary use, this parcel may be regarded as residential. When a parcel indicates commercial as its primary use, but university as its secondary use, the parcel is regarded as institutional.

Table 2: Reclassification of land use

Original Classification	Reclassification (*i*)
Private dwelling (e.g., apartment, triplex, condominium)	Residential (1)
Business service, Retail store, Restaurant, Hotel	Commercial (2)
School, University, College, Library, Church	Institutional (3)
Hospital, Governmental office (e.g., municipal office)	Industrial (4)
Park, Sport facility	Recreational (5)
Vacant (e.g., vacant - commercial)	Vacant (6)
Not classified elsewhere (e.g., common space, roadway)	Others (7)

Next, we may quantify land-use indicators. In this study, we develop area and numerical proportions of each land-use type as the selected indicators. A large proportion of the property area, and the number of observations of a particular land-use type, may indicate a predominant land-use type. More precisely, these two measures can be calculated as the proportion of total area and the proportion of total number of parcels, respectively:

$$AP_i = \frac{A_i}{\sum_{i=1}^{I} A_i} \quad \text{and} \quad NP_i = \frac{N_i}{\sum_{i=1}^{I} N_i}$$

where
AP_i = the Area Proportion of land-use type i
NP_i = the Numerical Proportion of land-use type i
A_i = the area of land-use type i in the catchment area
N_i = the number of observations of land-use type i in the catchment area
I = the number of different land-use types ($I = 7$ in our reclassification)

These land-use indicators were developed in the presence of multiple-story buildings, since our approach to identify trip generators/attractors is based on the land-use type itself within a specified catchment area, instead of estimating the actual transit demand from that area. The incremental changes of distance (e.g., 200, 300, 400, and 500 meters) are used to assess the impact of land use within each stop's catchment area.

One weakness of these measures is that they do not account for general travel demand generated based on the demographic and socioeconomic characteristics (e.g., population, employment, vehicle availability). While the accessibility measures do not give a direct indication of trip generation, these measurements give a good sense of the land-use types that are influencing transit demand in that service area. In addition, these are very straightforward indicators under limited availability of attributes (e.g., size and description of each parcel) and can be easily applied by practitioners if the accurate transit service area is defined.

3.5 Utilization of data by GIS and SQL

Figure 6 presents the use of data as well as the computational method at each level. ArcGIS 9.3, a commercial GIS with the Network Analyst extension (ESRI 1996), is used to support buffer analysis and selection of parcel-level land-use data, while SQL is used to manipulate AFC and GTFS data.
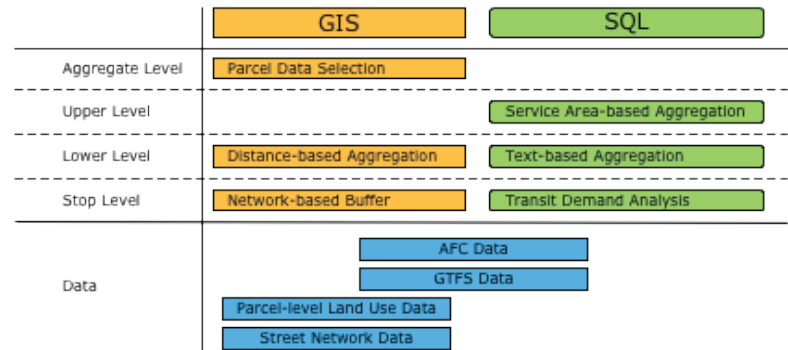
Figure 6: Utilization of various relevant data by GIS and SQL.

4 Data

This section describes the collection and processing of data used for the case study of Minneapolis–St. Paul, including a brief description of each data set, focusing on the attributes used in this study.

4.1 Data description

Automated fare collection (AFC) data: Go-To cards

The AFC data includes 2.17 million transactions for the full month of November 2008. Passengers must validate their transit pass/card ("Go-To card") each time they board a bus. Card readers on the buses are used to collect fares and record transactions. Each transaction has basic operational information: transaction date and time, route number, fare type, use type (e.g., indicating initial use or transfer), and the GPS location. Each Go-To card is assigned a unique special serial number (SSN), which can be used as the primary way of tracking the behavior of an individual passenger.

Google's General Transit Feed Specification (GTFS)

GTFS is an open format updated by hundreds of transit agencies in the United States and used by Google to incorporate transit information (e.g., routes, stops and schedules) into applications (GTFS 2010). Using the GTFS data allows us to use the detailed service schedule (488,105 rows on SEP08-Multi-Weekday-01 in stop_time.txt) and location of individual stops (14,601 stops in stops.txt) from November 2008.

Parcel-level land-use data

The parcel-level land use data from 2008 includes a standard set of attributes (e.g., address, description of land use, status of tax exemption) based on each tax parcel polygon (MetroGIS 2010). These data also provide the size of the polygon. Multiple tax parcels are sometimes represented by a single polygon, which implies a multiple-story building. One issue to be considered is that there may be multiple descriptions for a single tax parcel.

Street network data

The street network data in 2008 includes all roadways including local streets. The street network data is taken from the Census Bureau's TIGER files (US Census Bureau 2010). One important consideration with the TIGER files is the level of accuracy, including its precision in geographic coordinates and its representation of small segments such as walking paths.

4.2   Data preparation

Metro Transit schedules are defined separately for weekdays, Saturdays, and Sundays. The period from 3:00 a.m. on one day to 3:00 a.m. on the next day is used as a "service" day. For between-day variability, a set of typical weekdays, specifically Monday to Friday, November 17 to 21, is selected. In the data preprocessing, all AFC transactions that do not have latitude-longitude data or have suspicious route numbers (e.g., 59999) are eliminated from the data set. As a result of the fare policy that a passenger may take advantage of free bus-to-bus transfers within 2.5 hours, transactions that are not considered valid for our study include the following cases:

- Transactions for which the first transaction on a given day indicates a transfer transaction
- Transactions for which the duration between two successive transactions is more than 2.5 hours, but where the second indicates a transfer transaction

- Transactions for which the duration between two successive transactions is less than 2.5 hours, but where the second indicates an initial transaction

SQL is then used to ask for transactions satisfying these conditions. "Outliers" represent those transactions where the nearest stop location on the given route is greater than a pre-specified distance from the AFC transaction location (e.g., a quarter-mile is used in our study). Additional attributes such as the sequence of transactions of each cardholder and duration between transactions are calculated. We also generate a list of all stops served by each route, with or without directional information from the schedule. These lists are crucial to identify the boarding stop as well as an alternative alighting stop, in terms of finding the nearest stop in the opposite direction on each route (Lee et al. 2011).

5   Case study

5.1   Application to stop aggregation

A case study using the stop list from the Metro Transit GTFS data is performed. As shown in Table 3, the lower-level stop aggregation method reduces the network complexity. Compared with using the individual stops (14,601), the aggregated stops (7924) significantly reduce the number of stop groups. One important observation here is the high proportion of stop groups composed of only two stops (e.g., just including the opposite stop along a route).

Table 3: Lower-level stop aggregation using GTFS data for the Minneapolis-Sr. Paul area.

Number of stops	Number of groups		
	Distance-based	Text-based	Integrated
1	2361	5453	2117
2	5216	4490	5272
3	256	35	267
4	206	7	228
5	25	4	26
6	8		9
7	1	1	1
8	2	1	2
9			
10	2		2
Total	8077	9991	7924

   The upper-level stop aggregation method presented above is also tested along Route 6 (Figure 7).
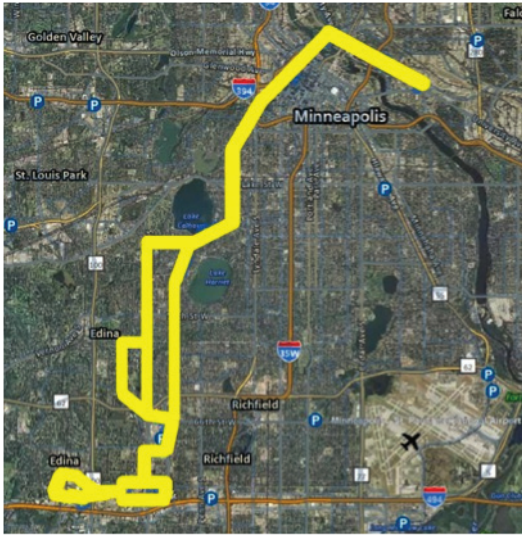
Figure 7: Map of Route 6, which runs north-south along Hennepin Avenue from the University of Minnesota, through downtown Minneapolis, through the Uptown transit center and Southdale shopping center, to the southern suburb of Edina.

Figure 8 demonstrates the whole procedure of stop aggregation. The results from two-level stop aggregation clearly indicate that stop groups at the upper level (162 groups in both directions) are more representative in terms of having an identical or similar transit service area. It is likely that these stop groups can be represented by a single node in our network representation.

Individual stop		Lower level		Upper level	
Total	334	Total	214	Total	162
		Individual	72		
Northbound	163	Northbound	38		
Southbound	163	Southbound	34		
Shared	8	Integrated	142		

Figure 8: Two-level stop aggregation: A single route is illustrated for two-level stop aggregation, based on a portion of Route 6.

5.2 Application to example locations

We explored land-use indicators for four example locations in Hennepin County (the west side of Minneapolis). As stated in the introduction, our goal is to investigate the interaction between transit demand and land-use types within the transit catchment area. For better visualization, each component of the methodology (the temporal distribution of transit demand, the definition of the stop's catchment area, and the land-use indicators of area and numerical proportion) is illustrated in a single row.

Figure 9 shows that the majority of transit demand at this first stop group is observed during the afternoon (4:00 p.m. to 6:00 p.m.). In the aggregate stop's catchment area, the two land-use indicators show the highest proportion in commercial, which is consistent with the expected pattern of persons leaving a workplace. It is also observed that the indicators change dramatically at 500 meters, which means that some residential areas are appearing between 400 and 500 meters from this stop group. A second stop group is shown in Figure 10, indicating that transit demand is much less concentrated during any specific hours. Obviously, the highest area proportion is institutional; however, two types, residential and institutional, are competitive in the numerical proportion. (Our analysis assumes that vacant land itself does not generate or attract transit demand; we do not include this classification in our analysis.) This is not only because the area proportion is largely institutional, but also because a lot of college and university campus housing (e.g., dormitories and sorority or fraternity housing) are counted as residential in our reclassification.

Two other examples, Figures 11 and 12, show that the majority of transit demand at these locations is observed during the morning (6:00 a.m. to 8:00 a.m.), with a little longer distribution toward the afternoon in Figure 12. The two land-use indicators are dominated by residential, which might be expected to have this kind of temporal distribution of transit demand, with commuters leaving home to travel to work.

These case studies provide evidence to support our assertion that the temporal distribution of transit demand at a stop's catchment area can be estimated with parcel-level land-use data. To further explore this concept, two locations in Ramsey County (compared with Hennepin County in the previous case study) are tested. We first identify the stop's catchment area, and then measure the two land-use indicators. Based on these results, the temporal distribution of transit demand appears to be similar to that observed in the Hennepin County. This procedure is also illustrated in two examples, shown in Figures 13 and 14.
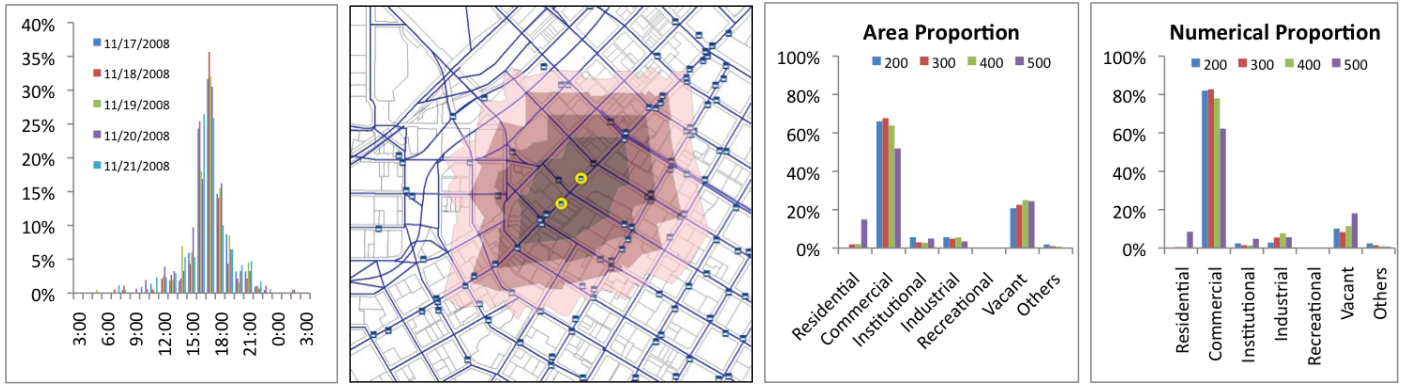
Figure 9: Day-to-day transit demand, stop's catchment area, and two land-use indicators (Location 1).
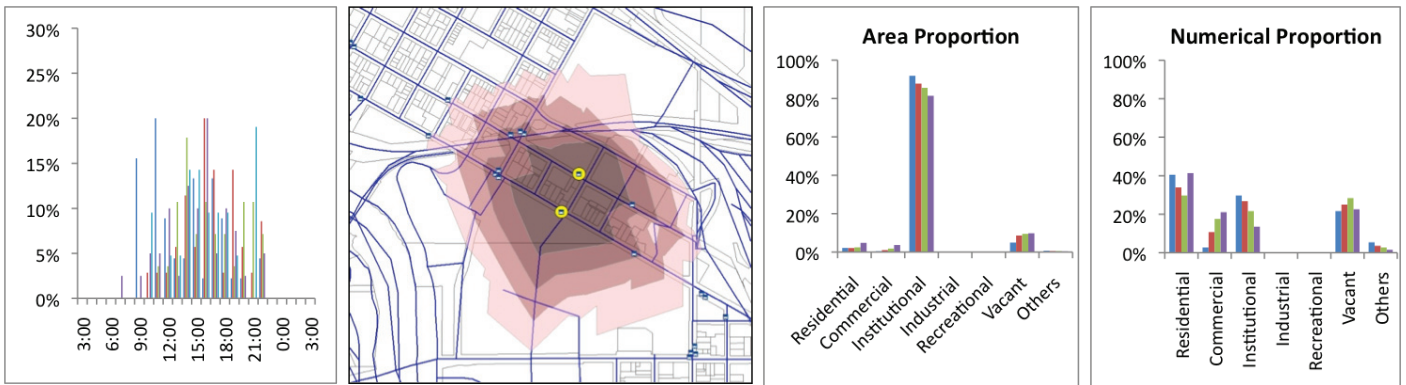

Figure 10: Day-to-day transit demand, stop's catchment area, and two land-use indicators (Location 2).
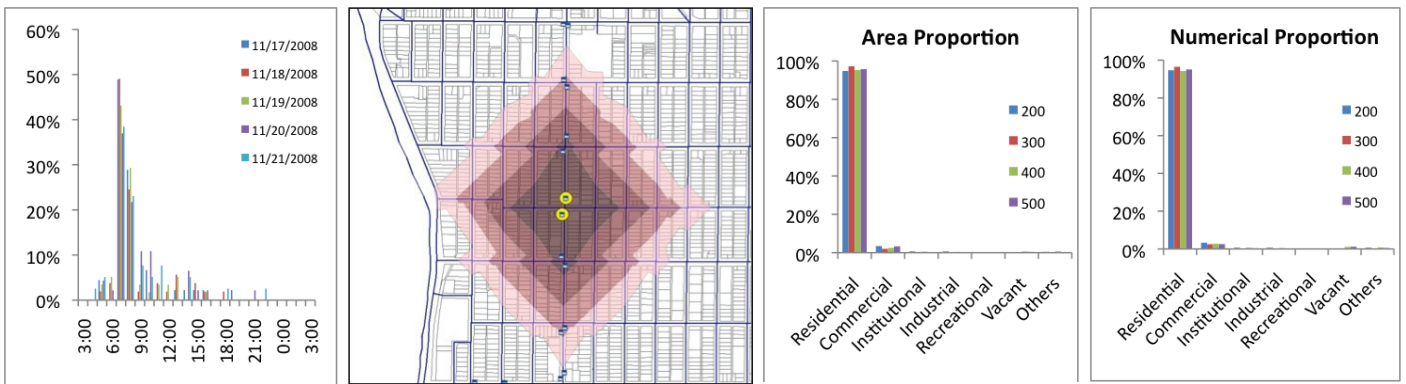

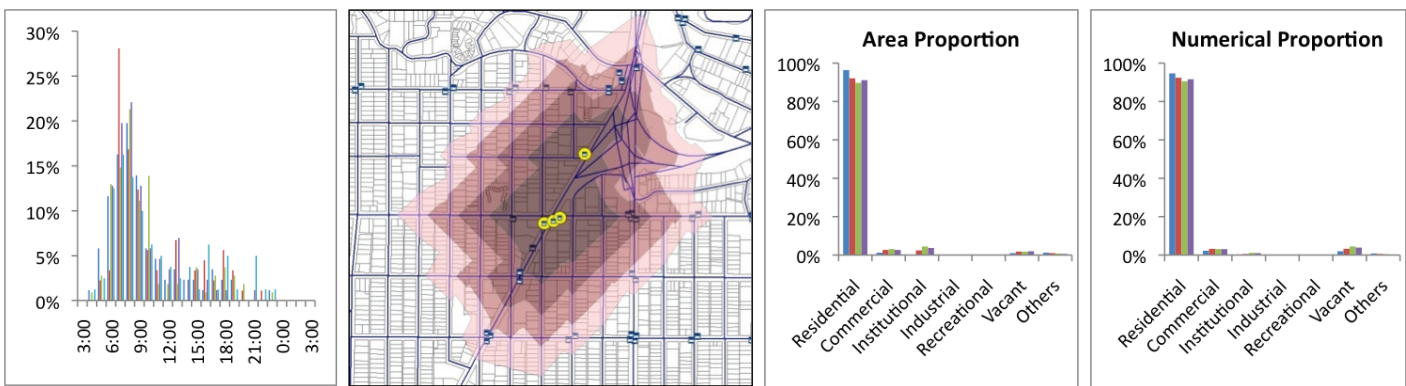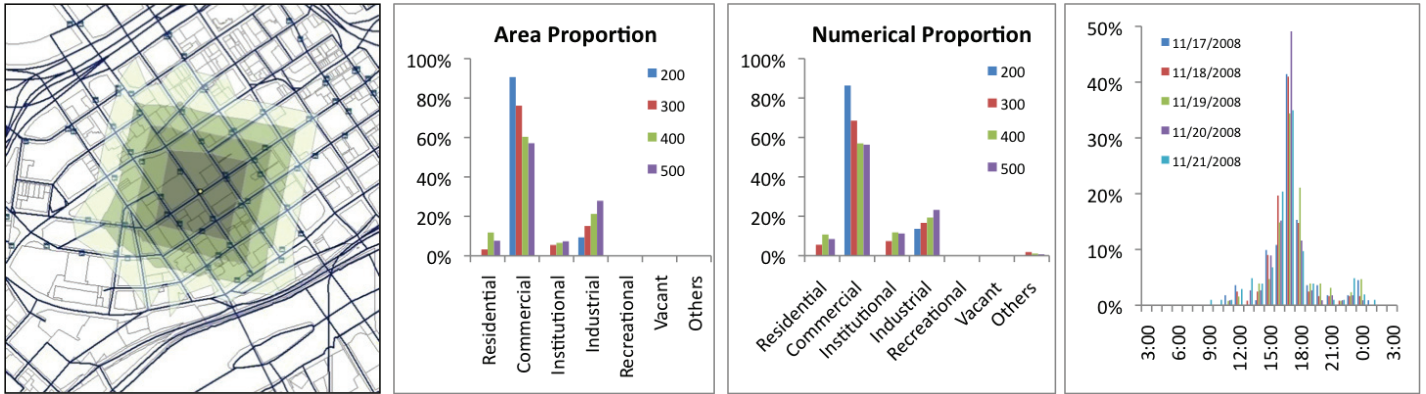**Figure 11:** Day-to-day transit demand, stop's catchment area, and two land-use indicators (Location 3).

Figure 12: Day-to-day transit demand, stop's catchment area, and two land-use indicators (Location 4).

Figure 13: Stop's catchment area, two land-use indicators, and day-to-day transit demand (Location 5).
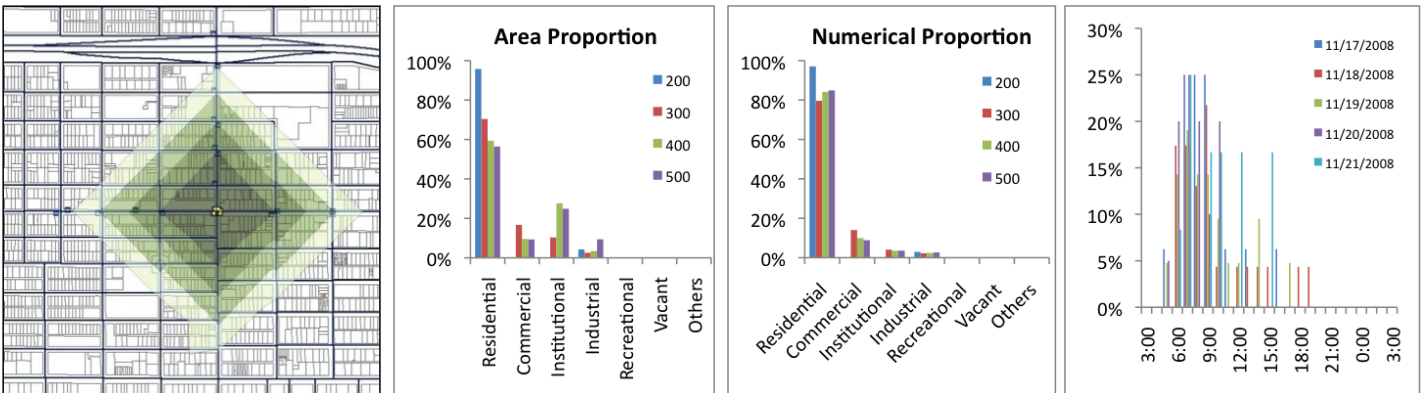

Figure 14: Stop's catchment area, two land-use indicators, and day-to-day transit demand (Location 6).

5.3   Limitations

Capturing accurate transit service areas is a challenge. While a street network-based buffer is a popular method for service area analysis, this method also has limitations in applying distance parameters and defining the service area. The TIGER street network includes streets (e.g., highway and ramp in Figure 15, on the left) that are not accessible to pedestrians, and the network does not easily represent the pedestrian-level network (e.g., shortcuts, sidewalks and pathways, open space, parking lots) (Upchurch et al. 2004; Cheng and Agrawal 2010).
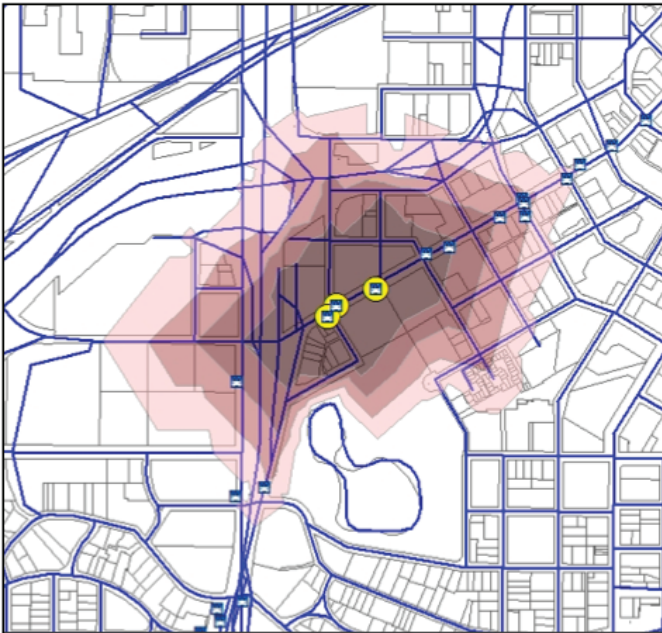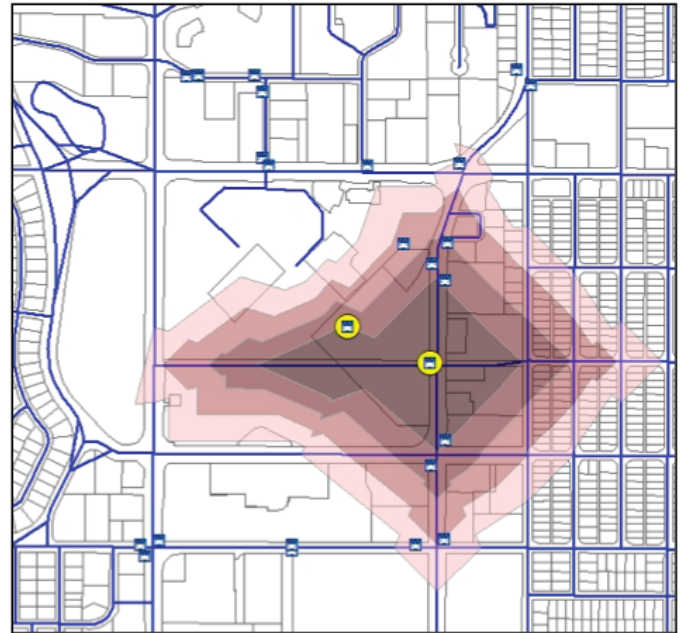
6   Conclusion and future work

6.1   Conclusion

The primary objective of this study was to develop and demonstrate a method to measure land-use types within a specified service area based on stop aggregation. This research helps in advancing our understanding of transit demand in the vicinity of bus stops through an analysis of detailed land-use types and appropriate knowledge of major trip generation and attraction for the Twin Cites region. Our stop aggregation model

Figure 15: Examples of street network buffer limitations.

The other limitation is that the potential land uses within a stop's service area may be also constrained by an isolated network (e.g., a park in Figure 15, on the left, or a shopping center, on the right) in the TIGER files. Although this park is in close proximity to several bus stops, a street network-based service area does not capture this area. The key problem seems to be the inaccuracy and lack of connectivity in the TIGER street network. Adding connecting lines in this case (e.g., based on an aerial photo or other knowledge) is needed to prevent these problems. One effort is to set up more accurate street network by some minor edits (e.g., adding a walking path) (Furth et al. 2007a).

may improve our understanding of the relationship between transit demand and land-use patterns. This method can better utilize and enhance AFC data to obtain a better understanding of travel patterns.

The proposed two-level stop aggregation method offers numerous potential applications for engineers and planners. First, this aggregation is a stronger foundation that can reduce the complexity of the transit network. The advantage of stop aggregation is that stop groups that serve common or similar land uses and activities can be represented by a single node. Use of GTFS in the stop aggregation method is another advantage because the GTFS data can provide information on trends in temporal and spatial transit characteristics to better estimate future patterns of demand. A set of metrics to better represent land-use types in a specified transit service area was also developed to connect to the transit demand data automatically col-

lected from various electronic technologies. By looking at these measures, this study builds the spatial and temporal knowledge of transit passenger demand at a specified service area. The procedure was applied to a case of Minneapolis–St.Paul metropolitan area using various relevant data and can be expected to be applicable to better understand the linkage between transit demand and land-use patterns.

6.2   Future Work

The methods employed in this study can be extended several ways. First, developing accurate network data sets (e.g., a pedestrian-level network) would not only define a stop's catchment areas of transit users, but also enhance the parcel-level data selection. Even though a sensitivity analysis (using different values of distance limits) has been conducted to measure land-use types and changes, more work could be done. Specifically, in the stop aggregation methods, one should consider different values of stop spacing in determining different distance parameters that depend on the service environment (e.g., urban and suburban). Another important point that might be pursued is how supply-side characteristics (e.g., peak and off-peak frequencies, local and express routes) can be reflected in the stop aggregation and the associated transit catchment areas.

7   Acknowledgments

References

Barry, J. J., R. Newhouser, A. Rahbee, and S. Sayeda. 2002. Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record: Journal of the Transportation Research Board*, 1817: 183–187.

Beimborn, E. A., M. J. Greenwald, and X. Jin. 2003. Accessibility, connectivity, and captivity: Impacts on transit choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1835: 1–9.

Cervero, R. and K. Kockelman. 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research A,* 31(3): 199–219.

Cervero, R., C. Ferrell, and S. Murphy. 2002. TCRP Research Results Digest 52: Transit-oriented development and joint development in the United States: A literature review. Transportation Research Board of the National Academies: Washington, DC.

Cheng, C. and A. W. Agrawal. 2010. TTSAT: A new approach to mapping transit accessibility. *Journal of Public Transportation,* 13(1): 55–72.

Crowley, David F., A. S. Shalaby, and H. Zarei. 2009. Access walking distance, transit use, and transit-oriented development in North York City Center, Toronto, Canada. *Transportation Research Record: Journal of the Transportation Research Board,* 2110: 96–105.

El-Geneidy, A. M., P. Tetreault, and J. Surprenant-Legault. 2010. Pedestrian access to transit: Identifying redundancies and gaps using a variable service area analysis. *Proceedings of the Transportation Research Board Annual Meeting*, Washington, DC.

Environmental Systems Research Institute (ESRI). 1996. *ArcView network analyst: Optimum routing closest facility and service area analysis*. ESRI, Inc.: Redlands, CA.

Fitzpatrick, K., K. Hall, D. Perkinson, L. Nowlin, and R. Koppa. 1996. *Guidelines for the location and design of bus stops.* Transit Cooperative Research Program Report No. 19. Texas Transportation Institute, Transportation Research Board, National Research Council, Washington, DC.

Furth, P. G. and D. S. Navick. 1992. Bus route O-D matrix generation: Relationship between biproportional and recursive methods. *Transportation Research Record: Journal of the Transportation Research Board,* 1338:14–21.

Furth, P. G. and A. B. Rahbee. 2000. Optimal bus stop spacing through dynamic programming and geographic modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 1731:15–22.

Furth, P. G., M. C. Mekuria, and J. L. SanClemente. 2007a. Stop spacing analysis using geographic information system tools with parcel and street network data. *Transportation Research Record: Journal of the Transportation Research Board,* 2034:73–81.

Furth, P. G., M. C. Mekuria, and J. L. SanClemente. 2007b. Parcel-level modeling to analyze transit stop location changes. *Journal of Public Transportation,* 10(2): 73–91.

General Transit Feed Specification (GTFS) Data Exchange. 2010. http://www.gtfs-data-exchange.com/ (Accessed December 2010).

Gutiérrez, J. and J. C. García-Palomares. 2008. Distance-measure impacts on the calculation of transport service areas using GIS. *Environment & Planning B: Planning & Design*, 35: 480–503.

Hess, P. M., A. V. Moudon, and M. G. Logsdon. 2001. Measuring land use patterns for transportation research. *Transportation Research Record: Journal of the Transportation Research Board,* 1780:17–24.

Hsiao, S., J. Lu, J. Sterling, and M. Weatherford. 1997. Use of geographic information system for analysis of transit pedestrian access. *Transportation Research Record: Journal of the Transportation Research Board,* 1604:50–59.

Kerrigan, M. and D. Bull. 1992. Measuring accessibility: A public transport accessibility index. *Environmental Issues: Selected Proceedings of the Seminar B PTRC Summer Annual Meeting.* PTRC Education and Research Services, 245–256.

Kimpel, T. J., K. J. Dueker, and A. M. El-Geneidy. 2007. Using GIS to measure the effect of overlapping service areas on passenger boardings at bus stops. *Urban and Regional Information Systems Association (URISA) Journal,* 19(1): 5–11.

Lee, S. G., and M. D. Hickman. 2011. Travel pattern analysis using smart card data of regular users. *Proceedings of the 90th Annual Meeting of the Transportation Research Board,* Washington, DC.

Lee, S. G., A. Khani, D. Tong, and M. D. Hickman. 2011. Aggregate level origin-destination estimation using smart card data and transit schedule. In *Proceedings of the 2011 TRF Annual Forum*, Proceedings of the 52nd Annual Forum. Long Beach, CA.

Loutzenheiser, D. R. 1997. Pedestrian access to transit: Model of walk trips and their design and urban form determinants around Bay Area Rapid Transit stations. *Transportation Research Record: Journal of the Transportation Research Board,* 1604: 40–49.

MetroGIS. 2010. http://www.metrogis.org/ (Accessed December, 2010).

Murray, A. T., R. Davis, R. J. Stimson, and L. Ferreira. 1998. Public transportation access. *Transportation Research D,* 3: 319–328.

O'Neill, W. A., R. D. Ramsey, and J. Chou. 1992. Analysis of transit service areas using geographic information systems. *Transportation Research Record: Journal of the Transportation Research Board,* 1364: 131–138.

O'Sullivan, S. and J. Morrall. 1996. Walking distances to and from light-rail transit stations. *Transportation Research Record: Journal of the Transportation Research Board,* 1538: 19–26.

Polzin, E. S., R. M. Pendyala, and S. Navari. 2002. Development of time-of-day-based transit accessibility analysis tool. *Transportation Research Record: Journal of the Transportation Research Board,* 1799: 35–41.

Schlossberg, M. and N. Brown. 2004. Comparing transit oriented developments based on walkability indicators. *Transportation Research Record: Journal of the Transportation Research Board,* 1887: 34–42.

Upchurch, C., M. Kuby, M. Zoldak, and A. Barranda. 2004. Using GIS to generate mutually exclusive service areas linking travel on and off a network. *Journal of Transport Geography,* 12(1): 23–33.

US Census Bureau. 2010. US Census TIGER files. http://www.census.gov/geo/www/tiger/ (Accessed December 2010).

Weinstein, A., B. Vanessa, I. Katja, and M. Schlossberg. 2007. How far, by which route, and why? A spatial analysis of pedestrian preference. *Proceedings of the 2007 Transportation Research Board Annual Meeting*, Washington, DC.

Zhao, F., L-F. Chow, M-T. Li, I. Ubaka, and A. Gan. 2003. Forecasting transit walk accessibility: Regression model alternative to buffer method. *Transportation Research Record: Journal of the Transportation Research Board*, 1835: 34–41.