SpAM

SpAM (Spatial Analysis and Methods) presents short articles on the use of spatial statistical techniques for housing or urban development research. Through this department of Cityscape, *the Office of Policy Development and Research introduces readers to the use of emerging spatial data analysis methods or techniques for measuring geographic relationships in research data. Researchers increasingly use these new techniques to enhance their understanding of urban patterns but often do not have access to short demonstration articles for applied guidance. If you have an idea for an article of no more than 3,000 words presenting an applied spatial data analysis method or technique, please send a one-paragraph abstract to rwilson@umbc.edu for review.*

Who Can Access Transit? Reviewing Methods for Determining Population Access to Bus Transit

Steven Biba
University of Texas at Dallas

Kevin M. Curtin
Germana Manca
George Mason University

Abstract

This article explores the use of Geographic Information Systems in determining transit-service areas. The traditional methods of determining the population that can access transit are briefly reviewed, and a new method is proposed. The parcel-network method takes advantage of the spatial and aspatial attributes of parcels, and the ability to easily determine network distances from parcels to bus stop locations. This parcel-network method avoids the well-known and unrealistic assumptions associated with the existing methods, and reduces overestimation of the population with access to transit, resulting in improved spatial precision and superior inputs to transit-service decisionmaking processes. The way in which this new method is performed is examined in detail. The study area consists of a section of the bus network within the Dallas metropolitan area. This article summarizes the work of the authors in a research article entitled, "A New Method for Determining the Population With Walking Access to Transit" and published in the International Journal of Geographical Information Science *(Biba, Curtin, and Manca, 2010).*

Introduction

It has long been recognized that Geographic Information Systems (GIS) are a useful tool for transportation modeling given the ability to realistically model linear and network features (Curtin, 2008a, 2008b, 2007). These features can, in turn, be used to determine the service areas—and the associated populations within those areas—for transit routes in order to perform transit planning (DART, 2002). Unfortunately, several commonly used methods consistently overestimate the population with access to transit (O'Neill, Ramsey, and Chou, 1992; Zhao et al., 2003), and therefore transit planners can lose confidence in the ability of GIS to provide reliable answers to questions concerning the changes in accessibility that will be achieved by changes in service.

This article reviews the traditional GIS methods for determining access to transit and presents an improved method for doing so—the parcel-network method. This method employs disaggregate cadastral data with network functionality to determine walking distances to transit facilities in order to more precisely estimate the population with access to transit. This method provides more conservative population estimates than the methods currently in use, it produces summary statistics and output datasets that cannot be generated through the other methods, and it provides a more flexible basis for further refinement of the transit forecasting and planning process. Examples are included using a set of data from the north Dallas region. A quantitative comparison of the results is made for the buffer, network-ratio, and parcel-network methods. This article summarizes the work in a research article entitled, "A New Method for Determining the Population With Walking Access to Transit" published in the *International Journal of Geographical Information Science* (Biba, Curtin, and Manca, 2010).

Traditional Methods for Determining Transit-Service Areas

Historically, transit-service area determination has been implemented within GIS by creating a distance buffer around the transit route, or around stops along that route (Ayvalik and Khisty, 2002; Hsiao et al., 1997; Peng et al., 1997), and estimating the population within that buffer based on an overlay of census polygons. The buffer method (or area ratio method) assumes a uniform distribution of population within the census polygon, which is frequently not the case. Moreover, this method implies that the entire population within the buffer has walking access to the transit route. Unfortunately, the buffer method has been shown to consistently overestimate the population within the service area since the actual walking distances within the buffer are greater than the Euclidean distances used to generate the buffer (Horner and Murray, 2004; O'Neill et al., 1992). The level of overestimation has been shown to be influenced by the size of the analysis zone, and therefore experts recommend that the smallest practical level of population aggregation be employed (Handy and Niemeier, 1997; Lee, 2005).

The Network-Ratio Method

O'Neill et al. (1992) used a refined method—termed the network-ratio method—to more accurately measure accessibility to transit services. This method considered the total length of the street network within analysis zones surrounding a transit route, and the length of the streets within those zones that are also within a specified network distance from the transit stops. The formula for computing population with access to transit with the network-ratio method is—

$$A_i = (W_i / M_i) \times P_i. \qquad (1)$$

Where: A_i = the population in analysis zone i with access to transit, M_i = the total length of street network in analysis zone i, W_i = the length of the street network within walking distance to transit in zone i, and P_i = the total population of zone i.

Although the network-ratio method eliminates the error associated with the assumption of uniform population distribution over census polygons, it does assume that population along a street is proportional to street length, and that there is a uniform distribution of population on every street. This assumption is particularly weak in mixed residential zones or zones with retail, industrial, and recreational activities, which are precisely the kinds of areas that are likely to have transit routes. Lastly, if the network-ratio method includes all roads within walking distance of a transit facility, highways and their associated frontage roads and off-ramps can lead to substantial error in the population estimates. In fact, research has shown that these limitations of the network-ratio method lead to consistent overestimation of the population with walking access to transit, although the errors are not as large as those seen with the buffer method (Zhao et al., 2003).

Although we do not review the information here, the definition of "access to transit" is a subject of considerable research. We point the readers to Biba, Curtin, and Manca (2010) for a detailed review of the literature regarding walking distances to transit. Suffice it to say that the methods described below can be employed with any distance determined to be appropriate for a particular transit access study.

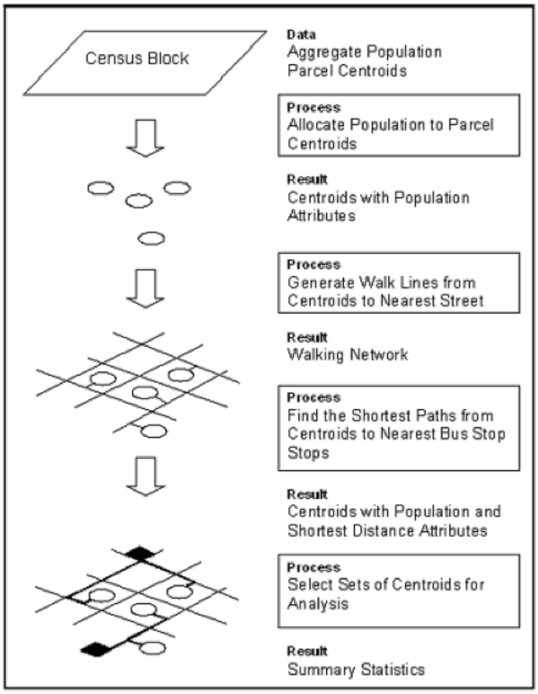
The Parcel-Network Method

The parcel-network method exploits both the superior precision of parcel databases and the network functionality that has been increasingly incorporated into GIS. Moreover, cadastral databases also include attribute information such as street addresses and land use or zoning categories (residential, commercial, and so on). Parcels provide a level of spatial accuracy that has not previously been exploited for transit access studies, and that accuracy allows for detailed modeling of walking access to transit facilities. The parcel-network method differs from traditional methods of determining access to transit in that it looks outward from the population locations (the parcels) to the transit features, rather than looking out from the transit features and making assumptions about what population lies within some distance of those features.

Implementing the parcel-network method requires four major steps, some of which have several sequential processes (exhibit 1). The first task is to apply demographic characteristics to the parcels; the second task creates a walking network from parcels to the transit facilities; the third task is to compute network walking distances from each parcel centroid to its nearest bus stop; and finally analysis is performed to assess the population that can access the transit facility across the walking network. These tasks represent a novel combination of parcel-based attribute imputation with network analytic techniques.

Exhibit 1

The Parcel-Network Method



Step 1. Demographic Attribute Transfer

In the absence of building footprints in the cadastral database (as was the case in our data), the centroid of the parcel is a reasonable representation of the starting point for pedestrians seeking transit. Thus, the first step in the parcel-network method is to generate centroid points for all parcel polygons in the study area. Although the centroid could fall outside of the parcel polygon if the shape were irregular, in practice there are very few residential polygons where this is the case. Pertinent attribute information (based on the census blocks or other supplementary data) can then be associated with the parcel centroids. This step highlights the flexibility of the parcel-network method in that it allows users a great deal of latitude in choosing how they model the characteristics of the population under consideration. That is, the population of employees at commercial or industrial parcels could be considered, the population of students at parcels containing educational institutions could be an input, or the residential population could be of primary interest.

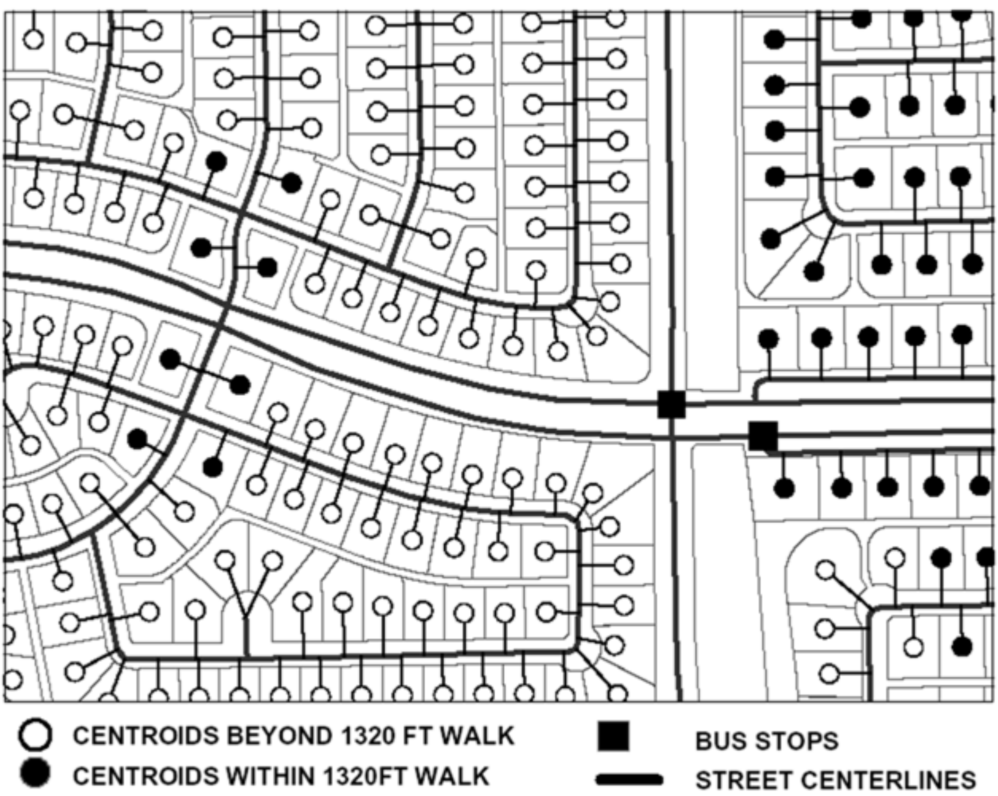
Since, in the research presented here, the object was to identify the total residential population with access to transit the total number of dwelling units per census block was computed, the population per dwelling was computed and transferred back to the parcel centroids, and the estimated population was computed for each parcel. Note that each parcel (and its centroid) is associated with a variable number of dwelling units. A parcel containing one single-family home would have only one dwelling unit associated with it. A parcel with multiple dwelling units (such as an apartment building) would be assigned a population in proportion to the total number of dwelling units in the census block. The result is a set of parcel centroids that now have reasonable population attributes. A wealth of other demographic data could be associated with parcel centroids in this way (income, car ownership, and so on) if a particular transit research project demanded it.

Step 2. Creating a Walking Network

Given that the street network does not connect to the parcel centroid points that now have associated attribute information, new links in that network must be generated in order to find the walking paths that pedestrians would follow. In order to do so an automated process was employed within the GIS to select each centroid, identify the point on the street network closest to that centroid, and generate a new network link connecting those two points. The result is a set of parcel centroids with population attributes and network connections to the walking network (exhibit 2). If one uses the addressed street to generate the walking path from the centroid it is virtually certain there will not be a barrier for a pedestrian.

Exhibit 2

Parcel Centroids and the Walking Network



Step 3: Computing Walking Distances

With the parcel centroids connected to the street network, an origin for a walking trip to transit has been defined. For the purposes of this research, the destinations are potential bus stops along the existing bus routes. In our study area potential bus stops are located at all intersections along a bus route.

The next step in the parcel-network method is to generate walking network distances between all of the origins and destinations, and then choose the smallest of those distances for each parcel centroid. There are extremely efficient shortest path algorithms implemented in industry standard GIS software products that can be used to populate such origin-destination matrices.

Once the shortest path to a bus stop is determined, the distance traveled and the closest bus stop can be transferred as attributes to the parcel centroids. At the conclusion of the first three stages of the parcel-network method, the result is a set of parcel centroids with population, the distance to the nearest bus stop, and the identifier for that bus stop.

Step 4: Determine Population With Transit Access

The fourth and final step in the parcel-network method is to use these data sources to determine the population with access to transit. We can now estimate the population in any distance band from a transit facility. In this study we chose to select those centroids where the walking distance is 0.25 mile or less to the bus stops of each distinct route. Based on this selection a summary of the population associated with each route can be made.

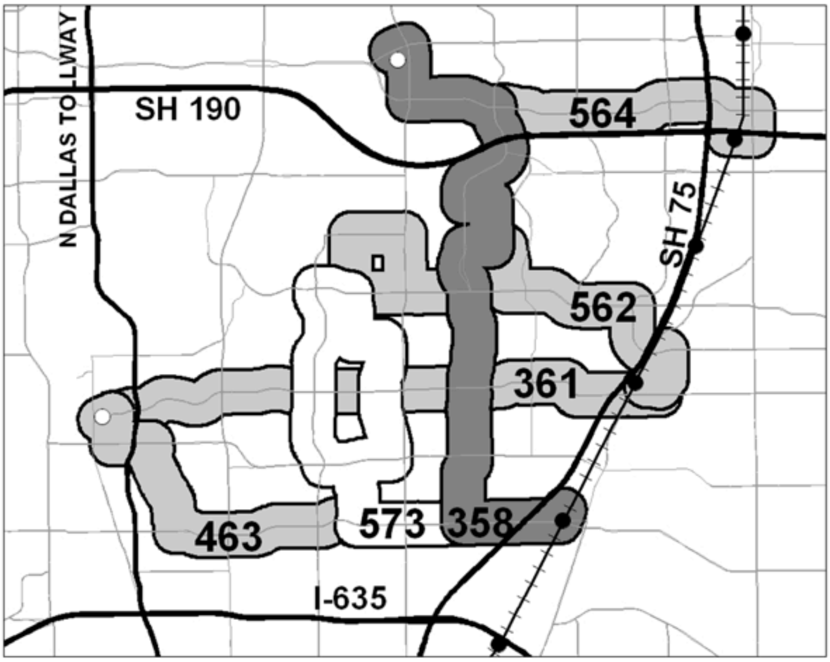
Data and Study Area

The study area for this research is an approximately 100-square-mile section of the Dallas Area Rapid Transit (DART) service area to the north of the city of Dallas in the communities of Richardson and Plano, Texas. We examine in detail six routes, chosen to represent the different types of route that DART manages. These include a dense multifamily corridor route (463); a transit express route (564); a university, multifamily, and commercial route (562); and several mixed multifamily and single-family neighborhood routes (358, 573, and 361) (exhibit 3). All of these routes serve at least one rail station and one transit center with the exception of routes 573 and 562.

Parcel centroids were reclassified as single family, multifamily, residential care facilities, or commercial parcels. Commercial centroids were included in creating the walking network, but they were assigned a zero dwelling count and therefore did not participate in the population allocation or contribute to the summary statistics. Residential care facilities, nursing homes, and group quarters were given a dwelling count of one if they had only a bed count. If they had a dwelling count, they were treated as any other multifamily facility. In the present study the centroids were reclassified and the number of dwelling units transferred using the land use polygon data layer, the development monitoring point file, aerial photography, various reference sources, and occasionally verification with field examinations.

Exhibit 3

Study Area Bus Routes



Results

The results are summarized in exhibit 4. As expected, both the buffer and network-ratio methods give larger population estimates than the parcel-network method in every case. More specifically, the buffer method always gives the largest estimate for each route. The average increase in estimate when comparing the buffer method with the parcel-network method is nearly 71 percent, although a single outlier affects this increase with an overestimation of 184 percent (this outlier on route 564 will be discussed in detail below). The elimination of this outlier results in an average overestimation of 48 percent.

Exhibit 4

Population Estimates

	Route Numbers					
	358	361	463	562	564	573
Buffer method	17,338	16,062	27,203	13,843	4,270	28,006
Overestimation	51%	53%	42%	54%	184%	41%
Network-ratio method	12,762	12,235	23,277	10,075	3,264	22,241
Overestimation	11%	16%	22%	12%	117%	12%
Parcel-network method	11,482	10,507	19,131	8,978	1,505	19,797

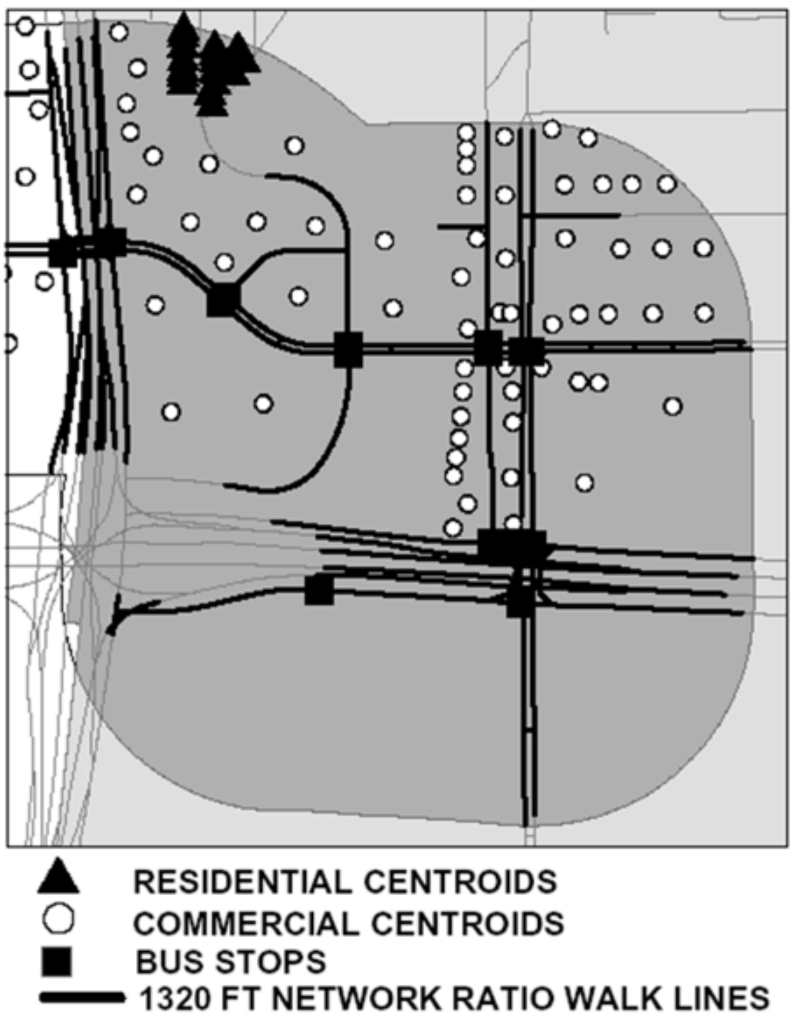
Note: All data computed within a 1,320-foot buffer.

The network-ratio method gives a smaller population estimate but still uniformly results in higher values than the parcel-network method. The average increase in estimate of 32 percent is also affected by the outlier route that influenced the buffer method, and the elimination of this route results in an average increase in estimate of 14.6 percent.

The overestimations given by the traditional methods can be dramatic. To demonstrate this, consider that in the study area being examined here, there are numerous cases where network links are close to transit facilities, but have no associated population (exhibit 5). In particular these network links are highway segments and arterials that are dominated by commercial or industrial land uses. The network-ratio method assigns population to these links even though no such population exists, resulting in an overestimation of the population with transit access.

Exhibit 5

Detail of Route 564



The problem is very apparent for Route 564, shown in detail in exhibit 5. The focus is on a single analysis zone associated with Route 564. The buffer method compares the area within a 0.25-mile buffer with the total area of the analysis zone. This ratio is then applied to the population of that analysis zone. In this case the area within the buffer zone is 16.5 percent of the total area of the census block group. Therefore, the buffer method suggests that 16.5 percent of the population of the block group has access to transit (369 people). Similarly, the network-ratio method compares the length of the streets within walking distance of the route to the total length of streets in the analysis zone. For the case being examined here, 16.8 percent of the streets in the analysis zone are within 0.25-mile walking distance of route 564. This would suggest that 376 people have access to transit.

As shown by the residential parcel centroids (represented with triangles on exhibit 5) the parcel-network method estimates that only a small population (32 people) actually lives within the 0.25-mile buffer in this census block. The other parcel centroids are commercial or industrial with no associated population. Of the 32 people, *none* has access to route 564 within a 0.25-mile walking distance. Therefore, the estimate from the parcel-network method is that *zero* people have walking access to this transit facility. This example highlights the severity of the overestimation that is possible using the buffer and network-ratio methods, and highlights the strength of the parcel-network method to accurately spatially identify the locations of potential transit users.

Conclusions and Future Research

The parcel-network method increases spatial precision because it relies on a discrete rather than a continuous allocation of attributes, unlike both the buffer method and the network-ratio method. The parcel-network method improves even further by allocating population to discrete points (parcel centroids), and no assumptions about population distribution over street length or analysis zone area are made.

The results above demonstrate that the parcel-network method can address well-documented issues in determining the population with access to a transit route; specifically the need to implement disaggregated population data (parcel data), and to model access via walking networks. The method eliminates the demonstrated overestimation that plagued previous methods due to their allocation schemes. By associating aggregate population data with parcel centroid points (rather than with areas or lines), and by creating a walking network that extends from the centroid of each parcel, the parcel-network method provides a degree of spatial precision that had not previously been possible without the use of data collected through survey research. In the future this method could be expanded to explore not only the physical access to transit, but also many other demographic characteristics (car ownership, income, age distribution, and so on) that could influence ridership patterns and therefore route optimality. Once the parcel-network method is implemented for a particular study area, the relative permanence of the parcel and street network data may allow for efficient update when new demographic data become available.

In addition to these improvements, the parcel-network method allows for analysis options that had not previously been available. Not only do we know the distance to the nearest transit facility but we can also generate the exact path that each (cost minimizing) transit user would follow. This path information could be used in several ways to improve transit accessibility. First, it could help in the study of neighborhood design in order to make improvements in existing pedestrian facilities through sidewalk improvements, elimination of barriers, or adding amenities. Second, the path information could be used to model changes to transit routes that would reduce total (or average) walking distances to the route for a chosen population.

The parcel-network method can also provide a more accurate starting point for application in other transit use models. A dataset prepared using the parcel-network method could be easily applied in transit use models that employ a distance decay formula in the estimation of transit use rather than transit accessibility. Perhaps most importantly, this article highlights the potential for the combination of highly detailed spatial information (in the form of cadastral databases) with network analysis techniques to advance research across GIScience.

Authors

Steven Biba is a former graduate student in the Department of Geography at the University of Texas at Dallas.

Kevin M. Curtin is an associate dean in the College of Science and associate professor in the Department of Geography and GeoInformation Science at George Mason University.

Germana Manca is an assistant professor in the Department of Geography and GeoInformation Science at George Mason University.

References

Ayvalik, Cemal K., and C. Jotin Khisty. 2002. "Heuristic Analysis of Impacts of Commuter Rail Station Consolidation on Pedestrian Access," *Transportation Research Record* 1793: 47–54.

Biba, Steven, Kevin M. Curtin, and Germana Manca. 2010. "A New Method for Determining the Population With Walking Access to Transit," *International Journal of Geographical Information Science* 24 (3): 347–364.

Curtin, Kevin M. 2008a. "Network Analysis." In *The Encyclopedia of Geographic Information Science*, edited by Karen Kemp. Thousand Oaks, CA: Sage: 310–313.

———. 2008b. "Network Data Structures." In *The Encyclopedia of Geographic Information Science*, edited by Karen Kemp. Thousand Oaks, CA: Sage: 314–317.

———. 2007. "Network Analysis in Geographic Information Science: Review, Assessment, and Projections," *Cartography and Geographic Information Science* 34 (4): 103–111.

Dallas Area Rapid Transit (DART). 2002. *Service Planning Five Year Action Plan*. Dallas: Dallas Area Rapid Transit.

Handy, Susan L., and Debbie A. Niemeier. 1997. "Measuring Accessibility: An Exploration of Issues and Alternatives," *Environment and Planning A* 29 (7): 1175–1194.

Horner, Mark W., and Alan T. Murray. 2004. "Spatial Representation and Scale Impacts in Transit Service Assessment," *Environment and Planning B: Planning and Design* 31 (5): 785–797.

Hsiao, Shirley, Jian Lu, James Sterling, and Matthew Weatherford. 1997. "Use of Geographic Information System for Analysis of Transit Pedestrian Access," *Transportation Research Record* 1604: 50–59.

Lee, B. Ho Yin. 2005. "Parcel-Level Measure of Public Transit Accessibility to Destinations." In *Access to Destinations*, edited by David Levinson and Kevin Krizek. Bingley, United Kingdom: Emerald Group: 215–242.

O'Neill, Wende A., R. Douglas Ramsey, and JaChing Chou. 1992. "Analysis of Transit Service Areas Using Geographic Information Systems," *Transportation Research Record* 1364: 131–138.

Peng, Zhong-Ren, Kenneth J. Dueker, James Strathman, and Janet Hopper. 1997. "A Simultaneous Route-Level Transit Patronage Model: Demand, Supply, and Inter-Route Relationship," *Transportation* 24 (2): 159–181.

Zhao, Fang, Lee-Fang Chow, Min-Tang Li, Ike Ubaka, and Albert Gan. 2003. "Forecasting Transit Walk Accessibility—Regression Model Alternative to Buffer Method," *Transportation Research Record* 1835: 34–41.