

Accessibility Indicators for Transportation Planning Using GIS

by

Ming-Sheng Lee*

Department of Civil and Environmental Engineering

University of California, Irvine

408 Verano Place

Irvine, CA 92612

E-mail: mingshel@ea.oac.uci.edu

and

Konstadinos G. Goulias

Department of Civil and Environmental Engineering

and

The Pennsylvania Transportation Institute

The Pennsylvania State University

University Park, PA 16802

Voice: (814) 863-7926

Fax: (814) 865-3039

E-mail: kxg2@psu.edu

ABSTRACT

In this paper a method to create GIS-based accessibility indicators is presented. The method allows to create person-by-person and store-by-store (disaggregate) accessibility indicators but also to derive zonal summary (aggregate) indicators that can be used in more traditional transportation planning applications. These indicators have also been used as explanatory variables in person-based transportation planning models illustrating the relationship between accessibility and shopping behavior, which in turn can be used in the trip generation models in the usual travel demand forecasting process. In this study the use of network shortest path, Gaussian function, parameter value 4.856, and employment intensity as attraction measures created an accessibility with the best behavioral foundation. The study also shows that building GIS-based accessibility indicators is feasible and provides better information than aggregate accessibility indicators. **Key words: GIS, accessibility, shopping travel**

December 1996

Paper to be Presented at the 76th Annual Transportation Research Board Meeting, January 12-16, 1997 and submitted for Publication in the Transportation Research Record.

**This work was done when the first author was a graduate student at the Pennsylvania State University*

EDITED VERSION OF THE ORIGINAL SUBMISSION

THE PROBLEM

Transportation planning methodology and practice can be greatly enhanced using Geographic Information Systems (GIS). This relatively new technology is used here to revisit the computation of accessibility measures that require more realistic representation of the transportation system than is currently used in traditional transportation planning models. Unlike past applications of accessibility, person-by-person and store-by-store accessibility indicators can be created today using the fine detail offered by GIS. These indicators, then, can be used as explanatory variables in travel behavior equations.

The concept of accessibility has been comprehensively employed in the literature as a better measure of transportation quality of service (Handy, 1993). Accessibility has generally been defined as the ease with which activity opportunities may be reached from a given location using a particular transportation system. An accessibility indicator (also called measure herein) incorporates the performance of a transportation system and the distribution of land-use activities in the study area, i.e., it includes an attractiveness (benefit) measure of each potential destination and weighs each destination by its associated travel cost. Since an accessibility index is a function of both land use patterns and the performance of the transportation system, it is a particularly appropriate criterion for evaluating the service provided by the transportation system. In addition, when person-by-person accessibility indicators can be created, access provided to specific and different categories of users can be studied by creating person-by-person measures of the quality of life (Wachs and Kumagai, 1973). Such an indicator may also reflect the overall cost of reaching work places, shopping centers, and social and recreational opportunities. These are a few potential advantages in favor of using accessibility indicators to identify and solve transportation problems.

Today, accessibility research can be found in three different transportation-related studies: in travel behavior research, in transportation /land-use system analysis, and in more general welfare and policy studies (Pirie, 1981). Although there has been a large body of research concerning the measurement and applications of accessibility indicators, the use of them has never been widespread. There are two reasons for this. First, there is no conceptually robust notion of accessibility. As Pirie (1981) stated "work is urgently needed in the area of setting standards or criteria for accessibility planning." In the existing literature, there is no agreement on the best operational definition and numerical measurement. Depending on the analysis intent and available data, researchers adopted various definitions and measurements of accessibility from their own standpoints (see the review by Lee, 1996). Second, the application of accessibility has not been tested rigorously. Its application is based on the assumption that people with different accessibility levels reveal different travel behavior. Existing research, however, failed to prove this. In addition, those analyzing the relationship between accessibility and travel behavior adopted indicators that are not appropriate for transportation planning. For example, activity density, the most frequently used indicator, does not take into account the performance of the transportation system.

GIS combined with detailed behavioral data offer a unique opportunity to revisit

accessibility with the potential to increase our understanding of this subject and provide us with a tool for better transportation planning. This paper summarizes an experimental study in which GIS-based accessibility indicators are built and used in statistical models describing shopping trip making behavior. These new person-by-person and store-by-store accessibility indicators (disaggregate measures) are also compared to those used in the past that are indicators based on summaries of fairly large geographic subdivisions (aggregate measures).

The rest of the paper is organized as follows. First, a general description of the data used is provided. These include a set of one day travel diary data, transportation network data, and business data. Then, the process of building GIS-based aggregate and disaggregate accessibility measure is presented. Finally, the effectiveness of these indicators, as explanatory variables in the travel demand models and a comparison with the aggregate accessibility indicators, is shown. The paper ends with a summary and conclusions.

DATA USED

The disaggregate accessibility indicators used in this research are computed at the most disaggregate level possible, i.e., they are person-by-person and store-by-store. To do this, exact locations of origins (e.g., a home) and all the potential destinations (e.g., all stores) of trips need to be included in the data. In this research, three groups of data are used: (a) The Centre Region Travel Patterns Survey (CRTPS), which is a one-day travel diary containing travel behavior data and the addresses of each interviewee's home and work location, (b) The American Business Lists⁷ data set, which includes all the businesses in the study area, and (c) A compilation of transportation system data that were created using the U.S. census of 1990 and statewide electronic maps.

Travel Behavior Data

The Centre Region Travel Patterns Survey (CRTPS) was carried out between April and June 1993. The original purpose was to help design a Transportation Demand Management (TDM) plan for the Centre County, Pennsylvania. Centre County is at the geographical center of Pennsylvania and contains approximately 120,000 residents. It also contains the Pennsylvania State University, which during the school year attracts approximately 40,000 students. According to development conditions, the core of Centre County called Centre Region can be described as a minor urban area, while other regions within the county are rural areas. *Travel patterns* is a term used to indicate the ways people make use of transportation facilities (i.e., descriptors of travel behavior). The Centre Region Travel Patterns Survey was indeed designed to investigate travel behavior and to estimate travel demand of the study area. The survey was divided in three parts. The first part contains various socio-demographic questions. In the second part, respondents were asked to record all information about each trip they made during a weekday using a travel diary (Personal Travel Log). The third part of the survey was

designed to examine the general attitudes regarding the transportation system. The first two parts of the survey are used in this study (for a detailed description of the instruments used, survey results, and the travel behavior variables, see Lee, 1996). The personal travel log is a record of both the geographic and chronological aspects of each trip made by a respondent. To more accurately depict this, for each trip made in the day of this survey, the respondents were asked to fill out questions regarding the origin and destination of the trip, the beginning and ending time of the trip, the beginning and ending odometer reading (if an automobile was used), as well as the purpose of the trip. The advantage of applying this data set to accessibility research is that it contains all addresses visited by a respondent. The total number of people with usable information for this analysis is 170.

Business Data

A complete business data base of the Centre County was acquired from the American Business Lists ⁷(ABL), a company that maintains the business data of the entire United States. This data set is compiled from the following sources.

Yellow Pages and Business White Pages: Since every business has a telephone, telephone directories are the major sources of this data set. According to the telephone directories, employees of ABL carry out telephone surveys to obtain business information and Yellow Page directories are used to update and add new businesses. In addition, Business White Pages are also surveyed to identify companies who choose not to be listed in the Yellow Pages. Before any information is entered in the data set, it is verified by telephone.

Annual Business Reports, Business Magazines, and Other Related Information: These sources provide in-depth information about the executives and directors, employment figures, sales volumes, and other vital statistics of publicly-traded companies. They also provide current information on mergers and acquisitions, executive changes, and financial results, on a daily and weekly basis.

Federal, State and Municipal Government Data: A wealth of information is available from governmental agencies. Most current directories and reports are used to enhance the content of the data set.

Postal Service Information: ZIP+4, Carrier Route, and NCOA: The most current postal data are used to improve addresses and to identify possible moves through the National Change of Address (NCOA) program. In addition, each NCOA change is verified by telephone to make sure the company is still in business. Some of the information in this business data set is crucial in identifying the exact locations of the businesses, such as addresses and postal ZIP codes. Other information, such as number of employees and sales volumes, has the potential to be used as a measure of a store=s attractiveness in

accessibility indicators.

The business data set includes: Company name, location and mailing address, ZIP codes (including ZIP+4), carrier route codes, type of business (Standard Industrial Code-SIC), number of employees, name and gender of the key decision maker, approximate annual sales volume, telephone numbers, fax numbers, credit rating codes, franchise/brand information, professional specialties, year established, and size of yellow page advertisement.

The behavior we focused on in this research is shopping. In order to analyze accessibility to shopping and its corresponding travel behavior, we need to select appropriate destinations for shopping trips. The business database uses the Standard Industrial Classification (SIC) codes to represent the business types. SIC was developed for use in the classification of establishments by types of employment activity. For purposes of this classification, an establishment is an economic unit, generally at a single physical location, where business is conducted or where services or industrial operations are performed (e.g., a factory, store, hotel, farm, and so on). The definition of an establishment used in the SIC is different from the commonsense definition of a company, which may be divided into several establishments. For example, a company may own a wholesale unit, a retail unit, and a management unit. In SIC, these three units are different establishments and have different corresponding codes. The structure of the classification makes it possible to categorize establishments in divisions, two-digit major groups, three-digit industry groups, and four-digit industry codes. The use of different classifying basis depends on the level of industrial detail considered most appropriate for analysis. For example Division G is RETAIL TRADE, Major Group 59 is MISCELLANEOUS RETAIL, Industry Group Number 594 is Miscellaneous Shopping Goods Stores, and Industry Number 5941, is Sporting Goods and Bicycle Shops. If further detailed classification is needed, the users may generate additional subdivisions within four-digit industry numbers, such as: 5941-15: Skiing Equipment Retail, 5941-41 : Bicycles Dealers, and so on.

The Division G (Retail Trade) includes establishments engaged in selling merchandise for personal or household consumption and rendering services incidental to the sale of the goods. Some of the important characteristics of retail trade establishments are: the establishment is usually a place of business and is engaged in activities to attract the general public to buy; the establishment is considered as retail in the trade; and the establishment sells to customers for personal or household use. According to these characteristics, it seems that all businesses listed under Division G may be the potential destinations of shopping trips. However, it is not necessary for all of the establishments in Division G to satisfy these characteristics. In this study a modification has been made for some establishment types to ensure that all destinations used for computing accessibility indicators actually attract shopping trips to their location.

ABL data are not 100% accurate due to the rapid changes in the business world (e.g., relocation and closings). The time of purchasing this data set was June 1995. At the time, the data had been updated till May 1995. A variety of methods have been used to

verify the data at hand and are illustrated in detail in Lee, 1996. The total number of businesses listed in this data set for Centre County, Pennsylvania, is 4,166. In addition, CRTPS was executed in 1993. In order to match the actual shopping opportunities in 1993, the businesses that were established in 1994 and 1995 have been deleted from the list of shopping opportunities in the models formulated in later sections.

Transportation System Data

In terms of geographic data, the boundary of the county, boundaries of different townships, and the highway network are necessary. These data were acquired from two main sources: TIGER/Line files and Pennsylvania Department of Transportation files. The geographic data files extracted from TIGER/Line files for the Centre County are: the boundary files of the county, the boundary files of all the minor civil divisions (townships, boroughs) within the county, and the street files of the county. All the files extracted were organized into two layers: the boundary layer and the street layer. The boundary layer forms the base map for this research, and the street layer is used for address-matching all the sample units in the travel pattern=s survey and the businesses in the business data.

The 1995 Pennsylvania Cartographic/Geographic Information System CD-ROM is the first endeavor for the Department of Transportation to share geographic and attribute data with other agencies. These data are a collection of digitized maps produced by the Computer Aided Mapping (CAMP) system, roadway attribute data from the Roadway Management System (RMS), and digital geographic information produced by the GIS Development Section. The geographic and attribute data in this data base are: County Centerline Files, County GIS Files, District Base Maps, Statewide Traffic Volume Map, Statewide Boundary Maps, Statewide Railroad Map, Statewide Airport Map, and Statewide Grid Map. The PennDOT files include not only geographic data files but also the attribute data files with information on access control, current Truck Annual Average Daily Traffic, current Annual Average Daily Traffic, daily vehicle miles traveled, facility type, functional class, number of traffic lanes, pavement surface type, truck percent, state route number, and total pavement width.

GIS-BASED ACCESSIBILITY MEASURES

To build aggregate or disaggregate accessibility measures that can be used in transportation planning we: (1) computed via network simulation the shortest path from each home origin of the survey participants to each potential shopping destination using an algorithm in a GIS environment, (2) all potential shopping destinations were chosen based on their SIC code, (3) whenever possible parameters were estimated for Centre County otherwise values were used from the literature, and (4) a comparison of different formulations of accessibility indicators in behavior analysis was made to choose one that provides better behavioral explanation. None of the respondents in our data base used transportation modes other than private automobiles for shopping trips. Therefore, the

accessibility measures built in this study are specifically defined as the automobile users' accessibility (also known as highway accessibility). However, the method can be easily expanded to other modes and/or multiple mode trips.

To build a simulated roadway network using GIS, we prepared a digitized map of the study area that includes: (a) The boundary files of the Centre County, (b) The boundary files of all the minor civil divisions (townships, boroughs) within the county, and (c) The roadway files of the county. The first two groups of boundary files formed the base map for this research. The base map provided coordinates to all the map features, which we would later add to the data base, such as points of respondents' home locations or centroids of the traffic analysis zones and lines of additional road segments. The roadway layer was used to address-match all the respondents in the travel patterns survey and store locations in the business data. An origin in the accessibility calculation represents the base point of measuring spatial separation to various opportunities.

Types of Accessibility Formulations

Similarly to the definition of accessibility, the numerical indicators of accessibility differ widely in the literature. The first effort to classify accessibility indicators is separating them into relative accessibility and integral accessibility. Classified by Ingram (1971), relative accessibility describes the separation between two places, while integral accessibility describes the separation between a given place and all others within a given study area served by a transportation network. Essentially, relative accessibility is the difficulty of making a particular trip. Integral accessibility measures the ease of reaching all potential destinations within a network. Both of these measures have their specific applications. Because human spatial movement is not possibly confined to one point, it is obvious that integral accessibility is more suitable for the study here. The most commonly used indicator formulations can be classified into simple measures of separation, gravity type measures, and isochronic measures.

Simple Measures of Separation

This kind of accessibility indicator simply uses the summation of spatial separation as the measure of accessibility. The spatial separation between two places can be measured by Euclidean distance (airline distance), a simplified network shortest path, travel time as reported in Census data, or generalized travel cost. This type of indicator may be expressed as:

$$A_i = \sum_{j=1}^n d_{ij}, \quad (1)$$

where A_i is the accessibility in origin i , d_{ij} is the spatial separation between i and j , and n is the number of destinations j . In addition, the separation term can be written in several

alternative formulations to incorporate the distance decay effect. The simplest form is the reciprocal function. This function may be expressed as:

$$A_i = \sum_{j=1}^n d_{ij}^{-k}, \quad (2)$$

where k is a constant to be estimated for the network in question, which means different values of k will be needed for different trip purposes.

An alternative function is the negative exponential function:

$$A_i = \sum_{j=1}^n \exp(-\beta d_{ij}), \quad (3)$$

where β is a parameter calibrated for a particular network.

Ingram (1971) has noted that both of these functions tend to decline too rapidly with respect to increasing spatial separation. He suggested a modified form of the Gaussian function:

$$A_i = \sum_{j=1}^n \exp(-d_{ij}^2 / \nu), \quad (4)$$

where ν is a constant determined for a study area. The advantage of using the Gaussian function is that it declines slowly at the region close to the origin and smoothly toward zero at a great distance. Guy (1983) proposed an alternative approach to obtain ν . It is shown that:

$$\nu = 2d_*^2, \quad (5)$$

where d_* is the distance from origin i at which accessibility is deemed to decline at the most rapid rate. The Gaussian function can be rewritten as:

$$A_i = \sum_{j=1}^n \exp[(d_{ij} / d_*)^2 / (-2)], \quad (6)$$

Gravity Type Measures

The gravity type indicators are probably the most popular. This type of indicator was originally created from an analogy with Newton's Gravitational Law. Devised by Hansen (1959), they weigh opportunities at a destination by the spatial separation from an origin to that destination. The following equation defines its general form:

$$A_i = \sum_{j=1}^n O_j (f(C_{ij})), \quad (7)$$

where O_j represents the opportunities at zone j , such as number of employees (also known as employment intensity), number of shops, etc., C_{ij} is the spatial separation

between i and j (can be measured by travel distance, travel time, or generalized cost), and $f(C_{ij})$ is the function of spatial separation, similar to the forms discussed in the simple separation subsection above.

Isochronic Measures

These indicators sometimes are called Accumulative-opportunities@ measures. They index the accessibility level of a place according to the number of opportunities that can be reached within a given travel time value Ax @. The numerical expression of the indicator can be considered as a particular case of gravity type (Koenig, 1980). When the travel time to a specific opportunity is less than x , then $f(C_{ij}) = 1$. Otherwise, $f(C_{ij}) = 0$. As Pirie (1979) has noted, the decision of a given value Ax @ is arbitrary. An example of this indicator type used by Hanson and Schwab (1987) follows:

$$A_i = \sum_{n=1}^{10} R_n / 0.5n, \quad (8)$$

where A_i is the isochronic accessibility indicator measuring the opportunities within 5 km of air line distance from an individual's home or work place. R_n is the number of establishments within the n th annulus, that is, between $0.5n$ km and $0.5((n-1))$ km from the i th individual's home. All of the parameters in this equation were decided arbitrarily according to the characteristics of the study area. Table 1 lists all indicator formulations and the symbols used in the models here. The longer the spatial separation is the worse the accessibility level becomes. A high value of separation summation indicators thus represents worse accessibility. When spatial separation is calculated with respect to the distance decay functions, however, a higher indicator value is associated with better accessibility level.

The ideal way to derive parameters in accessibility indicators is using data from a transportation study on spatial trip distribution. Unfortunately, such a study was never done in Centre County. Arbitrary assumptions are made and values from existing literature are used in this research. The only exception to this is the parameter used in the Gaussian function. This parameter, d_* , is the distance from home at which accessibility is deemed to decline at the most rapid rate. We acquired this distance from the CRTPS question that asks the respondents the distance to the nearest shopping center. This should be a reasonable assumption about the distance at which accessibility level most rapidly changes. We used the average value of the answers to that question as d_* . Travel speed 30 miles/hour was assumed to obtain the parameter value in terms of travel time. This is used when the spatial separation d_{ij} was measured by travel time. Table 2 lists the indicator groups, the assumed parameter values, and their sources.

Disaggregate Accessibility Measures

Figure 1 provides a simplified schematic example of disaggregate accessibility

calculation. For each respondent, a home-based accessibility measure uses his/her home location as the origin for calculating spatial separation to opportunities. Similarly, in the aggregate accessibility indicators calculation, instead of using the home location as an origin, the corresponding centroid of the traffic analysis zone where the respondent's home is located is used as the origin.

The intention of this study is to build disaggregate accessibility indicators and to do this we need the exact location of each respondent's residence. Since CRTPS contains the address of each respondent, the address-matching program in a GIS application can be used to perform this task. After address-matching, a point is placed on the base map to represent the location of a respondent's home. The destinations are the end points of spatial separation. The destination set may be a subset of all potential destinations when accessibility indicators are derived for each trip purpose/activity. Therefore, the selection of destination locations should be based on the trip purpose analyzed (shopping trips in this case). With a few exceptions, almost all the businesses listed under SIC code division G, retail trades, are included in the potential destination set, which for Centre County contains 490 store locations. Similarly, a point was created for each business on the digitized map using address-matching. Substantial work was expended in verifying the address ranges in the GIS database and the addresses in the business database. The major problem encountered has been the presence of rural routes in the addresses. Rural routes are temporary addresses assigned by the post office for mail delivery and they are not included in the TIGER/line files. Some new procedures have been defined within this study to correct the base TIGER/line files using tax parcel maps and to augment them with rural routes from local government information (for details see Lee, 1996).

Although we imported street data base from the TIGER/Line files for address-matching, we did not use it as the roadway network for calculating spatial separation. The TIGER/Line street data base does not contain attribute data for transportation research purposes, such as direction and functional classification. Therefore, we settled for the PennDOT road database, which includes mostly major roadways, a few township routes, and some local roads that are directly connected to the state route system. Nevertheless, additional local roads were added so that the network is closer to reality. Once we built the simulated network, we connected the points of origins and destinations to the network. By doing this, we could identify the shortest travel path from the origins, which are the home locations of each of the 170 respondents in the survey, to all potential destinations, which are the 490 shopping locations in our business database. These shortest paths, then, are utilized to represent the spatial separation between a respondent's home to the shopping places.

For each respondent, we calculated two kinds of shortest paths from a given residence to all shopping opportunities: one is based on shortest travel distance and the other is based on the shortest travel time. The travel time of each link was derived from the speed limit on each link. This choice is due to the virtual lack of congestion in Centre County when compared to other urban networks, indeed, travel times are not very different when network simulation is employed. The use of both travel time and travel

distance is done to reflect the fact that for a pair of origin and destination, the traveling path of shortest distance may be different from that of the shortest travel time. For example, a destination may be reached faster through a longer freeway than through a shorter local road. In the behavioral analysis, we will test which type has stronger behavioral interpretation.

Aggregate Accessibility Measures

The usual calculation of accessibility in transportation planning is based on Traffic Analysis Zones (in this case the census block group). Then, the zone centroid, an imaginary center of a zone, summarizing the network within the zone where a respondent resides, is used as the origin and the centroids of zones where shopping opportunities are located are used as the destinations. Figure 4 provides a schematic representation of the aggregate accessibility calculation. Two types of aggregate spatial separation measures were considered: one is calculated using the shortest travel distance and the other using the air line distance. Since almost all of the block groups within Centre County contain shopping opportunities, we use all the block group centroids as potential destinations.

BEHAVIORAL ANALYSIS

In this paper statistical analysis using accessibility measures focuses on home-based shopping trip frequency expressed as the probability of a given person to make a shopping trip in a day departing from home (the persons considered in this study made one or none one way shopping trip departing from home in a day). The model form used is LOGIT (Greene, 1993), in which the propensity to make a trip is explained statistically by a variety of socio-demographic variables (in the spirit of trip generation models used in transportation planning). Accessibility is used as one additional explanatory variable in the statistical model. More detailed analysis including travel distance and other trip purposes can be found in Lee, 1996.

We expect that high levels of accessibility are associated with a high propensity of trip making. This expectation is confirmed here in Table 3, where each row corresponds to one model including a different accessibility indicator. The making of a shopping trip in a weekday from home was found significantly related to home based accessibility as some t-tests indicate. The better the accessibility level a respondent faces, the higher the probability of making a shopping trip. Among the indicators, those with the Gaussian spatial separation equation are the most significant to home-based shopping (Table 3), especially when employment intensity is the attractiveness measure. This may be due to the fact that we estimated the parameter of Gaussian function from the CRTPS survey, thus providing for better behavioral credibility and statistical performance. In addition, the measures using distance of shortest path are more significant than those with shortest travel time. Since we calculate the travel time based on assumed speed limits

given by the functional class of each link, the shortest travel time estimates may not be close to actual field values for specific links. Inclusion of employment intensity (a proxy for goods variety) as an attractiveness measure leads to higher significance than that of the number of stores, which is also better than not including attraction variables at all.

It should also be noted that in Table 3, whenever the accessibility measures are proportional to the summations of the spatial separation, i.e., high values of these indicators represent low accessibility levels, their regression coefficients are negative as expected. However, the negative power function fails to capture the distance decay effect of our study area. Presumably, the negative power function together with the arbitrary parameter, -2, decrease the accessibility level too rapidly when spatial separation increases. Some “outliers” thus have extremely low values of accessibility. Their effects in the models cause the negative coefficients and low t-ratios of the accessibility indicators. Although the indicators with negative exponential function have positive coefficients, as expected, they did not capture the distance decay effect well enough to have significant t-ratios. Overall, Gaussian function can better capture how accessibility changes with respect to the change of travel impedance in the study area.

This relationship can be further confirmed when comparing the spatial theme maps of accessibility levels and trip frequency. Figure 3a shows where the respondents residences are and if they made a shopping trip. In Figure 3b, we present the geographic distribution of GE_G_DIS. It can be clearly seen that, with only a few exceptions, almost all respondents who made home-based shopping trips dwelled in the area with the highest accessibility level.

Aggregate (zone summaries) accessibility measures can be easily derived in traditional four-step travel models. In this section, we compare the analysis of the disaggregate measures designed in this study and aggregate measures used in past research. The result will provide guidance for practical applications. Two types of aggregate measures were considered: the measures calculated according to shortest travel distance and to air line distance. To achieve a high degree of aggregation commonly found in regional studies, the zoning system used here is based census block groups. The accessibility measures were calculated from the zone centroid where a respondent resided to all the other zones within Centre County. We use Gaussian measures with shortest travel distance as spatial separation, because these measures were proven more significantly related to trip making in the disaggregate accessibility analysis. Table 4 summarizes the results. Aggregate measures were found to significantly effect the probability of making a home-based shopping trip. Their significance is even higher than their counterparts, disaggregate measures. Within the aggregate category, measures based on shortest distance were better than those on airline distance. This result was not a surprise, since the number of home-based shopping trip frequency is either 0 or 1, the less variation an explanatory variable has, the higher its corresponding significance level in explaining behavior in the CRTPS data set. The aggregate approach assigned all respondents in a zone with the same value eliminating one source of variation. That is the reason why they are more significant than disaggregate measures. The disadvantage of

using an aggregate measure, however, can be seen clearly from Figures 4a and 4b. The aggregate accessibility is the same for all respondents residing within a zone, indifferently of their residence locations with respect to a more direct major road. Figure 4b shows that the disaggregate measure accounts for this difference. Furthermore, the finding that the measures, using a shortest path as spatial separation, are more significant than those with airline distance in aggregate analysis gives us confidence in using the former in all analyses (aggregate and disaggregate).

CONCLUSIONS

In this paper a method to create GIS-based aggregate and disaggregate accessibility indicators has been illustrated. The method allowed to create person-by-person and store-by-store accessibility indicators, called the disaggregate indicators, but also to derive aggregate indicators that can be used in more traditional transportation planning applications. These indicators have also been used as explanatory variables in disaggregate transportation planning models illustrating the relationship between accessibility and shopping behavior. These models can be used as inputs to the trip generation step in the usual travel demand forecasting process. In this study the use of the network shortest path, Gaussian function, parameter value 4.856, and employment intensity as opportunity measures created an accessibility with the best behavioral foundation. Most important, however, this study shows that building GIS-based accessibility indicators is feasible and provides better information than aggregate accessibility indicators.

The approach illustrated here, however, has not been problem-free. The major difficulties in measuring accessibility using GIS arose from problems associated with address-matching and insufficient road network data. There were two types of address-matching problems encountered: the mismatch of addresses and unrecognizable rural addresses. The former was solved by correcting the address data in the TIGER/Line files according to the tax parcel maps. The latter was solved through the help of the rural municipality governments and through mailing and telephone surveys (see Lee, 1996). Insufficient roadway data were augmented by adding segments, which are important for the calculation of accessibility. The study presented here is limited in many ways: (a) we worked with exclusively highway data, (b) rural routes are approximated by using major highways alone, (c) the calculation of shortest travel time paths is based on functional class-based speed limits, and (d) key parameter values were taken from the literature instead of been locally estimated/calibrated. These are four improvements that future research can follow as a natural extension of the study here.

The comparison of disaggregate to aggregate accessibility indicators here is used to simply show that the former can reveal finer patterns of variation in accessibility levels. Although statistically the two approaches may be equivalent, GIS-based theme maps gave us another chance to pinpoint discrepancies in the aggregate indicators. When study limitations dictate it, aggregate indicators can still be used in aggregate travel

demand models with the knowledge that they suffer from ecological fallacy. However, by using GIS, we can increase the efficiency of aggregate modeling and follow the trend of travel demand modeling in increasing disaggregation and detail. With the progress of GIS technology some travel diary databases now include for each trip made by each participant longitudes and latitudes of origins and destinations. Researchers and practitioners can take advantage of this to reduce labor and cost required in the approach presented in this paper. Given these facts, the disaggregate indicators should be considered in a favorable way over their aggregate counterparts.

In future applications, it is worth computing such indicators using multi-day travel diary data from a large metropolitan area. There are four reasons for this. First, the GIS problems faced here may be less frequent in the data set from a metropolitan area. Second, a metropolitan multi-day travel diary contains more respondents and richer behavioral information leading to potentially better travel demand models. Third, as shown in our models, accessibility measures are significant contributors in explaining travel behavior. Fourth, in behavioral equations “perceived” user accessibility, as shown in one of our formulations, may be more appropriate than “engineering” accessibility indicators. For example, a survey of perceived opportunity locations and travel time to each location may lead to more realistic behavioral equations.

The implications of the findings here for practice may lead to significant improvements in regional modeling and forecasting. One of the major issues in current practice is the estimation of the effects of added transportation capacity (Shunk, 1994). Accessibility measures capture these effects and then through models such as the home-based shopping trip model presented in this paper we can estimate the effects of added capacity. This is illustrated in a limited way here. In addition, given the evidence accumulated in this study on the power of GIS-based modeling we feel that regional councils and planning agencies should convert their transportation planning models to GIS-based transportation planning models and then choose to perform any aggregation scheme that operational constraints may dictate.

ACKNOWLEDGMENTS

Partial funding for this study was provided by the Mid-Atlantic University Transportation Center at Penn State University and portion of the data were provided by the Pennsylvania Department of Transportation. Four anonymous reviewers provided valuable comments and they are greatly acknowledged.

REFERENCES

- California Transportation Directions Policy Committee. *California Transportation Directions : Mobility for 2010*, Sacramento, CA: Caltrans, 1991.
- Dalvi M.Q. and Martin K.M. The Measurement of Accessibility: Some Preliminary Results. *Transportation*, 5, 1976, pp. 17-42.
- Greene W H. *LIMDEP User=s Manual and Reference Guide Version 6.0*, 1993.
- Guy C. M. The Assessment of Access to Local Shopping Opportunities: A Comparison of Accessibility Measures. *Environment and Planning B*, Vol. 10, 1983, pp. 219-238.
- Handy, S. L. *Regional Versus Local Accessibility: Variations in Suburban Form and the Effects on Non-Work Travel*. Doctoral Dissertation, 1993, University of California, Berkeley.
- Hansen W.G. How Accessibility Shapes Land Use. *Journal of the American Institute of Planners* , 25, 1959, pp. 73-76.
- Hanson S. and M. Schwab. Accessibility and Intraurban Travel. *Environment & Planning A*, Vol.19, 1987, pp. 735-748.
- Ingram D.R. The Concept of Accessibility: A Search for an Operational Form. *Regional Studies*, 5, 1971, pp. 101-107.
- Koenig J. G. Indicators of Urban Accessibility: Theory and Application. *Transportation*, 9, 1980, pp. 145-172.
- Leake G.R. and A. S. Huzayyin. Accessibility Measures and Their Suitability for Use in the Trip Generation Models. *Traffic Engineering and Control*, 20, 1979, pp. 566-572.
- Leake G.R. and A. S. Huzayyin. Importance of Accessibility Measures in Trip Production Models. *Transportation Planning and Technology*, Vol 6, 1980, pp. 9-20.
- Lee M. Analysis of Accessibility and Travel Behavior Using GIS. Master of Science Thesis. Department of Civil and Environmental Engineering, The Pennsylvania State University, May 1996. University Park, PA.
- Lists of 10 Million Business, American Business List⁷, Omaha, Nebraska, 1995
- Ortuzar J. and L. Willumsen, *Modeling Transport*, John Wiley & Sons, New York, NY
Pennsylvania Cartographic/ GIS CD-ROM, Pennsylvania Department of Transportation,

Harrisburg, PA, 1995.

Pirie G.H. Measuring Accessibility: A Review and Proposal. *Environment and Planning A*, Vol. 11, 1979, pp. 299-312.

Pirie G.H. The Possibility and Potential of Public Policy on Accessibility. *Transportation Research*, Vol. 15A, No.5, 1981, pp 377-381.

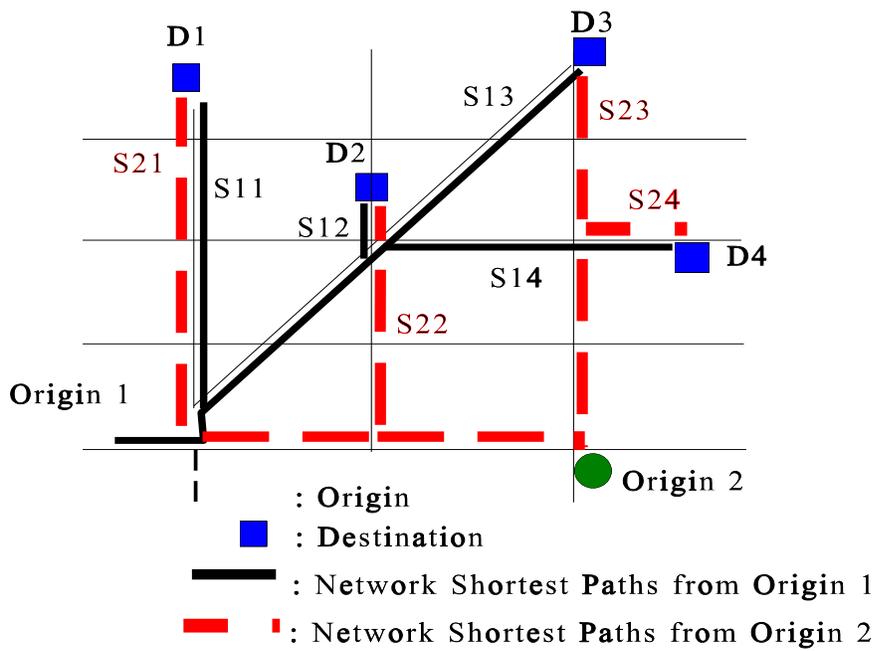
Shunk G.A. (1994) The Effects of Added Transportation Capacity - Conference Proceedings December 16-17, 1991. TMIP. DOT-T-94-12. U.S. DOT, Washington, D.C.

Standard Industrial Classification Manual, Executive Office of the President, Office of Management and Budget, 1987

TIGER/LineJ 1992 User's Guide, U. S. Bureau of the Census, Washington D.C., 1992

TransCAD Reference Manual, Version 2.0, Caliper Corporation, Newton MA, 1990.

Wachs M. and T. G. Kumagai. Physical Accessibility as a Social Indicator. *Social -economic planning science*, 7, 1973, pp. 437-456



$$A_i = \sum_{j=1}^4 O_j \cong f(S_{ij})$$

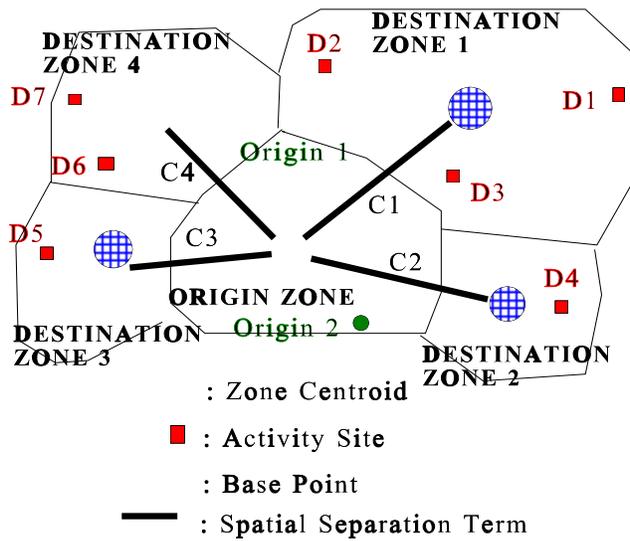
A_i = Disaggregate accessibility of origin i

O_j = Opportunities at destination j (can be employment intensity, number of stores, or constant 1 when we do not include opportunity measures in the calculation)

S_{ij} = Network shortest distance or travel time from origin i to destination j

$f(S_{ij})$ = Distance decay function (can be negative power, negative exponential, or Gaussian)

Figure 1 Measurement of Disaggregate Accessibility



$$A_i = \sum_{j=1}^3 (O_1 + O_2 + O_3) \cong f(C_1) + \sum_{j=4}^4 O_4 \cong f(C_2) + \sum_{j=5}^5 O_5 \cong f(C_3) + \sum_{j=6}^7 (O_6 + O_7) \cong f(C_4) = A_2$$

- A_i = Aggregate accessibility of origin i
- O_j = Opportunities at destination j
- C_k = Spatial separation from origin zone to a destination zone k (can be network shortest distance or Euclidean distance)
- $f(C_k)$ = Distance decay function

Figure 2 Measurement of the Aggregate Accessibility

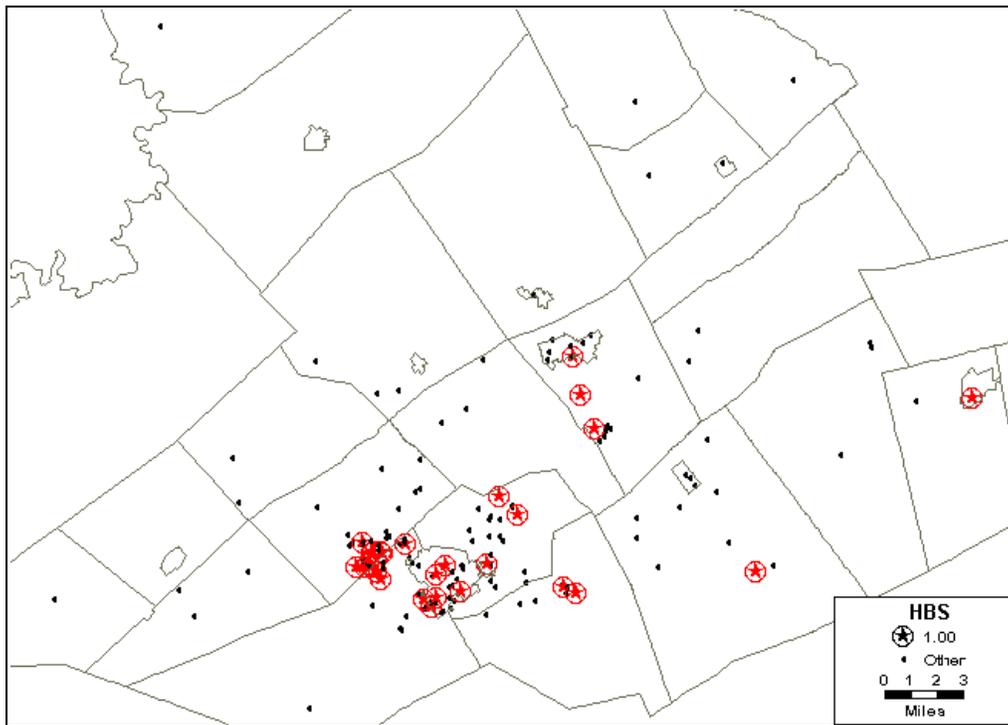


Figure 3a Residential Location of the Sample

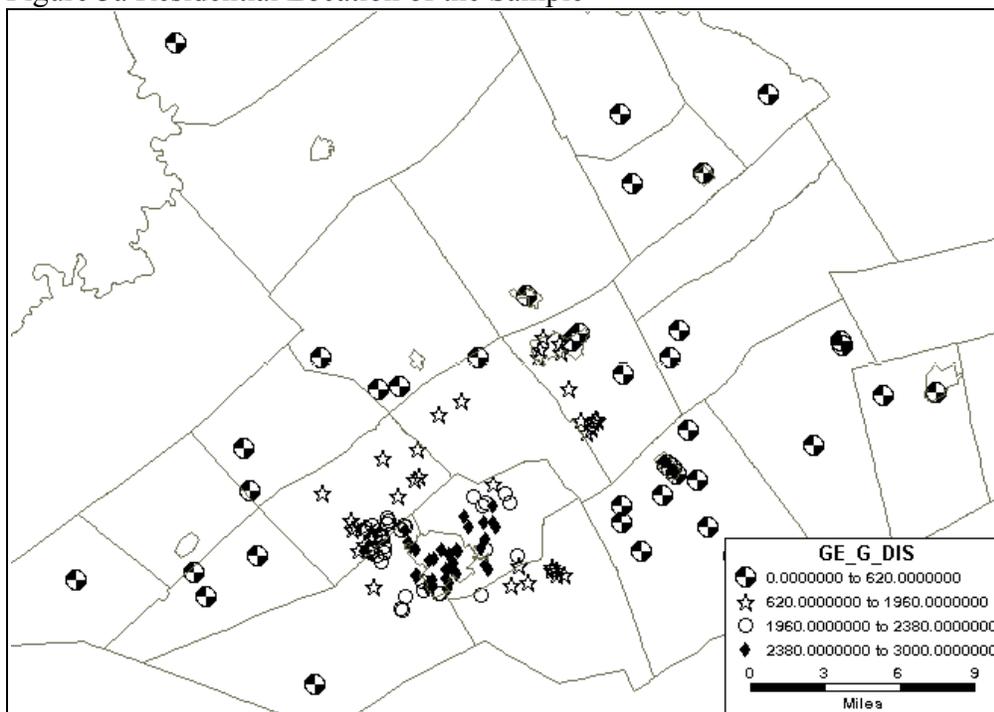


Figure 3b Mapping of GE-G-DIS

TABLE 1
Groups of Accessibility Indicator Formulations

Group	Types of Separation	Types of Opportunity Measure	Simplified Indicator Name
Separation	Shortest Distance^a	None^c	S_DIS
	Shortest Time ^b	None	S_TIME
Reciprocal	Shortest Distance	None	POW_DIS
		Employment Intensity ^d	GE_P_DIS
		Number of stores ^e	GS_P_DIS
	Shortest Time	None	POW_TIME
		Employment Intensity	GE_P_TIM
		Number of Stores	GS_P_TIM
Negative Exponential	Shortest Distance	None	NEG_DIS
		Employment Intensity	GE_N_DIS
		Number of Stores	GS_N_DIS
	Shortest Time	None	NEG_TIME
		Employment Intensity	GE_N_TIM
		Number of Stores	GS_N_TIM
Gaussian	Shortest Distance	None	GAU_DIS
		Employment Intensity	GE_G_DIS
		Number of Stores	GS_G_DIS
	Shortest Time	None	GAU_TIME
		Employment Intensity	GE_G_TIM
		Number of Stores	GS_G_TIM

^a Shortest Distance: Network shortest travel distance from an origin to a destination

^b Shortest Time: Network shortest travel time from an origin to a destination

^c None: Opportunity measures are not included ($O_j = 1$)

^d Employment Intensity: Number of employees at the shopping destination

^e Number of Stores: Number of stores at the shopping destination

TABLE 2
Assumed Parameter Values

Indicator Group	Parameter	Value	References
Reciprocal	k	2.0	Hansen (1959)
Negative Exponential	β	0.2	Dalvi and Martin (1976)
Gaussian	d^*	4.856 (Shortest Dis.)	CRTPS
	d^*	0.161 (Shortest Time)	CRTPS

TABLE 3
Probability to Make a Home-based Shopping Trip

Name	Definition ^a	Coeff.	t-ratio	Prob t \geq x	Goodness-of-fit Measures ^b
S_DIS	^c $\sum d_{ij}$	-1.74	-1.79	0.07	12.30
POW_DIS	$\sum d_{ij}^{-2}$	-0.03	-0.12	0.90	8.03
NEG_DIS	$\sum \exp(-0.2d_{ij})$	0.41	1.42	0.16	13.05
GAU_DIS	$\sum \exp[(d_{ij}/4.856)^2/(-2)]$	1.04	2.14	0.03	13.08
GE_P_DIS	^d $\sum \text{Emp}(d_{ij}^{-2})$	-0.03	-0.15	0.88	8.03
GE_N_DIS	$\sum \text{Emp}(\exp(-0.2d_{ij}))$	0.63	1.82	0.07	14.33
GE_G_DIS	$\sum \text{Emp}(\exp[(d_{ij}/4.856)^2/(-2)])$	0.98	2.18	0.03	13.26
GS_P_DIS	^e $\sum \text{Num}(d_{ij}^{-2})$	-0.03	-0.12	0.90	8.03
GS_N_DIS	$\sum \text{Num}(\exp(-0.2d_{ij}))$	0.46	1.48	0.14	13.23
GS_G_DIS	$\sum \text{Num}(\exp[(d_{ij}/4.856)^2/(-2)])$	1.02	2.13	0.03	12.98
S_TIME	^f $\sum t_{ij}$	-1.49	-1.60	0.11	11.22
POW_TIME	$\sum t_{ij}^{-2}$	-0.03	-0.12	0.91	8.03
NEG_TIME	$\sum \exp(-0.2t_{ij})$	26.81	1.61	0.11	11.26
GAU_TIME	$\sum \exp[(t_{ij}/0.161)^2/(-2)]$	1.29	1.86	0.06	11.95
GE_P_TIM	$\sum \text{Emp}(t_{ij}^{-2})$	-0.03	-0.14	0.89	8.03
GE_N_TIM	$\sum \text{Emp}(\exp(-0.2t_{ij}))$	26.77	1.71	0.09	11.70
GE_G_TIM	$\sum \text{Emp}(\exp[(t_{ij}/0.161)^2/(-2)])$	1.27	1.95	0.05	12.40
GS_P_TIM	$\sum \text{Num}(t_{ij}^{-2})$	-0.03	-0.12	0.90	8.03
GS_N_TIM	$\sum \text{Num}(\exp(-0.2t_{ij}))$	27.72	1.61	0.11	11.25
GS_G_TIM	$\sum \text{Num}(\exp[(t_{ij}/0.161)^2/(-2)])$	1.26	1.83	0.07	11.82

^a Note that origins are home locations and destinations are all shopping destinations

^b **Chi-Square {-2 [LL(β)-LL(c)]} with 7 d.f.**

^c d_{ij} : Network shortest travel distance from origin i to destination j

^d Emp: Employment intensity

^e Num: Number of stores at a shopping destination

^f t_{ij} : Network shortest travel time from origin i to destination j

TABLE 4
Home-based Shopping Models with Aggregate and Disaggregate Accessibility
Measures

Name	Definition ^a	Coeff.	t-ratio	Prob t \geq x	Goodness-of-fit Measure ^b
GDS_AG	(home BG ^c centroid, all BG centroids, Gaussian ^d , shortest distance ^e , none)	1.14	2.37	0.02	14.37
GES_AG	(home BG centroid, all BG centroids, Gaussian, shortest distance, employment intensity)	0.85	2.21	0.03	13.21
GSS_AG	(home BG centroid, all BG centroids, Gaussian, shortest distance, number of stores)	0.87	2.12	0.03	12.82
GDA_AG	(home BG centroid, all BG centroids, Gaussian, air-line distance, none)	1.52	2.05	0.04	12.89
GEA_AG	(home BG centroid, all BG centroids, Gaussian, air-line distance, employment intensity)	1.04	1.93	0.05	12.06
GSA_AG	(home BG centroid, all BG centroids, Gaussian, air-line distance, number of stores)	1.06	1.85	0.07	11.70
GAU_DIS	(home location, all shopping destinations, Gaussian, shortest distance, none)	1.04	2.14	0.03	13.08
GE_G_DIS	(home location, all shopping destinations, Gaussian, shortest distance, employment intensity)	0.98	2.18	0.03	13.26
GS_G_DIS	(home location, all shopping destinations, Gaussian, shortest distance, number of stores)	1.02	2.13	0.03	12.98

^a Definition: (origin, destinations, distance decay function, form of separation, opportunity measure)

^b Chi-Square $\{-2[LL(\beta)-LL(c)]\}$ with 7 d.f.

^c BG: Census Block Group

^d Parameter $d^* = 4.856$

^e Shortest distance: Network shortest travel distance