The ability to forecast pedestrian trips is critical, but to date, methods to do so have been time consuming and subjective. This research identifies characteristics that are indicators of pedestrian activity, and develops a method to use the characteristics in conjunction with local GIS data to determine the likelihood of pedestrian activity in Traffic Analysis Zones (TAZs) in Wake County, North Carolina. Such measures of the likelihood of pedestrian activity, often called Pedestrian Environmental Factors (PEFs), have been proven to be statistically significant in explaining travel behavior in Portland, Oregon and Montgomery County, Maryland.

The method developed in this project can analyze zonal characteristics and to develop a PEF describing the pedestrian friendliness of zones in a regional model. This method is better than pooling the knowledge of local officials, in that it is less subjective, more thorough, can describe varying levels of pedestrian activity, and it can forecast effects on the pedestrian environment that changes in the zonal characteristics may have.

Recognizing that specialized knowledge of the area is important this method includes a
review of the final PEFs by local officials, but after the characteristics and thresholds
have been adjusted to match areas of known pedestrian activity, those characteristics and
thresholds can then be used to forecast the effects of changes in the characteristics
without the need for professional review.
BIOGRAPHY

William Edwin Letchworth was born on February 15, 1980 in Wilson, North Carolina to Al and Frankie Letchworth. He attended kindergarten through 12th grade at the local public schools, and graduated from E.T. Beddingfield Jr. High School in May of 1998.

After completing high school, Will attended North Carolina State University where he discovered an interest in civil engineering, specifically in transportation. During the summers of his enrollment, Will worked with the North Carolina Department of Transportation and as an undergraduate research assistant, where he helped develop a project level traffic forecasting manual. He received a Bachelor of Science degree in Civil Engineering in May of 2002.

Directly after graduation with his undergraduate degree, Will continued his studies at N.C. State for a Masters of Science degree in Civil Engineering, with an emphasis in Transportation. His primary research involved the trip generation characteristics of neo-traditional developments. He plans to remain in the Raleigh area and work for a transportation consulting firm.

Will is registered as an engineering intern in North Carolina and is engaged to April Aycock.
ACKNOWLEDGMENTS

First and foremost, I would like to thank God for all that He has blessed me with. Next, I wish to thank my parents for their unconditional love and unending encouragement to excel. I would also like to thank my sister Dana, for being a wonderful friend and exact opposite.

Very special thanks are also due to my fiancé April. Without her love, support, understanding, and even editing, this thesis would not have been possible.

Many thanks are also due to the transportation professors at North Carolina State University. Dr. John Stone, my advisor, has been a great source of wisdom and knowledge. I have thoroughly enjoyed working with him for over 3 years and much of what I am as a transportation engineer can be directly credited to his teachings. My transportation knowledge has also been greatly expanded by Dr. Joe Hummer. Without him, my knowledge of traffic engineering would be far less, and I would certainly have had far less opportunities to don an orange vest and play in traffic. I have learned more from spending a few moments with him, than from spending days with others. Finally, many thanks are also due to Dr. Billy Williams. Dr. Williams is not only an excellent professor, he is also a great friend. While I only had the opportunity to take two classes with him, the many hours spent
standing in his doorway or over a plate or barbecue have expanded my knowledge of all things, not just transportation, and provided many hours of entertainment.

Finally I would like to thank some of the numerous individuals have proven invaluable to the completion of this thesis. The idea for this project can be directly credited to Leta Huntsinger of the Triangle Regional Model Service Bureau at the Institute for Transportation Research and Education. Her knowledge of transportation modeling continues to amaze me. Many thanks are due to Nita Bhave, also of the Triangle Regional Model Service Bureau for her extraordinary support in providing essential information and knowledge. Thanks are also due to Jin-Ki Eom, Marc Stanard, and Ed Johnson for their outstanding help.
TABLE OF CONTENTS

List of Figures viii
List of Tables x

1.0 Introduction 1
  1.1 Background and Problem Statement 1
  1.2 Scope and Objectives 5

2.0 Literature Review 7
  2.1 Chapter Introduction 7
  2.2 Methods of Analyzing Pedestrian Activity 7
  2.3 Characteristics Indicative of Pedestrian Activity 9
  2.4 GIS Data Available For Wake County 10
  2.5 Documented Relationships to Pedestrian Activity 17
  2.6 Characteristics Summary 22
  2.7 Chapter Summary 25

3.0 GIS-based Method 26
  3.1 Chapter Introduction 26
  3.2 GIS Method for Identifying Zones with High Pedestrian Activity 26
  3.3 Recalculating PEFs with Updated GIS Data 31
  3.4 Chapter Summary 31
4.0 Identifications of Zones and Calibration of Thresholds 32

4.1 Chapter Introduction 32

4.2 Zone Structure 33

4.3 Walking Trips in the Triangle Travel Behavior Survey 33

4.4 Statistical Correlation of Survey Data with Chosen Characteristics 37

4.5 Presence of Sidewalks 39

4.6 Population Density 42

4.7 Employment Density 45

4.8 Street Network Connectivity 48

4.9 Mixture of Land Use 50

4.10 Income 54

4.11 Selection of Best Characteristics 57

4.12 Chapter Summary 59

5.0 PEF Creation and Validation 61

5.1 Chapter Introduction 61

5.2 Characteristic Aggregation 61

5.3 Classification of PEFs 64

5.4 PEF Review 66

5.5 Adjustment of Pedestrian Activity Thresholds 68

5.6 Chapter Summary 73

6.0 Findings and Recommendations 74

6.1 Project Summary 74
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>Summary Findings</td>
<td>74</td>
</tr>
<tr>
<td>6.3</td>
<td>Conclusions</td>
<td>78</td>
</tr>
<tr>
<td>6.4</td>
<td>Recommendations</td>
<td>79</td>
</tr>
<tr>
<td>6.5</td>
<td>Chapter Summary</td>
<td>83</td>
</tr>
<tr>
<td>7.0</td>
<td>References</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Appendix A: Pedestrian Characteristics</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Appendix B: Wake County Land Use Codes</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Appendix C Entropy Calculation</td>
<td>94</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

1.1 Pedestrian Environmental Factor  
2.1 Wake County Street Centerline Data  
2.2 Wake County Tax Parcel Data  
2.3 Wake County Transit Data  
2.4 Sidewalk Locations in Wake County  
2.5 Traffic Analysis Zones in Wake County  
2.6 Percentage Walk Trips by Household Income  
3.1 GIS-Based Method for Identifying Zones with High Pedestrian Activity  
3.2 Classifications Aggregated into a PEF  
4.1 Triangle Travel Behavior Survey Walk Trips in Wake County  
4.2 View of Central Raleigh in Wake County  
4.3 Number of Zones by Number of Walk Trips  
4.4 Population Density vs Walking Trips in Wake County TAZs  
4.5 Sidewalk Density vs Number of Zones  
4.6 Rankings of Zones by Sidewalk Density  
4.7 Population Density vs Number of Zones  
4.8 Rankings of Zone by Population Density  
4.9 Employment Density vs Number of Zones  
4.10 Employment Density by TAZ  
4.11 Street network density by number of zones
4.12 Street Network Density  50
4.13 Entropy by TAZ  52
4.14 Land use entropy by TAZ  53
4.15 Percentage Walk Trips by Household Income  54
4.16 Distribution of Annual Income Across Wake County TAZs  55
4.17 Income Distributions Across Wake County  56
5.1 TAZ Number 74  63
5.2 Wake County PEF Distribution  64
5.3 PEF Distribution in Wake County  65
5.4 PEF Classifications in Central Raleigh  66
5.5 Classification Indicated by Ed Johnson of CAMPO  68
5.6 Sidewalk Density Distribution with New Classifications  69
5.7 PEFs with Adjusted Sidewalk Density  70
5.8 PEF with New Sidewalk Threshold Distribution  72
C.1 A Tract With an Entropy of 1  94
C.2 Parcel Buffer  97
C.3 Hypothetical TAZ Displaying Entropy Values  99
C.4 Methodology for Computing TAZ Entropy  100
LIST OF TABLES

2.1 Characteristics Related to Pedestrian Activity 9
2.2 Working Data Set 17
2.3 Characteristics Describing Intensity and Walking Quality 18
2.4 Relative Rankings for Density Dimension 23
2.5 Relative Rankings for Design Dimension 24
2.6 Relative Rankings for Diversity and Other Categories 24
4.1 Sidewalk Density Thresholds and Comparison to Survey Data 40
4.2 Walking Trips by Population Density 42
4.3 Population Density Thresholds and Comparison to Survey 44
4.4 Employment Density Thresholds and Comparison to Survey Data 46
4.5 Street Network Density and Comparison to Survey Data 49
4.6 Entropy Thresholds and Comparison to Survey Data 53
4.7 Average Yearly Household Income Threshold and Comparisons to Survey Data 55
4.8 Number of TAZs with High Surveyed Walking Trips by Characteristic 58
4.9 Number of TAZs with Medium Surveyed Walking Trips by Characteristic 58
5.1 Characteristic Thresholds 62
5.2 PEF Calculation 63
5.3 Classifications with New Sidewalk Density 69
5.4 Number of TAZs by PEF Category 71
5.5 Final Characteristics and Thresholds 73
1.0 INTRODUCTION

1.1 Background and Problem Statement

The ability to forecast pedestrian trips is critical, and was highlighted by recent guidance from the Federal Highway Administrator. “We expect every transportation agency to make accommodation for walking and bicycling a routine part of their planning, design, construction, operation, and maintenance activities (Blomberg, 2004). Additionally the 2000 Census indicated that almost two times as many people walked to work as took transit (Sicelhoff, 2004). As congestion in many parts of the country grows and the serious consequences of such growth become more apparent, including increased air pollution and decreased worker productivity, metropolitan areas are beginning to coordinate their transportation systems with their land use. To make informed transportation decisions, planners and engineers must be able to forecast the response of transportation demands based on the socioeconomic characteristics of the population, transportation system, built environment, and people’s travel choices, including walking. (Lawton, 2002). Furthermore, as the construction of new facilities becomes more difficult and expensive, developers are turning to alternative strategies to reduce fuel consumption, decrease air pollution, and improve traffic congestion. Land development strategies incorporating residences, employment, and shopping opportunities with pedestrian amenities are becoming more common in urban infill development and in suburban construction on “greenfield” sites with an emphasis on decreasing the proportion of trips taken in automobiles.
1.0 Introduction

However, current transportation planning software such as TransCAD do not explicitly contain models for pedestrian trips as they do for automobiles, trucks, buses, and commuter rail. Such models must be developed to address the growing need to predict pedestrian trips.

To consider pedestrians in a regional model, modelers must be able to identify areas with high pedestrian activity, answering the question “Where do people walk?” This is much like determining the network for automobile forecasting. In a general sense, this is done by identifying certain characteristics, often grouped into what Cervero calls the ‘3-Ds’, density, diversity, and design (Cervero, 1997). To date, this identification has been done in one of two ways: (1) by gathering a group of professionals and stakeholders and pooling their knowledge of the area, or (2) by sending a team into the field to evaluate each zone and give it a score, often called a Pedestrian Environmental Factor or PEF (Replogle 1990).

The first method was used for the 1999 update of the Triangle Regional Model (NCDOT, 1999). The Triangle Regional Model (TRM) uses TransCAD to analyze future options for transportation systems in the Triangle Region including Wake, Orange, Chatham, Durham, and parts of Franklin and Johnston counties in North Carolina. Socioeconomic and land use data, as well as the transportation infrastructure characteristics are used as inputs. To consider pedestrian activity, TRM modelers and local officials identified the areas with a high likelihood of pedestrian trips based on type of development, Traffic Analysis Zone (TAZ) location, presence of sidewalks, and proximity to other high density/mixed use TAZs. Then the modelers identified TAZs within a two-mile radius of the high access areas and
reduced vehicular trips to reflect the effects of pedestrian activity. TRM modelers indicate that the method to select pedestrian accessible zones is subjective and that forecasting the effects that changes to the environment might have on future travel patterns is difficult.

The Maryland-National Capital Park and Planning Commission (M-NCPCC) and the Portland METRO most prominently employ the second method. This method involves sending one or more teams of evaluators to every zone in the model and ranking it based on several characteristics such as street connectivity, sidewalk availability, ease of street crossings, building setbacks, and land use heterogeneity (Replogle, 1990). These rankings for each characteristic are then summed to create a Pedestrian Environmental Factor or PEF (Figure 1.1). These quantitative PEFs have been shown to be statistically significant when integrated into the trip generation and mode choice models in regional travel demand models (Cambridge, 1993).
1.0 Introduction

\[
\text{PEF} = \sum S_i
\]

Where \( S_i \) = score for pedestrian activity based on TAZ characteristics

\( 0 \leq S_i \leq 3 \) (0=none, 1=low, 2= medium 3= highest level or characteristic)

In the case of the Portland Metro, sidewalk continuity (sw), ease of street crossings (sc), local street characteristics (lsc) (ie:grid vs. cul-de-sac ie:connectivity), and topography (t) are evaluated for each zone and are given a score of zero, one, two, or three, with three being the most conducive to pedestrian activity, and one the least. Therefore the Portland Metro’s PEF takes the form

\[
\text{PEF} = S_{sw} + S_{sc} + S_{lsc} + S_t
\]

Figure 1.1: Pedestrian Environmental Factor

Replogle indicates that these models are at the forefront of considering pedestrian activity. If a PEF method were created for the Triangle Region, modelers could use quantitative PEFs measure to adjust vehicle trip generation rates and mode choice equations in the regional model.

While the Portland and Maryland PEFs provide a more systematic estimate of the pedestrian environment, problems still exist. The method used to rank zonal characteristics is
1.0 Introduction

subjective, because an accurate description of the pedestrian environment depends largely on the evaluator’s opinion and consistency. Estimating PEFs in this manner is also time consuming. For a model the size of the Triangle Regional model, which contains 2,315 zones covering over three counties, this process could take months, based on the usual TRM windshield surveys for housing condition and network characteristics. Additionally, forecasting the effects of changes to the built environment is difficult with such subjective measures. A better method is obviously needed.

Following Replogle’s suggestions and the opinions of TRM modelers that more rigorously measured quantitative factors should replace qualitative indicators, this thesis, using Wake County, NC as a guide, evaluates the available GIS data and literature to identify GIS-based indicators for walking trips, and subsequently proposes a method for assigning PEFs to traffic analysis zones using GIS methods. This GIS-based method will quickly and quantitatively identify and evaluate socioeconomic data and characteristics of TAZs that are related to pedestrian activity. The PEFs proposed for Wake County will be based on nationally recognized characteristics that other modelers could use to identify areas of high pedestrian in their regional models.

1.2 Scope and Objectives

This purpose of this thesis is to develop a quantitative GIS-based method for determining PEFs for Wake County TAZs. The specific objectives are:
1.0 Introduction

- To review the literature on zonal characteristics that are indicators of pedestrian activity
- To develop a GIS-based method to determine PEFs for Wake County TAZs using zonal characteristics
- To verify the GIS-based method and resulting Wake County PEFs using travel diaries and local professional opinions
- To recommend further research that can be taken to improve pedestrian modeling using GIS methods

The study area for the project includes the 1,345 zones in the Wake County portion of the Triangle Regional Model and the 265,315 property tax parcels in Wake County.

Only non-recreational walking trips will be studied in this project, because these walking trips substitute for trips taken in automobiles. Additionally these trips substitute for the cold start, short trips which pose a detriment to air quality.

The results of the GIS-model will be TAZ-based PEFs that will indicate the relative levels of non-recreational pedestrian activity in traffic analysis zones. Using these PEFs, Triangle Regional Modelers will be able to adjust trip generation and mode choice models to more accurately determine current and future vehicle traffic estimates.
2.0 LITERATURE REVIEW

2.1 Chapter Introduction

There are a variety of characteristics that are indicative of pedestrian activity. This chapter will highlight both socioeconomic characteristics and features of the built environment that have a proven relationship with pedestrian activity and can be identified using the GIS data available for Wake County.

2.2 Methods of Analyzing Pedestrian Activity

The U.S. Department of Transportation (Schwartz et al, 1999) highlights five methods of estimating pedestrian travel demand.

*Comparison Studies*  
Methods that predict non-motorized travel on a facility by comparing it to usage and to surrounding population and land use characteristics of other similar facilities.

*Aggregate Behavior Studies*  
Methods that relate non-motorized travel in an area to its local population, land use, and other characteristics, usually through regression analysis.

*Sketch Plan Methods*  
Methods that predict non-motorized travel on a facility or in an area based on simple calculations and rules of thumb about trip lengths, mode shares, and other aspects of travel behavior.

*Discrete Choice Models*  
Models that predict an individual's travel decisions based on characteristics of the alternatives available to them.
2.0 Literature Review

Regional Travel Models

Models that predict total trips by trip purpose, mode, and origin/destination and distribute these trips across a network of transportation facilities, based on land use characteristics such as population and employment and on characteristics of the transportation network.

As discussed in Chapter 1, the methods used by modelers in the Triangle Region generally fall into the category of sketch plan methods. However, to move into the Regional Travel Model category, a quantitative indicator of pedestrian activity needs to be created, involving the creation of a Pedestrian Environmental Factor. To date this has been done sending a team of evaluators to every zone in the model, but this project attempts to make this process more efficient and defensible by using geographic information systems (GIS).

2.3 Characteristics Indicative of Pedestrian Activity

The literature identifies more than 25 characteristics as being related to pedestrian activity (Appendix A). The characteristics (Table 2.1) are generally grouped into what are called the 3Ds of the built environment – density, diversity, and design (Cervero and Kockelman, 1997).
### Table 2.1: Characteristics Related to Pedestrian Activity

<table>
<thead>
<tr>
<th>Density</th>
<th>Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment density</td>
<td>Diversity of Land Use</td>
</tr>
<tr>
<td>Retail Store Density</td>
<td>Restaurants in Proximity to Businesses</td>
</tr>
<tr>
<td>Activity Center Density</td>
<td></td>
</tr>
<tr>
<td>Retail Intensity</td>
<td></td>
</tr>
<tr>
<td>Population Density</td>
<td></td>
</tr>
<tr>
<td>Residential Density</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td><strong>Other</strong></td>
</tr>
<tr>
<td>Availability of Sidewalks</td>
<td>Topography</td>
</tr>
<tr>
<td>Transit Shelters</td>
<td>Family Size</td>
</tr>
<tr>
<td>Building Setbacks</td>
<td>Income</td>
</tr>
<tr>
<td>Ease of Street Crossing</td>
<td>Student Populations</td>
</tr>
<tr>
<td>Street Connectivity</td>
<td></td>
</tr>
<tr>
<td>Parallel Transit Malls</td>
<td></td>
</tr>
<tr>
<td>Wide Sidewalks</td>
<td></td>
</tr>
<tr>
<td>Street Plantings and Benches</td>
<td></td>
</tr>
<tr>
<td>Light rail system</td>
<td></td>
</tr>
<tr>
<td>Low Speed Streets</td>
<td></td>
</tr>
<tr>
<td>Homes built between 1946 and 1973</td>
<td></td>
</tr>
<tr>
<td>Street Light Provisions</td>
<td></td>
</tr>
<tr>
<td>Walking Accessibility</td>
<td></td>
</tr>
<tr>
<td>Availability of Transit</td>
<td></td>
</tr>
</tbody>
</table>

While all of these characteristics have been mentioned in the literature as being related to pedestrian activity, they cannot all be identified for Wake County using GIS methods. This list must be reduced to characteristics that are both measurable using GIS methods and for which data are available to measure them in Wake County. This requires an investigation into the GIS data available for Wake County.
2.4 GIS Data Available For Wake County

GIS data are available for Wake County from three primary sources: the Wake County Property and Mapping website (http://www.wakegov.com/county/propertyandmapping/gisdigitaldata.htm), the Capitol Area Metropolitan Planning Organization (CAMPO), and from the Triangle Regional Model Service Bureau. The data include street centerline, tax parcel, transit, sidewalk, and traffic analysis zone data.

2.4.1 Street Centerline Data

Street centerline data are available for Wake County as a line layer. Referring to Table 2.1, street centerline data are related to street connectivity and low speed streets (if speed limit data is included). This data can also be used in conjunction with other GIS data to help measure the availability of sidewalks, ease of street crossings, and walking accessibility. The attributes of the data for Wake County, include the street name, length, and address identifiers. The street data file does not include speeds as data attributes, making it impossible to use speeds as a criterion for predicting pedestrian activity. While the Triangle Regional Model does contain a street data file with speed attributes, this file does not include all streets in Wake County, and the speed limits are based on roadway classification, not on actual field observations of the posted speed limit. Observation of the Wake County Property and Mapping website indicate that the street centerline data are updated regularly.
2.4.2 Tax Parcel Data

Of the attributes in Table 2.1, tax parcel data can be used to measure retail store density, activity center density, retail intensity, residential density, homes built between 1946 and 1973, diversity of land use, and restaurants in proximity to businesses. GIS-based parcel data are available for Wake County as a polygon layer. This data file contains information for 265,315 tax parcels in Wake County. In addition to parcel area and the year that any buildings on the property were built, the data contain an attribute code indicating the type of land use of the parcel. These 98 different land use codes (Appendix B) permit the identification of land use, such as neighborhood retail stores, and restaurants. The tax parcel
codes in the Wake County data will allow measurements of land use mix in Wake County TAZs. Land use mix is an important precursor to pedestrian activity. The availability of land use locations and type will be used in an innovative estimate of land use entropy, which is subsequently discussed.

![Wake County Tax Parcel Data](image)

**Figure 2.2: Wake County Tax Parcel Data**

### 2.4.3 Transit Data

Transit data can be used to indicate the presence of transit shelters, parallel transit malls, light rail systems, and the amount of transit availability. Transit availability is an important
indicator of pedestrian activity because transit riders usually walk to transit stops. Transit data for Wake County were obtained from the Triangle Regional Model Service Bureau in line and point layer format. This data include the current routes and stops for all bus routes serving Wake County, including those operated by North Carolina State University, the Triangle Transit Authority, and Capitol Area Transit. This information does not include headway or bus capacity information, but the data could be used to indicate comparative levels of transit availability in a TAZ by measuring route miles in a zone.

Figure 2.3: Wake County Transit Data
2.4.4 Sidewalk Data

Sidewalk data are critical to assessing pedestrian activity (Cervero and Kockelman, 1997). In 2000 the Capitol Area Metropolitan Planning Organization (CAMPO) completed a complete inventory of all the sidewalks in Wake County. In 2002 an update to this data was completed. The resulting GIS data contain the location of all sidewalks in Wake County and various attribute data including the adjacent street name, width, and condition of pavement.

Figure 2.4: Sidewalk Locations in Wake County
2.4.5 Traffic Analysis Zone Data

Traffic Analysis Zones (TAZs) are the smallest areas of analysis in a regional model. The TAZ data for the Triangle Regional Model were obtained from Triangle Regional Model Service Bureau officials. Various socioeconomic data for each TAZ in Wake County include: the TAZ area, population, number of jobs, number of households, number of university beds, and average income. TAZ data are important in two respects. First, the socioeconomic data in the TAZ files can be used to identify employment, population, and residential density, as well as diversity of land use, income, family size, income, and the presence of student populations. Second, all data not tied to the TAZs (street networks, sidewalk data, etc.) will be aggregated to the TAZ level.
2.0 Literature Review

2.4.6 Summary of Available Data

The list of characteristics that are related to pedestrian activity (Table 2.1) can be reduced to a working set for this research by eliminating those for which there are no Wake County GIS data (Table 2.2).
2.0 Literature Review

Table 2.2: Working Data Set

<table>
<thead>
<tr>
<th><strong>Density</strong></th>
<th><strong>Diversity</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment density</td>
<td>Diversity of Land Use</td>
</tr>
<tr>
<td>Retail Store Density</td>
<td>Restaurants in Proximity to Businesses</td>
</tr>
<tr>
<td>Activity Center Density</td>
<td></td>
</tr>
<tr>
<td>Retail Intensity</td>
<td></td>
</tr>
<tr>
<td>Population Density</td>
<td></td>
</tr>
<tr>
<td>Residential Density</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Design</strong></th>
<th><strong>Other</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability of Sidewalks</td>
<td>Topography</td>
</tr>
<tr>
<td>Transit Shelters</td>
<td>Family Size</td>
</tr>
<tr>
<td>Building Setbacks</td>
<td>Income</td>
</tr>
<tr>
<td>Ease of Street Crossing</td>
<td>Student Populations</td>
</tr>
<tr>
<td>Street Connectivity</td>
<td></td>
</tr>
<tr>
<td>Parallel Transit Malls</td>
<td></td>
</tr>
<tr>
<td>Wide Sidewalks</td>
<td></td>
</tr>
<tr>
<td>Street Plantings and Benches</td>
<td></td>
</tr>
<tr>
<td>Light rail system</td>
<td></td>
</tr>
<tr>
<td>Low Speed Streets</td>
<td></td>
</tr>
<tr>
<td>Homes built between 1946 and 1973</td>
<td></td>
</tr>
<tr>
<td>Street Light Provisions</td>
<td></td>
</tr>
<tr>
<td>Walking Accessibility</td>
<td></td>
</tr>
<tr>
<td>Availability of Transit</td>
<td></td>
</tr>
</tbody>
</table>

While the noted characteristics can be identified using the GIS data available for Wake County, it is important to choose characteristics that have documented statistical relationships with pedestrian activity.

2.5 Documented Relationships to Pedestrian Activity

Several studies have investigated how the built environment affects mode choice, trip rates, and vehicle miles traveled (VMT). Cervero and Kockelman (1997) studied the San Francisco Bay Area and found that density, land-use diversity, and pedestrian oriented
designs reduce trip rates and encourage non-auto travel in statistically significant ways. Compact development affected personal business trips the most, while within-neighborhood retail shops were most strongly associated with mode choice for work trips. Additionally, those persons living in neighborhoods with well connected street designs and restricted commercial parking averaged significantly less VMT. In light of the need to use variables to capture the many-sided dimensions of the built environment Cervero and Kockelman studied 13 major categories of characteristics and through factor analysis found that land use intensity (which captures the density dimension) and walking quality (which captures the design dimension) were the most correlated indicators of pedestrian activity. Within these two categories Table 2.3 shows the factors that have the most explanatory power when describing intensity and walking quality. A negative value indicates that as the characteristic grows larger, pedestrian activity decreases.

Table 2.3: Characteristics Describing Intensity and Walking Quality (Cervero and Kockelman, 1997)

<table>
<thead>
<tr>
<th>Factor loadings on</th>
<th>Intensity factor</th>
<th>Walking quality factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail store density</td>
<td>0.954</td>
<td></td>
</tr>
<tr>
<td>Activity center density</td>
<td>0.949</td>
<td></td>
</tr>
<tr>
<td>Retail intensity</td>
<td>0.874</td>
<td></td>
</tr>
<tr>
<td>Walking accessibility</td>
<td>0.802</td>
<td></td>
</tr>
<tr>
<td>Park intensity</td>
<td>0.806</td>
<td></td>
</tr>
<tr>
<td>Population density</td>
<td>0.796</td>
<td></td>
</tr>
<tr>
<td>Sidewalk provisions</td>
<td>0.851</td>
<td></td>
</tr>
<tr>
<td>Street light provisions</td>
<td>0.816</td>
<td></td>
</tr>
<tr>
<td>Block length</td>
<td>-0.746</td>
<td></td>
</tr>
<tr>
<td>Planted strips</td>
<td>0.728</td>
<td></td>
</tr>
<tr>
<td>Lighting distance</td>
<td>-0.724</td>
<td></td>
</tr>
<tr>
<td>Flat terrain</td>
<td>0.464</td>
<td></td>
</tr>
</tbody>
</table>
The factor loadings in Table 2.3 show that some characteristics are more closely related to walking than others and they suggest that characteristics with the highest factor loadings should be considered before those with lower factor loadings.

Cambridge Systematics used data from a 1985 home interview survey in the Portland area, results from the Portland regional travel forecasting model, and land use information to test the hypothesis that travel behavior is affected by land use patterns. They found that residents in neighborhoods with higher density, proximity to employment, grid pattern streets, sidewalk continuity, and ease of street crossings tend to make more pedestrian and transit trips. Cambridge Systematics also found that households in zones with PEF values from nine to 12 make nearly four times as many walk and bicycle trips as households located in zones with PEF values from four to eight. Additionally, Cambridge Systematics studied residential density and found that zones with four or more households per acre had over 19 percent transit and walk/bicycle trips. Finally they studied transit level of service, expressed as the quantity of employment accessible by no more than a 30-minute transit trip and found, in areas with greater than 160,000 jobs within a 30-minute trip, that 19 percent of trips were walk or bicycle trips. This study does not suggest which of the characteristics are more closely related to pedestrian activity than others, because it only examines the pedestrian environment of a zone as a whole, as described by the PEF.
Another study (TCRP 2003) reviewed various studies to investigate the impact of land use decisions on transportation outcomes. It found that in general at higher densities the use of alternative modes (particularly transit and pedestrian travel) is higher. To have the strongest correlation density must be considered as being inclusive of related phenomena, with higher densities there is a higher demand for commerce, which in turn creates a mix of development and the demand for pedestrian amenities and transit services. Citing a report by Dunphy and Fisher, walking and biking become more significant at higher densities, at 7 percent of daily trips at densities of 2,000 to 5,000 persons per square mile to 12 percent at 7,500 to 10,000, 28 percent at 10,000 to 49,999, and 46 percent at densities greater than 50,000 persons per square mile.

The 1989 Puget Sound longitudinal cohort study (Frank and Pivo, 1994) found population density to have a significant relationship with choice of the single occupant vehicle (SOV) and walking modes. Employment density appeared to have the strongest tie with mode choice overall. In terms of transit, the paper stated that several studies confirmed that the maximum distance to walk to a transit stop was one-half to three-quarters of a mile, and that high residential densities are more important than high employment densities in determining walking to access transit. This suggests that employment density may overall be more closely related to walking than population density, except around transit stops.

The TCRP report also examined studies on land use diversity, which can be measured in a variety of ways, including job/housing balance and entropy (Appendix C). It states that land
2.0 Literature Review

use diversity/mix may well be the most significant factor contributing to walking by freeing residents from total reliance on private vehicles. However, a simple job/housing balance measure may not be appropriate because such a measure does not take into account matching jobs with skill levels, and because an ideal job/housing ratio has yet to be determined. Another form of a quantitative measure of land use mix employed by several researchers is an entropy index. Accessibility and dissimilarity indices were also studied, with the accessibility index developed by Kockelman being the most influential urban form variable, and entropy being the second most influential variable.

The USDOT (1999) states that income is the most important influence on mobility and that persons in low income groups are more than twice as likely to make a walk trip than those in other income groups. Data from the 2001 National Household Transportation Survey indicates that persons living in households with less than a $10,000 yearly income make over 16 percent of their total trips by walking, but in households with over $25,000 in yearly income, that percentage drops to less than nine percent (Figure 2.7).
2.0 Literature Review

Figure 2.6: Percentage Walk Trips by Household Income

2.6 Characteristics Summary

It is evident from the literature that different studies rank different characteristics higher than others. For example in Cervero and Kockelman (1997) population density has a high factor loading, indicating that it is an important indicator of pedestrian activity. However, Frank and Pivo (1994) said that employment density had the strongest tie with mode choice overall. The TCRP (2003) report considers land use diversity/mix to be the most significant factor in contributing to walking, but the USDOT (1999) considers income to be the most important influence. Therefore, attempting to relatively rank the characteristics is difficult.
The report by Cervero and Kockelman (1997) is the most helpful when attempting to place a relative ranking on characteristics. While it only considered neighborhoods in San Francisco, the breadth of characteristics that the study covered allows for better relative rankings.

Considering the list of characteristics from the literature that are indicative of pedestrian activity (Table 2.1), the available GIS data for Wake County, and characteristics that have a measured statistical relationship with increased walking, Table 2.4 can be constructed to guide in the selection of characteristics for inclusion into a Pedestrian Environmental Factor for the TAZs in Wake County. Of the characteristics in Table 2.2 and considering the rankings in Table 2.3 ranked characteristics, Retail Store Density would be the highest, Retail Intensity second, and Population Density third. Of the remaining two characteristics, Employment Density is mentioned by Frank and Pivo as being the most important to mode choice overall, so it should be ranked higher than Residential Density in Table 2.4.

<table>
<thead>
<tr>
<th>Characteristic for PEF</th>
<th>Wake Co GIS Data Available</th>
<th>Statistical relationship</th>
<th>Ease of Measurement</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Store Density</td>
<td>Yes</td>
<td>Yes</td>
<td>Moderate</td>
<td>Cervero and Kockelman (1997)</td>
</tr>
<tr>
<td>Retail Intensity</td>
<td>Yes</td>
<td>Yes</td>
<td>Moderate</td>
<td>Cervero and Kockelman (1997)</td>
</tr>
<tr>
<td>Population Density</td>
<td>Yes</td>
<td>Yes</td>
<td>Easy</td>
<td>Cervero and Kockelman (1997)</td>
</tr>
<tr>
<td>Residential Density</td>
<td>Yes</td>
<td>Related to Population</td>
<td>Easy</td>
<td>Frank and Pivo (1994)</td>
</tr>
</tbody>
</table>

For the design dimension, Cervero and Kockelman indicate that Sidewalk Provision is the most important, while Block Length (one measure of street connectivity) would be second. Homes built between 1946 and 1978 was a characteristic not considered by Cervero and
Kockelman and that has been criticized by others as having a questionable statistical relationship with walking, so it should come last in a table of relative rankings.

**Table 2.5: Relative Rankings for Design Dimension**

<table>
<thead>
<tr>
<th>Characteristic for PEF</th>
<th>Wake Co GIS Data Available</th>
<th>Statistical relationship with walking</th>
<th>Ease of Measurement</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability of Sidewalks and Bike Paths</td>
<td>Yes</td>
<td>Yes</td>
<td>Medium</td>
<td>Cervero and Kockelman (1997)</td>
</tr>
<tr>
<td>Street Connectivity</td>
<td>Yes</td>
<td>Yes</td>
<td>Medium</td>
<td>Cervero and Kockelman (1997)</td>
</tr>
</tbody>
</table>

Land use diversity as measured by entropy is not considered by Cervero and Kockelman (Table 2.3), however, the TCRP report (2003) states that land use diversity is the most significant factor contributing to walking. Therefore it should be a highly ranked characteristic. A job/housing balance is mentioned in the TCRP report (2002) as being a poor measure of land use diversity.

In the “Other” category, the USDOT states that income is the most important influence on mobility, so that characteristic should be included in a PEF model (Table 2.6). Student populations have not been found to have a statistical relationship with walking.

**Table 2.6: Relative Rankings for Diversity and Other Categories**

<table>
<thead>
<tr>
<th>Characteristic for PEF</th>
<th>Wake Co GIS Data Available</th>
<th>Statistical relationship with walking</th>
<th>Ease of Measurement</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversity of Land Use (Entropy)</td>
<td>Yes</td>
<td>Yes</td>
<td>Hard</td>
<td>TCRP (2003), Kockelman (1996)</td>
</tr>
<tr>
<td>Restaurants in Proximity to Businesses</td>
<td>Yes</td>
<td>Yes</td>
<td>Hard</td>
<td>Pushkarev and Zupan (1975)</td>
</tr>
<tr>
<td>Job/Housing Balance</td>
<td>Yes</td>
<td>No</td>
<td>Easy</td>
<td>TCRP (2003)</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>Yes</td>
<td>Yes</td>
<td>Easy</td>
<td>USDOT (1999); NHTS Graph</td>
</tr>
<tr>
<td>Student Populations</td>
<td>Yes</td>
<td>No</td>
<td>Easy (With Surrogate)</td>
<td></td>
</tr>
</tbody>
</table>
While most of these characteristics have a statistical relationship with walking, some are much easier to measure than others. The easy characteristics to measure are Population Density, Employment Density, Residential Density, Availability of Transit, Homes built between 1946 and 1973, Income, and Student Populations. The harder characteristics to measure are Retail Store Density, Retail Intensity, Availability of Sidewalks, Street Connectivity, Diversity of Land Use, and Restaurants in Proximity to Businesses. Based on Table 2.4-2.6 and the desire to choose at least one measurement from each of the three dimensions of density diversity and design, as well as income, the following characteristics will be chosen for further evaluation and possible inclusion in the Wake County PEF: Employment Density, Population Density, Diversity of Land Use (Entropy), Street Connectivity, Sidewalk Availability, and Income will be further studied for Wake County.

2.7 Chapter Summary

Chapter 2.0 describes the various characteristics that indicate pedestrian activity. Considering the available GIS data for Wake County and the strength of the statistical relationship of the characteristics to walking, six characteristics are chosen to help indicate pedestrian activity in Wake County. Depending on the subsequent comparison using travel diaries, they will be included in a PEF for Wake County. Chapter 3.0 will discuss the method used to investigate these characteristics and to choose which characteristics will be incorporated into the PEF model for Wake County. Chapter 4.0 will subsequently set thresholds and test the relationship of these characteristics to walking activity in Wake County.
3.0 GIS-based Method

3.1 Chapter Introduction

Chapter 1.0 documented the need for a GIS-based method to identify zones with high pedestrian activity and Chapter 2 described various GIS-based data and characteristics that can be incorporated into a PEF. This chapter presents the method for developing a PEF and discusses the steps involved in using the method.

3.2 GIS Method for Identifying Zones with High Pedestrian Activity

Figure 3.1 shows the method developed in this thesis for identifying zones with high pedestrian activity. Step 1 defines the study area. Because GIS data are geographically referenced, the study area must be well defined to ensure a complete study. Typically, the study area will be the entire geographic area covered by a regional model. In this research the study area is limited to the Wake County portion of the Triangle Regional Model.

Step 2 obtains the GIS data available for the study area. All data that pertains to walking should be gathered. At the most basic level, all TAZs and their related socioeconomic data, streets, parcels, and sidewalk files should be obtained.

Step 3 chooses the characteristics than can be identified with the GIS data available for the study area. This step was performed in Chapter 2.0 of this report.
3.0 GIS-based Method

1. Define study area

2. Obtain available GIS data for study area

3. Choose characteristics that can be identified with GIS data available for the study area

4. Identify TAZs containing characteristics

5. Set ped activity thresholds for individual PEF characteristics

6. Classify TAZ (High, medium, low, n/a) using thresholds

7. Select characteristics that are best correlated with survey data, and local and national patterns

8. Aggregate selected characteristics to form Pedestrian Environmental Factor

9. Classify zonal PEFs into high, medium, and low

10. Have Pedestrian Environmental Factor reviewed by local officials

11. PEF classifications realistic?

YES

PEFs ready for inclusion into Regional Model and characteristics and thresholds ready for use in forecasting

NO

Survey data

Available literature

Statistical breakdown

Figure 3.1: GIS-Based Method for Identifying Zones with High Pedestrian Activity
3.0 GIS-based Method

Any number of characteristics can be chosen but efforts should be made to include at least one characteristic from each of the 3Ds and to choose those characteristics that have a documented relationship to walking. The list of characteristics in Chapter 2 can be used as a guide for future efforts.

After choosing the characteristics, Step 4 queries the GIS data to identify the areas that contain those characteristics. Because the TAZ is the smallest level of analysis in a regional model, data available on a finer grain, like parcel data, should be aggregated to the TAZ level.

Step 5, setting the pedestrian thresholds, is one of the most important, but most difficult steps. Very little literature exists describing specific thresholds relating to pedestrian activity because thresholds are often a matter of local and regional behavior. Even though individual characteristic classifications will later be aggregated into a zonal PEF, it is important to define high, middle, and low thresholds for each characteristic. The existing literature, local trends, and survey data should be used to define these thresholds if possible. In the absence of any quantitative data to form thresholds from the literature or local survey data, the distribution of a particular characteristic across all of the TAZs in the study area should be used. Standard classification schemes in ArcMAP GIS include: natural breaks, where the classes are based on the natural groupings of the data and breaks are placed where there are relatively big jumps in the data; quantile, where each class contains an equal number of
3.0 GIS-based Method

feature; equal interval, where each class has the same range; and standard deviation, where breaks are placed at standard deviations from the mean.

After developing the characteristic thresholds, the TAZs can then be classified according to the characteristics and levels (Step 6). These classifications should then be checked against local survey data for their relationship with walking. In the absence of detailed pedestrian survey data, they should be checked against general trends based on the knowledge of professionals in the area (Step 7).

The comparisons of TAZ characteristics with survey data should yield several characteristics that are better related to walking than others. Additionally, there will be some characteristics for which the available data are of better quality (ie; more detailed or frequently updated), more easily queried, or more justified by the literature. In Step 7, these characteristics should be chosen from the complete list and in Step 8 they should be aggregated into the total PEF for each zone. Unless specific local survey data exists, each characteristic should be weighted equally, and a pedestrian activity ranking of high should be equal to 3, medium 2, low 1, and none equal to 0 (Figure 3.2). Four or five characteristics should be sufficient to identify areas of pedestrian activity, as indicated by the Portland and Montgomery County models (Replogle, 1990). Care should be taken, however, to choose elements from each of the three ‘Ds’, density, diversity, and design.
For example, after applying these guidelines to Wake County and using the characteristics selected in Chapter 2.0 the Wake County PEF is given in figure 3.2.

\[
\text{PEF} = \sum S_i \\
\text{Where } S_i = \text{score for pedestrian activity based on TAZ characteristics}
\]

\[0 \geq S_i \leq 3\ (0=\text{none},\ 3=\text{highest level})\]

**TAZ ranked:**
- \(S_{sc}\) = Street Connectivity
- \(S_{lu}\) = Land Use Mix
- \(S_{pd}\) = Population Density
- \(S_{sw}\) = Sidewalk Availability

**Figure 3.2: Classifications Aggregated into a PEF**

In Step 9, based on the distribution of PEFs across all zones, the zones should be grouped into high, medium, and low pedestrian-friendly categories. Local officials, engineers, and modelers should then be consulted to verify the zonal PEFs. This helps ensure that no areas of high pedestrian activity are missed, and that the areas categorized as having high, medium, and low pedestrian activity do reflect these levels in reality.

If the classifications of TAZs generally fit into areas that local officials, engineers, and modelers identify as having pedestrian activity, the PEFs are ready for inclusion into the regional model. If the PEFs for areas with known pedestrian activity are low, or if the PEFs
3.0 GIS-based Method

are high in areas in which little walking takes place, the analyst should return to Step 5 and adjust the pedestrian activity thresholds for each individual characteristic.

3.3 Recalculating PEFs with Updated GIS Data

After local officials have approved the PEFs, the analyst can then use the characteristics that were aggregated into the PEF and their associated thresholds for future analysis. For example, if street connectivity is part of the PEF for an area, and updated street centerline files are obtained, steps four, eight, and nine can be performed to obtain an updated PEF for each zone in the study area without the need for professional review. However, if additional characteristics are to be included in the PEF, or the thresholds related to each characteristic are altered, all of the steps should be followed to ensure that the new PEFs match actual pedestrian travel patterns.

3.4 Chapter Summary

This chapter presented the steps for identifying zones with high pedestrian activity using GIS. A regional modeler can follow these steps to assign a pedestrian environmental factor to each TAZ in their model. Chapters 1.0 and 2.0 of this report followed Steps 1, 2, and 3 for Wake County. Chapters 4.0 and 5.0 of this report will follow Steps 4 through 10 to create a PEF for the TAZs in Wake County.
4.0 Identification of Zones and Calibration of Thresholds

4.1 Chapter Introduction

Chapter 3.0 presented the GIS-based method for identifying zones with high pedestrian activity. Steps 1 through 3 were detailed in Chapters 1.0 and 2.0 of this report. After some introductory comments about zonal structure and survey data for Wake County, this chapter performs Steps 4 through 7 of the method. Step 4 is relatively easy and identifies TAZs and their characteristics. However, the literature is sparse on the subject of thresholds that relate to more or less pedestrian activity. For example a threshold should answer the question: what is a high sidewalk density versus a medium sidewalk density? In an attempt to define the thresholds for the Triangle Region, published thresholds, natural breaks in the data, and the statistical properties of the characteristic distributions across the TAZs in Wake County will be compared to walking trips that were taken in Wake County and were reported in the Triangle Travel Behavior Survey. Step 6 attempts to adjust the thresholds to calibrate them for Wake County. Step 7 uses the results of the adjustment and calibration to identify PEF characteristics that represent density, diversity, and design and that best relate to walking trips reflected in the survey data.
4.2 Zone Structure

The quality of a PEF in describing a TAZ’s pedestrian activity depends greatly on the TAZ structure. For example, if a large TAZ had a small area of dense development where many walking trips took place. This would not be reflected on the TAZ level if the average density for the TAZ is low. Good modeling practice suggests that the density of development should be homogenous across the TAZ and that the number of trips generated by each zone should be comparatively equal. Capturing land use diversity, however, presents a problem. Good modeling practice states that TAZs should attempt to bound homogeneous activities (residential, commercial, industrial, etc.). Therefore, steps should be taken to include characteristics that are not affected by TAZ boundaries when considering land use diversity. In this project, the inclusion of the entropy characteristic and the means of calculating it across zone boundaries addresses this issue (Appendix C). Finally, TAZs should follow natural and geographic boundaries, which should enhance the identification of pedestrian friendly neighborhood designs because development patterns usually follow boundaries that could inhibit pedestrian activity. TRM officials have indicated that they followed these practices when constructing the TAZs used in the Triangle Regional Model. Hence, the zones are appropriate for PEF assessment.

4.3 Walking Trips in the Triangle Travel Behavior Survey

The Triangle Travel Behavior Survey was performed in 1995 to investigate the travel behavior of residents of the Triangle Region (Triangle Transit Authority, 1995). Part of this
4.0 Identification of Zones and Calibration of Thresholds

A survey was a travel diary that residents filled out whenever they made a trip over a 48-hour period. This information included non-recreational walking trips, and these trips were later coded so that the origin and destination TAZs could be determined.

Six hundred and forty six non-recreational walk trips were identified in the 1995 Triangle Travel Behavior Survey. Of these, 385 took place in Wake County, across 161 TAZs, representing approximately 5 percent of the total trips recorded by the survey for Wake County. In Figures 4.1 and 4.2 the zones showed in green contained between 1 and 2 walk trips (76 zones), those shown in yellow contained between 2 and 10 walk trips (73 zones), and the zones shown in red contained over 10 walk trips (12 zones) in the 48-hour reporting period. These classifications were made following the natural breaks in the number of trips by zone data (Figure 4.3) that occur between 2 and 3 and ten and eleven.

Central and North Raleigh are areas of higher concentrations of walking trips as shown in the plots of the survey data (Figures 4.1 and 4.2). Central and Southern, Eastern, and the extreme portions of North Wake County contain the least amount of pedestrian trips. Western Wake County in the Cary and Apex areas, contain fewer trips than central and Northern Raleigh, but more trips than Southern and Eastern Wake County.
Figure 4.1: Triangle Travel Behavior Survey Walk Trips in Wake County
4.0 Identification of Zones and Calibration of Thresholds

Figure 4.2: View of Central Raleigh in Wake County

Figure 4.3: Number of Zones by Number of Walk Trips
4.4 Statistical Correlation of Survey Data with Chosen Characteristics

While the characteristics chosen from the literature have been proven through other studies to have a correlation with increased pedestrian activity, unfortunately, none of the characteristics chosen from the literature exhibited more than a small correlation with the trips indicated in the Triangle Travel Behavior Survey. For example, Figure 4.4 shows a plot of TAZ sidewalk density versus the number of non-recreational walking trips in the TAZ.

![Figure 4.4: Population Density vs Walking Trips in Wake County TAZs](image)

The data points included in this graph are only for the zones that had at least one non-recreational walk trip recorded in the Triangle Regional Survey (161 zones). The linear least squares regression coefficient for the relationship between walking trips and sidewalk density
4.0 Identification of Zones and Calibration of Thresholds

was 0.0277. (This low correlation may be due primarily to the small number of surveys administered in the region, leading to the counter-intuitive conclusion that the presence of sidewalks does not relate to walking.) A total of 1500 surveys were administered in the Triangle, stratified by urban, suburban, and rural areas. Only 161 zones out of 1344 zones in Wake County contained one or more walking trips. While this sample size is sufficient for relating household characteristics to travel behavior, it is insufficient for relating TAZ characteristics to travel behavior. This small sample size led to a majority of zones having a small number of indicated walking trips, negating the indication of any effects that density, diversity, and design may have on an individual’s propensity to walk. This lack of statistical correlation with walking leads to use of the survey data as a guide to general geographic distribution of TAZs in Wake County where walking takes place, not as a definitive indicator of zones that do and do not have a high propensity for walking. By using characteristics that have previously been identified with increased pedestrian activity, at this point it is assumed that those characteristics can predict pedestrian activity in Wake County other methods must be used to determine the thresholds between characteristic classifications.

With the knowledge that thresholds cannot be developed using the survey data, the classification methods outlined in Step 5 in Chapter 3.0 will be used to classify the individual characteristics. The next six sections will follow Steps 4, 5, and 6, of the GIS-method shown in Chapter 3.0 using the plots of TAZs with survey trips in Wake County as a general guide to the geographic distribution of zones with a likelihood for pedestrian activity.
4.5 Presence of Sidewalks

Presence of sidewalks is one of the factors of the built environment that Cervero and Kockelman (1997) indicate is most closely associated with pedestrian activity. To date, most measures of sidewalk availability and connectivity were subjective or labor intensive requiring field visits or counting the number of block faces with sidewalks. Sidewalk density (linear distance of sidewalk in a zone divided by the zonal area) is easily measured by GIS and should be a comparative measure of the presence of sidewalks.

It should be noted that year 2002 sidewalk files are being compared to 1995 travel data. Unfortunately sidewalk GIS files did not exist before 2000. Since sidewalk inventory generally grows instead of declining, an over prediction of areas with high pedestrian activity can be expected.

None of the studied literature discussed the different levels, or thresholds, of sidewalk density and its relation to increased pedestrian activity. Therefore, local thresholds must be developed using the distribution of sidewalk densities and the survey data. It will be impossible to match the sidewalk density exactly to the survey data, due to the low correlation, but as discussed in Section 4.4 the general geographic trends should be upheld.
4.0 Identification of Zones and Calibration of Thresholds

Figure 4.5: Sidewalk Density vs Number of Zones

As expected, there are a large number of zones (542) with no sidewalks (Figure 4.5). These zones with no sidewalks will be grouped into the “none” category. It is very difficult to find areas where there are large differences in the distribution. Therefore a quantile classification (equal number in each group with 3 groups) was used to classify the TAZs with a sidewalk density greater than 0.1 ft/mi². Table 4.1 shows these classifications and how they relate to the zones with high numbers of surveyed walking trips.

Table 4.1: Sidewalk Density Thresholds and Comparison to Survey Data

<table>
<thead>
<tr>
<th>Sidewalk Density ft/mi²</th>
<th>Lower Threshold</th>
<th>Upper Threshold</th>
<th># of TAZs With High Surveyed Walking Trips</th>
<th># of TAZs With Medium Surveyed Walking Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (562 TAZs)</td>
<td>0</td>
<td>0.1</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Low (260 TAZs)</td>
<td>0.1</td>
<td>12000</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Medium (261 TAZs)</td>
<td>12000</td>
<td>33000</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>High (262 TAZs)</td>
<td>33000</td>
<td></td>
<td>8</td>
<td>20</td>
</tr>
</tbody>
</table>
Additionally, only 8 out of the 161 zones with survey trips did not contain sidewalks, 5 with a medium number of trips, and 3 with a low number.

Figure 4.6: Rankings of Zones by Sidewalk Density

The zones with a high sidewalk density have a similar geographic distribution to the zones with indicated walking trips (Figure 4.6). A large number of zones in central and northern Raleigh have high sidewalk density, zones in western Wake County have medium and low densities, and southern, eastern, and extreme northern Wake County have very few sidewalks.
With a lack of published sidewalk threshold data, and the degree to which the quantile thresholds match the survey data, these thresholds appear to be appropriate for the Triangle Region.

### 4.6 Population Density

Population density is also closely associated with increased pedestrian activity. It is also easily measured using the current TAZ attribute data by dividing the zonal population by the area of the zone. This is also one of the few characteristics where national information exists. Based on the 1990 National Personal Transportation Survey, Dunphy and Fisher calculated the daily person trips by mode for different population densities (Table 4.2)

**Table 4.2: Walking Trips by Population Density**

<table>
<thead>
<tr>
<th>Density Range (persons/mi^2)</th>
<th>Daily Person Trips</th>
<th>Percentage Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-99</td>
<td>0.24</td>
<td>6.4</td>
</tr>
<tr>
<td>100-249</td>
<td>0.24</td>
<td>6.2</td>
</tr>
<tr>
<td>250-499</td>
<td>0.29</td>
<td>7.3</td>
</tr>
<tr>
<td>500-749</td>
<td>0.21</td>
<td>5.4</td>
</tr>
<tr>
<td>750-999</td>
<td>0.26</td>
<td>6.7</td>
</tr>
<tr>
<td>1000-1999</td>
<td>0.23</td>
<td>6.0</td>
</tr>
<tr>
<td>2000-2999</td>
<td>0.28</td>
<td>7.1</td>
</tr>
<tr>
<td>3000-3999</td>
<td>0.29</td>
<td>7.6</td>
</tr>
<tr>
<td>4000-4999</td>
<td>0.3</td>
<td>7.6</td>
</tr>
<tr>
<td>5000-7499</td>
<td>0.36</td>
<td>9.4</td>
</tr>
<tr>
<td>7500-9999</td>
<td>0.45</td>
<td>12.4</td>
</tr>
<tr>
<td>10000-49999</td>
<td>0.95</td>
<td>27.8</td>
</tr>
<tr>
<td>&gt;50000</td>
<td>1.55</td>
<td>45.6</td>
</tr>
</tbody>
</table>
In Wake County in 1995, four zones (out of 1345) had greater than 50,000 persons per square mile, and 31 zones had a density of greater than 10,000 persons per square mile (Figure 4.7). Additionally, 165 zones have no population. A quantile breakdown using only the zones with greater than 1 person per square mile puts the areas with greater than 2415 persons per square mile in the high category. However according to the NPTS data walking begins to dramatically increase about 5000 persons per square mile.

Concurrent with the quantile threshold of 2400 persons per square mile, medium will be defined as between 2000 and 5000 persons per square mile. Anything less than 2000 persons per square mile will be considered low. Any zones with no population will be given a score of zero. Table 4.3 shows these thresholds and how they relate to the survey data.
A plot of the rankings of zones by population density shows that very few fall within the high category, and a large number of zones fall within the low category. None of the TAZs in the high category in terms of population density match with the TAZs in the high survey walking trip category. The TAZs in Northern Raleigh that fell in the high survey walking trip category.
walk trips category had low population density. However, there are many other zones in northern Raleigh that have a high population density, fitting the general geographic pattern of the survey data. In central Raleigh, the central core has a low population density, but the areas immediately surrounding the central core have a high population density, indicating that the walking trips in the survey may have been taken by individuals at work. Several zones in the central core had a large number of recorded walking trips in the survey. Like the sidewalk density, the medium and high population density fit the pattern of the recorded survey trips, with high densities in central and Northern Raleigh, and Western Wake County.

4.7 Employment Density

Employment density (the number of jobs in a TAZ divided by the TAZ area) is another characteristic that can be easily measured using the TAZ attribute data.

As expected, many of the zones have very low employment density, and a few of the zones have high employment density (Figure 4.9). The literature gives no guidance on how to divide employment density into high, medium and low ranges. A quantile breakdown divides the data at 0, 61, and 632 jobs per square mile. This does not seem reasonable, as the average density across all Wake County TAZs is approximately 2900 jobs per square mile.
4.0 Identification of Zones and Calibration of Thresholds

There are large changes in the data at 3000 and 6000 jobs per square mile. Those TAZs having less than 1 job per square mile will be considered to have no employment, between 1 and 3000 to be low, 3000-6000 to be medium, and zones with greater than 6000 jobs per square mile to have a high employment density (Table 4.4). Figure 4.10 shows that employment density is distributed very differently than the previous measures. There are many TAZs that have a low employment density and very few that have a high (87) or

![Figure 4.9: Employment Density vs Number of Zones](image)

Table 4.4: Employment Density Thresholds and Comparison to Survey Data

<table>
<thead>
<tr>
<th>Employment Density jobs/m²</th>
<th>Lower Threshold</th>
<th>Upper Threshold</th>
<th># of TAZs With High Surveyed Walking Trips</th>
<th># of TAZs With Medium Surveyed Walking Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (386 TAZs)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Low (814 TAZs)</td>
<td>1</td>
<td>3000</td>
<td>7</td>
<td>45</td>
</tr>
<tr>
<td>Medium (58 TAZs)</td>
<td>3000</td>
<td>6000</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>High (87 TAZs)</td>
<td>6000</td>
<td></td>
<td>5</td>
<td>12</td>
</tr>
</tbody>
</table>
medium (58) density. While many of the high density areas are clustered in central Raleigh, with an additional cluster in northern Raleigh, the medium density areas are somewhat scattered.

![Figure 4.10: Employment Density by TAZ](image)

**Figure 4.10: Employment Density by TAZ**

With respect to the walking trip survey data (Table 4.4), 5 out of 12 of the zones that have a high number of survey trips are also have a high employment density, but the remaining 7 zones all had a low employment density. Additionally, out of the 73 zones that had a medium number of walk trips, 52 or them had a less than 3000 jobs per square mile, placing them in the none to low categories.
4.8 Street Network Connectivity

Network connectivity is one characteristic in the design dimension that is often associated with high pedestrian activity. The more interconnected the street network, the more direct the paths that pedestrians can take to their destination. Dill (2004) offers several ways of measuring street connectivity, including street network density, connected node ratio, intersection density, and the link-node ratio. Because the nodes are not directly available in the GIS data, street network density and intersection density are closely correlated, and street network density is easily measured, the street network density measure was chosen. Street network density is measured by adding up the total street network distance in a zone, in miles, and dividing by the zonal area.

The distribution of street network density across the TAZs in Wake County is shown in Figure 4.11. A quantile breakdown suggests that below 4.75 miles per square mile be considered no street density and the other breaks be placed at 9.58 and 14.55 miles per square mile. However, this does not really fit the histogram of street network density vs number of TAZs (Figure 4.11). The number of TAZs begins to rapidly rise after 3 miles per square mile and evens out after approximately 7 miles per square mile. Then after 15 miles per square mile of street network, the number of TAZs begin to rapidly decline. Because these thresholds are similar to the quantile thresholds and are a better fit for the histograms, they will be used for the street network density. A plot of this breakdown (Figure 4.12) shows that it fits the general trend of the survey data. Nine out of the 12 zones identified as having a high number of survey walk trips have a high street network density and the remaining
three all had a medium street network density. The street network density plot fits the general trend of the survey data very well, with the high density zones in central, northern, and western Wake County.

![Street network density by number of zones](image)

**Figure 4.11: Street network density by number of zones**

**Table 4.5: Street Network Density and Comparison to Survey Data**

<table>
<thead>
<tr>
<th>Street Network Density mi/m²</th>
<th>Lower Threshold</th>
<th>Upper Threshold</th>
<th># of TAZs With High Surveyed Walking Trips</th>
<th># of TAZs With Medium Surveyed Walking Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (564 TAZs)</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Low (258 TAZs)</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Medium (263 TAZs)</td>
<td>7</td>
<td>15</td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>High (260 TAZs)</td>
<td>15</td>
<td></td>
<td>9</td>
<td>26</td>
</tr>
</tbody>
</table>
4.9 Mixture of Land Use

It is important to include each of the three ‘D’s’ (density, diversity, and design) in a pedestrian environmental factor in order paint a more complete picture of the pedestrian environment.
One measure of diversity is the entropy calculation (Kockelman, 1996). This equation measures the balance of land use in a specific area based on the following equation,

\[
\sum_{i}^{j} \left( P_i \times \ln(P_i) \right) / \ln(j)
\]

Where: 
- \( j \) = the number of land use types
- \( P \) = the proportion of land use in type \( i \)

Entropy, as a measure of land use mix, is patterned after the Second Law of Thermodynamics. A zone with perfectly balanced land use would have an entropy of 1, while a zone with homogenous land use would have an entropy of 0. The function and its computer implementation account for land use intensity as well as land use type and proximity.

It would be unrealistic to simply report the balance of land use within the TAZ, because the TAZs are of different sizes and because pedestrians can generally be thought to travel up to three quarters of a mile, which may be larger or smaller than a TAZ. To make the result more realistic, the entropy was calculated for each individual parcel, using the land use codes provided for each parcel in the Wake County parcel database, and all of the parcels that had their center within three quarters of a mile of that individual parcel. The parcels were then aggregated to the TAZ level, and an average entropy was calculated for each TAZ, weighted by the parcel size. For more details on this calculation please see Appendix C.
The distribution of land use entropy by TAZs is shown in figure 4.13.

A quantile breakdown of the data divides the entropy calculation into 4 groups (none, low, medium, and high) Less than .055 is considered having no entropy, between .055 and .133 is considered low entropy, between .033 and .251 is medium entropy, and greater than .251 is considered high entropy for Wake County (Table 4.6).
Table 4.6: Entropy Thresholds and Comparison to Survey Data

<table>
<thead>
<tr>
<th>Average Entropy</th>
<th>Lower Threshold</th>
<th>Upper Threshold</th>
<th># of TAZs With High Surveyed Walking Trips</th>
<th># of TAZs With Medium Surveyed Walking Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (337 TAZs)</td>
<td>0</td>
<td>0.055</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Low (336 TAZs)</td>
<td>0.055</td>
<td>0.133</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Medium (336 TAZs)</td>
<td>0.133</td>
<td>0.251</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>High (336 TAZs)</td>
<td>0.251</td>
<td>0.55</td>
<td>7</td>
<td>25</td>
</tr>
</tbody>
</table>

The plot of these divisions (figure 4.14) shows the various areas of high entropy around Wake County. Seven of the 12 zones with a high number of survey walk trips had a high land use entropy, and of the remaining 5, one had a medium entropy, 2 had a low entropy, and 2 had no entropy.

![Figure 4.14: Land use entropy by TAZ](image-url)
4.10 Income

Income is another characteristic where national trends exist. Information from the 2001 National Personal Transportation Survey shows that individuals with lower income make a larger percentage of walking trips (Figure 4.15).

The NPTS data naturally breaks at $20,000 annual household income and $40,000 annual household income. At incomes less than $20,000 a year, the percentage of walking trips rises dramatically, and at incomes higher than $40,000 the percentage of walking fluctuates around 7.2 percent. Table 4.7 and Figure 4.16 show how these thresholds relate to the surveyed walking trips from the Triangle Regional Travel Behavior Survey.
4.0 Identification of Zones and Calibration of Thresholds

Table 4.7: Average Yearly Household Income Threshold and Comparisons to Survey Data

<table>
<thead>
<tr>
<th>Average Yearly Household Income $</th>
<th>Lower Threshold</th>
<th>Upper Threshold</th>
<th># of TAZs With High Surveyed Walking Trips</th>
<th># of TAZs With Medium Surveyed Walking Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (200 TAZs)</td>
<td>0</td>
<td>20000</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Low (113 TAZs)</td>
<td>1</td>
<td>40000</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Medium (292 TAZs)</td>
<td>20000</td>
<td>40000</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>High (740 TAZs)</td>
<td>40000</td>
<td>40000</td>
<td>4</td>
<td>39</td>
</tr>
</tbody>
</table>

Figure 4:16: Distribution of Annual Income Across Wake County TAZs

The large number of TAZs (200) in the none category may indicate a problem with the income data. Some of the TAZs in Wake County will not have an average income, because they do not have any population. However, out of the 200 TAZs that have the income recorded as 0, 135 of them have a population. While this may indicate a problem with the
income data it also highlights a strength of the GIS based method. By using GIS data, the TAZs can easily be re-analyzed when updated information is obtained.

The income distribution throughout Wake County is also very different than that of the surveyed trips. Incomes are lowest in southern Raleigh, in rural portions of northeast, southwest, and western Raleigh, and in the central business districts of small towns surrounding Raleigh.

Figure 4.17: Income Distributions Across Wake County
4.11 Selection of Best Characteristics

After identifying the TAZs containing the characteristics from the literature (Step 4), determining their thresholds (Step 5) and classifying the TAZs based on the characteristics (Step 6), several characteristics must be chosen for inclusion into the PEF. The state of the practice in Maryland and Portland is to choose four characteristics. Referring to Cervero’s 3Ds, at least one characteristic should come from density, diversity, and design. The choices should be based on the relationship with surveyed walk trips.

For Wake County, street network density has the highest number of matches with the surveyed walking trips from the Triangle Regional Travel Behavior Survey (Table 4.8 and 4.9). Areas with high and medium street density also match the areas of high and medium surveyed walk trips with high densities in central and Northern Raleigh and low densities in Southern, Eastern, and the extreme portions of North Wake County.

Sidewalk density is second best with eight zones with a high number of survey trips having high sidewalk density. Compared with zones with a medium number of surveyed trips, sidewalk density matches the third best (tied with entropy). Even though sidewalk density is a measure of design, the high factor loading put on sidewalk availability by Cervero, the use of sidewalk availability by Portland and Montgomery County, and the frequency with which
4.0 Identification of Zones and Calibration of Thresholds

Table 4.8: Number of TAZs with High Surveyed Walking Trips by Characteristic

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sidewalk Density ft/m²</th>
<th>Population Density persons/m²</th>
<th>Employment Density jobs/m²</th>
<th>Street Network Density mi/m²</th>
<th>Average Entropy</th>
<th>Average Yearly Household Income $</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Low</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Medium</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>High</td>
<td>8</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 4.9: Number of TAZs with Medium Surveyed Walking Trips by Characteristic

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sidewalk Density ft/m²</th>
<th>Population Density persons/m²</th>
<th>Employment Density jobs/m²</th>
<th>Street Network Density mi/m²</th>
<th>Average Entropy</th>
<th>Average Yearly Household Income $</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>19</td>
<td>12</td>
<td>6</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Low</td>
<td>15</td>
<td>26</td>
<td>45</td>
<td>9</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>Medium</td>
<td>18</td>
<td>31</td>
<td>9</td>
<td>34</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>High</td>
<td>20</td>
<td>3</td>
<td>12</td>
<td>26</td>
<td>25</td>
<td>39</td>
</tr>
</tbody>
</table>

it is discussed in the literature as an indicator of pedestrian activity call for it to be included in the PEF equation for Wake County TAZs.

Entropy matched the third best with seven zones with a high number of survey trips having high entropy. While the areas with high and medium entropy across Wake County are somewhat different than the areas with high and medium surveyed walk trips, this measure matches well with the high walk trips, and it is the only measure of land use diversity available.
The fourth additional characteristic is population density. While employment density matches the best with respect to the number of zones with high surveyed walk trips, population density seems to fit the geographic distribution better, matches better with the zones with a median number of surveyed walk trips, and, according to Cervero, has more explanatory power than employment density.

Therefore the PEF for Wake County will be a function of:

- Street Network Density
- Sidewalk Density
- Land Use Entropy
- Population Density

and will be defined as a sum of the individual characteristic scores for each TAZ.

\[ PEF = S_{sd} + S_{sw} + S_e + S_{pd} \]

Where \( S_{sd}, S_{sw}, S_e, S_{pd} = (0, 1, 2, 3) \)

Thus, the higher the PEF, the more pedestrian activity the zone is likely to have.

4.12 Chapter Summary

Using the characteristics identified in Chapter 2.0 that could be identified using the Wake County GIS data, this chapter performs Steps 4 through 7 of the GIS method for determining zones with high pedestrian activity. After identifying the areas in Wake County containing
these characteristics, and comparing them with survey data from the 1995 Triangle Regional Travel Behavior Survey, street network density, sidewalk density, land use entropy, and population density were chosen for inclusion into a PEF to describe the pedestrian friendliness of zones in Wake County. Chapter 5.0 will aggregate these characteristics into the PEF, and compare these PEFs with areas of known pedestrian activity determined through discussions with local officials and modelers.
5.0 PEF Creation and Validation

5.1 Chapter Introduction

In Chapter 4.0, street network density, sidewalk density, land use entropy, and population density were chosen for inclusion into a PEF to describe the pedestrian friendliness of zones in Wake County. This chapter follows Steps 8 through 11 of the GIS method for determining zones with high pedestrian activity.

5.2 Characteristic Aggregation

In Chapter 4.0, street network density, sidewalk density, land use entropy, and population density were chosen for inclusion into a PEF to describe the pedestrian friendliness of zones in Wake County. Thresholds for the various levels (none, low, medium, and high) for each characteristic were developed using the literature and the distribution of each characteristic across the TAZs in Wake County (Table 5.1).

To aggregate the characteristics into a PEF, each classification must be given a score. An area with a high classification of a certain characteristic will be given a 3, with a medium classification a 2, with a low classification a 1, and without any level of the characteristic a 0. As with the Portland and Maryland models, and due to a lack of local data that suggests otherwise, each characteristic will be weighted equally.
Therefore:

$$PEF = S_{sd} + S_{sw} + S_e + S_{pd}$$

Where \( S_{sd}, S_{sw}, S_e, S_{pd} = (0, 1, 2, 3) \)

### Table 5.1: Characteristic Thresholds

<table>
<thead>
<tr>
<th>Sidewalk Density ft/mi(^2)</th>
<th>Population Density persons/mi(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lower Threshold</strong></td>
<td><strong>Upper Threshold</strong></td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>0.1</td>
</tr>
<tr>
<td>Medium</td>
<td>25000</td>
</tr>
<tr>
<td>High</td>
<td>50000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Street Network Density mi/mi(^2)</th>
<th>Average Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lower Threshold</strong></td>
<td><strong>Upper Threshold</strong></td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
</tr>
<tr>
<td>Medium</td>
<td>7</td>
</tr>
<tr>
<td>High</td>
<td>15</td>
</tr>
</tbody>
</table>

To illustrate the PEF scoring concept consider TAZ number 74. It has a medium sidewalk density, a medium street density, high entropy, and low population density resulting in a \((2 + 2 + 3 + 1)\) PEF of 8 (Table 5.2).
5.0 PEF Creation and Validation

Table 5.2: PEF Calculation.

<table>
<thead>
<tr>
<th>TAZ</th>
<th>Sidewalk Density</th>
<th>Street Density</th>
<th>Average Entropy</th>
<th>Population Density</th>
<th>Sidewalk Score</th>
<th>Street Score</th>
<th>Entropy Score</th>
<th>Population Score</th>
<th>PEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>11983</td>
<td>21.7</td>
<td>0.362</td>
<td>18127</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>71</td>
<td>13404</td>
<td>24.4</td>
<td>0.275</td>
<td>5253</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>72</td>
<td>30113</td>
<td>7.7</td>
<td>0.194</td>
<td>3622</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>73</td>
<td>18083</td>
<td>20.0</td>
<td>0.378</td>
<td>2467</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>74</td>
<td>32400</td>
<td>11.2</td>
<td>0.343</td>
<td>470</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>75</td>
<td>9443</td>
<td>16.5</td>
<td>0.327</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>76</td>
<td>0</td>
<td>12.1</td>
<td>0.364</td>
<td>2113</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>77</td>
<td>0</td>
<td>13.0</td>
<td>0.300</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>78</td>
<td>699</td>
<td>13.8</td>
<td>0.203</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>79</td>
<td>0</td>
<td>24.2</td>
<td>0.364</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>80</td>
<td>0</td>
<td>22.9</td>
<td>0.317</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>81</td>
<td>18403</td>
<td>16.2</td>
<td>0.156</td>
<td>2117</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>82</td>
<td>17762</td>
<td>14.4</td>
<td>0.138</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>83</td>
<td>1708</td>
<td>4.2</td>
<td>0.085</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 5.1: TAZ Number 74
5.3 Classification of PEFs

For ease of comparison the PEFs can be grouped into high, medium, and low categories (Step 9). Neither the Montgomery County nor the Portland modelers grouped their PEFs, but a high, medium, and low grouping is easier to present to local officials than 12 gradations (0-12). However, since none of the other planning agencies that use PEFs group their rankings, no guidance is available as to how to group the TAZs. Additionally, the distribution of PEFs across the TAZs in Wake County also provides few natural breaks to aid in classification (Figure 5.2).

![Wake County PEF Distribution](image)

**Figure 5.2: Wake County PEF Distribution**

Therefore, the PEFs will be divided evenly, at 4 and 8. A plot of the data (Figures 5.3 and 5.4) shows that zones in central Raleigh have a high PEF with an elongated area of high
PEFs stretching north along Wake Forest Rd. Figure 5.3 also indicates that the central business districts of the towns surrounding Raleigh have high PEFs. Western Wake County also has a high PEF in the Cary and Apex areas. Of the 12 zones with a high number of walk trips from the Triangle Regional Travel Behavior Survey, 6 had a high PEF and 6 had a medium PEF.

Figure 5.3: PEF Distribution in Wake County
5.0 PEF Creation and Validation

5.4 PEF Review

After classifying the PEFs into high, medium, and low groups, a map of these groups should be reviewed by local professionals that are familiar with the area. Leta Huntsinger, Program
Manager for the Triangle Regional Model Service Bureau, indicated that the central cores of Raleigh and the surrounding small towns should be rated high in terms of walkability. She was most familiar with the Cary area, and indicated that downtown Cary and Eastern Cary would be areas that should have a high PEF.

Ed Johnson, Director of the Capitol Area Metropolitan Planning Organization (CAMPO) was also consulted. He echoed Ms. Huntsinger’s comments that the central cores of the small towns surrounding Raleigh should be highly rated, as well as the downtown Raleigh area and the area around North Carolina State University.

Mr. Johnson also mentioned several specific areas that would have high, medium, and low amounts of pedestrian activity (Figure 5.5). His classifications indicate that some of the areas that the GIS method assigns a high PEF should in reality have a lower PEF. Investigations into the areas that he marked as Medium, show that most of the characteristics have the same classification (usually high or medium). Investigations into the areas that he marked as Low reveal that the sidewalk density classifications are usually in the high range, and one or two points higher than the street density, entropy, or population density measurements. This suggests that the sidewalk density thresholds may need adjustment.
5.5 Adjustment of Pedestrian Activity Thresholds

Following the method outlined in Chapter 3.0, the classification thresholds for sidewalk density were adjusted in an attempt to make the PEF values better match the zones that Mr. Johnson indicated had low and medium pedestrian friendliness. To lower the number of
zones in the high sidewalk density category the threshold between medium to high was raised to 50,000 ft per square mile and between low to medium to 25,000 ft per square mile (Table 5.3 and Figure 5.6).

Table 5.3: Classifications with new Sidewalk Density Thresholds

<table>
<thead>
<tr>
<th>Sidewalk Density ft/mi²</th>
<th>Lower Threshold</th>
<th>Upper Threshold</th>
<th># of TAZs With High Surveyed Walking Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (562 TAZs)</td>
<td>0</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>Low (439 TAZs)</td>
<td>0.1</td>
<td>25000</td>
<td>2</td>
</tr>
<tr>
<td>Medium (164 TAZs)</td>
<td>25000</td>
<td>50000</td>
<td>2</td>
</tr>
<tr>
<td>High (180 TAZs)</td>
<td>50000</td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 5.6: Sidewalk Density Distribution with New Classifications
This affected the areas with high sidewalk density that matched with areas of high surveyed walk trips very little, removing one zone from the high category and placing it in the middle category. This also reduced PEFs for the area that Mr. Johnson indicated as having a low likelihood for pedestrian activity from medium to low, and reduced some of the TAZs having a high PEF in the area where he stated there would be a medium amount of activity from high to medium (Figure 5.7)

Figure 5.7: PEFs With Adjusted Sidewalk Density
This change did not drastically change the number of zones in each PEF category (Table 5.4) or the overall geographic distribution of zones with a high PEF (Figure 5.8). The new thresholds for sidewalk density also did not change the number of TAZs that match with those TAZs having a high number of surveyed walk trips, with 6 having a high PEF and 6 having a medium PEF.

### Table 5.4: Number of TAZs by PEF Category

<table>
<thead>
<tr>
<th>PEF Category</th>
<th>Initial PEF</th>
<th>PEF With Adjusted Sidewalks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>513</td>
<td>537</td>
</tr>
<tr>
<td>Medium</td>
<td>588</td>
<td>602</td>
</tr>
<tr>
<td>High</td>
<td>244</td>
<td>206</td>
</tr>
</tbody>
</table>

The PEF for Wake County TAZs has been developed using nationally recognized characteristics with a documented relationship with walking. The thresholds for these characteristics have been determined based on national data, local surveyed walking trips, and the distribution of characteristics across the TAZs. After the characteristics were aggregated into a PEF for each zone, the PEFs were compared with professional assessments of the pedestrian friendliness of zones in Wake County. Based on how well the PEFs with the new sidewalk thresholds match the comments received from Ms. Huntsinger and Mr. Johnston, at this point these thresholds (Table 5.5) are assumed appropriate for the determination of areas with high pedestrian activity in Wake County.

Now that these thresholds have been developed and approved by local professionals, a modeler could use updated or forecasted data in GIS form to re-analyze the TAZs in Wake
county without the need for professional review. However if new methods for analyzing a
classification are developed, a new classification is used, or the thresholds are changed local
professionals should be consulted again to approve the new PEFs.
Table 5.5: Final Characteristics and Thresholds

<table>
<thead>
<tr>
<th>Sidewalk Density ft/mi²</th>
<th>Population Density persons/mi²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lower Threshold</strong></td>
<td><strong>Upper Threshold</strong></td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>0.1</td>
</tr>
<tr>
<td>Medium</td>
<td>25000</td>
</tr>
<tr>
<td>High</td>
<td>50000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Street Network Density</th>
<th>Average Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lower Threshold</strong></td>
<td><strong>Upper Threshold</strong></td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
</tr>
<tr>
<td>Medium</td>
<td>7</td>
</tr>
<tr>
<td>High</td>
<td>15</td>
</tr>
</tbody>
</table>

5.6 Chapter Summary

This chapter follows Steps 8 through 11 of the GIS method for determining zones with high pedestrian activity. After aggregating street network density, sidewalk density, land use entropy, and population density into a single PEF for each of Wake County’s 1345 TAZs, the PEFs were compared to assessments of the county’s pedestrian activities made by two local professionals. This comparison led to the adjustment of the thresholds for sidewalk density in order to make the PEFs more closely match with areas of known pedestrian activity.
6.0 SUMMARY FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

6.1 Project Summary

The ability to forecast pedestrian trips is critical, but to date, methods to do so have been time consuming and subjective. This project identifies characteristics that are indicators of pedestrian activity, and develops a method to use the characteristics in conjunction with local GIS data to determine the likelihood of pedestrian activity in Traffic Analysis Zones (TAZs) in Wake County, North Carolina. Such measures of the likelihood of pedestrian activity, often called Pedestrian Environmental Factors (PEFs), have been proven to be statistically significant in explaining travel behavior in Portland, Oregon and Montgomery County, Maryland.

6.2 Summary Findings

6.2.1 GIS-Based Method

The method developed in this project is a quick way to analyze zonal characteristics and to develop a PEF describing the pedestrian friendliness of zones in a regional model. Most zonal characteristics can be analyzed in less than 10 minutes using existing ArcGIS processes. The entropy calculation requires less than 10 minutes of setup, but the computer processes for Wake County took approximately 7 hours. The most time consuming parts of this method are obtaining the GIS data and choosing the
6.0 Findings and Recommendations

characteristics and thresholds from the literature that can be identified with the GIS data. Once these tasks have been accomplished (such as when updating the PEFs with new GIS data) all of the zones in a model can be re-analyzed in less than 4 hours.

This method is better than pooling the knowledge of local officials, in that it is less subjective, more thorough, it can describe varying levels of pedestrian activity, and it can forecast any effects on the pedestrian environment that changes in the zonal characteristics may have. Recognizing that specialized knowledge of the area is important, this method includes a review of the final PEFs by local officials, but after the characteristics and thresholds have been adjusted to match areas of known pedestrian activity, those characteristics and thresholds can then be used to forecast the effects of changes in the characteristics, without the need for professional review.

The GIS-based method is also superior to the method used by modelers in Portland, Oregon and Montgomery County, Maryland, in that it requires far less time than sending a group of observers into the field to rank traffic analysis zones in terms of pedestrian friendliness.

6.2.2 Need for Survey Data

Due to the lack of threshold data, and regional variability in travel behavior, it is paramount that survey data be available to compare with the characteristics indicative of
walking. As was shown for with the case study, data from a regional survey with random sampling intended to relate household characteristics to travel behavior is not sufficient to relate neighborhood characteristics to travel behavior.

6.2.3 Importance of Good GIS Data

The quality of the GIS data available for an area can greatly improve, or greatly hinder the use and accuracy of this method. County and state governments collect most GIS data. Within the counties included in the Triangle Regional model, the amount and descriptiveness of tax parcel, street, and sidewalk data varied greatly. If the method developed in this project were to be used over a multi-county area, the county that had the least descriptive GIS data would govern the use of all other GIS data for identification of pedestrian zones. Thus, this method is only as strong as its weakest link.

As seen through the literature review, some indicators have been more closely associated with increased walking than others. The presence of a sidewalk network is consistently at the forefront of discussions when involving pedestrian activity. Having a GIS database of sidewalks will greatly increase the predictive power of any GIS model. Low speed streets are also mentioned repeatedly in the literature. While roadway networks are usually coded in GIS format, speeds rarely are. The inclusion of speed information could also improve the predictive power of a GIS model. Finally, the coding of a land use type in the tax parcel data is paramount. The land use entropy calculation is a validated
measure of land use mix, but it requires knowledge of the land use type. This land use code need not be as descriptive as Wake County’s, but simply putting a code in one of the 6 categories included in the entropy equation would greatly aid in the identification of pedestrian zones.

6.2.4 Entropy

To date, entropy has been measured within an artificially defined area, such as Census blocks or TAZs and only on a small scale, generally for a few neighborhoods. The method to measure entropy developed in this thesis uses tax-parcel data in GIS form in conjunction with a computer program written in Matlab. This method can evaluate the land use mix irrespective of zonal boundaries and on a county-wide level, making it a more realistic measure and feasible to use for a regional model.

6.2.5 Findings for Wake County

The lack of survey data for Wake County presented the largest barrier to increasing PEF accuracy. Without survey data thresholds relating to the characteristics could not be determined and matching zones with high surveyed walking trips with specific characteristics and the final PEF was difficult. At best, the TAZs with a high number of surveyed walking trips could be matched with TAZs having a high level of a specific characteristic and the geographic distribution of areas with high and medium surveyed walking trips could be used as a general guide to areas of high pedestrian activity.
6.0 Findings and Recommendations

The GIS-based method indicated that central and northern Raleigh, western Wake County, and the downtown areas of small towns in Wake County had a high propensity for pedestrian activity. This follows closely with my knowledge of the area and with the comments from Ms. Huntsinger and the general comments of Mr. Johnson.

6.3 Conclusions

During this thesis, the researcher concluded the following:

- There are a variety of characteristics that are indicative of pedestrian activity and they generally fall into the ‘3Ds’ of Density, Diversity, and Design.

- Identifying these characters related to pedestrian activity is intuitively easy, yet difficult to implement quantitatively due to the lack of detailed travel diaries and published activity thresholds.

- Sidewalk and Street Connectivity, Land use Entropy, and Population and Employment density are more promising indicators of pedestrian activity than Retail Store and Residential Density, House Age, Restaurants in Proximity to Businesses, Student Populations, and Income.

- Identifying thresholds of pedestrian activity is extremely difficult because survey data is sparse and automatic statistical classifications and subjective adjustment sometimes gives unrealistic results.
• The proposed methodology can utilize a set of characteristics and the available GIS data to assign a Pedestrian Environmental Factor to traffic analysis zones in a regional model.

• The characteristics identified in this project could be used by any modeling agency, provided that the GIS data exists for that area. However, the thresholds developed in this project may not be appropriate for other areas.

• For Wake County, North Carolina, the PEFs developed by this method matched closely with the areas of high pedestrian activity identified by local professionals.

• Good GIS and survey data is critical for increasing the accuracy and validity of the PEFs developed using this method.

6.4 Recommendations

The following recommendations are for any modeling agency that wishes to include pedestrians in their regional modeling efforts.

• Use the GIS-based method outlined in this project to identify areas of high pedestrian activity using nationally recognized characteristics. Using GIS data can decrease subjectivity and greatly decrease the time required to identify these areas, as compared to sending a team of observers to rate every TAZ in a regional
6.0 Findings and Recommendations

model. The output should be in the form of a descriptor, indicating the pedestrian friendliness of a zone (PEF).

- Solicit input from local officials, modelers, and engineers to validate the determined areas of high pedestrian activity. Even if sufficient survey data is available to locally correlate the chosen characteristics with increased walking, the specialized knowledge of local individuals can be invaluable for double-checking model results.

- Structure regional surveys to be able to correlate neighborhood variables to increased walking. While this setup will still allow modelers to correlate household variables with trip generation and mode choice, the additional data will allow for determining which characteristics are locally predictive of increased walking and will allow the division of characteristic levels into high, medium and low categories.

- Examine other methods for collecting walking information, such as automatic pedestrian counters and collision reports involving pedestrians.

- Encourage regional standardization of GIS data, because travel demand models often cover more than one data-collecting jurisdiction. This will help to avoid the weakest link problem.

- Examine other methods for collecting GIS data. For example, satellite or aerial photography could easily be used to develop databases of sidewalk locations.
6.0 Findings and Recommendations

- Develop GIS databases of sidewalk locations. Sidewalks are consistently at the forefront of studies that correlate characteristics of the built environment and walking. Knowing their location is paramount.

- Include land use codes on the parcel information. This will allow for better investigations into the land use diversity in the study area.

- Include speed data on the street information. This will allow the identification of areas with low speed streets, which are mentioned throughout the literature as being favorable to pedestrian activity.

- Continue efforts to decrease the dependency of the entropy calculation on the parcel size. The current method reduces the dependency when aggregating the individual parcel entropies into a TAZ entropy by using a weighted average. However, the entropy calculation itself is still based on proportions of parcels within a certain area. Therefore large parcels carry the same weight as small parcels. This dependency needs to be removed, which could make land use entropy a more powerful measure of land use diversity.

- Continue investigations into which individual characteristics are more closely correlated with pedestrian activity. This could lead to different weightings for the various characteristics included in a PEF, making it a better descriptor.
Modelers in the Triangle Region, in addition to following the above recommendations, should:

- Expand the use of the GIS method to include all counties in the Triangle Regional Model. This will require the selection of a different set of characteristics because the remainder of the Triangle Region does not contain sidewalk files and the Chatham and Orange Counties do not include land use descriptors in their tax parcel data. However, the availability of street data, and the information that already exists in the TAZ structure will allow for this method to be used.

- Encourage the designers of the next Triangle Regional Survey to structure the survey in such a way that TAZ characteristics can be compared with walking trips. This will greatly aid in the selection of characteristics and in determining their thresholds.

- Make efforts to automate this process using the GIS-DK programming script. This could further reduce the time to calculate zonal PEFs and increase the ease of use.

To continue this research, the following should be done:

- Investigate combinations of characteristics, especially those in the same dimension such as employment and population density. These combinations could prove to be better at describing pedestrian activity than a single characteristic.
6.0 Findings and Recommendations

- Investigate probability models for predicting a zone’s pedestrian friendliness based on surveyed walking trips and zonal characteristics.

- Study applications of Bayesian statistics to help determine the chances of pedestrian activity given a specific PEF.

- Develop an automated method for measuring the accessibility of a zone, preferably on the parcel level and irrespective of zonal boundaries.

- Continue to develop automated methods to measure other characteristics as more GIS data becomes available.

6.5 Chapter Summary

The sixth chapter of this report reviews the project, lists and discusses findings and conclusions that derive from the project, and makes recommendations for future research.

The seventh chapter of this report is a list of the works referenced in this report.
7.0 References


Bhat, C. Lawton, K. “Passenger Travel Demand Forecasting” TRB A1C02 : Committee on Passenger Travel Demand Forecasting 2002.


USDOT. “Conditions and Performance Report Chapter 1 – Personal Mobility” 1999


Appendix A: Pedestrian Characteristics

A.1 Existing Models

In chapter 1.0 the Maryland-National Capital park and Planning Commission’s (M-NCPCC) model and the Portland METRO model were highlighted as being at the forefront of using indicators of pedestrian friendliness. The M-NCPCC model considers the availability of sidewalks, bicycle paths, and bus stop shelters, the extent of building setbacks from the street, and the heterogeneity of land use at a local level (Replogle 1990).

The Portland METRO model considers sidewalk continuity, ease of street crossings, local street characteristics (grid vs. cul-de-sac ie:connectivity), and topography (Cambridge, 1992). Ease of street crossing was evaluated based on the street width, the distance between signals, and the volume of traffic on the roadway. To rate sidewalk continuity, the sidewalk network was analyzed and gaps in the system were identified. Areas with high sidewalk continuity were rated higher than areas with no sidewalks or where significant barriers to sidewalk use were present. For street continuity, zones were examined to determine the degree to which streets were connected to each other. Zones with many dead-end streets and cul-de-sacs, scored lower than zones with a grid street network. Finally topography was analyzed. Zones with many steep slopes were given a lower rating than zones that are relatively flat.
Appendix A: Pedestrian Characteristics

In summary, the two models at the forefront of including pedestrian friendliness factors account for:

- Availability of Sidewalks and Bike Paths
- Bus Stop Shelters
- Building Setbacks
- Heterogeneity of Land Use
- Sidewalk Continuity
- Ease of Street Crossing
- Connectivity
- Topography

A.2 Additional Characteristics

However, these models, while at the forefront of considering characteristics of the built environment, do not consider all of the factors that can influence pedestrian activity. In fact, Cambridge Systematics identifies parallel transit malls, wide sidewalks, extensive street plantings and benches, transit shelters, and a light rail system that operates on surface streets as amenities that affect pedestrian behavior. Additionally, the socioeconomic characteristics of the residents (specifically smaller families with low income and students) and employment characteristics (high employment in a small area) are identified as having an impact on pedestrian activity. Finally, Cambridge Systematics studied residential density and transit level-of-service. While none of these factors are included in the Portland PEF, they could be included in the PEF for the Triangle Region.
Appendix A: Pedestrian Characteristics

Specifically related to transit, Scholssberg and Brown focus on neighborhoods that lie within one quarter and one half mile from a transit stop. Some of the primary factors they considered when attempting to rank these Transit Oriented Developments focus on intersection density, the density of dead end streets, higher than average development density, land use mix, roadway connectivity and design, and building design.

Low speed streets are also conducive to pedestrian travel. Speeds of 30 or 40 mph are appropriate on “traffic routes,” but on residential roads lower speeds around 20 mph are more appropriate. (Transport 2000 Trust, 2003) NCDOT Traditional Neighborhood Development Street design guidelines highlight low speeds and state that the upper limit of vehicle speed is 20 mph.

Berrigan and Troiano (2002) found that people who lived in urbanized areas in homes built prior to 1946 and between 1946 and 1973 were significantly more likely to walk than people living in homes built after 1973. They argue that home age is a useful proxy for neighborhood design; however the designs of neighborhoods built between 1946 and 1973 vary greatly and are not always consistent with neighborhoods built before 1946.

In one of the most comprehensive studies of how the built environment influences travel demand, Cervero and Kockelman studied several different characteristics and their influences on travel demand. Their built environment variables included, population density, employment density, accessibility to jobs, dissimilarity of land use, vertical
mixture, per developed acre intensities of land use, activity center mixture, commercial intensity, proportion of developed acres within ¼ mile of convenience stores or retail-service use, grid street design, four-way intersection proportion, number of blocks, number of dead ends and cul-de-sacs: average speed limit, street width, proportion of blocks with sidewalks, planting strips, street trees, overhead street lights, bicycle lanes, mid-block crossings, proportion of intersections with signalized controls; average block length, distance between overhead street lights, slope, off street parking, and on street parking. For each of these built environment variables the mean standard deviation, and range for the San Francisco area was indicated.
Appendix B: Wake County Land Use Codes

B.1 Type/Use

Residential
01 One Family
02 Two Family
03 Three Family
04 Four Family
05 Multi-Family
06 Residential w/ Bus. Use
07 Garden Apt.
08 Townhouse Apartment
09 Elevator Apartment
10 Rooming House

Bank Buildings
11 Bank Buildings
12 Drive-In only Banks

Bowling Centers
14 Bowling Alley

Special Type Buildings
16 Day Care
17 Club House
18 Bath House

Garage Type Bldgs.
19 Sales and Service
20 Service Garage
21 Parking Deck
22 Oil & Lube
23 Car Wash
24 Wand Car Wash

Gasoline Stations
25 Service Station
27 Other/Booth

Hotels/ Motels
28 Hotel/Motel w/ Full Facilities
29 Hotel/Motel w/ Limited Facilities
Appendix B: Wake County Land Use Codes

30 Motel - Ext. Stay
31 Hotel High Rise
32 Hotel/Motel - Ind.

Office Buildings
34 Typical Office
35 Ofc./Rtl/Res Conv.
36 Medical Types
37 Office/Apt.
38 Office Bldg. - HR

Restaurants
39 Restaurant
40 Fast Food
41 Plain Drive-In
42 Store Type Bldg.
43 Bar/Club
44 Cafeteria

Retail Type Buildings
46 Food Mart
47 Sgl. Tenant
48 Multi-Tenant
49 Supermarket
50 Discount Store
51 Department Store
52 Bulk Retail
53 Mall
54 Community Ctr.
55 Neighborhood Ctr.
56 Junior Anchor
57 Stores w/ Apts.
58 Stores w/ Offices
59 Stores w/Offices & Apts.
60 Junior Dept. Store

Other Finished Bldgs.
62 Airport Terminal
63 Vet Clinic
64 Clinic
65 Club
66 Church
67 Dormitory
68 Fire Station
Appendix B: Wake County Land Use Codes

69 Gymnasium  
70 Hospital  
71 Library  
72 Mobile Home  
73 Municipal Bldg.  
74 Nursing Home  
75 Funeral Home  
76 Retirement Home  
77 School  
78 Theatre  

Industrial Type Bldgs.  
79 Lt. Manufacturing  
80 Manufacturing  
81 Pharm. Plant  
82 Prefab Whse.  
83 Warehouse  
84 Bulk Dist. Whse.  
85 Flex Whse.  
86 Mini-Whse.  
87 Bottling Plant  
88 Chemical Plant  
89 Biological Plant  
90 R & D  
91 Hangar  
92 Power House  
94 Telephone Exch.  
95 Truck Terminal  
96 Laboratory  
98 Laundry  

B.2 Classification for Entropy Calculation  

Residential  
1-10, 67, 72  

Commercial  
11, 12, 16, 19-32, 39-64, 74-76  

Public  
68, 70, 71, 73, 77
Appendix B: Wake County Land Use Codes

Office
34-38, 66

Industrial
79-98

Entertainment
65, 14, 17, 18, 69, 78
Appendix C: Entropy Calculation

C.1 Introduction

Entropy is a measure of balance calculated by

\[-\sum (P_j \times \ln(P_j))/\ln(J)\]

where for calculating balance of land use: \( P_j = \) the proportion of land in the \( j \) use type
\( J = \) the total number of use types

The valued of the measure ranges from 0 to 1, with 1 representing a perfect balance of land uses. Kockelman separates the land use types into six categories: residential, commercial, public, office/research, industrial, and parks/entertainment.

<table>
<thead>
<tr>
<th></th>
<th>Res</th>
<th>Com</th>
<th>Pub</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Off</td>
<td>Ind</td>
<td>Ent</td>
</tr>
</tbody>
</table>

Figure C.1: A Tract With an Entropy of 1

Entropy was originally defined for the energy state of a system in the Second Law of Thermodynamics and is commonly used to quantify the uniformity of gaseous mixtures. The entropy measure was first used to quantify land use balance by Cervero (1989), in looking at suburban employment centers. Applied to land use it represents the balance of...
land use within a certain, predefined area. This area can obviously vary in size from very small to very large. Kockelman (1996) and Frank and Pivo (1995) this measure was applied to hectares within census tracts. That is, the balance of hectares within each census tract was analyzed.

For this analysis, a measure of entropy needs to be calculated for each TAZ (the smallest unit of analysis in a regional model) in Wake County. Therefore two main problems exist, TAZ size, and aggregation methods.

C.2 TAZ Size

TAZs vary in size (from .002 to 7.6 square miles), therefore individuals may walk across several TAZs or large TAZs may have artificially high entropy.

This is a problem that has not been dealt with before, because previous entropy analysis has been confined to census tracts. TAZs are generally very small in more dense areas such as CBDs, but it is in these areas that people are more likely to walk. Additionally, in the more suburban areas, TAZs are very large, so they could contain a large number of subunits, that if evenly balanced would result in high entropy. This is not realistic, because if these subunits were spread out, individuals would not necessarily walk in these areas. Therefore entropy needs to be measured based on the performance capabilities of
pedestrians, not based on boundaries set up for the purpose of regional modeling or census-taking.

For Wake County, the smallest subunit available within a TAZ is the parcel level. There are approximately 265,000 parcels in the Wake County dataset. A numerical value is attributed to each parcel describing the type of land use on that parcel. However, for some records (approximately 50,000), the type use is not recorded.

To tackle the problem of varying TAZ size we developed a computer program in Matlab to measure the entropy of a parcel based on a walking distance of \( \frac{3}{4} \) of a mile, the longest distance that pedestrians will generally walk. This method looks at every parcel in the dataset, and calculates the entropy using the type use of all parcels that have their center within \( \frac{3}{4} \) of a mile of the parcel being studied.
As mentioned before, some parcels do not have a type use code. However, these parcels still have entropy. To deal with this problem when calculating entropy for a specific parcel, the parcels within $\frac{3}{4}$ mile that do not have a type use code are disregarded. Therefore the sum of the proportions will always equal to 1 (or 0 in the case of no parcels within the $\frac{3}{4}$ mile having a type use code, therefore the entropy of the study parcel will be 0).

The result of this process is a measure of entropy for every parcel in Wake County.
C.3 Aggregation Methods

After calculating the entropy for each parcel, they must then be aggregated to the TAZ level to create an average entropy measure for each TAZ.

The intuitive method would be to simply average the entropies of all parcels within the boundaries of a TAZ. However, this creates a bias toward small parcels, the entropy of a large parcel factors the same into the average entropy as does a small parcel. To solve this problem a weighted average was computed. First, the parcels were assigned a number based on the TAZ that they fell within using ArcMap from ESRI. In the case of a parcel falling within two TAZs two records were created. Next, the area of each parcel was multiplied by its entropy. Then the parcel entropies multiplied by the area were summed by TAZ number. Finally, the sum was then divided by the total parcel area for each TAZ, resulting in a weighted average entropy for each TAZ.

The solution to this problem is best explained visually. Suppose you have a TAZ with 9 parcels. Each of the big parcels has an area of 1, and the small parcels have an area of .25.
The average entropy would be \((.1+.2+.3+.4+.5+.5+.5+.5+.6)/9 = .4\)

The weighted average would be
\[
\frac{(.1(1)+.2(1)+.3(1)+.4(1)+.5(.25)+.5(.25)+.5(.25)+.5(.25)+.6(1))}{(1+1+1+1+.25+.25+.25+.25+1)} = .35
\]

### C.4 Methodology for Computing TAZ Entropy

In the most basic form, this methodology can be broken down into 2 steps (Figure A.4). The inputs needed for this method are the TAZ layer for the study area, and a file containing a unique parcel identifier, land use code, parcel area, and x and y coordinates of the parcel’s centroid. In step 1, the computer code in Matlab is used to calculate the entropy for each parcel based on the parcels that are within \(\frac{3}{4}\) mile of the center parcel. In step 2, ArcMap is used to assign a TAZ number to each parcel then compute the weighted average entropy for each TAZ.
C.5 Benefits of Method

This computerized method for calculating zonal entropy offers several benefits beyond a more realistic measure of entropy. First, the computer program to calculate the entropy per TAZ is not restricted to a certain number of parcels. Updated parcel files with more, or less parcels are easily analyzed. Second, a minimal amount of data is required. Third, the computer program easily handles parcels that do not have a land use code, or have a code that is not in the pre-approved list of valid land use codes. These erroneous codes are simply excluded from the calculation. Finally, the land use codes are easily updated by modifying a few lines in the program in the event that the coding scheme is modified.
C.6 Entropy Script

The following is the MATLAB script used to calculate entropy for each parcel in the Wake County tax parcel database. The % sign indicates that the line is used for a comment.

% This Matlab script calculates land use entropy for tax parcels using data from a formatted, comma-separated, file. The script below is based on Wake County, NC land use types % The following column specifications are required for the input data file: % Column 1 - Parcel ID % Column 2 - Parcel Land Use Type % Column 3 - X-coordinate of parcel centroid % Column 4 - Y-coordinate of parcel centroid % Column 5 - Parcel area

% Set directory and read in data file
cd 'C:\Documents and Settings\Will\Desktop';
d=dlmread('ParcelArea.txt',';',1,0);
% Determine number of parcels in data file
rows=size(d,1);
% Initialize loop index and start timer to track progress
i=1;
tic;
% Initialize vectors to hold entropy values and errors
error=[];
entropy=[];
% Loop through each parcel
while i < rows+1,
    % The following if statement print timer value as the even 1000 parcels are processed
    if i/1000-round(i/1000)==0
        i
toc
    end
    % Read the coordinates of the current parcel
    x=d(i,3);
Appendix C: Entropy Calculation

```matlab
y = d(i,4);
% Initialize the vector to hold land use type proportions
p = [];
% Calculate vector of distances from current parcel to all others
dist = sqrt((d(:,3)-x).^2+(d(:,4)-y).^2);
% Create rox index of all parcels within 1/3 mile of current parcel
k1 = find(dist<3961);
% Create vectors of type and area for parcels within 1/3 mile of current parcel
type = d(k1,2);
area = d(k1,5);
s
k = find((type>0 & type<99) & (type~=13 | type==15 | type==26 | type==33 | type==45 |
   type==11 | type==12 | type==16 | (type>18 & type<33) | (type>38 & type<65) |
   (type>73 & type<77));
p = [p; sum(area(type2))/sum(area)];
type3 = find(type==68 | type==70 | type==71 | type==73 | type==77);
p = [p; sum(area(type3))/sum(area)];
type4 = find(type==66 | (type>33 & type<39));
p = [p; sum(area(type4))/sum(area)];
type5 = find(type>78 & type<99);
p = [p; sum(area(type5))/sum(area)];
type6 = find(type==65 | type==14 | type==17 | type==18 | type==69 | type==78);
p = [p; sum(area(type6))/sum(area)];
test = sum(p);
% The parcel ID and sum of area proportions is added to error vector if the sum
% of proportions does not equal 1
if abs(test-1)>0.0001
    error = [error; d(i,1), test];
end
% Create vector for land use categories
nz = find(p>0);
% Calculate the area-weighted land use entropy
entropy_i = -sum(p(nz).*log(p(nz)))/log(6);
else
    entropy_i = 0;
end
% Add the current parcel's entropy to the entropy vector and go to next parcel
entropy = [entropy; entropy_i];
i = i + 1;
end
% Create matrix of parcel ID's and corresponding entropy
final = [d(1:i-1,1), entropy];
% Write the entropy and error values to comma-separated files
dlmwrite('entropy.txt', final, ',')```
Appendix C: Entropy Calculation

dlmwrite('errors.txt',error,',')
toc