Using GIS and spatial modeling to examine active travel potential in a university town

Greg Rybarczyk, Ph.D.
Dept. of Earth and Resource Science
University of Michigan-Flint
OVERVIEW

- Introduction – Universities and Active Transportation
- Collecting the right data
- Exploratory Analysis + Spatial modeling
- Implications for stakeholders
INTRODUCTION

• **Background**
  - Universities are major trip generator-attractors
    - Traffic congestion
    - Personal Safety
    - Pollution
    - Attractiveness
    - Faculty retention

- Universities are microcosms of society
INTRODUCTION

- Universities and Active Transportation
  - Benefits of increasing walking, bicycling, and mass transit utilization among a university population
    - Reduce environmental externalities
    - Student recruitment
    - Beacon of sustainability
    - Educate next generation of planners/decision makers
INTRODUCTION

- Universities and Active Transportation
- University utilize Travel Demand Analysis (TDM)
  - Demand vs. Supply
  - Transit subsidies
  - Bicycle facilities
  - Traffic Calming
  - Programming
  - City-wide partnerships
  - Reducing parking demand
  - Courses
INTRODUCTION
What about spatial and aspatial factors at the trip-origin?

- Neighborhood conditions
- Family constraints
- Socioeconomic conditions
- Limited access to school
- Traffic conditions
- Distance
- Weather
- Social stigmas
- Etc…
INTRODUCTION

- **Research goals:**
  - Identify the factors that may cause a mode shift to active transportation at the trip start

- **Required tools:**
  - Survey instrument
  - GIS
  - Spatially explicit model
DATA

Survey Instrument

- Gender, university classification, home address
- Distance from UM-Flint
- Mode Share to & on UM-Flint campus
  - Auto
  - Bicycle
  - Bus
  - Walking
  - Scooter, etc.

- Conditions affecting mode choice
  - 13 Interventions to increase biking
  - 7 Barriers to decrease biking
  - 8 Interventions to increase walking
DATA

Neighborhood Context

- 2010 Environmental Protection Agencies “Smart Location Database”
- Census Block Group
  - Employment
  - Land-use Diversity
  - Demographics
  - Density
  - Design
- Transportation Environment
  - Bicycle and Pedestrian Crashes

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>Gross residential density (housing units per acre) on unprotected land</td>
</tr>
<tr>
<td>Diversity of land use</td>
<td>Employment and housing entropy</td>
</tr>
<tr>
<td>Urban design</td>
<td>Street intersections per square mile</td>
</tr>
<tr>
<td></td>
<td>High-speed road network density</td>
</tr>
<tr>
<td>Transit service</td>
<td>Aggregate transit service frequency, afternoon peak period</td>
</tr>
<tr>
<td></td>
<td>Distance to nearest transit stop</td>
</tr>
<tr>
<td>Destination accessibility by transit*</td>
<td>Working-age population within a 45-minute transit commute</td>
</tr>
<tr>
<td>Destination accessibility by car</td>
<td>Jobs within a 45-minute drive</td>
</tr>
<tr>
<td>Demographics</td>
<td>Percentage of households with no car, 1 car, or 2 or more cars</td>
</tr>
<tr>
<td></td>
<td>Percentage of workers that are low, medium, or high wage (by home and work locations)</td>
</tr>
<tr>
<td>Employment</td>
<td>Employment totals broken down by 5-tier classification scheme</td>
</tr>
</tbody>
</table>

Greg Rybarczyk, Ph.D.
**MODELING APPROACH**

- **Geocoding/Spatial Joining**
  - University Respondents & Smart Location Data

- **Exploratory Analysis**
  - Descriptive Statistics

- **Cluster Analysis**
  - Active travel & Predictors

- **OLS & GWR**
  - Global & Local model

- **Coefficient Mapping**

Greg Rybarczyk, Ph.D.
## EXPLORATORY ANALYSIS

### Distance Distribution of Mode Choice

<table>
<thead>
<tr>
<th>Distance</th>
<th>Walk %</th>
<th>Bike %</th>
<th>Car Pool %</th>
<th>Bus %</th>
<th>SOV %</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1 mile</td>
<td>68.6</td>
<td>5.9</td>
<td>0.0</td>
<td>0.0</td>
<td>25.5</td>
</tr>
<tr>
<td>1-5 miles</td>
<td>4.0</td>
<td>11.0</td>
<td>14.0</td>
<td>6.0</td>
<td>65.0</td>
</tr>
<tr>
<td>&gt; 5 miles</td>
<td>0.8</td>
<td>0.5</td>
<td>10.7</td>
<td>0.3</td>
<td>87.8</td>
</tr>
<tr>
<td><strong>OVERALL</strong></td>
<td>7.9</td>
<td>3.0</td>
<td>10.3</td>
<td>1.3</td>
<td>77.6</td>
</tr>
</tbody>
</table>

Greg Rybarczyk, Ph.D.
EXPLORATORY ANALYSIS

Bicycling facilitators for those who are > 5 miles from campus

Greg Rybarczyk, Ph.D.
# Exploratory Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Moran’s I</th>
<th>Pattern</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BikeWalkBus</td>
<td>1.355</td>
<td>Clustered</td>
<td>0.00</td>
</tr>
<tr>
<td>DistUMFlint</td>
<td>0.498</td>
<td>Clustered</td>
<td>0.00</td>
</tr>
<tr>
<td>Male(^a)</td>
<td>-0.038</td>
<td>Random</td>
<td>.513</td>
</tr>
<tr>
<td>Faculty(^b)</td>
<td>0.124</td>
<td>Clustered</td>
<td>.022</td>
</tr>
<tr>
<td>Student(^b)</td>
<td>0.099</td>
<td>Clustered</td>
<td>.068</td>
</tr>
<tr>
<td>% Households with Workers</td>
<td>1.301</td>
<td>Clustered</td>
<td>0.00</td>
</tr>
<tr>
<td>% Zero Car Households</td>
<td>1.300</td>
<td>Clustered</td>
<td>0.00</td>
</tr>
<tr>
<td>Res. Density</td>
<td>1.268</td>
<td>Clustered</td>
<td>0.00</td>
</tr>
<tr>
<td>Auto-orientated Facility Density</td>
<td>2.120</td>
<td>Clustered</td>
<td>0.00</td>
</tr>
<tr>
<td>Regional Job/Pop Diversity</td>
<td>1.492</td>
<td>Clustered</td>
<td>0.00</td>
</tr>
<tr>
<td>Bike/Ped Crash Density</td>
<td>0.839</td>
<td>Clustered</td>
<td>0.00</td>
</tr>
</tbody>
</table>

\(^a\) = control - female  
\(^b\) = control - staff

Greg Rybarczyk, Ph.D.
SPATIAL MODELING

Model Design
GWR coefficients

Coefficient Mapping
Interpolate \( \beta \)'s and \( t \)-values using IDW

GIS-Masking
Query significant \( \beta \)'s (95% C.I.)
SPATIAL MODELING

- Two different types of modeling approaches that can be implemented in GIS
  - OLS (Ordinary Least Squares)
    - Global regression model
    - One equation, calibrated using data from all features
    - Relationships are fixed
    - Does not account for spatial heterogeneity *(Wen et al., 2010)*
  - GWR
    - Local regression model
    - One equation for every feature, calibrated using data from nearby features

For each explanatory variable, GWR creates a coefficient surface showing you where relationships are strongest. *(Esri, 2010)*
**SPATIAL MODELING**

\[
GWR = y_i = \beta_{io} + \sum_{k=1}^{p} \beta_{ik}x_{ik} + \epsilon_i
\]

Where:

\(y_i\) = dependent variable at location \(i\) \((i = 1, 2, \ldots, n, \text{where } n \text{ is the number of observations})\),

\(x_{ik}\) = independent variable of the \(k\)th parameter at location \(i\),

\(\beta_{ik}\) = estimated \(k\)th parameter at location \(i\) for the GWR model,

\(\beta_k\) = estimated \(k\)th parameter for the OLS model,

\(\epsilon_i\) = error term at location \(i\), and

\(p\) = number of parameters
Typical GWR outputs include:

- Best-fit ($R^2$)
- Parameter Estimates (magnitude of influence, -+)
- T-values (distribution of significance, -+)

- Mapping only the parameter estimate can be misleading (where is it significant?)
- T-values have been displayed as contour lines over the parameter estimates

- Can be “messy” or differences may be too large to interpret
SPATIAL MODELING

- T-values overlaid onto parameter estimates (standardized coefficients)

- Provides a vague understanding of where estimates are statistically significant

Greg Rybarczyk, Ph.D. (Matthews, 2012)
Creative symbologies can be used to link estimates to t-values

(Mennis, 2006)
Where are the parameter estimates significantly affecting active mode choice potential?
**SPATIAL MODELING**

- **Integrating parameter estimates and statistical significance**

1. Create surface of local estimates ($\beta$’s) for each explanatory variable using Inverse Distance Weighting (IDW)

2. Create surface of $t$-values for each explanatory variable
   - Inverse Distance Weighting
   - $T$-values become the z’s

3. Classify $t$-value surface for statistical significance
   - Choose 3 classes and manually change ranges (break values) to -1.96 and 1.96

4. Choose “no-color” for the 1st and 3rd classes (transparent - unique). Choose white for the 2nd class (opaque)
   - This is used as a mask in ArcGIS

5. Produce bivariate color scheme to display parameter estimates

(Matthews et al., 2012)
SPATIAL MODELING

T-value partitions

1.96 is the approximate value of the 95\% point of the normal distribution used in statistics.

Other critical values can be queried:
99\%, 2.58
90\%, 1.645
SPATIAL MODELING

- Setting up the significance mask
We can now visualize where the parameter estimates may be affecting *local* active transportation mode choices, positively or negatively.
SPATIAL MODELING
SPATIAL MODELING

% Zero-car Households Parameter Estimates
Value
High: 60.82
Low: -2.26

Distance to Campus Parameter Estimates
Value
High: 19.20
Low: -7.70

Greg Rybarczyk, Ph.D.
SPATIAL MODELING

- GWR Model Residuals
**SPATIAL MODELING**

- **Overall Model Comparisons:**

<table>
<thead>
<tr>
<th>Overall Model Diagnostics</th>
<th>OLS</th>
<th>GWR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimated diagnostics of OLS and GWR models (n=520)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>OLS</td>
<td>GWR</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.473</td>
<td>.625</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.464</td>
<td>.587</td>
</tr>
<tr>
<td>Akaike's Information Criterion (AICc)</td>
<td>4671.26</td>
<td>4595.38</td>
</tr>
<tr>
<td>F statistic</td>
<td>45.71</td>
<td>16.35</td>
</tr>
<tr>
<td>Sigma</td>
<td>11.00</td>
<td>48.83</td>
</tr>
<tr>
<td>Residual sum of squares</td>
<td>231381.1</td>
<td>169547.67</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
IMPLICATIONS

- **Job/Housing diversity**
  - Positively influences north = increased mixed-use campus development?

- **Travel safety (crashes)**
  - Significant negative affect on active travel west of campus = educational programming efforts needed?

- **Household car density**
  - # cars per household increases has positive affect south of campus?

- **Distance**
  - Negative influences east of campus = closer park and ride locations?
  - No significant effect north

- **GWR outperformed OLS model**
IMPLICATIONS

- Universities are in the transportation business!

- Examining what will propel faculty, staff, and students to use mass transit, walking, and bicycling can have far ranging affects on the university and host

- Spatially explicit modeling approaches and novel symbolization techniques can highlight where “best-practices” are needed
Thank You