Predicting Drive Time to Care Sites using Great Circle Distance

Christopher L Richmond, Chase W Gatlin, and Mark F Guagliardo, PhD

Planning Systems Support Group
Office of the Assistant Deputy Undersecretary for Health for Policy and Planning
Veterans Health Administration
ESRI Health GIS Conference
October 18-19, 2010, Denver, CO
Overview

- **Background**
  - Travel time vs. straight line distance
  - VHA special case

- **Goal**

- **Methods**

- **Results**
  - Best national regression model
  - Maps of regression residuals
Background

- Network-based travel costs are best
- Not always available or attainable
- Various non-network estimates used
  - Cartesian distances
    - Euclidean distance (straight line)
    - Manhattan distance
  - Great circle distances
    - Haversine
    - Vincenty
Time – direct distance correlation gets mixed reviews in literature.
Goal

Develop the best possible regression model to predict drive time to VHA primary care stations from any residential point in the U.S. using x,y coordinates and easily obtainable residential area statistics.
Goal (cont.)

Rationale:

- Need quicker turnaround for travel time analysis in the VHA
- Smaller ad hoc analyses by other VA offices
Methods

Data

- **Origination points**
  - 500,000 geocoded VHA enrollee addresses (x,y)

- **Destination points**
  - 932 VHA primary care sites (x,y)

- **County and census tract attributes bundled by ArcGIS 9.3**
Methods

Travel time estimation

- Software Used: ESRI ArcMap Version 9.3
- Street Reference Data: StreetMap Premium North America Tele Atlas
- Centrus Group 1 Geocoder
- Drive time and drive distance calculated to the nearest VA Health Care Facility and included in each enrollee record
Methods

Great circle distance calculation

- Used SAS variable array processing to discover nearest of 932 sites
- distance = geodist(person_lat, person_lon, site_latit(i), site_lon(i), 'DM')
Methods

Great circle distance calculation

- 9.3% of the 500,000 enrollees had a different “nearest site” by vincenty distance versus drive time estimation

Drive distance - vincenty distance

- mean = 4.0 miles
- s.d. = 4.8 miles
- min = 0.0 miles
- max = 88 miles
Distance Type Differences Mapped
Methods

Predictor variables

- URH - a VHA indicator of rurality
- Census tract population
- Census tract population density
- Census tract percent minority
- Census tract percent owning home
- County crop acreage
Methods

Building the regression model

- SAS PROC GLMSELECT
  - Allows stepwise selection of predictor variables
  - Makes use of categorical predictors easy
  - Allows split-sample development of models
## Results

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>198082620</td>
<td>33013770</td>
<td>831211</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>498512</td>
<td>19799731</td>
<td>39.71766</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Tot</td>
<td>498518</td>
<td>217882351</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metric</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root MSE</td>
<td>6.30220</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent Mean</td>
<td>22.55120</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coeff Var</td>
<td>27.94616</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Square</td>
<td>0.9091</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj R-Sq</td>
<td>0.9091</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.64645</td>
<td>0.03787</td>
<td>17.07</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>vincentydist</td>
<td>1</td>
<td>1.55105</td>
<td>0.00090043</td>
<td>1722.56</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>URH_R</td>
<td>1</td>
<td>2.54448</td>
<td>0.02242</td>
<td>113.51</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>URH_H</td>
<td>1</td>
<td>10.42304</td>
<td>0.08092</td>
<td>128.80</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>PercentMin</td>
<td>1</td>
<td>-0.00385</td>
<td>0.00039966</td>
<td>-9.64</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>percentowned</td>
<td>1</td>
<td>0.02777</td>
<td>0.00044578</td>
<td>62.30</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>CROP_ACR97</td>
<td>1</td>
<td>-0.00000355</td>
<td>6.488861E-8</td>
<td>-54.74</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>
Residuals Mapped
Residuals Mapped
Residuals Mapped of Same Site Matches
Residuals Mapped of Same Site Matches
Residuals Mapped of Unmatched Site Assignments
Geographically Weighted Regression

- Using tool in ArcGIS 9.3
- Same independent variables and predictors
- Fixed kernel search distance of 100 miles
- Default bandwidth method (Akaike information criterion)
Geographically Weighted Regression

- R-square = 0.9377
- Regression parameters vary spatially, so not practical for our purposes
Next Steps

- Try some interaction terms
- Seek additional predictor variables
- Experiment with other GWR options
Limitations

- VHA regression formula may not be appropriate for other applications
Thank you.