Modeling the Rural Urban Interface in the South Carolina Piedmont:

T. Stephen Eddins (teddins@clemson.edu)
Lawrence Gering
Jeff Hazelton
Molly Espey
ABSTRACT

Continuing growth of America’s cities has created a zone where people who desire both rural and urban qualities reside. This zone is known as the rural/urban interface. Urban areas can provide greater job diversity and business opportunities than rural areas. They can also provide a greater chance of interacting with different people. Rural areas, on the other hand, can provide a sense of peace and tranquility and also numerous outdoor recreational opportunities. A mix between the rural area and the urban area might be the best place, if both types of land’s amenities are important factors in deciding on a location in which to reside. This rural/urban interface and can offer, for example, the atmosphere of a rural area with the job diversity of an urban area.

A model for defining the rural/urban interface for the Piedmont Region of South Carolina (Anderson, Greenville, Oconee, Pickens, and Spartanburg Counties) was developed, allowing the identification of these transitional zones. Landsat Thematic Mapper data provided a description of the current land use and land cover of the study area. Census data was collected to obtain housing density and income figures. Ground validation was conducted using GPS technology to verify the model. This validated model created a snapshot view of the interface zones of the Piedmont of South Carolina and provided the groundwork for potential dynamic models in the future.

KEYWORDS: Urban Rural Interface GIS Model Census
OBJECTIVES

- Geographically define the rural/urban interface region in the Piedmont Region of South Carolina.
- Utilizing the latest GIS software, expand upon previous studies.
- Create unbiased, qualitative field verification methods.

Using previous studies, 5 discrete rural/urban categories were developed. These categories were Urban, Partly Urban, Interface, Partly Rural, and Rural.

Screen captures from the PowerPoint presentation, along with the text that accompanied the slides, will be presented throughout this paper.

Figure 1: Five Discrete Rural/Urbm Categories Used in this Project
Outline

- Study Area and Municipalities
- Data Flow Diagram
- Creating the Model
  - Census Data
  - Land Use Data
- Field Verification
- Conclusions

Figure 2: Outline of the Steps to Create the Rural/Urban Interface
Figure 3: Study Area: Piedmont Region of South Carolina
MUNICIPALITIES

Major municipalities in the study area are shown below: Greenville, with a population of around 56,000, is the largest city within the study area. Spartanburg has a population near 40,000. Clemson, located in the Southwest corner of Pickens County, is where Clemson University is located.

Figure 4: Major Municipalities of the Piedmont Region of South Carolina.
Figure 5: Data Flow Diagram for the Definition of the Rural/Urban Interface in the Piedmont Region of South Carolina.
Data Sources and Types

In order to create a GIS layout, data files have to be available. These data files come from various data sources. The three main data sources for this project were the U.S. Census Bureau (www.census.gov), South Carolina Department of Natural Resources (http://www.dnr.state.sc.us/gisdata/), and the University of South Carolina’s GIS website (http://www.cla.sc.edu/gis/dataindex.html).

Blockgroup shapefiles were downloaded from census (the www.census.gov) website. The census block provides the statistical framework for a variety of census data. A blockgroup is a group of census blocks can contain one or more blocks. The blockgroup was chosen for this project because they provided a better representation of the study area than blocks. Housing density and median household income data were also downloaded from this census site. These files were exported in an Excel format and then saved in a database (DB4) format so that they could be brought into a GIS environment.

Landsat Thematic Mapper (TM) data were downloaded from the University of South Carolina’s GIS website. SCDNR supplied this data to the university. This satellite imagery data provided the land cover, in 17 different classifications, for the study area. This data was previously classified, but will be reclassified in a GIS environment so that the rural/urban interface categories can be used to query the data.

Auxiliary data were also downloaded for various steps in the project. Interstate, South Carolina Highway, and United States Highway shapefiles were downloaded from the University of South Carolina’s GIS website. Topographic maps, aerial photographs and municipality shapefiles were downloaded from SCDNR’s GIS website.
Projection

ArcGIS version 8.3, ArcView version 3.3, Terrasync, and GPS Pathfinder Office version 2.90 were the software programs that were used in this project. Two key layers of information, Census blockgroup shape files and Landsat Thematic Mapper data, were used in order to define the rural/urban interface. The blockgroup shapefiles were originally projected in UTM NAD 83 meters (Universal Transverse Mercator North American Datum 1983). In order to match the rest of the data, they had to be converted to UTM NAD 27 meters by using ArcToolbox. Once this file was converted, it could be opened using ArcMap.
Census Data: Housing Density

Housing density information was downloaded from the Census website. This step was repeated for each of the five counties in the study area. Manipulations of the data were the next step because the downloaded housing data did not provide housing density. The downloaded spreadsheet did give area (in square meters) and the number of housing units per blockgroup. The density was derived by the housing unit per blockgroup by the area (square meters). This number was then divided by 1,000,000 to give the housing density of each blockgroup per square kilometer. The main purpose of collecting this data is so it could be joined with the blockgroup shapefiles. In order to join two files together in ArcMap, they must share a similar column. The GeoID field of the spreadsheets was similar to the ID field of the blockgroups. There were a few characters that needed to be deleted in the GeoID columns before they matched perfectly.

After the database calculations and changes were applied to each of the five counties Excel spreadsheets, they were saved as a DB4 file. They were then added to an ArcMap layout along with the blockgroup shapefiles that were opened earlier. These databases could then be joined to the shapefiles so that future queries could be applied. From the table of contents, the name of each shapefile was right-clicked for each individual county. Scrolling down to “Joins and Relates,” and selecting “Join” presented a menu to complete the join function. From here, the appropriate database and the common fields joined. Now each individual blockgroup for each county had a unique housing density in square kilometers (Figure 6).

By using the “Select By Attributes” tool in ArcMap, the blockgroups could then be queried. The blockgroups for each county needed to be separated into one of the five categories for the project. These categories were urban, partly urban, rural/urban interface, partly rural, and rural. The Scenario I queries for the housing density data are given below:
### Table: Urban and Rural Density Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>density ≥ 71 units/km²</td>
</tr>
<tr>
<td>Partly Urban</td>
<td>density ≥ 46 units/km² and density ≤ 70 units/km²</td>
</tr>
<tr>
<td>Rural/Urban Interface</td>
<td>density ≥ 18 units/km² and density ≤ 45 units/km²</td>
</tr>
<tr>
<td>Partly Rural</td>
<td>density ≥ 5 units/km² and density ≤ 17 units/km²</td>
</tr>
<tr>
<td>Rural</td>
<td>density ≥ 0 units/km² and density ≤ 4 units/km²</td>
</tr>
</tbody>
</table>

After each query was run, the blockgroups within each county that met the criterion were selected and highlighted. A new layer was created for each category by using the “Create Layer From Selected” function on the table of contents for the original shapefile (Figure 7). These layers were exported so they could be accessed later. If they were just created and not exported, they could only be used in the current view of ArcMap. These 25 new shapefiles, (5 categories for each of the 5 counties) were then converted from a vector to a raster format. This conversion was done so that simple map algebra could be used to define the interface definition.

The “Spatial Analyst” extension had to be turned on under “Tools” and then *Extensions*, so that this conversion could take place. From the “Spatial Analyst” drop down menu, “Convert Features to Raster” was selected (Figure 8). One of the 25 shapefiles was selected, a field that contained the same information for each blockgroup was selected, and then the conversion was completed. The “County” field was selected in this process because it was exactly the same for every polygon in the shapefile. This allowed the entire output to have only one attribute. It made the entire raster output equal “1.”
Figure 6: Joining the Census Blockgroup Shapefiles with the Housing Density Data.
Figure 7: Selecting the Desired Blockgroups Based on Housing Density
Figure 8: Converting the Blockgroups to Raster.
Landsat Thematic Mapper: Satellite Imagery Data

Thematic Mapper Data from the South Carolina Department of Natural resources shows land classifications for the five counties in the study area. Seventeen different classifications were combined to make up this data. Subsets of these classifications were created to represent each of the five rural/urban categories. These subsets are presented below:

<table>
<thead>
<tr>
<th>Land Classifications for Landsat Thematic Mapper Data Used in Each of Five Rural/Urban Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
</tr>
<tr>
<td>High Density Urban</td>
</tr>
<tr>
<td>Low Density Urban</td>
</tr>
<tr>
<td>Evergreen Forest</td>
</tr>
<tr>
<td>Deciduous Forest</td>
</tr>
<tr>
<td>Mixed Forest</td>
</tr>
<tr>
<td>Cultivated/Grassland</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td>High Marsh</td>
</tr>
<tr>
<td>Low Marsh</td>
</tr>
<tr>
<td>Fresh Water Marsh</td>
</tr>
<tr>
<td>Deciduous Forested Wetland</td>
</tr>
<tr>
<td>Evergreen Forested Wetland</td>
</tr>
<tr>
<td>Bottomland Hardwoods</td>
</tr>
<tr>
<td>Wet Scrub/Shrub</td>
</tr>
<tr>
<td>Dry Scrub/Shrub</td>
</tr>
<tr>
<td>Sand beach</td>
</tr>
<tr>
<td>Barren/Disturbed</td>
</tr>
</tbody>
</table>

A reclassification was conducted in order to select each of the previous subsets for each category. The “Spatial Analyst” extension in ArcMap was marked to be turned on by selecting “Tools” and then “Extensions”. “Reclassification” was selected from the “Spatial Analyst” drop down menu (Figure 9). The original layer for each county was selected so that the subsets could
be created. By marking every land classification that was needed as a “1” and everything that
was not needed as a “0,” the new subsets were completed.

Figure 9: Reclassifying the Landsat Thematic Mapper Data.
Combining the Datasets

This reclassification can be used in conjunction with the newly created Census blockgroup raster dataset to complete the final steps of defining the rural/urban interface. Both of these datasets (they are called datasets because they are in raster format…they would be called shapefiles if they were in vector format.) were opened in ArcMap. Because the blockgroup dataset equaled “1” and the data that we need in the land use reclassification equaled “1,” a map calculation was used to add this data together and yield all of the areas that equaled “2.” The “Raster Calculator” was selected under the “Spatial Analyst” extension (Figure 10). The two datasets were added and then the calculation was evaluated. The output for this calculation yielded areas that equaled “1” (not needed) and “2” (needed).

The next step was to convert this calculation back to a vector format (Figure 11). Using the “Convert Raster to Features” under the “Spatial Analyst” extension, the raster dataset was converted to a shapefile. The next step was to use the “Select by Attributes” function to highlight all of the areas where the “Gridcode = 2.” A new layer was then created from the highlighted data by right clicking on the name of the new shapefile and then navigating to “Create Layer From Selected” (Figure 12). Finally, this new layer was exported and saved. Repeating the previous steps for each of the five categories in each of the five counties will give you the rural/urban interface for the Piedmont region of South Carolina (Figure 13).

To check to see how much of the total area that was defined by the model, an “Area” script in ArcView version 3.3 was used to calculate the area for all five of the rural/urban interface categories that were created. A shapefile was also collected from the University of South Carolina GIS website that contained each county in South Carolina. These processes included measuring the area of all the categories in a county and compare it to the total area of
the county itself. In order to do this, individual shapefiles for each of the five counties in the study area had to be created. Each county was highlighted; one at a time, and the “Create Layer from Selected” function was applied to each county. Then this new layer had to be exported so that a new shapefile could be created and accessed later if needed. Now this new exported county shapefile could be opened up in ArcView version 3.3 and the area could be calculated by using the “Area” script. This concluded the definition of the model.
Figure 10: Combining the Datasets.
Figure 11: Selecting the Desired Gridcode.
Figure 12: Completion of One Category for One County.
Figure 13: Rural/Urban Interface Categories for the Piedmont Region of South Carolina
Creating the Waypoints

After the model had been defined and the area had been calculated, ground verification was used to validate the model. Selecting random points over the study area and collecting a GPS position at each of the randomly assigned points completed this ground verification. The rural/urban interface is the area that this study places the most emphasis. The breakdown of the 100 random points over the five categories is below:

<table>
<thead>
<tr>
<th>Category</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>10</td>
</tr>
<tr>
<td>Partly Urban</td>
<td>15</td>
</tr>
<tr>
<td>Rural/Urban Interface</td>
<td>50</td>
</tr>
<tr>
<td>Partly Rural</td>
<td>15</td>
</tr>
<tr>
<td>Rural</td>
<td>10</td>
</tr>
</tbody>
</table>

A “Random Point Generator” script in ArcView version 3.3 was used to select the location of the points. This script placed any number (100 in this case) of randomly selected points in one selected polygon. A distance between each point can also be entered (500 meters in this case) when the script is being run. In order to get the script to run properly, a few geoprocessing functions needed to be applied to each category. By using the “Merge” tool under the “Geoprocessing Wizard,” one category for every county was combined into one shapefile that covered the entire study area. Eventually, this created one file for urban, one for partly urban, one for the interface, one for partly rural, and one for the rural category.

These five new shapefiles were comprised of many different polygons. This script only ran for one polygon at a time, so these shapefiles needed to be dissolved. The “Dissolve” tool under the “Geoprocessing Wizard” allowed the many polygons to be dwindled down to one. Then this dissolved shapefile, comprised of one polygon, was highlighted and the “Random
Point Generator” distributed the assigned number of points across the shapefile. This script created the points in a new file so they could be saved (Figure 14).

After the random points were assigned to each of the five categories, the five new point shapefiles were combined using the “Merge” tool under the “Geoprocessing Wizard.” Merging these files of 10, 15, 50, 15, and 10 points into one file of 100 points would allow for a bias in field verification to be prevented. This merge would create one file for all 100 of the randomly selected points. They would still have an rural/urban attribute attached to them, but it would not be present during the field verification (Figure 15). When the field verification was conducted, each point was located and then the category assigned to that location was recorded. This ground verification was compared with the 100 points that were created with the “Random Point Generator” for validation of the model.

Before the fieldwork could be started, the X and Y coordinates for each point needed to be calculated so they could be found in the field. Another script in ArcView created these coordinates. These coordinates were exported to an Excel file, and then saved as a .CSV file that is comma delimited. This file contained only two columns and no headings for the columns. These coordinates were set up with the X values in the first column and the Y values in the second column before they were saved.

This file was then opened in GPS Pathfinder Office Version 2.9 as a waypoint file. “ASCII import” for the “Waypoints” under the “File” menu was selected to open this file. This added all 100 points that were created earlier as a waypoint. These 100 waypoints were then transferred into the GPS unit by connecting the rover unit to the computer that was running Pathfinder Office. Selecting “Data Transfer” from the “Utilities” menu performed this transfer. For this study, a Trimble GeoXM rover unit was used to conduct the field verification.
Figure 14: Random Field Validation Points Prior to Combination

Figure 15: Random Field Validation Points After Combination.
Collecting the Waypoints

Road data was obtained from the USC GIS website so that a map showing the field verification points could be overlaid on the road system of each county. After creating this map, a printout was used to determine the route that needed to be taken to collect the points in the quickest manner. A couple of days were devoted to each of the counties in the study area to collect the field verification points. The number of points that were located in each county follows:

- Anderson: 15
- Greenville: 25
- Oconee: 13
- Pickens: 20
- Spartanburg: 27

These points were not evenly distributed over the counties because they were evenly distributed over the study area for each category. The GeoXM rover unit from Trimble proved to be a very valuable tool in completing this field verification. By using the navigation tool in the Terrasync (the GPS software package used with the rover unit), an arrow pointed directly to the waypoint that was desired. Once the waypoint was located, an “X” appeared on the screen of the rover unit.

Once the desired point was reached, the waypoint file was closed and another file was opened so that a GPS position could be collected with a rural/urban category attribute attached to it. The predetermined number of the point was also attached to this point. The number was created in the office using ArcGIS. The category that was attached to this point was determined by looking at the surroundings of the point.

A system was created so that the housing density of each point could be calculated. There is no way that the entire square kilometer could be measured at the center of each point, so
a subset of the point was taken. It was decided that at least 10 percent of the square kilometer needed to be measured, and the houses that were measured would be multiplied by 10 to get the total housing density for the point. Every house that lied within 178.41 meters of the point center was tallied. Basic mathematic concepts were the basis for this concept. One square kilometer is comprised of one million square meters. Ten percent of one million is 100,000 square meters. The use of a Bushnell Yardage Pro Rangefinder that gave measurements in both yards and meters was used at each point center. The area of a circle is $\pi r^2$. A circle with an area of 100,000 square meters was desired for this validation. The square root of 100,000 divided by $\pi$ (~3.14) equaled 178.41 square meters.

A center of the point was established and a piece of flagging was placed to the North to mark the starting and stopping point for the individual house tally. Each house that fell within the required boundary was tallied and the total number of houses was multiplied by 10 to get the total housing density for the point (Figure 16). The housing densities for each category were the same as the query that was created in the office. The queries that were used follow:

- **Urban:** $\geq 71$ houses/km$^2$
- **Partly Urban:** $\geq 46$ houses/km$^2$ and $\leq 70$ houses/km$^2$
- **Interface:** $\geq 18$ houses/km$^2$ and $\leq 45$ houses/km$^2$
- **Partly Rural:** $\geq 5$ houses/km$^2$ and $\leq 17$ houses/km$^2$
- **Rural:** $\geq 0$ houses/km$^2$ and $\leq 4$ houses/km$^2$

Once the housing density for each point was determined, a GPS position was collected with the appropriate category attached. After the point was collected, the data file was closed and then the waypoint file was reopened so that then next point could be found.
Each housing unit was measured within a 178.14-meter diameter of the validation point center. In this example, three housing units were measured. Three was multiplied by an expansion factor of 10. This equaled 30 housing unit per square kilometer. This would make this point fall in the interface category based on the criteria for the blockgroups presented earlier.

Figure 16: Representation of a Field Validation Point used to Verify the Rural/Urban Interface Model
Figure 17: Study Area Contingency Table Comparing Predicted and Actual Results from the Rural/Urban Interface in the Piedmont Region in South Carolina.

<table>
<thead>
<tr>
<th>Ground Truthing</th>
<th>Urban</th>
<th>Partly Urban</th>
<th>Interface</th>
<th>Partly Rural</th>
<th>Rural</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Partly Urban</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Interface</td>
<td>1</td>
<td>5</td>
<td>48</td>
<td>1</td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>Partly Rural</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>Rural</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>15</td>
<td>50</td>
<td>15</td>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>
Conclusions

The first objective of this project was to define the rural/urban interface in the Piedmont region of South Carolina. Along with this interface, four more categories were defined. These other categories were rural, partly rural, urban, and partly urban. The methods that were used to create this model were expanded to the entire Piedmont region of South Carolina.

The second objective was to develop a process of predicting the rural/urban interface using ArcGIS version 8.3. A raster-based method, combining census data and landcover data, was used to complete this objective.

The third objective was to improve the method of ground validation from a subjective method that may have contained biases to a method that was repeatable, with major aspects of bias removed. This repeatable method of validation was achieved by creating random points across the study area. Prior knowledge of the rural/urban category attached to these random points was hidden before the fieldwork was conducted. A fixed-radius plot was attached to these random points and the number of housing units was taken at each location. The data that were collected out in the field were multiplied by an expansion factor of 10 to convert to “per-square kilometer” values and compared to the predicted data that was generated in the office.

The rural/urban interface for the Piedmont region of South Carolina is just a small region relative to all of the regions of the United States that are experiencing urban development. However, the same definition processes can be applied to a much larger project in the future. The next step for the identification of the rural/urban interface in the Piedmont of South Carolina would be a predictive, dynamic model that would show the growth of the interface over time. This model could build on the previous studies and also utilize earlier census data. This census data, from decade to decade, can help give an idea of the past growth rates of the rural / urban
interface. Using growth rate data from the past, the future rural/urban categories may be predicted for the Piedmont of South Carolina.