Modeling Growth and Predicting Future Developed Land in the Upstate of South Carolina

Report submitted to the Saluda-Reedy Watershed Consortium
November 2007

SALUDA-REEDY WATERSHED CONSORTIUM

Craig E. Campbell, Jeffery Allen, and Kang Shou Lu

Strom Thurmond Institute, Clemson University

CLEMSON UNIVERSITY

Modeling Growth and Predicting Future Developed Land in the Upstate of South Carolina

(Abridged)

Abstract

From 1990 through 2000 the amount of developed land in an eight-county region of Upstate South Carolina grew from 222,745 acres to 576,336 acres. Under current practices and policies the amount of developed land is anticipated to grow to 1,523,667 acres by the year 2030. Where that growth takes place can have serious impacts and can affect the character of the region. The Upstate contains an abundance of natural, environmental, and cultural resources that could be at risk from unmanaged growth.

The Strom Thurmond Institute has had previous success modeling future growth of developed land for the area around Charleston, South Carolina. For this study a comparable model, with some improvements, was developed to predict where the growth is most likely to occur through the year 2030 for eight counties in the Upstate region of South Carolina. A geographic information system-based model was developed, combining a binomial logistic regression approach with expert information provided by informed participants from throughout the region. A map created from the output of the growth model shows what the pattern of developed land for the study area might look like by the year 2030. These results can give decision-makers better information from which to implement good growth policy for the future of the region.

Introduction

The Strom Thurmond Institute (STI) and the SC Water Resources Center (SCWRC) have shown success in producing a model for urban growth prediction. In a previous project, STI and SCWRC used geographic information systems to model and predict the spatial extent of future urban growth for the Charleston Tri-County area (Berkeley, Charleston, and Dorchester Counties) through the year 2030[1]. The prediction was based on the historical trends found in a NASA-funded 1973-1994 satellite image change detection study, assuming current policy constraints. The objective was to provide a model to give decision-makers better information from which to implement good growth policy for the BCD area as well as South Carolina.

In the Charleston study, it was found that while the population of the Tri-County region grew 41 % between 1973 and 1994, the urban area grew 256%. For the STI model for future growth it was anticipated that the population would grow another 49% and the urban area would increase by 247%.

---

[1] Modeling and Prediction of Future Growth in the Charleston Region of South Carolina: a GIS-based Integrated Approach. Jeffery Allen and Kang Lu. Conservation Ecology **8**(2): 2 (2003). [online] URL: http://www.consecol.org/vol8/iss2/art2 .

For the current study a comparable growth model was developed for the Upstate region of South Carolina. The results of this project should enable scientists and decision-makers to do a better job of planning for the future of the region.

The growth model was developed for the eight counties of the Upstate that make up the Saluda River-Reedy River Watershed: Greenville, Spartanburg, Pickens, Anderson, Laurens, Newberry, Abbeville, and Greenwood. This large area contains a large variety of landscapes and features, including mountains in the northern portions of Pickens and Greenville Counties, a chain of large lakes forming the western border of Pickens, Anderson, and Abbeville Counties, several river systems traversing the region from the northwest toward the southeast, and the two major cities of Greenville and Spartanburg. The region is crossed by several Interstate highways (I-85, I-26, I-385), and just beyond the study area lie the major metropolitan areas of Charlotte, NC to the northeast and Atlanta, GA to the southwest. Figure 1 shows a map of the study area.

The 8 counties of the study area cover 3,345,532 acres. The population for all 8 counties grew from 960,750 in 1990[2] to 1,108,017 in 2000[3], an increase of 15.33% in 10 years. That population is forecast to grow to 1,472,270 by the year 2030[4], an increase of 32.87% over 30 years. The breakdown by county for area and population is listed in Table 1.

Table 1: Total Area and Population for the 8 Counties in the Study Area.

County	Total Area (acres)	Population				
		1990	2000	1990-2000 Change (%)	2030	2000-2030 Change (%)
Greenville	510,073	320,167	379,616	18.6	521,990	37.5
Spartanburg	524,274	226,800	253,791	11.9	332,450	31.0
Pickens	327,316	93,894	110,757	18.0	154,610	39.6
Anderson	484,660	145,196	165,740	14.1	215,380	30.0
Laurens	461,945	58,092	69,567	19.8	92,310	32.7
Newberry	414,133	33,172	36,108	8.9	43,580	20.7
Abbeville	326,955	23,862	26,167	9.7	30,790	17.7
Greenwood	296,175	59,567	66,271	11.3	81,160	22.5
Total	3,345,532	960,750	1,108,017	15.3	1,472,270	32.9

[2] US Census data, via SC Department of Commerce SiteSCope CD, also via 1998 ESRI Data & Maps, CD 1.

[3] US Census data, via 2002 ESRI Data & Maps, CD 7.

[4] South Carolina Population Reports: South Carolina Population 2005 – 2030; Source: Office of Research and Statistics, Health and Demographics Division. Based on 2003 Census population estimates. A publication of the: South Carolina State Budget and Control Board, Office of Research and Statistics, Health & Demographics Division, 1919 Blanding St., Columbia, SC 29201.

This project utilized a binomial logistic regression approach to model future land use changes, based upon a historic land use change detection. Geographic variables having spatial attributes, including physical variables, accessibility factors, initial conditions, and policy constraints, were used in the model to predict urban transition probability.

The second important component used for the model was the incorporation of *Expert Group Input*; informed information contributed by knowledgeable representatives from each of the counties. It is believed that involvement of interested and knowledgeable persons is vital to the creation of a valid model. The expert group input was combined with the logistic regression model to create a more accurate and informed final model.

The modeling process used for this study does not model or predict the future population. The forecast population figures are predetermined at the outset, and serve as an input to the model. The population forecast, along with a ratio of developed land growth to population growth, determines, quantitatively, the final amount of developed land area before any modeling is performed. The GIS-based growth modeling allocates the growth geographically, identifying where that growth is most likely to occur.

The Model

To perform the modeling of future growth of developed land for this project, a GIS-based logistic regression model developed previously by STI was used. This model runs within the ESRI ArcView GIS 3.3 application. A brief overview of how the logistic regression model operates is given.

As with any GIS-based project, the most important and most time-consuming step is collection and preparation of the input data. To begin, appropriate data sets delineating the study area must be obtained or generated. A prerequisite for the growth model is a pair of quality GIS data sets depicting the developed land for the study area at two points in the past. This allows an analysis of the change in developed land over a given historical time period. Ideally, the second time point is as recent as possible to serve as an accurate starting point for the future modeling. It is best if the dates of the developed data coincide with those of the population data. Generally, the developed land data sets are often raster images that have been extracted from land cover data derived from remotely-sensed imagery.

A set of input variable geographic data sets is required. These data sets are geographic features that are believed to have had some influence on the growth in developed land observed between the two historic time points. Examples of input variables are Interstate highways and other roads, the slope of the land, and infrastructure services such as water lines and sewer lines. These input feature data sets generally are converted into raster data sets in which each cell in the raster represents the *distance to* that feature. Cells with a greater distance to a road, for example, might be less likely to develop than those closer to a road.

Some input variable features change over time. New roads and water lines are constructed; old schools are closed and new ones are built. Ideally it is desirable to have two versions of such variable data sets; one to use for establishing the correlation between that variable and the growth between the initial two historical dates, and the second for determining the probabilities for future growth. For example, if the developed land data sets are for 1990 and 2000, it would be ideal to have the roads as they were in 1990 to correlate with the growth observed between 1990 and 2000. But then for modeling the future growth it is desirable to have the most current version of the roads that is available. In some cases data for multiple dates may not be available, or the time and effort may be prohibitive. In other cases the variable does not change with time; for example slope.

Working with the two historic developed land data sets and the input variables, the model uses binary logistic regression to establish the correlation between each variable and the observed change in developed land. The result of the logistic regression analysis is used to generate a future "probability grid," using the most current available versions of the input variables. The value of each cell in the probability grid indicates the relative likelihood of that cell becoming developed. Cells that are already developed at the start are given a probability of 1.0. If proper steps are taken, cells that are to remain undeveloped, such as water, wetlands, or protected lands, can be given a probability value of 0.0. In between 0 and 1, cells with higher probability values are more likely to develop than those with lower probability values, or, temporally, they will develop before those with lower values.

Once the probability grid is complete, the amount of existing developed land, the future population forecast, and the ratio of developed land growth to population growth are used to calculate the desired developed land area at some point in the future. (See Equation 4 below under *Procedure*.) The GIS growth model then uses the probability grid to select cells, starting with the highest probabilities and working down, until the total area is equal to the desired future area.

Data

Geographic data layers prepared for input to the Upstate growth model are listed in Table 2.

As noted in the introduction, many input feature data sets for features believed to influence growth are converted into raster data sets in which each cell in the raster represents the distance to that feature and are entered into the growth model as *distance-to* grids. Two of the inputs (slope and population density) are already by nature in a form useable as input rasters, where distance is irrelevant.

Most of the input feature data sets listed in Table 2 are entered into the logistic regression analysis as *independent variables*. These are the variables that control or influence the growth observed. Two of the data sets listed in the table, protected lands and wetlands, are not used in the logistic regression at all, but are used to exclude future development in the "Urban Classification" phase of the growth model. The use of each input feature is noted in the table.

Use of the developed land data sets requires special explanation. In the first phase of the model the correlations are established between the independent variables and the observed growth between the two initial time periods (1990 and 2000 in this case). The second of the two developed land data sets (2000) represents the observed growth, and thus is the *dependent variable*. That observed developed land is controlled by, or dependent on, the independent variables. In this phase the first of the two developed land data sets (1990) is used as an independent variable, both in its native form, as the fact of a cell being developed at the first time point (1990) controls its being developed at the second time point (2000), and as a *distance-to* grid, because proximity to currently developed land may influence the likelihood of becoming developed. Then in the second phase of the model, when the future probability grid is generated, the second or most recent developed land data set (2000) is used as an independent variable, or input, just as the first set was used in the first phase. The first developed land data set (1990) is not used at all in the future phase of the model.

Selection of a geographic analysis extent was necessary prior to creation of all input data sets. Although the area for this study is the 8 Upstate counties listed, data sets were created for a slightly larger geographic area to eliminate or reduce possible edge effects. (For example, an Interstate highway passing just beyond the county boundary might have an affect on growth in the region within the study area, but if that highway is left out of the model that affect would be completely missed.) A ten-mile buffer was created beyond the 8-county area and a rectangular box was created around that ten-mile buffer. The resulting rectangular area, which encompassed parts of Georgia and North Carolina as well as additional South Carolina counties, was used for selection and extraction of all input data sets. Figure 2 shows the 10-mile buffer and the data-creation analysis extent in relation to the study area. The developed land rasters were extracted from STI land cover data (see below) using this rectangular analysis area. The native cell size of the STI data was 30 meters by 30 meters. The properties of the STI developed land rasters (cell size, extent, projection, datum, and units) were used as the basis for all other raster data sets created.

Two available land cover data sets were compared for use as the developed land input layers: the National Land Cover Data (NLCD) for 1992 and 2001 from the Environmental Protection Agency's (EPA) Multi-Resolution Land Characteristics Consortium[5] (MRLC) and a classification done by Clemson University's Strom Thurmond Institute (STI) for 1985, 1990, 1995, and 2000[6]. The STI data for 1990 and 2000 was chosen for this project. These dates corresponded with the dates for the census population figures. Developed land grids were created for the 1990 and 2000 STI data by extracting only the developed land class from the land cover data.

After minimal filtering of the data, the amount of developed land for the 8 counties was found to be 222,745 acres in 1990 and 576,336 acres in 2000. The breakdown of developed land in 1990 and 2000 is listed in Table 3. The map in Figure 3 shows the developed land in 1990 and 2000 for the 8-county study area.

---

[5] MRLC: An Innovative Partnership for National Environmental Assessment, Multi-Resolution Land Characteristics Consortium (MRLC), U.S. Environmental Protection Agency. [online] URL: http://www.epa.gov/mrlc/.

[6] Allen, J., S. Sperry, A. Pasula, V. Patki and K. S. Lu. 2005. Land Cover Classification and Land Cover Change Analysis for the Saluda-Reedy Watershed. Report submitted to the Saluda Reedy Watershed Consortium and Upstate Forever. Greenville, S.C.

Table 2: Geographic data layers prepared for input to the Upstate growth model.

Input Data Sets	Date	Function
Developed Land	1990	independent variable
Developed Land	2000	dependent variable, then independent variable
Interstate Highways	1990	independent variable
	2001	independent variable
U.S. Highways	1996	independent variable
Primary Highways	1996	independent variable
Secondary Highways	1996	independent variable
	2006	independent variable
Streets	1990	independent variable
	2000	independent variable
Highway Nodes	1996	independent variable
	2001	independent variable
Rivers & Lakes	na	independent variable
Incorporated Areas	1990	independent variable
	2000	independent variable
Water Lines	1998	independent variable
	2002	independent variable
Sewer Lines	1998	independent variable
	2002	independent variable
Public Schools	1990	independent variable
	2007	independent variable
Greenville, Spartanburg, & Anderson	na	independent variable
Lake Keowee	na	independent variable
Lake Hartwell	na	independent variable
Clemson University	na	independent variable
Slope	na	independent variable
Population Density	1990	independent variable
	2000	independent variable
Protected Lands	2006	exclude from growth
Wetlands	na	exclude from growth

Table 3: Developed Land Area for the 8 counties in the study area in 1990 and 2000.

	Developed Land (acres)		Change (%)
County	1990	2000	
Greenville	52,015	137,823	165.0
Spartanburg	43,456	130,710	200.8
Pickens	16,632	48,335	190.6
Anderson	49,296	107,055	117.2
Laurens	20,913	51,030	144.0
Newberry	13,968	35,373	153.2
Abbeville	11,373	28,297	148.8
Greenwood	15,092	37,712	149.9
Total	222,745	576,336	158.7

Growth Ratios

One indication of the intensity of new development is the ratio of the change in the amount of developed land to the change in population.

$$Growth\ Ratio = \frac{\%\ change\ in\ developed\ land}{\%\ change\ in\ population}$$, where *Equation 1*

$$\%\ change\ in\ developed\ land = \left[\frac{(area_2 - area_1)}{area_1}\right] \times 100\%$$ and *Equation 2*

$$\%\ change\ in\ population = \left[\frac{(population_2 - population_1)}{population_1}\right] \times 100\%$$ *Equation 3*

A growth ratio of 1:1 would not indicate a case of no growth, as is often mistakenly inferred. A 1:1 growth ratio would indicate that a population increase of 10% would be accompanied by a 10% increase in developed land. Any ratio greater than 1:1 indicates that the *per capita* growth of new developed land exceeds the *per capita* footprint of developed land to date. Note that the growth ratio has nothing to do with time; it is based simply on the changes in developed land and population over any selected period of time.

Growth ratios in excess of 10:1 have been reported in the U.S. in recent decades[7]. For the Charleston Tri-County region of South Carolina from 1973 to 1994 a growth ratio of 6.2:1 was found, and a ratio of 5:1 was used for a year 2030 future growth modeling project conducted by the Strom Thurmond Institute and the SC Coastal Conservation League.

The overall growth ratio for the Upstate 8-county area of this study from 1990 to 2000, using the figures from the minimally-filtered STI land cover data, was 10.36:1. The

[7] Rusk, D, Blair, J and Kelly E.D. (1997). Debate on the theories of David Rusk. Edited transcript of proceedings in The Regionalist 2(3): 11-29.

ratios for each county individually varied from this; some being higher (as high as 16.9:1 for Spartanburg County) and others being lower (as low as 7.3:1 for Laurens County). A future growth ratio of 5:1 was chosen for this modeling project. This was believed to be a conservative figure that would produce believable results. As with the historic county-to-county variation, if the future growth ratio for the entire region was 5 to 1, it would not be exactly 5.00 to 1 for each of the eight individual counties, but would vary above and below 5:1. A future growth ratio was calculated for each of the 8 counties, proportional to that observed from 1990 to 2000, so that the overall growth ratio for all 8 counties would be 5.00:1. These ratios are listed in Table 4. Thus, Spartanburg and Laurens Counties were given future growth ratios of 8.14:1 and 3.52:1, respectively. (In the final methodology, these individual county future ratios were not used, but they were used in trials where growth due to county population growth was confined to each county.)

Table 4: Future Growth Ratios for Upstate Counties if the Overall Growth Ratio was 5:1 and if growth stayed proportional to that observed from 1990 to 2000.

County	Growth Ratio
Greenville	4.29
Spartanburg	8.14
Pickens	5.12
Anderson	4.00
Laurens	3.52
Newberry	8.36
Abbeville	7.44
Greenwood	6.43
Overall	5.00

Procedure

Future Developed Land Area

Base and forecast population data and base developed land data, by county and overall, have been listed in Tables 1 and 3 above. The amount of future developed land area is entirely determined by the existing developed land, the population forecasts, and the future growth ratio chosen, according to the following equation:

$$A_2 = A_1\left(1 + R\left(\frac{(P_2 - P_1)}{P_1}\right)\right) \qquad \textit{Equation 4}$$

where

$P_1 = initial\ \ population$,

$P_2 = \ final\ \ population$,

$R = developed\ land\ growth / population\ growth\ ratio$,

$A_1 = initial\ developed\ area$, and

$A_2 = final\ developed\ area$.

This can be made clear by a hypothetical example. Assume that the current developed land is 1000 acres and a growth ratio of 5:1 has been chosen. If the population is forecast to increase by 10 percent, the growth ratio dictates that the developed area will increase by 5 times that, or 50 percent. Thus the developed land will increase by 500 acres and the final area will be 1,500 acres.

Given the developed land area for the 8 counties for the year 2000 of 576,336 acres and an overall growth ratio of 5:1, the predicted developed land by the year 2030 is 1,523,667 acres. If counties were modeled individually using the growth ratios from Table 4 above and developed land growth was limited to the county, 2030 developed area by county would be as listed in Table 5. Note that the overall figure for the 8 counties together is not equal to the sum of the individual counties because the overall growth factor is not equal to the average of the county growth factors.

Table 5: 2030 developed land targets (nominal) based on the growth ratios in Table 4 (5:1 overall) and growth limited to county boundaries.

County	2030 (acres)
Greenville	359,466
Spartanburg	460,579
Pickens	146,366
Anderson	235,201
Laurens	109,728
Newberry	96,542
Abbeville	65,467
Greenwood	92,169
Overall	1,523,667

The logistic regression model determines where the new development is most likely to occur.

Logistic Regression Model

Several approaches were tested and evaluated before selecting the methodology ultimately used for the logistic regression portion of the growth model. The eight-county study area is a very large region and there was concern about modeling it as a single area. Given the diversity of the region, some of the input variables vary widely not only in their contribution to the probability of development, but even in their existence. For example, some of the counties have no Interstate highways within their boundaries; some counties have lakes within or adjacent to their boundaries while others do not.

Attempts were made at modeling development regionally (all eight counties or all thirteen SC counties wholly or mostly within the rectangular study area) and modeling each county individually. The individual county approach forces the new development due to a county's population growth to remain in that county, while the regional approach allows development to be distributed freely throughout the region without regard to county boundaries.

Under both methods the new growth spread out across the region in a spindly pattern, following every county street into rural and mountainous areas, instead of clustering more densely around already-developed centers. Furthermore, under the individual county approach, while the future growth patterns matched fairly well across county borders in most areas, there were some regions of discontinuity. Compared to the mosaic of the individual county models, running the eight counties together alleviated some growth from Spartanburg, Greenville, and to a lesser extent northern Pickens Counties and redistributed it to the other, less developed counties.

While experimentation conducted in an attempt to eliminate the spindly pattern along county roads and redistribute new development into more likely areas did not produce the desired results, it did result in the creation of a new input variable grid. The result, referred to as *percent available land developed*, was a grid of the percent of available land in 1990 that had become developed by 2000, by block group. (Developable land was determined by subtracting the 1990 developed area from the total area, neglecting area that may not be developable due to water, wetlands, protection, slope, etc.)

(percent available land developed = 2000 developed land – 1990 developed land) / (total land - 1990 developed land)

It was discovered that when running the eight counties together the model is overwhelmingly controlled by a small set of the most influential variables. Using only the 10 input variables with the consistently highest correlations, the 2030 result was almost indistinguishable from the previous 8-county result produced using the full variable set. It is noted that this is the case in the 8-county model, and that different variables may become significant at more local levels.

Another finding was that classification of *distance-to* variables into discreet classes consistently increased the magnitudes of their correlations. Several of the continuously-varying variables were converted to classified variables and the new classified data sets were used for subsequent modeling.

Other tests showed that it is not detrimental, and may be beneficial, to remove input variables that are not contributing significantly. This information indicates that it is acceptable to apply judgment on exclusion of variables on a county-by-county basis, rather than to use a blanket application of the above results across the board for all counties.

In review, in one approach the model was run for each county independently, using target developed areas for each county derived from that county's 2000 developed area, the population forecast, and the selected ratio of developed area growth to population growth. (5:1 in this case). This kept all developed area growth due to a county's population growth within that county, not allowing for any development across borders. This may not be realistic, for example, in cases where businesses are expanding in one county but affected residents are living in another. The same model was also run for the eight counties as a single unit, using the target developed area for the whole region derived from the total 2000 developed area, the total population forecast, and the same growth ratio. This allowed spillover across county borders, probably giving a more realistic simulation, but not allowing for the influence of specific variables at the local level. It was determined that a reduced variable set was sufficient, and probably better, for the regional model.

Table 6: Strengths and Weaknesses of the Modeling Approaches

	Strengths	**Weaknesses**
Single-County Approach:	• retains influence of local variables	• discontinuities at county boundaries • restricts development due to population to within county boundary
8-County Approach:	• eliminates discontinuities across county boundaries • alleviates some growth from most heavily developed counties	• lose influence of local variables

The final process chosen was a hybridization of the two approaches to overcome the weaknesses and keep the strengths of each. The amount of development assigned to each county by the limited-variable set, 8-county model was tabulated. These predicted developed areas were then used as the target developed areas for running the model once more for each county individually. It is noted that this approach allowed the growth ratio for individual counties to stray from the nominal ratios derived from using a ratio of 5:1 for the eight counties as a unit. (See Table 4 under *Growth Ratios*.) Table 7 lists all 18 independent variables used and the eleven chosen for the reduced variable set.

The individual county models were re-run using the new target areas and the previously-generated probability grids, which incorporated the full input variable sets. Upon mosaicking the resulting developed land grids together, it was determined that the goals of eliminating discontinuities across county boundaries and keeping local variables in the modeling process had been satisfactorily realized.

Even using the two-step approach with the logistic regression model, the problem of the new growth following county streets into areas not expected to show significant development, rather than filling in and clustering around the more heavily-developed areas, persisted. It was found that artificially reducing the value of the correlation coefficient for the *distance-to-streets* variable produced favorable results.

The results from the hybridized modeling approach include the final purely logistic regression output for each county, which still has the undesirable spindly growth out the county streets, and the final output for that county, representing a modified logistic regression model with the weight of streets reduced.

Once the modeling for each county was complete, the developed area grids for each time period were extracted from the results. The output grids from the individual counties were then mosaicked together to create the modified logistic regression model predicted developed areas for the full 8-county region for each of 6 years, 2005 – 2030.

The procedure was repeated for each growth ratio; 4:1, 3:1, 2:1, and 1:1.

Table 7:

Full Independent Variable Set (18)	Reduced Independent Variable Set (11) (X indicates inclusion.)
Existing Developed	X
% Available Developed (classified)	X
Distance to Existing Developed (classified)	X
Slope (classified)	X
Distance to Incorporated (classified)	X
Distance to Water (classified)	X
% Available Developed	X
Distance to Schools (classified)	
Distance to Interstate Hwy (classified)	X
Distance to Sewer Lines (classified)	
Distance to County Streets	X
Distance to Water Lines (classified)	X
Population Density	X
Distance to Major Hwy	
Distance to Secondary Roads	
Distance to Highway Nodes	
Distance to Primary Roads	
Cost Distance to Greenville, Anderson, or Spartanburg	

Expert Group Input

The final improvement to be made to the model was the incorporation of input from *Expert Group* information. Representatives from each of the counties in the study area were invited to meet to review the results of the logistic regression model. Participants were encouraged to provide feedback and criticisms of the future predicted developed land maps presented, and then encouraged to provide their own versions of how they anticipated their county developing over the next 23 years. County input varied widely in the amount of information provided and in the detail and quality of information provided. Therefore incorporation of the county input information (*Expert Group Input*) was handled individually for each county.

Representatives from Pickens and Newberry Counties deemed the logistic regression-based maps of 2030 developed land to be realistic and acceptable, and as such chose that it was not necessary to provide further input. No combination with expert group information was necessary and the modified logistic regression output was used as-is for those counties.

In general, each set of Expert Group input was used to create a grid data set of future developed land and then given a temporal component. This was a new approach and an improvement over previous modeling projects, where any expert group map was simply a monolithic time-independent map used to uniformly modify the logistic regression model output. To introduce the temporal component, the expert group map was divided into 6 rings (classified) based on distance from existing (2000) developed land. It would be expected that the closest ring, given a value of 6, would be more likely to develop before the next further ring, value = 5, and so on. (In some cases, fractional values (x.5) were later introduced.) The classified expert group grids for each county were mosaicked to create a single expert classified grid for the 8-county region. Already-developed land (2000) was added to the expert grid and given a value of 7. The values in the expert classified grid were as follows: 7 = already developed in 2000; 6.5 – 1 represent the distance to already developed, where 6.5 was the closest or most likely to develop and 1 was the farthest, or least likely to develop.

The mock temporal maps were then turned into *expert group probability grids* to better facilitate hybridization with the logistic regression model. This was based on the principle that new development is more likely to occur adjacent to or near existing development, and thus the inner ring has the highest probability of becoming developed and the outer ring has the lowest probability. The rings were converted to probabilities between 1 and 0. The expert group probability grid was created by dividing the classified expert grid, with discreet values from 1-7 and 0, by 7. This generated a probability grid containing decimal values between 1 (already developed) and 0 (not developed by 2030).

Combination of the expert group probability grid with the logistic regression probability grid can be achieved in a variety of ways, but a standard weighted approach was chosen for this project. If a 90% logistic regression/10% expert group weighting is desired, for example, the equation used is

$(0.9 * P_l + 0.1 * P_e) = P_w$, where

P_l = probability from logistic regression,

P_e = probability from expert group prediction, and

P_w = 90/10 weighted probability.

The weight given to each county's Expert Group input was determined individually based on the nature of the data provided and the visual appearance of the predicted developed land. Weightings ranging from 10% to 50% expert group were tested. Weightings that were too high resulted in the appearance of harsh straight lines which looked very unnatural and unrealistic. Effort was made to reflect the expert group input in the final result while avoiding such an artificial appearance.

In addition to selecting an optimal weighting for each county, there were several areas of Greenwood and Abbeville Counties that received further attention. County officials strongly believed that there would be heavy development of some lakefront regions. When initial hybridization of the expert group map with the logistic regression model did not generate the anticipated growth, additional modifications were made to the expert group probability grids.

For seven counties the modified version of the logistic regression model, with the weight of the streets artificially reduced, was used for hybridization with the expert group probability grids. However, after viewing the initial logistic regression 2030 map and discussing the otherwise paucity of new development in southern Spartanburg County, a County official claimed "If there was ever a case for spindly growth, it is that area of southern Spartanburg County." Based on this statement, the unaltered logistic regression output was used for combination with the Spartanburg expert group probability grid.

Table 8 summarizes the weight given to the *Expert Group Map* for each county.

Using the expert group data and weighting combinations discussed above, a time-series future developed land prediction was generated for each county at each growth ratio from 5:1 to 1:1. The output grids for each year were extracted from each county and mosaicked together to create an 8-county future developed land prediction for each of the 5-year intervals. This was performed for each growth ratio, producing a total of 30 developed land grids (6 years x 5 ratios). **These are the final output grids for the project.**

Table 8: Weight given to *Expert Group Input* when combined with modified logistic regression output.

County	Weight	notes
Greenville	10%	
Spartanburg	20%	unmodified logistic regression output used
Pickens	0%	*No Expert Group Information*
Anderson	20%	
Laurens	10%	negligible contribution
Newberry	0%	*No Expert Group Information*
Abbeville	20%	Area along Lakes Secession/Russell was increased.
Greenwood	50%	Area along Lake Greenwood was increased.

Results and Discussion

The population of the eight counties in the study area is expected to grow to 1,472,270 by the year 2030. It is anticipated that the amount of developed land will grow to 1,523,667 acres, based on assumption of a 5:1 ratio of developed area growth to population growth. This ratio is believed to be conservative based on the historic trend for the study area and on growth ratios in other areas.

In this project a model was developed combining a modified logistic regression approach with expert group information to predict spatially where the expected development is most likely to occur. Results of the model, both quantitative and spatial, were extracted for every five years from 2005 through 2030 at growth ratios of 5:1, 4:1, 3:1, 2:1, and 1:1.

Spatial results of the modeling process are shown in the enclosed set of map figures. The results from the model run at a 5:1 growth ratio have been mapped at each of the six time points from 2005 through 2030. Maps showing the results obtained using the lower growth ratios have been produced for the years 2015 and 2030. Maps showing the baseline developed areas (1990 and 2000) are included as well.

Table 9 lists the developed area, in acres, for the full eight-county study area, predicted by the final version of the growth model, at each ratio of developed area growth to population growth from 5:1 to 1:1. The 1990 and 2000 developed areas are included also. This is the result from combining the modified logistic regression probability grids with the expert group map probability grids, running each county individually, using the area predictions from the 11-variable, 8-county model output as the input area targets for the time-series. Comparing these 8-county results to the initial target developed areas derived from the growth equation based on forecast population growth and growth ratio, the largest error is -0.05% for the year 2030 at the 5:1 ratio. (1,523,667 acres predicted by the growth equation.*) (*All calculations were performed using cell counts, then converted to acres.)

Table 9: Developed Area Predicted by the Final Model, Alternate Growth Ratios, 8 Upstate Counties

	Developed Area (Acres)				
Year	5:1	4:1	3:1	2:1	1:1
1990	222,745	222,745	222,745	222,745	222,745
2000	576,336	576,336	576,336	576,336	576,336
2005	720,280	691,546	662,702	633,922	605,113
2010	881,919	820,804	759,768	698,629	637,446
2015	1,043,692	950,353	856,758	763,323	669,827
2020	1,205,440	1,079,665	953,973	828,017	702,172
2025	1,367,441	1,209,152	1,050,938	892,736	734,493
2030	1,522,891	1,333,425	1,144,377	954,988	765,739

Breakdown for predicted developed area at the 5:1 growth ratio by county is given in Table 10.

Table 10a: Predicted Developed Area, 5:1 Growth Ratio, By County

Year	Greenville	Spartanburg	Pickens	Anderson	Laurens
1990	52,015	43,456	16,632	49,296	20,913
2000	137,823	130,710	48,335	107,055	51,030
2005	177,115	180,254	60,874	133,757	56,170
2010	203,580	222,957	78,175	169,879	70,792
2015	227,373	256,019	94,883	203,116	91,386
2020	248,476	282,814	111,048	233,013	116,449
2025	268,054	306,831	127,181	258,966	143,164
2030	286,441	328,991	142,937	281,982	168,646

Table 10b: Predicted Developed Area, 5:1 Growth Ratio, By County

Year	Newberry	Abbeville	Greenwood	Total
1990	13,968	11,373	15,092	222,745
2000	35,373	28,297	37,712	576,336
2005	38,507	29,512	44,090	720,280
2010	48,074	35,799	52,663	881,919
2015	62,328	45,721	62,866	1,043,692
2020	79,863	58,477	75,301	1,205,440
2025	99,979	72,790	90,475	1,367,441
2030	120,642	87,259	105,993	1,522,891

In the final version of the growth model, developed land growth due to each county's forecast population growth was not confined to that county. In Table 11 the final figures from the model for each county are compared with what they would have been had the growth been forced to stay within each county's boundaries. Note that the growth ratios for each county would not have been 5:1, rather the overall growth ratio is 5:1. See the previous discussion regarding growth ratios. The deviations listed in the table are not errors, but illustration of the variation allowed by using what is believed to be a more realistic approach in the modeling.

No less than 16 independent variables were employed in the logistic regression model, and others were derived from the basic variables. (Primary highways and highway nodes were derived from the highways data and the innovative variable *percent of available land developed* was derived from the existing developed land data.) Most of the spatial variables entered the model in the form of *distance-to-* grids. It was found that the results of the logistic regression model are overwhelmingly controlled by a small subset of the variables. Generally, the same variables tended to be the most significant in both the 8-county model and the individual county models, although the order varied from county to county when ranked by their correlation coefficients. In particular, the input variables that consistently displayed the highest correlation were *distance to developed* (classified), *slope* (classified), *% available developed* (classified), *distance to incorporated boundaries* (classified), *% available developed* (unclassified), *distance to streets*, and *distance to water lines* and *sewer lines* (classified). Of course *existing developed* always received the highest correlation coefficient because if a cell was developed in 2000 it remained developed in future years. It was interesting to see that the new derived variable, *% of available land developed*, was consistently one of the highest-ranking variables, whether classified or not. *Distance to public schools* was significant (mid to low) in 6 of eight counties, but not in the 8-county model. Contrarily, *distance to Interstate highways* was of mid-level significance in the 8-county model but only surfaced as significant in one individual county model. *Cost distance to Greenville, Anderson, or Spartanburg*, *distance to primary highways*, *distance to major highways*, and *distance to nodes* were consistently insignificant or of very low significance. Interestingly, these variables were never used in a classified form. *Population density* received a negative correlation for six counties and overall, indicating greater population density corresponds to lower probability of becoming developed, but it invariably appeared toward the bottom of the rankings. Upon examining several of the more local variables, *distance to Lake Hartwell* and *distance to Clemson* showed no significance in Anderson or Pickens Counties, the only cases where they might apply. Only *distance to Lake Keowee* showed a significant correlation in Pickens County.

Table 11: 2030 Developed Land (5:1 growth ratio), Final Model vs. Restricting New Growth to Counties

	2030 Developed Land if growth had been restricted to counties			2030 Developed Land, Cross-County Growth Allowed (Final Model)			Deviation by allowing cross-county growth	
County	acres	increase (%)	ratio	acres	increase (%)	ratio	acres	%
Greenville	359,466	160.8	4.29	286,441	107.8	2.88	-73,025	-20.3%
Spartanburg	460,579	252.4	8.14	328,991	151.7	4.89	-131,589	-28.6%
Pickens	146,366	202.8	5.12	142,937	195.7	4.94	-3,429	-2.3%
Anderson	235,201	119.7	4.00	281,982	163.4	5.46	46,781	19.9%
Laurens	109,728	115.0	3.52	168,646	230.5	7.05	58,918	53.7%
Newberry	96,542	172.9	8.36	120,642	241.1	11.65	24,099	25.0%
Abbeville	65,467	131.4	7.44	87,259	208.4	11.79	21,792	33.3%
Greenwood	92,169	144.4	6.43	105,993	181.1	8.06	13,824	15.0%
Total	1,523,314	164.3	5.00	1,522,891	164.2	5.00	-423	0.0%

Development of the Upstate Growth Model involved some trial-and-error experimentation and incorporated some judgment decisions, such as running the logistic regression model for counties individually or as a regional unit, inclusion or exclusion of independent variables, alteration of the weighting of one of the variables, and the weightings assigned to the expert group input. As such, this modeling was very much a hands-on process and probably does not lend itself to being easily portable and generalizable for use by operators without a thorough knowledge of how the model works and familiarity with the region being modeled.

Emphasis should be placed on the importance of input from knowledgeable sources in the community, both in the form of the expert information that can be provided regarding their knowledge of future growth and in their critical assessment of the model output as it is being developed.

The Upstate Growth Model can be used not only to determine where growth is likely to occur, but also what natural and economic resources might potentially be at risk from urbanization.