ARC Population and Employment Allocation Disaggregator - A Simple GIS-based Tool for Parcel-Scale Population Projection

Stephen Bourne, PE [1], Mike Alexander, AICP [2], and Wei Wang [2]

[1]PBS&J, 2198 Goodwood Blvd, Smyrna, GA 30080; PH (404) 895-0753; FAX (770) 426-5316; email: sfbourne@pbsj.com
[2] Atlanta Regional Commission, 20 Courtland St., NE, Atlanta, GA, 30303; PH (404) 463-3100; FAX (404) 463-3205; email: MAlexander@atlantaregional.com, WWang@atlantaregional.com

ABSTRACT

Continued development of urban areas requires careful planning to ensure adequate resources are provided for growing industry and population while sustaining a healthy environment. This paper will discuss the on-going development of a GIS-based software tool, developed by the Atlanta Regional Commission (ARC), for projecting future development of the Atlanta Metropolitan Region. With this tool, the ARC can model growth of projected populations at fine scales, and thereby estimate future demands for transportation infrastructure, water, and other utilities.

The Atlanta metropolitan area is a 22-county region in the northwest corner of Georgia, with a burgeoning population of approximately 5 million. By land area, Atlanta is the fastest growing city in the world with some of the highest commute times in the United States. Continued development strains Atlanta's resources more and more. The ARC is dedicated to unifying the region's collective resources to prepare for a prosperous future. It does so through professional planning initiatives, the provision of objective information and the involvement of the community in collaborative partnerships.

To help adequately plan for future growth, the ARC has developed the ARC Population and Employment Allocation Disaggregation tool, an ESRI ArcMAP extension. The tool employs a two-step disaggregation procedure. Starting with a single region-wide population and employment projection (2000-2050) provided by a Regional Economic Models, INC (REMI) model, the tool first disaggregates this projection to large planning regions called super-districts. The ARC then adjusts the planning level projections manually to match their expectation for growth – and those of Atlanta's stakeholder community; city governments, developers, etc. The tool then disaggregates the superdistrict-scale projections to parcel scale by using map-based factors such as major roads and expressway ramps to estimate likelihood of development and allocating new growth to the most likely areas first. When the projection is at parcel scale, statistics for demands on the travel and utility networks can be calculated and used in forecasting future travel demands on the transportation system. The tool also provides a calibration procedure that compares modeled

growth to actual growth and finds the set of model parameters that minimizes differences between the two.

INTRODUCTION

Continued development of urban areas requires careful planning to ensure adequate resources are available for growing industry and population while sustaining a healthy environment. This paper will discuss the on-going development of a GIS-based software tool, developed by the Atlanta Regional Commission (ARC), for projecting future development of the Atlanta Metropolitan Region. With this tool, the ARC can model growth of projected populations and employment at fine scales, and thereby estimate future demands for transportation infrastructure, water, and other utilities.

The Atlanta metropolitan area is a 20-county region in the northwest corner of Georgia, with a burgeoning population of approximately 5 million. By land area, Atlanta is the fastest growing city in the world with some of the highest commute times in the United States. Continued development strains Atlanta's resources more and more. The ARC is the municipal planning organization (MPO) dedicated to unifying the region's collective resources to prepare for a prosperous future. It does so through professional planning initiatives, the provision of objective information and the involvement of the community in collaborative partnerships.

This paper will discuss on-going development of the ARC Population and Employment Allocation Disaggregator tool in two sections, 1) algorithm description, and 2) technology description.

ALGORITHM DESCRIPTION

Step One: Disaggregation from Region-wide to Superdistrict scale.

The first disaggregation step starts with a regional projection provided by the ARCs Regional Economic Models, Inc (REMI) model, which includes 1) projections for 18 employment sectors, based on the North American Industry Classification System (NAICS) categories, and 2) projections for population by age, where there are 17 age categories. The NAICS employment projections are disaggregated to the 78 superdistricts by forcing the distribution of jobs across the superdistricts to follow the distribution observed in the base year – the first year of the time horizon.

The REMI model provides population projections as 17 population-by-age categories, which are first aggregated to a total population projection. To translate population into land use, the total population is translated to household size/income groups, where there are 6 household sizes (size is number of people in the household) and 4 income categories – giving a total of 24 household size/income categories. The households are also distributed across the superdistricts according to observed distributions in the base year.

Interactive Adjustment of Superdistrict-scale Projections

Once the tool has disaggregated the REMI projections to the superdistrict scale, users can make adjustments to the projections using the tool. The need for manual adjustments comes from 1) the tool assumes the base year distribution of jobs and housing will remain constant at the superdistrict scale over the projection time horizon, which may not be true, and 2) often, the ARC planners and the local stakeholders have accurate information about how development will occur in the short-range horizon (0-10 years) and medium-range horizon (10 -20 years). By allowing for manual adjustments, the tool can ingest this information. In the process, the tool can be used to build consensus on superdistrict-scale growth over the time horizon and can therefore raise confidence in ARC projections.

Of course, REMI-based regional totals of jobs and households – also know as control totals - must be matched. Manually adjusting projections will result in either a net increase or decrease in the total regional number of households or jobs. To conserve the control totals, the tool implements a re-balancing algorithm, which evaluates the total number of jobs or households moved by the user to a superdistrict and then compensates by removing a weighted fraction of jobs or households from the remaining superdistricts such that the total removed equals the total re-allocated.

Step Two: Disaggregation from Superdistrict to Parcel Scale

Once superdistrict scale disaggregation has been completed and approved, the second step of the algorithm is to further disaggregate projected growth to the parcel scale. At this point, the analysis becomes raster-based. In a year-by-year loop, the algorithm allocates projected growth first to employment land uses and then residential land-uses, emulating the actual development process, in which available land is occupied progressively over time. Figure 1 shows a schematic of how the algorithm proceeds through step two of the calculation.

The priority for allocation of new growth is driven by likelihood of development, which is evaluated for each raster cell according to a linear combination of map-based factors. For example, for a given scenario, the user might specify that commercial land use is most likely to develop in close proximity to interstate ramps, major roads, and in neighborhoods where commercial land use already exists. By bringing in map layers of interstates and major roads, the tool can evaluate a likelihood of development raster – called an L-raster. This L raster is calculated as the weighted sum of three rasters: 1) proximity to the closest major road, 2) proximity to closest freeway ramp, and 3) density of commercial land use in a neighborhood of raster cells around each grid cell. Figure 2 shows an example of an L-raster calculation for commercial landuse. The L-raster is the figure on the far right. Areas in blue represent the highest likelihood areas for commercial development to occur. New growth will be allocated to these areas first.

Each map factor is associated with a weight, which gives the importance of one factor relative to the other factors. The weights must add up to 1.0, and can be set by the user. The choice of map factors and their associated weights is decided by the user, making the tool flexible for doing what-if scenarios and identifying the set of factors that are most influential to new growth – see the calibration section of this paper for more detail.

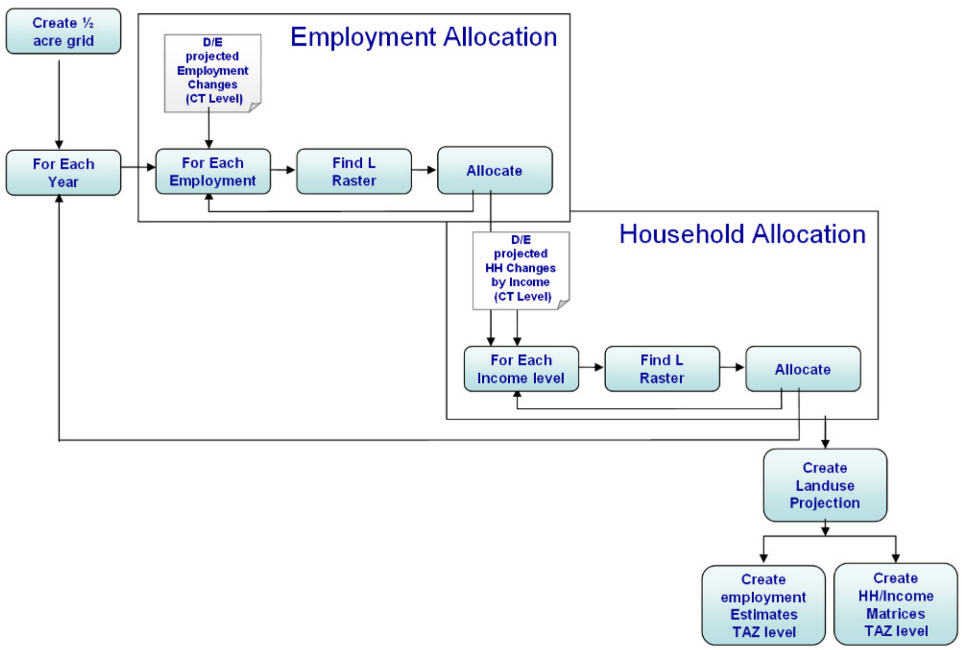

Figure 1: Schematic of Algorithm for Growth Modeling.

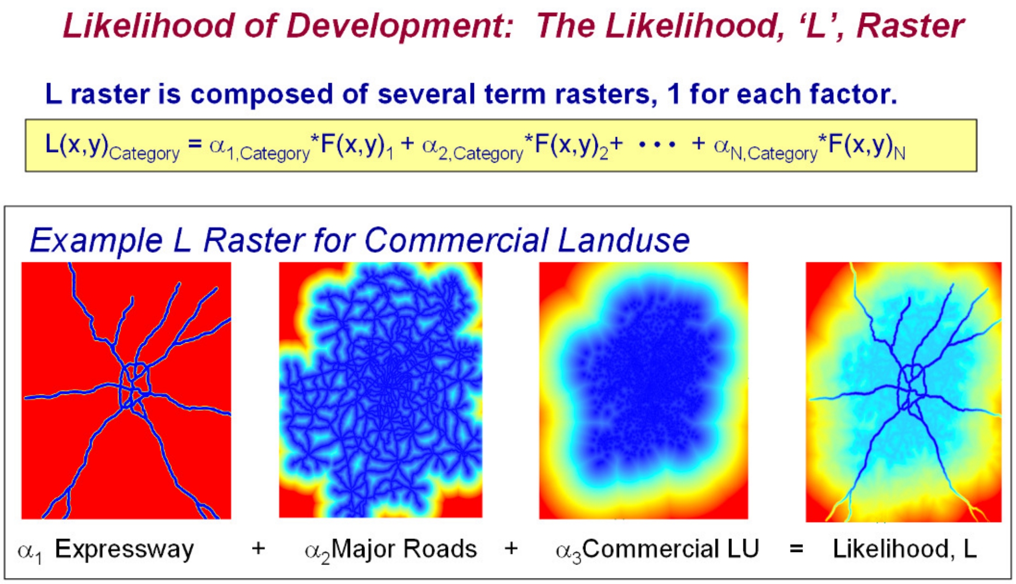

Figure 2: The L-Raster defines the likelihood of commercial or residential development at any given location by creating a weighted sum of user specified factors.

Translation from Employment and Households to Landuse

The superdistrict scale projections come in the form of numbers of new jobs and numbers of new households. As the algorithm allocates new growth by changing land use from vacant to developed, a suitable method for translating employment and households to land use is required. Using an ultimate build-out density map layer, the tool has an estimate at each location of the maximum density of jobs and households. The tool multiplies the number of new jobs or new households for any given year by the ultimate build-out density to arrive at a total number of acres required to contain the jobs or households.

Choosing the Raster Cell Size

In step two, the algorithm is converted from a vector-based to a raster-based calculation, where the vector-based shapes are converted to a grid. Using a raster-based approach is necessary because allocation during the second step must emulate the actual development process, where typically a developer will purchase and divide a large parcel into smaller ones and then develop multiple small parcels at a time in each given year. Converting to a raster-based analysis allows the algorithm to subdivide large parcels into smaller pseudo-parcels, in the form of raster grid cells.

An important aspect of converting to raster-based calculation is to choose the grid cell size of the rasters so that emulated development matches reality. Figure 3 shows the result of an analysis that found the size of actual changes in land use from 1999 to

2001. The rationale for the analysis is that changes in the land use polygons from year to year represent new development. We can set the grid cell size of the raster to match the most likely size of these changes. The implication is that 1) the growth modeling will be more accurate, as the chunks of land allocated to new development will match typical development, and 2) the calculation will proceed more quickly, as the basic grid cell size will be much larger than a single parcel.

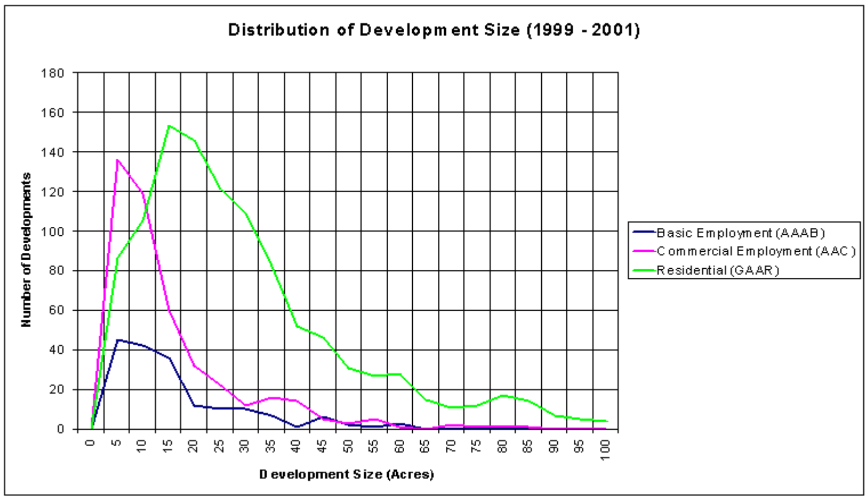

Figure 3: Distribution of Actual Landuse changes from 1999-2001.

For basic and commercial employment and residential landuse polygons, the difference in polygon sizes was calculated by essentially subtracting the 2001 polygons from the 1999 polygons. The distribution of the size of the changes was charted. From the Figure 3 on the preceeding page, we can see that for residential land use, the most likely size of new development is around 15-20 acres. Smaller average sizes for basic and commercial employment land use are found, 5-10 acres respectively. The grid cells size can be specified by the user in the ARC Population Disaggregation tool. The typical size used in our calculations has ranged from 500 ft (5.7 acres) to 1000 ft (23 acres).

Calculating Travel Demand Statistics

Once the algorithm has completed the two-step disaggregation process, a post-processing step is required to evaluate travel demand statistics for use in the ARC's travel demand model. The travel demand model requires employment and population estimates at the traffic analysis zone (TAZ) scale. The U.S. Census defines a TAZ as a special-purpose geographic entity delineated by state and local transportation officials for tabulating traffic related data from the decennial census, especially journey-to-work and place-of-work statistics. In the 13-county Atlanta metropolitan area, there are 1767 TAZs, with an average size of 1688 acres.

The tool calculates TAZ-scale totals for employment and households by income by translating the projected land use in each TAZ back to the NAICS employment and 24 category household size/income categories, ensuring that the distribution of jobs and housing matches that of the superdistrict.

Similar methods for aggregation of results to find demands for water, electricity, gas, and other resources can be added to the tool in the future.

Scenario-based Growth Modeling

In addition to the basic input data (map layers, REMI input, etc.), the user must specify the following to complete a run of the algorithm.

1. The time horizon of the analysis
2. The adjustments to superdistrict-scale projections
3. The map factors that influence development at the parcel scale for basic employment, commercial employment, and residential land uses.
4. The weights for the influential map factors

Collectively, this set of user specifications can be called a growth scenario. The ARC uses the tool to create multiple scenarios and compare the results. The intent can be to evaluate sensitivities to a certain map factor (eg. the effect of major roads), the effect of changing superdistrict scale projections to implement higher growth trends in one area versus another (eg. west Gwinnett County growing faster than east Gwinnett), and so on.

Calibration

An important question to answer in the course of growth modeling is whether the map factors chosen to define likelihood of development, and their respective weights, are reasonable. The tool implements a calibration algorithm to answer this question. The calibration process essentially consists of progressively adjusting weights of the factors until the difference in modeled and actual growth during an historic period is minimized.

The calibration algorithm requires a start year land use layer and end year land use layer. It calculates actual changes in land use by first converting the land use layers to raster form and then evaluating the number of grid cells that changed from vacant land to developed land. Then, the calibration procedure creates a set of possible weights for each map factor and runs the superdistrict to parcel scale disaggregation for every combination of those weights, evaluating the difference in modeled real land use change for each weight set.

The scale at which the comparison is made is dependent on the application. Currently, the comparison of actual and modeled growth occurs at the TAZ-scale, as the tool is being used for travel demand modeling. That is, the distribution of actual and modeled land use change is calculated for each TAZ. The region-wide comparison is made by finding the sum of squared errors (SSE) (ie. sum of differences between modeled and actual land use change) across all the TAZs for each of the three land use categories.

Simplex-based Calibration

The calibration method described above, where all combinations of possible weights are tested, is called the brute force method. Given it takes approximately 5 minutes to run a single growth modeling scenario on the ARC workstations, if there are as little as 3 map factors and the weight range is discretized into [0, 0.2, 0.4, 0.6, 0.8, 1.0], or six possible values, the number of runs of the scenario would equal $6^3 = 216$, which would require 18 hours to run. Given many more factors would normally be influential, and a finer discretization is preferred, a method for more quickly identifying the best set of weights is required.

The simplex calibration method is used to more quickly arrive at the best set of weights. Figure 4 illustrates the procedure. In the figure, 3 map factors are under consideration. The surface in the figure represents all the possible sets of weights that can be evaluated. Each black point on the surface represents a single set of weights. A triangle of neighboring points on the surface is called a simplex. First, the procedure evaluates the SSE for the 10, 8, 9 simplex. Point 10 - [0,1,0] results in the worst match between actual and modeled land use (largest SSE). The procedure then "flips" the triangle about the axis created by points 8 and 9, so that a new simplex is created of points 8, 9, and 6, and the SSE for point 6 is evaluated. The largest SSE is found to be point 9 for this simplex, and the triangle is again flipped to make a new simplex of points 8, 6, and 7. The procedure continues until a simplex is found where there is no option to flip the simplex triangle and get a lower SSE. The point with the lowest SSE is thus identified as the best possible set of weights.

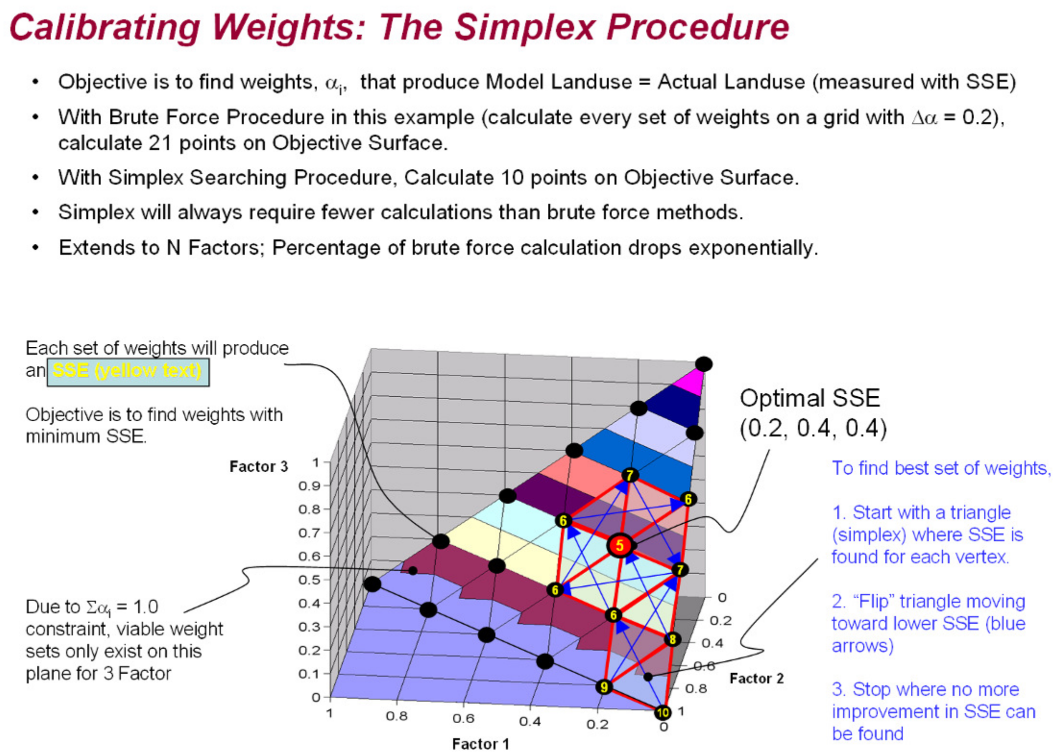

Figure 4: The simplex calibration procedure.

The simplex procedure reduces the calculation time substantially; savings increase as the number of map factors increases and the discretization of the weights becomes finer. A concern with using the simplex method is entering local minima in the SSE surface and not finding the global minimum as a result. To counter this, the algorithm is started from multiple simplexes. Moreover, once a minimum is found in the surface, the procedure "jumps" out of the minimum by moving randomly to a neighboring simplex and re-starting the search from this new point.

The calibration procedure is a powerful way of understanding which map factors are important in terms of new growth and which can be left out of the analysis. Moreover, the procedure provides some insight into the relative importance of each map factor. Finally, the calibration procedure provides some indication of how accurate the super-district to parcel scale disaggregation actually is. By comparing modeled growth to real growth, the ARC can understand how close these can actually come to projecting growth accurately, and conversely how much of the growth is not explainable with the growth model.

TECHNOLOGY DESCRIPTION

The ARC Population Disaggregation tool was created as an ESRI ArcMap 9.2 extension, so that it would work with the standard GIS platform used at ARC, the platform the majority of ARC staff are trained in and use on a regular basis. Figure 5 shows a screen shot of the tool. The major components of the tool are a windows explorer-like tree that describes the growth scenario and allows the user to enter inputs and run the model, a context-sensitive information panel in the center of the screen that responds to the user clicking nodes in tree with information about the node and options for what you can do with the node, and a map on the far right. The map is the standard ArcMap map, and can be manipulated just as the standard map can.

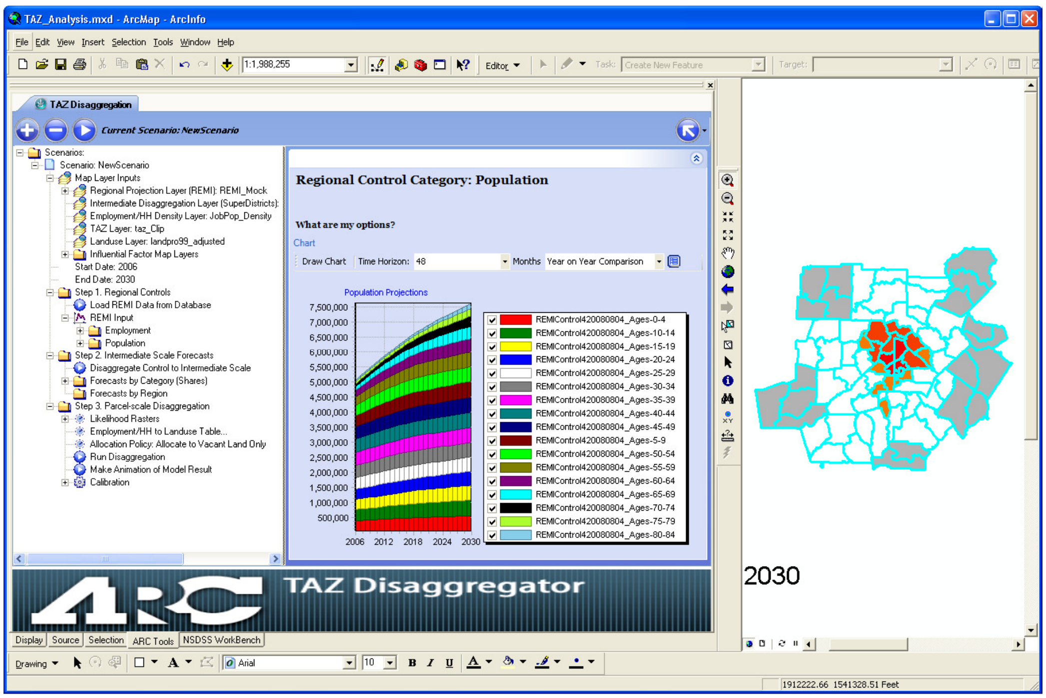

Figure 5: Screen shot of the ARC Population Disaggregator Tool.

Figure 6 shows a screenshot of the tree for a growth scenario. The tree is broken into four major sections 1) map layer inputs, 2) Step 1: Regional Controls, 3) Step 2: disaggregating from regional scale to superdistrict scale, and 4) Step 3: disaggregating from superdistrict scale to parcel scale. The user can create scenarios by clicking the scenarios folder at the top of the tree and clicking the "+" button on the toolbar. Once scenario is created, the user specifies which layers in the map should be considered the map layers to be used in the analysis (REMI layer, superdistrict layer, TAZ layer, etc.) and specifies the collection of map factor layers. The user can also enter the start and end date of the time horizon in this section of the tree.

Once the input data has been specified, the user can load the control data into the tool by clicking on the ***Load REMI Data from database*** node in the tree and choosing the load data option in the information panel. REMI data for the employment and population categories is loaded for the specified time horizon and can be charted as stacked time series charts (see figure 5). In step 2, the user then disaggregates the regional control time series to the super districts. The results are displayed in a table in the information panel by category and by superdistrict. The user can adjust the projections directly within these tables, or they can select the superdistricts on the map and adjust the projections up or down.

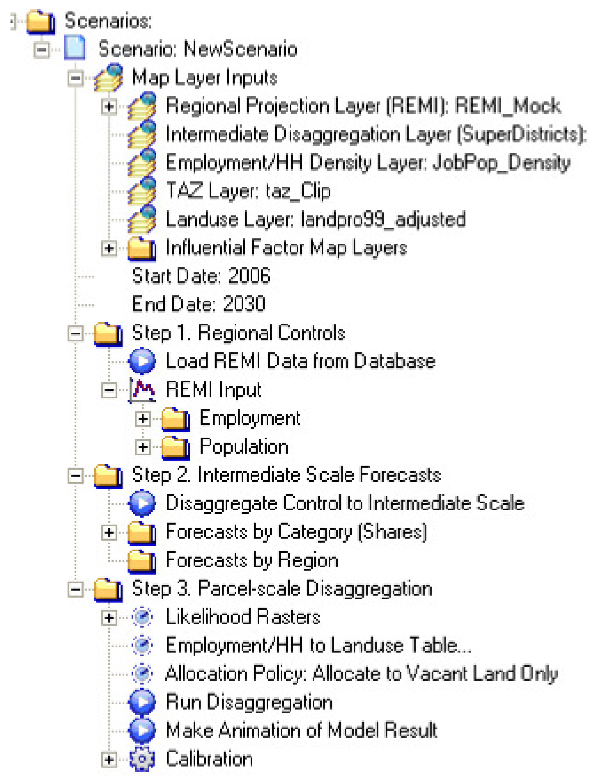

Figure 6: Scenarios for Growth Modeling are created in the Tree.

Finally, the user can run the disaggregation to parcel scale using step 3 in the tree. The user can set the weights for each factor in this part of the tree and set other modeling parameters. This disaggregation step is the most computationally intensive part of the algorithm and takes roughly 5 minutes on ARC workstations. The calibration procedure can also be run from this part of the tree if the user desires. Finally, an option for creating an .avi animation of the growth over the projection time horizon is included.

CONCLUSION

The ARC Population and Employment Allocation Disaggregation Tool is a powerful and flexible tool for GIS-based population projection. Using a combination of disaggregation steps that focus on 1) including the stakeholders in the projection process to build consensus, and 2) increasing accuracy in the modeling exercise through expert advice and calibration to historical pattern, the tool provides new insight into growth patterns in the Atlanta Metropolitan region.