



SMALL AREA SOCIO-ECONOMIC FORECAST TRAFFIC ANALYSIS ZONE DISAGGREGATOR DOCUMENTATION

Prepared By:
The ARC Research Division Staff and Stephen Bourne, PBS&J

Table of Contents

PURPOSE	3
METHODOLOGY	3
STEP ONE: DISAGGREGATION FROM REGION-WIDE TO SUPERDISTRICT SCALE	3
Interactive Adjustment of Superdistrict-scale Projections	4
STEP TWO: DISAGGREGATION FROM SUPERDISTRICT TO PARCEL SCALE.....	5
Translation from Employment and Households to Landuse.....	6
Choosing the Raster Cell Size	7
Calculating Travel Demand Statistics	8
SCENARIO-BASED GROWTH MODELING	9
CALIBRATION.....	9
SIMPLEX-BASED CALIBRATION	10
TECHNOLOGY DESCRIPTION	11
FORECAST DEVELOPMENT PROCESS	13
INTRODUCTION.....	13
SYSTEM REQUIREMENTS.....	14
INSTALLATION PROCEDURE.....	15
THE TAZ DISAGGREGATOR IS INSTALLED IN ARCMAP	17
UNINSTALLING THE TAZ DISAGGREGATOR	18
THE TAZ DISAGGREGATOR TOOL.....	19
THE TAZ DISAGGREGATOR TOOL PANEL.....	19
THEMXD FILE - THE BASIS FOR A TAZ DISAGGREGATION ANALYSIS	19
A TREE-BASED DESIGN	20
BUILDING A DISAGGREGATION SCENARIO.....	22
Step 1. Setting Map Input Layers and Projection Information	22
Step 2. Adding Proximity Factor Layers	25

Step 3. Setting the Likelihood Raster Weights	27
Step 4. Setting the Employment/Household to Land Use Map	29
Step 5. Setting the Allocation Policy	31
Step 6. Run the Analysis	32
CALIBRATION	34
SET MAP INPUT LAYERS.....	36
SET THE WEIGHT RESOLUTION	38
RUN THE CALIBRATION PROCESS	39
CALIBRATION RESULTS	40
UPDATING THE SCENARIO WITH THE MINIMUM SSE WEIGHT SET.....	42
CONCLUSION	43

PURPOSE

To complete the Small Area Socio-Economic Forecast, the ARC developed the ARC Population and Employment Allocation Disaggregation tool (TAZD), which is run as a tool extension in ESRI's ArcMap software. The purpose of the tool is to disaggregate the Regional Forecast to Traffic Analysis Zones for input into the Travel Demand Model. To do this, the TAZD employs a two-step disaggregation procedure. Starting with the single region-wide population and employment forecast (2020-2040) developed with the Regional Economic Models, INC (REMI) Policy Insight (PI) model, the tool first disaggregates this projection to large planning regions called super-districts, which are agglomerations of year 2000 Census Tracts sized to represent sub-county economic markets. 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 the parcel scale by using spatial factors such as major roads, expressway intersections, like land use, and environmentally sensitive areas to estimate the 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.

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

METHODOLOGY

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 Superdistrict by forcing the distribution of jobs across the Superdistrict 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 referred to 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 Superdistrict such that the total removed equals the total re-allocated.

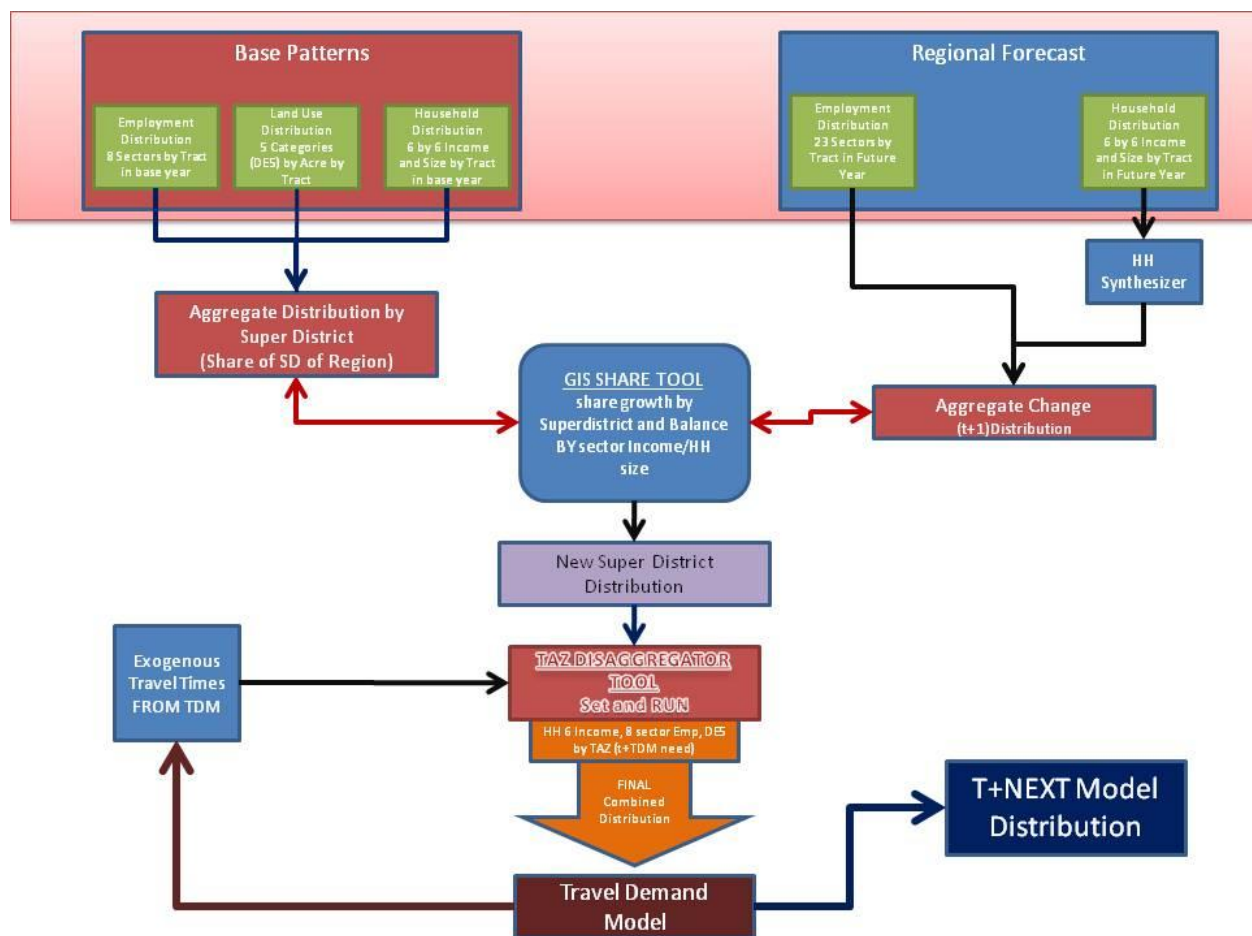


Figure 1

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.

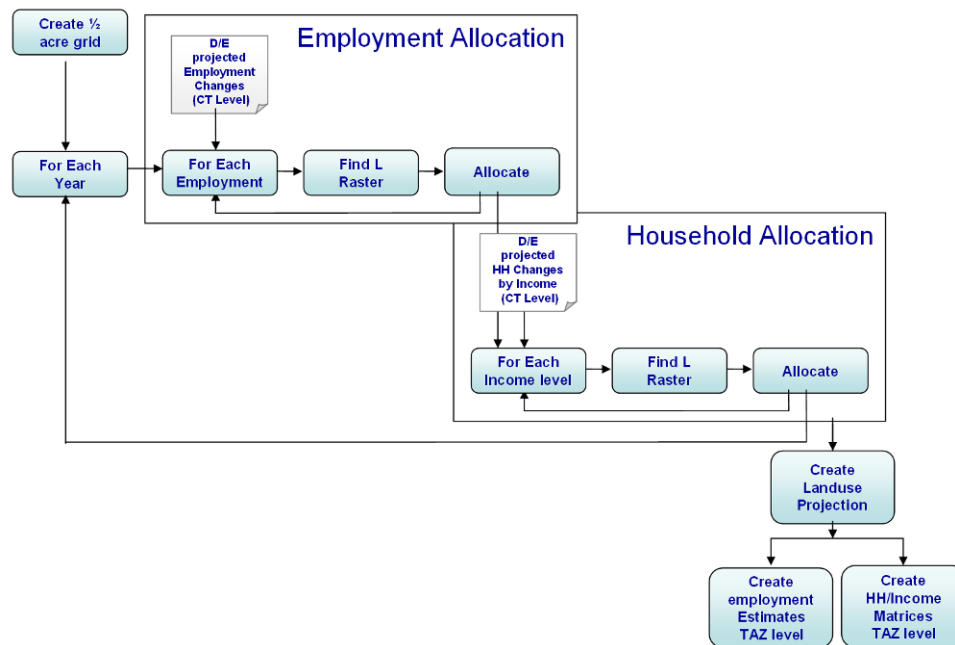


Figure 2

Figure 2: Schematic of Algorithm for Growth Modeling.

Likelihood of Development: The Likelihood, 'L', Raster

L raster is composed of several term rasters, 1 for each factor.

$$L(x,y)_{\text{Category}} = \alpha_{1,\text{Category}} * F(x,y)_1 + \alpha_{2,\text{Category}} * F(x,y)_2 + \dots + \alpha_{N,\text{Category}} * F(x,y)_N$$

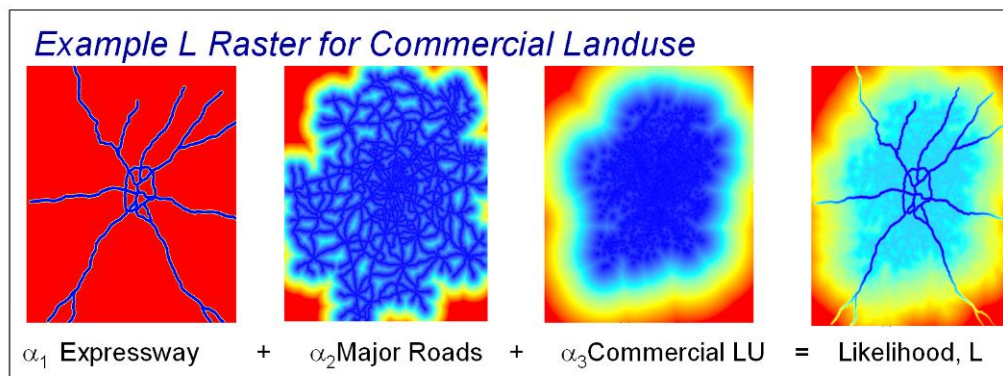


Figure 3

Figure 3: 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.

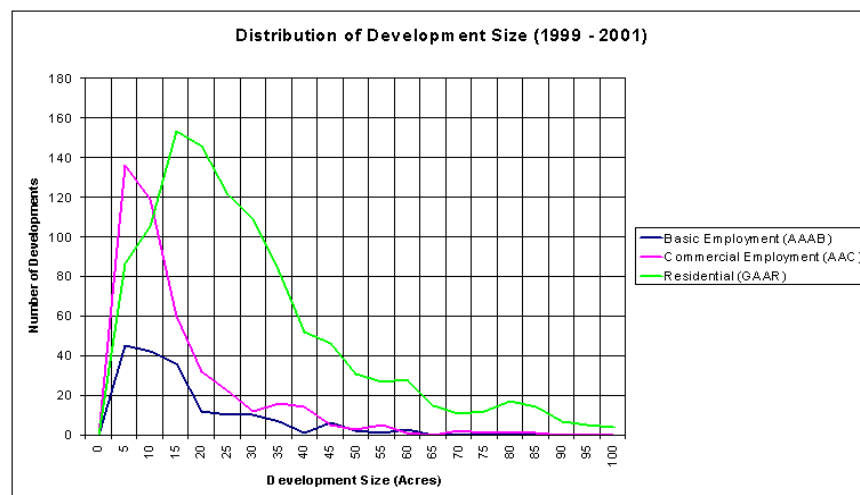
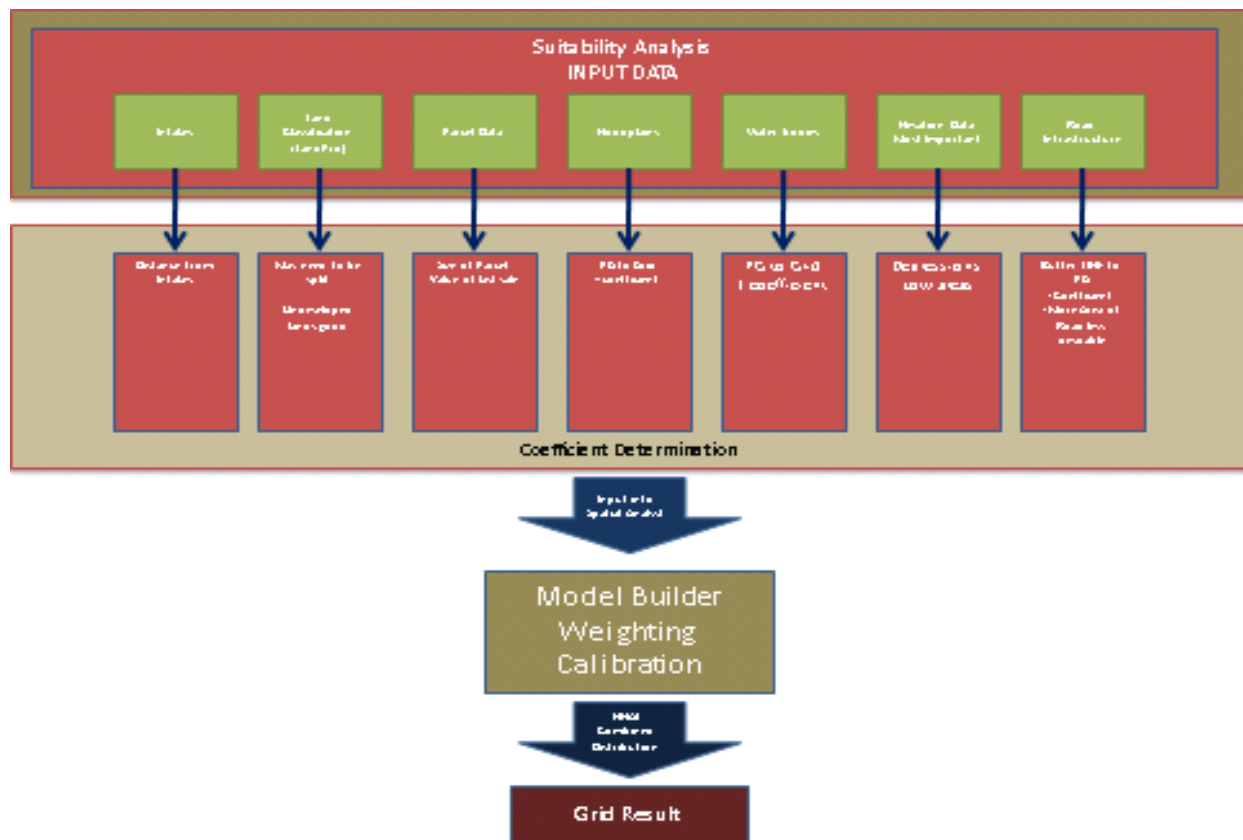


Figure 4

Figure 4: 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

preceding 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 (e.g. 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) (i.e. sum of differences between modeled and actual land use change) across all the TAZs for each of the three land use categories.

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.

Calibrating Weights: The Simplex Procedure

- Objective is to find weights, α_i , that produce Model Landuse = Actual Landuse (measured with SSE)
- With Brute Force Procedure in this example (calculate every set of weights on a grid with $\Delta\alpha = 0.2$), calculate 21 points on Objective Surface.
- With Simplex Searching Procedure, Calculate 10 points on Objective Surface.
- Simplex will always require fewer calculations than brute force methods.
- Extends to N Factors; Percentage of brute force calculation drops exponentially.

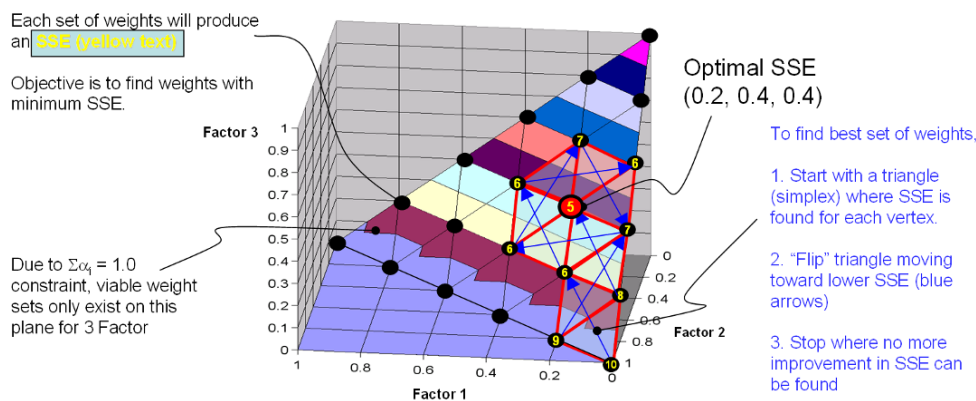


Figure 4: The simplex calibration procedure.

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.

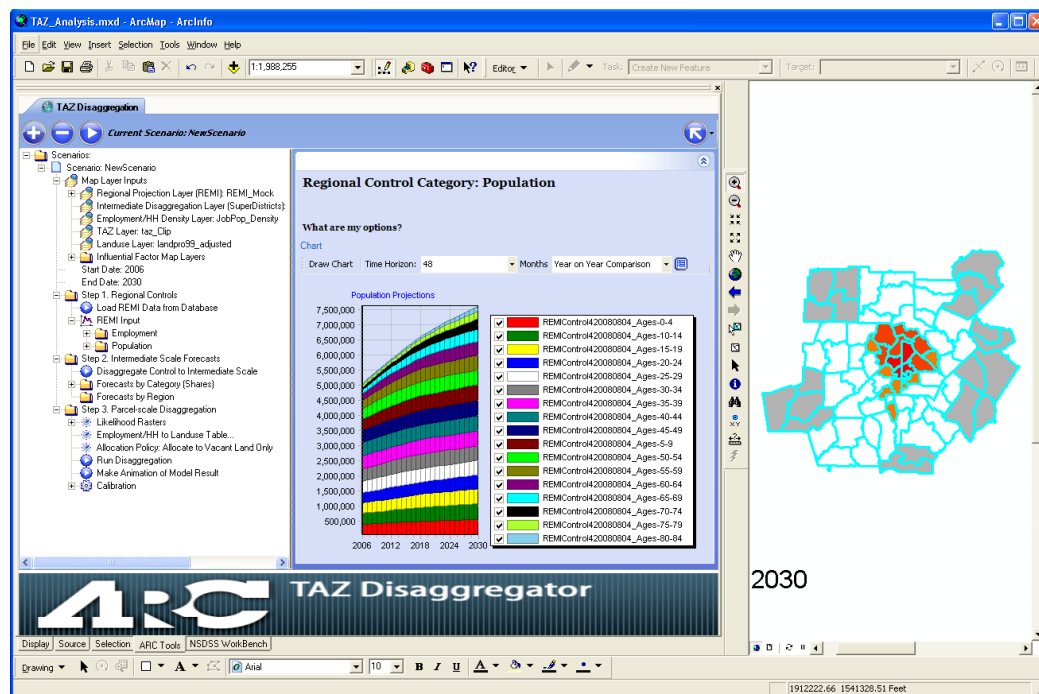


Figure 5: Screen shot of the ARC Population Disaggregator Tool.

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 Superdistrict on the map and adjust the projections up or down.

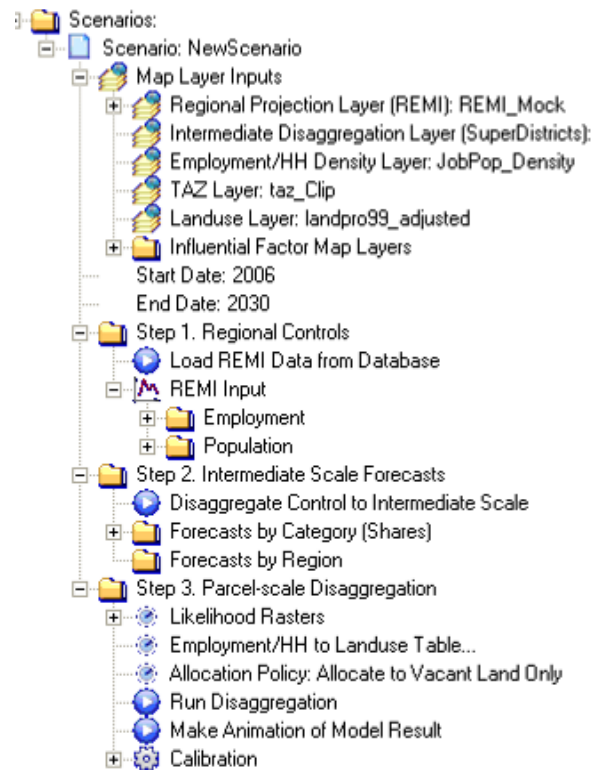


Figure 6: Scenarios for Growth Modeling are created in the Tree.

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.

FORECAST DEVELOPMENT PROCESS

INTRODUCTION

This user manual will describe the use of the TAZ Disaggregator in detail. An online version of this manual is accessible from the tools.

For more information about the intent, design process, database design, and software design of the TAZ Disaggregator, please refer to the following documents:

TAZ Disaggregator Conceptual Presentation

TAZ Disaggregator Requirements and Specifications Documents

These documents can be obtained through the Atlanta Regional Commission Planning Office.

The TAZ Disaggregator must be installed with the installation program called ARCTools.msi that is included in the installation CD or that you may have downloaded.

This chapter describes the steps necessary for installing the tools.

SYSTEM REQUIREMENTS

As the TAZ Disaggregator will be installed as an extension to ArcMap, its system requirements are the same as ArcGIS system requirements. The system you are installing the TAZ Disaggregator on must have the following requirements:

Hardware:

CPU Speed:

1.0 GHz recommended or higher

Processor:

Intel Pentium or Intel Xeon Processors

Memory/RAM:

512 MB minimum, 1 GB recommended or higher

Display Color:

Greater than 256 color depth

Swap Space:

300 MB minimum

Disk Space:

Typical 765 MB NTFS, Complete 1040 MB NTFS

Disk Space Requirements:

In addition, up to 50 MB of disk space maybe needed in the Windows System directory (typically C:\Windows\System32). You can view the disk space requirement for each of the 9.1 components in the Setup program.

Operating System:

Recommended:

Microsoft Windows XP Professional Edition

Optional:

Microsoft Windows XP Home Edition

Microsoft Windows 2000 Professional Edition

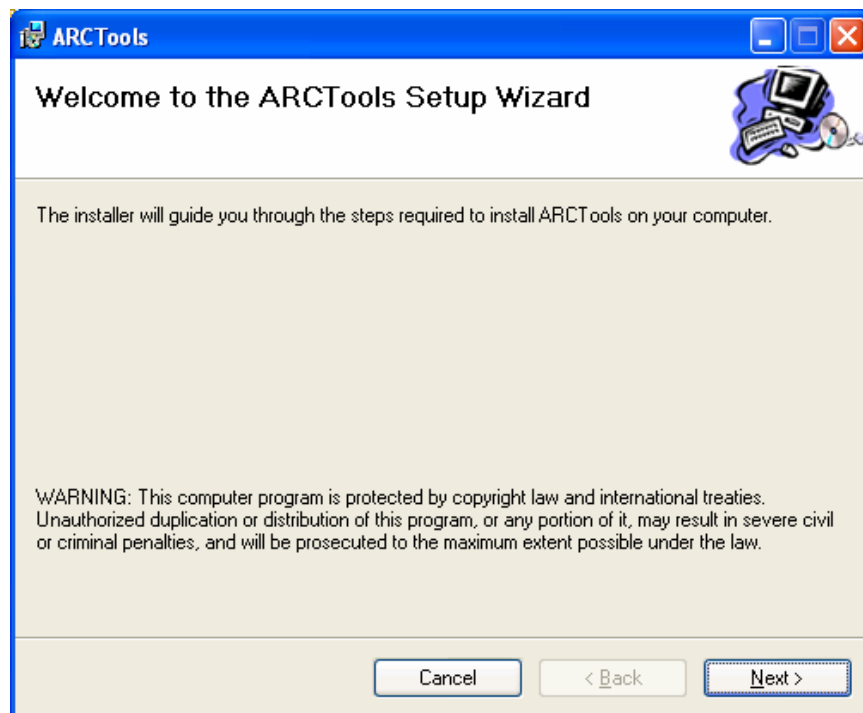
Software:

GIS:

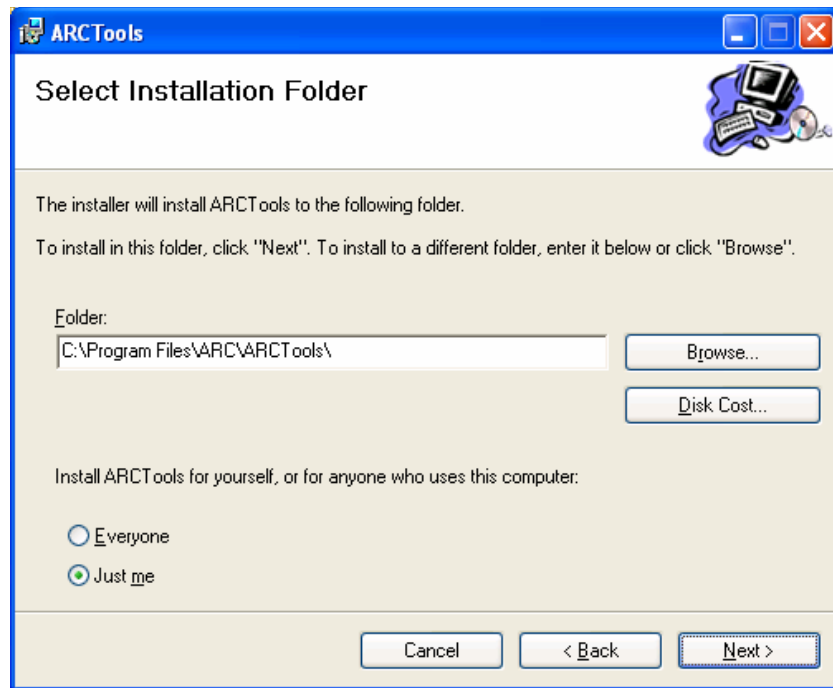
ArcGIS version 9.0 or higher, Minimum ArcView Version
Spatial Analyst Extension

INSTALLATION PROCEDURE

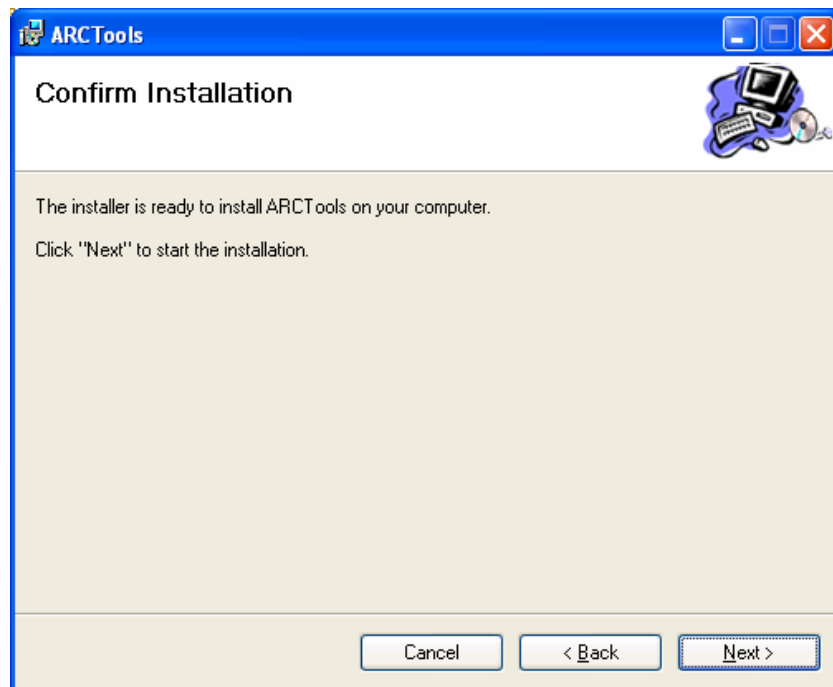
Step 1: To start the installation procedure, either insert the installation CD and the installation will begin automatically or find the ARCTools.msi file in Microsoft Windows Explorer and double-click it. The welcome screen will appear as seen below. Click Next to proceed.



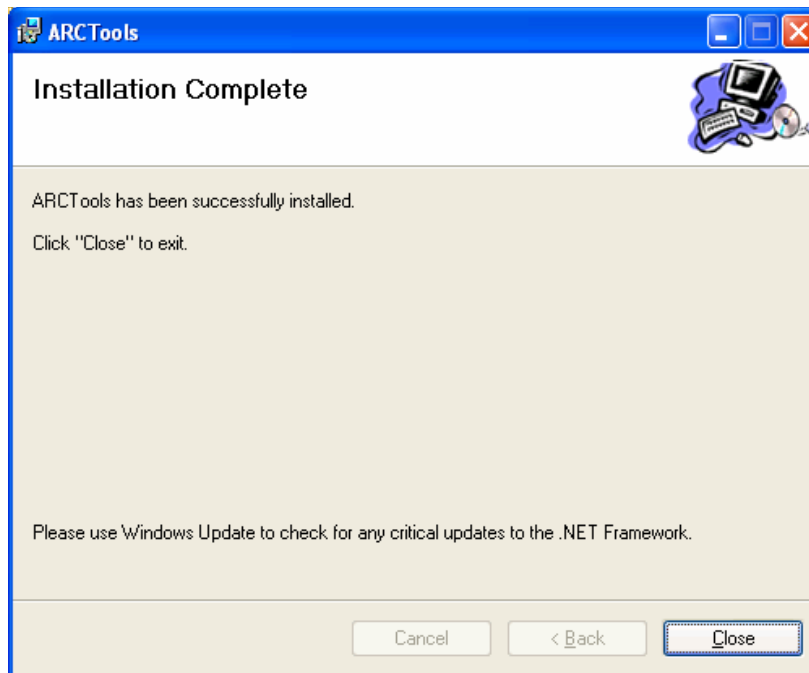
Step 2: The next screen will prompt you for an installation folder. The default folder is C:\Program Files\ARC\ARCTools\ as shown below. Select a different folder if preferred. Select to install the tool for Just me (you can install for Everyone if you want other users of the installation machine to use the tool).



Step 3: The next screen will ask to confirm that your installation preferences are correct. Select Next to install.

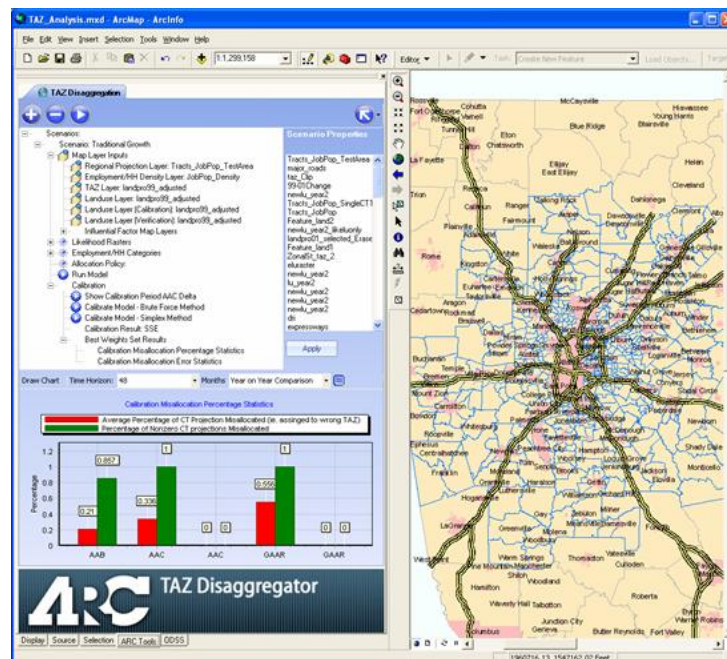


Step 4: If installation is successful, the next screen will inform you. Click Close to Exit.

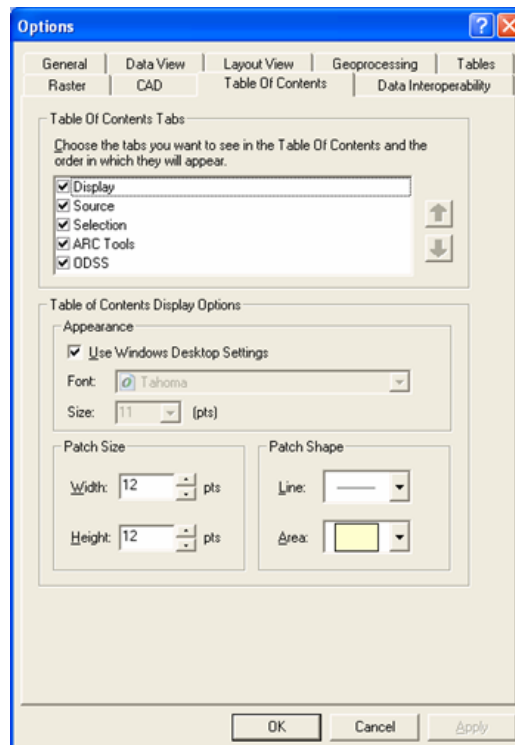


THE TAZ DISAGGREGATOR IS INSTALLED IN ARCMAP

To view the TAZ Disaggregator, open ArcMap. The ARC tools should appear as a table of contents tab as shown below.

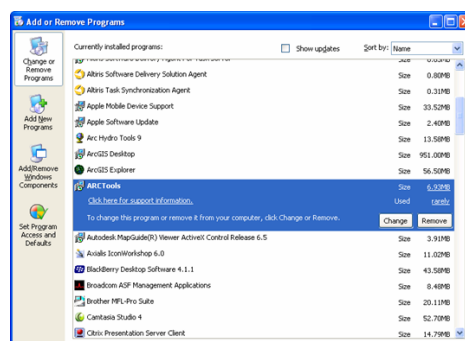


If you don't see the table of contents tab, go to Tools | Options menu item in ArcMap. Select the Table of Contents tab in the Options window. ARC Tools should be one of the items in the Table of Contents Tabs list. Make sure it is selected.



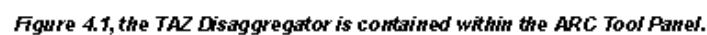
UNINSTALLING THE TAZ DISAGGREGATOR

In some cases, you may want to uninstall the tools. To uninstall, first make sure ArcMap is not running. Then, go to Start | Settings | Control Panel. Select Add or Remove Programs. The Add or Remove Programs window will appear as below. Select the ARCTools entry in the list and then click Remove. Follow the steps to remove the tools.



This chapter will describe the TAZ Disaggregator tool from a functional perspective. Each part of the tool will be described and its use illustrated. The next chapter provides a tutorial which shows through an example analysis how to use the tool.

The TAZ Disaggregator Tool resides as a tool panel inside ArcMap. To access the tools, click the ARC Tools Tab at the bottom of the Table of Contents. The panel is designed to contain all the tools you will need to use the TAZ Disaggregator. A full knowledge of ArcGIS is beneficial when using the tools, but not necessary.



To use the tools, it is necessary to have an MXD open in ArcMap. The MXD must be loaded with the necessary input layers including:

- Page 19

3. A polygonal **Density Layer** that gives the density of employment and households.
 - The attribute field that contains density for employment should be called “Density” and should contain number of jobs per acre.
 - The attribute field that contains density for households should be called “Density_HH” and should contain number of households per acre.
4. A **Landuse layer** that contains an integer attribute called DE5_Int that has the following values:
 - AAAB - Basic Employment: Index = 1
 - AAC - Commerical Employment: Index = 2
 - AU - Unusable Land: Index = 3
 - GAAR – Residential: Index = 4
 - VAC - Vacant Land: Index = 5

For calibration, the MXD must contain calibration and verification land use layers that have the same format as the standard land use layer.

Proximity factor features are also contained in map layers. At any time, the user can add a layer to the map and include this layer as a proximity factor. The features can be points, lines, or polygons.

The MXD file saves the various parts of the disaggregation analysis including the input map layers, generated rasters and map layers, and the input settings and chart results. Saving the MXD saves all this information.

A TREE-BASED DESIGN

The TAZ Disaggregator is centered on a windows explorer-like tree called the Disaggregator tree, which is used to build, run, and view the results of disaggregation analyses. The elements in the tree are referred to as treenodes, or just nodes.

Each disaggregation analysis is termed a scenario. A user may create various scenarios. In each scenario, the proximity factors that determine the likelihood of development can be weighted differently based on the mode of development that scenario emulates. Further, in each scenario, the way in which allocation of projected growth occurs can be set differently. For example, a traditional growth scenario might place highest weight on proximity to roadways and only allocated new development to vacant land, while a re-development oriented scenario might focus more on infill.

The properties of the scenario are organized in the tree. As shown in figure 4.2, there are various elements of the tool that support the tree - including the property list, map, toolbar, and chart.

The usage for each of these elements will be described later in this chapter and in the tutorial chapter.

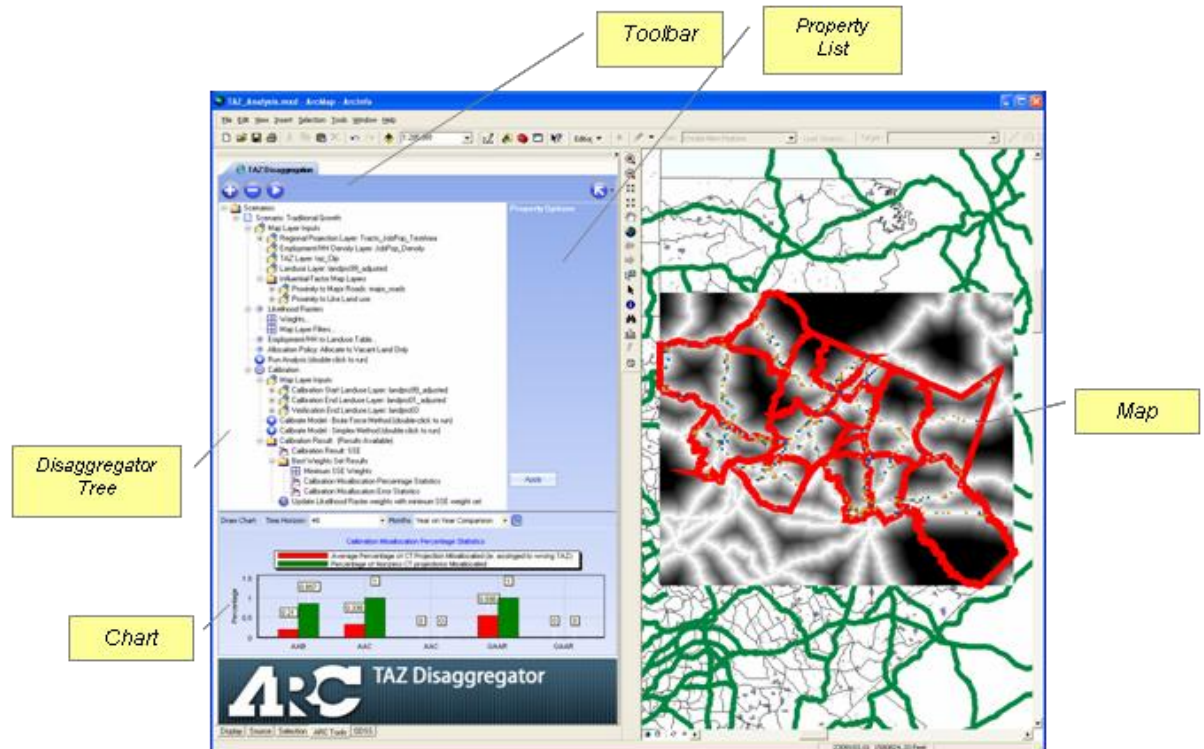


Figure 4.2, the TAZ Disaggregator is centered on the Disaggregator tree where scenarios are built and run. The toolbar, property list, map, and chart support input specification and input and result visualization.

BUILDING A DISAGGREGATION SCENARIO

The Disaggregation Tree contains scenarios as the root nodes (see figure 4.3). There can be several scenarios in one MXD.

Each scenario consists of map layer inputs, likelihood raster inputs (the weights for proximity factors), a table that maps employment/HH categories to land uses, and an allocation policy.

The general process for setting up and running a scenario is:

1. Set the map layers and projection information.
2. Add the proximity factor layers,
3. Set the likelihood raster weights for each land use.
4. Set the Employment/HH to Landuse Map.
5. Set the allocation policy,
6. Run the Analysis.

The disaggregation tree lists the inputs for the scenario in the order of the process steps. Each of these steps will be described in detail in the next few pages.

Calibration is an optional process that can be conducted for a scenario. The aim of calibration is to find the set of likelihood raster weights that result in modeled land development that best matches actual land development. The calibration process will be described after the general process steps.

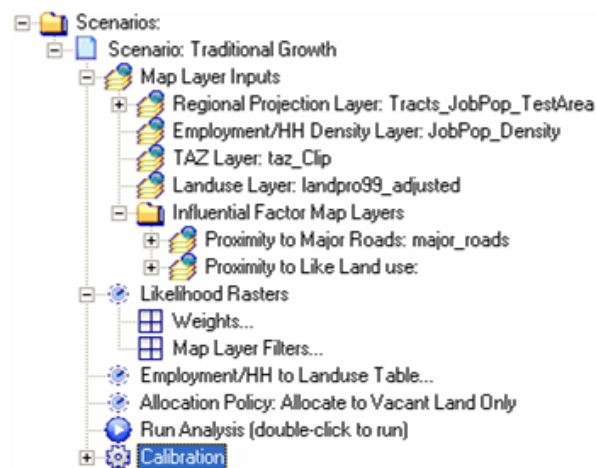


Figure 4.3. A disaggregation scenario is described in the tree. The components are the map layer inputs and projection information, the likelihood raster setting, the employment/HH to landuse table, and the allocation policy.

Step 1. Setting Map Input Layers and Projection Information

Map Layer Inputs (see figure 4.4)

To set map layers as inputs:

1. Click on the map layer you want to set in the tree (these are the nodes in the tree with the map layer icons).
2. On clicking, the properties list will be populated with all the layers in the MXD. Double-click (or click the layer and click the Apply button) the layer in the properties list that you'd like to use.
3. The tree will be updated with the layer you have selected.

Projection Attribute Fields (see figure 4.5)

The projection layer also contains the DRAM-EMPAL projections for each of the various employment and household (HH) categories. You will need to map the categories to the field in the projection layer. To do this:

1. Drill down in to the projection layer node in the tree. Note that there is a node for each of the employment and HH categories.
2. Click on a category. The Property Options list will be populated with the fields of the projection map layer.
3. Double-click on the name of the field that holds the projects for the selected category.
4. **Make sure to save the MXD to save your selections.**

Projection Horizon

The projection horizon is also set under the regional projection layer node. Set the horizon by clicking on the projection horizon node and entering the number of years in the projection horizon. Click the Apply button to update the projection horizon. A default horizon of 5 years is used if no horizon is set.

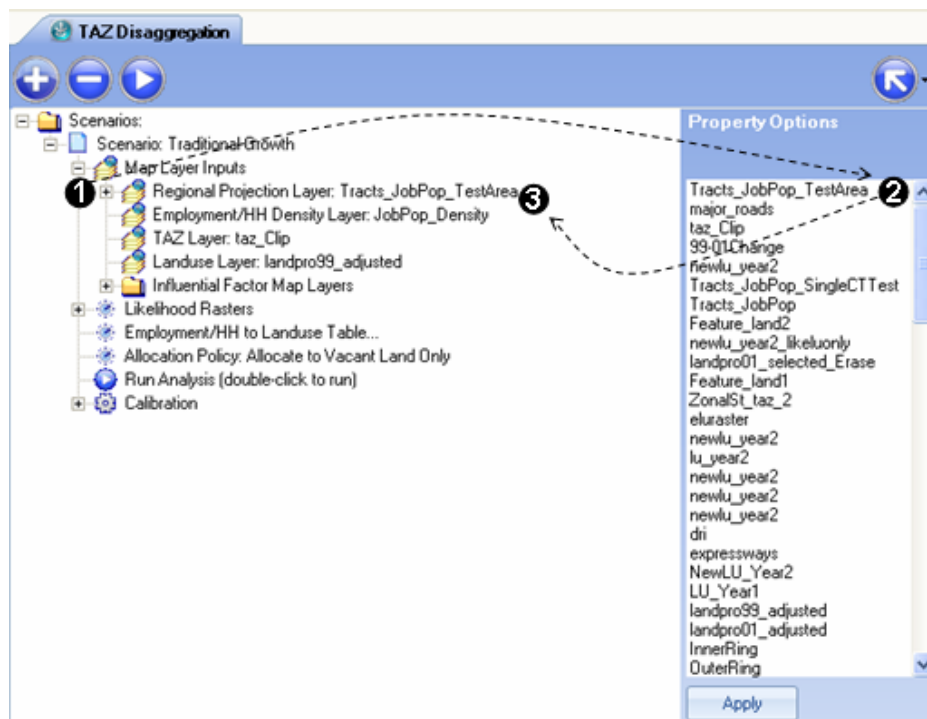


Figure 4.4, setting map layers is done by clicking the input property on the tree and then selecting the map layer to use from the properties list.

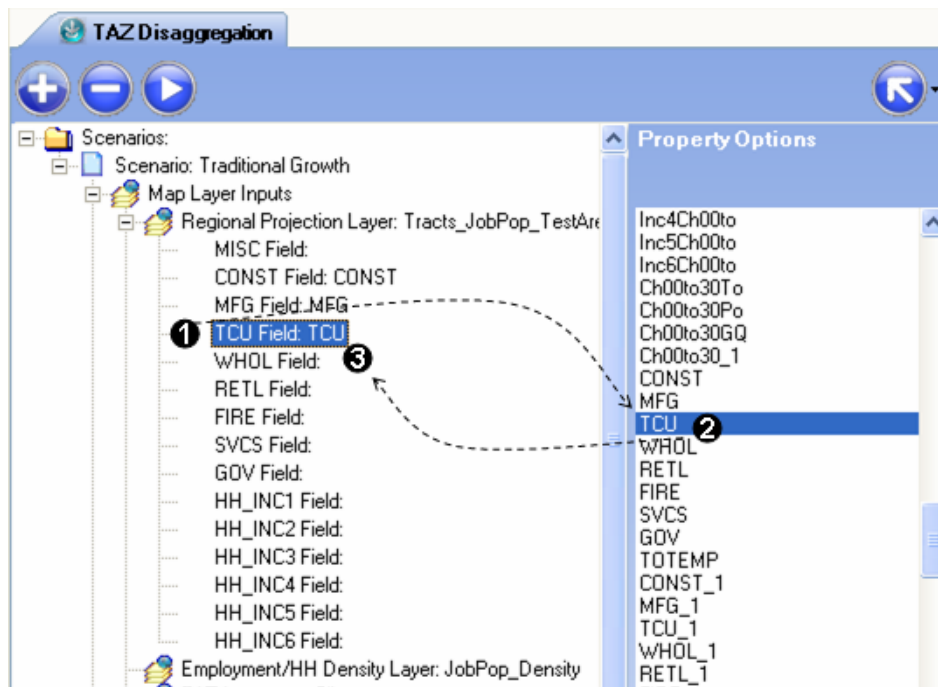


Figure 4.5. The projection layer should contain regional (census tract) projections for the various employment and household categories in the attribute fields. Map the attribute fields to the categories by selecting the category in the tree and then selecting the field in the property options list.

Step 2. Adding Proximity Factor Layers

To **add** a proximity Factor: (see figure 4.6)

1. Select the “Influential Factor Map Layers” node in the tree,
2. Click on the Add button in the toolbar,
3. A window will pop up asking for the name of the new factor, enter a name and click OK,
4. A window will pop up asking for the key of the factor. This is an up-to eight digit alphanumeric key that will be used by the tool internally to create raster layer names. Enter the key and click OK.
5. Set the Default Weight for the new factor. This weight is the weight that will be used when this factor is included in a likelihood raster. The user is free to set this weight as something other than the default in the likelihood raster part of the tree, but a default value is needed. To set the weight, click on the Default Weight property in the tree for the new factor. Type in the weight and click the Apply button. The weight will be updated in the tree.
6. The features in a proximity layer can optionally be filtered so that only a certain set of the features are used in creating the proximity raster. To set this you must first select the attribute field from the proximity factor layer and then select the value of the attribute that a filtered feature must have. Like the default weight, you can set the default field and value to use for each proximity factor. To do this, click on the Default Subset Field property in the tree. The properties list will be filled with the attribute fields present in the proximity factor map layer. Choose the field that you would like to use by double-clicking the field name in the list. Then, click on the Field Value property and enter value for the field you’d like to use. Click Apply to apply the value. The tree will be updated with the value. If you do not want to filter the proximity layer, don’t set any subset field or value.

There is one proximity factor that is not entered as a map layer. The proximity to Like Landuse factor is included by default with a new scenario. This factor can’t be specified by features in a user-entered map layer because the landuse is changing with each year according to allocation by the emulation algorithm. Therefore, no map layer needs to be associated with the proximity to Like Landuse layer.

To **remove** a proximity Factor, click on the factor and then click the remove button on the tree.

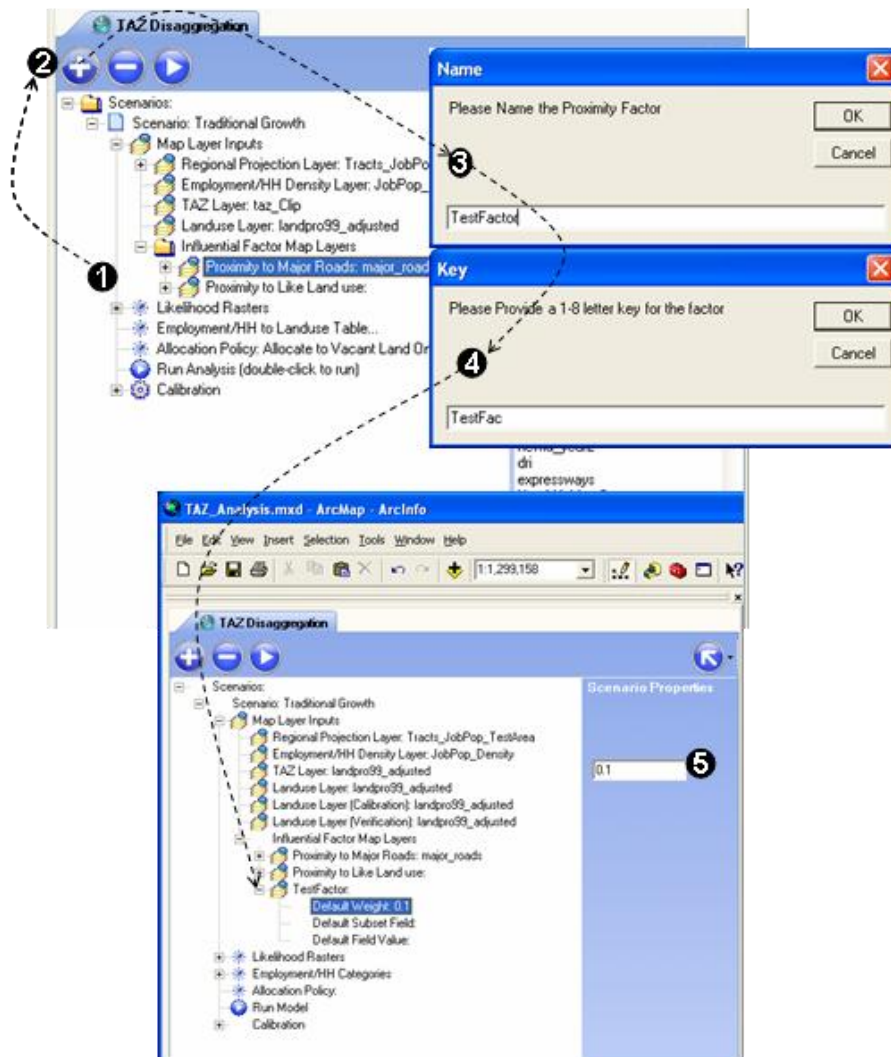


Figure 4.6, Adding a proximity factor requires selecting a layer from the map and then specifying the default weight and filter properties for the factor.

Step 3. Setting the Likelihood Raster Weights

Projections for each landuse type are allocated with likelihood, L, raster (see chapter 3 for a description of the L-raster). The L-raster gives the likelihood of development at a particular location. It is the superposition of a weighted set of N proximity rasters, F. The weights, α , are set using the Likelihood Raster portion of the tree.

$$L(x, y)_{landuse} = \sum_{i=1}^N \alpha_{i,landuse} F(x, y)_i$$

The weights are listed in a table (see figure 4.7) that is shown below the tree when the “Weights...” node is clicked. The table is organized with proximity factors as columns and landuses as rows. The total weight is shown in a column to the right of the table. The total weight is required to be 1.0 for a given landuse’s L-raster. If you have proposed a set of weights that does not sum to 1.0, the total will be colored red. You can enter the weights by simply typing them into the table. After you have typed in a weight, the total weight value will be updated.

Optionally, you can also set the filters for the proximity factor map layers within the Likelihood Rasters portion of the tree. To set the filters,

1. Click on the Map Layer Filters Node. A table with the proximity factor map field and values as columns and the landuses as rows will show below the tree.
2. Each filter has two inputs: the field in the proximity factor map layer and the value you want that field to be. There are therefore two columns in the table for each proximity factor - one for field and one for value. To change the field, click on the cell for the landuse and the field column for the proximity factor you want to change. When you do the properties list will be filled with the names of the fields in the proximity factor map layer.
3. Double-click on the field you want and the field will be changed in the table.
4. To enter field values, click on the cell you’d like to enter a value for and then enter the value with the text box in the property area. Click Apply to update the table.

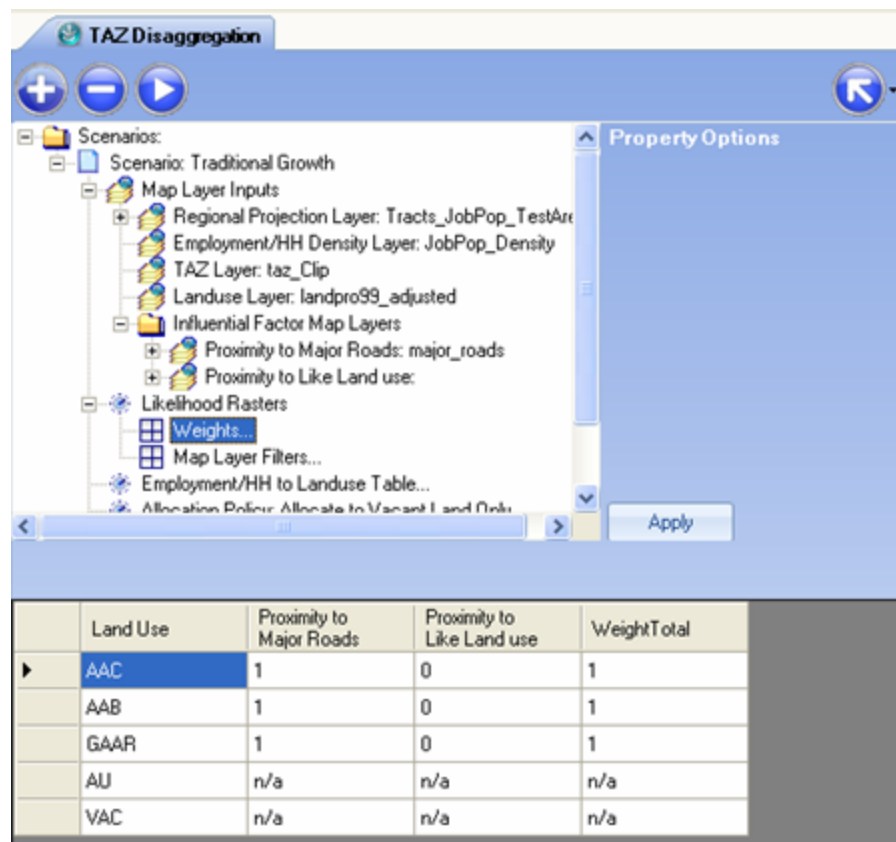


Figure 4.7, Each landuse has a different Likelihood raster, which is the superposed weighted sum of the proximity factors. Setting specific weights for proximity factors for each landuse is done in the Likelihood rasters section of the tree. To set the weights, click on the weights node. Enter the weights you'd like in the table that appears.

Step 4. Setting the Employment/Household to Land Use Map

Projections are reported from DRAM/EMPAL as either employment or household (HH) by income categories. Allocation is done according to landuse. As a result, each employment or household category must be mapped to a land use.

The category to landuse table is shown in the tree under the Employment/HH Categories portion of the tree (see figure 4.8). Each category and its associated landuse are shown in this tree.

The values of the category to landuse table are derived from a table called “EmpHHLUMap” that should be loaded to the MXD. This table must have two columns, one called DE5 and one called EmpHH. In the EmpHH column is the name of the employment or household category. In the DE5 column is the value of the corresponding DE5 landuse. If this table is not present in the MXD, the following default set is used:

- MISC: Retail - Commercial Employment (AAC)
- CONST: Retail - Commercial Employment (AAC)
- MFG: Retail - Basic Employment (AAAB)
- TCU: Retail - Basic Employment (AAAB)
- WHOL: Retail - Basic Employment (AAAB)
- RETL: Retail - Commercial Employment (AAC)
- FIRE: Retail - Commercial Employment (AAC)
- SVCS: Services - Commercial Employment (AAC)
- GOV: Services - Basic Employment (AAAB)
- HH_INC1: Retail - Residential (GAAR)
- HH_INC2: Retail - Residential (GAAR)
- HH_INC3: Retail - Residential (GAAR)
- HH_INC4: Retail - Residential (GAAR)
- HH_INC5: Retail - Residential (GAAR)
- HH_INC6: Retail - Residential (GAAR)

You can also change the category to landuse table by clicking on the property you want to change in the table and then entering the value with the properties list. Also, you can add new categories and take away categories by using the add and remove buttons in the toolbar.

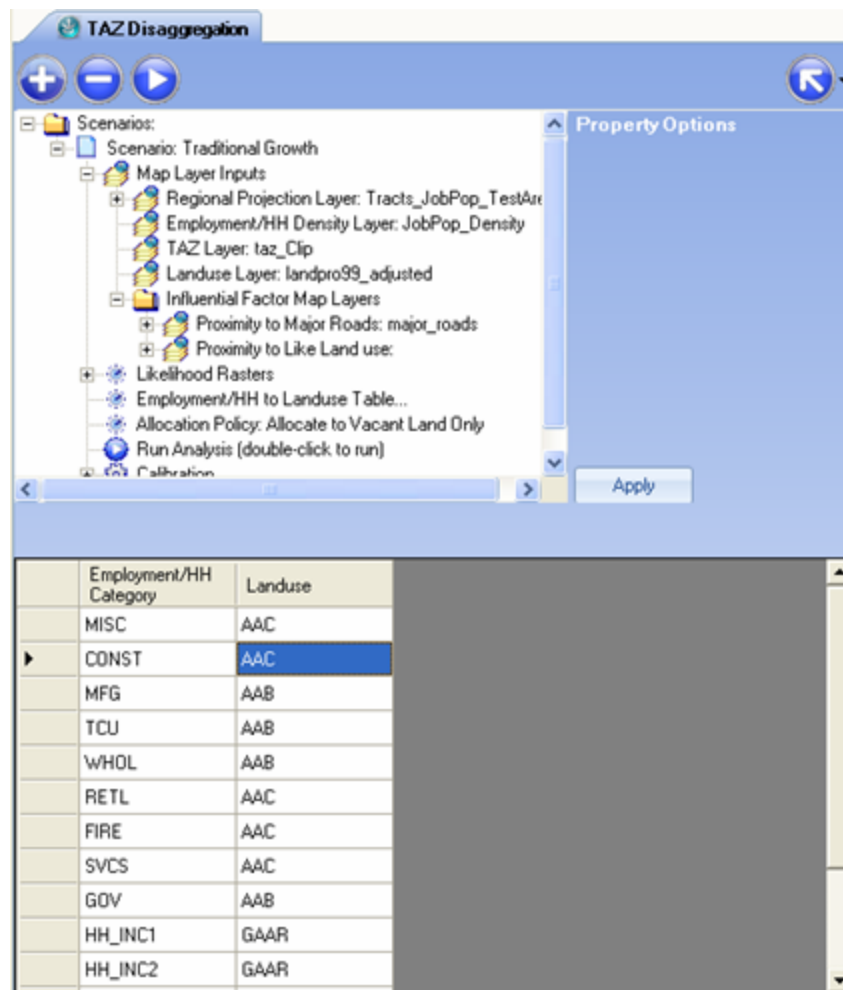


Figure 4.8, the Employment/HH to Landuse category table is accessed by clicking the Employment/HH to Landuse Table node in the tree. This table provides the link between projections, which are in employment or HH categories and landuse, which is how the projections are allocated.

Step 5. Setting the Allocation Policy

Allocation of projected new employment and households can proceed according to two policies, vacant land only, and re-development. With vacant land only, allocation is only considered for land that is classified as vacant. This means that even in situations where land is sparsely populated, it is not considered for new development. For the re-development option, all land is considered for re-development. Vacancy can be added as another proximity factor, meaning vacant land can be made more desirable by adjusting the weight for the vacant land proximity factor.

To set the allocation policy, (see figure 4.9)

1. Click on the allocation policy node.
2. Select either Allocate to Vacant Land Only or Allocate to All Land and then click apply.
3. The tree will be updated.

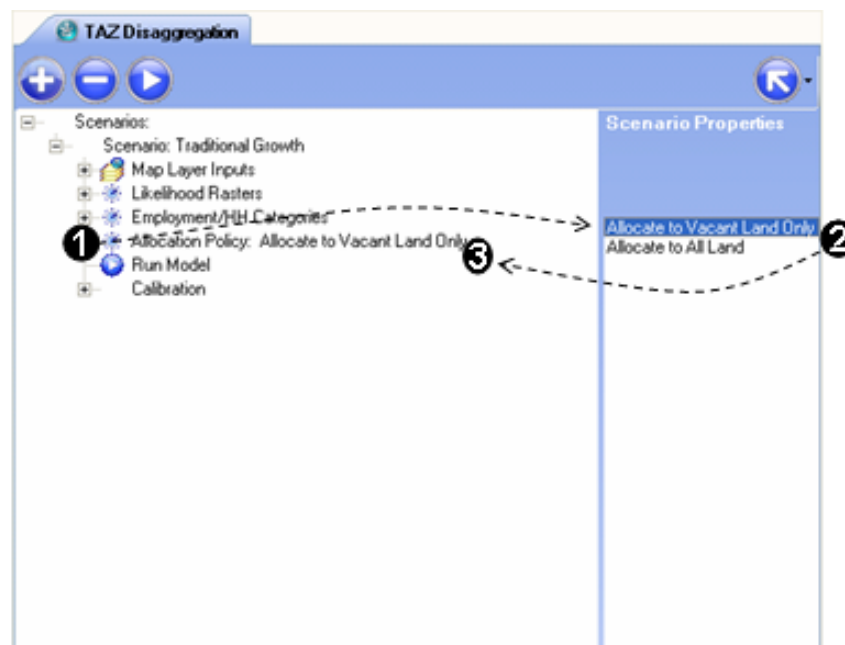


Figure 4.9, Set the allocation Policy by clicking on Allocation Policy and selecting the allocation policy from the properties list.

Step 6. Run the Analysis

To run the analysis: (see figure 4.10a)

1. Click on the Run Analysis node in the tree and then click the run button in the toolbar. You can also double-click the Run Analysis node.
2. As the tool runs, it will provide information about what is being processed in the status bar.
3. When the process is finished, a window will pop up indicating successful completion.

A quality control procedure is run when the analysis is run to ensure the various map layers, fields, and weights have been set correctly. The quality control procedure will inform you of what is missing if the analysis can not be run.

See chapter 2 for an explanation of the algorithm that is run.

The run will add the L-rasters to the map as the calculation proceeds. At the end of the run process, the raster showing the new landuse allocation will be added to the map (see figure 4.10b).

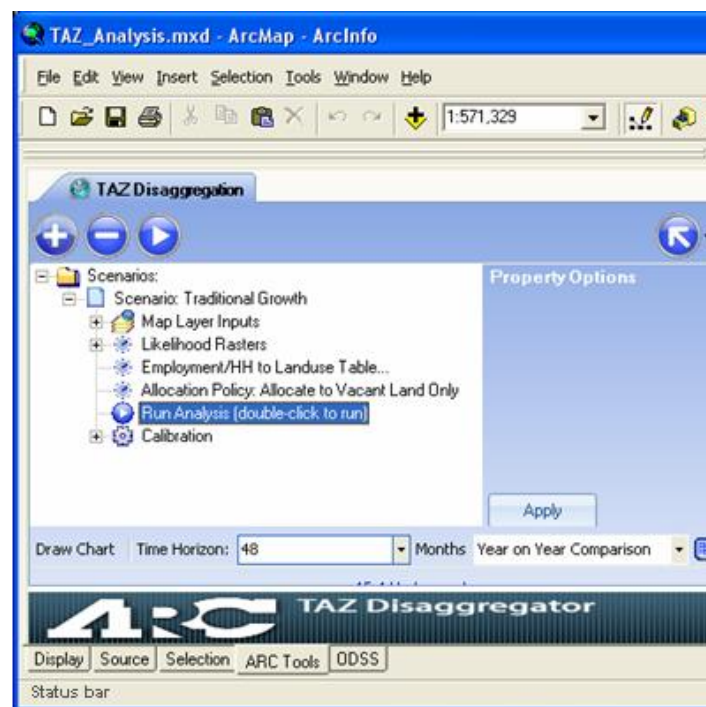


Figure 4.10a, Run the analysis by double-clicking the Run Analysis node in the tree.

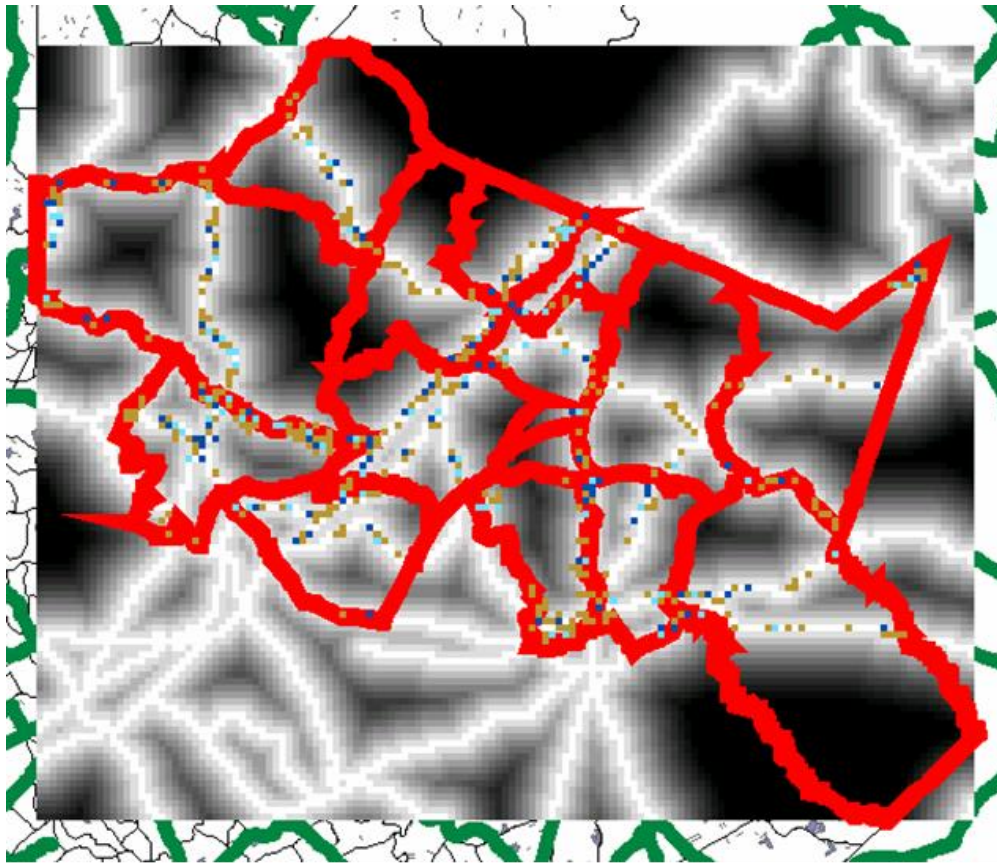


Figure 4.10b, The L-rasters (gray-scale rasters) are added to the MXD as the emulation algorithm runs. The new landuse raster (brown and blue raster) is added at the end of the emulation.

CALIBRATION

Calibration can be conducted for a scenario. Recall that the likelihood raster that is used to determine allocation of new employment and housing. The L-raster is the superposition of a weighted set of N proximity rasters, F, as shown in the following equation.

$$L(x, y)_{landuse} = \sum_{i=1}^N \alpha_{i,landuse} F(x, y)_i$$

The goal of calibration is to find the set of weights, α , that result in the closest match between actual growth and modeled growth.

The calibration process consists of two steps, **calibration** and **verification**. The calibration step emulates development from a start year to an end year. The calibration weights are the set of weights that results in the smallest TAZ-level sum of square errors (SSE) between the emulated development and actual development in the end year. The verification step uses the calibration weights and evaluates the SSE during a verification period, which starts at the end of the calibration period and ends at the verification end year. The verification SSE score provide a more accurate measure of the performance of the emulation algorithm because the weights are not found from the data in the verification period. The calibration SSE is almost always lower (more accurate) than the verification SSE score.

To find the calibration weights, two algorithms are available, brute force, and simplex.

The **brute force method** evaluates the SSE for all viable sets of weights (subject to a specified weight resolution (eg. 0.2 on a 0 → 1.0 range). The set of weights that results in the lowest SSE (during the calibration period) is selected as the best set of weights. The advantage of the brute force method is that it calculates the full SSE surface (the SSE for all possible sets of weights) and ensures the best set of weights is found. This best set of weights is also referred to as the global minimum.

The **simplex method** is a searching algorithm that looks for the minimum SSE weight set by iteratively adjusting the weights to move in the direction of maximum slope in the SSE surface. The advantage of the simplex method is that it doesn't need to calculate the SSE for every set of weights, only for those in the search path. The disadvantage is that the algorithm can stop at local minima in the SSE surface, and is therefore not guaranteed to find the global minimum. Several additions to the simplex algorithm ensure that this possibility is unlikely, but it cannot be guaranteed. This algorithm is described in more detail in chapter 3.

This section describes how to run the calibration process.

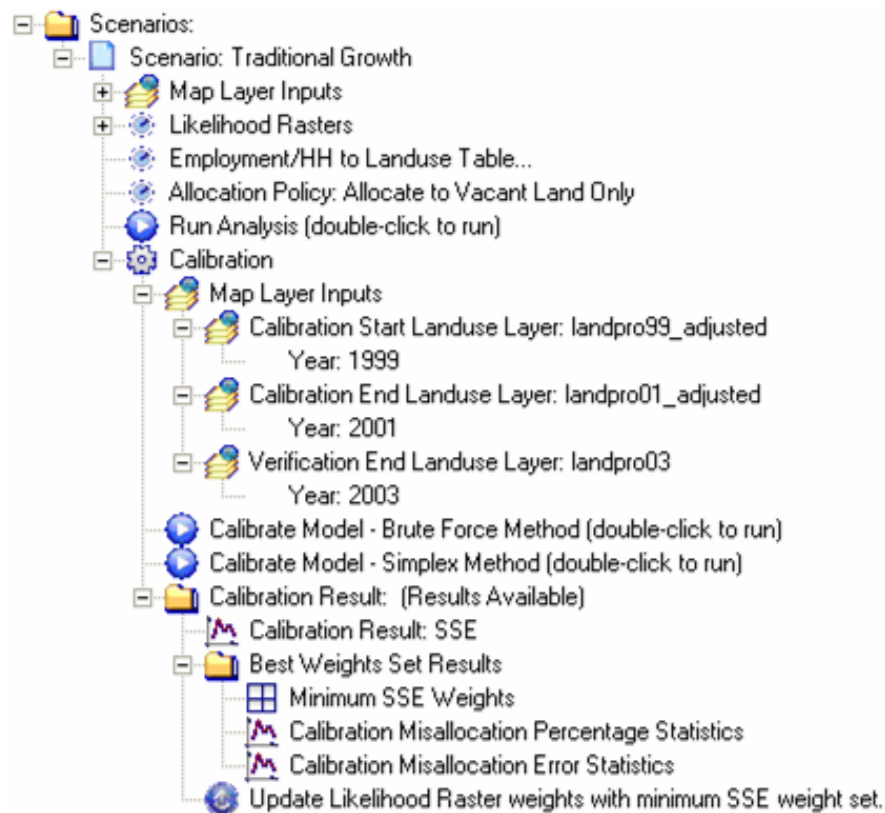


Figure 4.11, The calibration process is run with the calibration portion of the tree. The aim of calibration is to find the set of weights that results in the closest match between actual growth and modeled growth. The calibration process requires several additional map layers to correspond to the start and end of the calibration period and the start and end of the verification period. For simplicity, the start of the verification period is assumed to be the end of the calibration period.

Map Layers (see figure 4.12)

To run the calibration process, three land use layers are required:

- **Calibration Start Landuse** Layer: this layer is the landuse as the beginning of the calibration period
- **Calibration End Landuse** Layer: this layer is the actual landuse at the end of the calibration period. This layer is also used as the landuse at the start of the verification period.
- **Verification End Landuse** Layer: this layer is the actual landuse at the end of the verification period.

The specification of the landuse layers for the calibration process is done in the calibration portion of the disaggregation tree. Like the main tree, there is a Map Layer Inputs section. Set the various layers as you would the main map input layers (see step 1 of the building a disaggregation scenario).

Land use Year

An additional piece of required information is the year for the three land use layers. The years of each land use layer must be known for the emulation process because the emulation loop splits the total projection into yearly increments and allocates on a per year basis. For example, if the calibration start year is 2001 and the calibration end year is 2003, then the emulation algorithm will first allocate half of the new growth in 2001 - 2002 and then the other half in 2002 - 2003. The number of times the algorithm loops is critical because it will determine the pattern of emulated development.

To set the year for a landuse layer, drill down to the year property under each landuse layer in the tree and enter the year using the text box in the properties area and click apply. The year will be shown in the tree.

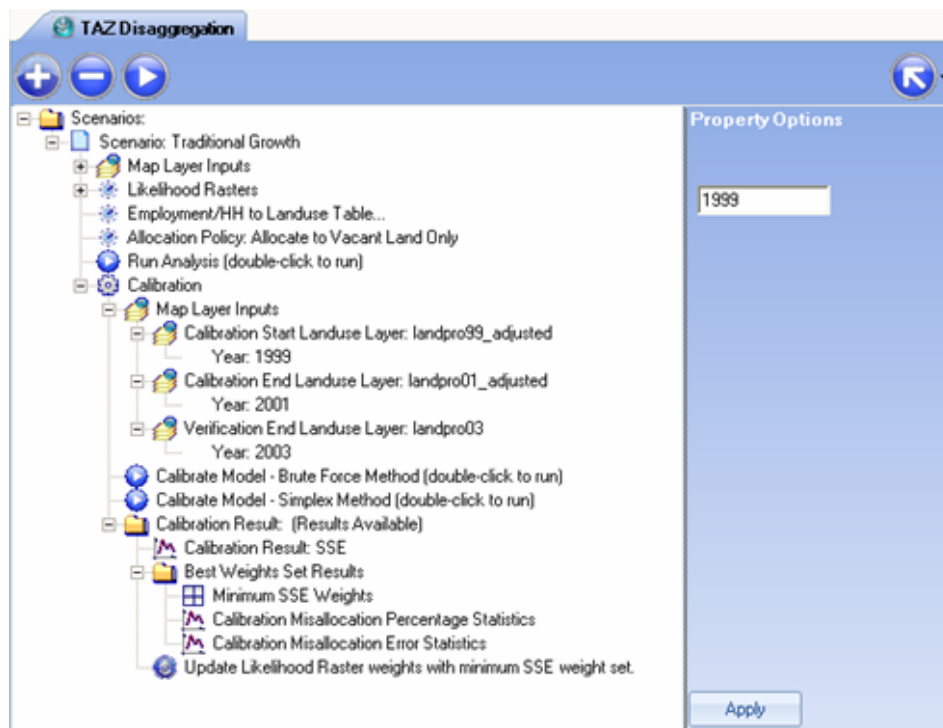


Figure 4.12, The calibration process requires three map layers, the calibration start Landuse, calibration end Landuse and verification end Landuse. Set these layers as you would layers in the general process map layer inputs (see step 1 of the general process). You must also set the years that the landuse layers correspond to. Set these by clicking the year node in the tree and then entering the year in the text box that appears.

SET THE WEIGHT RESOLUTION

The calibration process requires a resolution for the weights. For example, if the range of weights for a specific factor is 0.0 → 1.0, then with a 0.2 resolution, the calibration process will try 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0 as weights for that factor. The finer the resolution, the longer the calibration process will take, but the more accurate the solution will become.

To set the resolution, click on the weight resolution node in the tree. Enter the weight you'd like in the text box that appears and click apply (see figure 4.13).

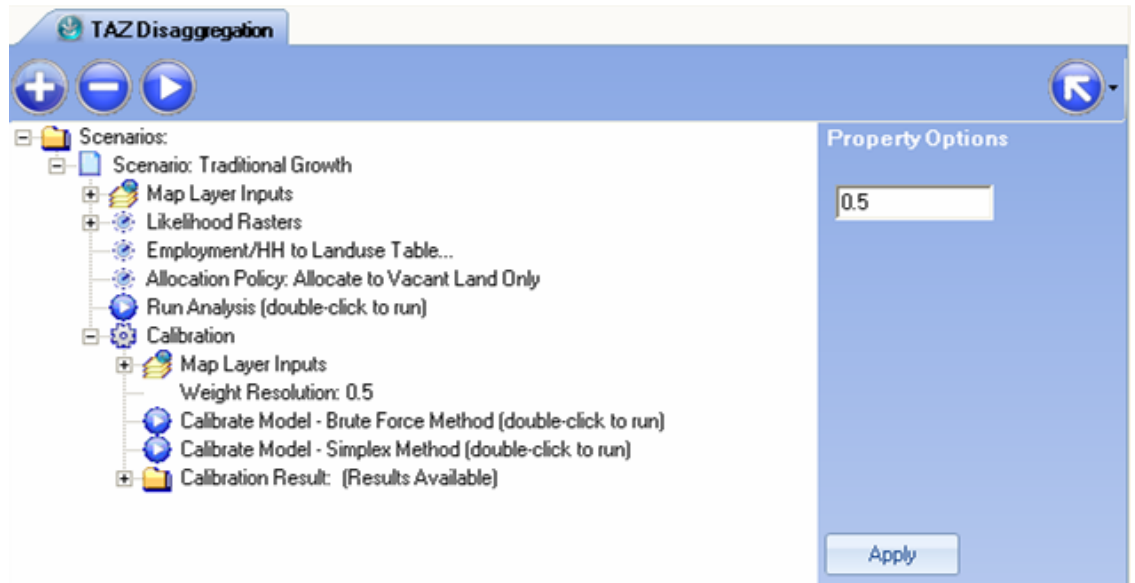


Figure 4.13, Set the weight resolution by clicking the weight resolution node in the tree and entering the weight you'd like in the text box that appears. Click Apply to update the weight resolution.

RUN THE CALIBRATION PROCESS

The calibration process will use the scenario inputs (regional projection layer, employment/HH density layer, proximity factors and associated filters, employment/HH category to land use map, and the allocation policy) as input for the calibration. The weights will be controlled by the calibration algorithm and the best set of weights reported at the end of the process.

If you'd like to run the brute force algorithm, click on "Run Brute Force" algorithm node and click the run button in the toolbar (or double-click the "Run Brute Force" node). If you'd like to run the simplex algorithm, click on the "Run Simplex" algorithm node and click the run button (see figure 4.14).

As the calibration process is run, the status bar will report the progress of the algorithm.

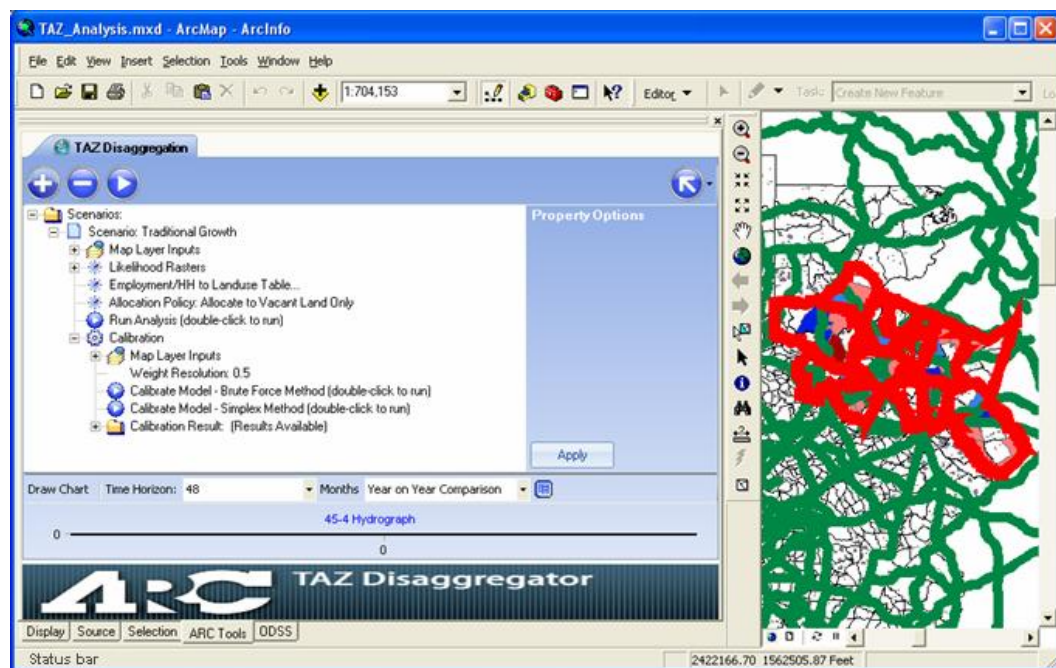


Figure 4.14, Run the calibration by double-clicking the run node in the tree for brute force or simplex method. The status bar reports the progress of the calibration as the run continues.

CALIBRATION RESULTS

The calibration process creates several results (see figure 4.15):

Calibration Result: SSE Chart - this chart shows the SSE for all the weight sets that were calculated. The best set of weights is the one with the lowest magnitude bar. The labels on the bars give weights for the proximity factors. For example, if the title given on the horizontal axis is Factor Weights: Major Roads, LikeLU, then a bar with 1,0 as a label means that that Major Road weight is 1.0 and the LikeLU weight is 0.0. Show the chart by double-clicking the Calibration Result: SSE node.

Minimum SSE Weights - This table gives the minimum SSE set of weights for each of the landuses. Show the weights by clicking the Minimum SSE Weights node.

Calibration Misallocation Percentage Statistics - this chart shows the accuracy that results with the minimum SSE weight set. Two series are shown in the chart,

- average percentage of CT Projections misallocated
- percentage of CT projections misallocated.

These two series show that even using the minimum SSE weight set, often a high percentage of the CT projections are misallocated to the wrong TAZ, and often 100% of the projections are misallocated in some way. It is important to view this chart to understand in an absolute sense how well the proximity factors can emulate development.

Show the chart by double-clicking the node.

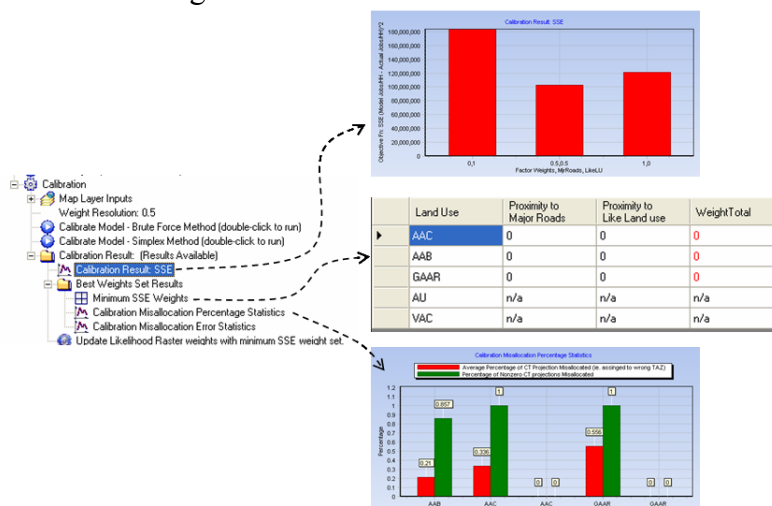


Figure 4.15. Calibration returns several results including the Calibration Result: SSE chart, Minimum SSE Weight Set, and the Calibration Misallocation Percentage Statistics Chart.

The calibration process also updates the modelstat and actualstat attribute fields in the TAZ map layer with the GAAR HH count for modeled and actual GAAR landuse. The GAAR HH count is calculated as the area of GAAR landuse multiplied by the HH density in those areas. Make sure that these fields are present in the layer. Both are double type attribute fields.

Figure 4.16 shows an error field, which has been calculated as ModelStat - ActualStat. Note from the result that a several TAZs are misallocated. Also note, that under this scheme, where grid cells are 1000' x 1000' and HH densities range from 9 to 13 per acre, 1 grid cell corresponds to about 200 households.

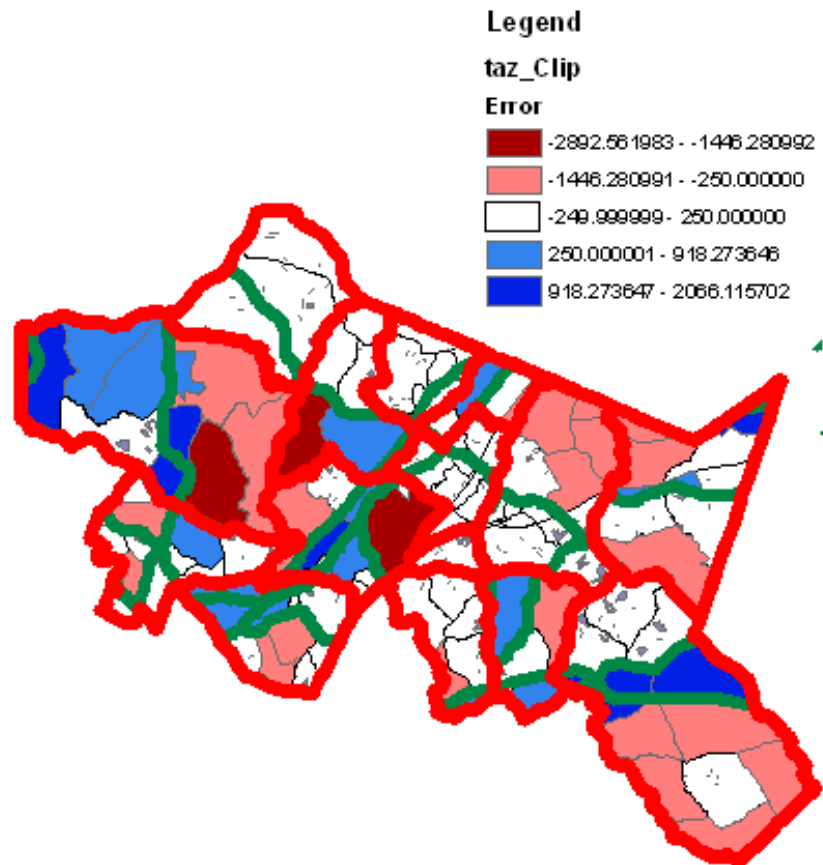


Figure 4.16, The calibration process updates the ModelStat and ActualStat attribute fields in the TAZ Map layer with number of households per TAZ according to the actual landuse and the modeled land use. This figure shows the difference between the model and actual results.

UPDATING THE SCENARIO WITH THE MINIMUM SSE WEIGHT SET

The scenario can be updated with the calibration weight set with the Update Likelihood Raster weights with minimum SSE weight set node under the calibration results portion of the tree. To update the scenario weights, double-click the node.

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.