Using a Neural Network to Evaluate Land-Use Change

David Nouza, Verne LaClair, and Matthew Schwab

Abstract

By applying an Artificial Neural Network to a selection of raster spatial layers (GRIDS) it is possible to model build-out within ArcGIS. Inputs include two land-use time steps, a mask layer, and user-selected socioeconomic, political, and environmental inputs–driving variables. For each cell in the study area the real-change between the two time steps is determined, and analyzed vis-à-vis the provided driving variables in order to produce a probability of land-use change layer. ArcObjects and Visual Basic are used to automate integration with the neural network. The spatial accuracy of model predictions is evaluated in a spreadsheet by charting the proportion of correct predictions (# predictions/# cells transitioning in the observed database) against a floating cell-window. The use of an artificial neural network to evaluate land-use change is applicable on a regional level.

Introduction

Conventional build-out analysis is a process that allows testing of existing land-use regulations and policies. It allows a glimpse at a possible future when all land has been developed to the maximum extent permitted. Modeling an intermediate level of change – for example transitioning ten percent of the available land – is problematic since it requires a subjective determination of land desirability. The Build-Out Modeling Tool (BMT) was developed to assist the analyst performing intermediate build-out analyses.

Our study area was the West Branch and Boyds Corner basins of the New York City watershed. The watershed is part of an extensive reservoir system that supplies the City. It is regulated by the City's Department of Environmental Protection (DEP). The study area was chosen to see how the BMT would perform in an area that was largely built-out and had numerous land-use regulations in place. We hypothesized that growth would be displaced from restrictive areas to less regulated areas.

There are two phases in any build-out analysis: depicting change and quantifying the impact of the changes. A geographic information system (GIS) can assist the analyst in both phases. By providing a map-based interface, the GIS can assist in exposing spatial relationships that exist between model inputs. By using an Artificial Neural Network (ANN) in conjunction with the GIS the relative impact of each input layer on these relationships can be evaluated using non-linear spatial correlation.

The ANN based BMT is designed to emulate the functionality of biological neurons in order to achieve a higher parallel processing potential for digital data. It is similar in structure to regression type land use change models that model relationships between potential drivers and occurrences of transitions. It differs in that it is capable of finding nonlinear solutions to fit data, and that it is somewhat tolerant of errors in data which may be present in land use/cover data derived from various sources. BMT currently employs a multi-layer perceptron with one or two hidden nodes depending on the number of input variables. Inputs are derived from raster spatial layers (grids), with cell values normalized between 0.0 and 1.0. Output is in the form of ArcGIS grids composed of cell locations that transitioned (coded as a 1), and those that did not change (coded as a 0).

Version 2 of BMT has been implemented in Visual Basic in the ArcGIS environment. Other elements that have been developed include enhancements that utilize MSU's Land Transformation Model (LTM) approach for formatting ESRI Grid outputs for input into a PC based version of the Stuttgart Neural Network Simulator (SNNS), and allocate exogenous growth prediction to the most suitable grid cells.

Method

The BMT processes land use changes in response to driving variables over a number of time steps. At the end of each time step, a land use map showing the probability of change for each cell is generated. The area of land whose land-use type is externally predicted to change is allocated to its new land use category through the use of this probability map (Figure 1). The land with a new land-use type is overlaid onto the initial land-use data set and a new land-use map is thus created. This new land use map then serves as the basis for the next set of driving variables and the creation of a new probability map of land use change.

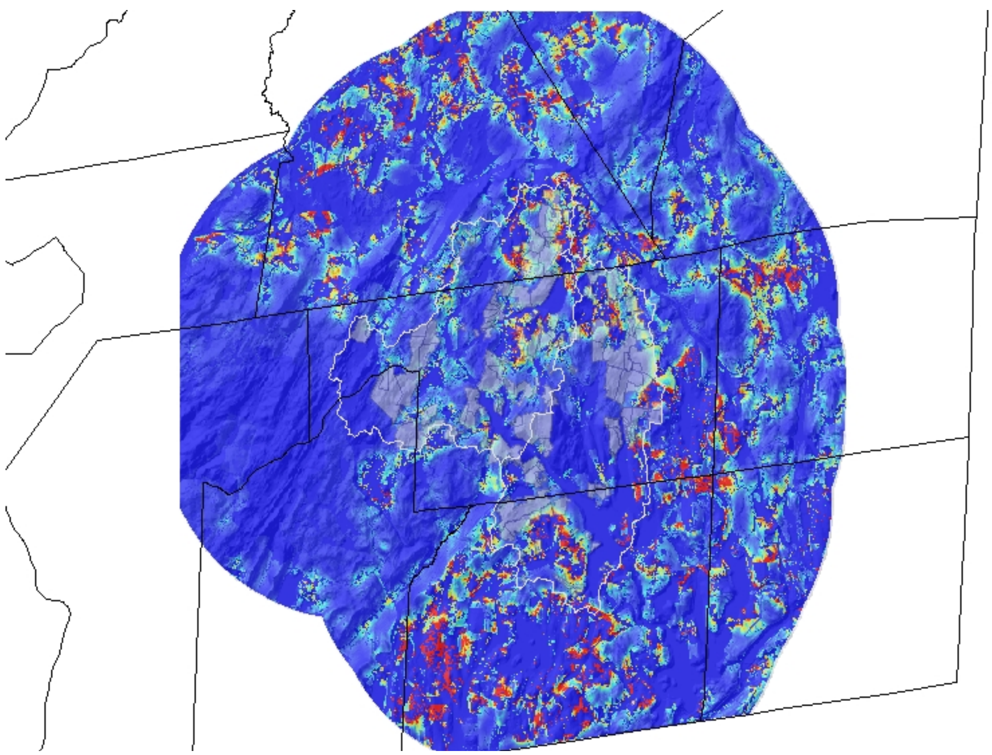

Figure 1. 2010 BMT probability map for West Branch & Boyds Corner Basin region. Red indicates high probability of growth, blue indicates low.

Land Use/Land Cover (LULC) classifications in the West Branch/Boyds Corner study area were used as input to train the BMT. In particular, a 1990 UMASS based LULC classification was provided by DEP and the 2001 Landsat 7 based LULC classification was developed in-house.[1]

The BMT was configured to use a select set of socioeconomic, political, and environmental spatial datasets. These driving variables have been proven through test runs on controlled datasets to influence urbanization patterns in the East of Hudson area of the City's watershed. ArcGIS, and in particular, Spatial Analyst was used to develop the driving variables. Through a process of several trial model runs and subsequent analysis, the model inputs found to significantly influence development patterns included the following:

Distance to Reservoirs	Slope
Distance to Existing Residential	Viewshed (Quality Views)
Distance to Commercial Centers	Zoning
Road Network Density	Urban Cell Density

Many other factors contribute to the complex interactions between economics, policy, culture, management, environment, and human behavior that influence land use change. These driving variables fairly represent factors that a traditional build-out analysis would consider, plus they could be generated and graphically displayed with reasonable research and interpretation efforts.

In addition to the selection of good model inputs, areas that are to be excluded from land use change must also be specified. For purposes of simplicity and model flexibility, the initial exclusionary layer was limited to the obvious areas such as open water, transportation right-of-ways, existing urban cells, designated forests, and state lands/parks.

Once the desired inputs had been developed, the neural network was used to generate and learn patterns of development in the region based on the two distinct land-use snapshots: 1990 and 2000. The overall "predictability" of the model was determined by using predetermined input parameters based on observed changes in land-use from 1990 to 2000 and comparing the neural nets output with known 2000 datasets. An Excel spreadsheet was then generated to document and chart the individual influence of each driving variable input on the model's performance. The proportion of correct predictions (# predictions/# cells transitioning in the observed database) was plotted against window size to chart the spatial accuracy of the model's predictions.

The study area was subdivided into a grid consisting of 10meter cells which facilitated the calculation of a predictability value for each cell in the grid. Each of the included driving variable inputs and the exclusionary layer were generated and snapped to the same extent as the LULC layers.

[1] Using Impervious Feature Extraction in conjunction with available digital tax parcel data within the study area, a 0.3-meter raster data layer was created with the tax parcel year-built as the attribute. For areas with no year-built (i.e. roads) a value of "1" was assigned, allowing them to be present, yet retain their original classification in the final data set. This technique addressed inconsistencies in delineation of urban areas in LULC classifications generated using different methods and provided the BMT with a more accurate representation of urban development.

Two independent scenarios were modeled with the same number of cells specified to undergo urban transition; 250,000 cells or 4.32% of all available cells in the study area. Identical driving variable inputs were used to eliminate influences from additional factors. The single difference being the addition of DEP acquired lands to the second scenario. The purpose of this exercise was to evaluate the effect of imposing development restrictions within the basins by means of parcel acquisition by the DEP.

Results

Model results indicated a strong correlation to proximity of existing urban cells and quality of views at the smallest scales, while slope and proximity to reservoirs maintained a positively increasing correlation with increasing site scale.

Predictability was found to be good as illustrated by Figure 2 below. The vertical arrow indicates that within a window of 50 pixels X 50 pixels (10m2 per pixel), or 0.5 square kilometer, the model had performed at a correct prediction rate of 52.5% when applied to the known LULC file.

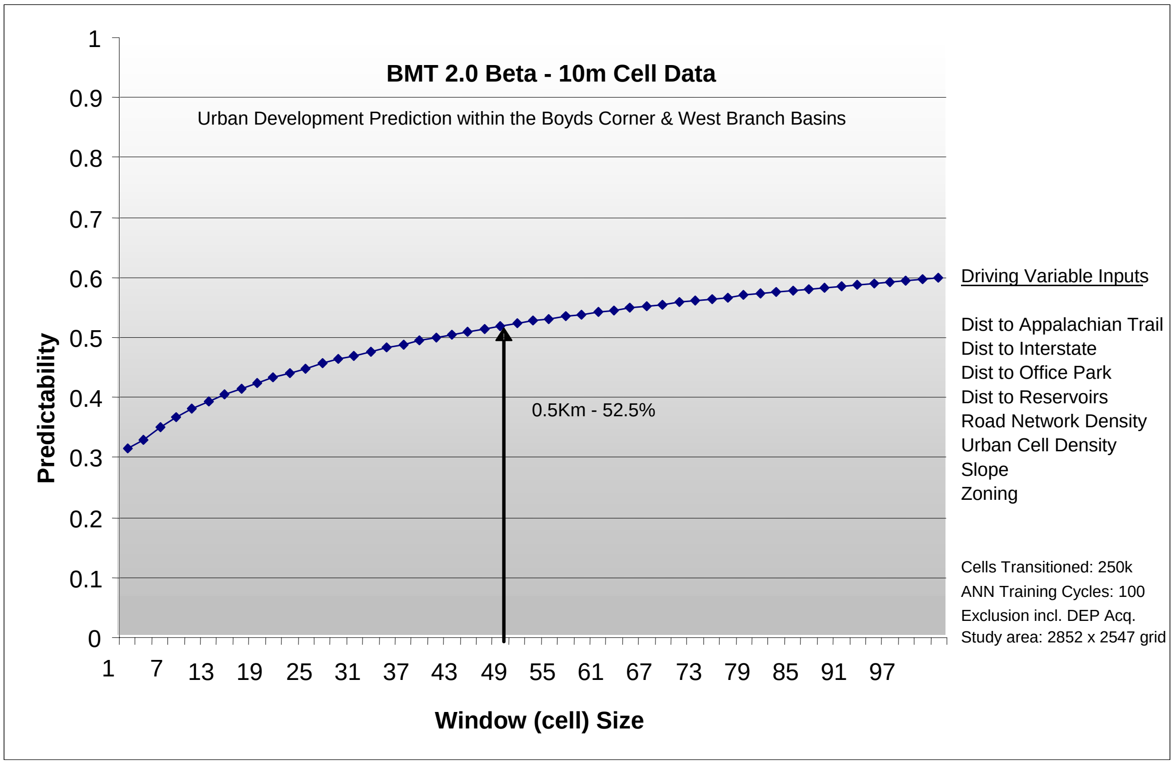

Figure 2. Model performance and predictability level.

A screen shot showing the 10 year projected development pattern is shown below together with shaded DEP lands. (Figure 3). Existing urban (2000), is shown as dark blue, and predicted urban areas within the basins (2010) are dark red.

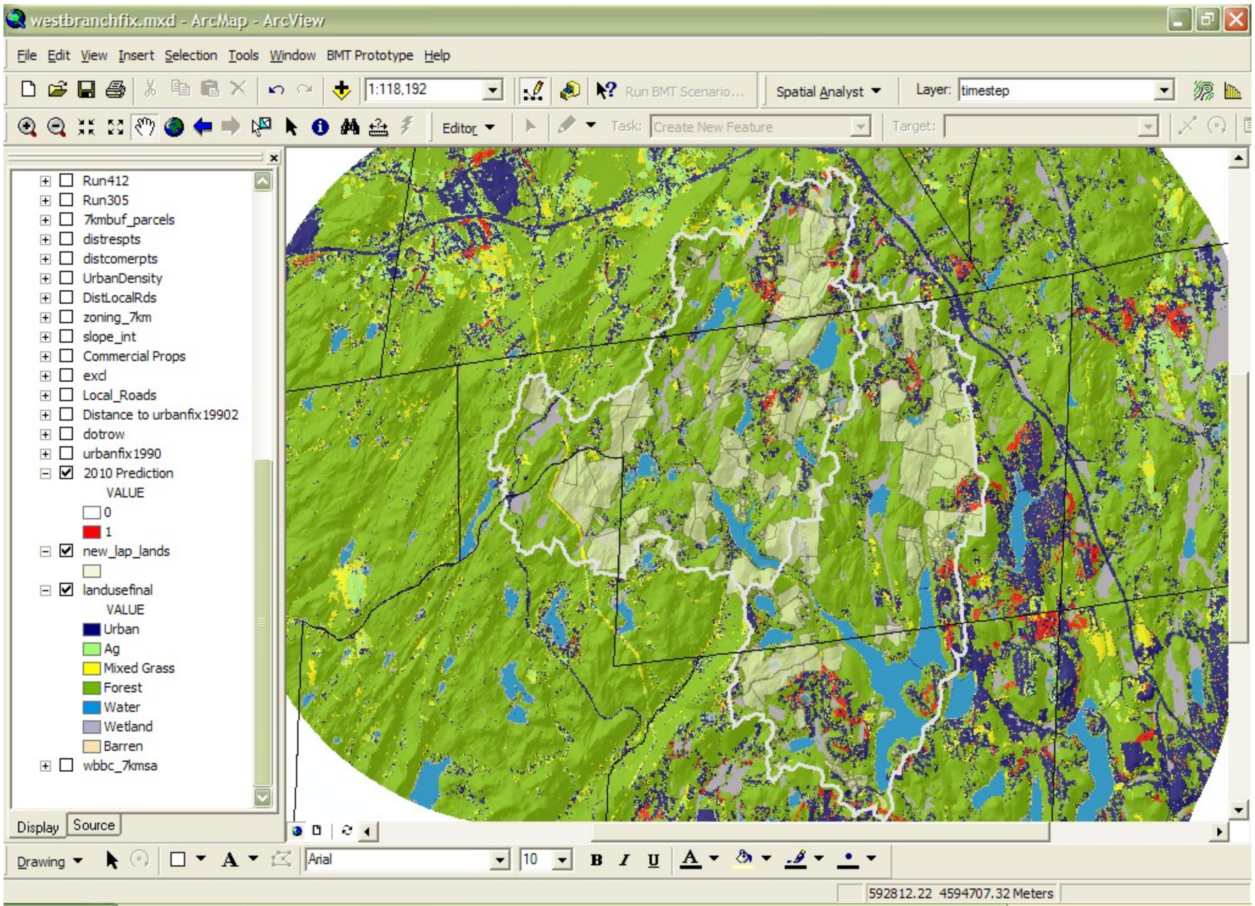

Figure 3. 2010 urban projections with DEP parcels shaded light green.

Discussion

The primary areas selected for growth within the basins are concentrated in two distinct regions. The first is on west side of the West Branch reservoir, in the southern part of the basin, and the other is directly opposite in the northern region of the Boyds Corner basin. A further breakdown of number of cells transitioned within the basins, as well as cells transitioned within DEP parcels is provided in Table 1 and charted in Figure 4. A range of transition values were selected and modeled to get a better understanding of the likelihood of cells within the basin to change to urban. This process also provided insight as to the proximity of probability values for cells in the basins to the critical threshold value, which is an indictor of where and at what scale the models predictability performance begins to degrade, or over generalize.

Date of run	cell size	DEP Acq Lands	Total Transitioned		Basin trans		DEP Acq. trans	
			Cells	Acreage	Cells	Acreage	Cells	Acreage
7/31/2003	25	Excluded	40000	6178	3494	540		
8/1/2003	25	Available	40000	6178	2211	341	735	114
5/21/2004	10	Excluded	100000	2500	12265	307		
5/20/2004	10	Available	100000	2500	16131	403	3228	81
6/25/2004	10	Excluded	250000	6250	33981	850		
6/24/2004	10	Available	250000	6250	42749	1069	12416	310

Table 1. Table of projected development amounts within study area.

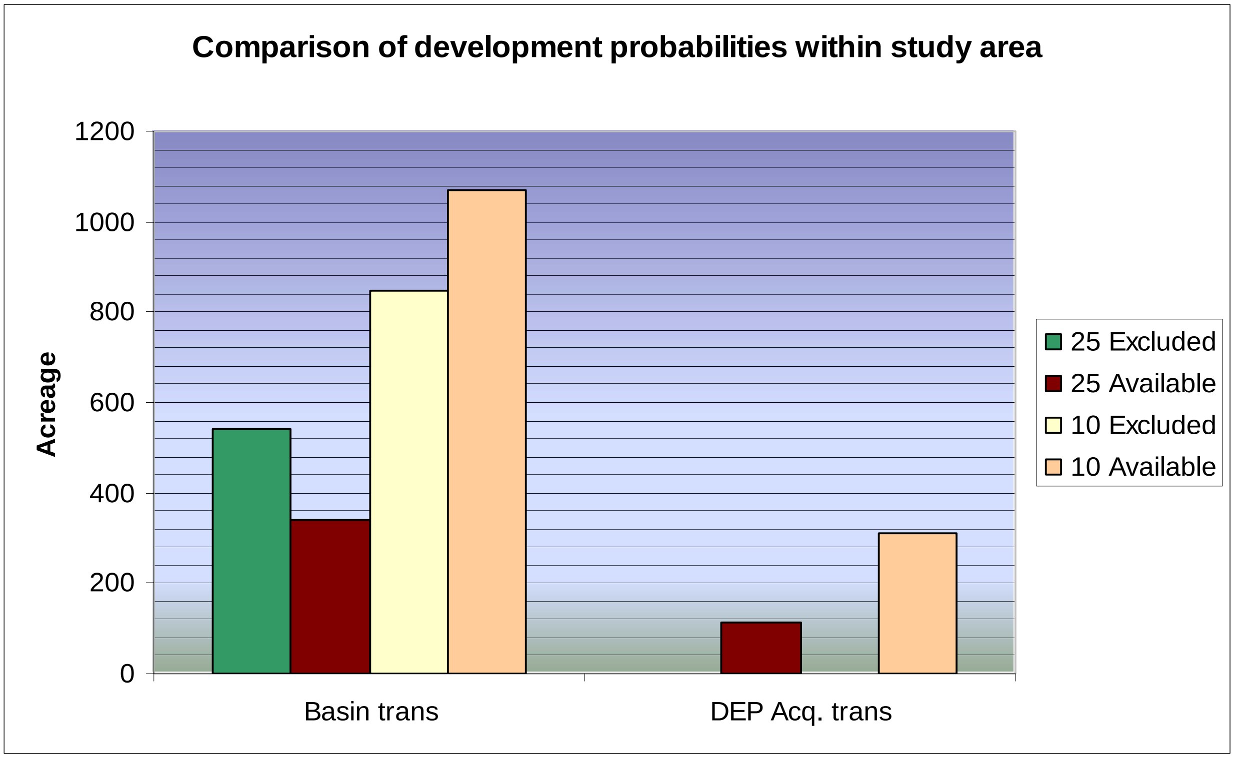

Figure 4. Chart comparing development predictions for excluding DEP parcels from development vs. Allowing DEP parcels to be open to development.

Of the 250,000 cells with the highest probability to change, 55,842 (22%) had values greater than or equal to 0.90. The critical threshold value, which is the least probable value of the 250,000 cells selected, was 0.657. These results indicate that the model is fairly precise at the 10meter cell size scale in identifying change from no change. When studying the effects of introducing additional restrictions on development within the study area, it was found that by removing development eligibility of select parcels, transition probabilities increased within the watersheds.

This is perhaps in contrast to expected results. However, given that the study area resides in a rapidly growing and highly developed region of the Hudson Valley, it may be an indication that “green space” is a positive driving variable for residential development.

A second observation of the results listed in Table 1 indicated that when a smaller cell size was used, the expected results were obtained. That is, excluding DEP parcels from development tends to drastically reduce predicted cell transition in the watershed. Performing a simple spatial calculation on these results, it was determined that the majority of projected development was occurring in parcels that are located near the edges of the basin boundary.

Several assumptions were made in order to keep the model simple. First, it was assumed that the patterns of influence for each driving variable remained constant since 1990. Second, the spatial

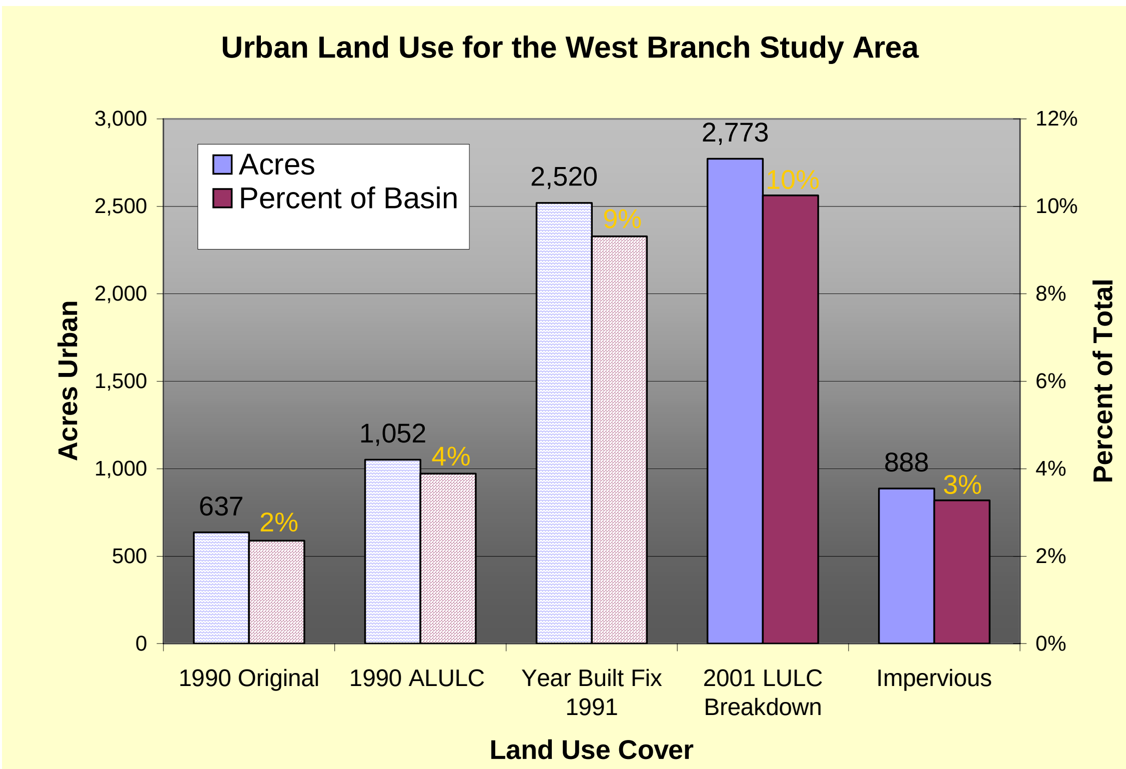

Figure 5. Land-Use Land-Cover results for study area.

relationships used to build the driving variables were assumed to be accurate and would remain constant throughout the projected time step (10 years). Third, the neural network was assumed to remain constant over time, thereby stabilizing the relative influence of development patterns over the projected time period. Lastly, the amount of urban transition specified for each model run is assumed to be static over the projected time step. Given information on new and planned infrastructure projects and additional temporal information about land use changes and population projections, it becomes possible to redefine a few of these assumptions in order to evaluate their effects on overall model performance. Ultimately the modeler would have sufficient information available to be able to minimize the number of assumptions necessary to obtain consistent and accurate modeling results for the area to be studied.

2001 Land Use/Land Cover Category	Study Area		WBR Basin	
	Acres	% of Total	Acres	% of Total
Urban	21,232	**15%**	2,773	**10%**
Agriculture	6,424	**5%**	113	**0%**
Brush	7,347	**5%**	918	**3%**
Forest	93,740	**66%**	19,343	**72%**
Water	5,886	**4%**	1,889	**7%**
Wetland	6,980	**5%**	2,004	**7%**
Barren	18	**0%**	13	**0%**
Total	141,627		27,054	

Figure 6. Break-out by land-cover type.

References

Land Transformation Model: http://ltm.agriculture.purdue.edu/ltm.htm

Pijanowski, B. C. , M. Bauer, K. Sawaya and B. Shellito. 2001. Using remote sensing to parameterize the Land Transformation Model for the Twin Cities. American Society of Photogrammetry and Remote Sensing (ASPRS) Meetings, St. Louis, Mo. April 2001. Internet: http://ltm.agriculture.purdue.edu/pubs/ASPRS 2001 FINAL.pdf

Pijanowski, B.C., D. G. Brown, G. Manik and B. Shellito. 2002. Using Neural Nets and GIS to Forecast Land Use Changes: A Land Transformation Model. Computers, Environment and Urban Systems. 26(6): 553 – 575. Internet: http://ltm.agriculture.purdue.edu/ceus.pdf

Pijanowski, B.C., B. Shellito, S, Pithadia and K. Alexandridis. 2002. Forecasting and assessing urban sprawl along in coastal watersheds along eastern Lake Michigan. Lakes and Reservoirs. 2002 7:271-285. Internet: http://ltm.agriculture.purdue.edu/Pijanowski.203.pdf

Stuttgart Neural Network Simulator: http://www-ra.informatik.uni-tuebingen.de/SNNS/

Authors Information

David Nouza
PAR Government Systems Corporation
Geospatial Software and Modeling
314 South Jay Street
Rome, New York 13440
dave_nouza@partech.com

Verne LaClair
PAR Government Systems Corporation
Geospatial Software and Modeling
314 South Jay Street
Rome, New York 13440
verne_laclair@partech.com

Matthew Schwab
New York City Department of Environmental Protection
71 Smith Avenue
Kingston, New York 12401
mschwab@dep.nyc.gov