Title: Quantifying decadal land use patterns in Macon County, NC, from 1885-2035 using historic imagery, census data, and a spatio-temporal interpolation model

Author: Ryan W. Kirk, Graduate Research Assistant, Department of Forest Resources, University of Minnesota, 1530 Cleveland Ave N, St. Paul, MN, 55108, phone: 612-624-4280, fax: 612-625-5215, e-mail: kirk0073@umn.edu

Abstract: This paper discusses preliminary results of an on-going land-use change study in Macon County, NC. Land use patterns at a one-hectare resolution were modeled for each decade between 1885 and 2035 using a variety of spatial and aspatial estimates of land use and an iterative, rule-based decision-tree model. Land use patterns of known dates were used to interpolate and extrapolate land use patterns using various combinations of simple logic rules, simple linear regression, and logistic regression. In all, it is estimated that Macon County lost over 18,000 ha of farmland over the past century, with a majority of that area being reforested and developed. Over the next 30 years, a slight loss of forest area is predicted as development pressures increase. Reconstructing historic land use provides a consistent land use data sequence for a vast array of analyses, a method of identifying errors in land-use classification schemes, and a framework for evaluating land use trajectories.

Introduction and Objectives:

Past land use data are crucial for understanding many current and future socio-natural dynamics (Foster, 1992). The modeling of land-use change has received substantial research attention in recent decades, yet uncertainty remains. Many studies that rely on two or a few dates of data do not capture the dynamic nature of land-use change. Similarly, the drivers of land-use change, both in a mechanistic and an empirical context, are not completely understood or fully incorporated into many studies. Furthermore, there is a distinct challenge in combining spatially-explicit land use data with aggregated estimates of the quantity of change. This project aims to improve spatio-temporal mapping through assessing and modeling 150 years of land-use change in the Southern Appalachian Mountains. The primary objectives are:

1) Reconstruct a logically consistent and detailed record of land-use change in the county over the past 120 years, and forecast change over the next 30 years.
2) Identify primary drivers of land-use change in the region.
3) Evaluate the use of a rule-based decision tree model for reconstructing and estimating land-use change in mountainous regions.

To meet these objectives, two general hypotheses were evaluated:

1) Terrain features (elevation, slope, aspect, distance to streams) and infrastructure features (distance to roads, distance to market centers, building densities) are significantly correlated with land-use categories and specific land-use trajectories in mountainous areas, although the extent of correlation varies over time for different categories.
2) Decadal historic land use patterns in mountainous areas can be reconstructed to at least 80% categorical accuracy based on sparse spatial data and more frequent ancillary data.

Background and Rationale:

Land-use change modeling frequently involves a two-step process. In the first, land-use change trajectories are defined through analysis of either existing land-use data or from research into the decision-making processes driving land use. Quantitative change detection analysis of land use data sets from two or more time periods is the most common approach applied (Coppin et al 2004). Second, these data are used to drive some model for estimating change. Manson et al (2005) identify seven classes of land-use change technical modeling approaches: equation, system, statistical, expert, evolutionary, cellular, and agent-based systems. Each of these approaches varies in complexity and applicability, but all are well established in the literature.

Several articles also highlight future trends in land-use modeling. In an editorial for a special issue of *Agriculture, Ecosystems and Environment*, Veldkamp and Lambin (2001) identify 4 key trends in land-use change modeling research:

1) "Modelling of drivers of land-use change;
2) modelling of scale dependency of drivers of land use change;
3) modelling progress in predicting location *versus* quantity of land-use change;
4) the incorporation of biophysical feedbacks in land-use change models."

The varieties of current land-use modeling methodologies have all provided valuable contributions to the field. This paper addresses two aspects of land-use change modeling not fully analyzed in other studies: 1) explore the relationship between land-use change and terrain features, and 2) reconstruct historic land use using spatial (historical photographs and maps) and aspatial (census) data. In addition, the development of a longer historical record (150 years) at a more detailed spatial and temporal resolution will provide insights into land-use change trajectories unavailable with studies relying on more sparse temporal or categorical data sets.

Study Area:

The study area is Macon County, NC, a 1300 km^2 county located in the Blue Ridge physiographic province of the Southern Appalachian Mountains (Figure 1). The area is characterized by steep topographic relief (550-1650 m elevation), mild winters, cool summers, abundant precipitation (130-180 cm/yr), and high primary productivity. The steep terrain combined with the relatively early European settlement has resulted in a unique land-use history, with agriculture expansion, forest harvesting, and development occurring at different periods and pacing than in other North American regions.

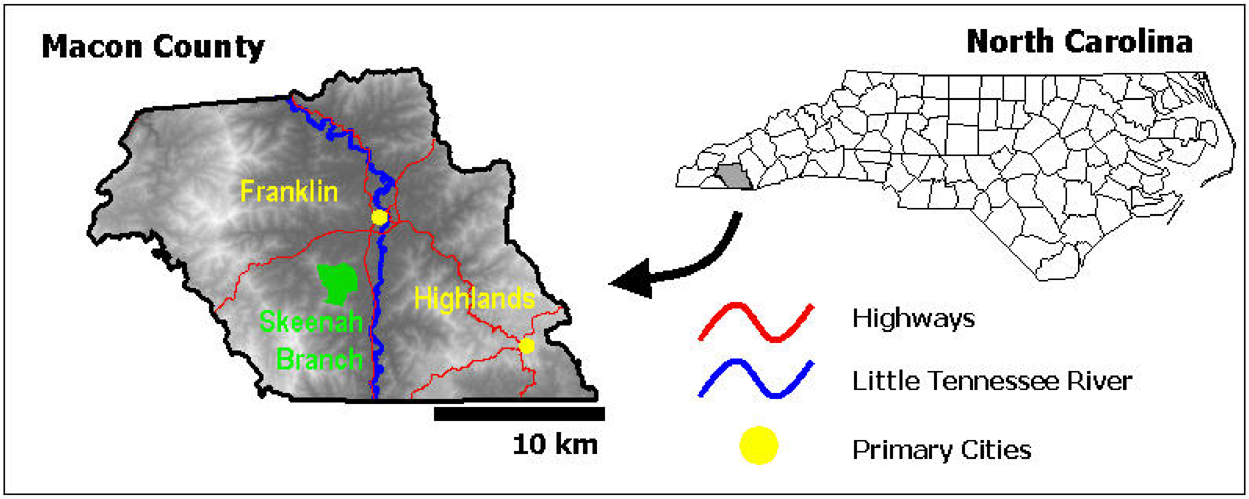


Figure 1: Nested study areas. Highways and primary roads are also included.

Macon County was formed in 1828, and the county boundary in its present form was finalized in 1886 (hence the starting date of this analysis). Between 1900 and 2000, forests increased from 74% to 90% of the total area, and agriculture acreage declined by 90% over the same period. Currently, the area faces increasing primary and secondary home development (Wear and Bolstad 1998; Cho and Newman 2005), increasing recreational activity, and steadily decreasing agricultural use (Figure 2). In this mountainous county and the surrounding region, terrain patterns provide clear drivers of and limitations on land-use change.

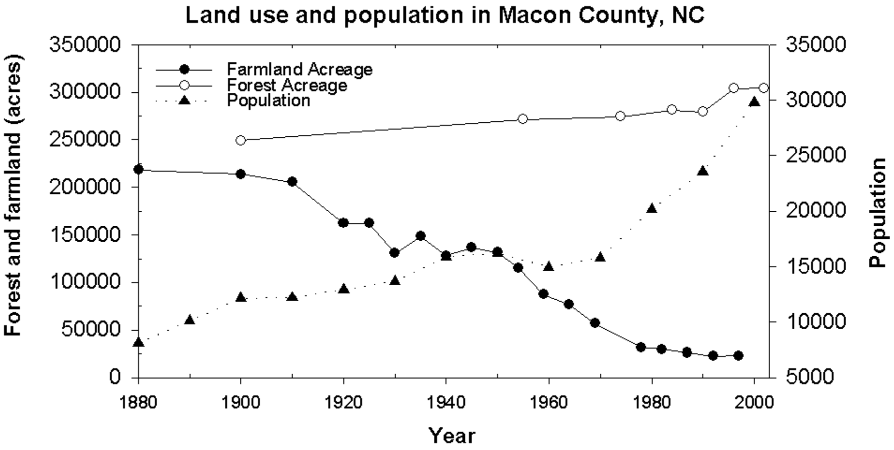


Figure 2: Macon County trends in forest and farmland area, and population.

Previous research in the area has identified several links between the mountain terrain and land use patterns. Bolstad et al (1998) accurately predicted forest composition using terrain variables. Wear and Bolstad (1998) used a simple terrain-based model to project building densities until 2003 with relatively high confidence. Turner et al (1996) concluded that landscape disturbance regimes are strongly influenced by terrain features and ownership. Similarly, Cho and Newman (2005) concluded that several terrain-related variables are key determinates of development probabilities. Separately, Gragson and Bolstad (2005) argue that the Southern Appalachians have fewer restraints on sprawl than other regions in the U.S., so estimating patterns and consequences is important for planning purposes. Combined, these positive research results clearly suggest a need for more detailed and consistent data sets for evaluating the effects of land-use on any number of ecological and economic patterns.

Methods:

The primary aim of this project is to develop a logically consistent land use data set at a relatively high level of spatial detail and over a long, continuous time period. To meet this objective, a GIS-based reconstruction of historic land use was conducted, and 30-year forecasts of change were estimated given recent trends and population projection estimates.

Two primary sets of data were collected and developed. First, historic spatial data of land-use was collected and assimilated into a GIS database. Second, aggregated, tabular estimates of land use patterns and population densities were compiled from various national and regional forest inventory summaries, and national and statewide agricultural and population censuses. It was postulated that spatial land use patterns in the mountainous county could be reconstructed for periods where no spatial data were available based on known landscape patterns and land-use change trends constrained by the tabular estimates.

A wide variety of historic spatial data were collected from a diverse range of sources for the entire county, with a more detailed temporal and categorical record being collected for Skeenah Branch, a 1500ha watersheds within the county. Building locations and densities were extracted from 1906, 1940 series and 1970 series USGS topographic maps (USGS 1906-1977), a 1927 soil survey maps, and orthorectified 1953, 1993, and 2004 aerial photographs. Road locations were collected from 1896 and 1906 USGS topographic maps, a 1927 soil survey map, 1953 aerial photographs, 1998 NCDOT data, and 2004 E911 program data. General forest area data were collected from a 1904 USGS report (Ashe and Ayers 1904), a 1960s USFS inventory map, and 1953 and 2003 aerial photographs. Detailed land cover was manually classified from 1953 and 2003 aerial photographs, and land cover classifications derived from Landsat images that were acquired from a 1993 Southern Appalachian Man and the Biosphere program (SAMAB, 1996) and a 2000 Coweeta Long-Term Ecological Research (LTER) program. In addition to these spatially-explicit data, aggregated land use data were collected from several U.S. Census of Population and Census of Agriculture reports, 5 separate USFS Forest Inventory reports dating to 1904, and regular North Carolina Department of Agriculture reports.

Raster-based modeling was conducted in two general phases. First, land use and land-use trajectory relationships to terrain variables (slope, elevation, terrain shape, and distance to stream) and infrastructural variables (distance to roads, travel distance to market centers, building density, and road density) were analyzed for the detailed watershed and countywide for periods with available data using post-classification change detection routines, simple linear regression, and logistic regression. Second, a series of decision tree models were developed that use the quantified drivers and available data as predictors.

Land-use trajectory relationships were evaluated using detailed reconstructions in the study watershed, Skeenah Branch (other watersheds were studied, but are not discussed here). Skeenah Branch is a dead-end watershed that feeds directly into the Little Tennessee River, which is the primary river in the region. Historic photos were acquired and orthorectified for each decade starting in the 1940s, and land use classes were manually identified via digitizing of vector data sets. Trends in change were quantified (Figure 3), such as a steady loss of pasturelands in the uplands, and steady densification of development near the watershed outlet.

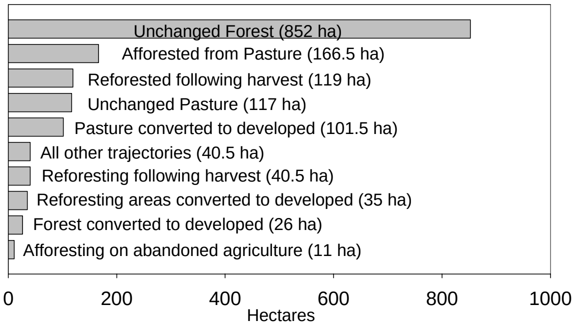


Figure 3: Land use trajectories in Skeenah Branch Watershed from 1940-2005.

This analysis was expanded to a countywide basis for selected data sets, including building density layers that were derived from the building point locations found on USGS topographic maps and historic soil survey maps. In the end, it was evident different land use classes could be most effectively modeled using unique predictor data sets or different logic rules. Figure 4 illustrates some of the key predictor variables.

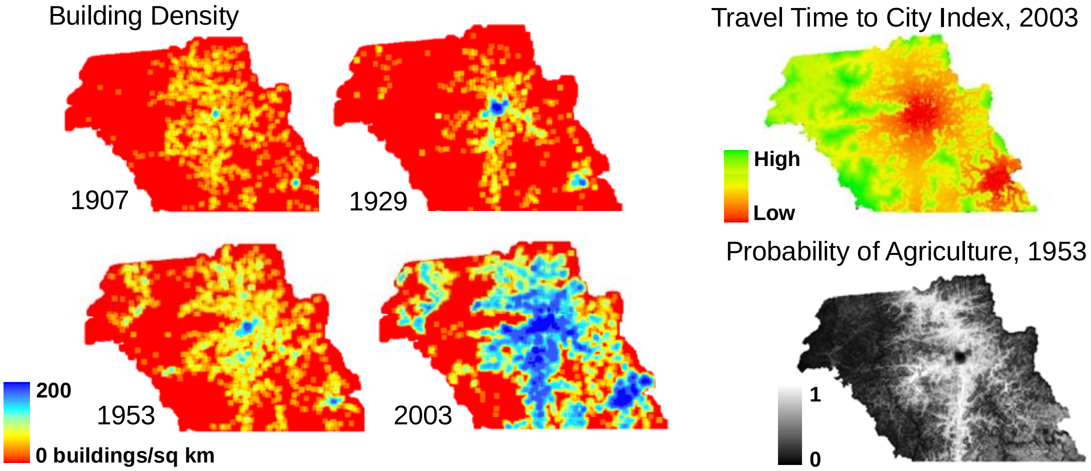


Figure 4: Key Predictors of land use and land use change. Probabilities such as found in the lower right image are built upon other important terrain and infrastructure attributes.

Once a general understanding of the land-use change trends were identified and quantified, the second phase was reconstructing and modeling the land use. Thus, the spatio-temporal interpolation "model" for this study consists of a combination of decision-rules and probability estimates using logic rules and logistic regression, each used based on a decision-tree selection criteria. Logic rules are defined by the sample data and from general trends identified in previous research studies. These rules either strictly force a specific land use change (e.g., once a cell was developed, it will not be changed again during the duration of this study) or provide a discrete choice (e.g., abandoned forest after 1925 will either reforest or be developed). Probability estimates are developed when no strict rule is available. Using the logic rule and probability estimates together, the total area of each land use class in the county can be constrained by the aggregate census and inventory estimates. In other words, the most likely cells for a given change are chosen first.

It was determined early in the project that modeling would be conducted in three phases that match epochal trends in the county's history. First, the period of most accurate spatial data (1955-2005) was reconstructed, which corresponds with the period of regular agriculture abandonment and steady population increase. Second, land use was "backcasted" for the oldest period (1945 to 1885). The peak agriculture period occurred during this period, population trends were relatively constant, and very little development occurred. With a relative dearth of spatial data, this backcasting period relied on logistic regression models to estimate spatial trends. Finally, the future period (2015 – 2035) was modeled using present trends, population projections, and a previous economic study that identified parcels with a high probability of being developed (Cho and Newman, 2005).

For the reconstruction of the known period (1955-2005), over 84% of the cells could be estimated using logic rules, in what effectively amounts to a process of eliminating unclassified cells. Table 1 shows the sequentially applied decision rules for the period.

Table 1: Logic rules applied to the fill in the known period (1955-2005)

Rule Order	**Description**
1	Once a cell is developed, it remains developed for the remainder of the period
2	A uniform water mask based on the 2005 data is applied to all periods.
3	If forested on the 1955 classification, the 1975 forest/non-forest classification, and the 2005 classification, then forested throughout the period.
4	If shrub in 1955 and shrub in 2005, then always shrub.
5	If on public land and transitional forest, then always returns to forest
6	If a cell is flagged as unchanging (e.g., airport, golf courses), then unchanged

For the remaining 16% of the cells, transition probabilities were calculated and weighted on select terrain variables. Based on those probabilities random cells were selected and assigned selected classes, with subsequent years getting updated accordingly. For example, if a transitional forest cell is identified for development, all subsequent decades are switched to developed.

The second phase of modeling consisted of backcasting the period of 1945-1885. For this period, the only spatial data available was building densities and road networks. Even given this lack of data, logic rules accounted for as much as 17% of the cells. These rules included the water mask, an "if forested and far from roads or on steep slopes, then remains forested" rule, and backcasting transitional forests into the agriculture that likely preceded it. Once logic rules were exhausted, a logistic model of the probability of agriculture was applied (Figure 4 above). The logistic model takes the generic functional form of:

Prob(ag) = f(elevation, slope, distance to nearest river, and building density of the most recent year, distance to nearest agriculture cell at previous time step)

The total area of farmland for a decade was determined by applying the percentage change of the U.S. Census of Agriculture (as summarized in Waisanen and Bliss 2002) estimate to the actual cell counts in the GIS data set. For example, the most accurate land cover classification in this study, the 1955 and 2005 manually digitized classifications, were both about 50% larger than the Census of Agriculture estimate. The probability of agriculture was iteratively increased until the desired number of cells was selected. Forest area and developed areas were filled in after the agriculture using standard transition probabilities.

During, the forecasting phase (2005-2035), up to 53% of the cells were classified using logic rules. Of the remainder, the first step consisted of identifying cells for development. Cho and Newman (2005) create an economic model to identify the probability that specific land parcels in the county would be developed. These probabilities were used to select a number of parcels proportional to the U.S. Census of Population projections. The remaining cells were filled using transition probabilities similar to the previous phases.

Data management and visualization was completed with ArcGIS 9.x. Exploratory modeling was conducted in IDRISI Kilimanjaro. Final modeling and analysis used Python scripting in conjunction with the ArcGIS 9.x Geoprocessor model and the R, GDAL, ShapeLib, and numeric Python (NumPy) Open-Source utilities.

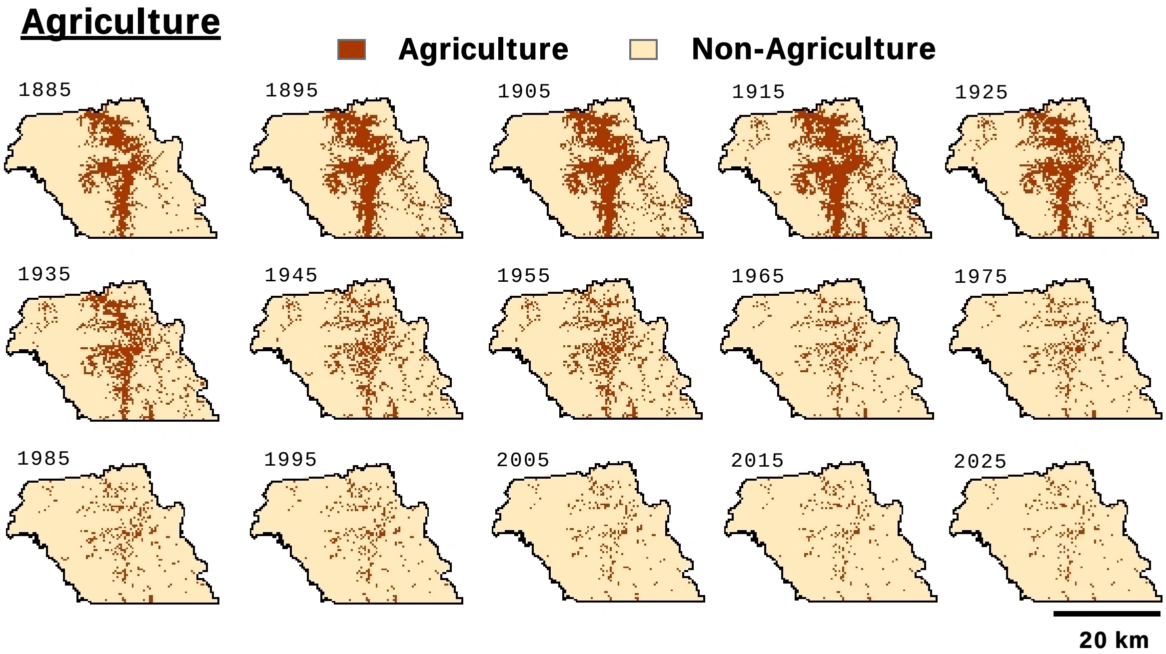
Results:

Based on the iterative model, land use patterns for 6 generalized land use classes were modeled for each decade between 1885 and 2035 (Table 2). In general, developed areas steadily increased over the previous century and are expected to continue over the next 30 years.

Table 2: Cell Counts (1 cell = 1 ha) across all years and class.

Year	Forest	Agriculture	Developed	Transitional Forest	Shrub	Water
1885	109513	24060	81	0	0	986
1895	101843	31730	81	0	0	986
1905	100041	33532	81	0	0	986
1915	94436	39131	88	0	0	985
1925	94426	31486	98	7645	0	985
1935	94413	25738	111	13393	0	985
1945	101774	19370	126	12197	188	985
1955	109976	15612	175	6540	1351	986
1965	111113	11314	175	8378	2674	986
1975	113594	10077	756	6143	3084	986
1985	117817	8801	2532	867	3637	986
1995	117130	7392	4077	1326	3729	986
2005	115285	5928	6587	2481	3373	986
2015	115122	5656	8260	1313	3303	986
2025	114576	5296	10410	184	3188	986
2035	112821	4880	12660	194	3099	986

Forest area declined at the beginning of the 20th century, but as agriculture started declining, large areas of reforestation occurred. Transitional forest (defined as abandoned pastures or fields) declined drastically. In all, it is clear that land-use change in Macon County has been dynamic and non-linear for all 6 land-use classes considered. Figure 5 illustrates the 150-year change for two of the classes, and figure 6 contains full classifications for 3 sample years.


Agriculture

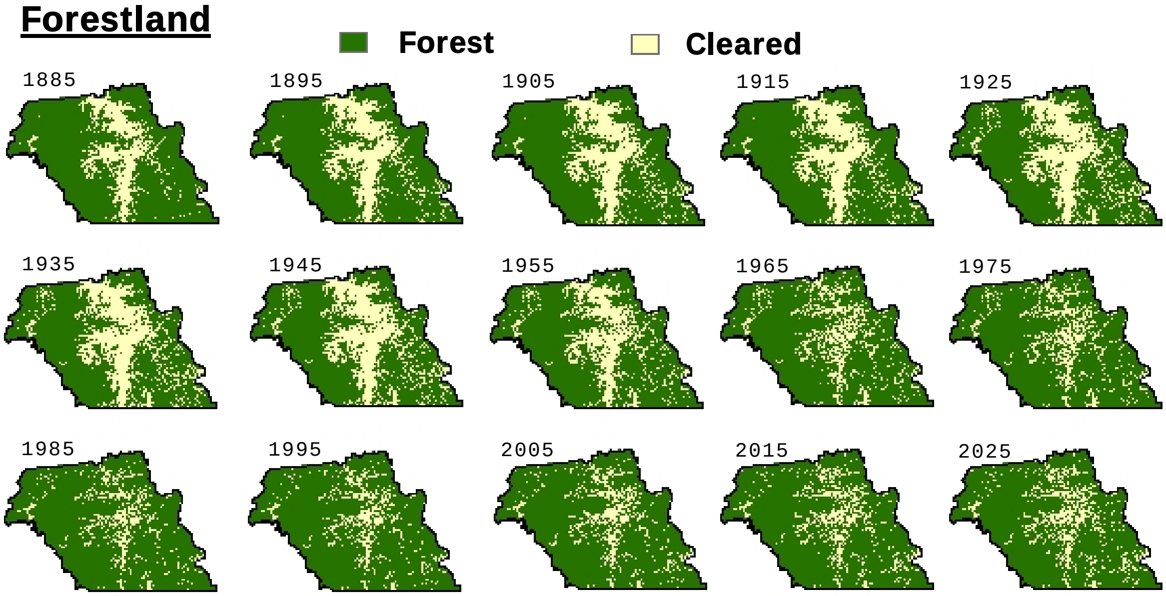

Forestland

Figure 5: Complete sequence estimates for two of the six final land use classes.

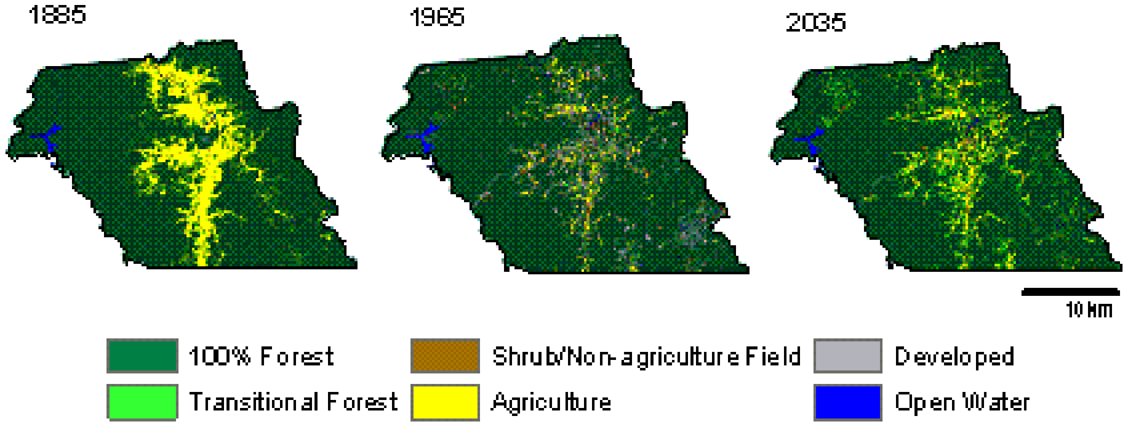


Figure 6: Complete classification for select years. Note that there was no spatial data available for these particular dates.

Formal accuracy assessments are presently being conducted as part of this on-going work. Anecdotally, however, several of the key trends of the Skeenah Watershed were captured with the historic reconstruction. Agriculture was correctly estimated for the main valley in the watershed in the early half of the 20th century, and then was abandoned, starting with the steeper slopes. The model also predicts that a majority of the watershed will be developed in the next 30 years, a trend that has already begun. Thus, initial assessment is positive.

Discussion:

In all, this multi-phase, iterative modeling effort relied heavily on the detailed understanding of the land-use change processes in the region and in the realization that mountainous terrain has a very strong influence on landscape patterns. Where reconstruction and logical consistency are important, a decision-rule model appears quite valuable. For situations where such terrain constraints aren’t available, or for identifying driving factors across regions, a less deterministic model might be warranted.

While GIS proved to be a valuable framework for analysis, several GIS-based modeling implications were identified. The large variety of input data sources, scales, and accuracy provided distinct challenges for analysis. Total forest area measurements of the most accurate spatial data differed by as much as 7% from the inventory-based estimates from the same period. Similarly, the census estimates of agricultural areas differed from the mapped area by up to 10,000 ha for the same year. This is clearly one of the major challenges in using spatially-aggregated and sampling-based estimates of area in a spatially-explicit context. In this study, the mapped areas were deemed to be “truth”, and the aggregate data were used for the proportional change between decades rather than exact numbers for a specific decade. This approach of proportional selection appeared to effectively capture the epochal trends.

Another big area of concern was the large differences between the spatial data. Satellite-based classification and digitizing-based classifications focused on fundamentally different factors. For example, it is common with digitizing to split roads down the center, effectively ignoring the impervious surface. Pixel-based satellite imagery classification algorithms, on the other hand, are strongly influence by bright linear features such as roads. In addition, several of the historic maps that were not cartometrically correct, such as the 1904 USGS Ashe/Ayers map, were not usable in the spatial modeling. However, they were valuable for understanding the general trends of land-use change. This information is valuable in defining decision rules based on “expert” knowledge.

A final issue relating to the disparate data sources and accuracies is the spatial, temporal, and categorical detail. In this study, all land use data were generalized into 6 broad classes. Aggregation is clearly the most common solution to overcoming the scale-mismatch issues. While the categorical detail was coarsened here, this study was able to produce a moderately high level of spatial detail (one hectare cells across a 134,000 ha county) and a nearly unheard of temporal sequence (16 straight decadal classifications). This was due in part to the high level of detail in the core data sets, to the breadth of data acquired, and to the identifiable trends of the specific region.

In all, the aim of this project was to reconstruct a logically consistent land cover data set over a long temporal period in order to facilitate a variety of natural resource-related analyses. As such, capturing the core trends of the dynamic land-use change were paramount. Overall, a decadal scale analysis appeared appropriate for capturing those key epochal trends while maintaining an acceptable level of spatial, categorical, and modeling error.

Acknowledgments: Data development and research funds were provided as part of the NSF-sponsored Coweeta Long Term Ecological Research (LTER) program.

References:

Agarwal, C., G. M. Green, J. M. Grove, et al. 2002. A review and assessment of land-use change models: dynamics of space, time, and human choice. USDA U.S. Forest Service Northeastern Research Station General Technical Report NE-297.

Ayers H. B., and W. W. Ashe. 1905. The Southern Appalachian forests. USGS. Washington, DC.

Bolstad, P. V., W. Swank, and J. M. Vose. 1998. Predicting Southern Appalachian overstory vegetation with digital terrain data. *Landscape Ecology* 13:271-283.

Cho, S.H., and D. H. Newman. 2005. Spatial analysis of rural land development. *Forest Policy and Economics* 7: 732-744.

Coppin, P., I. Jonckheere, K. Nackaerts, B. Muys. 2004. Digital change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sensing*. 25(9): 1565-1596.

Delcourt, H. R. and W. F. Harris. 1980. Carbon Budget of the Southeastern U.S. Biota: Analysis of Historical Change in Trend from Source to Sink. *Science*. 210:321-323.

Foster, D. R. 1992. Land-use history (1730-1990) and vegetation dynamics in central New England, USA. *Journal of Ecology*. 80:753-772.

Gragson, T. L., and P. V. Bolstad. 2006. Land use legacies and the future of Southern Appalachia. *Society and Natural Resources* 19:175-190.

Irwin, E. G., and J. Geoghegan. 2001. Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems and Environment*. 85: 7-23.

Kline, J. D. 2003. Characterizing Land Use Change in Multidisciplinary Landscape-Level Analyses. *Agricultural and Resource Economics Review* 32:103-115.

Kline, J. D. 2005. Predicted future forest- and farmland development in western Oregon with and without land use zoning in effect. USDA U.S Forest Service Pacific Northwest Research Station Research Note PNW-RN-548.

Manson, S. M., B. L. Turner, II and J. Geoghegan. 2005. State of the art in describing future changes in ecosystem services: forecasting land cover change. In Millennium Ecosystem Assessment: Scenarios Working Group Report. eds. H. Mooney and A. Cropper. Geneva, Switzerland, Island Press.

[SAMAB] Southern Appalachian Man and the Biosphere. 1996. The Southern Appalachian assessment. Social/cultural/economic technical report: report 4 of 5. Atlanta: U.S. Department of Agriculture, Forest Service, Southern Region.

Turner, M. G., D. N. Wear, and R. O. Flamm. 1996. Public and private disturbance regimes in the Southern Appalachians. *Ecological Applications* 6:1150-1172.

U.S. Geological Survey. Various Maps. 1:24,000. 7.5 Minute Series. Washington D.C.: USGS, 1892-1991.

Veldkamp, A., and E. F. Lambin. 2001. Predicting land-use change. *Agriculture, Ecosystems and Environment*. 85: 1-6.

Waisanen, P. J. and N. B. Bliss. 2002. Changes in population and agricultural land in conterminous United States counties, 1790 to 1997. *Global Biogeochemical Cycles* 16:1137.

Wear, D. N., and P. V. Bolstad. 1998. Land use changes in southern Appalachian landscapes: spatial analysis and forecast evaluation. *Ecoystems* 1:575-594.

Wear, D. N., and R. O. Flamm. 1993. Public and private disturbance regimes in the Southern Appalachians. *Natural Resources Modelling* 7:379-97.

Wear, D. N. and J. G. Greis. 2002. *Southern forest resource assessment: Summary report.* Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Statement.