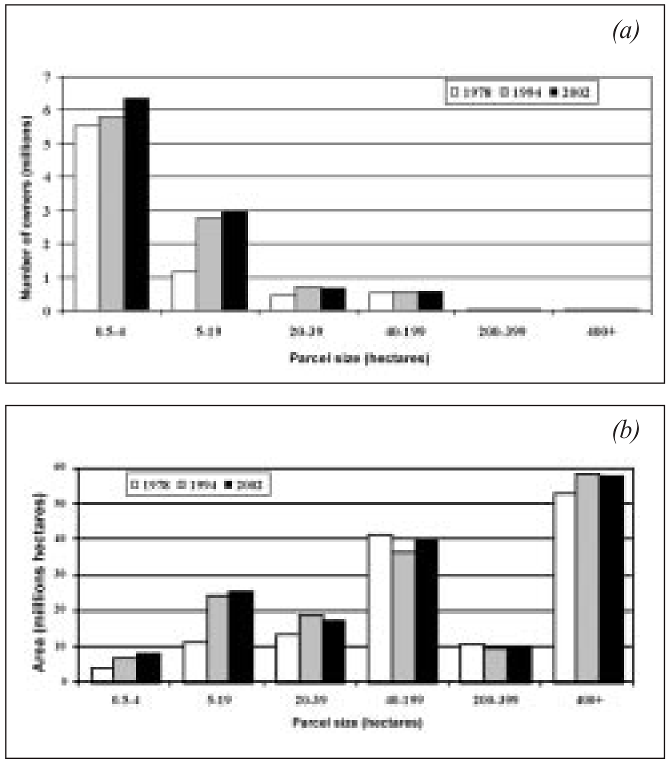
Assessment and Mapping of Forest Parcel Sizes

Brett J. Butler and Susan L. King[1]

Abstract.—A method for analyzing and mapping forest parcel sizes in the Northeastern United States is presented. A decision tree model was created that predicts forest parcel size from spatially explicit predictor variables: population density, State, percentage forest land cover, and road density. The model correctly predicted parcel size for 60 percent of the observations in a validation data set (weighted kappa = 0.45). This decision tree model was used to create a map representing the average forest parcel size across the region.

Figure 1.—*Size of forest landholdings as a function of (a) number of owners and (b) area of forest land owned in the United States in 1978, 1994, and 2002.*



Introduction

Our Nation's forest resources are becoming increasingly constrained by social factors. Wear *et al.* (1999) found a negative relationship between population density and harvesting probability, with harvesting probabilities approaching zero as population densities approached 58 people/km^2 (150 people/mi^2). One factor correlated with harvesting activity is the area of forest land owned by individual forest landowners. Preliminary results from the National Woodland Owner Survey (NWOS) (Butler and Leatherberry 2003) suggest that parcel size also is correlated with ownership objectives, management activities, and future land use intentions.

Over the past 25 years, the number of private forest landowners in the United States increased from 7.8 million to 10.6 million (Birch *et al.* 1982, Butler and Leatherberry 2003). Most of this increase comes from landowners who own less than 20 ha (50 ac) of forest land (fig. 1). This increase in the number of landowners with smaller parcels is known as forest parcelization. As Sampson and DeCoster (2000) stated, forest parcelization is a major factor leading to "a growing crisis in maintaining sustainable private forests" (Samson and DeCoster 2000, 6). The financial difficulties of managing smaller parcels of forest land (i.e., economies of scale) (Row 1978) are partially responsible for the loss of working forest lands; the changing characteristics of the landowners (Egan and Luloff 2000), including changing ownership objectives and management practices, is another contributing factor.

To date, most information about forest parcel size across broad geographic regions has been produced in a tabular format (e.g., Birch 1996). Information about forest parcel size could be improved by providing the data in a spatially explicit (i.e., map) format to enable visual inspection of spatial patterns and combining with other spatial data sources to conduct further analyses. Estimation techniques must be employed because no nationally available spatially explicit data sources on forest parcel size exist.

[1] Research Forester and Operations Research Analyst, respectively, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, 11 Campus Boulevard, Suite 200, Newtown Square, PA 19073. Phone: 610–557–4045; fax: 610–557-4250; e-mail: bbutler01@fs.fed.us.

By combining point-based estimates of forest parcel size with spatially explicit ancillary data, we model forest parcel size across the Northeastern United States and generate a forest parcel map. In previous research (King and Butler, in press), we explored linear regression, neural network, and decision tree techniques for generating this model. We found the linear regression models had low predictive power (i.e., poor R^2 values) and required transformations of multiple variables, including the dependent variable. The model building process associated with the neural networks was complex and difficult to repeat, and the results were difficult to interpret. The decision tree models were easy to implement, produced results that were easy to interpret, and appeared to have relatively high predictive power. Because of these factors, we opted to use decision tree models for this study. The accuracy and goodness-of-fit of the decision tree modeling technique are tested using a validation data set and quantified using a confusion matrix and a weighted kappa statistic. Potential to improve these methods and future research directions also are discussed.

Methods

To generate a forest parcel map for the Northeastern United States, we created a decision tree model that predicted parcel size as a function of a set of independent variables. The Northeastern United States was selected because of data accessibility. Similar data have been collected for all the contiguous States, and the methods employed in this study can be expanded to this broader area.

The dependent variable was derived from the U.S. Department of Agriculture Forest Service's Forest Inventory and Analysis program's NWOS program (Butler and Leatherberry 2003). The NWOS randomly selects private forest landowners from across the United States and collects information, including size of landholding. The landowners were selected by assessing the forest cover at a random set of points across the United States; for those points determined to be located in forested areas, the ownership of record was determined through tax records. For this study, we used 764 points representing landowners surveyed in 2002 in the Northeastern United States.

The sizes of the forested landholdings were grouped into three categories: 0.5–19 ha (1–49 ac), 20–399 ha (50–999 ac), and greater than or equal to 400 ha (≥ 1,000 ac). These categories are a simplified version of the categories used in previously published research (e.g., Birch *et al.* 1982, Birch 1996, Butler and Leatherberry 2003). The more detailed groups reported in earlier reports are collapsed to aid in producing results that are more accurate, more consistent, and easier to interpret.

Predictor variables needed to have a theoretical relationship to forest parcel size, be spatially explicit, and be publicly available. The independent variables were percentage forest land cover, percentage urban land cover, percentage agricultural land cover, population density, housing density, change in population levels, distance to roads, distance to U.S. Census-defined cities and places, road density, a population gravity model, and State. Land cover data were summarized over 6.25 ha (15.5 ac) areas based on the Multi-Resolution Land Characteristics Consortium's National Land Cover Data (Vogelmann *et al.* 2001). Population, housing, city, and place data were derived from U.S. Census data (U.S. Commerce Department 2000). Population and housing density were summarized for each U.S. Census-defined block group. Road data were obtained from Bureau of Transportation Statistics (U.S. Department of Transportation 1998). Parcels closer to urban areas are more likely to be influenced by urban land use processes; we included a gravity model that quantifies the "influence" of these population centers across the landscape (equation 1) (Kline and Alig 2001).

$$\text{Gravity Index}_i = \sum_{j=1}^{3} \frac{(\text{Population}_k)^{0.5}}{\text{Distance}_{ij}} \tag{1}$$

where:

i = sample point; and

j = the three cities having the largest influences, as defined by this equation.

Due to high Pearson correlation coefficient values with other predictor variables, the percentage agricultural land cover and housing density variables were excluded in favor of other variables that had higher predictive powers and more direct interpretations.

Each independent variable not in the proper format was converted into a surface or grid layer and reprojected to an Albers projection using Geographic Information System (GIS) software (Environmental Systems Research Institute 2001). The value for each independent variable at each sample point with known forest parcel size was extracted using the GIS software. The resulting data set formed the basis for building and validating the decision tree model.

The decision tree was generated using the Chi-squared Automatic Interaction Detector algorithm (Kass 1980; SPSS 2002). Seventy-five percent (573) of the data points were used to develop the model; 25 percent (191) of the data points were used to validate the model. The validation data were used to create a confusion matrix and generate a weighted kappa statistic (κ) (Congalton and Green 1999). The confusion matrix displays the observed versus predicted parcel sizes for each point. The κ statistic condenses the confusion matrix into a single statistic and represents the proportion of agreement beyond chance with 1 representing perfect agreement, 0 representing no agreement, and a negative value indicating agreement less than expected.

The decision tree model will vary depending on the data used to train the model. We produced multiple models by randomly dividing the data set into subsets of various training and validation groups. Although population density was consistently selected as the first variable to split the data set, and overall model accuracy did not vary appreciably, subsequent splitting of variables varied among models. The model presented herein was selected because of its relatively high predictive power and the inclusion of theoretically consistent variables. In subsequent efforts, we will investigate techniques for producing a final model based on the convergence of multiple models.

The final decision tree model was used to generate a parcel map by translating the model results into a series of if-else logic statements. The logic statements were applied on a pixel-by-pixel basis using a GIS software package.

Results

The final decision tree model (fig. 2) included population density, State, percentage forest, and road distance variables to predict forest parcel size. At the highest level, population density was the best predictor of forest parcel size, with areas of higher population density more likely to have smaller forest parcels. When the population level was between 6 and 18 people/km^2, the State variable was the best predictor of whether the forest parcel would be large (≥ 400 ha) or medium (20–399 ha). At population densities greater than 18 people/km^2, the percentage of the area covered by forest was indicative of whether the forest parcel was medium or small (0.5–19 ha).

The decision tree model was translated into a series of 13 if-else logic statements. By applying these statements to the spatially explicit predictor variables, we produced a map (fig. 3).

This decision tree model predicted 60 percent of the validation data correctly (table 1) yielding a weighted kappa statistic of 0.45 [95-percent interval = (0.35, 0.55)]. Of the misclassification errors made, 88 percent were misclassification of points to one of the adjacent categories.

Figure 2.—*Decision tree model of forest parcel sizes in the Northeastern United States. Population density is in units of people/km^2, and road distance is measured in meters.*

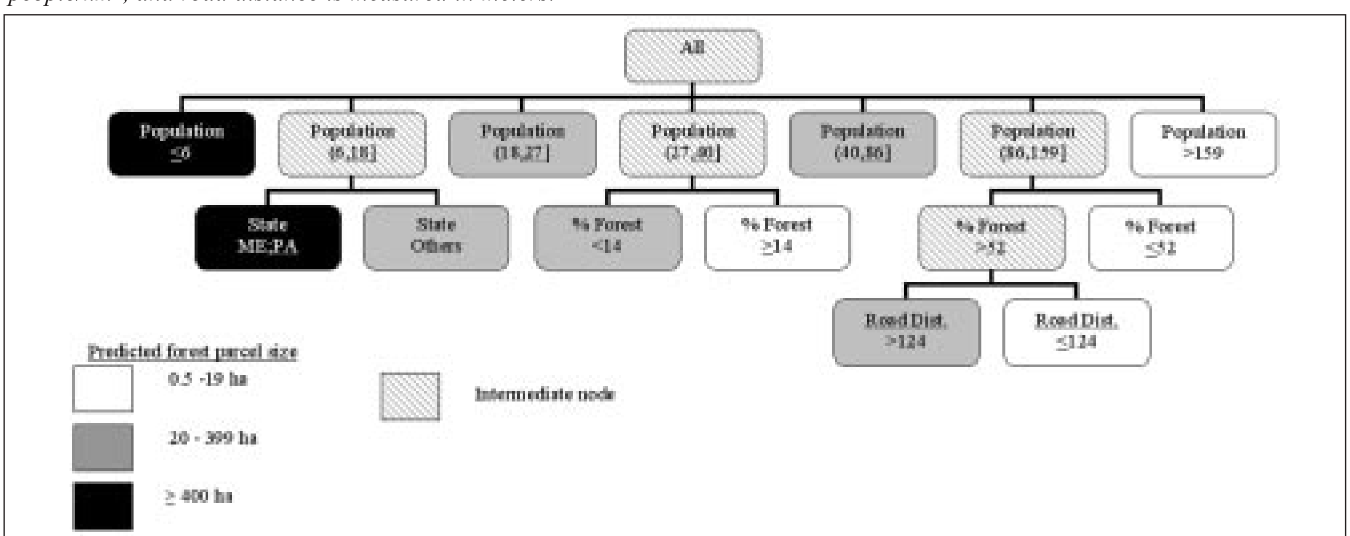
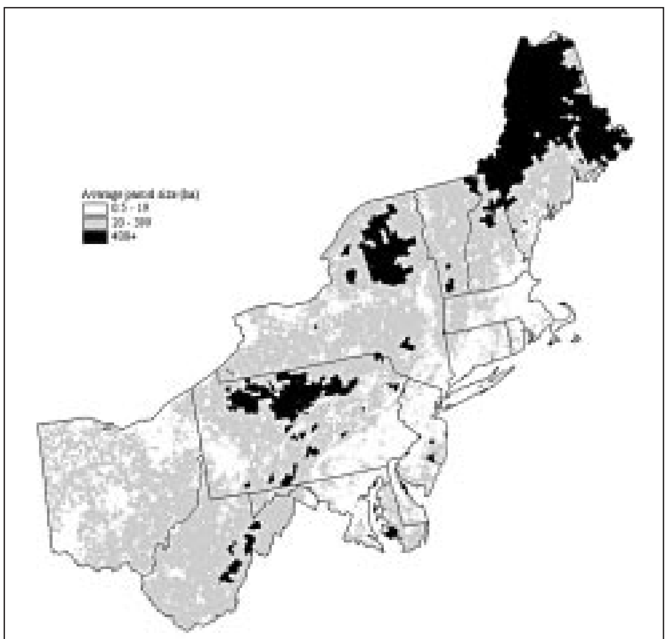


Table 1.—*Confusion matrix representing observed and predicted ownership categories based on a closest-neighbor estimation technique (numbers in parentheses represent column percentages).*

Predicted	**Observed**						
	Federal	**State**	**Local**	**Forest industry**	**Other corporate**	**Family**	**Total**
Federal	42	3	2	1	7	23	78
	(56.8)	(1.0)	(2.7)	(0.2)	(1.5)	(1.1)	(2.3)
State	0	139	3	20	26	102	290
	(0.0)	(45.1)	(4.1)	(4.9)	(5.7)	(4.9)	(8.5)
Local	1	8	6	4	11	25	55
	(1.4)	(2.6)	(8.1)	(1.0)	(2.4)	(1.2)	(1.6)
Forest industry	0	13	3	251	36	65	368
	(0.0)	(4.2)	(4.1)	(61.5)	(7.9)	(3.1)	(10.8)
Other corporate	2	21	11	40	157	204	435
	(2.7)	(6.8)	(14.9)	(9.8)	(34.6)	(9.7)	(12.7)
Family	21	111	35	80	183	1391	1821
	(28.4)	(36.0)	(47.3)	(19.6)	(40.3)	(66.1)	(53.2)
Nonforest	8	13	14	12	34	293	374
	(10.8)	(4.2)	(18.9)	(2.9)	(7.5)	(13.9)	(10.9)
Total	74	308	74	408	454	2103	3421

Figure 3.—*Average size of forest parcels in the Northeastern United States as predicted by a decision tree model.*



Conclusions

The final decision tree model (fig. 2) makes sense intuitively. This model found population density to be the most powerful predictor of parcel size and shows the expected negative relationship. Having larger parcel sizes predicted in Maine and Pennsylvania versus the other States is logical considering the strong forest industries in those States and the States' development and population patterns. Percentage forest had the expected relationship with parcel size with areas of more forest being more likely to have larger forested parcels at a given population level. Road distance, although only of tertiary importance in the model, also exhibited the expected relationship to parcel size; areas of higher road density were more likely to have smaller forested parcels.

The resulting forest parcel size map (fig. 3) also seems reasonable. The areas with the highest probabilities of having larger forested parcels are located in areas with higher concentrations of forest industry or large public landholdings.

Northern Maine is an example of the former, and north-central Pennsylvania and the Adirondack region of New York are examples of the latter.

In addition to refining the decision tree model, we want to explore some geospatial modeling techniques, such as indicator kriging, before we attempt to implement this study on a broader geographic scale. Further refinement of accuracy assessment techniques will also be explored, including error/probability mapping.

Forest parcel size is an omnipresent factor influencing how and why land is used. Modeling of forest parcel size will yield more insight into the forest parcelization phenomenon and the impact of these trends on forest fragmentation, harvesting practices, and other trends.

Literature Cited

Birch, T.W. 1996. Private forest-land owners of the United States, 1994. Resour. Bull. NE-134. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 183 p.

Birch, T.W.; Lewis, D.G.; Kaiser, H.F. 1982. The private forest-land owners of the United States. Resour. Bull. WO-1. Washington, DC: U.S. Department of Agriculture, Forest Service. 64 p.

Butler, B.J.; Leatherberry, E.C. 2003. Preliminary results from the national woodland owner survey, 2002. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. http://www.fs.fed.us/woodlandowners.

Congalton, R.G.; Green, K. 1999. Assessing the accuracy of remotely sensed data: principles and practices. New York: CRC Press. 137 p.

Egan, A.F.; Luloff, A.E. 2000. The exurbanization of America's forests: research in rural social science. Journal of Forestry. 98(3): 26–30.

Environmental Systems Research Institute, Inc. 2001. ArcInfo, version 8.3. Redlands, CA: Environmental Systems Research Institute.

Kass, G.V. 1980. An exploratory technique for investigating large quantities of categorical data. Applied Statistics. 29(2): 119–127

King, S.L.; Butler, B.J. [In press.] Generating a forest parcel map for Madison County, NY. In: Beavers, M., Barrett, T.M., eds. Proceedings, 2003 systems analysis in forest resources symposium; 2003 October 7–9; Stevenson, WA. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Kline, J.D.; Alig, R.J. 2001. A spatial model of land use change in western Oregon and western Washington. Res. Pap. PNW-RP-528. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 24 p.

Row, C. 1978. Economies of tract size in timber growing. Journal of Forestry. 76(9): 576–582.

Sampson, R.N.; DeCoster, L. 2000. Forest fragmentation: implications for sustainable private forests. Journal of Forestry. 98(3): 4–8.

SPSS, Inc. 2002. SPSS Answer Tree, version 3.1. Chicago, IL: SPSS.

U.S. Department of Commerce, Census Bureau. 2000. Census 2000 TIGER/Line Data. Washington, DC: U.S. Department of Commerce, Census Bureau. http://www.census.gov/geo/www/tiger/. [Date accessed unknown.]

U.S. Department of Transportation, Bureau of Transportation Statistics. 1998. Geographic data technology Dynamap/1000 street shapefiles, version 9.0. Washington, DC: U.S. Department of Transportation, Bureau of Transportation Statistics. http://www.bts.gov/programs/geographic_information_services/download_sites/gdt/maindownload.html. [Date accessed unknown.]

Vogelmann, J.E.; Howard, S.M.; Yang, *et al.* 2001. Completion of the 1990s national land cover dataset for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering & Remote Sensing. 67: 650–652.

Wear, D.N.; Liu, R.; Foreman, J.M.; Sheffield, R.M. 1999. The effects of population growth on timber management and inventories in Virginia. Forest Ecology and Management. 118: 107–115.