Available online at www.sciencedirect.com
SCIENCE DIRECT
Remote Sensing of Environment 103 (2006) 56–66
Remote Sensing of Environment
www.elsevier.com/locate/rse

A model-based approach to estimating forest area

Ronald E. McRoberts

St. Paul, MN, USA

Received 23 July 2004; received in revised form 24 March 2006; accepted 25 March 2006

Abstract

A logistic regression model based on forest inventory plot data and transformations of Landsat Thematic Mapper satellite imagery was used to predict the probability of forest for 15 study areas in Indiana, USA, and 15 in Minnesota, USA. Within each study area, model-based estimates of forest area were obtained for circular areas with radii of 5 km, 10 km, and 15 km and were compared to design-based estimates based on inventory plot data. Precision estimates for the circular areas were also obtained using variance formulae developed for this application that incorporated spatial correlation among model predictions for individual pixels. The model-based estimates were generally comparable to the design-based estimates. The advantages of the model-based approach are that maps and small areas estimates may be obtained and the necessity of releasing exact plot locations for user-specific applications is alleviated.

Keywords: Logistic regression model; Landsat Thematic Mapper; Indiana, USA; Minnesota, USA

1. Introduction

Traditionally, large-scale, natural resource inventory programs have used data collected from ground plots to respond to the user question “How much?” by reporting plot-based estimates of natural resource attributes for states or provinces, counties, or municipalities. Increasingly, users are also asking “Where?” and are requesting access to plot data for estimation for their own areas of interest (AOI). When data requests do not require exact plot locations, there are few constraints on data access. However, if exact locations are required, then several issues must be considered. First, revealing exact locations may entice users to visit the plots to obtain additional information, thus artificially disturbing the sampling location and contributing to bias in inventory estimates. Second, plots may be located on private land, and while landowners usually permit access by inventory field crews, they generally prohibit additional access. In these situations, user visits to plot locations may jeopardize future access by inventory field crews. Third, revealing the exact plot locations may violate constraints on the release of proprietary information. Thus, if exact plot locations are required for a user’s analysis, policy constraints may prohibit the inventory program from accommodating the user’s data request.

In response to the “Where?” question, the Forest Inventory and Analysis (FIA) program of the USDA Forest Service has initiated local, regional, and national mapping efforts. Although the objectives of these efforts have been to map the spatial distributions of forest attributes, generally they have not included investigations of whether maps may be used to obtain unbiased and precise areal estimates of those attributes. For the latter objective, estimates obtained using only data for the plots located in the AOI have been necessary. If unbiased and sufficiently precise areal estimates of forest attributes could be obtained by aggregating mapping unit predictions, then several advantages would accrue. First, release of plot locations would be unnecessary for estimation for users’ AOIs. Second, because mapped values would be based on aggregated data from multiple plots, proprietary information would not be released. Third, estimation would be possible for small areas for which the number of plots is insufficient for plot-based estimation. Fourth, efficiencies would be gained by simultaneously addressing both the “How much?” and “Where?” questions.

Nelson et al. (2005) compared forest area estimates based on sample plot data from two national inventories of the USA to comparable estimates from four satellite image-derived maps.

E-mail address: mcroberts@fs.fed.us.

0034-4257/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2006.03.005

The sample plot data were collected for non-federal lands by the National Resources Inventory (NRI), a program of the USDA Natural Resources Conservation Service, and for all forestland ownerships by the FIA program. The satellite image-derived maps included a nominal 1991 AVHRR forest cover type map of 1-km spatial resolution (Zhu & Evans, 1994), a nominal 1992–1993 AVHRR land cover map of 1-km spatial resolution from the National Atlas (2003), the Vegetation Continuous Fields (VCF) percent tree canopy cover data obtained from MODIS imagery with a 500-m spatial resolution (Hansen et al., 2003), and the National Land Cover Dataset (NLCD) obtained from nominal 1992 Landsat TM imagery with 30-m spatial resolution (Vogelmann et al., 2001). They found that the forest cover type map produced non-federal forest area estimates that were most similar to NRI and FIA estimates and that for the VCF data, a minimum canopy cover threshold of 25% produced national estimates for all ownerships that were most similar to FIA estimates. However, they found that the 25% threshold produced large deviations between state-level VCF and FIA estimates. They concluded that it would be inappropriate to use forest/non-forest maps created from VCF products to estimate forest area for states or smaller geographic areas.

Nelson et al. (2005) also reviewed and reported European comparisons of forest inventory and satellite imaged-derived estimates of forest area (Häme et al., 2001; Kennedy & Bertolo, 2002; Päivinen et al., 2001; Schuck et al., 2003). Their conclusion was similar to that for comparisons in the USA: While satellite image-derived estimates may be acceptable for large geographic areas, they have limited utility for smaller geographic areas. Although the European forest map was calibrated to match countrywide inventory estimates of proportion forest area, and the forest cover types for the USA were mapped on pixels with forest density estimates exceeding per-state thresholds based on inventory estimates, neither map was specifically designed to produce estimates of forest area, particularly for small areas.

Many investigators have mapped forest cover using classification techniques and coarse spatial resolution AVHRR data. Mayaux and Lambin (1995) note that the advantages of the coarser resolution maps are greater data availability and a spatial resolution that more closely matches large AOIs, while the disadvantage is a loss of spatial detail for smaller areas. Mayaux and Lambin (1995), Czaplewski and Catts (1992), and Walsh and Burk (1993) describe methods for correcting for misclassification bias.

Kennedy and Bertolo (2002) used AVHRR data with a maximum likelihood approach to calculate the probability of forest. They used unsupervised classification to select clusters of homogeneous 2×2 AVHRR pixel blocks and the Coordination of Information on the Environment (CORINE) data set, a Landsat image-based interpretation of land cover for computing forest area, to train the 2×2 pixel blocks. Inventory-based estimates of forest area were used to calibrate pixel values so that resulting map-based estimates matched inventory-based estimates for counties, regions, and provinces. They predicted forest area proportion for each image pixel as a weighted average of observed proportion forest for the classes where the weights were probabilities of class membership obtained from maximum likelihood analyses.

Alternative approaches to area estimation have included both mixture modeling and regression analyses (Häme et al., 2001). DeFries et al. (2000) used a linear mixture model approach to predict continuous fields of land cover categories from end member classes derived from Landsat Multispectral Scanner system data. Thomas et al. (1993) used a goodness of fit approach to investigate the number of ground cover components and spatial averaging for estimating woodland area by spectral mixing. Cross et al. (1991) identified spectral signatures of pure forest and non-forest cells for coarse resolution imagery and then used a linear mixture model to predict proportions of forest and non-forest by decomposing the spectral values of mixed resolution cells into signature components.

Both Magnussen et al. (2000) and Moisen and Frescino (2002) compared multiple approaches for predicting forest attributes. Magnussen et al. (2000) evaluated predictions of forest cover type proportions obtained from a maximum likelihood classifier and three models using predictors based on proportions of TM clusters obtained from unsupervised classification. Among the models, one was based on neural networks and two were variations of linear models. The neural networks approach yielded lowest mean absolute deviations, while the maximum likelihood approach was better for predicting non-vegetated cover types. Moisen and Frescino (2002) evaluated predictions of two discrete and four continuous forest attributes obtained from linear models, general additive models, classification and regression trees, multivariate adaptive regression splines, and artificial neural networks. They concluded that the multivariate adaptive regression splines and the artificial neural networks approaches were marginally superior.

Regression modeling has been a popular international approach for use with satellite imagery to map a variety of forest attributes: large area volume and above ground biomass in Finland (Tomppo et al., 2002); hardwood and conifer cover in Oregon, USA (Maiersberger et al., 2001); height and basal area in Scotland (Puhr & Donoghue, 2000); biomass in Brazil (Steininger, 2000); volume in British Columbia, Canada (Gemmell, 1995); age and structure in the Pacific Northwest of the USA (Cohen et al., 1995); age in Estonia (Nilson & Peterson, 1994); age in Colorado, USA (Nel et al., 1994); biomass in England (Danson & Curran, 1993); volume in New Brunswick, Canada (Ahern et al., 1991); and suites of forest inventory variables in Finland (Tomppo, 1987, 1988).

Numerous investigators have used regression techniques to estimate or enhance estimates of forest area. For example, Deppe (1998) used a regression approach to enhance estimation of forest area in Brazil and Bolivia. The combined use of regression techniques, double sampling, and fine and coarse resolution imagery has received considerable attention. Iverson et al. (1989) classified Landsat TM imagery with respect to forest and non-forest using forest inventory data and aerial photographs and then used linear regression to estimate the relationship between the amount of forest landscape in the classified TM imagery and AVHRR spectral values. Nelson (1989) used linear regression, ratio of means, and mean of ratios estimators with AVHRR-GAC 4-km data and Landsat MSS 80-m data. Zhu and Evans (1994) used ground sample data to classify finer resolution Landsat TM imagery with respect to a forest attribute