Remote Sensing of Environment 151 (2014) 149–156

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Post-classification approaches to estimating change in forest area using remotely sensed auxiliary data

Ronald E. McRoberts

Northern Research Station, U.S. Forest Service, Saint Paul, MN, USA

ARTICLE INFO

Article history:
Received 4 September 2012
Received in revised form 22 March 2013
Accepted 26 March 2013
Available online 16 October 2013

Keywords:
Model-assisted regression estimator
Model-based estimator
Bootstrap

ABSTRACT

Multiple remote sensing-based approaches to estimating gross afforestation, gross deforestation, and net deforestation are possible. However, many of these approaches have severe data requirements in the form of long time series of remotely sensed data and/or large numbers of observations of land cover change to train classifiers and assess the accuracy of classifications. In particular, when rates of change are small and equal probability sampling is used, observations of change may be scarce. For these situations, post-classification approaches may be the only viable alternative. The study focused on model-assisted and model-based approaches to inference for post-classification estimation of gross afforestation, gross deforestation, and net deforestation using Landsat imagery as auxiliary data. Emphasis was placed on estimation of variances to support construction of statistical confidence intervals for estimates. Both analytical and bootstrap approaches to variance estimation were used. For a study area in Minnesota, USA, estimates of net deforestation were not statistically significantly different from zero.

1. Introduction

The Land Use, Land Use Change Forestry (LULUCF) sector plays a vital role in the global greenhouse gas (GHG) balance. Although the approximately 13 million hectares (ha) of forest that are converted to other land uses annually worldwide account for as much as 25% of anthropogenic GHG emissions (Achard et al., 2002; FAO, 2005, p. 13; Gullison et al., 2007), the LULUCF sector also has the greatest potential to remove GHGs from the atmosphere.

Carbon accounting includes assessment of the scale of GHG emissions from the forestry sector relative to other sectors. The gain–loss approach to carbon accounting is the most commonly used approach for estimating GHG emissions for national measurement, reporting, and verification (MRV) systems under the auspices of the Intergovernmental Panel on Climate Change (IPCC) (Giardin, 2010). With this approach, the net balance of additions to and removals from a carbon pool is estimated as the product of the rates of land use area changes and the responses of carbon stocks for those land use changes. Remote sensing-based approaches to estimating rates of forest area change have been emphasized as an important tool for monitoring changes in forest area (GOFC-GOLD, 2010, chap. 2). Further, good practice requires that the uncertainty in estimates of forest area change should be reported, regardless of the method used to obtain the estimates (Köhl, Baldauf, Plugge, & Krug, 2009; Watson, 2009).

Remote sensing-based change detection methods include two primary categories, trajectory analyses and bi-temporal methods. Trajectory analyses use time series of three or more images to assess not only the type and extent of change but also the trends and temporal patterns of change over time. Bi-temporal methods entail the analyses of images for two different dates and can be further separated into two subcategories. With post-classification, two forest/non-forest classifications constructed separately using two sets of forest/non-forest training data and two images are compared to estimate change, whereas with direct classification, a single classification of change is constructed using a single set of forest change training data with data for two images. Although trajectory analyses produce more detailed information such as type and timing of change and direct classification focuses explicitly on the change categories of interest, both methods have rather severe data requirements. With trajectory analyses, an extensive time series of imagery is typically required (Kennedy, Cohen, & Schroeder, 2007; Zhu, Woodcock, & Olofsson, 2012). With direct classification, large numbers of change observations may be necessary for training the classifier and/or assessing accuracy, a difficult task when rates of change are small and change observations are acquired using equal probability sampling designs. The advantage of post-classification is that the data requirements are much less severe. The disadvantage is that two sets of classification errors must be accommodated, although forest/non-forest classification errors are often less frequent than forest change classification errors.

The overall objective was to estimate parameters related to forest area change using multiple approaches to inference. Response variables of interest included gross deforestation, defined as loss of forest area; gross afforestation, defined as gain in forest area including reforestation; and net deforestation defined as the net result of gross deforestation and gross afforestation. For a study area in northeastern Minnesota in the United States of America (USA), two datasets were used, observations of forest/non-forest for national forest inventory (NFI) plots and corresponding summer Landsat imagery for the years 2002 and 2007. Because

0034-4257/$ – see front matter. Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.rse.2013.03.036

the combined dataset included few observations of forest area change, only post-classification approaches were used. An intermediate technical objective was to estimate areal population means, $\hat{\mu}$, and variances, $V\hat{a}r(\hat{\mu})$, for proportion forest for each year. The final technical objective was to use the two sets of estimated means and variances, one set for each of 2002 and 2007, to construct approximate 95% confidence intervals for estimates of parameters related to forest area change between the two years.

A nonlinear logistic regression model was used to estimate the relationship between forest/non-forest observations and Landsat spectral information, and the analyses included investigations of the effects on estimates of means and variances using different combinations of spectral variables in the model. Both a probability-based, model-assisted regression estimator and a model-based estimator were used.

2. Data

The study area was defined by the portion of the row 27, path 27, Landsat scene in northeastern Minnesota, USA, which was cloud-free for the two image dates, 16 July 2002 and 30 July 2007 (Fig. 1). The Landsat Thematic Mapper (TM) spectral data were transformed using the normalized difference vegetation index (NDVI) transformation (Rouse, Haas, Schell, & Deering, 1973) and the three tasseled cap transformations (TC_{green}, TC_{bright}, TC_{wet}) (Crist & Cicone, 1984; Kauth & Thomas, 1976) for each image. These four transformations were used as independent variables when constructing models of the relationship between ground and remotely sensed data.

Ground data were obtained for plots established by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service which conducts the NFI of the USA. The program has established field plot centers in permanent locations using a sampling design that is regarded as producing an equal probability sample (McRoberts, Bechtold, Patterson, Scott, & Reams, 2005). Each FIA plot consists of four 7.32-m (24-ft) radius circular subplots that are configured as a central subplot and three peripheral subplots with centers located at distances of 36.58 m (120 ft) and azimuths of 0°, 120°, and 240° from the center of the central subplot. Centers of forested, partially forested, or previously forested plots are estimated using global positioning system (GPS) receivers, whereas centers of non-forested plots are verified using aerial imagery and digitization methods.

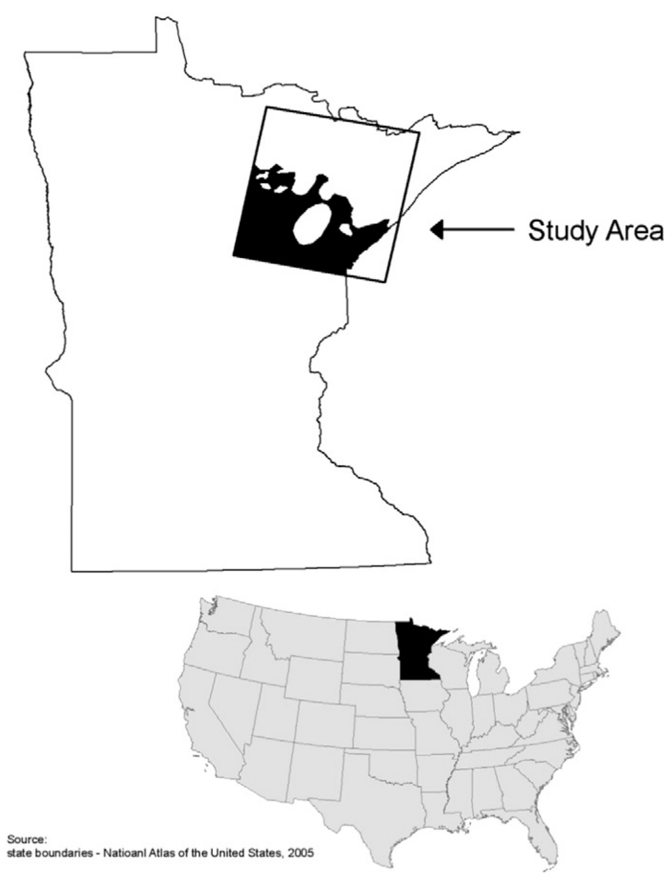

Fig. 1. Study area in northeastern Minnesota, USA.

Data were available for 249 FIA plots measured in both 2002 and 2007. Field crews visually estimate the proportion of each subplot that satisfies the FIA definition of forest land: minimum area of 0.4 ha (1.0 ac), minimum crown cover of 10%, minimum crown cover width of 36.6 m (120 ft), and forest land use. Field crews also observe species and measure diameter at-breast-height (dbh) (1.37 m, 4.5 ft) and height for all trees with dbh of at least 12.7 cm (5 in.). Growing stock volumes are estimated for individual measured trees using statistical models, aggregated at subplot-level, expressed as volume per unit area, and considered to be observations without error. For this study, data for only the central subplot of each plot were used to avoid dealing with spatial correlation among observations for subplots of the same plot. Doing so resulted in little loss of information, because the correlation among observations for subplots of the same plot was greater than 0.85. Subplot-level proportion forest and volume data were combined with the values of the spectral transformations for pixels containing subplot centers. For future reference, the term *plot* refers to the central subplot of each FIA plot cluster.

Two concerns must be addressed when constructing datasets using the FIA plot data and Landsat imagery. First, because the smaller 168.3-m^2 plots may not adequately characterize the larger 900-m^2 TM pixels, observations for the four plots that were not completely forested or completely non-forested were deleted from the analyses. Second, because FIA field crews classify plots with respect to land use, not land cover, plots whose tree cover has been removed are still classified as forest if trees are expected to regenerate and forest land use is expected to continue. Thus, observations of land cover for plots with forest land use but no measurable volume were considered to be missing at random and were also deleted from the analyses. These two data issues are discussed in detail in Section 4.1. Following deletions, land cover observations for 199 plots remained.

3. Methods

3.1. Inferential assumptions

All analyses were based on three underlying assumptions: (1) a finite population consisting of N units in the form of square, 900-m^2 Landsat pixels, (2) a sample of n population units in the form of pixels that contain FIA plot centers, and (3) availability of auxiliary data in the form of the Landsat spectral transformations for all pixels. In the following sections, the terms *population unit* and *pixel* are used interchangeably.

For areal assessments, the objective is typically to estimate the area for a class of the response variable. Because the estimate of class area is simply the product of total area which is usually known and the estimate of the class area proportion, the focus of this study was estimation of the proportion, in this case proportion forest which was denoted μ. Thus, the analytical objective was construction of an approximate 95% confidence interval for $\hat{\mu}$ expressed as,

$$\hat{\mu} \pm 2 \cdot \sqrt{V\hat{a}r(\hat{\mu})}, \quad (1)$$

where $V\hat{a}r(\hat{\mu})$ is the estimate of the variance of $\hat{\mu}$.