ISPRS Journal of Photogrammetry and Remote Sensing 142 (2018) 292–300

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions

Ronald E. McRoberts[a,*], Stephen V. Stehman[b], Greg C. Liknes[a], Erik Næsset[c], Christophe Sannier[d], Brian F. Walters[a]

[a] Northern Research Station, U.S. Forest Service, Saint Paul, MN, USA
[b] Forest and Natural Resources Management, State University of New York, Syracuse, NY, USA
[c] Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway
[d] Systèmes d'Information à Référence Spatiale (SIRS), Villeneuve d'Ascq, France

ARTICLE INFO

Keywords:
Intepreter error
Bias
Precision
Greenhouse gas inventory
Gain-loss method

ABSTRACT

The gain-loss approach for greenhouse gas inventories requires estimates of areas of human activity and estimates of emissions per unit area for each activity. Stratified sampling and estimation have emerged as a popular and useful statistical approach for estimation of activity areas. With this approach, a map depicting classes of activity is used to stratify the area of interest. For each map class used as a stratum, map units are randomly selected and assessed with respect to an attribute such as forest/non-forest or forest land cover change. Ground observations are generally accepted as the most accurate source of information for these assessments but may be cost-prohibitive to acquire for remote and inaccessible forest regions. In lieu of ground observations, visual interpretations of remotely sensed data such as aerial imagery or satellite imagery are often used with the caveat that the interpretations must be of greater quality than the map data. An unresolved issue pertains to the effects of interpreter error on the bias and precision of the stratified estimators of activity areas.

For a 7500-km^2 study area in north central Minnesota in the United States of America, combinations of forest inventory plot observations, visual interpretations of aerial imagery, and two forest/non-forest maps were used to assess the effects of interpreter error on stratified estimators of proportion forest and corresponding standard errors. The primary objectives related to estimating the bias and precision of the stratified estimators in the presence of interpreter errors, identifying factors and the levels of those factors that affect bias and precision, and facilitating planning to circumvent and/or mitigate the effects of bias. The primary results were that interpreter error induces bias into the stratified estimators of both land cover class proportion and its standard error. Bias increased with greater inequality in stratum weights, smaller map and interpreter accuracies, fewer interpreters and greater correlations among interpreters. Failure to account for interpreter error produced stratified standard errors that under-estimated actual standard errors by factors as great as 2.3. Greater number of interpreters mitigated the effects of interpreter error on proportion forest estimates, and a hybrid variance estimator accounted for the effects on standard errors.

1. Introduction

Two approaches to greenhouse gas emissions accounting are common, the *stock-change* approach and the *gain-loss* approach (IPCC, 2006, Volume 4, Chapter 2, p. 2.10; GFOI, 2016, p. 22). With the stock-change approach, mean annual emissions are estimated as the mean annual difference in carbon stocks between two points in time (IPCC, 2006, Volume 4, Chapter 4, Section 4.2.1.1; GFOI, 2016, Chapter 3). For countries with established forest ground sampling programs such as national forest inventories, the stock-change approach is fairly easy to implement. However, for countries with remote and inaccessible forests, the stock-change approach may be prohibitively expensive. For these countries, the gain-loss approach may be a more feasible alternative. With this approach, emissions are defined to be the net balance of additions to and removals from a carbon pool and are estimated as the product of the areas of "human activity causing emissions", characterized as *activity data*, and the responses of carbon stocks for those activities, characterized as *emission factors* (IPCC, 2006, Volume 1, Chapter 1, Section 1.2; GFOI, 2016, pp. xvii, 22)

Estimation of areas of activities often relies on remote sensing-based

* Corresponding author.
E-mail address: rmcroberts@fs.fed.us (R.E. McRoberts).

https://doi.org/10.1016/j.isprsjprs.2018.06.002
Received 23 February 2018; Received in revised form 27 May 2018; Accepted 2 June 2018
0924-2716/ Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).

land cover or land cover change maps (Olofsson et al., 2013, 2014; Ban et al., 2015). Of importance, however, estimating areas of these activities by simply adding the areas of map units assigned to activity classes, a practice characterized as *pixel counting*, is a biased procedure because it does not account for map classification errors. Stratified sampling and estimation is a statistically rigorous alternative. With stratified sampling, map classes are used as strata, and within-stratum samples are selected using simple random or systematic sampling designs. Sample unit observations of land cover or land cover change are then used as reference data, and the stratified estimators are used to estimate the areas of activity classes of interest (Olofsson et al., 2013, 2014). In the absence of reference data error, the stratified estimators are unbiased and more precise than simple random sampling estimators (Chen and Wei, 2009).

Reference data in the form of ground observations are often considered optimal, although Foody (2009, 2010) notes that even ground reference data are subject to error. However, regardless of error, acquisition of ground reference data for remote and inaccessible regions may be prohibitively expensive, if not logistically infeasible. For these situations, reference data in the form of visual interpretations of remotely sensed data are often used, albeit with the stipulation that such reference data are of greater quality than the map data with respect to factors such as resolution and accuracy (Mannel et al., 2006; Stehman, 2009; Olofsson et al., 2013; Pengra et al., 2015; Tsendbazar et al. 2015; Boschetti et al., 2016; GFOI, 2016, pp. 125, 139). However, even if visual image interpretations are of greater quality than the map data, they cannot be assumed to be without error. For five trained interpreters of stereo aerial photography, Næsset (1991) reported that interpretations of crown coverage for structurally homogenous Norwegian boreal forests differed substantially among interpreters and among different times of year for the same interpreter. For the same forest conditions, Næsset (1992) reported that interpretations of broad tree species groups by 12 professional, trained interpreters using stereo aerial photography produced only 31–79% agreement with field reference data. For five trained interpreters of videography, Powell et al. (2004) reported interpreter disagreement of almost 30% for five land cover classes in the Brazilian Amazon, two of which were forest-related classes. Thompson et al. (2007) reported errors of 30% when aerial imagery was used to classify boreal forest stands into coniferous, deciduous, and mixed classes in Ontario, Canada. For three trained interpreters, Sun et al. (2017) reported that despite among-interpreter consistency, manual interpretations of Google Earth and other fine resolution imagery were not as reliable as ground measurements for seven land cover classes in Central Asia. In summary, reference data in the form of visual interpretations of remotely sensed data, even by well-trained professional interpreters, are subject to substantial interpreter disagreement and error.

If the reference data are *imperfect* in the sense of being subject to error, then the stratified estimators may be biased, sometimes substantially biased despite only small errors (Foody, 2009, 2010, 2013). Although the effects of imperfect reference data on estimators of class proportions and areas have been at least partially addressed, little has been reported on the effects of imperfect reference data on variance estimators. Compliance with the IPCC good practice guidance for greenhouse gas inventories requires not only avoiding over- and/or under-estimates but also reduction of uncertainties (IPCC, 2006, Volume 1, Chapter 1, Section 1.2; GFOI 2016, p. 15) with the obvious caveat that uncertainties cannot be reduced unless they are first correctly estimated. In particular, correct estimation of uncertainty requires incorporation of the effects of imperfect reference data into variance estimators (Olofsson et al., 2014).

The objectives of the study were fivefold: (1) to assess the effects of imperfect reference data on the bias and precision of stratified estimators of land cover class proportions; (2) to characterize conditions that affect the magnitudes of bias and precision; (3) to develop a variance estimator that incorporates the effects of interpreter error; (4) to illustrate the effects of interpreter error on bias and precision with inventory ground data and visual interpretations of aerial imagery using two forest/non-forest maps; and (5) to facilitate planning for estimation of activity data. Because the ultimate objective is an estimate of the area of a land cover class, and area can be expressed as the product of the class proportion and the total population area which is usually known, the focus of the study was estimation of the class proportion.

2. Data

2.1. Study area

The study area was the 7583 km^2 of Itasca County in north central Minnesota in the United States of America (USA) (Fig. 1). Land cover includes water, wetlands and approximately 80% forest consisting of mixtures of pines (*Pinus* spp.), spruce (*Picea* spp.), and balsam fir (*Abies balsamea* (L.) Mill.) on upland sites and spruce (*Picea* spp.), tamarack (*Larix laricina* (Du Roi) K. Koch), white cedar (*Thuja occidentalis* (L.)), and black ash (*Fraxinus nigra* Marsh.) on lowland sites. Forest stands in the study area are typically naturally regenerated, uneven-aged, and mixed species.

2.2. Forest inventory data

Data were obtained for 310 ground plots established by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service which conducts the NFI of the USA. The plots were established in permanent field locations using a quasi-systematic sampling design that is regarded as producing an equal probability sample (McRoberts et al., 2010; Mountrakis and Xi, 2013) and were measured between 2014 and 2016. Field crews visually estimate the proportion of each plot that satisfies the FIA definition of forest land: (i) minimum area 0.4 ha (1.0 ac), (ii) minimum tree cover of 10%, (iii) minimum width of 36.58 m (120 ft), and (iv) forest land use. All field crews are well-trained, tested on their ability to assess plot variables, and hence well-qualified to distinguish forest from non-forest based on the FIA definition. A small number of plots in three categories were deleted and considered to be missing at random (Rubin, 1987): (i) plots with mixtures of forest and non-forest cover, (ii) plots with forest use but with no tree cover due to conditions such as recent harvest, and (iii) to the degree possible, plots with non-forest use but with tree cover of which parks and rural residential areas are examples. Plot centers were estimated using global positioning system receivers with sub-meter accuracy. The field crew, plot-level, forest/non-forest observations were used as reference data to produce estimates of proportion forest that served as the standard for comparison for estimates based on visual interpretations of aerial imagery. They also served as the basis for assessing map and interpreter accuracies.

2.3. Percent tree canopy cover datasets

The *Global Forest Change* (GFC) dataset is based on cloud-free, composite, annual growing season Landsat 7 Enhanced Thematic Mapper Plus data (Hansen et al., 2013). For 30-m × 30-m pixels, the dataset includes predictions of maximum percent tree canopy cover in the range of 0–100% for vegetation taller than 5 m for the year 2010. The 2011 *National Land Cover Database* (NLCD) includes percent tree canopy cover values in the range of 0–100% for 30-m × 30-m pixels (Homer et al, 2015). Each of the two datasets was used to construct a forest/non-forest map (Section 3.2) which then facilitated stratified sampling and estimation (Section 3.3).

2.4. Aerial imagery

Aerial imagery was obtained from the Farm Service Agency of the U.S. Department of Agriculture through the National Agriculture