Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation

Pontus Olofsson a,⁎, Giles M. Foody b, Stephen V. Stehman c, Curtis E. Woodcock a

a Department of Earth and Environment, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA
b School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK
c Department of Forest and Natural Resources Management, State University of New York, 1 Forestry Drive, Syracuse, NY 13210, USA

Abstract

The area of land use or land cover change obtained directly from a map may differ greatly from the true area of change because of map classification error. An error-adjusted estimator of area can be easily produced once an accuracy assessment has been performed and an error matrix constructed. The estimator presented is a stratified estimator which is applicable to data acquired using popular sampling designs such as stratified random, simple random and systematic (the stratified estimator is often labeled a poststratified estimator for the latter two designs). A confidence interval for the area of land change should also be provided to quantify the uncertainty of the change area estimate. The uncertainty of the change area estimate, as expressed via the confidence interval, can then subsequently be incorporated into an uncertainty analysis for applications using land change area as an input (e.g., a carbon flux model). Accuracy assessments published for land change studies should report the information required to produce the stratified estimator of change area and to construct confidence intervals. However, an evaluation of land change articles published between 2005 and 2010 in two remote sensing journals revealed that accuracy assessments often fail to include this key information. We recommend that land change maps should be accompanied by an accuracy assessment that includes a clear description of the sampling design (including sample size and, if relevant, details of stratification), an error matrix, the area or proportion of area of each category according to the map, and descriptive accuracy measures such as user’s, producer’s and overall accuracy. Furthermore, mapped areas should be adjusted to eliminate bias attributable to map classification error and these error-adjusted area estimates should be accompanied by confidence intervals to quantify the sampling variability of the estimated area. Using data from the published literature, we illustrate how to produce error-adjusted point estimates and confidence intervals of land change areas. A simple analysis of uncertainty based on the confidence bounds for land change area is applied to a carbon flux model to illustrate numerically that variability in the land change area estimate can have a dramatic effect on model outputs.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Land use or land cover change (referred to as “land change” for the reminder of the article) impacts on a very diverse array of environmental properties and processes. The effects of a land change may be felt across a broad spectrum of environmental systems including the atmospheric, hydrologic, geomorphologic and ecologic. Deforestation may, for example, act as a source of carbon to the atmosphere, lead to enhanced soil erosion, reduce the extent of habitat and so to species declines and contribute to displacement of human populations. Land change is, therefore a critical variable in relation to two environmental issues of great societal concern: climate change and biodiversity loss. Land change can be a cause and a consequence of climate change and is a variable of greater impact than climate change (Skole, 1994). Land change is, for example, the single most important variable affecting ecological systems (Chapin et al., 2000; Vitousek, 1994) and the greatest threat to biodiversity (Sala et al., 2000). The importance of land change is evident in the growth of interest in land change science (Turner et al., 2007) and so there is consequently considerable interest in land cover and a need for accurate information on land cover and its dynamics. Indeed the central role of land surface change to a vast array of contemporary concerns is reflected in its role as an underpinning feature of the current grand challenges for the geographical sciences articulated recently by the US National Academy of Sciences (CSDCSND, 2010). Remote sensing has the potential to provide accurate information on land cover but numerous problems may be encountered and the adequacy of this information has been questioned (Townshend et al., 1992; Wilkinson, 1996, 2005).

In many applications the main focus is the area of a land cover class or its gross change over time (gross change refers to the total area of
gain or total area of loss of a land cover or land use class, for example, total area of forest cover gain, whereas net change refers to the difference between the gains and losses of a class, for example forest gain subtracted from forest loss). The importance of area has been evident throughout the entire history of satellite remote sensing. For example, the area of land planted to wheat was a central component of the Large Area Crop Inventory Experiment (LACIE) in the 1970s (Erickson, 1984). Numerous studies have investigated changes in the extent of land cover classes such as forests (DeFries et al., 2002; Olofsson et al., 2011), deserts (Eklundh & Olsson, 2003), lakes (Smith et al., 2005), savannas (Branstrom et al., 2008), crop lands (Fuller et al., 2012; Liu et al., 2005), shrublands (McManus et al., 2012) and urban areas (He et al., 2010; Lizarazo, 2010; Schneider & Woodcock, 2008).

In some instances there is also considerable interest associated with the agencies and consequences of changes, such as associated in studies of burned areas (e.g. Vivchar, 2011) or ice sheets (McCabe et al., 2011). Interest in land cover and land changes remains and may be expanded further as a result of major policy related issues. This activity may be linked to a variety of areas of policy, from that seeking to counter-narcotics (e.g. Taylor et al., 2010) to those underpinning major issues in contemporary environmental science. For example, member states of the European Union are obligated to maintain the extent of key habitats under the Habitats Directive (EC, 2011). The goal of the United Nations Collaborative Programme on Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (UN-REDD), established in 2005, is to reduce greenhouse gas emissions from deforestation while maintaining sustainability (UN-REDD, 2008). This calls for area estimates of deforestation, accompanied by a statement of the uncertainty of the estimates (e.g. confidence intervals) – as clearly stated in the UN-IPCC Good Practice Guidance for Land Use, Land Use Change and Forestry (IPCC, 2003, p. G-2):

“Estimates should be accurate in the sense that they are systematically neither over nor under true emissions or removals, so far as can be judged, and that uncertainties are reduced so far as is practicable”

Accurately quantifying the extent of a land cover class or its amount of change over time is, however, a non-trivial task. Recent estimates of the extent of global urban land cover, for example, differ by an order of magnitude (Potere & Schneider, 2007). Similarly estimates of land changes, such as forest change which is central to carbon accounting (Kuemmerle et al., 2011; Olofsson et al., 2011) are often subject to substantial error (Skole & Tucker, 1993). Addressing key environmental science concerns at regional to global scales requires accurate information on the extent of land cover classes and their change over time, information that is often based on remote sensing.

The area of land change may be obtained directly from change maps determined from remote sensing data. Such maps are typically produced by an image classification analysis. Singh (1989) reviews many of the methods for mapping land change from remote sensing. Popular approaches are typically based either on the comparison of the radiometric properties of images acquired at different times such as change vector analysis (Lambin & Strahler, 1994) or on the differences in land cover maps obtained from the images such as from a post-classification comparison analysis. The latter methods are particularly popular but in many studies of change via a post-classification comparison analysis, the accuracy of change is not evaluated (e.g. Kuemmerle et al., 2009; Xian & Crane, 2005) as researchers instead assess the accuracy of the individual classifications (i.e., the classification for each date). As will be illustrated below, even when the two classifications are highly accurate it is possible that the accuracy of the change map will be low and that the area computed from a change map could be badly biased.

Pixel counting involves determining the number of pixels allocated to a map class and multiplying this number by the area of a pixel to obtain the mapped area of the class. This approach is a simple way to compute area but in the presence of asymmetric classification errors pixel counting is biased for the true proportion of area (Czaplewski, 1992; Gallego, 2004; Stehman, 2005). The bias resulting from applying pixel counting to obtain the area of a class is labeled as “measurement bias” rather than as “estimator bias” because a pixel count represents a complete census of the region and therefore is not a sample-based estimator (Särndal et al., 1992; Stehman, 2005). A variety of approaches exist to estimate area using information obtained from an accuracy assessment sample (Card, 1982; Czaplewski, 1992; McRoberts, 2011; Stehman, 2009). Card (1982) introduced the stratified estimator for area estimation that is the primary focus of this article. Gallego’s (2004) review of area estimation provides an excellent summary of many of the area estimation options, including a critique of pixel counting and an overview of estimators combining ground and remote sensing information, as well as a review of methods for small area estimation applicable when interest lies in small geographic regions that receive few sample units. Stehman (2009) noted that many of the area estimators previously proposed in the literature could be unified under the framework of “model-assisted” estimation in which “a model motivates the form of the estimator, but inference depends on the sampling design” (Lohr, 1999, p. 147) and so design-based inference (Särndal et al., 1992; Stehman, 2000) is still invoked. McRoberts (2011) re-invigorated recognition of the importance of estimating standard errors and confidence intervals to provide a complete inference when estimating area using information from remote sensing (see also McRoberts, 2010). Although in this article we limit attention to design-based inference, the existence of model-based inference should be acknowledged.

Despite the availability of methods to refine class extent estimates for classification errors, the pixel counting approach continues to be widely used. Consequently, the full potential of the accuracy assessment data is not being utilized. In fact, it is not being used at all for estimating area in a pixel counting approach. An accuracy assessment does more than indicate the accuracy of the map – it provides sample data that can be used to avoid the measurement bias of pixel counting and to decrease the standard error of the estimated area. As such, the accuracy assessment results should not be the final step of the quality evaluation but an integral part of the overall analysis of accuracy and area.

The problems encountered in estimating the area and accuracy of change are well known in remote sensing. Similarly, methods to explicitly address the concerns have been discussed openly in the literature for decades, but the community has yet to adopt these methods as part of routine practice. Here, we re-state some basic principles to help ensure that the vast potential of remote sensing as a source of information on change may be realized more fully and thus contribute constructively to major international research and policy priorities (e.g. deforestation and UN-REDD).

1.1. Objectives

The objective of this article is to present a simple strategy for using the information obtained for map accuracy assessment to estimate area of a land cover class or of land change, and to construct confidence intervals that reflect the uncertainty of the area estimates obtained. The rationale for presenting this strategy is that even though the methods we advocate date back to at least (Card, 1982), the remote sensing community has yet to consistently adopt these good practices. Our evaluation of common practice of land change accuracy assessment is based on peer-reviewed land change papers published during 2005–2010 in Remote Sensing of Environment and the International Journal of Remote Sensing. Our review of these articles revealed that many published assessments did not provide the full information to address land change accuracy and area estimation objectives (Section 2). To remedy the common shortcomings of land
change assessments, we present an analysis that makes full use of the map and accuracy assessment data by: 1) estimating accuracy (e.g., user’s, producer’s, and overall accuracies), 2) estimating area of land change using the accuracy assessment sample data to adjust area for map classification error, and 3) estimating standard errors or confidence intervals for the error-adjusted area estimates. Several numerical examples are provided to show how to produce these estimates using simple estimators that can be applied to popular simple random, systematic, and stratified random sampling designs (Section 3). Lastly, we illustrate the importance of accounting for uncertainty of land change area estimates in applications that use these area estimates as inputs. Specifically, we show how the outputs of a carbon flux model change with respect to the variability of the land change area estimates input to the model (Section 4).

We focus on applications in which a map of land change is produced. This map is subjected to an accuracy assessment based on a spatially explicit (e.g., per pixel) comparison between the map class and the reference class for a sample of spatial units (cf. Stehman & Czaplewski, 1998). The reference class is defined as the best available determination of the ground condition at a specified location. The reference class label is assumed to be correct, but it is well-known that reference classification error often occurs and can impact greatly on evaluations of land cover and land cover change by remote sensing (Foody, 2010). For simplicity, we limit the attention to an accuracy assessment unit that is a pixel or other equal area spatial unit. We assume a hard classification scheme for both the map and reference classification, where a hard classification is defined as one in which each pixel is assigned fully to only one class. The literature also provides extensive guidance on other important issues connected to the data such as the sample size and class definitions that are beyond the scope of this current article (Stehman, 2012; Strahler et al., 2006).

1.2. Accuracy and uncertainty

Accuracy is defined as the degree to which the map produced agrees with the reference classification. The most commonly used measures of accuracy are the following (see Liu et al., 2007 for a comprehensive review of other accuracy measures):

(i) **Overall accuracy** is simply the proportion of the area mapped correctly. It provides the user of the map with the probability that a randomly selected location on the map is correctly classified.

(ii) **User’s accuracy** is the proportion of the area mapped as a particular category that is actually that category “on the ground” where the reference classification is the best assessment of ground condition. If a “user” employs the final change map for locating a particular area of land change, the user’s accuracy gives the conditional probability of that map location actually having changed. User’s accuracy is the complement of the probability of commission error.

(iii) **Producer’s accuracy** is the proportion of the area that is a particular category on the ground that is also mapped as that category. The producer’s accuracy provides the “producer” of the final land change map with the conditional probability of a particular location of actual land change appearing as land change on the map. Producer’s accuracy is the complement of the probability of omission error.

(iv) The kappa coefficient of agreement is often used as an overall measure of accuracy. Kappa purportedly incorporates an adjustment for “random allocation agreement”, but the validity of such an adjustment is arguable and numerous articles have questioned the use of kappa (Foody, 2002; Pontius & Miliones, 2011; Stehman, 1997). Adjusting for random allocation agreement has no relevance to area estimation so the following analysis and discussion will not include kappa. We recommend that kappa should not be used in the assessment of the accuracy of land change maps.

Because accuracy measures are typically estimated from a sample, these estimates are subject to uncertainty. The uncertainty of an estimate can be represented by computing its standard error or by reporting a confidence interval. A confidence interval provides a range of values for a parameter taking into account the uncertainty of the sample-based estimate.

The manner in which uncertainty is addressed depends on the inference framework employed. In this article, we will use design-based inference (Särndal et al., 1992) in which the uncertainty associated with the estimator is defined as the variability of the estimates over the set of all possible samples that could have been obtained for the chosen sampling design and population sampled. The standard error computed for the particular sample selected is an estimate of the variability over the set of all possible samples. Other sources of uncertainty will often be present. For example, error in the reference class label, geo-location error, and mis-matched classification legends may generate additional uncertainty. But we address only the variability attributable to sampling in our uncertainty analysis.

2. Literature review

As stated in the objectives (Section 1.1), an accuracy assessment of a land change map should include the following information: 1) estimates of accuracy of change; 2) estimates of land change area that adjust the map “pixel count” area for classification error; and 3) confidence intervals associated with the accuracy and land change area parameter estimates. Methods for achieving these three outcomes have been extensively published. For example, formulas for estimating accuracy and the standard errors of these accuracy estimators may be found in Card (1982) and Stehman and Foody (2008). Area estimation has recently received considerable attention, with various approaches reviewed by Gallego (2004), McRoberts (2010), McRoberts (2011), and Stehman (2009). The stratified estimator of area we highlight (Section 3) has been in use at least since Card (1982). The stratified estimator is applicable to sampling designs commonly used in accuracy assessment, namely simple random, systematic, and stratified random sampling. When applied to simple random or systematic sampling (i.e., designs without stratification), the stratified estimator has been historically referred to as a “poststratified” estimator to distinguish between using strata in the sampling design (i.e., stratified sampling) in contrast to using strata in the estimator (i.e., poststratified estimation).

As a guide to popular practice, we surveyed articles published between 2005 and 2010 in two major journals, Remote Sensing of Environment and the International Journal of Remote Sensing, that have a track record of publishing land change articles (see Table S1, Supplemental material). The articles were categorized into four different classes in relation to the way accuracy and area estimation of land change were reported: 1) no accuracy measures presented, 2) accuracy measures presented but not for the accuracy of change (i.e., the accuracies of the maps produced at each date are provided but a direct assessment of change accuracy is not conducted), 3) accuracy measures for change presented, and 4) accuracy measures presented and accuracy information used for calculation of adjusted change area and/or confidence intervals. It should be evident that the four classes are ordered in terms of the rigor of the accuracy assessment with the latter class comprised of articles that adhere most closely to good practice.

Our evaluation of the articles published in the two journals revealed that most land change studies do not take full advantage of the information available from accuracy assessment. Of the 57 publications identified as conducting a land change study (mainly deforestation), 8 did not publish any accuracy measures at all, 24 did not report an accuracy assessment of change (although accuracy of single date maps may have been provided), 16 reported an accuracy assessment of change, and 9 studies provided the accuracy measures and the information needed to compute the error-adjusted area estimate of change. Only three articles included this last step of presenting an
estimate of land change area that was adjusted according to the accuracy assessment data. So while the problems of estimating area from imperfect classifications and methods to address them have been discussed in the remote sensing literature for at least 30 years (e.g. Card, 1982) it appears that methods used and conclusions drawn in many studies fail to make full use of the available data.

A post-classification analysis to obtain change (described in Section 3) was used in 31 of the reviewed papers. Of these 31 articles, 5 did not report any accuracy assessment, 23 had accuracy assessments performed on the individual land cover classifications used for the overlay (i.e. single-date accuracy) but not for the final change map, and only 2 assessed the accuracy of the final change map. As demonstrated in Section 3, accuracy measures produced for the individual single date land cover maps may not be indicative of the accuracy of the change map. The overall accuracy of a change map constructed by overlaying two land cover classifications is the product of the overall accuracies of the pre-comparison maps if the classification errors are independent (Fuller, 2003; van Oort, 2007). However, the assumption of independent errors would generally be untenable because locations that were difficult to classify correctly at one date would likely be difficult to classify correctly at another date (Congalton, 1988). Thus reporting overall accuracy for both dates is unlikely to be informative of the overall accuracy of the post-classified change map.

The common practice of reporting error matrices demonstrates how well established the notion of accuracy assessment is in the remote sensing community. Performing an accuracy assessment is not a trivial task and requires time and resources. That researchers invest the effort to assess the accuracy of maps and construct error matrices is encouraging, but these assessments often are confined to the single date land cover maps compared to predict change rather than to the gross change map itself. As shown in Sections 2 and 3, once an accuracy assessment for land change has been performed and an error matrix has been created, estimating the error-adjusted land change area and confidence interval is not difficult but yields valuable information on change that may be very different from the results obtained solely from the map (i.e., pixel counting).

3. Estimating land change area: methods and examples

In Section 1, the key information and computations needed to produce a complete and rigorous report of accuracy of a land change map and estimation of area are documented. Two examples are provided. The first example is presented to illustrate in detail the calculations needed to make use of the accuracy data for estimating area and associated confidence intervals (Section 2). The second example shows the analysis of accuracy and area when change is determined by overlaying land cover classifications for the two dates bracketing the change period, in this case classifications of forest and non-forest (Section 3), but the approach is applicable to any land change such as those highlighted in Section 1.

3.1. Estimating accuracy and area of change

Suppose the objective is to assess the accuracy of a map with \( q \) categories and to estimate the area of a particular map category. A sample of assessment units (e.g., pixels) is selected by simple random, stratified random (with the map classes as strata), or systematic sampling. A sample error matrix is constructed where the map categories \((i = 1, 2, ..., q)\) are represented by rows and the reference categories \((j = 1, 2, ..., q)\) by columns (Table 1). Note that in some presentations of an error matrix (e.g. Card, 1982), the rows and columns are reversed. The basic principles of the methods outlined below still apply to such situations but accommodation for the switch in row and column contents is required.

Table 1 illustrates the common practice of reporting the error matrix in terms of sample counts. A more informative presentation of the error matrix is in terms of the unbiased estimator of the proportion of area in cell \( i,j \) of the error matrix:

\[
\hat{p}_{ij} = W_i \frac{n_{ij}}{n_i}
\]

where the total area of the map is \( A_{tot} \), the mapped area of category \( i \) is \( A_{ni} \) (subscript \( m \) denotes "mapped"), and the proportion of the area mapped as category \( i \) is \( W_i = A_{ni} / A_{tot} \). The error matrix in terms of estimated area proportions is shown in Table 2. An advantage of the presentation given in Table 2 is that accuracy and area estimates can be computed directly from the error matrix.

Because of classification error, the mapped area proportions given by \( A_{ni} / A_{tot} \) are usually biased when the objective is to estimate the true proportion of area of category \( i \) as determined from the reference classification. Instead of obtaining the area directly from the map classification, an area estimator can be based on the reference classification of each sample unit. The area proportions for each reference-defined category \( j \) are estimated from the column totals (\( \hat{p}_{j} \)) in Table 2. An unbiased estimator of the total area (based on the reference classification) of category \( j \) is then:

\[
\hat{A}_j = A_{tot} \times \hat{p}_j
\]

(2)

Eq. (2) can be re-expressed in an expanded alternate form that more clearly reveals the estimator as a stratified estimator:

\[
\hat{A}_j = A_{tot} \sum_i W_i \frac{n_{ij}}{n_i}
\]

(2')

This stratified estimator can be viewed as an "error-adjusted" estimator of area because it includes the area of map omission error of category \( j \) and leaves out the area of map commission error. The estimated standard error of the estimated area proportion is (Cochran, 1977)

\[
S(\hat{p}_j) = \sqrt{\frac{\sum_i W_i^2 \frac{n_{ij} (1 - n_{ij})}{n_i - 1}}{A_{tot}}}
\]

(3)

The standard error of the error-adjusted estimated area is

\[
S(\hat{A}_j) = A_{tot} \times S(\hat{p}_j)
\]

(4)

An approximate 95% confidence interval for \( \hat{A}_j \) is,

\[
\hat{A}_j \pm 2 \times S(\hat{A}_j)
\]

(5)

where the margin of error is defined as the z-score (\( z \) is a percentile from the standard normal distribution) multiplied by the standard error (i.e., the \( \pm \) part of the confidence interval), and the value of the z-score depends on the confidence level (for 95% confidence, \( z = 1.96 \) which is approximated here to 2 for simplicity of presentation).

Table 1

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>q</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( n_{11} )</td>
<td>( n_{12} )</td>
<td>...</td>
<td>( n_{1q} )</td>
<td>( n_{1} )</td>
</tr>
<tr>
<td>2</td>
<td>( n_{21} )</td>
<td>( n_{22} )</td>
<td>...</td>
<td>( n_{2q} )</td>
<td>( n_{2} )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>q</td>
<td>( n_{q1} )</td>
<td>( n_{q2} )</td>
<td>...</td>
<td>( n_{qq} )</td>
<td>( n_{q} )</td>
</tr>
<tr>
<td>Total</td>
<td>( n_{1} )</td>
<td>( n_{2} )</td>
<td>...</td>
<td>( n_{q} )</td>
<td>( n )</td>
</tr>
</tbody>
</table>
Eqs. (1)–(5) are applicable to simple random, systematic, or stratified random sampling. These estimators are the usual stratified estimators if stratified random sampling is implemented. If applied to simple random or systematic sampling, the estimators are poststratified estimators in which the strata are incorporated via the estimator instead of via the sample selection as is the case for a stratified design (Card, 1982; Cochran, 1977; Särndal et al., 1992). Although the area of estimated change is based on the reference classification of the sampling units, the map is still an important component of the area estimation approach because of the role of \( W_i \) in the area estimator and the importance of the stratification defined by the map classification when the strata are used in the sampling design or used in a post-stratified area estimator. The standard error formula (Eq. 3) for the poststratified estimator is an approximation if the sampling design is systematic. This approximation is usually expected to overestimate the true standard error because, for each stratum, the formula applied is based on simple random sampling, and the simple random approximation typically overestimates the standard error if the sampling design is systematic (Wolter, 2007).

Although our primary focus is estimating the area of land change, we briefly review the estimation of accuracy for stratified random sampling. User’s accuracy requires data only from within a given stratum so it can be computed directly from the sample counts. But over-sampling User’s accuracy requires data only from within a given stratum.

### Table 2

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>( q )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \hat{p}_{11} )</td>
<td>( \hat{p}_{12} )</td>
<td>( \hat{p}_{1i} )</td>
<td>( \hat{p}_1 )</td>
</tr>
<tr>
<td>2</td>
<td>( \hat{p}_{21} )</td>
<td>( \hat{p}_{22} )</td>
<td>( \hat{p}_{2i} )</td>
<td>( \hat{p}_2 )</td>
</tr>
<tr>
<td>( q )</td>
<td>( \hat{p}_{qi} )</td>
<td>( \hat{p}_{qi} )</td>
<td>( \hat{p}_{qi} )</td>
<td>( \hat{p}_q )</td>
</tr>
<tr>
<td>Total</td>
<td>( \hat{p}_{1} )</td>
<td>( \hat{p}_{2} )</td>
<td>( \hat{p}_{q} )</td>
<td>1</td>
</tr>
</tbody>
</table>

The accuracy assessment sample (\( n_i \)) constructed from the accuracy assessment sample of a change map by Jeon et al. (in press), Class 1 is deforestation, and classes 2 and 3 are no change classes of forest (class 2) and non-forest (class 3). Map categories are the rows while the reference categories are the columns.
which gives a final land change area estimate with a margin of error (at approximate 95% confidence interval) of

\[
\hat{A}_1 \pm 2 \times \sqrt{\frac{\hat{A}_1}{C_6}} = 45,651 \pm 21,502 \text{ ha}.
\]  

The confidence interval quantifies the uncertainty associated with the sample-based estimate of the area of deforestation. Taking this uncertainty into account, the true area of deforestation could be as low as 241,149 ha or as high as 67,153 ha at the 95% level of confidence. Note that even though the confidence interval for the area of deforestation is wide, it does not include the value for the map area of deforestation which is 22,353 ha, highlighting the need to adjust the area obtained from pixel counting by taking into account the information contained in the error matrix.

Estimates of accuracy for this change map, based on the stratified random accuracy assessment sampling design, can be obtained by applying Eqs. (6)–(8) to the estimated error matrix in Table 4:

\[
\hat{U}_1 = \hat{p}_{11} = \frac{1}{3} \sum_{j=1}^{3} \hat{p}_{1j} = 0.012 \div 0.013 = 0.969
\]

\[
\hat{P}_1 = \hat{p}_{11} + \frac{1}{3} \hat{p}_{21} = 0.012 \div 0.026 = 0.480
\]

\[
\hat{O} = \sum_{j=1}^{3} \hat{p}_{j} = \hat{p}_{11} + \hat{p}_{22} + \hat{p}_{33} = 0.012 \div 0.59 + 0.34 = 0.944.
\]

The effect of incorrectly ignoring the stratified design when estimating accuracy is illustrated by computing the estimates directly from the error matrix in Table 3.

### Table 4

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
<th>User’s</th>
<th>Producer’s</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.012</td>
<td>0.0004</td>
<td>0.013</td>
<td>0.97±0.03</td>
<td>0.48±0.23</td>
<td>0.94±0.04</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.006</td>
<td>0.595</td>
<td>0.038</td>
<td>0.639</td>
<td>0.93±0.03</td>
<td>0.99±0.01</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.007</td>
<td>0.003</td>
<td>0.337</td>
<td>0.347</td>
<td>0.97±0.03</td>
<td>0.88±0.04</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.026</td>
<td>0.598</td>
<td>0.376</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.3. Example 2: accuracy and area estimation for post-classification change analysis

In this second example, we illustrate the accuracy and area estimation calculations when change is determined from a post-classification change analysis. Olofsson et al. (2011) mapped changes in forest cover between 2005 and 2010 for the country of Romania. The map was obtained using a post-classification comparison of two forest/non-forest maps, one from 2005 and one from 2010. Accuracy of each of these two forest/non-forest maps was assessed based on a stratified random sample of 57 × 57 m sample units. The sample for assessing the 2005 forest/non-forest map was selected separately from the sample for assessing the 2010 classification. The reference land cover was determined using high resolution imagery in Google Earth™ and Landsat imagery. Both forest/non-forest maps were highly accurate yielding overall accuracies of 96%, user’s accuracies between 92% and 99% and producer’s accuracies over 96% (Table 5). However, as noted in Section 2, the error matrices of Table 5 are often a common endpoint of accuracy assessment of post-classification change analyses, yet these single date error matrices do not provide information relevant to assess the accuracy of gross changes (i.e., gross gain and gross loss of forest).

The two maps were then overlaid to produce the 2005–2010 change map and obvious errors in the forest change categories were manually corrected. An accuracy assessment of the change map was performed using a stratified random sample that was independent of the two samples selected for assessing the accuracy of the 2005 and 2010 forest/non-forest maps (Table 6). The reference land cover of each sample was determined using Google Earth™ and Landsat imagery. The equations of Section 1 were applied to estimate the error-adjusted area of deforestation: \( A_1 = 2 \times S(T_1) = 119,420 \pm 57,550 \text{ ha} \). The area of deforestation calculated from the change map ("pixel count") was 154,159 ha, and this area falls within the 95% confidence interval of the error-adjusted estimated area. This example differs from the example in Section 2 in that the commission error of change becomes prominent in this post-classification change assessment. The area of commission error is removed by the error-adjusted estimate and the

### Table 5

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>Total</th>
<th>W</th>
<th>User’s</th>
<th>Producer’s</th>
<th>Over</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>166</td>
<td>10</td>
<td>176</td>
<td>0.30</td>
<td>0.94±0.04</td>
<td>0.96±0.04</td>
<td>0.97±0.02</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>180</td>
<td>183</td>
<td>0.70</td>
<td>0.98±0.02</td>
<td>0.98±0.02</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>169</td>
<td>190</td>
<td>359</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>160</td>
<td>13</td>
<td>173</td>
<td>0.33</td>
<td>0.92±0.04</td>
<td>0.99±0.02</td>
<td>0.97±0.02</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>189</td>
<td>190</td>
<td>0.67</td>
<td>0.99±0.02</td>
<td>1.0±0.02</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>161</td>
<td>202</td>
<td>363</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
resulting error-adjusted estimated change area (119,420 ha) is smaller than the mapped area (154,159 ha).

Even though the two forest/non-forest maps were highly accurate with user’s accuracies of about 95% (Table 5), the user’s accuracy of the deforestation class in the change map was only 51% (Olofsson et al., 2011), indicating that the forest change obtained by post-classification was inaccurate.

The results of a post-classification comparison are often summarized by a “from-to” change matrix that provides the area of the different possible land cover transitions from one land cover category to another or transitions to one category from another. The example presented in this subsection involved a comparison of two land cover maps (i.e., two dates) of only two classes (forest and non-forest) so there are four possible “cells” in the change matrix (two no change “transitions” and two changes, forest to non-forest, and non-forest to forest). If instead land cover maps containing several classes and/or representing more than two dates are compared, a more elaborate “from-to” transition matrix would be required. For change in one time interval (two time points), the number of elements in the from-to change matrix is the square of the number of land cover classes. For example, overlaying two annual IGBP MODIS Land Cover Products (Friedl et al., 2010), with 17 classes mapped would have produced a change matrix with a staggering 289 elements. To avoid the problems inherent in mapping change by overlaying two maps, the producers of the MODIS Land Cover Product strongly recommend against post-classification comparisons of their products for inferring change (Friedl et al., 2010, p. 177).

It should be noted, however, that situations might arise where post-classification approaches are necessary such as when a more current map is compared to a historical baseline map or when classification training data for the change classes are insufficient for a direct classification of change. Further, post-classification approaches for change detection do have the potential of producing accurate results (Singh, 1989).

### 3.4. Additional context

To provide additional exploration of the accuracy and area estimation issues, we identified from the literature several articles for which the data were available to conduct the complete analysis described in Section 1. A total of 15 accuracy assessments were extracted from these articles. In all cases, the accuracy assessment sampling design was stratified random with the strata defined by the map classes.

The first analysis of these examples examines the relationship between the error-adjusted estimate of area and the mapped area of land change (Fig. 1). To facilitate the comparison of different studies, the error-adjusted area estimates and associated 95% confidence intervals are reported relative to the mapped area of change (i.e., the error-adjusted area estimates are divided by the mapped area of change). Therefore, a scaled value of 1.0 in Fig. 1 would indicate that the mapped area and error-adjusted estimated area are equal whereas a value of 3 would indicate that the adjusted area estimate is three times the mapped extent. Similarly, if the confidence intervals (scaled by map area) include the value of 1.0, the mapped area of change falls within the confidence interval estimated from the error-adjusted area of change suggesting no significant difference.

### 4. Propagating the uncertainty of land change area estimates: a sensitivity analysis of modeled terrestrial carbon flux

The effect of the uncertainty of land change area estimates on applications using these area estimates is illustrated using the results from three terrestrial carbon flux studies that use estimates of land change obtained by remote sensing (Olofsson et al., 2010, 2011). This effect is demonstrated via a simple sensitivity analysis in which the land change area input to a carbon flux model is varied to examine the effect on the output of the model. The land change area values used as input to the model were the error-adjusted estimated area and the upper and lower 95% confidence bounds of the change area from the error-adjusted approach (and the map change area in the case of Romania). The confidence bounds for area of change reflect uncertainty attributable to sampling variability, and this lone source of variation is incorporated in the carbon flux uncertainty analysis. Our intent is to provide a simple, easily conducted analysis to illustrate the potential impact of
variation in the estimated land change area on studies that use such information. More sophisticated error propagation techniques would be required to incorporate multiple sources of uncertainty.

In the first example (Fig. 2a), the area of deforestation for a portion of New England (USA) was estimated by direct classification for two time periods, 1990–2000 and 2000–2005 (Jeon et al., in press). The area estimates were converted to annual rates of change, and these rates were then input to a carbon book-keeping model (e.g. (Houghton et al., 1983)) to estimate the annual terrestrial carbon flux. The same methodology was applied in the second example based on Olofsson et al. (2010) in which the impact of land use change on the terrestrial carbon flux of the country of Georgia was investigated. For the Georgia study (Fig. 2b) the area of deforestation was estimated for a single time period, 1990–2000. For the third example (Fig. 2c), Olofsson et al. (2011) quantified the carbon implications of forest restitution in Romania following the collapse of the Soviet Union. Rates of forest harvest in Romania were estimated for 1990–2000, and 2005–2010. The estimated annual rates of forest change input to the terrestrial carbon model to generate Fig. 2 are listed in Table 7.

The annual terrestrial carbon flux estimated using a model varies greatly because the uncertainty of the estimated area of deforestation used as input to the model is substantial. In Fig. 2a, using the lower confidence bound for the area of deforestation would result in New England being considered a carbon sink throughout the century whereas using the upper bound of deforestation results in a century-long carbon source. The confidence interval for the area of deforestation in the Georgia study is even broader and the uncertainty in land change area estimates has a substantial impact on the predicted terrestrial carbon budget. Using just the mapped area of land change in the Georgia study is even broader and the resulting carbon fluxes (Fig. 2b) predicted by the model consequently vary greatly. Using the upper bound to predict carbon flux results in sink of $-0.46 \text{Tg C}$ in 2007 while the lower bound generates a sink of $-0.09 \text{Tg C}$ as reference, the anthropogenic emissions in Georgia were $1.6 \text{Tg C}$ in 2007 (UN, 2007) which translates into terrestrial carbon offsets of 6% and 28%, respectively (“carbon offset” in this context refers to the percentage of the anthropogenic carbon emissions that is offset by the terrestrial carbon sink). Using the upper confidence bound results in a substantial carbon source starting in 2020 and reaches a maximum of almost $0.5 \text{Tg y}^{-1}$ in 2060 which is about a third relative to the anthropogenic emissions in 2007 and would result in a 33% increase in the total carbon emissions (anthropogenic and terrestrial) of Georgia. This result sharply conflicts with the result obtained using the lower confidence bound of deforestation area which leads to predicting a carbon sink that would persist until 2060 before leveling off at close to zero.

The modeled annual carbon flux of Romania (Fig. 2c) resulting from using different area estimates as model input varies less than the modeled fluxes observed for the New England and Georgia studies. The land change mapped in the Romania study was assumed to be forest harvest where the logged forest was assumed to regrow whereas in New England and Georgia, the mapped change was assumed to be permanent deforestation. As forests regrow following forest harvest, the impact on the carbon budget is less uniform though the margin of error of the 2005–2010 logging area is large (48% of the area). Carbon sequestered in the regrowing forest counterbalances the carbon released from the harvested wood. This is not the case in New England and Georgia.

These examples (Fig. 2) demonstrate that the uncertainty in land change area estimates has a substantial impact on the predicted terrestrial carbon budget. Using just the mapped area of land change in further estimation of important ecosystem variables such as terrestrial carbon flux, may result in estimates that are very different from reality. The examples shown have illustrated situations in which both the magnitude and direction of the carbon flux may vary depending on which estimated values are used. Quantifying the uncertainty of land change area estimates is thus essential. The results of the carbon flux examples highlight the importance of an accuracy assessment to inform uncertainty analyses for applications that use land change area estimates as key inputs.

5. Conclusions

The mapped area of deforestation, forest harvest or any other land change is likely to be different from the actual area because of map classification error even if the change map has a high overall accuracy.

Fig. 2. The annual net flux of terrestrial carbon for (a) New England, (b) Georgia and (c) Romania associated with changes in forest harvest, deforestation, and forest expansion. A carbon book-keeping model was run using the error-adjusted estimated rates of deforestation ((a) also includes the mapped rate) including the lower and upper confidence bounds. All rates were estimated until 2005 (a), 2000 (c) and 2010 (c), and kept constant in the model until 2100.
The confusion matrix used in accuracy assessment provides information on the magnitude of the classification errors that allows an adjustment to be made in the area estimator. A variety of methods may be used to achieve the error-adjustment and here the focus has been on using a stratified estimator that can be applied to a stratified design as well as to simple random and systematic sampling designs (in which case the estimator is called “poststratified”). The error-adjusted area using the stratified estimator avoids the potential measurement bias associated with the mapped area obtained from pixel counting. Estimating the standard error of the stratified area estimator quantifies the uncertainty attributable to sampling variability, and this uncertainty can then be taken into account when estimated land change area is used as input to estimation of ecosystem variables such as the carbon flux between land and atmosphere. The examples provided in Section 4 illustrated that substantial differences in carbon flux predictions could result when uncertainty of the area estimates is accounted for. Knowledge of the uncertainty of the estimates may, however, greatly enhance the understanding of carbon fluxes and inform policy activity linked to them.

Unfortunately, the current status of reporting of accuracy assessment results often falls short of providing the necessary information to fully use the accuracy assessment data. Accuracy measures such as the overall accuracy and kappa are commonly published, but these are insufficient to fully address the area estimation and uncertainty information needs (Section 2). Clearly the remote sensing community understands and recognizes the importance of accuracy assessment because time and resources are often allocated to conduct costly assessments of change maps. Given the proper information from the accuracy assessment, it is not difficult to produce error-adjusted area estimates and confidence intervals (Section 1). The information that should be presented as a routine and an integral part of reporting accuracy assessment includes:

1. a clear description of the sampling design used to collect the accuracy assessment data
2. the error matrix expressed in terms of estimated area proportions (see e.g. Tables 2 and 4)
3. the mapped area or proportion of mapped area of each class
4. estimates of the basic descriptive measures overall, user’s, and producer’s accuracies of change (see Section 1)
5. mapped areas should be adjusted to eliminate bias attributable to map classification error and these error-adjusted area estimates should be accompanied by confidence intervals to quantify the sampling variability of the estimated area (see Section 3.1).

Including this information in the description of methods and reporting of results will enhance the use of the accuracy assessment data beyond the current common practice and lead to better estimates of area, accuracy, and the uncertainty associated with these estimates. It will also allow remotely-sensed data products to be employed more effectively in later analyses such as the estimation of carbon fluxes that are often central to addressing major scientific questions and contributing to policy initiatives.

### Table 7

<table>
<thead>
<tr>
<th>Study</th>
<th>Mapped</th>
<th>Adjusted</th>
<th>Margin of error (95% Cl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE 90-00</td>
<td>4235</td>
<td>10,219</td>
<td>±5302</td>
</tr>
<tr>
<td>NE 00-05</td>
<td>6268</td>
<td>5427</td>
<td>±1577</td>
</tr>
<tr>
<td>Georgia 90-00</td>
<td>1645</td>
<td>1695</td>
<td>±1420</td>
</tr>
<tr>
<td>Romania 90-00</td>
<td>14,067</td>
<td>15,122</td>
<td>±5397</td>
</tr>
<tr>
<td>Romania 05-10</td>
<td>30,832</td>
<td>23,884</td>
<td>±11,510</td>
</tr>
</tbody>
</table>

### Acknowledgment

This research was funded by USGS Award Support for SilvaCarbon and NASA award NNX11AJ79G to Boston University, and USGS Cooperative Agreement G12AC0221 to State University of New York. We thank the anonymous reviewers for the comments that improved the manuscript.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.rse.2012.10.031.

### References


