The effects of omission errors on area and area change estimates

Pontus Olofsson*[,1], Paulo Arevalo[1], Andres Espejo[2], Carly Green[3], Erik Lindquist[4], Ronald E. McRoberts[5], Maria J. Sanz[6]

1. Department of Earth & Environment, Boston University, 685 Commonwealth Avenue, Boston, MA 02215, USA (olofsson@bu.edu)
2. The World Bank, 1818 H Street NW, Washington DC (aespejo@worldbank.org)
3. Environmental Accounting Services, 50 Charles Crt, Lake Hawea, New Zealand (carly.green@enviroaccounts.com)
4. Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy (Erik.Lindquist@fao.org)
5. Department of Forest Resources, University of Minnesota, 1530 Cleveland Avenue North, St. Paul, MN 55108, USA (mcrob001@umn.edu)
6. Basque Centre for Climate Change, Sede Building 1, Scientific Campus of the University of the Basque Country, 48940 Leioa, Spain (mj.sanz@bc3research.org)

*Corresponding author

Original manuscript submitted Jan 23 2019 to Remote Sensing of Environment
Revised manuscript submitted Jun 15 2019 to Remote Sensing of Environment

Olofsson, P.; Arévalo, P.; Espejo, A.B.; Green, C.; Lindquist, E.; McRoberts, R.E.; Sanz, M.J. 2020. **Mitigating the effects of omission errors on area and area change estimates.** REMOTE SENSING OF ENVIRONMENT. 236. DOI (**10.1016/j.rse.2019.111492**).

Highlights

- Literature recommendations related to sampling-based estimation are augmented.
- Omissions of land change in maps can introduce large uncertainty in area estimates of land change.
- If stratifying by map class, omissions of change classes tend to carry large area weight.
- Substrata in forest strata that are unlikely to contain error mitigate effects of omissions.
- Increasing sample size or constructing efficient stratifications mitigate effects of omissions.

Abstract

Information on Earth's land surface and change over time has never been easier to obtain, but making informed decisions to manage land well necessitates that this information is accurate and precise. In recent years, due largely to the inevitability of classification errors in remote sensing-based maps and the marked effects of these errors on subsequent area estimates, sample-based area estimates of land cover and land change have increased in importance and use. Area estimation of land cover and change by sampling is often made more efficient by *a priori* knowledge of the study area to be analyzed (e.g., stratification). Satellite data, obtained free of cost for virtually all of Earth's land surface, provide an excellent source for constructing landscape stratifications in the form of maps. Errors of omission, defined as sample units observed as land change but mapped as a stable class, may introduce considerable uncertainty in parameter estimates obtained from the sample data (e.g., area estimates of land change). The effects of omission errors are exacerbated in situations where the area of intact forest is large relative to the area of forest change, a common situation in countries that seek results-based payments for reductions in deforestation and associated carbon emissions. The presence of omission errors in such situations can preclude the acquisition of statistically valid evidence of a reduction in deforestation, and thus prevent payments. International donors and countries concerned with mitigating the effects of climate change are looking for guidance on how to reduce the effects of omission errors on area estimates of land change. This article presents the underlying reasons for the effects of omission errors on area estimates, case studies highlighting real-world examples of these effects, and proposes potential solutions. Practicable approaches to efficiently splitting large stable strata are presented that may reduce the effects of omission errors and immediately improve the quality of estimates. However, more research is needed before

further recommendations can be provided on how to contain, mitigate and potentially eliminate the effects of omissions errors.

1. Introduction

Remote sensing data suitable for thematic mapping of land surface features – primarily data from the Landsat and Sentinel-2 satellites – are now routinely available free of cost (Woodcock et al., 2008; Wulder et al., 2019). Greater levels of pre-processing by space agencies in combination with powerful open source software and computing platforms (Gorelick et al., 2017) have made it easier than ever to produce maps of land cover and change in land cover and/or use (referred to as land change throughout the article). Still, translating spaceborne measurements of reflected sunlight or backscattered longwave radiation into a set of discrete map classes of complex land surface conditions is inherently complicated and results are bound to be imperfect. Classification errors are inevitable, and their magnitude and distribution will determine the quality and interpretation of a remote sensing-based map (McRoberts, 2011). The communication of map quality within the remote sensing community has traditionally been done by an accuracy assessment based on a comparison of map labels and independent reference observations acquired for locations selected by probability sampling (Stehman, 2000). A reference observation is the most accurate available assessment of the true condition on the land surface. A probability sample allows for inference for various map accuracy measures for the entire population which, in the case of remote sensing-based mapping, is the collection of map units comprising the study area (Stehman, 1997).

Map accuracy assessments grew in importance during the 1980s with the availability of digital remotely sensed data, classification algorithms and processing power, while maps constructed by manual interpretation of remotely sensed imagery – often accepted as correct – became rarer (Congalton, 1991). Assessments were initially focused primarily on overall map accuracy (Congalton, 2004), but the literature from the 1980s and early 1990s highlighted the need for class-specific accuracies such as user's and producer's accuracy (Card, 1982; Congalton, Oderwald, & Mead, 1983; Foody, 1992). Measures of class-specific accuracy enable a more comprehensive investigation of the map quality, especially for rarer classes such as of those of land change. However, often overlooked in the earlier accuracy-themed remote sensing literature is the notion that an analysis confined to map accuracy – being it overall or class-specific – merely indicates the level of map incorrectness (McRoberts, 2011; Olofsson, Foody, Stehman, & Woodcock, 2013). Attempts to estimate the area of a specific map class by methods that sum values for map units assigned to that map class ("pixel-counting") is a biased procedure that produces erroneous area estimates because the effects of classification errors are ignored (GFOI 2016, p. 125). A situation where the effects of errors of omission and commission offset each other is possible but unlikely and cannot be assumed.

While communicating statistically defensible estimates of areas of land cover and land change is of interest to the remote sensing community at large, the paradigm of area estimation has gained additional attention because of the interest in reducing emissions from deforestation and forest degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries (REDD+) negotiated under the UNFCCC (United Nations Framework Convention on Climate Change).. Countries can voluntarily report emissions and removals of carbon dioxide equivalents associated with land use change to REDD+ result-based finance initiatives such as the Forest Carbon Partnership Facility Carbon Fund, Amazon Fund, bilateral programs, etc. The objective of these efforts is to incentivize management of climate change mitigation by providing results based payments to countries providing evidence of reductions of emissions where "evidence" is in the form of inventory estimates using procedures that comply with the good practice guidelines stipulated by the Intergovernmental Panel on Climate Change (IPCC, 2006).

The IPCC identifies two main approaches to inventories: the stock-change approach and the gain-loss approach. The former estimates emissions or removals as the difference in national carbon stocks at two points in time (GFOI, 2016, p. 22). Because the approach is based on estimates of national carbon stocks, an established national forest inventory or other large-scale sampling programs is typically required for implementation of the stock-change approach. For countries without established national forest inventories, which is often the case in tropical countries, the gain-loss approach may be the only alternative (McRoberts et al., 2018). The gain-loss approach estimates net carbon emissions or removals as the sum of gains and losses in carbon pools occurring on areas of land subject to *REDD+ activities* that emit or remove carbon (GFOI, 2016, p. 23). The five REDD+ emission reduction activities are (1) deforestation, (2) forest degradation; (3) conservation, (4) enhancement of forest carbon stocks, and (5) sustainable management of forests (GFOI, 2016, p. 26). The areal extent of the REDD+ activities are referred to as *activity data*. Because activity data are needed for entire countries, and because deforestation is commonly mapped using satellite data, remote sensing is very likely to provide the main source of activity data.

Good practice for reporting activity data and emissions inventories is based on two criteria: (i) "should be accurate in the sense that they are neither over- nor underestimated as far as can be judged", and (ii) "and precise in the sense that uncertainties are reduced as far as practicable" (IPCC, 2006, Volume 1, Chapter 3). Estimating activity data based on pixel-counting in maps, even if accompanied by an assessment of map class-specific accuracy, fails to satisfy these criteria. In particular, pixel-counting is a biased estimator in the sense that on average, it does not produce the true value because of map classification errors (GFOI, 2016, p. 125). Instead, what is needed are confidence intervals for the area estimates which enable quantification of the uncertainty of estimates. REDD+ countries' first submissions of forest reference levels were based solely on pixel-counting, but from 2016 and onwards many countries have chosen to report area estimates and associated uncertainties obtained using methods developed for monitoring and reporting REDD+ activities that are consistent with IPCC good practice (Espejo & Jonckheere, 2017).

Applications in the context of REDD+ have gained attention in recent years, but the importance of statistical properties such as bias and uncertainty are not confined to the REDD+ context but to all remote sensing-based mapping applications. Still, at least up until 2010, bias and uncertainty were largely ignored in the remote sensing literature: an assessment of all articles related to mapping of land change published in *Remote Sensing of Environment* and *International Journal of Remote Sensing* for 2005-2010 showed that all but a few articles failed to include this information (Olofsson et al., 2013). We are not aware of a formal analysis of remote sensing articles published after 2010, but we hypothesize that the situation has changed and that area estimates reported in the literature are more frequently produced from sample data. Several articles, published in remote sensing journals since 2010, have described the need, use and guidance of estimation protocols (McRoberts & Walters, 2012; Olofsson et al., 2014; Stehman, 2013). The evolution of the literature described here is incomplete and omits important earlier contributions to the topic of accuracy and area estimation: Card (1982), complete with equations and numerical examples, made use of an unbiased estimator for estimation of area and map accuracy. Biging, Colby, & Congalton (1998); Macleod & Congalton (1998) discussed issues related to landscape stratifications in sampling-based estimation of accuracy of mapped land change – much of which is related to the topic of this paper. Gallego (2004) provided an excellent review of approaches to area estimation, including a critique of pixel-counting approaches. Additional important contributions that deserve recognition are Congalton & Green (2009); Foody (2002); Stehman & Foody (2008) among many others.

While the remote sensing fire community has primarily focused on validating and comparing burned area products by means of estimated map accuracy, issues similar to those discussed in this paper are discussed in the burned area literature. Topics of discussion include approaches to optimizing stratifications and sample allocations for accommodating omission errors and increasing precision of estimated accuracy (Boschetti et al., 2006; Boschetti, Stehman, & Roy, 2016; Padilla, Olofsson, Stehman, & Tansey, 2017), and proper identification of errors by addressing issues of geolocation and the use of high temporal and spatial resolution reference data (Csiszar, Morisette, & Giglio, 2006).

A contribution that is frequently cited in the remote sensing literature and used extensively within REDD+ is Olofsson et al. (2014), which presents methods for estimating areas of land change and associated confidence intervals, and recommends the use of a map of land cover and land change to define strata for use with a stratified random sampling approach. Activity data are often required at annual or bi-annual intervals (GFOI, 2016); intervals at which the extent of land change tend to be very small relative to stable land cover classes, even in tropical countries that experience relatively large rates of land change. In such cases, applying the stratified random sampling recommended in Olofsson et al. (2014) is likely to create a situation with one or a few very small strata (e.g., deforestation and/or forest regrowth) and one or two very large strata (e.g., intact forest). While defining strata based on map change classes is recommended because it facilitates targeted sampling to ensure sufficient statistical representation of land change (e.g., deforestation and reforestation), small strata that represent areas of interest in combination with a very large stratum of much lower interest, is potentially problematic. The problem arises when change is observed in the reference data at sample locations in the much larger stable (non-change) land cover stratum. Such omissions of land change in the map used to stratify the study area – characterized as *omission errors* – tend to carry large area weights and may result in area estimates with large uncertainty that are very different from mapped areas. The result is an adverse effect on the overall acceptance of the analysis and the ability to detect, in a statistically significant way, variations in the rates of land change over time.

The objectives of this article are to document and explain situations in which omission errors carry large area weights and to propose approaches to mitigate their effects. We also review case studies from various REDD+ countries.

2. Problem statement

2.1 The effects of omission errors on area estimates

Large omission errors are the result of an inefficient stratification which, in turn, often result in large margins of error (i.e., large uncertainties), large differences between the mapped and estimated areas, and wide confidence intervals. Of importance, however, is the recognition that large differences between mapped and estimated areas do *not* mean that the estimation process is erroneous or should be avoided in preference to pixel-counting, but simply mean that the map used to stratify the study area contains classification errors and that the stratification is inefficient.

Consider the error matrices in Table 1. In this hypothetical example, a change map has been constructed showing that stable forest occupies 80% of the study area, deforestation accounts for 0.5% of the mapped area, and non-forest accounts for the remaining 19.5%. A sample of 500 map units has been selected by stratified random sampling using the map classes as strata, and reference conditions have been observed at each sample unit. The sample units were allocated following the recommendations in Olofsson et al. (2014) for area estimation such that 50 were selected in the deforestation stratum and the rest allocated to the other two strata in proportion to their sizes. Seven sample units with forest or non-forest observations were present in the deforestation stratum (i.e., deforestation commission errors in the map; cells shaded blue), and two units observed as deforestation were found in the forest stratum (i.e. deforestation omission errors in the map; cells shaded red). If the sample count for error matrix cell *i*,*j* is denoted n_{ij}, the total number of sample units in stratum *i* is n_{i+} (the plus sign that replaces *j* indicates a sum across the columns in the matrix) and W_i is the weight of stratum *i* defined as the area proportion of the stratum relative the total study area, the estimated area proportion for cell *i*,*j* is

$$\hat{p}_{ij} \ = W_i \times n_{ij} \div n_{i+} \qquad (1)$$

Table 1. *Error matrix expressed as sample counts (upper) and estimated area proportions (lower). Map labels at sample locations are represented by rows and reference observations by columns.*

	Reference					
Stratum	*Defore-*	*Non-*	*Forest*	**Total**	Str. area [ha]	Str. weight, W_i

		station	forest				
M	*Deforestation*	43	2	5	**50**	5,000	0.005
a	*Non-forest*	0	81	9	**90**	195,000	0.195
p	*Forest*	2	10	348	**360**	800,000	0.800
	Total	**45**	**93**	**362**	**500**	**1,000,000**	**1**

M	*Deforestation*	0.0043	0.0002	0.0005	**0.005**	5,000	0.005
a	*Non-forest*	0	0.176	0.020	**0.195**	195,000	0.195
p	*Forest*	0.0044	0.022	0.773	**0.800**	800,000	0.800
	Total	**0.0087**	**0.198**	**0.793**	**1**	**1,000,000**	**1**

The lower error matrix in Table 1 contains the estimated area proportions. The area of deforestation estimated to be mapped correctly is 0.43% of the study area (shaded green; $\hat{p}_{11} = W_1 \times n_{11} \div n_{1+} = 0.005 \times 43 \div 50 = 0.0043$) while the two omissions of deforestation represent an area of 0.44%. Consequently, the omission error of deforestation in the map is larger than the area correctly mapped as deforestation. Note that the area of the commission error is very small. From Eq. 1, it is obvious that the strata weights (W_i) have a large effect on the area represented by the errors. The commission error is small (0.07%) because it occurs in the small deforestation stratum (W_1 = 0.005); even a doubling of the number of commission errors would still only represent 0.14% of the study area. Likewise, the area of the omission errors (0.44%) is large because the errors occur in a large stratum (W_3 = 0.8).

To estimate the area of deforestation, we can apply either a model-assisted regression estimator (McRoberts & Walters, 2012; Särndal, Svensson, & Wretman, 1992) or a stratified estimator (Cochran, 1977; Olofsson et al., 2013). When the sample and map data have been tabulated as in Table 2, the former becomes a bias-adjusted estimator, which subtracts the commission error and adds omission error from the mapped area of deforestation (Eq. 2), and the latter a direct estimator that sums the area of deforestation estimated from the reference observations (Eq. 3) (Stehman, 2013). When applied to the sample data expressed as estimated area proportions ($\hat{p}_{ij}$) in Table 1, both the direct estimator and the bias-adjusted approaches yield the same area estimate (Stehman, 2013):

$$\hat{p}_{j=1} = \hat{p}_{1+} - (\hat{p}_{12} + \hat{p}_{13}) + (\hat{p}_{21} + \hat{p}_{31}) = 0.0087, \qquad (2)$$

$$\hat{p}_{j=1} = \hat{p}_{11} + \hat{p}_{21} + \hat{p}_{31} = 0.0087. \qquad (3)$$

Multiplied by the total study area, the estimated area of deforestation is 8,744 ha, which is considerably larger than the mapped area of 5,000 ha, even though only two omission errors were observed. This is a common situation as shown in Section 2.2. Again, of importance, if sample data have been collected following good practices, *the estimated area is not wrong even if very different from the mapped area*. Keep in mind that all maps have errors and that the use of an unbiased estimator accommodates the effects of map classification errors. However, the presence of errors will affect the width of confidence intervals for the estimates – the larger the errors, the greater the uncertainty. A confidence interval at the 95% confidence level for the deforestation estimate in our hypothetical example is calculated as (Olofsson et al., 2014, Eq. 10; modified from Cochran, 1977, Eq. 5.56):

$$\hat{p}_{j=1} \pm z(0.975)\ \mathrm{SE}(\hat{p}_{j=1}) = \hat{p}_{j=1} \pm 1.96 \left[\sum_{i=1}^{3} W_i \frac{\hat{p}_{i1} - \hat{p}_{i1}^2}{n_{i+} - 1}\right]^{1/2} = 0.0087 \pm 0.0063. \quad (4)$$

Multiplying by the total map area gives an area estimate of deforestation with a 95% confidence interval of $8{,}744 \pm 6{,}289$ ha (i.e., a margin of error of $6{,}289 \div 8{,}744 = 72\%$). The numerator expression of Eq. 4, $W_i(\hat{p}_{i1} - \hat{p}_{i1}^2)$ for $\mathrm{i} = 1, 2, 3$ does not directly include information about commission errors for deforestation as opposed to the omission. Also, the multiplication by W_i suggests that a large stratum weight further exacerbates the effects of the omission on the confidence interval. Hence, a large omission for deforestation will result in a wide confidence interval around the area estimate. In addition to omissions and strata weights, the sample size has a direct effect on the width of the confidence interval. Because the denominator includes the within-strata sample sizes, a larger sample size in the forest stratum would have reduced the uncertainty in the deforestation area estimate. Accordingly, it is possible to counteract a less efficient stratification by increasing the sample size. But the collection of sample data can be a costly and time-consuming process as opposed to constructing a more efficient stratification (Section 3).

2.2 Examples from countries

Central to REDD+ are forest reference levels (FRLs). Countries that participate in REDD+ result-based finance initiatives need to submit a reference level expressed in tons of emitted carbon dioxide equivalents over a historical reference period to which future estimates of emissions are compared for assessing a country’s performance in implementing REDD+ activities (GFOI, 2016). Approximately 70% of countries that submitted a FRL to the UNFCCC in 2018 and 90% of the countries that have submitted their FRLs to the Forest Carbon Partnership Facility of the World Bank provided estimates of activity data with uncertainty quantified using confidence intervals (Espejo & Jonckheere, 2017) following the recommendations in GFOI (2016) and Olofsson et al. (2014). This represents an important milestone for increased transparency in the UNFCCC reporting framework. Uncertainties reported by many of these countries have been affected by omission errors that carry large weight. Figure 1 shows the margin of error for deforestation area estimates for the ten countries in Table 1. These countries are working with the Forest Carbon Partnership Facility (FCPF) to implement Emission Reduction Programs (ER Programs) as a first step in their national implementation of REDD+. The ER program stipulates the requirement to estimate uncertainty related the the activity data (and emission factors and subsequent total FRL and ex-ante emission/removal estiamtes). . From Figure 1, it is obvious that greater uncertainties are associated with small deforestation proportions. While less obvious, larger forest strata tend to result in larger errors. As illustrated below, if the stratum corresponding to a mapped activity (deforestation in this case) carries a very small weight while the forest stratum is large, activity omissions will prevent precise estimation of its area. For the FCPF to make decisions on REDD+ results-based payments to countries, it is essential that estimates of emissions from REDD+ activities are significantly less than both the reference level and previous estimates, but large uncertainties in consecutive estimates make such decision-making difficult, if not impossible

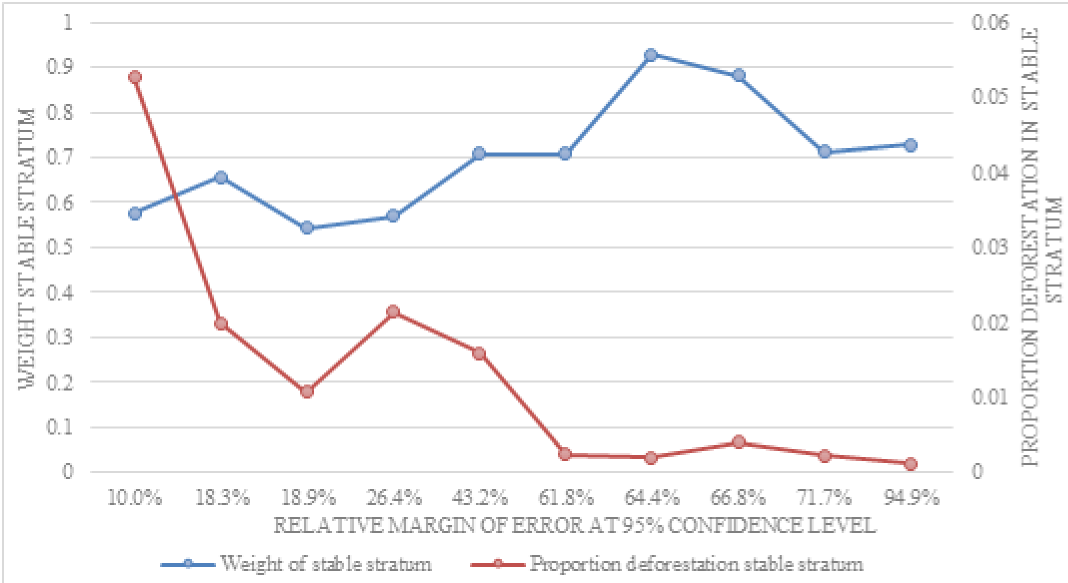


Figure 1. *Relative margin of error at 95% confidence level for deforestation estimates per weight of stable stratum and proportion of deforestation in this stratum*

Table 2. *Difference between mapped areas and area estimates of deforestation and relation to relative margin of errors.*

Case	Mapped [ha]	Estimated [ha]	Relative difference	MoE at 95% confidence
Chile, CF ER program (1997-2008)	21,933	16,512	33%	62%
Chile, CF ER program (2008-2014)	3,644	5,091	-28%	95%
Congo, CF ER program (2003-2012)	157,212	86,590	82%	64%
Congo, CF ER program (2013-2016)	70,930	57,781	23%	67%
Congo, National FREL (2000-2012)	127,000	145,000	-12%	72%
Costa Rica, CF ER program (2001-2011)	222,417	280,602	-21%	26%
Cote D'Ivoire, CF ER Program (2000-2015)	499,655	469,329	6%	7%

Ethiopia, ISFL ER program (2000-2013)	130,296	477,743	-73%	43%
Ghana, CF ER program (2000-2010)	579,990	356,077	63%	18%
Ghana, CF ER program (2012-2015)	790,090	653,428	21%	10%
Madagascar, CF ER program (2005-2015)	575,035	425,154	35%	19%
Madagascar, Easter Humid Ecoregion (2005-2013)	1,930,936	2,119,993	-9%	35%
Mexico, Yucatan CF ER program (2007-2011)	148,089	85,690	73%	67%
Suriname, National FREL (2000-2009)	24,784	35,816	-31%	17%
Suriname, National FREL (2009-2015)	60,362	65,419	-8%	13%
Vietnam, CF ER Program (2000-2005)	177,802	153,705	16%	20%
Vietnam, CF ER program (2005-2010)	124,147	127,618	-3%	22%

The complete list of the area estimates for deforestation that were used to construct Figure 1 is presented in Table 1. Area estimates, mapped (pixel-counted) areas, the difference between mapped and estimated areas, and the margins of error are presented. For example, the Carbon Fund Early Reduction program in Chile reported a dramatic reduction in deforestation from 16,512 ha in 1997-2008 to 5,091 ha in 2008-2014, but with margins of error of 62% and 95% respectively, it is not obvious that the estimates are significantly different as illustrated in Figure 2. Additional analysis is required in this case to determine if the two estimates are significantly different (using, for example, a two sample *t*-test (Rice, 1995, p. 387)). FRL developed by countries are typically estimated over an historical time period of around 10 years borcken into two change periods or around 5 – 7 years. Reporting of results (i.e. comparison of actual reductions achieved when compaed to the FRL baseline) is required at annual or bi-annual intervals (GFOI, 2016). Reporting at such high temporal frequencies has proven difficult because the areas of deforestation and other relevant REDD+ activities tend to be very small at annual intervals (Arevalo, Woodcock, & Olofsson, 2019a). Hence, decreasing the reporting intervals to obtain a larger number of consecutive area estimates would increase the uncertainty of the estimates and further exacerbate the problem of determining if a reduction of deforestation has occurred.

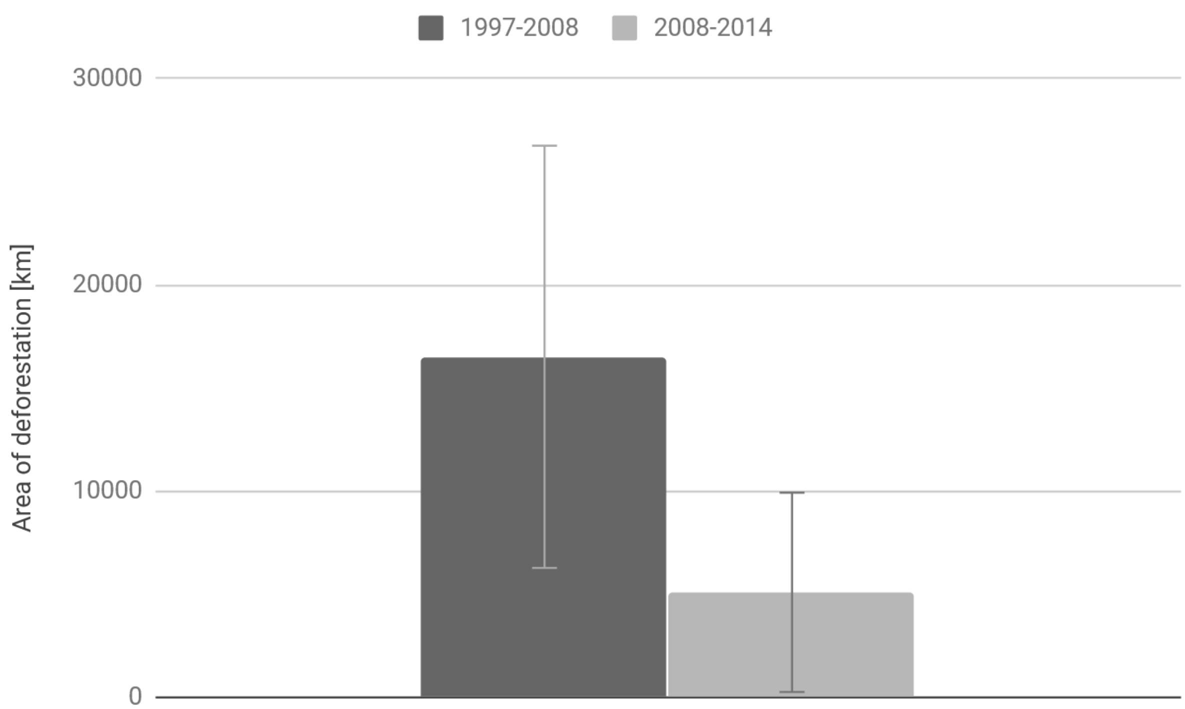


Figure 2. *The estimated area of deforestation with 95% confidence intervals in the Carbon Fund Early Reduction Program in Chile between 1997-2008 and 2008-2014.*

To further illustrate the issue of omission errors, a more detailed example from the FRL submitted to the UNFCCC by Republic of Congo in 2016 is shown below. The FRL targets the REDD+ activity “reducing emissions from deforestation”. The activity data used for constructing the FRL were estimated from sample data collected using stratified random sampling with the stratification constructed from a classification of Landsat data, SPOT 5 data and very fine resolution imagery available in Google Earth. The stratification includes stable forest, stable non-forest and forest cover loss. The error matrices are presented in Table 3 – the strata weights for forest cover loss and stable forest of 0.4% and 71% of the study area indicate that sample units observed as forest cover loss in the stable forest stratum will have a marked effect on area estimates and confidence intervals.

Table 5. *Error matrices expressed as sample counts (upper) and estimated area proportions (lower) submitted by Republic of Congo to UNFCCC for estimation of a FREL 2000-2012. Map labels at sample locations are represented by rows and reference observations by columns.*

		Reference					
	Stratum	*Forest c. loss*	*Non-forest*	*Forest*	**Total**	Str. area [ha]	Str. weight, W_i
Map	*Forest c. loss*	145	7	47	**199**	127,000	0.0037
	Non-forest	0	182	29	**211**	9,673,000	0.2835
	Forest	1	40	419	**460**	24,326,000	0.7128
	Total	**146**	**229**	**495**	**870**	**34,126,000**	**1.000**

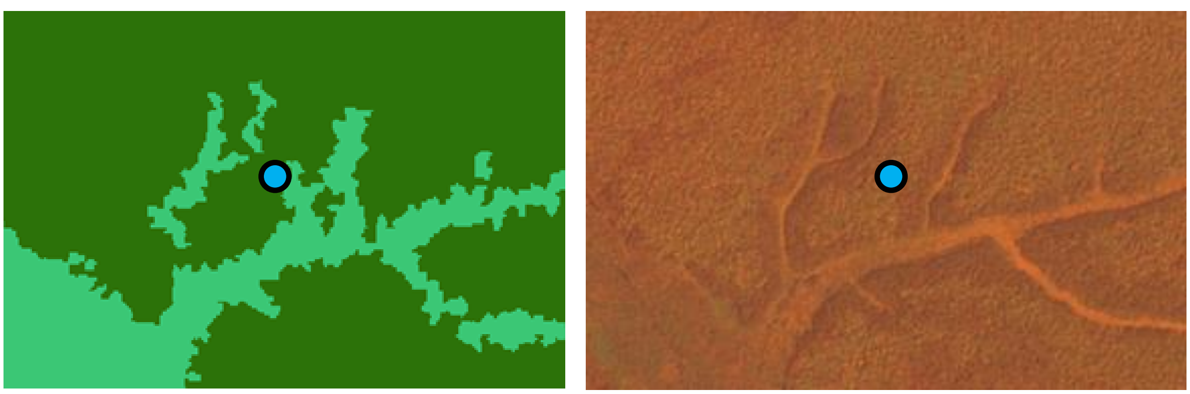
Ma	*Forest c. loss*	0.0027	0.0001	0.0009	**0.0037**	127,000	0.0037
	Non-forest	0	0.2445	0.0390	**0.2835**	9,673,000	0.2835

p *Forest*	0.0015	0.0620	0.6492	**0.7128**	24,326,000	0.7128
Total	**0.004**	**0.307**	**0.689**	**1.000**	**34,126,000**	**1.000**

A total of 870 sample units were selected with about half allocated to the stable forest stratum and the rest split between the forest loss and stable non-forest strata. Only a single omission error was observed, but it represents an area according to Eq. 1 of $\hat{p}_{31} = W_3 \frac{n_{31}}{n_{3+}} = 0.72 \frac{1}{460} =$ 0.15% of the study area, which is almost half of the area of the forest cover loss stratum. In comparison, the 47 + 7 commission errors in the forest loss stratum represent an area of only $\hat{p}_{12} + \hat{p}_{13} = W_1 \frac{n_{12}+n_{13}}{n_{1+}} = 0.0037 \frac{7+47}{199} = 0.10\%$. In this case, the very large number of commission errors "offset" about two thirds of the area of forest loss that was omitted in the map, which results in area estimate (0.43%) that is relatively close to the mapped area (0.37%) of forest loss. The large errors results in an uncertain estimate: expressed in hectares, applying a stratified estimator to the sample data yields an area estimate for forest loss of 145,420 ha and a 95% confidence interval of 104,092 ha, i.e. a margin of error of 72%. Because the estimate is a reference level to which future area estimates of forest loss will be compared, a wide confidence interval will make it difficult to determine if reductions of forest loss occur in the future.

The Republic of Congo example highlights the importance of sample data that represent the best possible assessment of the land surface conditions. The omission of forest loss in Table 3 was observed at a sample location in an area of terra firme and wetland forests with no signs of human intervention (Figure 3). While a loss of forest cover was observed in the reference data (Landsat and Sentinel-2), it is unclear if the loss event was the result of anthropogenic deforestation. It is essential to determine if sample data are collected to estimate *deforestation*, which entails a change in land-use, or simply *forest cover loss*, which includes, in addition to deforestation events, a temporary loss of forest cover. An area estimate for deforestation based on the sample data in Table 3 is 92,538 ± 7,877 ha (i.e., 8.5% margin of error) *if* assuming that the sample unit in question is stable forest, and that all other observations of forest cover loss are deforestation. Such a big difference in area and precision between deforestation and forest cover loss when the only difference in the sample data is the reference label of a single sample unit is not satisfactory. In situations as illustrated in this example, approaches are needed that mitigate the effects of errors. Such approaches are discussed in the next section.

Figure 3. *Sampling unit labelled as deforestation in the forest stratum. Left, overlaid over the forest cover change map (dark green is* forest *and light green* wetland forest*). Right, overlaid over December 2015 Sentinel 2 image in false color (4,6,11).*



Furthermore, the example emphasizes the importance of providing correct reference labels. A single incorrect label may introduce considerable uncertainty as illustrated in this example. Olofsson et al. (2014) recommends three independent reviewers to break ties and an indication of the level of confidence in provided labels – here, we augment Olofsson et al. (2014) by a recommendation to perform a "post-interpretation" to review each of the labels of the sample units in a team effort to identify and correct 1) clerical errors, 2) misinterpretations of reference conditions and 3) errors due to positional accuracy (i.e. a mismatch between map and reference units). The team should consist of at least the sample interpreters and a senior land cover expert. If a very large sample has been collected and a post-interpretation review is not possible, we recommend an approach based on hybrid-inference to incorporate the effects of interpreter errors into the analysis as illustrated in McRoberts et al. (2018).

3. Methods

3.1 Approaches to mitigate the effects of omission errors

From Eq. 1, we can conclude that the magnitude of the omission error depends on the weight and sample size of the stratum in which in the error occurred (the forest stratum in Table 2). If a more efficient stratification could be constructed such that the forest stratum weight could be reduced, the effects of the omission would be reduced. An arbitrary split of the forest stratum would achieve a reduction of the stratum weight but the expected number of omission errors would be proportional to the reduction of the stratum. Such an approach would not reduce the weight of the total omission error. Instead, what is needed is a disproportionate split of the forest stratum into a small substratum that ideally contains all the omission errors and a larger substratum that is free of omission errors. However, obtaining spatial information to achieve such a split is not straightforward. First, we need to distinguish between pre- and post-stratification approaches.

Pre-stratification, or just stratification, is a division of the study region into subregions serving as strata that are non-overlapping, and together comprise the whole region; stratified random sampling consists of simple random sampling within each stratum (Cochran, 1977, p. 89). Because land change tends to comprise small proportions of the landscape, stratified random sampling has the advantage of facilitating sufficient statistical representation of activities of interest, even if rare (Olofsson et al., 2014). Another benefit of stratified sampling is that any desired strata can be constructed provided they are exhaustive and non-overlapping. Therefore, any available information on the likely location of omission errors can and should be used to define strata. An attractive solution, exemplified in the remote sensing literature, is the use of buffer strata to mitigate the effects of omission errors (Arevalo et al., 2019a; Bullock, Olofsson, & Woodcock, 2018; Potapov et al., 2017; Tyukavina et al., 2013). A spatial buffer in this context is an area mapped as forest around pixels mapped as land change (forest loss in Figure 4). An example from a study area in Madre De Dios, Peru, is shown in Figure 4. The map data were extracted from a global map of forest cover change (Hansen et al., 2013), and a buffer (black) of three pixels of forest (green) around all pixels of forest loss (red) was constructed.

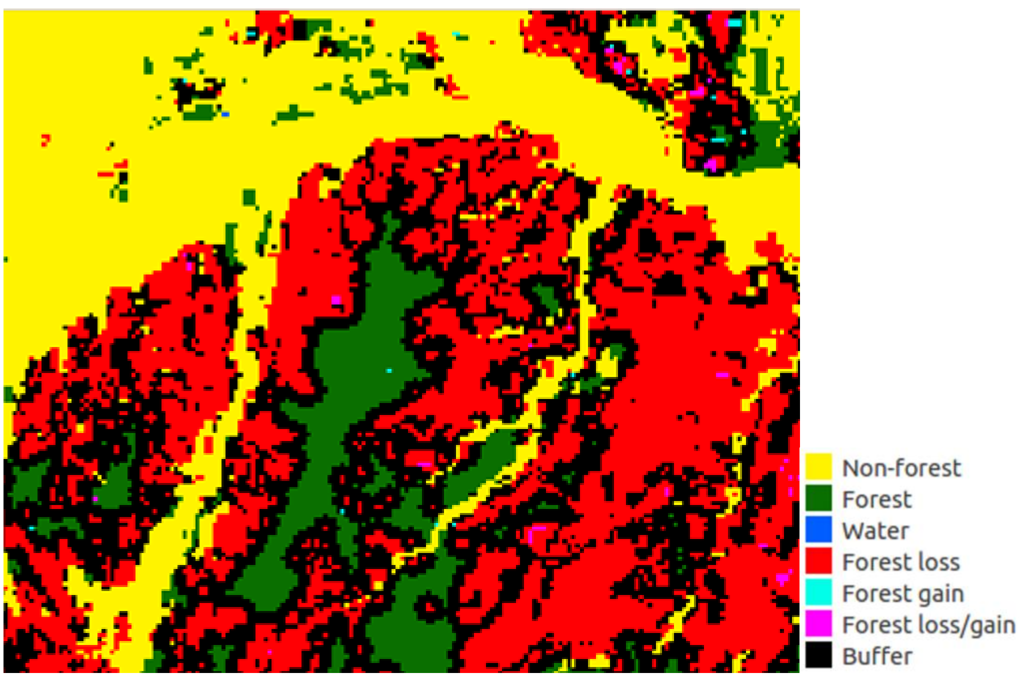


Figure 4. *A buffer stratum created from strata corresponding to the classes of a global change map.*

The hypothesis behind incorporating a spatial buffer into the stratification is that omissions of change typically occur in close proximity to areas of mapped change, while areas mapped as stable forest at larger distances from mapped change are unlikely to contain omissions. Because the buffer stratum in most situations will be much smaller in size than the forest stratum, Eq. 1 indicates that omission errors in a buffer stratum will carry considerably less area weight. Note that the effectiveness of a buffer stratum will decrease with decreasing weight of the forest stratum. Similarly, the effectiveness will decrease with increasing weight of the change strata because this will result in a larger buffer stratum.

The power of using buffer strata to reduce the weight of omission errors was illustrated in Arevalo et al. (2019a) who aimed at estimating the area of conversion between IPCC land categories across the Colombian Amazon at biennial intervals 2000-2016. Independent samples were collected for each biennial interval by stratified random sampling. For each of the biennial stratifications, the forest stratum had a weight of about 0.88 while the forest-to-pasture-conversion stratum (the main carbon-emitting activity) had a weight of only 0.001 on average. Because of the very large difference in strata weights, omissions of forest-to-pasture-conversion in the forest stratum carried a very large area weight. Each sample contained 1,050 units, of which 50 were selected from the forest-to-pasture-conversion stratum and 400 from the forest stratum. In one of the seven samples, a single error of omission was observed but it represented an area proportion of $0.88 \times (1 \div 400) = 0.0022$ or 114 Mha. In comparison, the area estimated as correctly classified as forest-to-pasture-conversion was 40 Mha. In other words, the area of omitted deforestation was three times larger than the area of correctly mapped deforestation! In one bi-annual interval, a single omission error resulted in a confidence interval for the area estimate of forest-to-pasture-conversion that included zero. A lower confidence interval bound less than zero indicates that the deforestation estimate for that was not significantly different from zero preventing further analysis of carbon emissions. However, because the authors could foresee the issue of the omission errors after the maps had been constructed, a buffer stratum of three pixels around each forest-to-pasture-conversion pixel in the forest stratum was constructed for each of the stratifications. The use of buffer strata is, as illustrated above, potentially effective in applications that involve area estimation of rare phenomena. Creating buffers is easy and independent of the approach used to create the initial stratification. In Arevalo et al. (2019), the use of buffer strata resulted in a decrease of the half width of the confidence interval of the area estimates of deforestation by 53 to 98%.

3.2 Simulation of optimal buffer size

An arbitrary buffer size of three pixels around areas of mapped deforestation in the forest stratum was used by Arevalo et al. (2019a), a two pixel buffer was used by Bullock et al. (2018), and a one pixel buffer was used by Potapov et al. (2017) and Tyukavina et al. (2013). It is not straightforward to recommend how to define a buffer stratum to contain omission errors as a buffer's effeciancy depends on the balance between its weight and the number of errors captured. A larger buffer will capture more omission errors thus redcuing the probability of errors occurring in forest stratum but a larger stratum carries a larger weight which increases the impact of the errors on the variance. In an attempt to investigate the impact of size, an omission error probability was calculated for each pixel in a study of the deforestation dynamics of the Colombian Amazon (Arevalo et al., 2019a). The omission probability is based on the cumulative sum of Ordinary Least Square residuals fit over Landsat surface reflectance time series (Arevalo, Woodcock, & Olofsson, 2019b). Each omission probability above 95% was assumed to be an omission error, and thirty buffer strata were created by increments of one pixel. While the number of omission errors captured by the buffer increased with its size, it was found that the number of omission errors in the buffer relative all omission errors in the study area decreased with increasing buffer size. To simulate the impact of varying buffer sizes on the standard error of the deforestation area estimate, the number of omission errors in the forest stratum in the sample data ($n_{F,o}$) for different sample sizes was assumed to be $n_{F,o} = n_F \ \frac{N_o - N_B}{N}$, where n_F is the sample size in the forest stratum, N the total number of pixels of the study area (520,239,684 pixels), N_o the total number of omission errors in the study area (184,050 pixels; we assume that all omission errors were contained by the 30 m buffer), and N_B the total number of omission errors in the buffer stratum ($N_{B=1} = 14{,}763$ and $N_{B=30} = 184{,}050$). The estimated area of the deforestation omitted was assumed to be $\hat{\rho}_{F,o} = W_F \ \frac{n_{F,o}}{n_F}$ where W_F is the weight of the forest stratum. The estimated area of the omission error in the buffer stratum $\hat{\rho}_{B,o}$ was assumed to be difference between $\hat{\rho}_{F,o}$ for different buffer sizes and $\hat{\rho}_{F,o}$ without any buffer. A standard error (Olofsson et al., 2014, Eq. 10) of the deforestation area estimate was calcualted as $\mathrm{SE}(\hat{\mu}_D) = (W_h \hat{p}_h - \hat{p}_h^2) \div (n_h - 1)$ for buffer sizes of 1 to 30 pixels, and for a n_F of 500 to 2,000 in increments of 250. The strata in addition to Forest and Buffer were Non-forest and Deforestion, neither of which contained any omissions of deforestation. The result is shown in Figure 5, with the buffer size yielding the smallest standard error represented by a diamond.

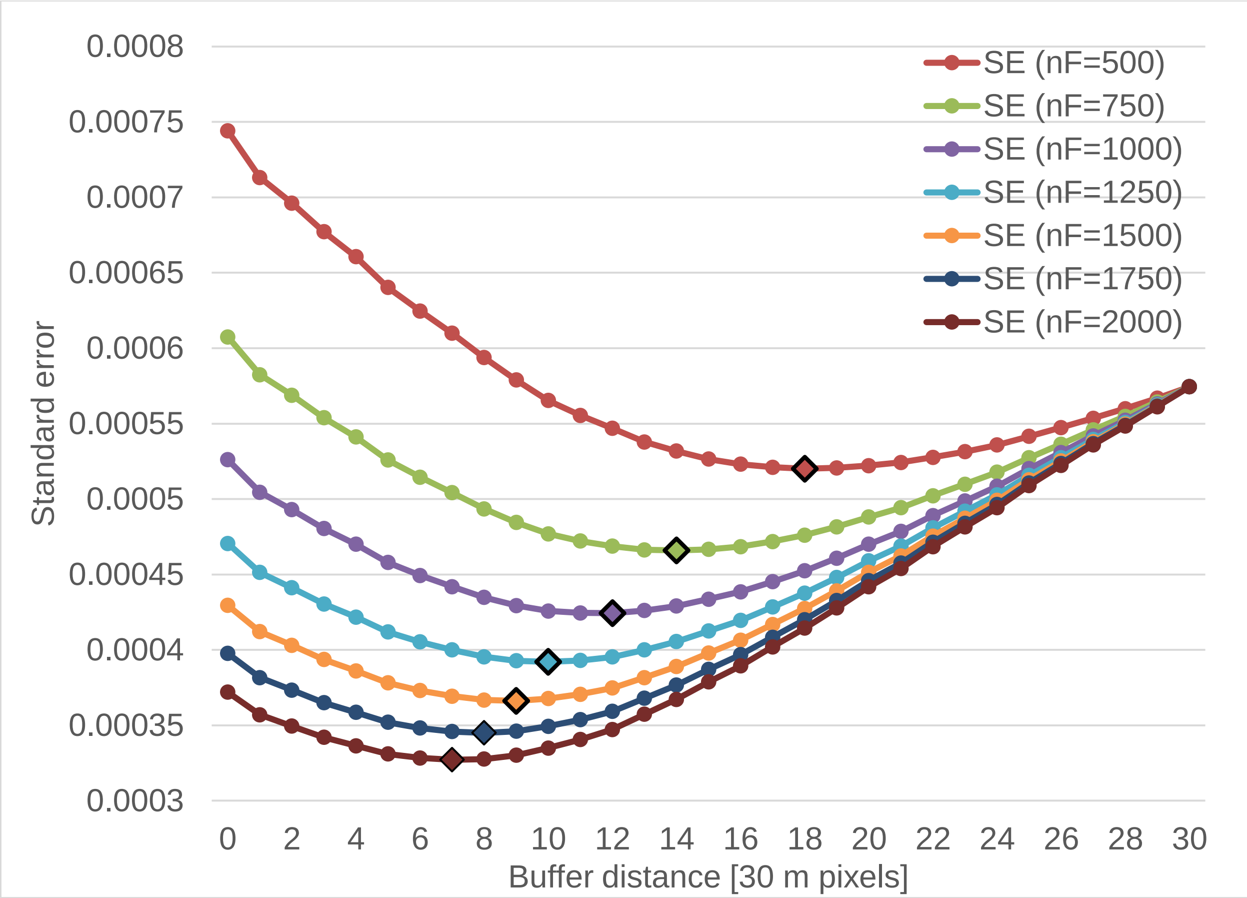


Figure 5. *The standard error of the deforestation area estimate for different sizes of the buffer stratum and for different sample sizes in the forest stratum; diamonds represent the buffer size that gives the smallest standard error.*

Post-stratification refers to a stratification of the study area that is independent of the selection of the sample and applied subsequent to the collection of sample data (Cochran, 1977, p. 134). A common and effective application of post-stratification is the use of a forest/non-forest map in combination with a forest inventory for estimation of forest area (McRoberts, Wendt, Nelson, & Hansen, 2002). Forest inventories are often based on ground plots selected by systematic sampling; stratifying the inventoried area into forest and non-forest will most likely increase precision for estimates of forest area without increasing the sample size. For situations and estimation objectives more relevant to this paper, we typically do not have sample data selected by simple systematic or random sampling but by stratified random sampling, and the feature to be estimated is often a rare phenomenon such as deforestation rather than the area of forestland. In this context, post-stratification is expected to be less relevant. But, post-stratifying the study area will never erode the precision of estimates but will at worst not add anything to the analysis.

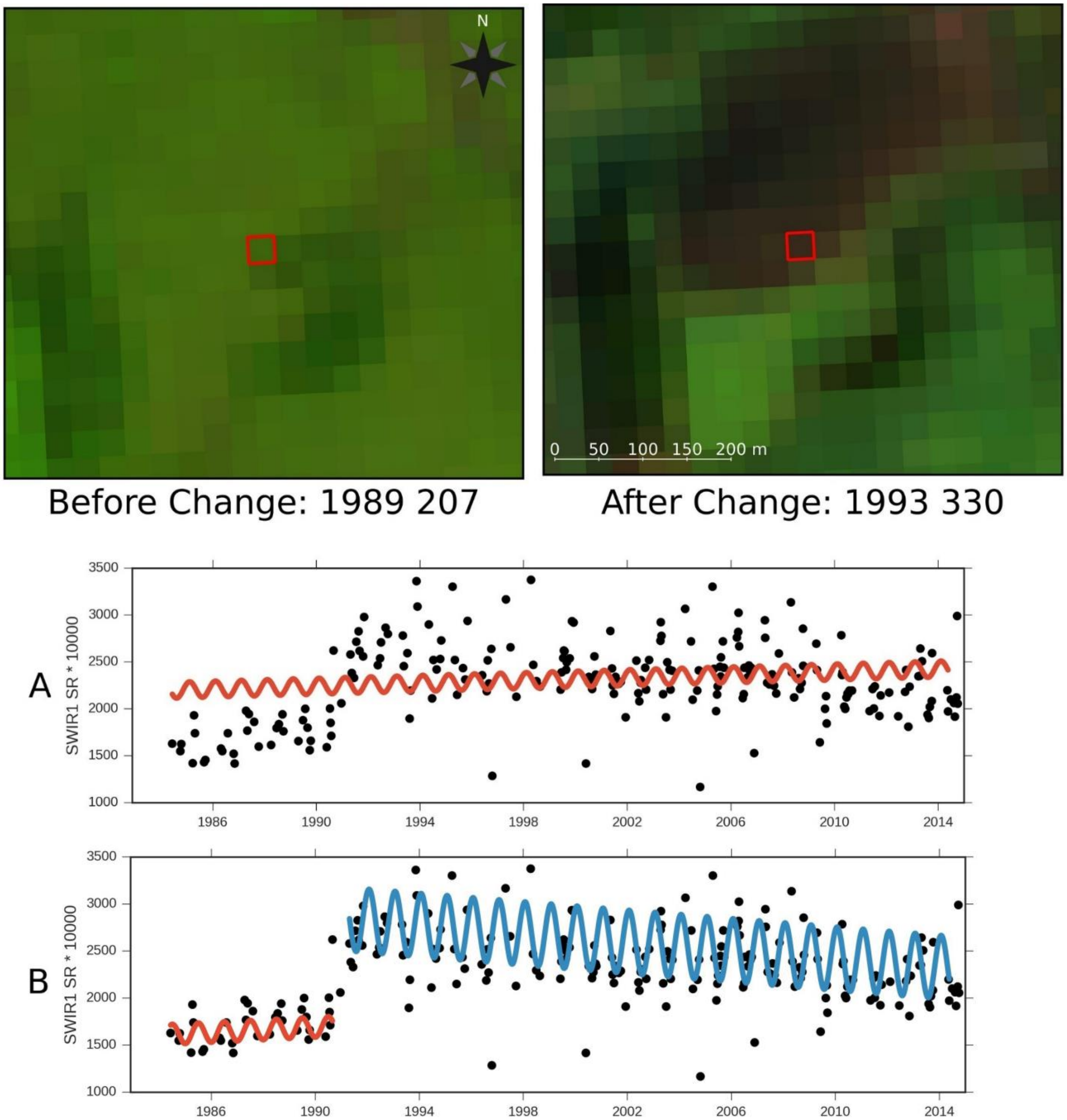
4. Results and Discussion

Figure 5 shows that the buffer sizes used in the published literature of one, two or three pixels are likely smaller than the optimal size, which in this study was found to be seven pixels for a sample size of 2,000 sample units and twelve pixels for 1,000 units allocated to the forest stratum. Note that the simulation results reflect the circumstances in Arevalo et al. (2019a), with weights of the forest and buffer strata ranging from 0.885 and 0 without buffer, to 0.833 and 0.052 for a 30-m buffer, respectively. In situations with more prevalent deforestation and/or less forest area, the optimal buffer size will be different. If the land category to be estimated is larger assuming the number and distribution of omission errors remain the same as in this simulation, a smaller buffer stratum will be optimal. If increasing the size of the deforestation in the simulation such that the deforestation buffer doubles in size compared to the simulation in Figure 5 while reducing the weight of the forest stratum to 0.6, the optimal buffer size is six pixels for a sample size in the forest stratum of 1,000 units (for 2,000 units the optimal buffer is four pixels). A further increase of the deforestation stratum such that buffer is quadrupled shifts the optimal buffer size to three pixels, while a six-fold increase in the deforestation buffer shifts the optimal size to two pixels. Accordingly, the optimal buffer size will depend on the size of forest and deforestation strata, and the sample size in the two strata. A situation such as that of Arevalo et al. (2019a) with a large forest stratum (0.9 weight) and a very small deforestation stratum (0.001 weight), a large buffer of at least ten pixels is recommended while smaller buffer strata are recommended if larger deforestation strata are used.

Buffer strata are not he only means to contain omission errors. We hypothesize that further methods to reduce the effects of omission errors (or errors in general) will be based on the output of the algorithms used to construct the stratification. Providing general guidelines is therefore more complicated, but it is likely that metrics can be extracted that indicate the lack of fit between model and observations for most automated mapping approaches. The larger the residuals, the greater the likelihood of errors. For monitoring algorithms based on comparing predictions to time series of Earth observations (e.g. Verbesselt, Hyndman, Newnham, & Culvenor, 2010; Zhu, Woodcock, & Olofsson, 2012), such metrics are readily available. For example, consider the situation in Figure 6, which shows a deciduous forest pixel in the state of Massachusetts in the USA. A deforestation event as a result of urbanization occurred in the pixel as evidenced by the increase in short wave reflectance in 1991 (Figure 6A and B). Figure 6A shows a time series of Landsat observations of shortwave infrared surface reflectance; a prediction model (red irregular line) is fit by the YATSM algorithm (Holden, 2015) to the initial observations in the time series and updated and compared to subsequent observations (black dots) to detect change on the land surface. If a change is detected, the prediction breaks and a new prediction is initiated when sufficient observations are available after the change event. The result is one or more time series segments at each pixel. The segments are classified together with training data to characterize the timing and the to/from land covers. An omission error occurs in the pixel in question because the prediction model in Figure 6A fails detect the deforestation event in 1991, resulting in a single segment that is incorrectly classified as forest.

Correct monitoring of the land surface should have resulted in Figure 6B with the red segment classified as forest and blue as forest-to-urban. By analyzing the residuals of the observations and predictions, information about the likelihood of omission errors can be obtained. An interesting issue that needs discussing is when to use the type of information generated by analyzing residuals in a time series-based approach to change monitoring. Instead of using the information to stratify the study area, a map maker could re-process the pixels identified as likely omission errors to improve the quality of the map. Using the information at different stages in the workflow will be or more less efficient and result in more or less precise estimates – more research is needed to provide guidance for such decisions.

Finally, in addition to the approaches illustrated in this section we want to reemphasize the importance of a post-interpretation review of the sample data to eliminate clerical errors and misinterpretations. Creating stratifications that are more efficient and performing residual analyses might all be in vain if the sample data are erroneous.

558 

559 **Figure 6.** *A deforestation event mapped by the YATSM algorithm. Figure 5A shows a time series*
560 *of Landsat observations of shortwave infrared surface reflectance (black dots) and the YATSM*
561 *prediction model (red squiggly line) failing to detect the event, as opposed to the model in Figure*
562 *5C.*
563

5. Conclusions

Omission errors – especially sample units observed as land change that occur in larger strata corresponding to stable land cover classes – have been shown to have a profound adverse effect on area estimates. In the REDD+ context, the effect has been found to be especially problematic because the areas of REDD+ activities linked to results-based payments, typically deforestation, tend to be very small relative to the large forest stratum, and discerning a reduction in deforestation by comparing area estimates over time is difficult. The issue of omission errors is not confined to REDD+ but applies to any remote sensing-based mapping application that aims at estimating rare phenomena on the land surface. In this article, we augmented the recommendation in Olofsson et al. (2014) of constructing strata that correspond directly to map classes by recommending a split of larger strata (typically the forest stratum) into a smaller substratum that is likely to contain the omissions of the activities of interest and a larger substratum that is unlikely to contain omission errors. While not always sufficient to resolve the issue, constructing a stratum corresponding to a buffer around activities prior to sampling is a simple but potentially effective way to contain omission errors. The optimal size of the buffer will vary with the weights of the activty data and forest strata and sample size but a buffer of at least three pixels is likley to be optimal. Post-stratification in this context is likely to be less efficient but it will not result in a decrease of the precision of estimates. Further approaches that need more exploration are based on the analysis of residuals of models and observations, which are likely to contain valuable information about the likelihood of omitted activities. We call upon the research community to employ, explore and document these and other approaches to create more efficient stratifications in sample-based estimation of land surface activities such as deforestation, forest degradation, and forest expansion.

Acknowledgements

This research was funded by support from the NASA Carbon Monitoring System and USGS/SilvaCarbon to Boston University (PI Pontus Olofsson). M.J. Sanz was supported by the Spanish Government through María de Maeztu excellence accreditation MDM-2017-0714. Carly Green was supported by the Australian Government as part of its contribution to the Global Forest Observations Initiative.

References

Arevalo, P., Woodcock, C. E., & Olofsson, P. (2019a). Continuous monitoring of land change activities and post-disturbance dynamics from Landsat time series: a test methodology for REDD+ reporting. *Remote Sensing of Environment*, *In press*. https://doi.org/10.1016/j.rse.2019.01.013

Arevalo, P., Woodcock, C. E., & Olofsson, P. (2019b). Spatial representation of the likelihood of errors in maps of land change. *Remote Sensing of Environment*, *in review*.

Biging, G., Colby, D., & Congalton, R. (1998). Sampling Systems for Change Detection Accuracy Assessment. In R. Lunetta & C. Elvidge (Eds.), *Remote Sensing Change Detection Environmental Monitoring Methods and Applications* (pp. 281–308). Chelsea, MI: Ann Arbor Press.

Boschetti, L., Brivio, P. A., Eva, H. D., Gallego, J., Baraldi, A., & Gregoire, J.-M. (2006). A sampling method for the retrospective validation of global burned area products. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(7), 1765–1773.

Boschetti, L., Stehman, S. V., & Roy, D. P. (2016). A stratified random sampling design in space and time for regional to global scale burned area product validation. *Remote Sensing of*

Environment, *186*, 465–478.

Bullock, E. L., Olofsson, P., & Woodcock, C. E. (2018). Monitoring Tropical Forest Degradation using Spectral Unmixing and Landsat Time Series Analysis. *Remote Sensing of Environment*, *In press*. https://doi.org/10.1016/j.rse.2018.11.011

Card, D. H. (1982). Using map category marginal frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering and Remote Sensing*, *49(12)*(3), 431–439.

Cochran, W. G. (1977). *Sampling Techniques*. New York, NY: Wiley.

Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, *37*, 35–46.

Congalton, R. G. (2004). Putting the Map Back in Map Accuracy Assessment. In R. S. Lunetta & J. G. Lyon (Eds.), *Remote Sensing and GIS Accuracy Assessment* (pp. 1–12). Boca Raton: CRC Press.

Congalton, R. G., & Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. *Trends in Ecology and Evolution*. Boca Raton, FL: Lewis Publications.

Congalton, R. G., Oderwald, R. G., & Mead, R. A. (1983). Assessing Landsat classification accuracy using discrete multivariate statistical techniques. *Photogrammetric Engineering and Remote Sensing*, *49*(12), 1671–1678.

Csiszar, I. A., Morisette, J. T., & Giglio, L. (2006). Validation of active fire detection from moderate-resolution satellite sensors: the MODIS example in northern eurasia. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(7), 1757–1764.

Espejo, A., & Jonckheere, I. (Eds.). (2017). *Proceedings: Technical Workshop on Lessons learned from Accuracy Assessments in the context of REDD+*. Rome: FAO HQ.

Foody, G. M. (1992). On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, *58*(10), 1459–1460.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*, 185–201.

Gallego, F. J. (2004). Remote sensing and land cover area estimation. *International Journal of Remote Sensing*, *25*(15), 3019–3047.

GFOI. (2016). *Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative* (2nd ed.). Rome: Food and Agriculture Organization.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27.

Hansen, M. C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S. a, Tyukavina, A., … Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–3.

Holden, C. E. (2015). Yet Another Time Series Model (YATSM). Zenodo. 10.5281/zenodo.17129. Retrieved from https://github.com/ceholden/yatsm

IPCC. (2006). *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. (H. S. Eggleston, L. Buendia, K. Miwa, T. Ngara, & K. Tanabe, Eds.). IGES, Japan.

Macleod, R. D., & Congalton, R. G. (1998). Quantitative comparison of change-detection algorithms for monitoring eelgrass from remotely sensed data - ScienceBase-Catalog. *Photogrammetric Engineering and Remote Sensing*, 207–216.

McRoberts, R. E. (2011). Satellite image-based maps: Scientific inference or pretty pictures? *Remote Sensing of Environment*, *115*(2), 715–724.

McRoberts, R. E., Stehman, S. V., Liknes, G. C., Næsset, E., Sannier, C., & Walters, B. F. (2018). The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *142*, 292–300.

McRoberts, R. E., & Walters, B. F. (2012). Statistical inference for remote sensing-based estimates of net deforestation. *Remote Sensing of Environment*, *124*, 394–401.

McRoberts, R. E., Wendt, D. G., Nelson, M. D., & Hansen, M. H. (2002). Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment*, *81*(1), 36–44.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V, Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, *148*, 42–57.

Olofsson, P., Foody, G. M., Stehman, S. V, & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, *129*, 122–131.

Padilla, M., Olofsson, P., Stehman, S. V., & Tansey, K. (2017). Stratification and sample allocation for reference burned area data. *Remote Sensing of Environment*, *203*, 240–255.

Potapov, P. V, Siddiqui, B. N., Iqbal, Z., Aziz, M. T., Zzaman, B., Islam, A., … Hansen, M. C. (2017). Comprehensive monitoring of Bangladesh tree cover inside and outside of forests, 2000-2014. *Environmental Research Letters*.

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury

Press.

Särndal, C. E., Svensson, B. H., & Wretman, J. H. (1992). *Model assisted survey sampling*. Springer.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*, 77–89.

Stehman, S. V. (2000). Practical Implications of Design-Based Sampling Inference for Thematic Map Accuracy Assessment. *Remote Sensing of Environment*, *72*(1), 35–45.

Stehman, S. V. (2013). Estimating area from an accuracy assessment error matrix. *Remote Sensing of Environment*, *132*, 202–211.

Stehman, S. V, & Foody, G. M. (2008). Accuracy Assessment, 297–310.

Tyukavina, A., Stehman, S. V, Potapov, P. V, Turubanova, S. A., Baccini, A., Goetz, S. J., … Hansen, M. C. (2013). National-scale estimation of gross forest aboveground carbon loss: a case study of the Democratic Republic of the Congo. *Environmental Research Letters*, *8*(4), 044039.

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, *114*(1), 106–115.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., & Cohen, W. B. (2008). Free Access to Landsat Imagery. *Science*, *320*(May), 1011–1012.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., … Zhu, Z. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, *225*, 127–147.

Zhu, Z., Woodcock, C. E., & Olofsson, P. (2012). Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment*, *122*, 75–91.