Increasing the efficiency of forest clearing estimation in the Legal Amazon using targeted sampling

Mark Broich [1]
Stephen V. Stehman [2]
Matthew C. Hansen [1]
Peter Potapov [1]
Yosio E. Shimabukuro [3]

[1] South Dakota State University – SDSU/GIScCE
Box 506B - Wecota Hall - Brookings - SD, USA
Mark.Broich@sdstate.edu
Matthew.Hansen@sdstate.edu
Peter.Potapov@sdstate.edu

[2] State University of New York – SUNY/ESF
1 Forestry Drive - Syracuse - NY, USA
svstehma@syr.edu

[3] Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
yosio@dsr.inpe.br

Abstract. Sampling designs are commonly used to estimate deforestation in the humid tropics. In addition to deforestation estimates, sampling designs provide estimates of precision. In this study, precision and efficiency of two sampling designs estimating the loss of intact forest in the Brazilian Legal Amazon are quantified and compared. The systematic design, proposed for the UNFAO Forest Resource Assessment 2010, utilizes sample blocks at every degree Lat/Long intersection. The targeted design uses MODIS indicated deforestation for stratification and sample block allocation into areas of deforestation hotspots. PRODES data available for the entire study area were used to quantify per sample block deforestation, which allowed rigorous comparison of sampling designs. While the systematic design had error margins equal to 18% of the true value, the targeted design provided more precise estimates (within 11.5%) and required less than half of the samples used in the systematic design. The targeted design proposed here can provide effective and timely, annual estimates utilizing MODIS detected deforestation in combination with the existing PRODES methodology to interpret sample blocks. The implementation of the targeted design can make an important contribution to policy-support and assessments of policy effectiveness in the form of precise, cost effective, and timely annual deforestation estimates.

Keywords: Deforestation, Landsat, MODIS, PRODES, UNFAO Forest Resource Assessment 2010.

1. Introduction

1.1 Quantification of humid tropical deforestation

Providing robust, repeatable, cost effective, and timely information on deforestation for humid tropical forests is still challenging, despite methodological advances and improved data availability. Coarse resolution remote sensing data are inadequate for accurate deforestation mapping (Morton et al. 2005) as most deforestation occurs at sub-pixel scale (Hansen et al. 2008a). High spatial resolution data, on the other hand, allow for more accurate quantification of deforestation area. Wall-to-wall mapping and sampling are the two main approaches that use high-spatial resolution data for estimating deforestation.

The Brazilian National Institute for Space Science's (INPE) Monitoring the Gross Deforestation in the Amazon Project (PRODES) provides annual wall-to-wall deforestation maps of the Brazilian Legal Amazon (INPE, 2008). PRODES delineates deforestation using high spatial resolution remote sensing imagery and a semi-automated interpretation approach (Shimabukuro et al. 1998). As high data volumes and long processing time precludes timely reporting, PRODES provides an annual preliminary approximation of deforestation in the Brazilian Legal Amazon (BLA) by mapping a subarea of the BLA. The subarea consists of the Landsat footprints that contained approximately 90% of deforestation in the previous year (INPE, 2008), resulting in a sample of ~35% of the study area. However, such an approach that maps only a portion of the study area without using a probability-based sampling frame can not quantify the precision of the estimate.

Statistically rigorous sample-based approaches have been discussed as a cost and time efficient alternative if the objective is to estimate rather than explicitly map deforestation (Stehman et al. 2005). Sample-based methods provide a measure of precision that bounds the final estimate, something not possible with non-validated map products. However, they need to be effective in precisely estimating deforestation, which does not occur spatially random but concentrated. Various studies estimating deforestation over large areas used stratified sampling approaches (FAO, 1996; FAO, 2001, Achard et al. 2002). All of these studies utilized expert opinion to divert more samples into areas with higher expected deforestation. A systematic design consisting of small data blocks sampled at every degree Lat/Long intersection has been proposed for the upcoming UNFAO Forest Resource Assessment 2010 (FRA 2010, (Ridder 2007). In order to minimize the area sampled and yet provide precise deforestation estimates, Hansen et al. (2008a) developed a stratified sampling design for the humid tropical forest biome using an automated MODIS forest clearing indicator product to target deforestation.

1.2 Study area

Our area of interest was the humid tropical forest biome (Olson et al. 2001) within the BLA. The Amazon forest within Brazil is of great interest as it is "one of the world's most important bioregions, harboring a rich array of plants and animal species and offering a wealth of goods and services to society "(Foley et al. 2007). Annual forest loss in the BLA has been estimated to range between 11.500 km^2 and 27,400 km^2 in the past ten reporting years (INPE 2008). Within the humid tropical forest biome (hereafter referred to as the humid tropics) inside the BLA, we delineated our study area as the area that was mapped by PRODES between 2000 and 2005 (hereafter referred to as the PRODES mapped area; figure 1). This area featured wall-to-wall PRODES deforestation data and complete cover ancillary information in the form of MODIS deforestation data (Hansen et al. 2008c). The available combination of information and the concentration of substantial deforestation in the region made it an ideal testing ground for the comparison of sampling designs.

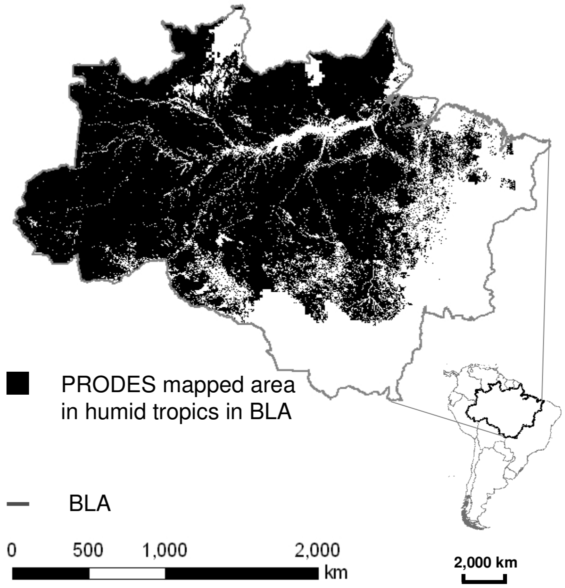

Figure 1: Study area: PRODES mapped area (2000-2005) within the humid tropics in the BLA, Brazil, South America.

1.3 Objectives

Our overarching aim was to demonstrate the effectiveness of sampling designs for estimating deforestation within the area inside the Brazilian humid tropics as mapped by INPE's (2008) PRODES product between 2000 and 2005. This is achieved by determining and comparing the precision of stratified and systematic designs, both using PRODES data to quantify deforestation within sample blocks. Effectiveness was defined as the ability of a given design to archive high precision based on a small sample size.

2. Methods

2.4 Data

The main data source used in this study was PRODES. The PRODES model quantifies clearing of the remaining intact forests of the BLA (INPE, 2008). We defined our area of interest as the aggregate of PRODES forest, PRODES deforestation identified within the 2000-2005 interval, and PRODES cloud coverage (figure 1; Appendix Table 1). Our area of interest was divided into 11,630 sample blocks each measuring 18.5 by 18.5 km. Area of forest loss within each sample block was defined as the aggregate of PRODES classes that represent deforestation between the nominal 2000 to 2005 study interval (Appendix Table 1). The area of these classes added up to 109,000 km^2 of deforestation. This number is smaller than the official PRODES deforestation area reported by INPE (2008) as we excluded uncertain PRODES-deforestation classes and deforestation outside the humid tropical biome that made up our study area.

MODIS-deforestation data at 500m resolution (Hansen et al. 2008c) within our area of interest was upscaled to the sample block scale (Hansen et al. 2008a) to serve as a continuous change indicator utilized in the stratified design.

2.5 Stratified design

The stratification was based on the MODIS indicated change per block. Cut points between strata were determined based on the Dalenius-Hodges rule. Low, medium and high change likelihood strata were defined as 0 – 1%, 2 – 7%, and >= 8% per sample block

MODIS indicated change, respectively (figure 2). We allocated a total of 150 sample blocks to the three strata based on the Neyman optimal allocation formula driven by per stratum MODIS-indicated variance. We used PRODES deforestation in planning calculations to derive the design's precision. We also assessed the precision achievable when using stratum specific regression estimators based on MODIS indicated change.

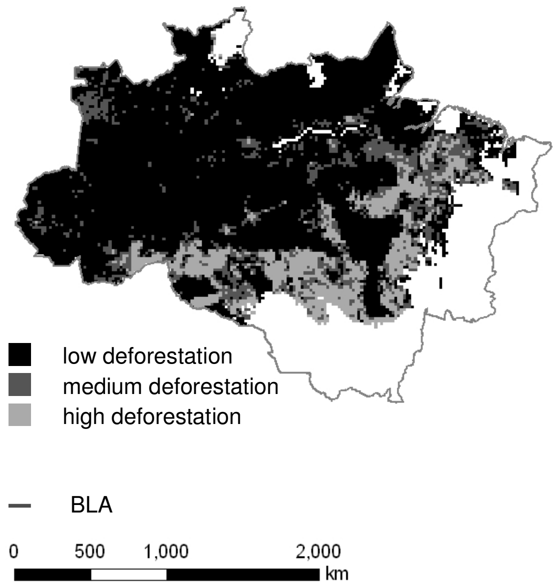

Figure 2: Sample blocks covering the PRODES mapped area divided into three strata MODIS deforestation indicator data.

2.6 Systematic design

For the systematic design we took sample blocks that were located at the one degree Lat/ Long intersection, which mimics the systematic design proposed for the FRA 2010 (Ridder 2007). In addition, we took sample blocks at every half degree Lat/ Long intersection to assess whether the increased number of sample blocks would translate into a substantial improvement in precision. We determined the precision of one degree and half degree grid samples as a function of all possible sample realizations within the study area.

2.7 Comparison between estimation methods

We compared the precision of a simple random sample (SRS) with 150 blocks, the stratified, and the systematic designs. The comparison of the stratified design with the SRS was used to show the gain in precision archived by the stratification. In addition, the usefulness of a MODIS-based regression estimator was assessed. Further, we compared the precision of the aforementioned designs with the precision of the systematic designs. Precision was defined as the margin of error at the 95% confidence level expressed as percentage of the true deforestation.

3. Results

The stratified designs sampled 1.5% of the study area while the one degree and half degree systematic designs sampled 3% and 11% of the study area, respectively. The distribution of the blocks among the strata is shown in Table 1. The table also shows how the study area is divided between the strata and the percentage of true deforestation within each stratum. The precision expected from the different designs is shown in Table 2.

Table 1: Number and percentage of sample blocks per stratum for the stratified and systematic designs, per stratum percentage of study area, and percentage of true deforestation.

Strata	% of study area	% of true deforestation	Number of sample blocks (% of sample blocks)		
			Stratified designs	Systematic designs	
				One degree	Half degree
Low change	74	8	26 (17%)	248 (96%)	971 (74%)
Medium change	14	22	20 (13%)	45 (14%)	186 (14%)
High change	12	70	104 (69%)	32 (10%)	153 (12%)
Total	100	100	150 (100%)	325 (100%)	1,310 (100%)

Table 2: Precision of the different sampling designs; * marks use of regression estimators.

Designs	Simple random sampling	Stratified designs	Systematic designs	
			One degree	Half degree
% Precision	34.3	13.4 and 11.5*	18.1	6.6

4. Discussion

The MODIS-stratified design that used MODIS to guide sample allocation effectively allocated samples into the high change stratum where the majority of the deforestation is concentrated (Table 1). The design was more than twice as precise as the SRS design (Table 2). The use of stratum specific, MODIS-based regression estimators further improved precision to 11.5% (Table 2), which is comparable to the precision obtained by FAO (2001), Achard et al. (2002), and Hansen et al. (2008a).

The precision of the one degree Lat/Long grid systematic design was lower than the precision of the stratified design even though it used twice as many sample blocks (Tables 1 and 2). The precision of the half degree Lat/Long systematic design provided a large improvement over the precision of the one degree grid systematic sample and had the highest precision calculated in this study (Table 2). Yet, most of the sample blocks were located in an area where only a small fraction of the total deforestation occurred (Table 2). The half degree systematic design also used more than eight time as many sample blocks as the stratified design, which was the most efficient design of this study.

Our study illustrated that the combination of the targeted sampling design with Shimabukuro's et al. (1998) semi-automated PRODES methodology to interpret sample blocks can provide precise estimates of deforestation by sampling only 1.5% of the PRODES mapped area. Our design may provide an efficient alternative to the current preliminary PRODES estimation procedures as it requires the interpretation of a much smaller area. Additionally, our probability sampling design has the advantage that it estimates the precision of the estimate, something not possible with the preliminary PRODES estimation procedure. Lastly, our targeted sampling approach can easily be repeated for a given year as the annual MODIS deforestation product can be made available with little effort.

5. Conclusions

The precision of deforestation estimates in the area mapped by PRODES as achieved by different sampling designs varies as a function of the design and the number of sample blocks used. With this application to the PRODES mapped area, we demonstrated the effectiveness of a targeted sampling design that employed MODIS at the design and analysis stage, to precisely estimate deforestation at sub-continental scale. The achieved precision was as high as the precision achieved by other studies (FAO, 2001; Achard et al. 2002; Hansen et al. 2008a) while requiring half of the sample blocks necessary for the systematic design proposed for the FAO's FRA 2010.

Updating the design with annual MODIS deforestation probability maps would allow annual estimates. Image pairs for sample blocks would be classified using the semi-automated PRODES interpretation protocol (Shimabukuro et al. 1998). Such a combined approach would provide precise annual deforestation estimates long before the wall-to-wall PRODES product would be completed. Our probability-based sample approach would sample from the entire area of interest, unlike the preliminary PRODES approximation that focuses only on the change hot spots of the previous year. Using a design based inference framework, our approach would also provide information on the precision of the estimate (Stehman, 2000), which is not possible with the preliminary PRODES estimation procedure. At the same time, we expect our product to correspond well with the preliminary PRODES approximation and the wall-to-wall map.

This study provides a blueprint for a robust, cost and time effective targeted sampling design for estimating deforestation in the area mapped by annual PRODES. Implementing the abovementioned targeted sampling methodology would make an important contribution to policy-support and assessments of policy effectiveness.

Acknowledgements

Support for this work was provided by NASA's Land Cover and Land Use Change program under grand NNG06GD95G.

Appendix

Table 1: Classes in the PRODES data (INPE, 2008) and their aggregation into PRODES mapped area and PRODES deforestation 2000-2005 as defined in this paper

ID	CLASS	Type	PRODES mapped area	PRODES deforestation 2000 - 2005
1	d2002_5	*deforestation uncertain if w/in 00-05*	X	
2	d2001_3	*deforestation uncertain if w/in 00-05*	X	
3	d2004_6	*deforestation uncertain if w/in 00-05*	X	
4	d2004_2	*deforestation*	X	X
5	d2002_0	*deforestation*	X	X
6	nao_floresta2_2003	*non-forest*		
7	nao_floresta	*non-forest*		
8	d2003_1	*deforestation*	X	X
9	d2005_7	*deforestation uncertain if w/in 00-05*	X	X
10	d1997_0	*non-forest*		
11	d2005_8	*deforestation uncertain if w/in 00-05*	X	
12	d2004_0	*deforestation*	X	X
13	d2000_0	*non-forest*		
14	d2005_0	*deforestation*	X	X
15	d2005_4	*deforestation*	X	X
16	hidrografia2004	*non-forest*		
17	nao_floresta2_2004	*non-forest*		
18	d2005_3	*deforestation*	X	X
19	d2004_7	*deforestation uncertain if w/in 00-05*	X	
20	d2002_1	*deforestation*	X	X
21	d2005_1	*deforestation*	X	X
22	hidrografia	*non-forest*		
23	d2004_1	*deforestation*	X	X
24	d2003_2	*deforestation*	X	X
25	d2002_4	*deforestation uncertain if w/in 00-05*	X	
26	d2000_2	*non-forest*		
27	d2003_6	*deforestation uncertain if w/in 00-05*	X	
28	d2003_0	*deforestation*	X	X
29	d2005_2	*deforestation*	X	X
30	nao_floresta2004	*non-forest*		
31	d2000_3	*non-forest*		
32	d2003_5	*deforestation uncertain if w/in 00-05*	X	
33	d2001_4	*non-forest*		
34	residuo2003	*deforestation*	X	X
35	residuo2004	*deforestation*	X	X
36	d2001_0	*deforestation*	X	X
37	d2004_3	*deforestation*	X	X
38	floresta	forest	X	X
39	*backg (no data)*	*no data*		
40	*n2002*	*no data*	X	
41	*n2001*	*no data*	X	
42	*n2000*	*no data*	X	
43	*n2004*	*no data*	X	
44	*n2005*	*no data*	X	
45	*n2003*	*no data*	X	
46	*n1997*	*no data*	X	

Citations and References

Achard, F., Eva, H.D., Stibig, H.J., Mayaux, P., Gallego, J., Richards, T., & Malingreau, J.P. (2002). Determination of deforestation rates of the world's humid tropical forests. **Science**, v. 297, p. 999-1002.

FAO. (1996). Forest Resources Assessment 1990: survey of tropical forest cover and study of change processes. **Forestry Paper No. 130**, Chapter 2.1.1 “Sampling design”. http://www.fao.org/docrep/007/w0015e/w0015e00.htm (accessed Dec 10, 2008).

FAO. (2001). Global Forest Resources Assessment 2000: main report. **Forestry Paper No. 140**. ftp://ftp.fao.org/docrep/fao/003/Y1997E/FRA%202000%20Main%20report.pdf (accessed Dec 10, 2008).

Foley, J.A., Asner, G.P., Costa, M.H., Coe, M.T., DeFries, R., Gibbs, H.K., Howard, E.A., Olson, S., Patz, J., Ramankutty, N., & Snyder, P. (2007). Amazonia revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin. **Frontiers in Ecology and the Environment**, v. 5, n. 1, p. 25-32.

Hansen, M.C., Stehman, S.V., Potapov, P.V., Loveland, T.R., Townshend, J.R.G., DeFries, R.S., Pittman, K.W., Arunarwati, B., Stolle, F., Steininger, M.K., Carroll, M., & DiMiceli, C. (2008a). Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. **Proceedings of the National Academy of Sciences**, v. 105, n. 27, p. 9439-9444.

Hansen, M., Roy, D., Lindquist, E., Adusei, B., Justice, C., & Altstatt, A. (2008b). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. **Remote Sensing of Environment**, v. 112, n. 5, p. 2495-2513.

Hansen, M., Shimabukuro, Y.E., Potapov, P., Pittman, K., (2008c). Comparing annual MODIS and PRODES forest cover change data for advancing monitoring of Brazilian forest cover. **Remote Sensing of Environment**, v. 112, n. 10, p 3784-3793.

INPE. (2008). Monitoramento da Floresta Amazônica Brasileira por Satélite. **Projeto PRODES**, http://www.obt.inpe.br/prodes/ (accessed Dec 10, 2008).

Morton, D.C., DeFries, R.S., Shimabukuro, Y.E., Anderson, L.O., Espirito-Santo, F.D.B., Hansen, M., & Carroll, M. (2005). Rapid assessment of annual deforestaion in the Brazilian Amazon using MODIS data. **Earth Interactions**, v. 9, paper 8. [online journal].

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., & Kassem, K.R. (2001). Terrestrial ecoregions of the worlds: A new map of life on Earth. **Bioscience**, v. 51, n. 11, p. 933-938.

Ridder, R.M. (2007). Global Forest Resource Assessment 2010. Options and Recommendations for a Global Remote Sensing Survey of Forests. **Forest Resources Assessment Programme. Working Paper 141**. Rome.

Shimabukuro, Y.E., Batista, G.T., Mello, E.M.K., Moreira, J.C., & Duarte, V. (1998). Using shade fraction image segmentation to evaluate deforestation in Landsat Thematic Mapper images of the Amazon Region. **International Journal of Remote Sensing**, v. 19, n. 3, p. 535-541.

Stehman, S.V., Sohl, T.L., & Loveland, T.R. (2003). Statistical sampling to characterize recent United States land-cover change. **Remote Sensing of Environment**, v. 86, n. 4, p. 517-529.

Stehman, S.V., Sohl, T.L., & Loveland, T.R. (2005). An evaluation of sampling strategies to improve precision of estimates of gross change in land use and land cover. **International Journal of Remote Sensing**, v. 26, n. 22, p. 4941-4957.