

## Supporting Text

### CLAS Selective Logging Analysis

The CLAS uses high spatial resolution satellite data for large-scale studies of forest disturbance. CLAS is a processing approach that involves: (i) automated atmospheric correction of satellite data; (ii) automated decomposition of each satellite image pixel into subpixel fractional cover measurements of live forest canopy, forest debris, and bare substrates; (iii) automated cloud, cloud-shadow, water, and deforestation masking; (iv) automated pattern-recognition algorithms for forest disturbance mapping; and (v) directed manual auditing of initial results to increase final product accuracy. The full description of CLAS was provided by Asner *et al.* (1), along with a detailed uncertainty analysis and validation effort. An abbreviated overview of the CLAS methodology is summarized below, which follows Fig. 4.

The current version of CLAS uses raw Landsat Enhanced Thematic Mapper Plus (ETM+) satellite imagery and applies sensor gains and offsets to convert from image digital numbers (DN) to top-of-atmosphere radiances. The radiance data are sent to an automated version of the 6S atmospheric radiative transfer model to estimate apparent surface reflectance for each image pixel (2). The 6S program is integrated into the CLAS processing stream and uses monthly averages of aerosol optical thickness (AOT) and water vapor (WV) values from the Moderate Resolution Imaging Spectrometer (MODIS) sensor onboard the National Aeronautics and Space Administration Terra spacecraft.

CLAS uses spectral mixture analysis to decompose each satellite image pixel (30 × 30 m) into fractional cover estimates of photosynthetic vegetation (PV), nonphotosynthetic vegetation (NPV), and bare substrate (0-100% cover). It is fully automated and uses the Monte Carlo Unmixing (AutoMCU) approach to derive uncertainty estimates of the subpixel cover fraction values (3-5). The method uses three spectral endmember “bundles” to unmix each image pixel using the following equation:

$$\rho(\lambda)_{\text{pixel}} = \Sigma [C_e \cdot \rho(\lambda)_e] + \varepsilon = [C_{\text{pv}} \cdot \rho(\lambda)_{\text{pv}} + C_{\text{npv}} \cdot \rho(\lambda)_{\text{npv}} + C_{\text{substrate}} \cdot \rho(\lambda)_{\text{substrate}}] + \varepsilon, \quad [1]$$

where  $\rho(\lambda)_e$  is the reflectance of each land-cover endmember (e) at wavelength  $\lambda$ , and  $\varepsilon$  is an error term. Spectral reflectance bundles [ $\rho_{\text{pv}}(\lambda)$ ,  $\rho_{\text{npv}}(\lambda)$ , and  $\rho_{\text{substrate}}(\lambda)$ ] were developed from field studies and space-based spectroscopic measurements taken by the Hyperion sensor onboard the Earth Observing-1 satellite (1).

CLAS includes a series of automated masks to exclude clouds, water bodies, cloud shadows, nonimage and nonforest areas (e.g., pasture, urban, and agriculture) from the processing stream (Fig. 4). The thermal channel of the ETM+ sensor (band 6) is used as an initial mask for clouds. Water bodies are masked by locating pixels in the calibrated Landsat reflectance data in which bands 1-4 (blue, green, red, and near-infrared) have a negative slope. Nonimage areas containing zero values are also masked. Cloud shadows are identified by using the rms error image that results from the AutoMCU processing (4, 6). Other AutoMCU outputs serve to mask nonforested areas as well as residual clouds and cloud shadows, and additional adjustments are then made to the brightness of “intact” forest areas in 55-km<sup>2</sup> subsets of the imagery, also described in detail by Asner *et al.* (1).

The above-described steps result in maps of forest areas adjusted for atmospheric effects and with nonforest areas, including clouds, masked out. These forest maps are input into an

image differencing procedure where pairs of AutoMCU subpixel fractional cover images, separated by  $\approx 1$  year, are used to create “change images” of PV (forest canopy) and NPV (surface woody and senescent vegetation material) that indicate areas of relative canopy disturbance or recovery. Forest disturbances in these images always have reductions in PV with concomitant increases in NPV fractional cover. A pattern-recognition analysis is then applied to narrow the total area of disturbed forest areas to “probable logging events.”

Logging activity results in: (i) low-intensity forest disturbances from tree-felling gaps; (ii) moderate-intensity linear features from skid trails along which felled trees are dragged by tractors or skidders; and (iii) high-intensity points of damage called log decks, where logs are loaded onto trucks for transportation (7-10). The log decks are connected by logging roads, seen as linear features causing large reductions in the fractional cover of PV, to local roads or rivers for transportation to sawmills or markets. CLAS identifies points (e.g., treefall gaps and log decks) and linear features (e.g., skid trails and logging roads) of recent disturbance occurring in forested areas. Log decks are automatically detected by searching for pixels where PV decreases significantly in a 30-m pixel centered on a  $7 \times 7$  pixel kernel [4.41 hectares (ha)]. A positive detection is flagged when pixels with large PV reduction are surrounded by three concentric rings of incrementally greater PV cover surrounding the target pixel.

The strategy for detecting decks works well in areas logged at higher intensities, as the decks tend to be abundant. However, in areas where the logging is more haphazard, or where the roads themselves also function as loading zones, individual log decks are not always distinguishable. Skid trails are a typological feature of selective logging practices, and they are the single-most ubiquitous surface feature found in harvested areas (11, 12). The presence of skid trails is quantifiable based on large decreases in PV fractional cover in linear or near-linear patterns (1). A moving  $6 \times 6$  pixel (3.24-ha) kernel is applied to each PV-change image to enhance linear features in the N-S, E-W, NE-SW, and NW-SE directions (Fig. 4). The number of directions in which the linear features are arranged, the spatial density of these linear features, and the presence or absence of logging decks are calculated for each location.

After the linear feature and log deck pattern-recognition steps are completed, CLAS automatically integrates the various results to identify contiguous pixel clusters of probable logging events. Logged areas are refined using a moving kernel approach. A base kernel of  $7 \times 7$  pixels (4.41 ha) and four  $3 \times 3$ -pixel (0.81-ha) subset kernels, one located at each corner of the base kernel, are used. The base kernel begins at each logging node and tests for a specific set of pattern-related criteria, as listed in Asner *et al.* (1). If the area in question tests positive, the analysis kernel is moved to its  $7 \times 7$ -pixel neighbors to the north, south, east, and west, which are then each tested against the same criteria (Fig. 4). The input layers and specific criteria tested within the base and subset kernels are described by Asner *et al.* (1).

Initial maps of probable logging events are visually checked and audited. False positives and negatives are manually removed and added by using audit criteria divided into high- and low-damage, obvious, and nonobvious categories. These categories were identified after extensive review of logging events identified in the field. A total of 24,378 polygons were included in the final logging maps (Table 3).

The final CLAS output is a map of logging polygons, within which canopy damage is quantified in each pixel by using the criterion that the PV must have decreased (Fig. 5). This results in logging polygons that have relative amounts of canopy damage quantified on a pixel-by-pixel basis. For the analysis of logging-to-deforestation dynamics, we simply used the polygons and ignored the within-polygon information on canopy damage (Fig. 5). This is

justified here, because we are interested in quantifying the landscape-level overlap between logging and deforestation, and because the deforestation data are provided in polygon format by the Brazilian government (below). For the forest gap fraction analysis, we use gap statistics taken from within each polygon, as described later.

### **Program for Monitoring Deforestation in the Brazilian Amazon (PRODES) Deforestation Maps**

The PRODES of the Brazilian National Institute for Space Research (INPE) provides maps of deforestation obtained from the same Landsat system used in CLAS. The PRODES deforestation maps are provided in polygon format at a scale of 1:250,000. Deforestation data from the PRODES processing of Landsat ETM+ images are subject to a 4% error, as published on the official web site of the program (13, 14). The maps are freely available on the World Wide Web at [www.obt.inpe.br/prodes](http://www.obt.inpe.br/prodes). These constitute the best deforestation maps available for the Brazilian Amazon and are widely regarded as the most accurate source of high-resolution deforestation information in the country (15). We compiled the PRODES digital deforestation maps for the years 2000-2004.

### **Logging-Deforestation Analysis**

Interactions between the CLAS logging and PRODES deforestation products were analyzed in a geographic information system (GIS). The total area included in the analysis was  $\approx 2,030,637 \text{ km}^2$  and spanned the Brazilian states of Pará, Mato Grosso, Rondônia, and Acre in the years 2000-2004. Logged areas were included in our analyses if they were not obscured by clouds in any study year (2000-2004). Application of this criterion, which resulted in the removal of 11-16% of the original logged area (Table 4), ensured that our results would not be biased by discontinuities in the time series. The PRODES deforestation layers were the cumulative overlay of any area that had been deforested at any time during the study years. In the final step, all logging and deforestation spatial data, along with state municipality borders, were intersected in the GIS and exported as a data spreadsheet for tabulation of final logging-deforestation statistics at the scale of state municipality.

We quantified our error in three areas: (i) atmospheric correction and auditor uncertainty associated with the CLAS processing of the logging data; (ii) error associated with the PRODES processing of the deforestation data; and (iii) logging-deforestation overlap registration error. Asner *et al.* (1) showed that the error caused by uncertainty in atmospheric correction of the Landsat imagery (aerosol and water vapor) was only  $\pm 0.7\%$  (Table 7). In addition, the error caused by auditors was found to be  $\pm 12.8\%$ . The resulting rms error for the atmospheric correction and auditor uncertainty was  $\pm 12.8\%$ .

Image registration error occurred when spatial data obtained from different satellite images were mosaicked and/or juxtaposed by using slightly different reprojection transformations in the GIS. Initial overlap analyses using 48 randomly located sample locations spread throughout the study area showed that spatial data from different path/rows in each annual mosaic of Landsat images, among annual mosaics for the same path/row, and between CLAS logging and PRODES deforestation layers had a mean registration error of 96 m, with a standard deviation of 55 m. To simulate the impact of registration error on our logging-deforestation results, logging images for one logging-year/deforestation-year combination were shifted a

distance approximately equal to the mean misregistration plus one standard deviation ( $\approx 150$  m) in eight cardinal directions, N, NE, E, SE, S, SW, W and NW, and the differences in logging/deforestation overlap were tabulated. This analysis resulted in a rms annual registration error (weighted by state logged area) of  $\pm 0.7\%$  (Tables 7 and 8). This value was then used to calculate the rms of annual registration error for the overlap between logging and each subsequent cumulative deforestation data, 1, 2, 3, and 4 years later. These were estimated at  $\pm 0.7\%$ ,  $\pm 1.0\%$ ,  $\pm 1.2\%$ , and  $\pm 1.4\%$ , respectively. These sources of error were combined with all other errors described above to calculate an overall error in the logging-deforestation overlap, which averaged 13.45% (Table 7).

### **Conversion of Photosynthetic Vegetation to Canopy Gap Fraction**

Forest canopy gap fraction, or the fractional canopy cover in the upward-pointing hemisphere from any given ground location, has long served as a central measure of canopy structure in forest ecosystems. Canopy gap fraction largely determines photosynthetic rates, canopy energy and water balance, primary production, mammal and insect dynamics, and even the probability of fire (16-21). It is therefore highly advantageous to convert a remotely sensed radiometric measurement of fractional PV cover to the traditional canopy gap fraction measure. The two measures are not the same. First, PV fraction is a planar metric, whereas canopy gap fraction is hemispherical in nature (4, 12). Moreover, strong adjacency effects between satellite pixels (caused by interpixel light scattering) result in a nonlinear component to vegetation mixture modeling with multispectral data. This effect is maximum at forest gap values greater than  $\approx 85\%$ .

Asner *et al.* (4, 6) developed a set of equations relating PV cover fraction derived from the AutoMCU subroutine of CLAS to field-based measurements of forest canopy gap fraction. We have continued to improve the PV-gap fraction relationship, this time via a more extensive comparison of PV fractions from CLAS to field-based forest canopy gap fractions collected across a wide range of low-, medium- and high-impact logging sites (4, 11, 12). We have found that no single equation can easily represent the nonlinear shape of the PV-gap relationship [e.g., as discussed in Asner *et al.* (6)], and thus we moved to a look-up table approach based on the relationship presented in Fig. 6. In a densely forested region, remotely sensed PV values  $< 67.5\%$  equate to total canopy opening or 100% gap fraction. Again, this is caused by strong interpixel adjacency (light-scattering) effects, that is, satellite observations of a canopy opening, such as a logging deck, contain minimal but consistent PV signatures contributed by neighboring closed-canopy pixels (4, 10). When embedded spatially within a forest mosaic of high PV values, a single pixel containing little to no green vegetation has a PV value  $\gg 0\%$  caused by interpixel scattering of light (especially in the near-IR). This interpixel scattering effect diminishes in importance as the local mosaic of pixel values becomes less diverse ( $68\% < PV < 90\%$ ). Canopy gap fraction values then decrease linearly as PV values increase from 70% to 90% (4). At PV values  $> 90\%$ , the relation again changes slope across sites of differing logging intensity (4), as shown in Fig. 6. Our experience indicates that the relationship does not saturate completely even at PV values  $> 90\%$  (6). Based on these effects, we used the look-up table approach to convert PV estimates to forest canopy gap fraction for each pixel throughout the study region.

### **Canopy Damage and Closure Following Logging**

A  $1 \times 1$ -km window was used to calculate the AIGF for all logged areas by using the formula:

$$AIGF = \frac{\sum f_i}{n}, \quad [2]$$

where  $f_i$  is each logged pixel's gap fraction value within each  $1\text{-km}^2$  area, and  $n$  is the total number of logged pixels within that area. Because forest gap was greater in areas having been disturbed by logging (i.e., felling gaps, skid trails, and logging decks), the AIGF provides a consistent estimate of canopy damage within logged forest, where a higher AIGF value indicates a more open canopy and thus more disturbance throughout the forest as a whole. Using this information, all logged areas within each study state and for each study year were divided into 10 harvest AIGF classes: class 1 was 0-10% AIGF, class 2 was 11-20% AIGF, class 3 was 21-30% AIGF, and so on (Fig. 1 and Table 5).

To understand the time dependence of forest canopy closure, logged areas subjected to post logging deforestation (clear-cutting) were removed from our analysis by excluding pixels with a gap fraction value of 100%. Postlogging canopy closure was then assessed at the Amazon scale by using three methods: First, comparisons of the gap fraction distributions within each AIGF class were made 1 year before, immediately after, and 1-2 years after logging (Fig. 7). This approach allowed for a quick visual understanding of the distribution of forest gap damage within the 10 AIGF classes. Second, postlogging forest canopy closure was assessed by using the means and standard deviations of logged pixels whose individual gap fraction values immediately after logging fell within the thresholds defined above for the AIGF classes. The means and standard deviations used in this analysis were calculated from >180,000 pixels per AIGF class. The mean and standard deviation gap fraction values 1 year before, immediately after, and 1-2 years after logging were then calculated, and the results are reported in Fig. 8. Third, we calculated the area-integrated gap fraction over the entirety of each AIGF class at the scale of the Amazon for 1 year before, immediately after, and 1-2 years after logging, as presented in Table 6. This approach, although providing similar temporal gap recovery trends, differ importantly from those in method (2), because the integration process merges the values of all logged pixels within each AIGF class to calculate a single area-integrated value, and thus it provides no information on the variation of the individual pixels within the forest at that time. For understanding the recovery of the logged forest within each gap class as a whole unit, the AIGF approach is the most informative.

These assessments were conducted only for year-2000 logging, because 2-year postlogging data were not available for the 2001 and 2002 logging maps. AIGF variation in intact forest, measured to compare with that of logging and canopy closure, was assessed by following the AIGF values of up to 8.76 million intact forest pixels for 3 years. Intact-forest AIGF variation increased with larger gap fraction classes (Fig. 7), but in all cases, was far less than that from logging. This same pattern is visible at the scale of individual logged pixels (Fig. 8). Multiple Kolmogorov-Smirnov analyses (480 pairs) were used to identify significant differences ( $P > 0.05$ ) between the gap fraction distributions of each AIGF class calculated by the method (1) described above, before, immediately after, and 1 and 2 years after logging. Nearly all gap fraction distributions were significantly different from one another. This was found both between different gap classes and between the same gap classes of the four study states. The only exceptions were for gap class 9 (80-89%), where a significant difference was not

found between logged and 1-year post logging gap fraction distributions in Acre, and for the Para/Rondonia and Acre/Rondonia comparisons in which AIGF class 10 immediately after logging and 2 years after logging.

Forest areas that underwent higher-intensity logging and were thus in higher AIGF classes exhibited steeper slopes of canopy recovery within the first and second years after logging (Table 6). As expected, considering the forest's AIGF recovery, individual pixels with higher logging-related canopy damage recovered at faster rates (Fig. 8).

Because overall logging intensity, by definition, occurs at a spatial scale greater than an individual pixel (e.g., one tree fall gap, logging deck, or section of skid trail), we used AIGF classes to estimate harvest intensities throughout the Amazon study area. We considered that the primary source of error for calculation of AIGF class proportions was the scale over which AIGF was calculated. To generate error estimates based on different scales, we recalculated gap class proportions over the state of Mato Grosso in year 2000 ( $n > 15$  million logged pixels) by using eight unique square kernels with edge lengths ranging from 250 to 2,000 m in increments of 250 m. The standard deviation of the eight calculations per gap class was then used to generate uncertainty estimates for each gap class. The standard deviations for the proportions of gap classes 1-10 were 3.65, 2.26, 0.54, 0.56, 0.76, 0.39, 0.14, 0.27, 0.24, and 0.1%, respectively. For example, the application of these uncertainty estimates to Amazon year 2000 estimates would show an area of  $4,264 \pm 722$ ,  $1,470 \pm 150$ , and  $16 \pm 19$  km<sup>2</sup> for gap classes 1, 5, and 10, respectively.

### **Is Logging a Precursor to Deforestation?**

To compare the probability of deforestation after logging to the probability of deforestation from intact forest, we developed a procedure that randomly selected a large number of points (up to 1,000, depending on the total number of available logged or intact pixels) within logged and deforested areas. This was performed separately for each of 10 5-km zones (0-4.9 km, 5-9.9 km, etc. ...) extending outward to 50 km from all major roads. The GIS coverage of major roads was obtained from the Instituto do Homem e Meio Ambiente da Amazônia (Fig. 9). The entire procedure was run at the state scale by using year-2000 logging and intact forest maps and compared with year-2004 deforestation maps from PRODES (13). The pixel size of all input data was  $100 \times 100$  m. Areas having incomplete temporal PRODES or CLAS logging coverage (caused by cloud interference or Landsat image problems) between the 2000 and 2004 study years were not included in the analysis, nor were any areas within state or federally protected areas. The final area, at the Amazon scale, included in this analysis was 757,514 km<sup>2</sup>. The decision to have a maximum distance limit of 50 km from the nearest major road was based both on previous analyses that have shown that the majority of deforestation occurs within this 50-km buffer zone (22, 23), and the results of our analyses that showed that  $\approx 95\%$  of year-2000 logging in the Amazon study area occurred within 15 km of the nearest road (Fig. 10).

A  $\chi^2$  analysis was used to statistically compare the probability of year-2000 logging and intact forest having been deforested by the year 2004. First, we scaled the state-level results to the entire study region by weighting the state-level deforestation proportion results by the area logged per distance class from each road in each state. The  $\chi^2$  analysis was then run separately for each distance-to-road class. The results of this analysis, presented in Fig. 3, show that logged areas are significant precursors to deforestation, and that there is a strong distance interaction,

with closer distances to major roads having increased probabilities of deforestation for both logged and intact forest areas.

## References

1. Asner, G. P., Knapp, D. E., Broadbent, E. N., Oliveira, P. J. C., Keller, M. & Silva, J. N. M. (2005) *Science* **310**, 480-482.
2. Vermote, E. F., Tanre, D., Deuze, J. L., Herman, M. & Morcrette, J. J. (1997) *IEEE Trans. Geosci. Rem. Sens.* **35**, 675-686.
3. Asner, G. P. & Lobell, D. B. (2000) *Rem. Sens. Environ.* **74**, 99-112.
4. Asner, G. P., Keller, M., Pereira, R., Zweede, J. C. & Silva, J. N. M. (2004) *Ecol. Appl.* **14**, S280-S298.
5. Asner, G. P. & Heidebrecht, K. B. (2002) *Intl. J. Rem. Sens.* **23**, 3939-3958.
6. Asner, G. P., Knapp, D. E., Cooper, A. N., Bustamante, M. M. C. & Olander, L. P. (2005) *Earth Interact.* **9** (7), 1-31.
7. Stone, T. A. & Lefebvre, P. (1998) *Intl. J. Rem. Sens.* **19**, 2517-2524.
8. Souza, C., Roberts, D. & Monteiro, A. L. (2005) *Earth Interact.* **9** (19), 1-25.
9. Souza, C. & Barreto, P. (2000) *Intl. J. Rem. Sens.* **21**, 173-179.
10. Asner, G. P., Keller, M., Pereira, R. & Zweede, J. (2002) *Rem. Sens. Environ.* **80**, 483-496.
11. Pereira, R., Zweede, J., Asner, G. P. & Keller, M. (2002) *For. Ecol. Manage.* **168**, 77-89.
12. Asner, G. P., Keller, M. & Silva, J. N. M. (2004) *Global Change Biol.* **10**, 765-783.
13. Instituto Nacional de Pesquisas Espaciais (2005) *PRODES: Assessment of Deforestation in Brazilian Amazonia* (Brazilian National Institute for Space Research, São Jose dos Campos, Brazil).
14. Câmara, G., Valeriano, D. M. & Soares, J. V. (2006) *Metodologia para o Cálculo da Taxa Anual de Desmatamento na Amazônia Legal* (Brazilian National Institute for Space Research, São Jose dos Campos, Brazil).
15. DeFries, R. S., Asner, G. P., Achard, F., Justice, C., Laporte, N., Price, K., Small, C. & Townshend, J. (2005) in *Tropical Deforestation and Climate Change*, eds. Moutinho, P. & Schwartzman, S. (Amazon Institute for Environmental Research, Belem, Brazil), p. 131.
16. Bray, J. R. (1956) *Ecology* **37**, 598-600.
17. Brokaw, N. V. (1985) *Ecology* **66**, 682-687.
18. Murray, K. G. (1988) *Ecol. Monog.* **58**, 271-298.
19. Shelly, T. E. (1988) *Biotropica* **20**, 114-119.
20. Mulkey, S. S. & Percy, R. W. (1992) *Funct. Ecol.* **6**, 719-729.
21. Holdsworth, A. R. & Uhl, C. (1997) *Ecol. Appl.* **7**, 713-725.
22. Alves, D. S. (2002) *Intl. J. Rem. Sens.* **23**, 2903-2908.
23. Nepstad, D., Carvalho, G., Barros, A. C., Alencar, A., Capobianco, J. P., Bishop, J., Moutinho, P., Lefebvre, P., Silva Jr., U. L. & Prins, E. (2001) *For. Ecol. Manage.* **154**, 395-407.