A New Sampling Scheme for Tropical Forest Monitoring Using Satellite Imagry

Frédéric Achard, Tim Richards, and Javier Gallego

Abstract.—At the global level, a sampling scheme for tropical forest change assessment, using high resolution satellite images, has been defined using sampling units independent of any particular satellite sensor. For this purpose, a sampling frame has been chosen a hexagonal tessellation of 3,600 km^2.

The stratification has been performed to reduce the variance of the estimated variable. Areas of current (1992-1997) deforestation in the moist zone of the tropical belt have been identified by international consultation. The resulting "hot spot" maps and a forest map derived from NOAA AVHRR 1-km resolution data are used to stratify the sampling population and to determine sampling rates.

Appropriate tropical forest policies must be supported by better informed perspectives on the role those resources play in the global environment. Of major concern is the issue of forest distribution and its rate of change. New information technologies make possible the development of more advanced systems of data gathering and analysis, which can accurately inform us of ongoing changes in the forests of the tropics.

The first phase of the TREES Project was dedicated to the development of global forest cover assessment techniques at the level of the tropical belt. This baseline inventory of tropical forests concerns in order of priority the rain forest biome and adjacent seasonal formations (Malingreau *et al.* 1995).

The concept of TREES was to make wall-to-wall coverage using highly repetitive observations at low resolution. A multi-annual set of AVHRR data was acquired over the whole tropical belt during the period 1991-1994. These data were analyzed to derive a first global tropical forest base map at 1-km resolution. Other investigators have already chosen to map tropical forest area mainly from coarse spatial resolution AVHRR data, most often by using some finer resolution satellite data as a reference data source (e.g., Stone *et al.* 1994). The main advantages of the TREES approach are the daily availability of data (near daily coverage) and the spatial resolution (1.1 km in the case of AVHRR) that is more commensurate to the geographical scale of the monitoring exercise.

Achard and Richards, Global Vegetation Monitoring Unit, Gallego, Agriculture and Regional Information Systems Unit, Space Application Institute of the Joint Research Centre of the European Commission, 21020 Ispra, Italy.

However, coarsening the spatial resolution leads to a loss of spatial detail at a rate that depends on the spatial structure of the landscape (Woodcock and Strahler 1987). Consequently the use of coarse spatial resolution maps will effect specific methods, such as the estimation of cover type areas, the validation of results, and the assessment of the product accuracy.

The TREES global assessment has been calibrated using high resolution image maps at selected sites (Mayaux and Lambin 1995). The calibration method developed during the first phase of the project called for the use of Landsat TM data and the design of a specific statistical technique for the production of forest cover figures (Mayaux *et al.* 1998). Ecologically important parameters such as forest cover percentage and fragmentation index have been used as the main variables for the stratification and correction (regression) steps.

OBJECTIVES FOR THE DEVELOPMENT OF A MONITORING SYSTEM

The activities of the first phase were conducted with a view to developing a "prototype" of an operational tropical forest monitoring system: TREES II phase. The operational tropical forest monitoring system should on a regular basis :

- produce relevant and accurate information on the state of the tropical forest resources,
- analyze such information in terms of deforestation trends and possible impacts.

Multi-scale Strategy: Low Resolution Monitoring and High Resolution Sampling

The TREES methodology starts with the identification of "hot spot" areas (Myers 1993) where deforestation is most active, using the combination of individuals' expertise and a range of remote sensing and ground reference data. Potential indicators such as spatial fragmentation (Jeanjean and Achard 1997) or population increase (Lambin and Ehrlich 1997) are currently being tested to develop a more automatic (i.e., not based on subjective knowledge) detection procedure of the rapid change areas.

Because of the perceived difficulty in deriving accurate measurements for forest change assessment from low resolution data alone (D'Souza and Malingreau 1994, Eva *et al.* 1995), it was decided to estimate the area percentage and the forest change from interpretations of high resolution imagery. A new sampling scheme was deemed indispensable to work with high resolution satellite imagery in an affordable way.

Knowing that the estimation accuracy depends on the variance of the parameter to be measured, further conditions have been defined for this new scheme (to be respected as much as possible):

- To facilitate the implementation of the system, it was decided to use a sampling frame independent of any particular satellite sensor, so that any sensors can be used if suitable to the target; however, observed units may be linked to a particular sensor, considered as best adapted at the date of the survey.
- A preliminary stratification on the forest cover change parameter should be performed to reduce the variance of the estimated variable.
- Unbiased estimators and error variance may be computed from the sample.

METHODOLOGICAL DEVELOPMENTS

Hot Spot Identification

Definition of an Deforestation "Hotness" Index Per Region

Areas of current (1992-1997) deforestation in the moist zone of the tropical belt have been identified by international consultation (Achard *et al.* 1998). The resulting "hot spot" maps and the 1-km resolution forest map are used to stratify the sampling population and to determine sampling rates during the sampling phase.

Deforestation Hotness Index.—In some regions the hot spots cover very large areas (Brazil and Guyanas), while in other regions they cover relatively small areas where deforestation is presumably more concentrated (West Africa and Insular Southeast Asia). To correct for such discrepancies between the heterogeneous sizes of the hot spot areas and their "effective" deforestation rate, a hotness index was defined using the FAO deforestation figures for the period 1990-1995 (FAO 1997). This hotness index is the ratio between the hot spot area of the region and the related deforested area in the period 1990-1995.

$$\text{Hotness(region)} = \frac{\text{Deforestation 1990} - \text{95(region)}}{\text{Hot spot area(region)}}$$

Table 1.—*Deforestation hotness index per region of the tropical belt*

Region	Deforestation 1990-1995 (1000 ha)	Hot spot area of region (1000 ha)	Hotness Index (%)
Central America	4,794	18,400	26.1
Panamazon and Andes	8,764	47,100	18.6
Brazil and Guyana	12,880	112,900	11.4
Continental South-East Asia	5,911	38,300	15.4
Insular South-East Asia	9,401	33,200	28.3
Central Africa	5,699	31,500	18.1
Madagascar	650	6,000	10.8
West Africa	2,459	3,600	68.3

Sampling Scheme

A sampling scheme was developed to address the measurement of the recent forest area changes using high resolution satellite imagery. The statistical sampling will provide an estimate of the precision of the estimator.

Selection of a Sampling Frame

Previous regional or continental monitoring exercises have used the Landsat or SPOT scenes reference systems as a basis for sample unit selection (FAO 1996, Gallego 1998). Such Earth observation satellite-based systems usually locate the remote sensed scenes in a regular manner on the basis of a system of paths (orbits) and rows (arbitrary latitudinal divisions). Such satellite scenes reference systems have the drawback of having overlapping units with an overlap degree that changes with latitude. They can provide a convenient sampling frame around the Equator, but they become increasingly complex polewards. Because the estimation of tropical forest cover change has to be seen in a global context (attention is now being given to Eurasian and North American forests), a system that would be compatible among continents had to be developed.

For this purpose, a sampling frame is constructed that has near-equal area units. The sampling frame is based on hexagon tesselations on a sphere (Thuburn 1997, White *et al.* 1998). The basic geometry can be visualized by looking at a standard soccer ball. Various polyhedral grid systems have been developed for different purposes (White *et al.* 1992, Olsen *et al.* 1998).

This type of geometry has a number of useful characteristics, particularly that the sample frame is made up of nominally equal area and equal shape units, which share a consistent spatial topology with their neighbors. As a result of this, the units all fit together and there are no overlaps or gaps between them. A scheme with equal probabilities of sample unit selection can therefore be devised.

The average size of the hexagons has been tuned to 3,600 km^2, which is comparable to the size of the smallest satellite images (SPOT images are approximately 60 x 60 km). The sampling frame is a finite set of 12,532 points, which are the centers of the hexagons that intersect with the study area by more than 10 km^2. The sampling probability is proportional to the area of the intersection of the hexagon with the study area

Stratification of the Sampling Frame

Stratification reduce the varia of the estimate variable usedin the procedure (forest fragmentation, ecophysiographic units etc.). The absence of adequate maps of forest cover change is a major limitation in producing a good stratification. The hot spot map produced by the expert consultation (Achard *et al.* 1998) ha been used for the stratification. Five strata have been defined (table 2).

The only limitation of the somewhat "subjective stratification" (the hot spot map is derived from individual expertise) is that the estimate of the variances is not as good as when derived from an optimal stratification (i.e., the sampling error is higher).

Eight regions have been identified as areas of particular interest and are used to tune the sample size in each region. The distribution of the sampling frame by strata and subcontinent is given in table 3. The three main strata are stratum 5 (non-forest), stratum 4 (forest without hot spot), and stratum 1 (hot spot area percentage > 50 percent in forest domain).

Table 2.—*Definition of the sampling strata*

		% hot spot area in sampling unit			
		No hot spot	<20%	20-50%	>50%
Percent Forest cover (AVHRR) in sampling unit	No forest (0%)	**Stratum 5**			
	<10%		**Stratum 3**		
	>10%	**Stratum 4**	**Stratum 3**	**Stratum 2**	**Stratum 1**

Table 3.—*Distribution of the sampling frame by strata*

Region	Strata 1	2	3	4	5	Total
Central America	48	38	50	351	164	651
Panamazon and Andes	143	54	98	963	347	1,605
Brazil and Guyana	363	64	77	998	367	1,869
Continental South-East Asia	129	55	90	414	1,273	1,961
Insular South-East Asia	124	50	53	1,038	324	1,589
Central Africa	103	50	62	737	491	1,443
Madagascar	46	2	35	7	153	243
West Africa	10	2	12	168	563	755
Other Regions	0	0	1	320	2,095	2,416
Total	966	315	478	4,996	5,777	12,532

Systematic Sampling by Points on a Regular Grid

A systematic sample of points is based on a rectangular grid in an equal area projection with several replicates. A hexagon is selected if a point of the sampling grid falls inside the hexagon and inside the study area. This provides us with a way of sampling the hexagons of the tesselation with a probability proportional to the area of the hexagon that falls inside the study area. The steps for sampling are:

- Select the equal area projection and the sampling grid. We have chosen the Lambert Cylindrical equal area projection and a grid of 600 km x 500 km.
- Select a pattern of points in the grid cell, which we shall repeat across the grid. We have selected 12 points that will generate 12 replicates. Figure 1 gives the location of the points that generate the six first replicates, for which the replicate number is given randomly.

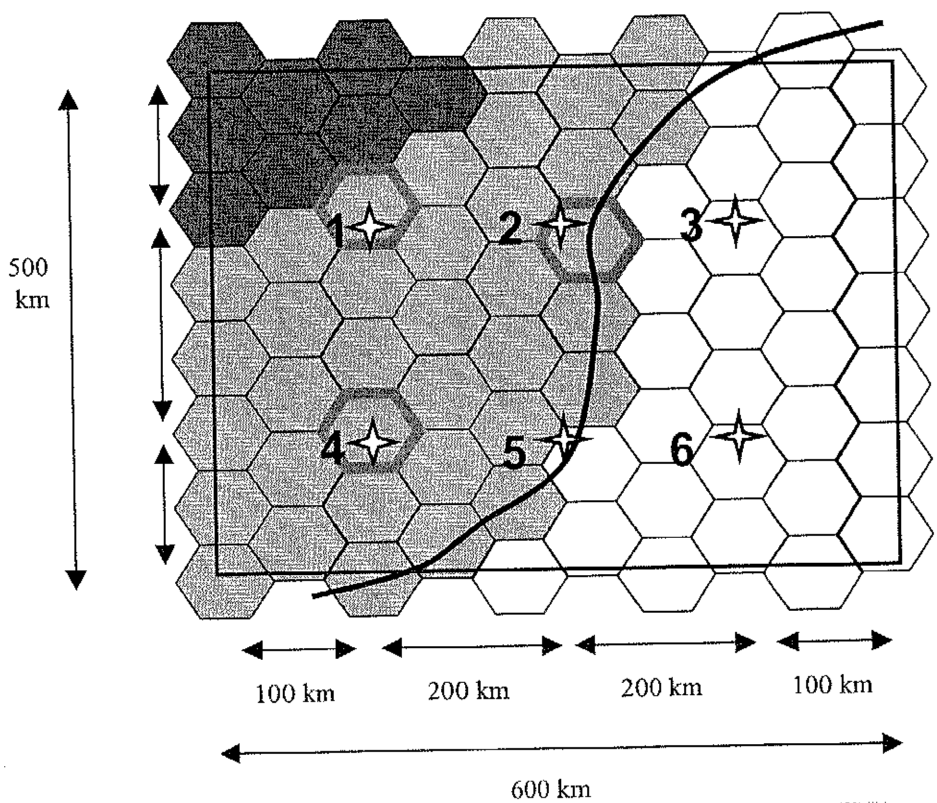

Figure 1.—*Systematic sampling grid with an example of six replicates.*

- Decide the number of replicates that will be kept for each stratum and region. This determines the sampling intensity. The number of replicates per region and stratum have been tuned to be roughly proportional to the hotness index (per region) and with an increasing intensity for hot spot areas (per stratum): see table 4.
- Repeat the pattern of points across the study area for a pre-sample of hexagons. Each point of the pre-sample will be labeled with a number of replicate k.
- If the point k falls in a hexagon that belongs to stratum h and region g, the hexagon is selected if $k \leq m_{hg}$, where m_{hg} is the number of replicates for stratum h and region g. When $m_{hg} = 0.5$ (or $m_{hg} = 0.25$), it means that every other (or every four) rectangular grid is used.
- The total target sample size has been set to 95 (as a feasible target linked to availability of resources). A final linear tune of all replicate numbers was made to accommodate that latest figure.

The resulting number of sampling units per stratum and region is given in table 5.

Location and Size of Observation Units

Analysis performed on forest maps for Eurasia suggest that the optimal (in terms of information/ cost ratio) size of observation units (sites) is not the same everywhere. The analysis was based on the comparison between use of full Landsat –TM scene (180 x 180 km) and quarter Landsat –TM scene (90 x 90 km) with the hypothesis that image acquisition and interpretation costs for a full TM scene are about twice the costs for a quarter TM scene. Smaller sites are suitable for relatively homogeneous areas (high spatial correlation), and larger sites are preferable for more heterogeneous areas (Gallego *et al.* 1998).

Table 4.—*Number of replicates per stratum in each region (sampling intensity)*

	Hotness Index	Number of replicates per strata (h)				
Region (g)		**1**	**2**	**3**	**4**	**5**
Central America	26.1	6	3	2	0.5	0.25
Panamazon and Andes	18.6	6	3	2	0.5	0.25
Brazil and Guyana	11.4	3	2	2	0.5	0.25
Continental South-East Asia	15.4	6	3	2	0.5	0.25
Insular South-East Asia	28.3	6	3	2	0.5	0.25
Central Africa	18.1	6	3	2	0.5	0.25
Madagascar	10.8	3	2	1	0	0
West Africa	68.3	12	6	4	2	0.25
Other Regions		0	0	0	0	0

Table 5.—*Number of sampling units per stratum / region /size*

	Total per Region	Number of sampling units per strata (Full/Quarter)				
Region	**(F/Q)**	**1**	**2**	**3**	**4**	**5**
Central America	1/5	0/3	0/0	0/0	1/1	0/1
Panamazon and Andes	7/15	4/8	1/1	0/2	2/3	0/1
Brazil and Guyana	8/12	8/5	0/1	0/1	0/4	0/1
Continental South-East Asia	6/9	4/5	1/1	1/0	0/3	0/0
Insular South-East Asia	7/6	7/1	0/1	0/0	0/4	0/0
Central Africa	9/5	4/0	1/3	2/0	2/2	0/0
Madagascar	0/2	0/1	0/0	0/0	0/0	0/1
West Africa	1/2	0/0	0/0	1/0	0/2	0/0
Total per strata	39/56	27/23	3/7	4/3	5/19	0/4

The information presently available is insufficient to measure spatial heterogeneity of deforestation patterns. We adopted a site size rule based on two assumptions:

- Deforestation will be more heterogeneous where the forest pattern is more heterogeneous;
- Deforestation is more heterogeneous in hot spots.

We consider first the Landsat scenes that have centers closest to the centers of the sampled hexagons. Each selected full Landsat scene *j* defines a cluster T_j of hexagons, for which the Landsat scene center is closest to their centers. The rule selects a full Landsat scene if one of the two following conditions is valid:

- (hot spot area of T_j) x (hotness index of the region) > 250,000 ha
- (standard deviation of forest percent in hexagons within cluster T_j) x (total area of T_j) > 600,000 ha

Otherwise, a quarter Landsat scene is selected, again with a criterion of minimal distance between the center of the hexagon and the center of the quarter Landsat scene. Thresholds for the site size rule were selected after plotting both composite parameters to get a rough proportion of one full scene for two quarter scenes.

Once the sampling scheme is applied, the full or quarter Landsat scenes associated with the sampled hexagons are selected as observational units. The resulting number of full and quarter scenes is given in table 5.

Note the size of the grid cells (600 km x 500 km) and the maximum number of replicates within cells have been selected to ensure that an observational unit will not be selected twice, i.e. than no more than one hexagon in a given cluster is selected.

DISCUSSION ON THE NEW SAMPLING SCHEME

Implementation

This new sampling scheme has been developed to estimate the forest change from interpretations of high resolution satellite imagery in an affordable way. The scheme has to be tested now against implementation constraints such as data availability (partial or total absence).

The non-forest area will be sampled with a very small rate (four sampling units). Images will be acquired for sites with zero percent forest only on specific sample units when there is a doubt on that figure. Otherwise, they will generate a zero value for estimation.

To tackle the issue of non-availability of satellite imagery (for example, in the case of permanent cloud cover), a list of additional replicates can be produced (by increasing the number of replicates in table 4) to allow replacement of missing scenes. The selection should follow a random ranking list within the corresponding stratum. The original sample will remain in the list as missing data.

In some cases, it will be necessary or preferable to reduce the size of the observation units:

(a) because of missing good quality imagery over large areas (cloud coverage)
(b) because existing knowledge would allow replacement of a full scene by one or two quarter scenes (homogeneity of the parameter)
(c) because of the need for higher resolution data for more accurate change measurement (SPOT imagery).

The effect is that data will be partially missing. In case (b), a prediction is used instead of a full measurement. In all cases, the replacing scene should be centered on the original scene location and with the largest overlap.

A simple method for interpretation of the imagery is being developed. Some complex or non-standard aspects can be accepted for the pre-processing level, as long as they are transparent for the final change matrices.

Estimation

In the estimation phase, the selection probability of an observation unit will be taken into account. Applying the site size rule, for each hexagon there is a linked observation unit (full Landsat scene or quarter Landsat scene) and a relative cluster T_j of hexagons. The probability that a specific observation unit is selected is the sum of probability that the corresponding hexagons are selected. Each observation unit does not belong necessarily to a single sampling stratum, but its sampling probability is known. This is a situation of unequal probability sampling rather than stratified sampling.

ACKNOWLEDGMENTS

The authors would like to thank Anthony R. Olsen, U.S. Environmental Protection Agency, Corvallis, OR, USA, for his review of the paper and his very helpful comments.

LITERATURE CITED

Achard, F.; Eva, H.; Glinni, A.; Mayaux, P.; Richards, T.; Stibig, H.J. 1998. Identification of deforestation hot spot areas in the humid tropics. Luxembourg: European Commission. TREES Series B, N°4, EUR 18079 EN. 100 p.

D'Souza, G.; Malingreau, J.P. 1994. NOAA-AVHRR studies of vegetation characteristics and deforestation mapping in the Amazon Basin. Remote Sensing Reviews. Vol.10: 5-35.

Eva, H.D.; D'Souza, G.; Malingreau, J.P. 1995. Potential of ATSR-1 data for detection of clearings within dense humid tropical forest. International Journal of Remote Sensing. 16: 2071-2079.

FAO. 1996. FRA 1990: Survey of tropical forest cover and study of change processes. Rome: FAO, Forestry Paper 130. 152 p.

FAO. 1997. State of the world's forests - 1997. Rome: FAO. 200 p.

Gallego, F.J. 1998. Double sampling for area estimation and map accuracy assessment. In: Mowrer, H.T.; Congalton, R.G., ed. Quantifying spatial uncertainty in natural resources: theory and applications for GIS and remote sensing. Chelsea, MI: Ann Arbor Press.

Gallego, F.J.; Feunette, I.; Carfagna, E. 1998. Optimising the size of sampling units in an area frame. In: Proceedings of GeoENV 98; Valencia; 1998 November 18-20.

Jeanjean, H.; Achard, F. 1997. A new approach for tropical forest area monitoring using multiple resolution data. International Journal of Remote Sensing. 18: 2455-2461.

Lambin, E.F.; Ehrlich, D. 1997. The identification of tropical deforestation fronts at broad spatial scales. International Journal of Remote Sensing. 18: 3551-3568.

Malingreau, J.P.; Achard, F.; D'Souza, G.; Stibig, H.J.; D'Souza, J.; Estreguil, C.; Eva, H. 1995. AVHRR for global tropical forest monitoring: the lessons of the TREES Project. Remote Sensing Reviews. 12: 29-40.

Mayaux, P.; Lambin, E.; 1995. Estimation of tropical forest area from coarse spatial resolution data: a two-step correction function for proportional errors due to spatial aggregation. Remote Sensing of Environment. 53: 1-16.

Mayaux, P.; Achard, F.; Malingreau, J.P. 1998. Global tropical forest area measurements derived from coarse resolution maps at a global level: a comparison with other approaches. Environmental Conservation. 25 (1): 37-52.

Mayaux, P.; Janodet, E.; Blair-Myers, C.M.; Legeay-Janvier, P. 1997. Vegetation Map of Central Africa, Scale 1/5M. TREES Publications Series D, N°1, EUR 17322. Luxembourg: European Commission.

Myers N. 1993. Tropical forests: the main deforestation fronts. Environmental Conservation. 20: 9-16.

Olsen, A.R.; Stevens, D.L.; White, D. 1998. Application of global grids in environmental sampling. Computing Science and Statistics. in press.

Stone, T.A.; Schlesinger, P.; Houghton, R.A.; Woodwell, G.M. 1994. A map of the vegetation of sSouth America based on satellite imagery. Photographic Engineering and Remote Sensing. 60: 541-551.

Thuburn, J. 1997. A PC-based shallow water model on a hexagonal icosahedral grid. Monthly Weather Review. 125: 2328-2347.

Woodcock, C.E.; Strahler A.H. 1987. The factor of scale in remote sensin. Remote Sensing of Environment. 21: 311-332.

White, D.; Kimerling, J.; Overton, W. 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. Cartography and Geographic Information Systems. 19: 1, 5-22.

White, D.; Kimerling, A.J.; Sahr, K.; Song, L. 1998. Comparing area and shape distortion on polyhedral-based recursive partitions of the sphere. International Journal of Geographical Information Systems. in press.