Forest Area Estimation Using Sample Surveys and Landsat MSS and TM Data

F. Deppe

Abstract

Sample surveys data, and Landsat Multispectral Scanner System (MSS) and Landsat Thematic Mapper (TM) image data were used to establish forest area estimates within a test site area in the state of Rio Grande do Sul, southern Brazil. Two methods were applied: direct expansion and regression estimator methods. For the former, an area frame sampling scheme was used and, for the latter, estimates were corrected by means of classification results. Reference data were compiled for two periods through aerial photography interpretation (API) and direct field observations. The direct expansion method produced rapid and independent area estimates while results from the regression estimator method showed narrower confidence intervals, indicating higher accuracy. Mean relative efficiency figures for forest classes were 3.63 using the MSS data set and 7.73 using the TM data set. Findings of this study demonstrated the potential and advantages of using Landsat image data for forest area estimation.

Introduction

Forest area estimates are essential in resources inventories, especially for forest management and planning activities, e.g., in production planning, making marketing decisions, and determining policies. Also, there is a constant demand for up-to-date and accurate information for area estimates at a local, regional, and national scale. Area estimation established by integration of sample survey (based on area frame sampling) and digital classification of satellite images is one remote sensing application that shows great potential in terms of feasibility (e.g., operationability, time, and costs). Remotely sensed data can be used to provide information at a whole range of spatial, spectral, radiometric, and temporal resolutions, including global coverage, depending on the sensors used. Therefore, analysts and users should be aware of the characteristics and type of data used for specific applications (e.g., agriculture, forestry, and regional inventories), in order to meet different aims in different regions of interest. However, regardless of the application, there are some general difficulties related to remote sensing data that need to be taken into account (Bierlaire *et al.*, 1993). First, the spectral and spatial resolution of the sensors can impose some limitations; second, problems are introduced by the effects of the atmosphere; and third, there are problems related to the ambiguities in data interpretation.

Area crop estimation using remote sensing data was carried out on an operational basis by the National Agricultural Statistics Service of the U.S. Department of Agriculture between 1980 and 1987 with very satisfactory results (Allen and Hanuschak, 1988). Other studies were carried out in Canada (Ryerson *et al.*, 1985), Australia (Daubin and Beach, 1981), and Lybia (Latham *et al.*, 1983). In Europe, one of the uses of remote sensing for area estimation can be illustrated by the MARS (Monitoring Agriculture with Remote Sensing) project (Meyer-Roux, 1992), undertaken by the Satellite Applications Institute (Joint Research Centre, Ispra). The long-term project was established to provide information for the Directorate General VI (Agriculture) and the Statistical Office (EUROSTAT) of the European Union, in terms of up-to-date and accurate data for estimating the vegetation condition for the member states, in order to detect and analyze problems and to take appropriate measures (Bierlaire *et al.*, 1993; Meyer-Roux, 1992; Annoni and Gallego, 1992; Sharman, 1992).

In this study, two separate area estimates were produced using different methods. The first is referred to as a direct expansion method where sample surveys, i.e., ground observations, of land-cover classes are taken based on an area frame sampling scheme. The observations at the sample units can be taken with high accuracy; however, the direct expansion estimates will include high sampling errors. The second is referred to as a regression estimator method where estimates are produced based on (1) digital classification of the satellite imagery and (2) direct expansion estimates. With digital classification, sampling errors are avoided because the whole study area is classified on the satellite image. Care should be taken because, as a result of digital image classification, pixels can be incorrectly classified. This applies to the situation where reflectance properties of the land-cover classes are not sufficiently different from each other. Area estimates produced from pixel counts will contain these errors. However, the regression estimator method, which takes the two sources of information (i.e., from the direct expansion method and digital classification), uses a double sample in order to produce a regression estimate. This procedure has proved to be more accurate, and the regression estimator method is used to increase the accuracy of the sample survey. The degree to which it will be increased depends upon the correlation between the sample survey results and the digital classification results in the sample units.

Centre for Research in Remote Sensing and Meteorology, Federal University of Rio Grande do Sul, Campus do Vale, Av. Bento Gonçalves, 9500, Prédio 44202, P.O. Box 15044, CEP 91501 - 970 - Porto Alegre, RS, Brazil (deppe@if1.if.ufrgs.br).

Photogrammetric Engineering & Remote Sensing, Vol. 64, No. 4, April 1998, pp. 285–292.

0099-1112/98/6304–285$3.00/0

Objectives

The general objective of this study was to investigate the application of remote sensing to forest resources management and planning for the purpose of establishing area estimation and area change measurements of forest (natural forest and forest plantations) and other land-cover classes (pasture, bare soil, river/reservoirs, and rice fields) in the state of Rio Grande do Sul (RS), southern Brazil. The specific objectives were

- to establish a sample survey procedure based on an area frame sampling scheme in order to produce independent area estimates and sample survey reference information,
- to establish area estimates of forest cover and other land-cover classes using Landsat MSS and TM image data,
- to investigate the potential for improving area estimates accuracy through the use of classified satellite data using the regression estimator method,
- to determine area changes in forest and other land-cover classes between 1975 and 1989 based on area estimates generated through the regression estimator method, and
- to investigate the feasibility of applying the regression estimator method in future projects.

Data Sources and Test Site Area

Data sources consisted of (1) Landsat 1 Multispectral Scanner (MSS 1 - Orbit/Path 237/81) acquired on 16 June 1975; (2) Landsat 5 Thematic Mapper (TM 5 - Orbit/Path 221/81), Quadrant A, acquired on 28 August 1989; (3) aerial photographs, 23- by 23-cm format, black and white, 1:110,000 scale; and (4) 1:50,000-scale topographic maps.

The test site area selected covers an area of 42 km by 42 km (1,764 sq km) and is located in the state of Rio Grande do Sul (RS), southern Brazil. The area is characterized by a development of new settlements in both suburban and rural areas with an increasing number of new commercial and industrial establishments. Agricultural enterprises in the area include activities such as cattle and sheep grazing, cropping, and forestry (mainly for paper pulp production). Other activities included coal mining industry. Figure 1 illustrates the test site location. The UTM (Zone 22) northings and eastings are Top Left 6699,000 - 406,000; Bottom Left 6657,000 - 406,000; Top Right 6699,000 - 448,000; and Bottom Right 6657,000 - 448,000.

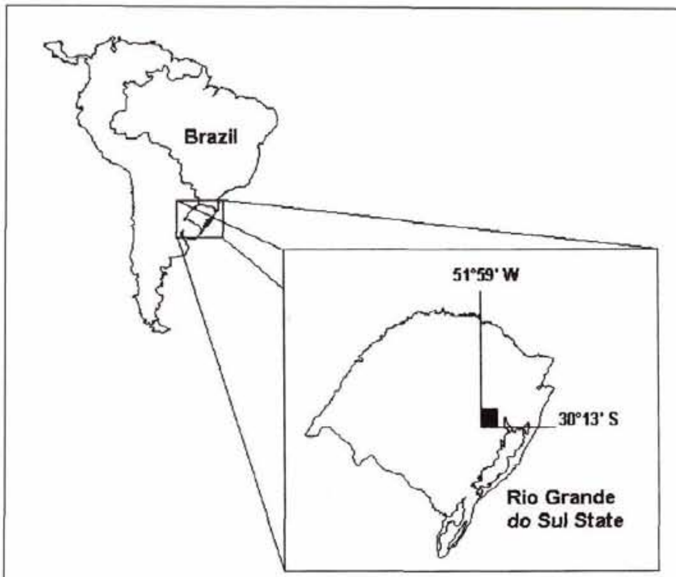

Figure 1. Location of the test site area.

Area Estimates through the Direct Expansion Method

The application of the direct expansion method has been investigated in several studies, for example, Alonso *et al.* (1991), Gallego and Delince (1991a), Gallego and Delince (1991b), and Gangkofner *et al.* (1990). The primary objective of this method was to obtain independent area estimates and related statistical data (e.g., variance estimate) in order to evaluate the findings and to use them in further work, such as the regression estimator. The decision to stratify or not will depend on the objective of the survey and on the land-cover feature characteristics over the area of interest. As pointed out by Gallego and Delince (1991a), satellite images are useful as a means of establishing a stratification procedure. Additionally, other information can be used, such as topographic maps and statistical, soil, phenologic, and climatic data. Examples of stratification procedures used in different countries in the five pilot regions of the MARS project and in the U.S.A. (NASS-USDA method) are given by Gallego and Delince (1991b). It can be pointed out that the efficiency of the stratification is not always acceptable (Gangkofner *et al.*, 1990). With regard to the number, size, and shape of the sample units, there is no single standard guideline to be followed; however, research is being done in this field. For example, Alonso *et al.* (1991) investigated the use of sample units with square segments (700 metres by 700 metres), and sample units with irregular segments (approximately 50 ha). For crop area estimates in most of the regions of the European Union (EU), an appropriate size for sample units is suggested to be between 25 and 100 ha, depending on the average field size and the homogeneity of the land use. A general rule for agricultural zones is that each sample unit should embrace on average 20 to 30 fields (Gallego and Delince, 1993b).

Sample Surveys

The main task of the sample surveys employed was to map field boundaries and to record land-cover classes within the sample units with the aid of imagettes produced with Landsat image data. For the historical sample survey (i.e., the 1975 period), aerial photographs were used as surrogate information and aerial photographic interpretation (API) techniques were applied to produce the necessary information. For the contemporary sample survey (i.e., the 1989 period), direct field observations were carried out.

An area frame sampling scheme was adopted as a framework for the sample surveys. The strata were comprised of 36 sub-populations (stratum) and each of these are seven km by seven km in size. In order to define the finite population of sample units, the Universal Transversal Mercator (UTM) coordinate system grid was imposed over each stratum. Thus, each stratum was divided into 49-sq-km units; therefore, the size of the sample units was one sq km. This sample unit size had been adopted in other studies and proved satisfactory (Bunce and Heal, 1984; Taylor and Eva, 1993). A systematic unaligned sampling approach was used to spatially select 36 sample units (i.e., one sample unit of one sq km was randomly chosen from each stratum), resulting in a sample fraction of 2.04 percent (Figure 2). This approach allows a uniform distribution of the sample units and enables the production of unbiased estimates (Webster and Oliver, 1990). Examples of the use of lower sample fractions can be found in Gangkofner *et al.* (1990).

A sample survey documentation data set, produced with the aid of Landsat images, was used to record and map field information. The methodology adopted is similar to the one suggested by Perdigão (1991) and has been used in other studies (e.g., Taylor and Eva, 1993). The documentation consisted of (1) Landsat MSS (1975) false-color composite imagettes (bands 4, 5, and 7) of the one-sq-km sample units and

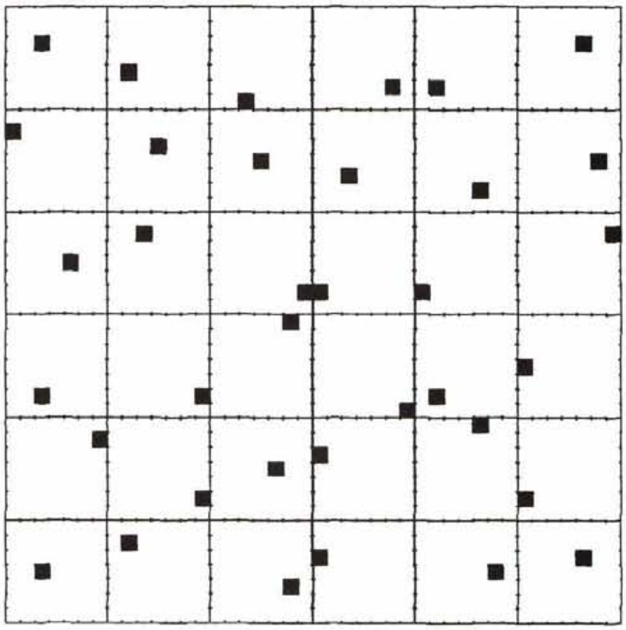

Figure 2. Area frame sampling scheme illustrating the distribution of the sample units of the test site area (42 km by 42 km, UTM coordinates: top left 406,000; 6699,000, bottom right 448,000; 6657,000). The filled squares represent the one-sq-km sample units.

surrounding areas (0.5-km border) to 1:10,000 scale; (2) Landsat TM (1989) false-color composite imagettes (bands 2, 3, and 4) of the one-sq-km sample units and surrounding areas (0.5-km border) to 1:10,000 scale; (3) 1:10,000-scale topographic map extracts (paper photocopies) of the sample units and surrounding areas (matching the imagettes); (4) transparent overlays on which to draw field boundaries within the sample units; and (5) forms on which to annotate field numbers and cover types within each of the sample units.

Sample Survey Data Sets

The field boundaries contained in the transparent overlays resulting from API and field observations were digitized in order to obtain areal measurements of the fields (i.e., land-cover classes) within the sample units. The 72 sample units (i.e., 36 for each data set) were digitized using the Tydic digitizing module of the Spans Tydac GIS software (Intera Tydac, 1991). To overcome the errors introduced in the digitization process and also due to the imprecision in drawing the boundaries of the sample units, the areas of the fields in each sample unit were weighted to the true total area in relation to the digitized area. A database was created enabling the establishment of (1) total area of each sample unit, (2) total area of each land-cover class within each sample unit, and (3) total area in the test site.

The Direct Expansion Method

The area of each land-cover class and the standard error were calculated by the direct expansion method (Gallego and Delince, 1991b) using the sampling survey and producing unbiased area estimates. The estimator for each land-cover class was its mean proportion per sample unit, for the sample units contained within the total area. This was given by the equation

$$\bar{y}_c = \frac{1}{n}\sum_{i=1}^{n} y_{ic} \tag{1}$$

where n is the number of sample units and y_{ic} is the proportion for a specific land-cover class in the i^{th} sample unit. If D is the total area, the total land-cover area Z_c is

$$Z_c = D\bar{y}_c. \tag{2}$$

If N is the total population of sample units within the test site from which the sample was taken and it is equal to the test site area divided by the area of individual sample units, the standard error S.E.(Z_c) and 95 percent confidence interval (C.I.95%) were calculated as follows:

$$\text{S.E.}(Z_c) = D\sqrt{\left[\left(1-\frac{n}{N}\right)\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(y_{ic}-\bar{y}_c\right)^2\right]} \tag{3}$$

$$\text{C.I.}_{95\%} = Z_c \pm 1.96\ \text{S.E.}\ (Z_c) \tag{4}$$

Area Estimates through the Regression Estimator Method

Area estimation with the regression estimator method consists of correcting the estimated average of a variable Y as a function of the results obtained from an auxiliary variable X (Alonso *et al.*, 1991). In other words, for a particular land-cover feature, the Y variable is the area determined as a result of a ground survey, and the X variable is the proportion of pixels resulting from the image classification for that particular land-cover feature. A regression strategy is then established for the two variables and the estimated areas are the result of the application of the regression parameters. In this method, pixel counts for each class extracted from the classified image are converted to proportions of the sample unit areas, and then regressed against the equivalent proportions extracted from the sample survey. The regression estimator method is used to produce unbiased or nearly unbiased estimates through the use of classified satellite images with the combination of ground survey data.

In order to evaluate the results of the regression estimator method, relative efficiencies (REs) can be established as the ratio between the variance of the ground survey area estimate and the variance of this estimate when it has been corrected with the use of satellite images. The RE can provide an economic evaluation of the whole approach. Thus, it will be economic if the RE reaches a threshold that can be computed when the cost of an increase in the sample size is compared with the cost of image analysis (Gallego and Delince, 1991a). As pointed out by Gallego and Delince (1993a), the regression estimator and the use of remote sensing data provide an improvement in the statistical precision. Additionally, remote sensing provides information on the spatial location of the land-cover features. Giovacchini and Brunetti (1992), investigating the use of the regression estimator method in Italy, developed a cost methodology where several components are considered (e.g., the cost of preparation of the sample units, ground survey, image acquisition, image classification, and the cost of project management). Results demonstrated the economic advantages offered by the use of remote sensing for estimating crop areas.

Area estimation of the main crops over pilot regions for the MARS project were determined for member states of the European Union in 1988, 1989, 1990, and 1991. The results from these surveys carried out in France, Spain, Belgium, Italy, and Germany are given in Gallego and Delince (1993a), including the relative efficiencies for the different crops surveyed. Values ranged from 0.99 to 6.2, thus demonstrating the advantage of using remote sensing.

Digital Image Pre-processing

Pre-processing techniques implemented were image destriping, geometric correction and resampling, image registration, and radiometric correction. Both image data sets were geometrically corrected using a first-order linear transformation

TABLE 1. CONFUSION MATRIX FOR NON EQUALLY WEIGHTED AREA CLASSIFICATION: MSS BANDS NIR, RED, GREEN

Image Data	Reference Data						Total	User's Accuracy
	Natural F.	Eucalyptus	Acacia	Pasture	Bare Soil	River/Res.		
Natural F.	15	8	3	16		1	43	34.9%
Eucalyptus	4	25	4	14		2	49	51.0%
Acacia	1	4	8	1	1		15	53.3%
Pasture	15	9	1	313	18		356	87.9%
Bare Soil	2			10	9	2	23	39.1%
River/Res.	1			1		24	26	92.3%
Total	38	46	16	355	28	29	512	
Producer's Accuracy	39.5%	54.3%	50.0%	88.2%	32.1%	82.8%	Overall Accuracy	76.95%

Kappa = 0.53611, var [Kappa] = 0.00685

and were resampled to a 25-metre pixel size with the use of the nearest-neighbor algorithm. Additionally, the Landsat MSS data set was also resampled using a bilinear interpolation, producing output images with smoother edges. This image data set was used to generate the sample survey documentation (previous section). Image registration was carried out using a second-order non-linear transformation. The RMS errors for image correction and registration were all less than one pixel.

Radiometric corrections were carried out in order to standardize both image data sets to the same meaningful units, by converting the DNs to numbers representing physical values (i.e., reflectance). Because the Landsat MSS and TM sensors had different relationships between the physical units recorded and digital output, the influence of the sensor response on incident electromagnetic energy needed to be removed (Markham and Barker, 1987). The parameters and methods used here were given by Markham and Barker (1986; 1987), who present an extensive review of the radiometric properties of Landsat sensors. The radiometric procedure involved the conversion of the raw DNs into spectral radiance and subsequently to reflectance values (i.e., at-satellite reflectance). Different methods of radiometric correction can be applied. One example is that illustrated by Robinove (1982), which takes into account spectral irradiances from a pre-solar irradiance scale. However, Markham and Barker (1987) argued that the approach adopted here allows a further normalization for sensor bandpass differences (by considering the spectral distribution of sunlight) and also allows the reduction of between-scene variability (by compensating for solar zenith angle and Earth-sun distance effects).

Digital Image Classification

The digital image classification procedures were carried out in both image data sets (i.e., Landsat MSS bands 4, 5, and 7 and Landsat TM bands 1, 2, 3, 4, 5, and 7) and aimed to

- Classify the image areas equivalent to the sample survey units which were mosaicked into a single image file using the Landsat MSS data set (1975 period),
- Classify the image areas equivalent to the sample survey units which were mosaicked into a single image file using the Landsat TM data set (1989 period),
- Produce Landsat MSS and TM classified images of the total test site area, and
- Produce statistical parameters for area estimation.

The classified mosaics and images were used in the regression method. Statistics included per-class per-sample unit pixel counts extracted from each of the classified mosaics. Per-class pixel counts were extracted from the classified images of the test site area. These were all used to calculate the regression parameters and to obtain the regression estimator. The maximum-likelihood algorithm was adopted and training information was gathered from the historical and contemporary sample survey data sets, and were used to produce training statistics for the classifier. Area weighted classifications were adopted and a threshold confidence interval of 99 percent was applied (pixels with a probability of less than 1 percent of being included in a specific class were assigned as unclassified). The image processing software used was Erdas version 7.5 (Erdas, 1991).

Feature selection procedures were adopted in order to select the best band combination data sets. Jefferies Matusita (JM) Distance parameters, overall classification results, and Kappa statistics were used. In addition, the regression parameters were included in the feature selection analysis. The analysis indicated that all three MSS and all six TM bands would be required for optimal class discrimination.

In order to assess the classifier and to obtain a measure of accuracy, confusion matrices were generated using the sample surveys data as "reference image data" and the classified mosaic images as "image data." Among the parameters produced from the confusion matrices were (1) error of omission or producer accuracy, (2) error of commission or user accuracy, and (3) overall accuracy. Additionally, a Kappa coefficient of agreement and its variance (Rosenfield and Fitzpatrick-Lins, 1986; Hudson and Ramm, 1987) were calculated (Tables 1 and 2).

The Regression Estimator Method

The regression estimator approach used in this study consisted of correcting the sample surveys area estimates with the classified satellite data. The relationship between the sample surveys and image data was established as a function of a linear regression. The regression parameters were then used to produce area estimates using proportion of the per-class pixel counts classified for the whole test site area. To produce the regression estimator parameters for the classified mosaics, a pixel summation for each class and for each sample unit was extracted. The pixel summation was then converted into the proportions of their respective sample units (i.e., each sample unit is 40 by 40 pixels for a total of 1600 pixels). For each class, least-squares linear regression was used to correlate the image sample units with the equivalent sample units determined by the sample surveys.

Data input for the calculation of the regression estimator parameters included the per-class per-sample unit pixel counts from the mosaic-classified images and the per-class pixel counts from the test site area from the classified images. The parameters are defined as follows (Taylor and Eva, 1993). The class population average ($\overline{p}_c$) is the ratio between the total classified pixel count for each class over the test site area and the total area of the test site. It is given by

$$\overline{p}_c = \frac{\text{total class pixel count}}{\text{total area}}. \quad (5)$$

The average areal proportion for each class ($\overline{c}_c$) is established with the use of the regression parameters slope (m) and intercept (b) from that class as follows:

Table 2. Confusion Matrix for Non Equally Weighted Area Classification; TM Bands TM1, TM2, TM3, TM4, TM5, TM7

	Reference Data								
Image Data	Natural F.	Eucalyptus	Acacia	Pasture	Bare Soil	River/Res.	Rice Field	Total	User's Accur.
Natural F.	30	7		7	1			45	66.7%
Eucalyptus		48	4	6	1	1		60	80.0%
Acacia		5	19	4				28	67.9%
Pasture	22	13	1	248	15	4	14	317	78.2%
Bare Soil			1	6	16		1	24	66.7%
River/Res.	1					24		25	96.0%
Rice Field				5			8	13	61.5%
Total	53	73	25	276	33	29	23	512	
Producer's Accuracy	56.6%	65.8%	76.0%	89.9%	48.5%	82.8%	34.8%	Overall Accur.	76.76%

Kappa = 0.60409, var [Kappa] = 0.00544

$$\bar{c}_c = \bar{p}_c\, m + b. \qquad (6)$$

The regression estimate of the total areal extent for a particular class (E_c) is calculated as the product of the average areal proportion for each class ($\bar{c}_c$) and the total area of the test site (N_c). The total area of the test site is divided by the size of the sample units, which, in this case (i.e., one sq km), do not change N_c. The unit adopted was hectares (ha) in order to conform with the sample surveys figures. The estimate is

$$E_c = N_c\bar{c}_c. \qquad (7)$$

In order to calculate the standard errors, the estimated variance of the regression estimator has to be established. In this case, the variance (var(E_c)) which defines the precision of the estimate is given by

$$\text{var}\,(E_c) = N_c^2 s^2 \left[\frac{1}{n} + \frac{(\bar{p}_c - \bar{p})^2}{\sum_{i=1}^{n} (p_i - \bar{p})^2}\right]. \qquad (8)$$

The class average pixel count is given by

$$\bar{p} = \frac{\sum_{i=1}^{n} p_i}{n} \qquad (9)$$

where N_c is the total area of the test site in hectares, s^2 is the residual mean square of the regression, $\bar{p}_c$ is the class population average, $\bar{p}$ is the class average pixel count, and p_i is the class pixel count per unit area for the i^{th} sample unit.

The standard error of the class area estimate (S.E.(E_c)) and the 95 percent confidence interval (C.I.$_{95\%}$) become

$$\text{S.E.}(E_c) = \sqrt{\text{var}(E_c)} \qquad (10)$$

$$\text{C.I.}_{95\%} = E_c \pm 1.96\text{S.E.}(E_c) \qquad (11)$$

In order to determine the relative success of the regression estimator, the per-class relative efficiency (RE) was established by the ratio of the variance of the sample survey area estimate to the variance of the area estimate using image data. The steps adopted to establish the regression estimation results are illustrated in Figure 3.

Area Estimation Results

Table 3 contains the area estimates, 95 percent confidence intervals (in percentage area), and the coefficients of variation obtained from both the direct expansion and the regression estimator methods, when using 1975 data, i.e., historical sample survey and Landsat MSS data sets. Additionally, coefficients of determination and relative efficiencies (REs) are shown as a result of applying the regression estimator. Similar data are given in Table 4 except that contemporary sample survey and Landsat TM data sets were used.

The area estimates obtained through the regression estimator using the Landsat MSS data set showed narrower 95 percent confidence intervals for all the classes when compared with the area estimates obtained from the direct expansion method. No class intervals or their confidence intervals (represented in terms of ha) were greater than the actual area estimates obtained from the regression estimator. The larger confidence interval found is related to the Acacia class (i.e., 53.58 percent), and the lowest is related to the Pasture class (i.e., 5.66 percent). In the case of the area estimates obtained from the direct expansion method, the confidence intervals for all the classes, with the exception of Natural Forest, Pasture, and Bare Soil, were very large. The reduction in the confidence intervals was marginal for the Natural Forest class and substantial for the other classes as follows: Natural Forest, from 48.61 percent to 45.50 percent; Eucalyptus, from 63.76 percent to 43.79 percent; Acacia, from 120.35 percent to 53.58 percent; Pasture, from 11.98 percent to 5.66 percent; Bare Soil, from 43.37 percent to 21.99 percent; and River/Reservoir, from 98.82 percent to 11.50 percent.

The coefficient of variation decreased as a result of using the regression estimator for all the classes, ranging from 2.89 percent (for Pasture) to 27.33 percent (for Acacia). Therefore, this demonstrates that the regression estimator produced improved area estimates. By comparing these figures, it can be seen that the direct expansion and the regression estimator for the classes Natural Forest and Pasture are very similar. The classes Eucalyptus and Bare Soil were underestimated by the direct expansion method, and River/Reservoir and Acacia were overestimated. Relative efficiencies indicate improvements (higher values) in the regression estimator for all the classes, especially for Acacia, Pasture, and River/Reservoir.

The area estimates obtained from the regression estimator using the Landsat TM data set also showed narrower 95 percent confidence intervals for all the classes when compared to the area estimates obtained from the direct expansion method. The confidence intervals decreased for all the classes. No class intervals or their confidence intervals (represented in terms of ha) were greater than the actual area estimates obtained from the regression estimator. The larger 95 percent confidence interval found is related to the Rice Field class (i.e., 75.50 percent), and the lowest is related to Pasture class (i.e., 7.50 percent). In the case of the area estimates obtained from direct expansion, the confidence intervals for all the classes with exception of Pasture and Bare Soil were very large. The reduction in the confidence intervals was substantial for all the classes as follows: Natural Forest, from 53.88 percent to 30.61 percent; Eucalyptus, from 52.57 percent to 27.59 percent; Acacia, from 107.60 percent to 67.54 percent; Pasture, from 16.01 percent to 7.50 percent; Bare Soil, from 35.86 percent to 31.42 percent; River/Reservoir,

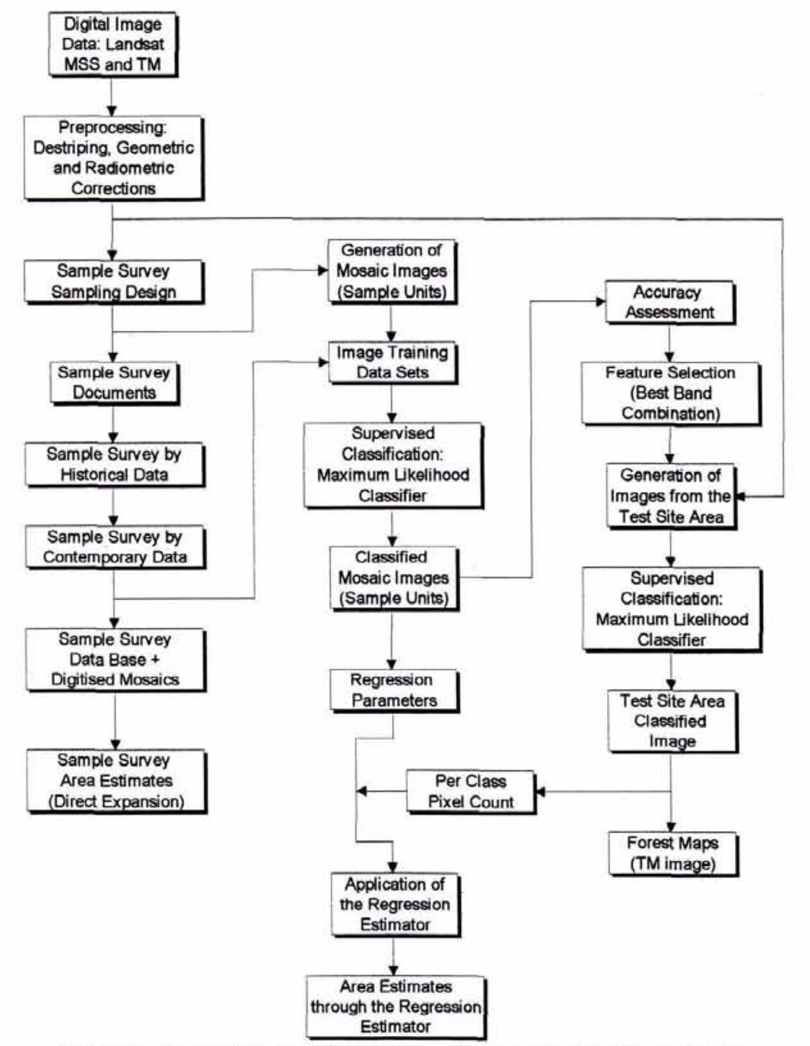

Figure 3. Area estimation flowchart; from image input to outputs.

from 85.25 percent to 24.55 percent; and Rice Field, from 100.31 percent to 75.50 percent.

The coefficient of variation decreased as a result of using the regression estimator for all the classes, ranging from 3.83 percent for Pasture to 38.52 percent for Rice Field. Again, as with the Landsat MSS data set, this demonstrated that area estimates were improved by the application of the regression estimator. By comparing the figures, it can be seen that the estimates for the classes Natural Forest, Eucalyptus, and Bare Soil from both the direct expansion and regression estimator are similar. The classes Acacia, River/Reservoir, and Rice Field were overestimated, whereas Pasture was underestimated by the direct expansion method. REs indicate improvements (higher values) in the regression estimator for all the classes, especially Eucalyptus, Acacia, Pasture, River/Reservoir, and Rice Field.

When using the regression estimator with the Landsat MSS data set, the coefficients of determination were considered low for the purpose of the regression estimator for the classes Natural Forest and Eucalyptus (i.e., 0.258 and 0.385).

TABLE 3. 1975'S AREA ESTIMATES BY DIRECT EXPANSION AND REGRESSION ESTIMATOR; MSS DATA SET (BANDS GREEN, RED, NIR)

	Direct Expansion of Sample Survey Data			Regression Estimator of Image Data				
Classes	Area (ha)	C.I.(95%) (%)	C.V. (%)	Area (ha)	C.I.(95%) (%)	C.V. (%)	R. sq.	R.E.
1 Natural Forest	13034.00	48.61	24.82	13022.19	45.50	23.22	0.258	1.3
2 Eucalyptus	9859.29	63.76	24.53	12366.54	43.79	22.34	0.385	1.6
3 Acacia	7111.86	120.35	61.40	5935.76	53.58	27.33	0.875	8.0
4 Pasture	119472.30	11.98	6.11	119720.99	5.66	2.89	0.734	3.8
5 Bare Soil	14019.39	43.37	22.13	19085.29	21.99	11.22	0.657	2.9
6 River/Reservoir	10439.45	98.82	46.34	5792.70	22.54	11.50	0.983	58.8

C.I. = Confidence Interval, C.V. = Coefficient of Variation (Std. E./Area)*(100), R.E. = Relative Efficiency

TABLE 4. 1989'S AREA ESTIMATES BY DIRECT EXPANSION AND REGRESSION ESTIMATOR; TM DATA SET (BANDS TM1, TM2, TM3, TM4, TM5, TM7)

	Direct Expansion of Sample Survey Data			Regression Estimator of Image Data				
Classes	Area (ha)	C.I.(95%) (%)	C.V. (%)	Area (ha)	C.I.(95%) (%)	C.V. (%)	R. sq.	R.E.
1 Natural Forest	14573.09	53.88	27.49	16701.22	30.61	15.62	0.631	2.7
2 Eucalyptus	22155.84	52.57	26.82	20693.01	27.59	14.08	0.780	4.6
3 Acacia	8412.32	107.60	54.90	3396.83	67.54	34.46	0.937	15.9
4 Pasture	97734.91	16.01	8.17	105515.30	7.50	3.83	0.769	4.3
5 Bare Soil	13059.97	35.86	18.30	14209.39	31.42	16.03	0.223	1.3
6 River/Reservoir	11634.07	85.25	43.50	7860.44	24.55	12.52	0.966	29.8
7 Rice Field	6328.35	100.31	51.18	4307.31	75.50	38.52	0.769	4.3

C.I. = Confidence Interval, C.V. = Coefficient of Variation (Std. E./Area)*(100), R.E. = Relative Efficiency

In the case of using Landsat TM data set, the coefficient of determination was low for Bare Soil (i.e., 0.223). The regression relationships for these classes were 1.84E-03, 6.94E-05, and 4.23E-03, respectively. For the above classes, it can be concluded that the regression estimator was unable to produce acceptable area estimates due to the poor relationship between the ground and image observations. The main reason for this could be related to spectral confusion. However, when analyzing the relative efficiencies (RE), it can be observed that the regression estimator generated a reduced variance mean, thus indicating the advantage of using the satellite image data for area estimation for all classes. However, despite the RE, the regression estimator results cannot be used for those classes which had low coefficients of determination. In view of this, the regression estimator results were considered adequate for the following classes: (1) using the Landsat MSS data set: Acacia, Pasture, and River/Reservoir; and (2) using the Landsat TM data set: Natural Forest, Eucalyptus, Acacia, Pasture, River/Reservoir, and Rice Field. Consequently, the classes subject to analysis of changes were Acacia, Pasture, and River/Reservoir. In order to define the significance of changes, standard errors and 95 percent confidence intervals were considered. If the confidence intervals of the differences between 1989 and 1975 (unpaired observations, i.e., where the variances from each time period are summed) were larger than the regression estimator area, the difference or change in area was considered to be significant. The only class presenting significant area changes was Pasture (Table 5).

The application of the regression estimator method produced acceptable area estimation results only for the classes Acacia, Pasture, and River/Reservoir when using the Landsat MSS data set. On the other hand, when using the Landsat TM data set, the regression estimator produced acceptable results for the classes Natural Forest, Eucalyptus, Acacia, Pasture, River/Reservoir, and Rice Field, thus indicating the superiority of the TM image for class discrimination. However, statistically significant changes were confirmed only for the Pasture class. Changes in forest could not be statistically confirmed.

Comments, Conclusions, and Recommendations

The methodology employed to acquire sample survey data proved to be appropriate for generating reference data and area estimates using the direct expansion method. It proved to be robust, particularly when seeking independent and rapid forest area estimates. It was demonstrated that the accuracy of the area estimates increased with the area of the classes.

Regarding the utilization of the Landsat MSS images, two aspects have to be considered. First, the scale of the aerial photographs used to produce the reference historical sample survey data set should be greater (e.g., 1:25,000) for a better determination of field boundaries and their correct identification and labeling. Second, as expected, the limited spectral information provided by the Landsat MSS in the infrared region of the electromagnetic spectrum represents a drawback for vegetation studies. Additionally, the coarse spatial resolution could interfere in the spectral information content (in comparison with Landsat TM spatial resolution). Thus, the classification accuracies were reduced due to the limitations imposed by the historical sample survey data set and MSS sensor characteristics. The analysis of the Landsat MSS classification results indicated that further investigation is needed before making conclusive statements with regard to its utilization for forest applications in the test site area. On the other hand, the Landsat TM images produced better classification results. The reasons for this are related to the improved resolutions (spatial, spectral, and radiometric) and the use of reference sample survey data acquired through direct field observations.

The combination of sample survey data with the classified image data resulted in area estimates with increased ac curacy. The advantage of using Landsat image data to establish area estimates through the regression estimator method, when compared to the direct expansion method using sample survey data, was clearly demonstrated. Relative efficiencies (REs) for forest classes when using the MSS image data ranged from 1.3 to 8.0. When using the TM image data, the REs were larger and ranged from 2.7 to 15.9. However, area estimate results for the Natural Forest and Eucalyptus classes, when using Landsat MSS image data, were not considered due to low R. sq. Changes in forest (calculated as the differences between areas estimated by the regression estimator method using Landsat MSS and TM data) were restricted to the Acacia class and were not significant.

Finally, there is a need for further research into the several tasks applied in this study. Moreover, the investigation of the use of remote sensing to produce area estimates for forest remains open in the state of Rio Grande do Sul (RS), Brazil. The findings of this study demonstrated the potential and advantages of using Landsat image data for forest area estimation and established a base line to carry out further investigations. With reference to the sample surveys, further

TABLE 5. REGRESSION ESTIMATOR AREA STATISTICS - CHANGES BETWEEN 1989 AND 1975 DATA SETS

	Regression Estimator Area (ha)	Std. Error (ha)	C.I. 95% (ha)	Remark
Acacia	−2538.93	2000.62	3921.21	Not Signif.
Pasture	−14205.69	5318.24	10423.76	Signif.
River/Reservoirs	2064.74	1188.57	2329.60	Not Signif.

Std. Error = Standard Error, C.I. = Confidence Interval (95%)
Signif. = Significant

investigation is needed in the use of different sample sizes and larger sampling rate. These will increase the number of fields embraced by a single sample unit and, therefore, will reduce sampling errors and variance. Digital image classification procedures have to be investigated in order to increase accuracy results and, therefore, improve the spectral discrimination between the forest classes. This would further improve the area estimates obtained through the regression estimator and would be desirable particularly for the Landsat MSS image data (to improve the relationship between sample survey and image data). As an alternative, it is suggested that principal component analysis techniques and multi-temporal images should be used. Further research is needed in order to investigate and establish a robust approach for area change analysis when using area estimates obtained through the regression estimator method. This would increase the potential of using regression estimator data in forest monitoring applications.

Acknowledgments

The author wishes to acknowledge the National Remote Sensing Centre Limited (NRSC Ltd), England, UK, for supporting the sample surveys activities; Dr. John C. Taylor for his constant help during the study; Dr. C. Bird and Dr. G. Thomas, Silsoe College, England, UK, for their suggestions and comments; Dr. Vitor Haertel, director of the Centre for Research in Remote Sensing and Meteorology (CEPSRM), Federal University of Rio Grande do Sul (UFRGS), Brazil, for supporting the sample surveys activities and providing data sources; and CNPq/Ministry of Science and Technology, Brazil, for sponsoring the research.

References

Allen, J.D., and G.A. Hanuschak, 1988. *The Remote Sensing Applications Program of the National Agricultural Statistics Service: 1980–1987*, NASS Report SRB-88-08, National Agricultural Statistics Service, Washington, D.C.

Alonso, F.G., S.L. Soria, and J.M.C. Gonzalo, 1991. Comparing two methodologies for crop area estimation in Spain using Landsat TM images and ground-gathered data, *Remote Sensing of Environment*, 35:29–35.

Annoni, A., and F.J. Gallego, 1992. Software for crop area estimation with remote sensing (MARS-PED) and quick digitization of sample units (TTS), *Proceedings of the Conference on the Application of Remote Sensing to agricultural statistics* (F. Toselli and J. Meyer-Roux, editors), Institute for Remote Sensing Applications, Joint Research Centre, Ispra, Villa Carlotta, Belgirate, Lake Maggiore, Italy, 26–27 November 1991, pp. 117–124.

Bierlaire, P., C. Silber, and M. Sharman, 1993. Operational near-real time monitoring of European agriculture using satellite and vectorial information, *Proceeding of the Fourth European Conference and Exhibition on Geographical Information Systems*, (J. Harts, H.F.L. Ottens, and H.J. Scholten, editors), Genova, Italy, 29 March — 1 April 1993, I:344–353.

Bunce, R.G.H., and O.W. Heal, 1984. Landscape evaluation and the impact of changing land use on the rural environment: the problems and an approach, *Planning and Ecology* (R.D. Roberts and T.M. Roberts, editors), Chapman and Hall, London, pp. 164–187.

Daubin, K.W., and D.W. Beach, 1981. Crop monitoring using Landsat, *Proceedings of the 2nd Australasian Landsat Conference*, September 1981, Canberra, Australia.

Erdas, 1991. *Field Guide*, Erdas Inc., Atlanta, Georgia.

Gallego, J., and J. Delince, 1991a. Stratification for acreage regression estimators with remote sensing, *Manual for the Use of Software for Agricultural Statistics Using Remotely Sensed Data* (A. Annoni, F. Dicorato, and J. Stakenborg, editors), Commission of the European Communities, Institute for Remote Sensing Applications, Joint Research Centre, Ispra, Varese, Italy.

———, 1991b. Crop area estimation through area frame sampling and remote sensing, *Manual for the Use of Software for Agricultural Statistics Using Remotely Sensed Data* (A. Annoni, F. Dicorato, and J. Stakenborg, editors), Commission of the European Communities, Institute for Remote Sensing Applications, Joint Research Centre, Ispra, Varese, Italy.

———, 1993a. Regression estimates with remote sensing, *Euro Courses; Remote Sensing Applied to Agricultural Statistics: Regional Inventories* (F.J. Gallego, editor), Commission of the European Communities, Institute for Remote Sensing Applications, Joint Research Centre, Ispra, Athens, 29 March - 2 April 1993.

———, 1993b. Area estimation by segment sampling, *Euro Courses; Remote Sensing Applied to Agricultural Statistics: Regional Inventories* (F.J. Gallego, editor), Commission of the European Communities, Institute for Remote Sensing Applications, Joint Research Centre, Ispra. Athens, 29 March - 2 April, 1993.

Gangkofner, U., A. Relin, and S. Saradeth, 1990. Regional inventory 1988/89 in Bavaria, *Proceedings of the Conference on the Application of Remote Sensing to Agricultural Statistics* (F. Toselli and J. Meyer-Roux, editors), Institute for Remote Sensing Applications, oint Research Centre, Ispra. Villa Andrea Ponti, Varese, Italy, 10–11 October 1989, pp. 87–96.

Giovacchini, A., and A. Brunetti, 1992. Agricultural statistics by remote sensing in Italy, an ultimate cost analysis, *Proceedings of the Conference on the Application of Remote Sensing to Agricultural Statistics* (F. Toselli and J. Meyer-Roux, editors), Institute for Remote Sensing Applications, Joint Research Centre, Ispra. Villa Carlotta, Belgirate, Lake Maggiore, Italy, 26-27 November 1991, pp. 87–93.

Hudson, W.D., and C.W. Ramm, 1987. Correct formulation of the Kappa coefficient of agreement, *Photogrammetric Engineering & Remote Sensing*, 53:421–422.

Intera Tydac, 1991. *SPANS (Spatial Analysis System)*, Intera Tydac Technologies, Inc.

Latham, J.S., D.C. Ferns, J.E. Colwell, R. Reinhold, and E.H. Jebe, 1983. Monitoring the changing aereal extent of irrigated lands of the Gefara Plain, Lybia, *Advance in Space Research*, 2:57–68.

Markham, B.L., and J.L. Barker, 1986. Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures, *EOSAT Landsat Technical Notes*, 1:3–7.

———, 1987. Radiometric properties of U.S. processed Landsat MSS data, *Remote Sensing of Environment*, 22:22–71.

Meyer-Roux, J., 1992. The development of the agriculture project of the EEC, *Proceedings of the Conference on the Application of Remote Sensing to Agricultural Statistics* (F. Toselli and J. Meyer-Roux, editors), Institute for Remote Sensing Applications, Joint Research Centre, Ispra. Villa Carlotta, Belgirate, Lake Maggiore, Italy, 26–27 November 1991, pp. 31–39.

Perdigão, V., 1991. *Ground Survey Documents Based on High Resolution Satellite Imagery 1990 Campaign*, Publication No. S.P.I.91.29, Institute for Remote Sensing Applications, Joint Research Centre, Ispra, Varese, Italy.

Robinove, J.C., 1982. Computation with physical values from Landsat digital data, *Photogrammetric Engineering & Remote Sensing*, 48: 781–784.

Rosenfield, G.H., and K. Fitzpatrick-Lins, 1986. A coefficient of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 52:223–227.

Ryerson, R.A., R.N. Dobbins, and C. Thibault, 1985. Timely crop area estimates from Landsat, *Photogrammetric Engineering & Remote Sensing*, 51:1735–1743.

Sharman, M., 1992. Action II: monitoring vegetation with AVHRR: purpose, principles and products, *Proceedings of the Conference on the Application of Remote Sensing to Agricultural Statistics* (F. Toselli and J. Meyer-Roux, editors), Institute for Remote Sensing Applications, Joint Research Centre, Ispra, Villa Carlotta, Belgirate, Lake Maggiore, Italy, 26–27 November 1991, pp. 145–158.

Taylor, J.C., and H.D. Eva, 1993. Operational use of remote sensing for estimating crop areas in England, *Towards Operational Application, Proceedings of the 19th Annual Conference of the Remote Sensing Society* (K. Hilton, editor), Chester College.

Webster, R., and M.A. Oliver, 1990. *Statistical Methods in Soil and Land Resource Survey*, Oxford University Press, Oxford, UK.

(Received 17 April 1996; accepted 15 July 1996; revised 18 June 1997)