Available online at www.sciencedirect.com

SCIENCE DIRECT

Remote Sensing of Environment 97 (2005) 39 – 51

Remote Sensing of Environment

www.elsevier.com/locate/rse

Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery

Kati J. Salovaara[a,*], Sirpa Thessler[b], Riffat N. Malik[c], Hanna Tuomisto[a]

[a]*University of Turku, Department of Biology, 20014 Turku, Finland*
[b]*Finnish Forest Research Institute, 00170 Helsinki, Finland*
[c]*Environmental Sciences Programme, Fatima Jinnah Women University, The Mall, Rawalpindi 46000, Pakistan*

Received 16 September 2004; received in revised form 8 April 2005; accepted 9 April 2005

Abstract

A lack of spatially and thematically accurate vegetation maps complicates conservation and management planning, as well as ecological research, in tropical rain forests. Remote sensing has considerable potential to provide such maps, but classification accuracy within primary rain forests has generally been inadequate for practical applications. Here we test how accurately floristically defined forest types in lowland tropical rain forests in Peruvian Amazonia can be recognized using remote sensing data (Landsat ETM+ satellite image and STRM elevation model). Floristic data and a vegetation classification with four forest classes were available for eight line transects, each 8 km long, located in an area of ca 800 km^2. We compared two sampling unit sizes (line transect subunits of 200 and 500 m) and several image feature combinations to analyze their suitability for image classification. Mantel tests were used to quantify how well the patterns in elevation and in the digital numbers of the satellite image correlated with the floristic patterns observed in the field. Most Mantel correlations were positive and highly significant. Linear discriminant analysis was used first to build a function that discriminates between forest classes in the eight field-verified transects on the basis of remotely sensed data, and then to classify those parts of the line transects and the satellite image that had not been visited in the field. Classification accuracy was quantified by 8-fold crossvalidation. Two of the tierra firme (non-inundated) forest types were combined because they were too often misclassified. The remaining three forest types (inundated forest, terrace forest and Pebas formation/intermediate tierra firme forest) could be separated using the 500-m sampling units with an overall classification accuracy of 85% and a Kappa coefficient of 0.62. For the 200-m sampling units, the classification accuracy was clearly lower (71%, Kappa 0.35). The forest classification will be used as habitat data to study wildlife habitat use in the same area. Our results show that remotely sensed data and relatively simple classification methods can be used to produce reasonably accurate forest type classifications, even in structurally homogeneous primary rain forests.

Keywords: Landsat ETM+; Satellite imaging; Peruvian Amazonia; Rain forest; Vegetation classification; Wildlife habitat

1. Introduction

Western Amazonian rain forests have traditionally been divided into two major forest types: inundated forests and non-inundated (tierra firme) forests. Inundated forests are structurally very heterogeneous, which has been recognized to depend on factors such as the frequency, depth and duration of inundation (Kalliola et al., 1992; Kubitzki, 1987; Prance, 1979). The tierra firme forests, in contrast, are structurally rather uniform. Until very recently, little attention was paid to the possibility that tierra firme forests might also represent ecologically different forest types. However, now there is increasing evidence that they do consist of a heterogeneous mosaic of floristic communities, and that the floristic differences between sites are related to differences in edaphic site conditions, especially soil nutrient content (Duivenvoorden, 1995; Duque et al., 2002; Gentry, 1988; Phillips et al., 2003; Ruokolainen et al., 1997; Ruokolainen &

* Corresponding author.
E-mail address: kati.salovaara@utu.fi (K.J. Salovaara).

0034-4257/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2005.04.013

Tuomisto, 1998; Tuomisto et al., 1995, 2003a,b,c; Vormisto et al., 2000).

Consequently, Amazonian tierra firme forests can be classified to different site types, each of which has a more or less characteristic floristic composition. In boreal and temperate forests, such site types are recognised in the field using indicator species (Cajander, 1926; Ellenberg, 1988) whose edaphic requirements are sufficiently well known. It has recently been suggested that pteridophytes (ferns and fern allies) could be used as forest site type indicators in Amazonian rain forests, and indeed it has been found that the floristic patterns of pteridophytes are highly correlated with those of other plant groups (Ruokolainen et al., 1997; Ruokolainen & Tuomisto, 1998; Tuomisto et al., 1995; Vormisto et al., 2000). This linkage between soils, terrestrial pteridophytes and canopy plants suggests that it may be possible to use remotely sensed data to separate rain forest types that have been recognized based on their indicator species composition.

Floristic and edaphic heterogeneity is important for land-use planning and conservation in Amazonian forests and may also be relevant for many other organisms beside plants. For example, herbivorous animals depend on the quality and quantity of plant resources available in the habitat and thus site productivity and floristic composition may influence their abundance and distribution. Therefore broad-scale vegetation surveys have potentially wide applications in resource management and conservation as well as in basic ecological research in primary rain forests.

However, vegetation surveys in tropical forests are more demanding and time-consuming than in temperate forests because plant species diversity is extremely high, their taxonomy is poorly known, and remote areas are logistically difficult to work in. As a result, it is rarely feasible to collect field data that cover the area of interest sufficiently and thus interpolation must be made between widely scattered field study points. Environmental reference maps or other ancillary data that could help in vegetation mapping and spatial interpolation are only rarely available, especially at a relevant spatial scale. Remote sensing data is often the best if not the only available source of such information (Moran & Brondizio, 1994; Tuomisto, 1998; Tuomisto et al., 2003c). For example, in our study area in Peruvian Amazonia the only available thematic map providing reasonable detail is a geological map that is based on very limited field-verification (Sanchez et al., 1999).

Although perennial cloud cover often poses problems for passive remote sensing in tropical areas (Asner, 2001), remotely sensed data have been successfully applied in separating areas of dense forest from degraded forest and non-forest land cover types (e.g., Achard et al., 2001; Laporte et al., 1995; Skole & Tucker, 1993). Seasonally inundated forests and swamps have also been successfully distinguished from non-inundated forests (de Grandi et al., 2000; Kalliola et al., 1992; Lobo & Gullison, 1998; Podest & Saatchi, 2002; Tuomisto et al., 1994) and some of the forest types traditionally recognized by local tribes have been found to be distinct in satellite images (Shepard et al., 2004).

Nevertheless, the majority of Amazonia (e.g., more than 80% of Peruvian Amazonia; Salo et al., 1986) is covered by relatively uniform-looking tierra firme forests. The differences between tierra firme forest types that grow on different soil types have been considered to be too subtle to be discernible by remote sensing. For example, several studies using Landsat TM data to classify tierra firme forests growing on sand and clay soils concluded that it was not possible to separate them (Foody & Hill, 1996; Hill, 1999; Hill & Foody, 1994). In spite of this, our earlier studies have linked the variation visible in Landsat satellite imagery with floristic patterns and environmental heterogeneity in Amazonian tierra firme forests (Thessler et al., in press; Tuomisto et al., 1995, 2003a,b), although it has also been suggested that the vegetation patterns visible in satellite images may be related to forest dynamics, such as pest outbreaks, disturbance or dispersal limitation, rather than to edaphic variation (Condit, 1996).

Numerous factors affect the potential success of forest classification using satellite images. One challenge is to select the best image features to be used as the basis of the classification. For the purposes of vegetation mapping, the near and mid infrared bands contain more relevant information than the visible bands (e.g., Singh, 1987; Thenkabail et al., 2004), but the most efficient features vary from one application to another. The spatial resolution of the study is also critical for classification success. The size of the pixel windows used should be selected relative to the size of the forest patches that are to be mapped. The spatial scales of vegetation patterns in Amazonian forests range from local tree fall gaps (tens of meters) through topographic hill–valley patterns (hundreds of meters) to landscape-wide patches of different forest types (kilometres) (Tuomisto et al., 1995, 2003a,b). Local heterogeneity within forest types can be considered noise in landscape-scale vegetation mapping and consequently more accurate classifications have been obtained when larger pixel windows (e.g., Hill & Foody, 1994; Rajaniemi et al., 2005; Tuomisto et al., 2003a, b) or segments (Hill, 1999; Lobo & Gullison, 1998) have been used. Use of pixel windows or segments also decreases the effect of misregistration of field plots and errors in satellite image rectification, both of which can cause classification error (Foody, 2002; Powell et al., 2004). However, if the spatial resolution of the study is too coarse, the classification can be blurred by high variation within pixel windows, as one window may include several vegetation types.

The number and distinctness of the forest classes to be mapped partly determine how accurate a classification can be achieved. In continuous tierra firme forest, easily recognizable limits between forest types are rare. This makes the allocation of sites to discrete vegetation classes difficult and often quite subjective, and therefore care

should be taken in defining the forest classes to be mapped (Foody, 2002; Nagendra, 2001). The availability of floristic field data helps in this task.

We have previously conducted a landscape-level vegetation classification in Peruvian Amazonia for the purposes of wildlife habitat studies (Salovaara et al., 2004). Eight 8-km long line transects were inventoried for their pteridophyte flora. On the basis of these floristic data, the forests were divided into four main types: inundated forests, terrace forests, Pebas formation forests and intermediate tierra firme forests. The pteridophyte species composition in the three non-inundated forest types indicated that they differ in soil fertility: terrace forests have clearly poorer soils than the two other forest types and Pebas formation forests have richer soils than intermediate tierra firme forests (Salovaara et al., 2004). Since tree species composition reacts to these kinds of soil differences (e.g., Clark et al., 1998; Duque et al., 2002; Phillips et al., 2003; Ruokolainen et al., 1997), the spectral characteristics of the forest are expected to do so as well, because they are mostly determined by the forest canopy.

A mammal census was conducted simultaneously with the vegetation survey in the eight floristic transects and in an additional four transects that lack floristic field data in order to compare the structure of mammal communities between forest types. Obtaining a vegetation classification for all twelve transects is indispensable for the animal ecological study. In the present study the objective was to test whether a Landsat ETM+ satellite image can be used to extend the vegetation classification to the additional four transects and, if so, to carry out the classification. Although there is an immediate practical need for the classification, it is also of more general scientific interest to test how accurately the forest types recognized on the basis of floristic data can be detected from remotely-sensed data.

2. Material and methods

2.1. Study area

Fieldwork was conducted in an extensive (800 km^2) uninhabited area of primary rain forest in northeastern Peruvian Amazonia along the Yavarí–Mirín river (Fig. 1). The Yavarí–Mirín is a tributary of the Yavarí river, which forms the border between Peru and Brazil. The area is within the proposed Yavarí conservation unit and has extremely high species diversity (Pitman et al., 2003 and references therein). Although restricted selective logging and other

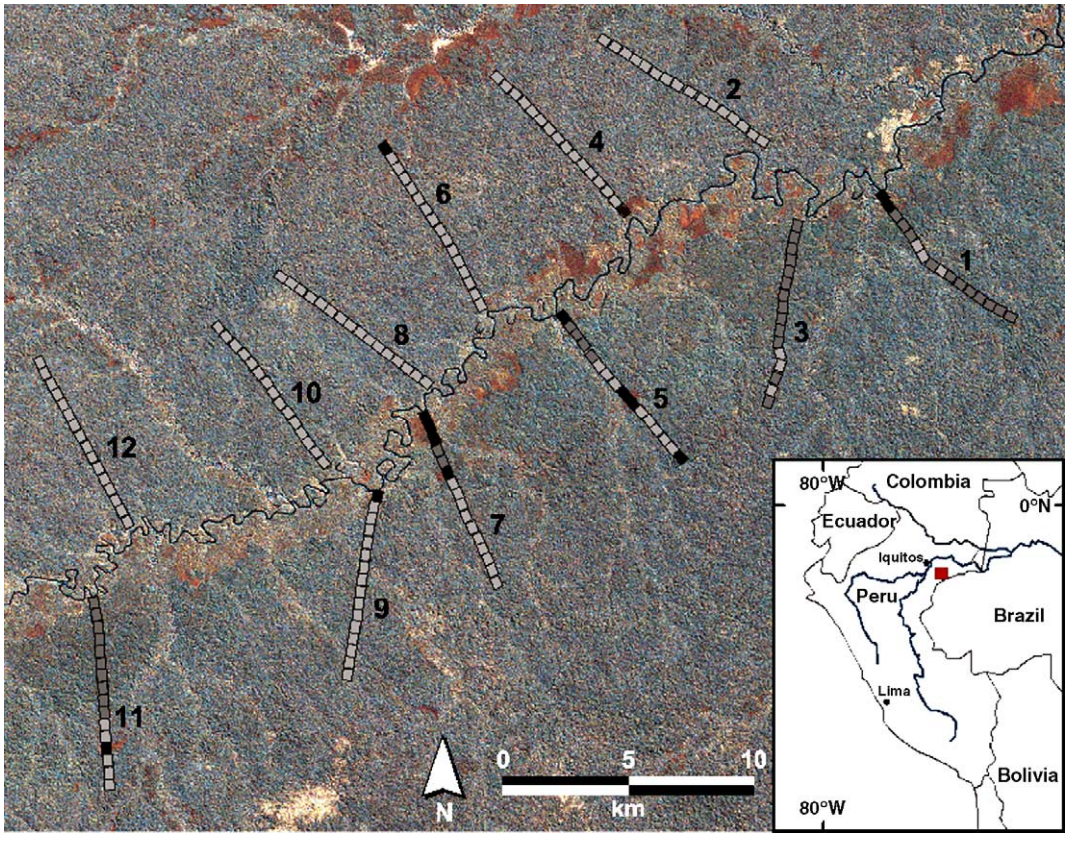

Fig. 1. Color composite (bands 4, 5 and 7 allocated to red, green and blue, respectively) of the Landsat ETM+ satellite image covering the study area in NE Peruvian Amazonia. Each square in the line transects represents one 500 × 400 m sampling unit. The three forest classes are shown as follows: inundated forests in black, terrace forests in dark gray and Pebas formation/intermediate tierra firme forests in light gray. Classification of the 500-m sampling units in transects 4–8 and 10–12 was based on floristic data (Salovaara et al., 2004), whereas transects 1, 2, 3 and 9 were classified in the present study with the help of remotely sensed data. The red square in the reference map indicates the location of the study area.

small-scale extractive activities take place in the area, they have not caused visible changes in the forest structure.

Elevation in the area ranges from 100 m to ca 180 m above sea level and terrain varies from practically flat to hilly. Most of the river basin is covered by non-inundated tierra firme forest. Areas close to the Yavarí–Mirín river channel are inundated seasonally and the floodplains of the major creeks sporadically after heavy rains. Palm swamps dominated by *Mauritia flexuosa* L. f. are patchily distributed throughout the landscape, mainly close to the river and major creeks. The vegetation and geomorphology of the area are described in more detail in Salovaara et al. (2004). The nearest weather station is in the city of Iquitos, ca 150 km northwest of the study area. There the mean monthly temperature varies between 25 and 27 °C and mean annual precipitation is about 3100 mm. The highest long-term mean monthly precipitation is in March (350 mm) and the lowest in June (180 mm), but years when the driest month receives less than 100 mm of rain are not uncommon (Marengo, 1998).

2.2. Data

Twelve 8-km long transects (Fig. 1) were cleared in the study area to allow a floristic survey and census of large-bodied herbivorous mammals (ungulates, primates, rodents) to be conducted. A vegetation classification for eight of the transects was produced with the help of floristic data (pteridophyte species abundances) collected in contiguous 2-m wide and 100-m long sampling units along the transects in March–April 2002 (Salovaara et al., 2004). The field work had to be ended prematurely because of logistical difficulties, leaving four transects (numbered 1, 2, 3 and 9 in Fig. 1) without floristic data. The present paper reports the use of satellite-derived data to assign the vegetation along these four transects to the forest classes recognized by Salovaara et al. (2004).

A Landsat Enhanced Thematic Mapper (ETM+) image covering the study area (path 005, row 63, acquisition date 25 June 2001) was used in the analysis. The image was obtained as geometrically and radiometrically corrected (level 1G) data in digital number (DN) values. The image was rectified using 1st order polynomial rectification. Ground control points (GCPs) were obtained in the field with a hand-held Global Positioning System (Garmin GPS 12) and from reference points read from a map that was based on Landsat MSS (Multispectral Scanner) images (IFG, 1984). Although the map manufacturer reports a relatively low spatial accuracy (125 m) for these maps, they have been useful in our earlier applications. A total of 12 GCPs were used as control points and additional four as check points, whose root mean square error was <1 pixel, or less than 30 m. River channels and oxbow lakes are the only usable landmarks in the region and a somewhat larger scene than the one shown in Fig. 1 was used for rectification to allow more even dispersion of GCPs.

All transects were georeferenced in the field by taking GPS readings at 500-m intervals. To improve the accuracy of the GPS data, a minimum of two repeated measurements were taken on different dates at each point and averaged. When the transects were overlaid on the Landsat ETM+ their locations were deemed sufficiently accurate (<90 m error) by visual estimation.

No topographical maps covering the study area exist, but a digital elevation model of the Shuttle Radar Topography Mission (SRTM DEM) is freely available (http://srtm.usgs.gov/) and was acquired as unfinished data in 2003. The SRTM DEM is based on the C-band radar data and has a horizontal resolution of 90 m. However, the study area is relatively flat and has only minor topographical variation, so topographical correction at this spatial resolution cannot be accomplished using SRTM data. We also tested whether the radiance values of the satellite image were correlated with incidence angle calculated using the DEM and known sun elevation, but no statistically significant correlation existed, and consequently topographical correction was not deemed necessary. No atmospheric correction was applied to the Landsat ETM+ image, because only one scene was used and the lack of atmospheric data for the time of image acquisition would have prevented reliable within-scene correction (Song et al., 2001).

Two different window sizes were used to extract digital number (DN) values from the satellite image. A balance needed to be found between solving two problems: larger pixel windows are more often mixtures of more than one forest type, but smaller windows are more severely affected by local noise. The smaller windows used here were 5×5 pixels (150×150 m) in size, and the larger windows ca 12×15 pixels (360×450 m). A window size 5×5 was selected because smaller pixel windows (3×3 and single pixels) were found to perform poorly by Hill & Foody (1994), and the larger window size is comparable to that used by Tuomisto et al. (2003b). The smaller pixel windows were used in analyses including the 200-m long floristic sampling units from the transects and the larger windows in analyses of the 500-m field sampling units. The image sampling units were made slightly shorter than the corresponding field sampling units in order to avoid including the same image pixels in two adjoining pixel windows.

The mean and standard deviation of the DN values within the pixel windows were extracted from the Landsat ETM+ image for bands 2, 4, 5 and 7. The ratio of bands 4 and 5 was also extracted. Bands 1 and 3 were not used, because they had either pronounced striping (band 1) or coherent noise (band 3), which originate from the Landsat detector. It was not possible to replace the faulty image, because no other cloud-free Landsat ETM+ scenes were available from the time period when the field study was conducted. Thermal band 6 was also excluded and the river channel was masked out to exclude pixels representing open water.

The DN values inside the 5×5 pixel windows were obtained by filtering. A moving window kernel of 5×5

pixels was applied over the image and the DN value of the pixel at the mid-point of each 200-m long transect sampling unit was used as the image data corresponding to that sampling unit. The 12×15 pixel windows were delineated manually on the image so that the long axis of each pixel window was aligned with and centered on the corresponding 500-m transect sampling unit (Fig. 1). Elevation data were extracted from the SRTM DEM. The value of the pixel containing the mid-point of each 200-m sampling unit was used as the elevation data corresponding to that sampling unit. For the 500-m sampling units, the mean and standard deviation of elevation were extracted using the same pixel window size as in the case of the DN values (360×450 m, corresponding to 4×5 pixels in the DEM).

3. Data analyses

3.1. Mantel tests

All the analyses were run separately using the 200-m and 500-m sampling units ($N=317$ for the 200-m sampling units, $N=127$ for the 500-m sampling units). Mantel tests were run to analyze how well the patterns in the digital numbers of the satellite image and in elevation correlated with the floristic patterns observed in the field. The Mantel test computes a correlation coefficient (Mantel's r) between two resemblance matrices, in the present case a floristic resemblance matrix and a satellite-derived resemblance matrix. Mantel's r is similar to Pearson's product–moment correlation, except for the important difference that correlations are not calculated between the original variables, but between resemblance matrices based on these variables. Resemblance matrices consist of pair-wise comparisons of all sampling units using a resemblance measure that is appropriate for the data at hand (Legendre & Legendre, 1998). The statistical significance of the Mantel correlation coefficient is estimated with a Monte Carlo permutation test; in the present case 999 permutations were used to test for significance at the $p<0.001$ level.

The floristic resemblance matrices were calculated using the Bray–Curtis index, which is a measure of the degree of difference in species abundances between two sites (Legendre & Legendre, 1998):

$$D(x_1, x_2) = \frac{\sum_{j=1}^{p} |y_{1j} - y_{2j}|}{\sum_{j=1}^{p} |y_{1j} + y_{2j}|}, \text{ where} \quad (1)$$

x_i = sampling unit i
y_{ij} = abundance of species j in sampling unit i
p = total number of species.

Several Euclidean distance matrices were constructed using the spectral features and elevation data, i.e., the means and standard deviations of DN values, ratio of bands 4 and 5, and elevation in the pixel windows. Each variable was used on its own to construct a Euclidean distance matrix and additional Euclidean distance matrices were constructed using a combination of several variables. In the latter case, the combination of variables to be included was selected using multiple regression on distance matrices with backward elimination and the individual variables were standardized to give them equal weight before computing the distances.

In the multiple regression on distance matrices, the dependent matrix was the floristic distance matrix, and the independent matrices were the Euclidean distance matrices based on each spectral and elevation feature separately. Initially, all independent distance matrices were included. At each iteration of the backward elimination, the independent matrix with the least significant coefficient of partial determination was excluded, until the coefficients of all matrices remaining in the model were statistically significant at the $p<0.05$ level after Bonferroni correction. The statistical significances were estimated by Monte Carlo permutation, using either 999 permutations (in the case of 500-m sampling units) or 99 permutations (in the case of 200-m sampling units).

4. Discriminant analysis

Discriminant analysis was used to build a discriminant function, which consists of a linear combination of explanatory variables (in this case, spectral features and elevation) that best discriminates among a number of predefined classes (in this case, the forest types of Salovaara et al., 2004). The parameters of the model are estimated on the basis of a training data set consisting of field-verified sampling units with known class membership (the eight classified floristic transects). The resultant model can then be used to predict class membership for unvisited sampling units (the four transects lacking floristic data) and the entire study area (Legendre & Legendre, 1998).

The training data set consisted of the existing floristic classification of the 100-m long vegetation sampling units from the eight transects that were inventoried for pteridophytes (Salovaara et al., 2004). Because the size of the sampling units was different between the vegetation and image data, some adjustments were necessary. When a 500-m image sampling unit contained more than one forest type, it was allocated to the forest type to which most of the corresponding 100-m vegetation sampling units belonged. In the case of ties in the 200-m image sampling units, preference was given to the forest type that was rarer in the entire data set. Consequently, 2% of the 200-m sampling units and 9% of the 500-m sampling units were mixed pixel windows that contained more than one forest type. This can increase misclassification (Smith et al., 2003), so the percentage of misclassified sampling units was calculated

separately for the mixed and the non-mixed image sampling units.

The 500-m section closest to the river on transect 4 had been classified as intermediate tierra firme forest on the basis of floristic data (Salovaara et al., 2004), but this was clearly a misclassification, because the area was observed to be inundated in the rainy season. This error could distort the discriminant model and cause false misclassifications, especially because the number of sampling units belonging to the inundated forest class was low. Therefore, the transect section in question was transferred to the inundated forest class prior to the following analyses. The main results obtained with the uncorrected data are also reported for the sake of comparison.

A total of eleven spectral and elevation features were available for model building (Table 1), so the first step in the analysis was to select the combination of features that would give the best classification accuracy with the fewest features. This was done by a stepwise discriminant analysis using backward elimination. At each step, the feature that contributed least to the discriminatory power of the model was removed, until only one feature remained. Discriminatory power was measured with Wilks' Lambda, which is the spectral (and elevation) variance among the sampling units that is not explained by the model (Legendre & Legendre, 1998). Every step produced one feature combination that was validated by testing how well it classified the sampling units whose class memberships were already known. The classification accuracies obtained in this manner are unrealistically high because the same sampling units were used for both model building and testing, but their purpose was only to help in selecting the optimal model for each sampling unit size.

The feature combination with the highest classification accuracy was then selected; when two or more models performed equally well, the one with the fewest features was selected. Once the best feature combination had been found, the accuracy of the corresponding discriminant model was assessed by 8-fold cross-validation. Seven of the field-verified transects were used to build the discriminant model and the remaining transect was used as the test data set. Each of the field-verified transects was used as the test data set in turn and classification accuracy was derived as the percentage of correctly classified sampling units across all the eight runs. This method of cross validation was used because the ultimate purpose of the discriminant analysis was to classify the sampling units of the uninventoried transects and a realistic estimate of classification accuracy was needed for a situation where the nearest field-verified sampling units are several kilometres away.

Table 1
Selection of feature combinations and corresponding classification accuracy using the entire data set for building and validating the discriminant function

	200-m sampling units	500-m sampling units
Number of sampling units	317	127
Included features	m2, m4, m5, m7, sd2, sd4, sd5, sd7, m4/m5, elevation	m2, m4, m5, m7, sd2, sd4, sd5, sd7, m4/m5, elevation, sd of elevation
Selected feature combination	Elevation, m7, m4/m5, m5	m7, m4, m5, elevation, sd4, sd5, sd2, m2
Classification accuracy	73.8%	92.1%

Elevation was obtained from the SRTM DEM and DN values from a Landsat ETM+ satellite image. The numbers show which wavelength bands were used: m stands for mean and sd for the standard deviation of the DN values of the band in question. Values were extracted from pixel windows of 150 × 150 m for the 200-m sampling units and 360 × 450 m for the 500-m sampling units. Total degrees of freedom were $n-1$, within classes $n-3$ and between classes 2.

Classification results and accuracies were summarized in error matrices and by Kappa values. An error matrix shows the number of correctly classified image sampling units in the diagonal and the number of image sampling units that were erroneously either included in (lower triangle) or excluded from (upper triangle) each class. Producer's accuracy indicates the percentage of image sampling units that were assigned to the correct class. User's accuracy gives the percentage of image sampling units assigned to a certain class that actually belong to that class according to the floristic classification.

The kappa coefficient (κ) was calculated from the error matrix Eq. (2). Kappa indicates to what extent classification accuracy is due to true agreement of the field data and the classified data, and to what extent it could have been achieved by chance (Lillesand et al., 2004). Its value varies between 0 and 1 and a value of 0.5 means that there is 50% better agreement than expected by chance alone.

$$\kappa = \frac{N\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r}(x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} * x_{+i})}, \text{ where} \quad (2)$$

r = number of the rows/columns in the error matrix
x_{ii} = number of observations in the cell ii (row i and column i)
x_{i+} = marginal totals of row i
x_{+i} = marginal totals of column i
N = total number of observations.

Once the model building, cross validation and accuracy evaluation of the discriminant models were completed for the 200-m and 500-m image sampling units, the models were applied to obtain classifications of transects 1–3 and 9. Classification of the entire study area was only produced using the larger sampling unit size.

For the classification of the entire study area the band and elevation layers that produced the highest classification accuracy were filtered by a mean and/or standard deviation

kernel of 12 × 15 pixels and the pixel data were imported to the discriminant model. The data from the eight classified transects was then used as a training data set for predicting the forest classes for the filtered image pixels. The produced forest classification was filtered with a majority kernel of 5 × 5 pixels. Since the field sampling may not have covered all of the variation present in the study area, the pixels with values outside the sampled ranges of spectral and elevational features were removed from the classification. This was done by plotting the DN values of a pair of spectral features for all the pixel windows. A convex object (i.e., a polygon formed as if a string is stretched around the field-verified pixels) was then drawn around the field-verified pixels and all pixel windows that were outside the convex object were masked out from the image (see Thessler et al., in press). This was repeated for all pairs of spectral and elevational features. The excluded areas were presented as unclassified pixels in the produced map. Preprocessing of the Landsat image and GPS data was done using the Erdas Imagine 8.6 and ArcGIS programs. Resemblance matrices and Mantel tests were computed with Le Progiciel R (Casgrain & Legendre, 2001), and multiple regressions on distance matrices were run with Permute!. Both programs are available through the www site <http://www.bio.umontreal.ca/legendre/indexEnglish.html>. The discriminant analyses were run in the SAS 8.2 (SAS Institute Inc., 1999).

5. Results

5.1. Simplifying the vegetation classification

The original classification of Salovaara et al. (2004) produced four main forest classes separated by floristic characteristics: inundated forests, terrace forests, Pebas formation forests, and intermediate tierra firme forests. All the current analyses were first run using these four classes, but it turned out that Pebas formation forests and intermediate tierra firme forests were difficult to separate from each other on the basis of remotely sensed data. The overall classification accuracy for the four-class classification was 66% with the 500-m sampling units and even lower (49%) with the 200-m sampling units.

The intermediate forests covered 16% of the area sampled in the field and were mainly found in small patches surrounded by Pebas formation forest. As these two forest types were also floristically more similar to each other than to the other forest types, it seemed appropriate to combine them. All the analyses reported below are, therefore, based on a three-class classification:

1) inundated forests; including both river floodplain and swamp forests
2) terrace forests of tierra firme; non-inundated forests growing on relatively nutrient-poor loamy soils
3) Pebas formation/intermediate tierra firme forests; non-inundated forests growing on relatively nutrient-rich clayey to loamy soils.

5.2. Mantel tests

The results of the Mantel tests between the floristic distance matrices and the corresponding image feature distance matrices are shown in Table 2. Mantel correlations were calculated both using each image feature individually and using those feature combinations that were retained in the multiple regression on distance matrices after backward elimination. Of the individual spectral features, the mean DN values of band 7 showed the highest Mantel correlation with both sampling unit sizes. The mean DN values of the other bands also showed relatively high correlation coefficients, as did elevation and the ratio of bands 4 and 5. The correlation coefficients obtained with the 200-m sampling units were invariably, and usually considerably, lower than those obtained with the 500-m sampling units. Those distance matrices that combined several features resulted in the highest correlation coefficients: 0.47 and 0.30 with the 500 and 200-m sampling units, respectively.

5.3. Building of discriminant models

Table 1 presents the best feature combinations selected by stepwise discriminant analysis using backward elimi-

Table 2
Matrix correlations as measured by a Mantel test between floristic distances and distances based on remote sensing data (the DN values of a Landsat ETM+ satellite image and a digital elevation model) in NE Peruvian Amazonia

Spectral or elevation features	Mantel's *r*	
	500-m sampling units	200-m sampling units
m2	0.28***	0.11***
m4	0.25***	0.14***
m5	0.24***	0.14***
m7	0.33***	0.18***
sd2	0.12**	0.00
sd4	0.21***	0.05*
sd5	0.15**	0.11**
sd7	0.05	0.02
Ratio of bands 4 and 5	0.23***	0.17***
Elevation	0.26***	0.24***
sd of elevation	0.22***	–
m2, m4, m5, m7, elevation, sd of elevation	0.47***	–
m4, m7, sd5, elevation	–	0.30***

Floristic distances were based on the Bray–Curtis index. Spectral and elevational distances were based on the Euclidean distance computed from the mean (m) or standard deviation (sd) of the DN values of one or more ETM+ bands and/or elevation within each pixel window (150 × 150 m for the 200-m sampling units, 360 × 450 m for the 500-m sampling units).
* $P<0.05$.
** $P<0.01$.
*** $P<0.001$.

nation and the overall classification accuracy when these features were used to classify sampling units into floristically defined forest types. The same features were used in the 8-fold cross-validation. In the cross validation, classification accuracy varied considerably between sampling unit sizes and depending on which transect was used as the test data set (Table 3). For the 200-m sampling units, accuracy ranged between 53% and 95% (overall accuracy 71%) and for the 500-m sampling units between 63% and 100% (overall accuracy 85%). The Kappa coefficient was 0.35 in the classification of the 200-m sampling units and reached 0.62 in the classification of the 500-m sampling units.

The error matrix of the 200-m sampling units shows a low classification accuracy for the terrace forest class: producer's accuracy was 47% and user's accuracy 25% (Table 4). Terrace forest sampling units were often erroneously classified to the Pebas formation/intermediate tierra firme class. Pebas formation/intermediate tierra firme forest showed the highest producer's and user's accuracy values (74% and 88%, respectively) and inundated forests showed intermediate values (70% and 61%, respectively).

A higher classification accuracy was obtained when using the 500-m sampling units (Table 5). For the terrace forest class, producer's accuracy was 75% and user's accuracy 48%. Pebas formation/intermediate tierra firme forest again showed the highest producer's and user's accuracy values (88% and 95%, respectively) and inundated forests showed intermediate values (73% and 89%, respectively).

The total number of 500-m sampling units was 127 and 11 of these were mixed pixel windows containing two forest types. Of the 11 mixed sampling units, four (36%) were misclassified, whereas 15% of the 116 sampling units located in the forest class interior were misclassified. There were seven mixed plots among the 317 200-m sampling units, and 57% of these were misclassified, whereas 29% of the sampling units that included one forest type were misclassified.

When the 500-m section closest to the river in transect 4 was not transferred to the inundated forest class but was retained as Pebas formation/intermediate forest, the classification accuracies were consistently lower than with the corrected data set. In case of the 500-m sampling units, the selected features were the same as in the analyses described above, but classification accuracy was slightly lower. Overall accuracy decreased from 85% to 84%. Producer's accuracy of inundated forest decreased from 73% to 60%, whereas in terrace forest it was 81% and in intermediate/tierra firme forests 86%. For these same forest types the user's accuracies were 67%, 50% and 95%, respectively. With the uncorrected 200-m data the feature m4/m5 was replaced by m4 in the model and the overall accuracy was 68%, as compared to 71% with the corrected data. Thus, the corrected data were used also in the subsequent classifications of the entire study area and the four transects that lacked floristic data.

Table 3
Eight-fold crossvalidation of the discriminant models when using Landsat ETM+ DN values and SRTM DEM elevation features to classify sampling units to floristically defined forest types in NE Peruvian Amazonia

	200-m sampling units	500-m sampling units
Overall accuracy	70.7%	85.0%
Range	52.5–95.0%	62.5–100.0%
Standard deviation	16.4	14.5
Kappa coefficient	0.35	0.62

The selected feature combinations in Table 1 were used in the analyses. Each one of the eight floristic transects was used as the test data set in turn and the table summarizes the results for the eight runs. Corresponding error matrices are shown in Tables 4 and 5.

Table 4
Error matrix of the classification using spectral and elevation features to assign the 200-m sampling units ($N=317$) to floristically defined forest types in NE Peruvian Amazonia

Actual class	Predicted class				Producer's accuracy (%)
	In	Te	P/I	Row total	
Inundated forest	**23**	3	7	33	*69.7*
Terrace forest	2	**18**	18	38	*47.4*
Pebas formation/ intermediate tierra firme forest	13	50	**183**	246	*74.4*
Column total	38	71	208	317	
User's accuracy (%)	*60.5*	*25.4*	*88.0*		

The diagonal shows the number of correctly classified sample units for each class. Producer's accuracy is the percentage of the sampling units predicted to belong to the correct class, and user's accuracy is the percentage of the sampling units predicted to belong to a particular class that actually belong to that class. The best linear discriminant function based on the feature combination indicated in Table 1 was used. The error matrix was calculated by summing the classification results of eight separate runs where each of the transects was used as a test data set in turn.

5.4. Classification of four transects and the study area

Overall, in the classification of transects 1–3 and 9, which lacked floristic field data, the results for both sampling unit sizes agreed relatively well. Only 20% of the 200-m sampling units were assigned to a different forest

Table 5
Error matrix of the classification using spectral and elevation features to assign 500-m sampling units ($N=127$) to floristically defined forest types in NE Peruvian Amazonia

Actual class	Predicted class				Producer's accuracy (%)
	In	Te	P/I	Row total	
Inundated forest	**8**	2	1	11	*72.7*
Terrace forest	0	**12**	4	16	*75.0*
Pebas formation/ intermediate tierra firme forest	1	11	**88**	100	*88.0*
Column total	9	25	93	127	
User's accuracy (%)	*88.9*	*48.0*	*94.6*		

See Table 4 for more detailed explanation.

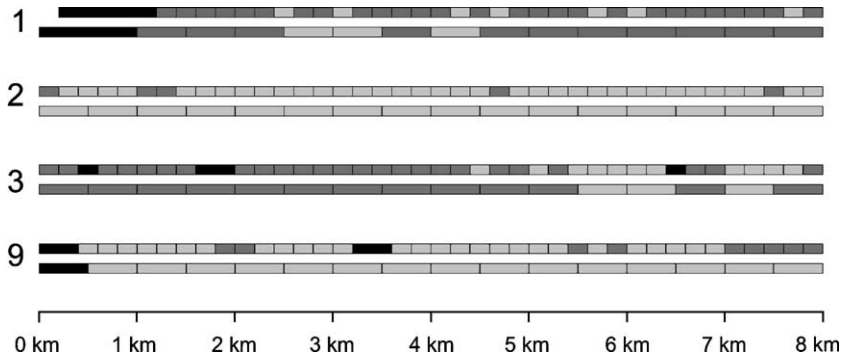

Fig. 2. Classification of the four transects lacking floristic field data into three vegetation types in NE Peruvian Amazonia. The results are shown separately for the 200- and 500-m sampling units. The classification is based on a linear discriminant model using the best combination of spectral and elevation features for each sampling unit size, as indicated in Table 1. The results for the 200-m sampling units are shown above those of the 500-m sampling units. The sampling units assigned to the three forest types are shown as follows: inundated forests in black, terrace forests in dark gray and Pebas formation/intermediate tierra firme forests in light gray. The first 200-m unit of transect 1 was covered by the river mask and was excluded from the analyses.

class than that of the corresponding 500-m unit (Fig. 2). In the case of these four transects, classification accuracy cannot be formally estimated because neither field data nor existing vegetation maps are available. However, the accuracy is probably close to that obtained in the model validation for the eight field-inventoried transects when each transect was excluded in turn. This yielded an overall classification accuracy of 85% and Kappa value of 0.62 for the 500-m sampling units.

Because less accurate results were obtained in all analyses for the 200-m sampling units than for the 500-m units, the latter sampling unit size can be considered preferable for use in future wildlife studies. In Fig. 1 the classified 500-m sampling units are placed on top of the Landsat ETM+ color composite. It is apparent that the obtained forest classification corresponds well to the visually observable patterns in the satellite image. Large swamp forests (in a reddish tone) were all correctly classified to inundated forest, as were river floodplain forests, even though the latter are less clearly visible in the image. An abrupt transition from darker green terrace forest to lighter green Pebas formation/intermediate tierra firme forest is visible in transects 3 and 11 and somewhat less clear in transects 1 and 7.

The forest classification map for the entire area is presented in Fig. 3. Generally, the classification results agree rather well with the patterns seen in the Landsat color composite in Fig. 1. The inundated forests are concentrated

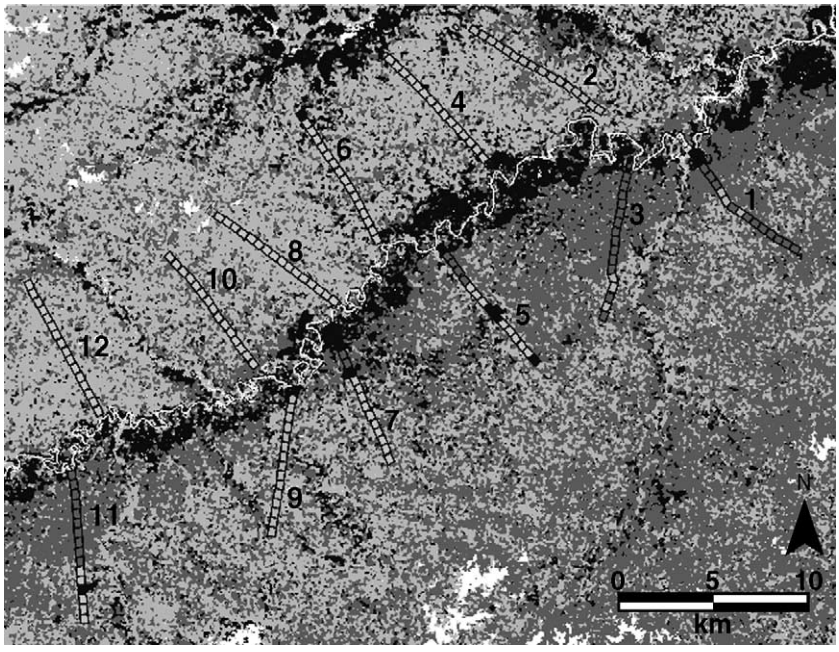

Fig. 3. Forest classification of the Landsat ETM+ satellite image shown in Fig. 1. The line transects whose forest classes were defined by floristic field data (transects 4–8 and 10–12) or by remotely-sensed data (transects 1–3 and 9) are overlaid. The three forest classes are shown as follows: inundated forest in black, terrace forest in dark gray and Pebas formation/intermediate forest in light gray. White areas are either open water or areas that were unclassified because their spectral and elevational features were outside the ranges covered by field sampling.

along the river and major creeks. The terrace forests are concentrated on the southern side of the river, whereas Pebas forests/intermediate tierra firme forests are more common on the opposite side of the river. Only 4% of the study area remained unclassified, because the values of the spectral and elevational features were outside the ranges covered by field sampling. These areas seem to correspond to the light-colored areas in the Landsat image (Fig. 1) and probably represent areas of open vegetation, e.g., large gaps produced by strong storms.

6. Discussion

Both the Mantel tests and the discriminant analyses gave clearly better results with the 500-m sampling units than with the 200-m units. Increasing the size of the sampling unit increases the risk that more than one vegetation types exist within it. It was indeed confirmed that the sampling units located at transition between two forest types were more often misclassified than those in the forest class interior. However, the classification error caused by this tendency was more than compensated for by the ability of the larger sampling units to dampen the effect of local noise (e.g., tree fall gaps). These results are well in line with the study of Hill & Foody (1994), who concluded that low-pass filtering increased the separability of Amazonian forest types.

The classification accuracy achieved in the present study was higher than we had expected, considering the fact that the forests were structurally rather similar, and that many previous studies have obtained low accuracies. For the 500-m sampling units the overall classification accuracy of the final discriminant model was 85% and the lowest producer's accuracy for an individual forest class was 73%. There are no generally accepted limits on how accurate a classification should be in order to qualify as reliable, but usually an overall accuracy exceeding 85% is considered reasonable, often with the additional criterion that the accuracy should not be lower than 70% for any class (Foody, 2002; Thomlinson et al., 1999; Smits et al., 1999). These recommendations are meant for general use in remote-sensing applications, which in most cases intend to separate structurally distinct classes, e.g., forest and open field, from each other.

The user's accuracy for the terrace forest class was only 48%, indicating that half of the sampling units assigned to that class from transects 1, 2, 3 and 9 (Fig. 2) could in fact belong to another forest type, most probably to the Pebas formation/intermediate class. There were, indeed, three 500-m sampling units in the middle part of transect 1 and two units in the latter part of transect 3 that based on field observations most closely resembled the Pebas formation/intermediate forest class. Nonetheless, they were assigned to the terrace forest class in the discriminant analysis. Except for these five sampling units, the classification of the four transects conformed well both to the general vegetation patterns observed in the field and to the patterns visible in the Landsat ETM+ image.

It is not straightforward to compare the present results to those of earlier studies because the numbers and definitions of the recognized vegetation classes are unique in each study. When more vegetation classes are recognized, the probability of erroneous class assignments increases. The present study only included three forest classes. On the other hand, vegetation classes that are structurally clearly different are intrinsically easier to separate than structurally similar ones. All of the forest types in this study were structurally quite homogeneous, closed-canopy rain forest, even though there is some structural heterogeneity within the inundated forests. The spatial scale and sampling design also vary among studies, but unfortunately information on sampling design or measures of accuracy are not always provided, making comparisons with other studies difficult. With these reservations in mind, some comparisons can be attempted.

Few other studies have tried to separate floristically defined forest types within Amazonian primary tierra firme forests. The only other studies that employed forest classes comparable to ours (tierra firme on nutrient-poor clay soils vs. moist sand or sandy clay soils) concluded that these forest types could not be separated from Landsat TM images (Foody & Hill, 1996; Hill, 1999; Hill & Foody, 1994). However, tierra firme forests on sand or sandy clay could often be identified visually (Foody & Hill, 1996), which was also the case for terrace forests and Pebas formation/intermediate tierra firme forests in our study. Foody & Cutler (2003) achieved a high classification accuracy (96%) for nine floristically defined forest classes in lowland dipterocarp forests in Borneo when using Landsat TM data. However, parts of the study area were heavily impacted by logging, which could cause the classes to be structurally and not just floristically dissimilar. Furthermore, no independent test data was used in accuracy assessment.

Separation of Amazonian inundated and tierra firme forests has been very successful. Lobo & Gullison (1998) were able to classify seasonally flooded lowland forests and seasonal evergreen dense forests 100% correctly using a Landsat TM satellite image in Bolivia and reached an overall classification accuracy of 94–98% with eight land cover types (water, swamps and several savanna and tropical forest classes). Hess et al. (2003) mapped inundated and non-inundated forests with 63–90% producer's and user's accuracies in the central Amazon basin from dual-season radar imagery.

Successional stages of moist tropical forests have also been classified quite accurately. Lu et al. (2003) achieved an overall accuracy of 78% (range of producer's accuracy 58–99%) when classifying three successional and one mature forest class in Brazilian Amazonia. Thenkabail et al. (2004) discriminated fallows and successional stages of tall forest (eight classes in total) in Cameroon with an accuracy of 96% using hyperspectral Hyperion images, but classifica-

tion accuracy was only 42% when a Landsat ETM+ image was used. Unfortunately their validation method was not reported. General classes of non-forest, mature forest and regenerating forest in Cameroon were discriminated with a 64–79% accuracy using low resolution NOAA AVHRR data (Lucas et al., 2000).

Even higher accuracies have been achieved in tropical areas when non-forest land cover types have been included. Vieira et al. (2003) reported an overall accuracy of 81% (range 60–100%) in a classification of seven secondary forest, cropland and pasture classes using a Landsat ETM+ -satellite image in Brazilian Amazonia. Trisurat et al. (2000) used supervised classification of Landsat TM data to discriminate grassland and six forest classes (e.g., mixed deciduous, dry evergreen and tropical rain forest) with an overall accuracy of 79% (range 50–100%) in Thailand.

Although this is not an exhaustive review of previous classification studies in tropical forests, these results illustrate the point that classification accuracy in the present study can be considered satisfactory, given that two of the classified forest types were non-inundated primary forests that lacked any apparent structural dissimilarities.

6.1. Applications

The methods used in the present study allowed the separation of floristically defined tierra firme forest classes from Landsat ETM+ images with a reasonable accuracy. This type of habitat data can offer valuable information for sustainable resource use and biodiversity conservation, where vegetation types can be used as surrogates for modeling the distributions of species and communities (Ferrier, 2002; Foody, 2003; Kerr & Ostrovsky, 2003). It is especially useful in complex tropical rain forests, where the distribution patterns of individual species are poorly known and other vegetation mapping methods are rarely applicable.

The tierra firme classes in the present study differ in soil fertility (see Salovaara et al., 2004), and soils can also influence herbivore abundances by a number of mechanisms. For example, plants may invest more in defense against herbivores when growing on poor soils (Coley et al., 1985; Fine et al., 2004; Janzen, 1974), and vegetation productivity may also be lower on poorer soils (Peres, 2000). The results of the present study together with the mammal census data from the same area will make it possible to test this hypothesis in the future by comparing population densities of large-bodied herbivorous mammals between forest types.

For the purposes of such wildlife habitat studies the classification accuracy (85%) and spatial resolution (500 m) of the present study can be considered sufficient. Large-bodied mammals generally have large home ranges and consequently small-scale habitat patterns are less relevant for them than for species that move in less extensive areas. Furthermore, the boundaries between western Amazonian biotopes are generally gradual rather than abrupt (Tuomisto et al., 1995). Thus, defining exact locations for the boundaries is inevitably somewhat arbitrary.

The classification presented here is rudimentary in the sense that only two tierra firme types were separated, even though both are internally heterogeneous (Salovaara et al., 2004). However, the two recognized tierra firme types are floristically clearly distinct, and apparently represent the two extremes of the soil fertility gradient in the study area (Salovaara et al., 2004). The intermediate tierra firme forest, which was lumped with Pebas formation forest here, was found in relatively small patches in the study area. Its recognition would probably be of only minor relevance for the habitat use of wide-ranging mammals. In any case, the classification presented here is more advanced than the generalized inundated forest vs. tierra firme dichotomy that has been frequently applied in ecological studies in Amazonia.

Studies using comparable methods to ours in the Yasuní area in Ecuador (Tuomisto et al., 2003a) and in the Sucusari area in NE Peru (Tuomisto et al., 2003b) yielded even higher Mantel correlations between the spectral and floristic distance matrices than the present study did. This suggests that a shorter ecological gradient was covered in the present study than in the two earlier studies. When the difference between the poorest and the richest soils included in a study is smaller, the floristic differences are also smaller, which can be expected to lead to smaller differences in canopy reflectance. Nevertheless, positive and significant Mantel correlations in all three areas indicate that the congruence between the spectral and floristic patterns is a recurrent phenomenon in different parts of Western Amazonia.

Acknowledgemets

We thank Glenda Cárdenas for collecting the floristic data, Tuuli Toivonen for help with the initial part of the image analysis, and Kalle Ruokolainen for constructive comments on the manuscript. We are also grateful to four anonymous reviewers whose comments improved the manuscript. The study was funded by the Academy of Finland, Helsingin Sanomain 100-vuotissäätiö and Ella and Georg Ehrnrooth's Foundation. Riffat Malik's research fellowship at the Department of Geography, University of Turku, was funded by the Finnish Centre for International Mobility (CIMO). The research permits for the field work were kindly granted by the Peruvian Instituto Nacional de Recursos Naturales (INRENA), Lima.

References

Achard, F., Eva, H., & Mayaux, P. (2001). Tropical forest mapping from coarse spatial resolution data: Production and accuracy assessment issues. *International Journal of Remote Sensing*, *22*, 2741–2762.

Asner, G. P. (2001). Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, *22*, 3855–3862.

Cajander, A. K. (1926). The theory of forest types. *Acta Forestalia Fennica*, *29*, 1–108.

Casgrain, P., & Legendre, P. (2001). *The R package for multivariate and spatial analysis, version 4.0 d5*. Montreal: Département de sciences biologiques, Université de Montréal, http://www.bio.umontreal.ca/legendre/indexEnglish.html (Available on the WWWeb site).

Clark, D. B., Clark, D. A., & Read, J. M. (1998). Edaphic variation and the mesoscale distribution of tree species in a neotropical rain forest. *Journal of Ecology*, *86*, 101–112.

Coley, P. D., Bryant, J. P., & Chapin, F. S. (1985). Resource availability and plant antiherbivore defense. *Science*, *230*, 895–899.

Condit, R. (1996). Defining and mapping vegetation types in mega-diverse tropical forests. *Trends in Ecology and Evolution*, *11*, 4–5.

de Grandi, G. F., Mayaux, P., Malingreau, J. P., Rosenqvist, Å., Saatchi, S., & Simard, M. (2000). New perspectives on global ecosystems from wide-area radar mosaics: Flooded forest mapping in the tropics. *International Journal of Remote Sensing*, *21*, 1235–1249.

Duivenvoorden, J. F. (1995). Tree species composition and rain forest-environment relationships in the midlle Caqueta area, Colombia, NW Amazonia. *Vegetatio*, *120*, 91–113.

Duque, A., Sánchez, M., Cavelier, J., & Duivenvoorden, J. F. (2002). Different floristic patterns of woody understorey and canopy plants in Colombian Amazonia. *Journal Tropical Ecology*, *18*, 499–525.

Ellenberg, H. (1988). *Vegetation ecology of Central Europe*. Cambridge: Cambridge University Press.

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, *31*, 331–363.

Fine, P. V. A., Mesones, I., & Coley, P. D. (2004). Herbivores promote habitat specialization by trees in Amazonian forests. *Science*, *305*, 663–665.

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

Foody, G. M. (2003). Remote sensing of tropical forest environments: Towards the monitoring of environmental resources for sustainable development. *International Journal of Remote Sensing*, *24*, 4035–4046.

Foody, G. M., & Cutler, M. E. J. (2003). Tree biodiversity in protected and logged Bornean tropical rain forests and its measurement by satellite remote sensing. *Journal of Biogeography*, *30*, 1053–1066.

Foody, G. M., & Hill, R. A. (1996). Classification of tropical forest classes from Landsat TM data. *International Journal of Remote Sensing*, *17*, 2353–2367.

Gentry, A. H. (1988). Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Annals of the Missouri Botanical Garden*, *75*, 1–34.

Hess, L. L., Melack, J. M., Novo, E.M.L.M., Barbosa, C. C. F., & Gastil, M. (2003). Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote Sensing of Environment*, *87*, 404–428.

Hill, R. A. (1999). Image segmentation for humid tropical forest classification in Landsat TM data. *International Journal of Remote Sensing*, *20*, 1039–1044.

Hill, R. A., & Foody, G. M. (1994). Separability of tropical rain-forest types in the Tambopata–Candamo reserved zone, Peru. *International Journal of Remote Sensing*, *15*, 2687–2693.

IFG. (1984). *Mapa planimétrico de imágenes de satélite 1:250,000. Perú*. Neu Isenburg: Institute for Applied Geosciences.

Janzen, D. H. (1974). Tropical blackwater rivers, animals, and mast fruiting by the Dipterocarpaceae. *Biotropica*, *6*, 69–103.

Kalliola, R., Salo, J., Puhakka, M., Rajasilta, M., Häme, T., Neller, R. J., et al. (1992). Upper Amazon channel migration: Implications for vegetation perturbance and succession using bitemporal Landsat MSS images. *Naturwissenschaften*, *79*, 75–79.

Kerr, J. T., & Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing. *Trends in Ecology and Evolution*, *18*, 299–305.

Kubitzki, K. (1987). The ecogeographical differentiation of Amazonian inundation forests. *Plant Systematics and Evolution*, *162*, 285–304.

Laporte, N., Justice, C., & Kendall, J. (1995). Mapping the dense humid forest of Cameroon and Zaire using AVHRR satellite data. *International Journal of Remote Sensing*, *16*, 1127–1145.

Legendre, P., & Legendre, L. (1998). *Numerical ecology* (2nd English ed.). Amsterdam: Elsevier Science BV.

Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2004). *Remote sensing and image interpretation* (5th ed.). USA: John Wiley and Sons Inc.

Lobo, A., & Gullison, E. (1998). Mapping the tropical landscape of Beni (Bolivia) from Landsat TM imagery: Beyond the "forest/non-forest" legend. In F. Dalmeier, & J. Comiskey (Eds.), *Forest Biodiversity Research, Monitoring and Modelling* (pp. 159–181). Paris: UNESCO.

Lu, D., Moran, E., & Batistella, M. (2003). Linear mixture model applied to Amazonian vegetation classification. *Remote Sensing of Environment*, *87*, 456–469.

Lucas, R. M., Honzak, M., Curran, P. J., Foody, G. M., & Ngueles, D. T. (2000). Characterizing tropical forest regeneration in Cameroon using NOAA AVHRR data. *International Journal of Remote Sensing*, *21*, 2831–2854.

Marengo, J. A. (1998). Climatología de la zona de Iquítos, Perú. In R. Kalliola, & S. Flores Paitan (Eds.), Geoecología y desarrollo amazónico - Estudio integrado en la zona de Iquítos, Perú. *Annales Universitatis Turkuensis Ser A II*, *114* (pp. 35–57). Sulkava: University of Turku.

Moran, E. F., & Brondizio, E. (1994). Integrating Amazonian vegetation, land-use, and satellite data. *Bioscience*, *44*, 329–338.

Nagendra, H. (2001). Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, *22*, 2377–2400.

Peres, C. A. (2000). Effects of subsistence hunting on vertebrate community structure in Amazonian forests. *Conservation Biology*, *14*, 240–253.

Phillips, O. L., Núñez Vargas, P., Lorenzo Monteagudo, A., Peña Cruz, A., Chuspe Zans, M. E., Galiano Sánchez, W., et al. (2003). Habitat association among Amazonian tree species: A landscape-scale approach. *Journal of Ecology*, *91*, 757–775.

Pitman, N., Vriesendorp, C., Moskovits, D. (Eds.) (2003). *Perú: Yavarí. Rapid biological inventories report 11*. Chicago: The Field Museum.

Podest, E., & Saatchi, S. (2002). Application of multiscale texture in classifying JERS-1 radar data over tropical vegetation. *International Journal of Remote Sensing*, *23*, 1487–1506.

Powell, R. L., Matzke, N., de Souza Jr., C., Clark, M., Numata, I., Hess, L. L., et al. (2004). Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sensing of Environment*, *90*, 221–234.

Prance, G. T. (1979). Notes on the vegetation of Amazonía: III. The terminology of Amazonian forest types subject to inundation. *Brittonia*, *31*, 26–38.

Rajaniemi, S., Tomppo, E., Ruokolainen, K., & Tuomisto, H. (2005). Estimating and mapping pteridophyte and Melastomataceae species richness in Western Amazonian rain forests. *International Journal of Remote Sensing*, *26*, 475–493.

Ruokolainen, K., Linna, A., & Tuomisto, H. (1997). Use of Melastomataceae and pteridophytes for revealing phytogeographic patterns in Amazonian rain forests. *Journal of Tropical Ecology*, *13*, 243–256.

Ruokolainen, K., & Tuomisto, H. (1998). Vegetación natural de la zona de Iquitos. In R. Kalliola, & S. Flores Paitan (Eds.), *Geoecología y desarrollo amazónico - Estudio integrado en la zona de Iquítos, Perú, Annales Universitatis Turkuensis Ser A II*, *114* (pp. 253–365). Sulkava: University of Turku.

Salo, J., Kalliola, R., Häkkinen, I., Mäkinen, Y., Niemelä, P., Puhakka, M., et al. (1986). River dynamics and the diversity of Amazon lowland forest. *Nature*, *322*, 254–258.

Salovaara, K. J., Cárdenas, G. G., & Tuomisto, H. (2004). Forest classification in an Amazonian rainforest landscape using pteridophytes as indicator species. *Ecography*, *27*, 689–700.

Sanchez, A., Chira, J., Romero, D., De la Cruz, J., Herrera, I., Cervante, J., et al. (1999). *INGEMMET Serie A: Carta Geológico Nacional 1:100000, Boletin no. 132*. Instituto Geológico Minero y Metalúrgico, Peru.

SAS Institute Inc. (1999). *SAS System version 8.2*.

Shepard Jr., G. H., Yu, D. W., & Nelson, B. W. (2004). Ethnobotanical ground-truthing and forest diversity in the western Amazon. *Advances in Economic Botany*, *15*, 133–171.

Singh, A. (1987). Spectral separability of tropical forest cover classes. *International Journal of Remote Sensing*, *8*, 971–979.

Skole, D., & Tucker, C. (1993). Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978 to 1988. *Science*, *260*, 1905–1910.

Smith, J. H., Stehman, S. V., Wickham, J. D., & Yang, L. (2003). Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment*, *84*, 342–349.

Smits, P. C., Dellepiane, S. G., & Schowengerdt, R. A. (1999). Quality assessment of image classification algorithms for land-cover mapping: A review and a proposal for a cost-based approach. *International Journal of Remote Sensing*, *20*, 1461–1486.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using Landsat TM data: When and how to correct atmospheric effects. *Remote Sensing of Environment*, *75*, 230–244.

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., Legg, C., & De Dieu, M. J. (2004). Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests. *Remote Sensing of Environment*, *90*, 23–43.

Thessler, S., Ruokolainen, K., Tuomisto, H. & Tomppo, E. (in press). Mapping gradual landscape-scale floristic changes in Amazonian primary rain forest by combining ordination and remote sensing. *Global Ecology and Biogeography.*

Thomlinson, J. R., Bolstad, P. V., & Cohen, W. B. (1999). Coordinating methodologies for scaling landcover classifications from site-specific to global: Steps toward validating global map products. *Remote Sensing of Environment*, *70*, 16–28.

Trisurat, Y., Eiumnoh, A., Murai, S., Hussain, M. Z., & Shrestha, R. P. (2000). Improvement of tropical vegetation mapping using remote sensing technique: A case of Khao Yai National Park, Thailand. *International Journal of Remote Sensing*, *21*, 2031–2042.

Tuomisto, H. (1998). What satellite imagery and large-scale field studies can tell about biodiversity patterns in Amazonian forests. *Annals of the Missouri Botanical Garden*, *85*, 48–62.

Tuomisto, H., Linna, A., & Kalliola, R. (1994). Use of digitally processed satellite images in studies of tropical rain forest vegetation. *International Journal of Remote Sensing*, *15*, 1595–1610.

Tuomisto, H., Ruokolainen, K., Kalliola, R., Linna, A., Danjoy, W., & Rodriguez, Z. (1995). Dissecting Amazonian biodiversity. *Science*, *269*, 63–66.

Tuomisto, H., Poulsen, A.D, Ruokolainen, K., Moran, R. C., Quintana, C., Celi, J., et al. (2003a). Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. *Ecological Applications*, *13*, 352–371.

Tuomisto, H., Ruokolainen, K., Aguilar, M., & Sarmiento, A. (2003b). Floristic patterns along a 43-km long transect in an Amazonian rain forest. *Journal of Ecology*, *91*, 743–756.

Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003c). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, *299*, 241–244.

Vieira, I. C. G., de Almeida, A. S., Davidson, E. A., Stone, T. A., de Carvalho, C. J. R., & Guerrero, J. B. (2003). Classifying successional forests using Landsat spectral properties and ecological characteristics in Eastern Amazônia. *Remote Sensing of Environment*, *87*, 470–481.

Vormisto, J., Phillips, O. L., Ruokolainen, K., Tuomisto, H., & Vásquez, R. (2000). A comparison of fine-scale distribution patterns of four plant groups in an Amazonian rainforest. *Ecography*, *23*, 349–359.