Dissertationes Forestales 59

Remote sensing of floristic patterns in the lowland rain forest landscape

Sirpa Thessler

Department of Geography
Faculty of Mathematics and Natural Sciences
University of Turku

Academic dissertation

To be presented, with the permission of the Faculty of Mathematics and Natural Sciences of the University of Turku, for public criticism in the Tauno Nurmela Auditorium on March 20th 2008 at 12.00 o'clock noon.

ABSTRACT

Land use and conservation planning urgently need information on floristic variation over large rain forest areas. Floristic variation can not be inventoried in every location and of all the flora, thus inventory is limited in sample sites of a group(s) of indicator species and modelled to predict the floristic composition of non-inventoried sites using spatially continuous information on the environment. Modelling is, however, practicable only if the dimensions of species data can be drastically reduced to a surrogate of floristic composition. The aim was to explore whether remote sensing can be applied to study and map the spatial variation of surrogates in lowland old-growth rain forest.

I studied three surrogates: 1) number of species in ecological categories, 2) vegetation / forest type classification, and 3) species composition, summarized as the scores of three ordination axes. The understorey Melastomataceae and pteridophytes, and tree and palm species were used as indicator species. Landsat TM or ETM+ - satellite images and the SRTM digital elevation model were used as a proxy of environmental variation. The prediction methods included a k nearest neighbour method and linear discriminant analysis. The study areas were located in eastern Ecuador, in north-eastern Peru and northern Costa Rica.

The main finding was that floristic patterns in lowland rain forest, expressed as vegetation classes, ordination axis scores or the number of species in ecological categories, can be predicted on the basis of remotely sensed data and field observations. The accuracy of the predictions depended on feature selection and weighting and on spatial resolution. The k-nn method proved to be a promising method in predicting floristic variation when it was expressed as a continuous variable, such as ordination axis scores or number of species. It also performed better than linear discriminant analysis in distinguishing forest classes using satellite image data.

Keywords: discriminant analysis, surrogate of floristic variation, forest classification, k nearest neighbours, satellite image, tropical forest

ACKNOWLEDGEMENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is a summary of the following articles, referred to in the text by their Roman numerals.

I Rajaniemi, S., Tomppo, E., Ruokolainen, K. & Tuomisto, H. 2005. Estimating and mapping pteridophytes and Melastomataceae species richness in Western Amazonian rain forest. International Journal of Remote Sensing 26: 475-494.

II Thessler, S., Ruokolainen, K., Tuomisto, H. & Tomppo, E. 2005. Mapping gradual landscape-scale floristic changes in Amazonian primary rain forest by combining ordination and remote sensing. Global Ecology and Biogeography 14: 315-325.

III Salovaara, K., Thessler, S., Malik, R. & Tuomisto, H. 2005. Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. Remote Sensing of Environment 97: 39-51.

IV Thessler, S., Sesnie, S., Ruokolainen, K., Tomppo, E., Ramos Bendaña, Z. S. & Finegan, B. in press. Detecting floristically defined rain forest types with Landsat TM imagery and tree species data in Northern Costa Rica. Remote Sensing of Environment.

TABLE OF CONTENTS

INTRODUCTION

Deforestation is advancing in tropical rain forest areas and is causing growing concern over biodiversity loss in these extremely species-rich forests. Resources for species conservation are limited and therefore it would be important to be able to plan the conservation efforts to be as efficient as possible. One basic requirement for efficient conservation planning is knowledge about the spatial distribution of species and species communities (Griffiths et al. 2000, Margules & Pressey 2000, Myers et al. 2000, Kerr & Ostrovsky 2003, Schulman et al. 2007).

The standard procedure for acquiring such knowledge is the establishment of field plots for quantitatively inventorying the vascular plant flora. An approach based on full species inventories, however, is not feasible in practically any tropical forest areas, as they are difficult to access and have a high number of taxonomically poorly known species. Floristic inventories in tropical rain forests have therefore been restricted to just part of the flora, most commonly trees (self-supporting woody plants above a certain diameter limit, often 10 cm at breast height).

Investigators from the Amazon Research Team at the University of Turku, Finland, have advocated the use of taxonomically defined groups of understorey plants, specifically terrestrial pteridophytes and the predominantly shrub family Melastomataceae, as indicator species for more general floristic patterns (Ruokolainen et al. 1997). The species distributions of both trees and the suggested indicator groups appear to react rather similarly to a number of soil characteristics (Ruokolainen et al. 1997, Ruokolainen & Tuomisto 1998, Ruokolainen et al. 2007). This suggests that the plant species of tropical rain forests form similar kinds of species associations as they have been found to do in temperate and boreal forests. Accordingly, the use of just part of the flora as an indicator of more general floristic patterns seems justified. Higgins & Ruokolainen (2004) have in fact reported that taxonomically defined parts of the flora may serve better as floristic indicators than structurally defined parts, such as trees exceeding a certain diameter.

Whichever approach one adopts in restricting the work of floristic inventory, it is clear that one cannot inventory the whole area of interest in any relevant exercise of conservation planning. What is needed are methods for predicting the species composition for non-inventoried sites. Such predictions can hardly proceed by modelling the occurrence of individual species. Rain-forest plant communities are simply too species-rich in relation to the realistically achievable number of inventory plots (Guisan & Zimmermann 2000, Ferrier 2002). Even if the modelling effort were not limited by the quantity of data, the resulting map layers of predicted distributions for tens or even hundreds of species would be difficult to deal with without some post-modelling classification or ordination.

At least in tropical rain forests, it thus appears promising to follow the suggestion of Ferrier (2002), who recommended predicting the properties of communities rather than the distributions of individual species. Variation in communities has been modelled by classifying sampling sites into categories with similar species composition (e.g. vegetation classes) or by dividing species into groups with similar habitat requirements, and by predicting their distributions. The site categories are assumed to be internally homogeneous in their species composition, and one category is assigned to one location. Thus the modelling provides a single map layer showing the distribution of the categories. The species groups are modelled as the proportion or number of species belonging to the group in a given location; thus while members of several species groups can occur in a single

location. The modelling of the species groups then provides a series of map layers (Ferrier et al. 2002).

Variation in species composition can also be described in terms of ordination axis scores (Öhmann & Gregory 2002). Ordination is a set of dimension-reducing methods commonly used in ecology. It arranges sample sites along a few ordination axes in such a way that floristically similar sites are located close together in ordination space (McCune & Grace 2002). The scores of each axis can then be predicted for non-visited sites on the basis of environmental data (Öhmann & Gregory 2002). A relatively similar method is to model and predict floristic distances calculated in a site-by-site matrix. By knowing the environmental characteristics of all the grid cells in the study area, the floristic distances of the field-sampled grid cells can be predicted across the study area. Floristic distances cannot be mapped directly but may be ordinated or clustered for mapping purposes (Ferrier et al. 2007). The floristic patterns expressed by ordination axis scores or by floristic distances in a site-by-site matrix are both continuous variables; they show which places are similar in species composition and which ones differ, but do not indicate what species are present (Ferrier 2002, Tuomisto & Ruokolainen 2006).

The number of species and a variety of diversity indices (e.g. Shannon's diversity index) are commonly computed to describe the diversity of a single community sample (α-diversity, Whittaker 1972). The problem with α-diversity is that it tells nothing of the identity of the species which make up the diversity value. In conservation, species are the object of concern; thus an index which is not sensitive to species identity can have only limited value. The spatial overlap of areas with a high species richness in different plant families is also often low (Williams-Linera et al. 2005, Tchouto et al. 2006, but see Villasenor et al. 2007). Knowledge of α-diversity has therefore relatively little value in conservation and/or land use planning. However, species richness may serve as a surrogate of floristic variation if it is calculated for ecologically defined species groups. Changes in the number of species in ecological categories reflect changes in the environment, and thus indirectly indicate variation in species composition (Faith & Walker 1996a, 1996b).

Vegetation classes, ecologically defined species groups and ordination axis scores summarize the multidimensional, complex data of species communities in terms of one or few variables, which can be called surrogates of floristic variation. The surrogates may be predicted for non-visited locations by combining field data on species and spatially continuous information on the environment, such as topography, climate, soils, land-cover type or spectral characteristics of satellite images (Nagendra & Gadgil 1999, Guisan & Zimmermann 2000, Ferrier 2002, Ferrier & Guisan 2006). The relevant question is then to determine which parts of this environmental information correlate with (plant) species distributions in such a way that successful predictive models can be made (Ferrier & Watson 1997).

In tropical rain forest areas, data on soils and climate do not appear particularly promising for the purpose of predicting the distribution of species or communities. Field observations on soil and climate are often limited to relatively few locations, and soil and climate values for large areas are thus interpolated on the basis of rather distantly placed field observations. Also, variation in climatic variables within lowland rain forest areas may well be too low for any relevant local scale predictions. Thus remotely sensed data appear to represent the best, if not the only, source of spatially continuous environmental information with sufficient spatial accuracy and coverage (Guisan & Zimmermann 2000).

However, satellite images can be used as predictors of floristic variation only if it can be related statistically to the variation in pixel values of satellite images. Such a link has been

found even in structurally relatively homogeneous lowland tropical forests (Thenkabail et al. 2003, Tuomisto et al. 2003a, 2003b). Satellite images, aerial photographs and derived digital elevation models (DEM) have also increasingly been used to predict tropical plant communities (Ter Steege et al. 2000, Funk et al. 2005). Clearly different forest categories, such as forest in white sand soil, flooded forests and non-inundated *tierra firme* forests in Amazonia, have been observed by remote sensing (Kalliola et al. 1991, Tuomisto et al. 1994, Ruokolainen et al. 1998, De Grandi et al. 2000), but early remote sensing studies were unsuccessful in distinguishing floristic differences in lowland areas within these large main forest categories (Hill & Foody 1994, Paradella et al. 1994, Foody & Hill 1996, Hill 1999). Several recent studies, however, have reported relatively high classification accuracies for rain forest types at landscape scale (Thenkabail et al. 2003, 2004, Foody & Cutler 2006, Sesnie 2007).

In my thesis work I applied three surrogates for multidimensional data of species composition in order to predict spatial floristic variation in tropical lowland rain forests by integrating remote sensing and field observations. The surrogates were as follows: 1) species richness of ecological categories (I); 2) species composition described by three-dimensional ordination solution (II); and 3) forest types defined by species composition, either alone or in combination with successional status (III and IV). The predictions were based on Landsat Thematic Mapper (TM) or Enhanced Thematic Mapper (ETM+) satellite images and Shuttle Radar Topography Mission (SRTM) digital elevation model. The predicted floristic variation concerned species composition of understorey pteridophytes and Melastomataceae (I, II, III) or tree and palm species composition and successional status (IV). I tested the prediction success of k nearest neighbour (k-nn) method and discriminant analysis in remote sensing based classification, and k-nn in prediction of continuous variables. I searched for appropriate spatial scale, feature combinations and procedures for reliable accuracy assessment in data-poor environments.

The general aim was to determine whether remote sensing can be applied in the detection and mapping of spatial variation of floristic patterns in lowland rain forests for purposes of conservation and land use planning. The detailed research questions were:

1) Can species composition, floristically defined vegetation classes and species richness of ecological categories be predicted in tropical rain forests based on field and remotely sensed data (I-IV)?

2) Can understorey indicator species be employed in remote sensing of floristic patterns (I-III)?

3) Can knn method predict and classify floristic patterns of rain forests accurately (I, II and IV) and does image segmentation and feature extraction from the segments increase classification accuracy in a fragmented landscape (IV)?

PREDICTING FLORISTIC VARIATION IN RAIN FORESTS

Challenges in predicting floristic variation

The use of satellite image data for predicting rain-forest floristic variation is complicated by two main factors. First of all, there is no consensus as to the main forces that control the distribution of species and species communities in tropical rain forests. Secondly, there are several technical problems with satellite images.

Different hypotheses have been presented as to the main driving forces of species turnover at landscape scale (some tens of square kilometres). Hubbell's so-called neutral hypothesis (Hubbell 2001) maintains that species composition, at least within relatively homogeneous habitats, is random and controlled by dispersal. Since the dispersal of species is spatially limited, floristic patterns must be spatially autocorrelated. In several studies such a spatial autocorrelation has indeed been found, supporting the neutral hypothesis (Hubbell 2001, Oliveira & Nelson 2001, Condit et al. 2002, Valencia et al. 2004). Accordingly, the patterns of satellite images have also been suggested to reflect forest dynamics (disturbances, pest outbreaks, limited dispersal) rather than edaphically determined floristic patterns (Condit 1996).

Others have found neither such spatial autocorrelation nor any indication of environmental control over species distribution patterns; they have consequently suggested that at least Neotropical rain forests are more or less uniform in plant species composition over large areas, with the same relatively few dominant tree species (Duivenvoorden 1996, Pitman et al. 1999, 2001, Duivenvoorden et al. 2002). If these views are valid, predicting floristic composition and mapping its variation would be close to useless. Species composition would either be under constant and unpredictable change (Hubbell 2001), or it would be identical over very large areas (Pitman 2001).

Clear environmental discontinuities have naturally been recognised. The distinction between forests in white sand soil, flooded forests and non-inundated *tierra firme* forests has long been recognised. It is only quite recently, however, that environmental and floristic differences within these main categories have been reported (Tuomisto & Poulsen 1996, Ruokolainen et al. 1997, 1998, Clark et al. 1999, Potts et al. 2002, Tuomisto et al. 2002, Phillips et al. 2003, Tuomisto et al. 2003a, Paoli et al. 2006, Ruokolainen et al. 2007). Patterns of tropical tree and understorey species composition have also been found to correlate with patterns of chemical and physical characteristics of soil and topography (Ter Steege et al. 1993, Clark et al. 1995, 1998, Vormisto et al. 2000, Duque et al. 2002, Potts et al. 2002, Tuomisto et al. 2003a, 2003b, Paoli et al. 2006, Ruokolainen et al. 2007). These studies support the hypothesis that floristic patterns are at least to some extent determined by environmental factors. This in turn means that floristic patterns can be predicted for unvisited sites by knowing the spatial variation in environmental characteristics.

The poor availability of satellite images for tropical areas is a common and familiar problem. Most tropical rain forest images are plagued by clouds and haze, and in some areas there may be time gaps of several years in the availability of high-quality images. Landsat TM satellite images also often suffer from systematic scan-line noise (Crippen 1989). Striping is caused by the different responses among the sixteen detectors in the satellite sensor, while banding results from differences between adjacent forward and reverse scan lines of all sixteen detectors (Lillesand & Kiefer 2000). Quite recently a systematic across-path east-west radiometric gradient has been found to be common within single Landsat TM scenes (Collett et al. 1998, Toivonen et al. 2002, 2006). The gradient has been explained as resulting from the effects of topography (shadows), anisotropic reflectance and atmospheric scattering. In lower latitudes the scanning line is often close to the solar azimuth angle, which seems to strengthen the phenomenon. The gradient was present in all bands and in almost all of the 49 images analysed, but was stronger in visible light bands than in infrared bands (Toivonen et al. 2006). Both scan-line noise and the artificial radiometric gradient become apparent in large and relatively uniform tropical

forest areas, where images need strong stretching to allow visualisation of vegetation patterns (Crippen 1989, Toivonen et al. 2006).

Remote sensing of floristic differences

Vegetation classes

Land-cover mapping is one of the most common applications of remote sensing, whether at a global scale (Bartholome & Belward 2005, Mayaux et al. 2005), a continental scale (Mayaux et al. 1999, Roberts et al. 2003, Eva et al. 2004, Mayaux et al. 2004, Stibig et al. 2007) or a regional one (Pedroni 2003). In these studies the separation of vegetation classes was mainly based on the physiognomic characteristics of forests, on macro-climatic conditions and on topography. The species composition of the classes was not considered, and the validity of the different classes as surrogates for floristic variation thus remained untested.

Remote sensing studies on the stages of tropical forest succession have been numerous, but have been motivated by the quantification of carbon sources and sinks, rather than the idenfication of species turnover through succession. A relatively low correspondence between regenerative age (class) and spectral information has usually been reported (Sader et al. 1989, Lucas et al. 2000, Lu 2005 and references therein, Kuplich 2006). Recently, hyperspectral images (Thenkabail et al. 2004), Landsat ETM+ images (Vieira et al. 2003) and radar images combined with optical images (Kuplich 2006) have provided promising results in discriminating between secondary forest classes. Vieira (2003) and Kuplich (2006) also linked information on dominant tree species to the forest classes distinguished in the remotely sensed data. Some forest types traditionally recognised by indigenous peoples have also been classified from satellite images (Shepard et al. 2004, Hernandez-Stefanoni et al. 2006).

Only few remote sensing studies have attempted to identify floristically defined old-growth rain forest types within the main environmental discontinuities, such as swamps or *tierra firme* forest. Early attempts at distinguishing floristically different old-growth forest types have not been very successful (Hill & Foody 1994, Paradella et al. 1994, Foody & Hill 1996, Hill 1999). More recently, relatively high overall accuracies have been achieved in classifying forest classes differing in their species composition. In Borneo, Foody (2003) ordinated nine undisturbed and logged rain-forest classes using a self-organising map (SOM) neural network and data on commercially valuable tree species, and discriminated them with an accuracy of 96 % from the Landsat TM image. This classification accuracy, however, was probably an overestimate, since all the data (24 field inventory plots) was used in both training and testing the classification. In a more recent study by Foody & Cutler (2006), the SOM neural network was applied separately to tree species data and Landsat TM data in grouping field plots; a high correspondence (83 %) was found between the two partitionings. The classification and modelling of vegetation communities has also been improved by integrating spectral information from satellite images with (remotely sensed) data on topography, climate and soil (Ferrier et al. 2002, Sesnie et al. in press).

Number of species in ecologically defined categories

Species may first be assembled into groups that share similar habitat requirements, after which the proportion of species belonging to the group is modelled and predicted on the basis of environmental gradients (Ferrier & Watson 1997, Ferrier et al. 2002 and references therein, Funk & Richardson 2002). Faith & Walker (1996b) have also suggested that the number or proportion of species in each group may also be used to estimate unknown variation in an environmental characteristic. Such studies have not directly employed spectral information from satellite images, but many of the environmental factors used (e.g. topography, vegetation map) can be derived by remote sensing.

The spectral data of satellite images are mainly influenced by the reflectance of the forest canopy. The number of species in ecological categories can be predicted by remote sensing if the ecological characteristics that were used to define the species groups also control the distributions of forest canopy species. The predicted numbers of species in ecological categories indicate underlying soil and topographic characteristics, which in turn may be reflected in the species compositional differences of the canopy and in the pixel values of satellite images. This approach was employed in Paper I.

The distributions of species groups have been much less studied in the tropics than the distributions of vegetation classes. To the best of my knowledge, there has been no previous study in which the number or proportion of species in ecologically defined groups has been predicted in rain-forest areas on the basis of the spectral information obtained from a satellite image. An obvious explanation for the lack of such studies is the scarcity of autecological knowledge of rainforest plant species.

Compositional variation and dissimilarity

Floristic variation may be predicted and mapped as a continuous variable by summarising species composition using some form of ordination and then predicting the ordination axis scores using environmental (possibly remotely sensed) information (Öhmann & Gregory 2002, II). The main compositional gradients may be visualised as mapping the scores of each axis separately (Öhmann & Gregory 2002) or all three axes simultaneously as an RGB image (II). This approach has not been applied in the tropics except in Paper II.

Alternatively, a relationship can be modelled between a site-by-site matrix of compositional dissimilarity and matrices of environmental and geographical distances. The compositional dissimilarity of a pair of field plots can be described as a function of their relative position on environmental gradients and in geographical space. The output is a matrix of compositional dissimilarity for every pair of grid cells, as predicted using environmental and spatial data (Ferrier et al. 2007). The spectral features of a satellite image analysis could be incorporated into the model along with any other abiotic environmental variables (Ferrier et al. 2002). This approach has been employed in the tropics, but remotely sensed information was not directly employed in the predictions (Faith & Ferrier 2002, Ferrier et al. 2007).

Table 1. Environmental characteristics of the study areas (Sanford Jr et al. 1994, Marengo 1998, Lips et al. 2001).

Study Area	Location	Area (km^2)	Rainfall per year (mm)	Rainfall of driest month (mm)	Average monthly temperature (˚C)	Range of elevation (m)
Yasuní	Eastern Ecuador	670	2850	130	26	200-300
Yavarí	North-eastern Peru	800	3100	180	25-27	100-180
Río San Juan	Northern Costa Rica	2600	4000	150	25-27	20-350

MATERIALS AND METHODS

Study areas

The study was conducted in three areas in lowland tropical wet forests in western Amazonia and Costa Rica: 1) Yasuní National Park in eastern Ecuador (I and II), 2) the proposed conservation unit of Yavarí-Mirín in north-eastern Peru (III) and 3) the biological corridor of La Selva - Río San Juan in northern Costa Rica (IV) (Fig. 1, Table 1).

The Amazonian study areas, Yasuní and Yavarí, are mostly old-growth lowland rain forests with little evidence of human disturbance. They are mainly covered by non-inundated (tierra firme) forest and to a lesser extent by seasonally inundated forest and palm swamps. Tierra firme forest is broadly defined as evergreen forest in undulating lowland terrain. Tierra firme forests on white sand soil and bamboo-dominated forests are routinely distinguished and also rather well-documented as clearly distinct forest types within tierra firme (Anderson 1981, Encarnación 1985, Nelson 1994). Other subdivisions of tierra firme forests do exist but are much more speculative, such as the Brazilian concepts of dense forest and open forest with or without palms (Pires et al. 1985). The Costa Rican study area (henceforth Río San Juan) is a highly fragmented mosaic of old-growth and regrowth rain forest patches, plantations and agricultural land. The largest patches of old-growth forest are found inside protected areas.

Field data and analyses of species data

The field sampling varied between the study areas (Table 2). The field plots in all three study areas were geolocated in the field using a handheld GPS or high-resolution (1:2000) maps (11 old-growth plots in Río San Juan, paper IV).

In Río San Juan (IV), tree species > 30 cm and palm species > 10 cm in dbh (diameter at breast height) were identified and their diameter were measured in the old-growth forest plots. The trees > 30 cm in dbh were assumed to form a canopy and thus to have the greatest effect on the spectral reflectance of satellite images. The regrowth forests were

visually classified in the single-regrowth forest class on the basis of tree height and canopy closure, but no field measurements were conducted.

In Yasuní (I and II), Melastomataceae (mainly shrubs and small trees) and pteridophyte species (terrestrial or low-epiphytic ferns and fern allies, max. 2 m above ground level) and in Yavarí (III) pteridophyte species were used as indicator species of more general floristic patterns. The number of Melastomataceae and pteridophyte species is considerably lower and they are faster and thus less expensive to inventory than trees, which facilitates geographically extensive and floristically representative field sampling (Ruokolainen et al. 1997, Vormisto et al. 2000, Tuomisto et al. 2003b, Ruokolainen et al. 2007).
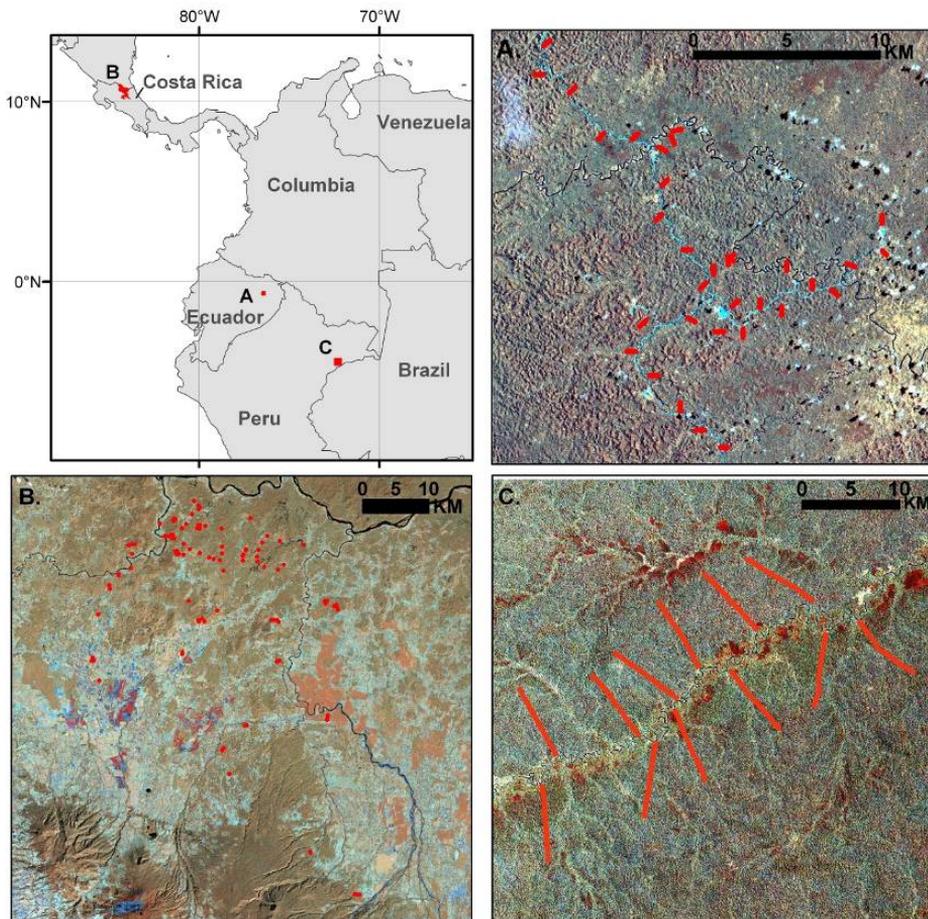
Figure 1. Location of study areas and field plots overlaid (in red) in the Landsat TM (maps A and C) or Landsat ETM+ (map B) satellite image using spectral bands 4, 5 and 7. Map A shows the Yasuní study area in Ecuador; map B the Río San Juan - La Selva study area in Costa Rica; map C the Yavarí study area in Peru.

Table 2. Field variables and sampling date and design in the different study areas. The last two rows summarize the analysis of the field data and the derived variables that were predicted on the basis of satellite image information. In the Río San Juan study area the field data was collected only from old-growth plots (N=52).

Study area	Yasuní	Yavarí	Río San Juan
Field sampling	340 plots: 34 transects (500 m x 5 m) subdivided in 10 plots (50 m x 5 m)	635 plots: 8 transects (8000 m x 2 m) subdivided in 80 plots (100 m x 2 m)	104 plots (50 m x 50 m)
Time of field work	April 1996, April 1997, March 1998	March - April 2002	1997, May – October 2003, January-April 2004
Identified species	Melastomataceae, pteridophytes	Pteridophytes	Tree species > 30 cm in dbh, Palm species > 10 cm in dbh
Other field measurements	- Soil samples - Topographic position*	-	Diameter
Analysis of field data	- Soil based classification - NMDS	Hierarchical clustering	Hierarchical clustering
Number and type of field variables	- Nbr of species in ecological categories (12) - Nbr of species in taxonomic groups (3) - Shannon's diversity index - Scores of ordination axes (3)	4 vegetation types: 1) Terrace forest, 2) Pebas formation** 4) Intermediate *tierra firme* forest ** 3) Inundated forest	4 forest types: 1) *Pentaclethra-Carapa* forest ** 2) Palm forest 3) Mixed forest** 4) Regrowth forests

* valley bottom/flatland, slope or hill top, ** classes combined in the final classification

In Yasuní (I), the species were classified into four ecological categories according to the highest density of individuals per category: poor or rich soil species (as indicated by soil cation content), floodplain species and swamp species. The number of species of Melastomataceae, pteridophytes and both of these groups combined in each category resulted in 12 field variables (4 categories by 3 species groups). Four additional field variables were obtained from the total number of species of Melastomataceae, pteridophytes, and both combined as well as Shannon's Diversity index (Table 2).

In the other study of Yasuní area (II), variation in plant species composition was expressed in a three-dimensional ordination space through a non-metric multi-dimensional scaling (NMDS) -ordination. This allowed summarizing the original 334-dimensional site-by-species data as a much lower number of dimensions, so that it became feasible to both

predict and visualize the predicted NMDS axis scores for unvisited locations on the basis of satellite image data.

In Yavarí and Río San Juan, the species data were categorized through hierarchical clustering. In Yavarí (III), the clustering resulted in four vegetation classes, defined by species composition (Salovaara et al. 2004). In the final prediction of the floristic category for unvisited sites, the two floristically most similar forest types were combined because we were unable to separate them on the basis of remotely sensed data (Table 2).

In Río San Juan (IV), the old-growth forest plots were clustered into three classes using species-specific importance values (Ramos Bendaña 2004). Two of the three old-growth forest types were floristically similar, *Pentaclethra macroloba* -dominated forests, and these were combined in the final classification. The final classification consisted of the regrowth forest class, which mainly included abandoned pastures and tree plantations, and the two old-growth forest classes (Table 2).

Remotely sensed data and pre-processing

The sufficiently cloud-free Landsat TM and ETM+ satellite images closest to the fieldwork dates were selected and preprocessed (Table 3). The images were georeferenced using a first-order polynomial model and resampled to a pixel size of 30 m. If clouds or cloud shadows were present, they were digitised and masked out. Field-plots lying in the masked areas were excluded from the analyses. A digital elevation model from the Shuttle Radar Topography Mission (SRTM DEM) was acquired for the Yavarí and the Río San Juan study areas. It is based on C-band radar data and has a horizontal resolution of 90 m. The data were downloaded free of charge from http://srtm.usgs.gov/ in 2003.

The radiance values captured by satellite sensors are affected by water vapour, aerosols and other atmospheric constituents as well as topography and sun position (Lillesand & Kiefer 2000). An atmospheric correction using the S6 atmospheric model (Vermote et al. 1997) was performed for the Yasuní image to correct atmospheric effects on pixel values (I and II). The correction, however, was considered risky, since the aerosols' optical depth was estimated using the same forest areas that were being studied. It has also been argued that the atmospheric correction is unnecessary when atmospheric measurements are not available for the study area, only a single satellite image scene is analysed, and training and test data sets are collected from the same scene (Song et al. 2001). Due to these reasons we did not perform the atmospheric correction for the satellite images from Yavarí (III) and Río San Juan (IV) study areas. At any rate, the correction had almost no effect on the predictions in Yasuní.

Table 3. Basic information on the satellite images and applied preprocessing.

Study Area	Satellite image	Path / row	Date	Pixel size (m)	Preprocessing
Yasuní	TM5	9 / 60	Sept. 1995	30 x 30	Atmospheric correction, destriping, rectification
Yavarí	ETM+	5 / 63	June 2001	30 x 30	Rectification
Río San Juan	TM5	15 / 53	Jan. 2001	30 x 30	Rectification, topographic normalisation

Topographic normalisation was not considered essential for the images from Yasuní and Yavarí, where topographic variation was limited (Table 1). In Río San Juan area the topography seemed to have considerable impact on the pixel values, and the image was corrected using the Minnaert topographic normalisation with bandwise calculated constant *k* (Minnaert & Szeicz 1961, Colby 1991). The topographic normalisation had only a slight effect on the classification accuracy.

Extraction of features from satellite image

The spectral value of a pixel is always a mixture of reflectances from varying sources, and variation exists between neighbouring pixels even in homogeneous land cover. Likewise location errors due to image rectification and georeferencing of field plots can increase prediction errors when single pixel values are used in image analysis. A single pixel of a Landsat image is so small (30 m × 30 m) that its spectral characteristics can be greatly affected by the structural and dynamic factors of the forest, thus complicating the relationship between satellite image and field data on species composition. Such factors include the size and inclination of a single large tree canopy, as well as tree-fall gaps and other structural changes that have appeared during the time gap between the acquisition of the satellite data and the field work. The effects of these uncertainties were reduced by extracting the spectral features and elevation either from pixel windows (all papers) or from segments (IV).

The size of a pixel window varied between the studies (Table 4) and was determined by the field sampling design and/or by testing different pixel window sizes. The pixel windows of geographically close plots may be overlapping, especially when transect line sampling is applied. This problem of dependence between neighbouring plots was avoided by specifying a minimum distance in selecting nearest neighbours (plots) in the error estimation of k-nn predictions in Yasuní (I and II). Thus the spectrally nearest neighbours for a pixel under estimation were specified as geographically further away than a minimum distance of 200 m. In Yavarí the number of field plots (N=635) was greater than in Yasuní (N=340), and the field plots were combined into larger units (200 m and 500 m long sampling units) to avoid overlap between neighbouring pixel windows (IV). In Río San Juan the field plots were located in such a way that the distance between plots was at least 150 m. Accordingly, a window of 5x5 pixels could be used.

The use of large pixel window increases the probability that a window will contain two or more land cover types, potentially reducing classification accuracy (Hill 1999). This problem can be avoided by segmentation, whereby a satellite image is divided into regions (segments) that are spatially continuous and internally homogeneous in their image features (Narendra & Goldberg 1980). We tested segmentation in the Río San Juan study area (IV), because segmentation is expected to increase classification accuracy especially in fragmented areas. The segmentation was performed using a modified implementation of "segmentation with directed trees" (Narendra & Goldberg 1980, Pekkarinen 2002). This method employs a type of watershed algorithm, in which the image is first divided into plateau and edge regions. All the plateau areas are then labelled, and labels for the edge regions are sought in the direction of the local edge gradient.

The spectral and elevation features were extracted from a Landsat TM or ETM+ satellite image and from the SRTM DEM as a mean and variance or standard deviation of pixels belonging to the window or segment. We also tested several band ratios (Table 4).

Table 4. Spatial unit of feature extraction, extracted features and features used in the prediction tasks. The table also indicates whether the predictions were based on the k-nn method or the discriminant analysis (DA), and the size of the field data used. M = mean; SD = standard deviation; var = variance of pixel values of image bands indicated by their number.

Prediction task	Unit of feature extraction (pixels)	Nbr and size (m) of plots	Method	Extracted features	Employed features
Species richness, Yasuní	pixel window (7 x 7)	335, 50 x 5	k-.nn	M1-M5, M7, SD1-SD5, SD7, M4/3	All 12 features, weighted
Axes scores, Yasuní	pixel window (7 x 7)	335, 50 x 5	k-nn	M1-M5, M7	All 12 features, weighted
Vegetation classes, Yavarí	pixel window (5 x 5)	317, 200 x 2	DA	M2, M4, M5, M7, SD2, SD4, SD5, SD7, M4/5, elevation	Selected 4 features (M7, M4/M5, m5, elevation)
	pixel window (12 x 15)	127, 500 x 2	DA	M2, M4, M5, M7, SD2, SD4, SD5, SD7, M4/5, elevation, sd of elevation	Selected 8 features (M2, M4, M5, M7, SD2, SD4, SD5, elevation)
Forest types, Río San Juan	pixel window (5 x 5)	103, 50 x 50	k-nn,	M1, M4, M5, M7, M1/M4, M1/M5, M1/M7, M4/M7, M4/M5, M5/M7, var1, var4, var5, var7, elevation	All 15 features, weighted
	pixel window (5 x 5)	103, 50 x 50	DA	Same as above	Canonical variables computed from M1, M4, M5, M7, M1/M4, M1/M5, M1/M7, M4/M7, M4/M5, M5/M7 and var7,
	segment	102, 50 x 50	k-nn	Same as above	All 15 features, weighted
	segment	102, 50 x 50	DA	Same as above	Canonical variables computed from M1, M4, M5, M7, M1/M4, M1/M5, M1/M7, M4/M7, M4/M5, M5/M7, var7 and elevation

Prediction methods and accuracy assessment

The non-parametric k-nn method was employed to predict continuous (I and II) and discrete (IV) field variables as follows: we first searched the k nearest neighbours by calculating Euclidean distance in feature space from the pixel lacking field data to each pixel for which field data were available. The three nearest neighbours (three pixels most similar in terms of spectral features and elevation) were then selected for the pixel among those pixels with field data. The predicted continuous field variable was calculated as a weighted mean of the k nearest neighbours. The weights were calculated as inversionally proportional to the squared Euclidean distance. In the case of discrete variables the forest type was obtained that had the highest sum of weights among the nearest neighbours (Tomppo 1991, 1996). The spectral and elevation features were also weighted in calculating distances to the k nearest neighbours. The weights of the features were determined using an optimization based on a genetic algorithm (Tomppo & Halme 2004, Tomppo et al. 2007). The algorithm minimized the root mean square error (RMSE) of the predictions for the continuous field variables or the value of 1-Kappa of the classification.

The field plots of Yavarí and Río San Juan were classified using a linear discriminant analysis (III and IV). The discriminant model searches the linear function of explanatory variables (spectral features and elevation) that best separates predefined classes. The parameters of the discriminant model are estimated on the basis of a field data set consisting of field-verified sampling plots with known class membership; the resulting model can be used to predict class membership for non-visited locations (Legendre & Legendre 1998). In Río San Juan (IV) we ran a canonical discriminant analysis prior to the linear dicriminant analysis. It forms linear combinations of spectral features and elevation, so-called canonical variables that best summarize between-class variation. The number of canonical variables was the number of forest types minus 1. All the features that significantly contributed to discriminating among forest types, according to the F test ($p < 0.05$), were included in the canonical discriminant analysis. In Yavarí the features were selected by a stepwise discriminant analysis with backward elimination (Table 4).

The accuracy of the predictions was assessed by comparing them to the field observations using a one-leave-out cross-validation (all papers) and an independent test data set (I and IV). In the cross-validation, either one plot (I, II and IV) or all the field plots of one transect (III) were excluded in turn and the rest of the field plots were used to predict the field variable for the excluded plot(s). Error matrices and derived overall perceptual accuracy and Kappa scores were employed as measures of unit-level classification accuracy. The unit was either the pixel window or the segment. We also compared the statistical significance of Kappa scores resulting from two different error matrices in Paper IV (Congalton & Green 1999). In the case of the continuous variables, RMSE, bias and standard deviation of bias were calculated to estimate the unit-level accuracy of the predictions. The predictions of the continuous variables were also compared against a null model, where predictions were calculated using randomly selected neighbours rather than spectrally nearest ones.

Other methods

We identified the areas that were not covered by the spectral characteristics of the field plot locations in order to avoid extrapolation (I and III). Extrapolation would force all the pixels into the predefined range of the field variable values, whether or not their spectral characteristics are represented among the training sites. Pixels that were beyond the ranges of spectral features corresponding to the field sampling units were defined for one pair of spectral bands in turn. We plotted all the pixel values of two bands and drew a convex hull around the pixels that were verified in the field. Pixels that were outside the convex hull in at least one pair of image bands were masked out from the image. This procedure increased the reliability of the predictions and the resulting map. It will also help to direct field sampling in the future to areas representing unexplored spectral characteristics. These areas can be assumed to be more likely to represent vegetation not previously verified in the field.

We employed the Mantel test to study correlations between remotely sensed features and the species composition recorded in the field (III). This test computes a correlation coefficient between two sites-by-sites (dis)similarity matrices with the same dimensions (Legendre & Legendre 1998). The Mantel tests were run between the matrices of floristic similarities (expressed as the Bray-Curtis dissimilarity index) and the distances of spectral and elevation features (expressed as Euclidean distances). The statistical significance of the correlations was estimated with a Monte Carlo permutation test.

Spatial autocorrelation in species richness and composition can be high with short geographical distances. We studied autocorrelation of the field variables by calculating semivariograms (I). The semivariograms were also used to define the minimum distance used in selecting the nearest neighbours in the error estimation of k-nn predictions in Paper I.

RESULTS AND DISCUSSION

Can floristic patterns be predicted and mapped accurately by remote sensing?

Species composition

The species composition of Melastomataceae and pteridophyte species that was summarized in the three axes of the NMDS ordination was predicted for Yasuní using the k-nn method (II). The accuracy of the predictions was difficult to assess simply on the basis of the RMSE values, since the axis score values cannot be observed in the field and do not have any concrete meaning. The map produced tells us which sites are similar in species composition and which differ, but we do not know which species actually occur at the sites. An axis score of 2, for instance, is arbitrary unless it is related to scores for other sites. To assess the accuracy of our predictions, we compared the RMSE values of the predictions to a null model based on randomly selected nearest neighbours; in other words, the null model represented a pure guess at axis scores. This comparison revealed that species composition can be predicted relatively accurately: the pooled RMSE for the predictions of the three ordination axis scores based on the spectral information of the Landsat TM was always lower than the pooled RMSE of the predictions based on the null model. The predictions for individual axis scores were mostly more accurate than the null model, which was run

100 times. Least accurate among the three axes were the predictions for the scores of axis 2, related to drainage.

Predicting species composition in tropical forests is rarely attempted without relating it to vegetation classes. Recently, Ferrier (2000) predicted compositional dissimilarity on the basis of environmental data (e.g. climatic gradients) and geographical distances, whereas we used the spectral information provided by the satellite image to describe environmental variation. The generalised dissimilarity modelling (GDM) employed by Ferrier et al. (2007) can also take advantage of remotely sensed data, although spectral information from satellite images has not yet been utilised (Ferrier 2002, Ferrier et al. 2007). Ferrier's group and ours used different methods in prediction but also employed different approaches. We first summarised the data on a lower number of dimensions by ordination and then predicted the compositional differences expressed by the axis scores for unvisited locations. Ferrier et al. (2007) first predicted the compositional similarity values for unvisited sites; the multidimensional data of predicted compositional similarities can then be summarised e.g. by ordination.

The predicted floristic ordination scores of three axes were visualised as a RGB colour composite, with the axes represented in red, green or blue. By this means we were able to summarize compositional variation in a single map layer. The visual inspection of the predicted map and the original Landsat TM image also confirmed the interpretation of the axes and helped to locate areas of mispredictions. The number of field plots in inundated forest and swamps was low and therefore structural and floristic variation of these forest types were not well represented in the field data. This reduced the prediction accuracy for axis 2 and appeared as mispredictions on the map as well as high RMSE values (II).

The advantages of this approach are that the compositional differences are predicted as a continuous variable without any a priori classification. The methods used summarize the multidimensional data in the format of a single map layer, showing gradual changes in species composition between sites. Gradual floristic changes between vegetation types have often been considered difficult in traditional vegetation mapping (Powell et al. 2004). Such continuous information on compositional differences can easily be applied in conservation planning and biodiversity modelling (Ferrier 2002). The approach can also help to locate future field sampling in areas with a distinct species composition compared to areas sampled during earlier field campaigns. The drawback, however, is that information on axis scores is in many cases more difficult to interpret in practical applications compared to vegetation classes.

Vegetation types

The classification accuracies of three floristically defined vegetation types in Yavarí (III) and Río San Juan (IV) were relatively high (overall accuracies of 85 % and 91 %, respectively) when estimated by the one-leave-out cross-validation. The classification of three old-growth forest types and one regrowth type in Río San Juan was also reasonable (82.5 %) when the k-nn method was used, especially when it is kept in mind that in both areas we classified structurally relatively similar land cover categories. In both areas, two of the three forest classes distinguished in the studies represented previously unrecognised classes of *tierra firme* or dense old-growth forest. The class accuracies of the most accurate classification were also at an acceptable level in both cases (>73 % in Yavarí and >88 % in Río San Juan). In Yavarí the user's accuracy of terrace forest class was still low, 48 %, partly due to great differences in the number of sampling units between vegetation types.

In Yavarí only the linear discriminant analysis was used to classify forest types, but in Río San Juan we tested both canonical discriminant analysis and the k-nn method. The k-nn method resulted in higher classification accuracy and Kappa scores than canonical discriminant analysis in classifying two old-growth forest types and one regrowth type, but the differences in Kappa scores were not statistically significant.

The classification accuracies can also be considered high when compared to those rare studies which have distinguished two or more old-growth rain forest types. Within *tierra firme*, high classification errors have mainly been reported (Hill & Foody 1994, Foody & Hill 1996, Hill 1999) but flooded and *tierra firme* forest have been discriminated accurately (Lobo et al. 1998, Hess et al. 2003). Discriminating among secondary forest stages has proved to be difficult (Sader et al. 1989, Lucas et al. 2000, Lu et al. 2003, Lu 2005 and references therein, Kuplich 2006), until recently with the availability of new satellite sensors and classification methods (Vieira et al. 2003, Thenkabail et al. 2004).

The forest or vegetation classification presented has the advantage that the classes are defined by their species composition, and thus express the main floristic differences in the study area. Such maps certainly have a place in the practical work of forest management and nature conservation, as they are easy to interpret and use. All the vegetation / forest type maps still provide information at a relatively general level, no matter how good is spatial accuracy or thematic discrimination. Information on floristic variation is always lost in vegetation classification. Firstly, each pixel is forced into a single class even though in natural landscapes changes between habitats are mostly gradual. Secondly, the classes are also assumed to be internally homogeneous, although species are distributed patchily within the vegetation classes defined. This is probably because species respond to different environmental gradients from those that have been used in defining the vegetation classes, or because of biological and historical factors such as competition, local extinction and dispersal. Finally, the vegetation classification does not necessarily take into account that some classes are more similar in species composition than the others; misclassifications are normally seen as equally erroneous, whether for example a forest type is misclassified as another, floristically and ecologically relatively similar forest type or as a very different, non-forested class (Faith & Walker 1996a, Guisan & Zimmermann 2000, Ferrier 2002).

Species richness of ecologically defined species groups

Numbers of Melastomataceae and pteridophyte species in three ecological categories (poor soil species, rich soil species and swamp forest species) were predicted fairly accurately compared to the null model. The predictions for species richness in the fourth ecological category, floodplain forest, showed the lowest accuracies, due apparently at least in part to the small number of field plots in that habitat type. The derived maps showed non-random spatial patterns for all the predictions. Visual comparison with the original Landsat TM image revealed that these patterns corresponded to at least elevational variation (I).

Differences in numbers of species in the ecological categories were interpreted as indirectly approximating compositional differences. Changes in the numbers of species in ecological categories indicate an environmental gradient. For example the number of poor soil species decreases with the gradient of soil fertility. The patterns of species richness in ecological categories were interpreted as reflecting topographical patterns and related edaphic patterns. These same environmental gradients in turn influence the composition of species, and thus variation in the number of species in ecologically defined species groups indirectly reflects differences in species composition. Similar argumentation has been used

by Faith & Walker (1996b) in the selection of conservation areas on the basis of indicator species number and environmental data. The number of species in ecological categories may also be related to the vegetation classification. For example sites with a high number of species preferring poor soils can be classified as poor soil *tierra firme* forest. Similarly, swamps will probably have a high number of swamp species. From these perspectives the relatively high accuracy of predictions for number of species in ecological categories was expected. The results (I) were also well in line with the results of Papers II-IV and with those studies that have found a relationship between species composition or floristically defined forest classes and the spectral information provided by satellite images (Thenkabail et al. 2003, Tuomisto et al. 2003a, 2003b).

A common problem in predicting the distributions of individual species is that many of species sampled are present only in a few field plots. This problem is solved by predicting species richness in ecological categories. The difficulties of the approach are that the habitat requirements of a species may be unknown, environmental information is often scarce and each category is predicted as a separate layer. We were able to classify species into ecological categories on the basis of soil sample data collected together with species data. However, soil data rarely exist and soil sampling increases the resources needed in field sampling, thus diminishing the applicability of the approach.

We also predicted the number of species in taxonomic groups and Shannon's diversity index, but the predictions were not accurate. These results corresponded to those that have reported low matrix correlation coefficients between species richness and spectral values of Landsat TM images (Tuomisto et al. 2003a, 2003b). In any case, species richness and diversity indices alone are of limited value in conservation planning.

Factors influencing the accuracy of predictions

Success in prediction naturally depends on the strength of the relationship between species and remotely sensed predictors, but it is also influenced by many other factors, such as the spatial resolution of the data and the methods used (Hill & Foody 1994, Ferrier & Watson 1997, Lobo et al. 1998, Kleinn et al. 2002).

Feature selection

The importance of feature weighting was highlighted in the results for Río San Juan (IV), where an increase in the Kappa score of over 20 % was obtained by the weighting of the spectral features and elevation in the k-nn classification. In the other prediction tasks (I-II) the effect of feature weighting in the k-nn method varied from slight increase to a 5.6 % decrease in RMSE. The prediction accuracy of four out of the 16 variables predicted decreased due to feature weighting in estimating species richness in Yasuní. The weights for the features were computed for all the field variables at once, and the solution with the lowest sum of RMSE values was selected. The results would probably have been improved if the weighting had been performed separately for the numbers of species in taxonomic groups and in ecological categories. The selection of features for the linear discriminant analysis also increased the proportion of correctly classified plots by 1.6 - 9.5 %.

In general, the mean values of spectral bands computed within a pixel window or segment showed higher matrix correlation coefficients with species composition (III), and they had higher discriminant power between forest types (III and IV) than the standard

deviations of spectral values. The mean pixel values for Landsat band 7 showed the highest matrix correlation with species composition in Yavarí; it was followed by the other middle infrared (MIR) band 5 and the near infrared (NIR) band 4 (III). Of the spectral bands of Landsat TM, MIR bands 5 and 7 showed highest discriminant power between vegetation types (IV). These results agree with earlier findings showing significant matrix correlations between the MIR and NIR bands and pteridophyte and Melastomataceae species composition in primary rain forests (Tuomisto et al. 2003a, Tuomisto et al. 2003b). Both the MIR and the NIR wavelengths are sensitive to vegetation moisture content and soil moisture, and thus have discriminated effectively between vegetation types along moisture gradients (Kaufman & Remer 1994, Thenkabail et al. 2003), but also between forest successional stages (Boyd et al. 1999, Vieira et al. 2003).

A vegetation index employing MIR and NIR bands (NIR-MIR/NIR+MIR) has been suggested as an alternative to the commonly used NDVI (Normalised difference vegetation index) in areas where atmospheric haze is often prominent, such as the tropics (Kaufman & Remer 1994). We were unable to compute either the NDVI or the simple ratio of Landsat bands 3 and 4 in Yavarí and Río San Juan because of prominent striping in band 3 (III and IV). In Yasuní the NDVI did not reduce the RMSE of the species richness predictions (I). The simple ratio of NIR (Landsat band 4) and MIR (Landsat band 5) in turn proved to have relatively high discriminant power between vegetation and forest types (III and IV), and a high matrix correlation coefficient with pteridophyte species composition (III).

Elevational variation was relatively low in all the study areas, but elevation may be considered important auxiliary information. Variation in species composition and species richness in the ecological categories were interpreted to be explained mainly by topography or by edaphic factors (drainage, soil fertility) (I and II). Elevation also showed relatively high matrix correlation with floristic composition in Yavarí and had high discriminant power between vegetation types (III). This was expected in the light of earlier findings that have reported high correlation coefficients between floristic and topographic patterns (Clark et al. 1995, Tuomisto et al. 2003b, Cayuela et al. 2006, Sesnie et al. in press,). Edaphic features and forest structure also correlate with topography (Clark et al. 1995, Tuomisto et al. 2003b, Valencia et al. 2004), and this correlation may help to discriminate forest types using the elevation feature. In Río San Juan, however, elevation did not vary significantly between forest types (IV).

Freely available SRTM DEM covers all the tropical areas, and provides elevation information at a spatial resolution of 90 m and with an average vertical error less than 16 m. The spatial and vertical resolution of the data may not be optimal for lowland areas, but is often the best elevation data available for the tropics. The vertical accuracy of the SRTM model can be increased by averaging elevation information over larger pixel units (Kellndorfer et al. 2004, Walker et al. 2007).

Spatial and thematic resolution

The appropriate spatial resolution was searched for in the vegetation type classification in Yavarí (III) and in the species richness predictions in Yasuní (I). Both studies suggested that a larger pixel window increases classification accuracy. Congruent results have also been reported earlier by Hill & Foody (1994) and Hill (1999).

The problem of mixed pixel units was indicated in the classification results for Yavarí. The 500-m sampling units containing several vegetation types (as shown by the field sampling, which was done at a finer scale than the final classification) were more

commonly misclassified than sampling units that were located within a single forest type (III). The confusion caused by combining more than one vegetation type within a pixel window may be reduced by segmentation. Thus segmentation may be seen as most useful in heterogeneous and especially in fragmented landscapes, where abrupt changes between land cover types are typical.

In the fragmented Río San Juan area, segmentation seemed to quite effectively separate agricultural land from forested land. Within the forests, however, despite topographic normalisation patches also corresponded to the shadows of hills. Against our expectations, based on earlier studies (Lobo 1997, Lobo et al. 1998, Hill 1999), segmentation did not reduce classification errors in Río San Juan. The reason may be found in the sampling design: the field plots were located in forest interiors, at least 150 m from the forest edge, so that the mixing of different land cover types in windows of 5×5 pixels was mostly avoided.

Accuracy obviously depends on what is predicted and how detailed level of discrimination is aimed at. For example discrimination between forest and non-forest may be expected to be easier than discrimination between forest types. In Yavarí (III) and Río San Juan (IV) we sought a level of compositional differences between vegetation types that can be distinguished by means of remotely sensed data. In both cases the two floristically most similar classes were combined in order to raise classification accuracy to an acceptable level. The combined forest types in Río San Juan were both floristically and structurally (dbh and stem density) the closest pair of forest types in our data. In Yavarí we did not have field data on forest structure.

Satellite images

The recently reported east-west radiometric gradient (Collett et al. 1998, Toivonen et al. 2002, 2006) needs to be considered, at least if floristic variation is being predicted over large areas. The existence of the gradient means that field sampling needs to be relatively dense, as only the spectral values of nearby field plots can be used in predicting floristic differences. The field plots that are used in the prediction should be within ca. 30 km; further away, variation in pixel values due to the gradient may override the influence of vegetation (Schulman et al. 2007). Our study areas were relatively small in size. In Yasuní, the maximum length in an east-west direction was ca. 25 km and in Río San Juan it was about 37 km. Thus the risk that the nearest neighbours for any pixel in the k-nn prediction would have been situated more than 30 km in an east-west direction was low in Río San Juan and non-existent in Yasuní. In future studies it will possible and perhaps even advisable to limit the search distance to perhaps 30 km in selecting the nearest neighbours in the k-nn method. This can easily be done with the k-nn method; a search distance limit in a north-south direction has also been operationally used in applying the k-nn method in the national forest inventory of Finland (Katila & Tomppo 2001). No such easy way to limit the effect of the radiometric gradient on predictions exists in discriminant analysis.

Prediction accuracy may be increased by using other image data instead of, or in addition to, Landsat TM or ETM+. Some other prediction method than k-nn or discriminant analysis might also improve prediction success. Hyperspectral images (Hyperion) showed a remarkably better discrimination of fallow and of primary and secondary forest classes than the broadband Landsat ETM+ image (Thenkabail et al. 2004). The spatial resolution of hyperspectral satellite images is comparable to that of Landsat images, but they register over 200 narrow bands along the 400-2500 nm range, which should improve the

identification of complex vegetation features (Lillesand & Kiefer 2000). Hyperspectral images are also less prone to saturation, which is known to cause problems for instance in predicting biomass in the tropics (Thenkabail et al. 2004).

New satellites, providing images with high spatial resolution, have also quite recently been launched, for example IKONOS and Quickbird. At the same time a panchromatic, high-resolution band has been incorporated in many medium-resolution images (e.g. Landsat ETM+, SPOT5). Comparisons of classifications performed with the hyperspatial IKONOS image (pixel size of 4 m) and the Landsat ETM+ have yielded controversial results. The IKONOS image performed better than the Landsat ETM+ in classifying nine classes of primary and secondary rain forest and fallows (Thenkabail et al. 2004), while three floristically defined tropical forest classes along a moisture gradient were separated more accurately by the Landsat ETM+ than by the IKONOS (Thenkabail et al. 2003). The use of a higher spatial resolution satellite needs a different approach (e.g. use of textural features), because use of spatial units larger than one Landsat pixel has shown increased accuracy in this study (I and III) and in those of Hill (1994, 1999) and Lobo (1997).

The most commonly used multispectral, medium resolution Landsat images still have several advantages compared to the relatively new hyperspectral and hyperspatial images. The Landsat images have large (185km × 185 km) scene size, whereas the hyperspectral image strip is commonly only a few kilometres wide and the scene size for instance of IKONOS is only 11 × 11 km. The availability of hyperspectral and hyperspatial images in archives is poor and the prices of images are high, whereas inexpensive, even free Landsat images from the 35-year Landsat mission are widely available through internet-based data archives (Fuller, 2006, Trigg et al. 2006). Yet another drawback of the commonly used IKONOS is that it does not have an MIR band, which has proved to discriminate well between vegetation types in papers III and IV and in other studies (Boyd et al. 1999, Thenkabail et al. 2003, Tuomisto et al. 2003a, 2003b, Vieira et al. 2003).

Due to the failure of the Landsat ETM+ scan line corrector in May 2003 (Howard & Lacasse 2004) and the ending operational lifetime of the Landsat TM5 (Trigg et al. 2006), other images are needed to replace Landsat ones. Optical, medium-resolution, multispectral but less widely used satellite images in tropical forests include SPOT 5 (4 bands, spatial resolution from 5 to 20 m, scene size of 60 km), ASTER (9 bands in visible, NIR and MIR wavelengths, spatial resolution of 15-90 m, scene size 60 km) and IRS-LISS3 (4 bands, spatial resolution of 23.5 m, scene size of 141 km) as well as on-request basis working ALI (9 bands at spatial resolution of 30 m, scene size of 37 km × 42 or 185 km). The list is not complete; for example the China-Brazil Earth Resource Satellite (CBERS) may be interesting in the remote sensing of Amazonia and other lowland tropical rain forest areas. CBERS-2 contains three sensors, of which the high resolution CCD camera registers four bands in visible and infrared ranges at a resolution of 20 m.

Prediction methods

Linear discriminant analysis is readily available in many commercial softwares and is easy to use. Unsurprisingly, it is one of the most commonly used classification methods in remote sensing studies of tropical forests (Trisurat et al. 2000, Pedroni 2003, Thenkabail et al. 2004). The k-nn method has been tested in the analysis of tropical vegetation only in papers I, II and IV, but is employed widely and also in operative use in satellite-image-based forest inventories (Tomppo 1996, Nilsson 1997, Tomppo et al. 1999, Gjertsen et al. 2000, Franco-Lopez et al. 2001, Tomppo et al. 2001, Reese et al. 2003, McInerney et al.

2005, Koukal et al. 2007, McRoberts et al. 2007) and in land cover and non-forest/forest classifications (Franco-Lopez et al. 2001, Haapanen et al. 2004) in the boreal and temperate zone. The k-nn method performed better than canonical discriminant analysis in the comparison in paper IV, but the differences in Kappa scores were not statistically significant. The k-nn method appears promising for the analysis of floristic variation in rain forest because it is non-parametric, in the sense that it does not make any assumptions on statistical distributions of the variables used or the response of field variables to spectral features. This makes the k-nn method suitable for many prediction tasks, but accurate k-nn predictions also require that the field sampling cover variation both in field variables and in the spectral features of the satellite image (Katila & Tomppo 2001).

In real-world data sets, the relationship between compositional turnover and environmental gradients is rarely linear (Ferrier et al. 2007). Thus non-parametric and nonlinear methods, such as the k-nn method, artificial neural networks and decision tree classifiers, have shown high accuracies in predicting species diversity and discriminating land cover classes or stages of forest succession (Foody & Cutler 2003, Ingram et al. 2005, Foody & Cutler 2006, Kuplich 2006, Sesnie et al. in press). This supports the conclusion that more flexible methods, which do not make assumptions on distributions of variables or expect a linear relationship among variables, are needed to increase the accuracy of information on the spatial variation of rain forest vegetation.

Procedure of accuracy assessment

It has been suggested that the accuracy of classification should be confirmed using several accuracy measures, since different measures may lead to different and even conflicting interpretations and conclusions (Foody 2002, Stehman 1997). We assessed the accuracy of the classifications using error matrices and derived overall classification accuracies and Kappa scores (III and IV). The k-nn predictions of continuous variables were validated by calculating the RMSE and bias of the predictions as well as by comparing the prediction accuracy to the null model (I and II). When the whole study area was classified (I-III), the classification results were also compared visually to the original satellite image.

The reliability of the accuracy estimate still depends on how the assessment is performed and on the kind of test data used. We used mainly one-leave-out cross-validation but also an independent data set for error estimation. There are problems related to one-leave-out cross-validation, and alternative methods have been proposed e.g. a model-based method developed by Kim and Tomppo (2006). In our case, however, I believe that the one-leave-out cross-validation indicates the level of the prediction error. The accuracy of the predictions was assessed using both cross-validation and an independent test data set in predicting the number of species in ecological categories (I) and in classifying vegetation or forest types (III; results not shown, and IV). In all cases, the accuracy estimates provided by the independent test data set followed those of the cross-validation. Actually, in most cases the cross-validation was more conservative and estimated lower accuracies for the predictions than the use of the test data (N in the test data sets between 25 and 80). In addition, the partitioning of a small field data set may cause difficulties in forming a relationship between the field variables and the spectral features. It may also be difficult to cover the ranges of variation in spectral values and field variables in both the train and the test data set, because the number of plots in some vegetation types becomes low. In the tropics, field data sets are often small due to logistical difficulties and the high cost of field

work. As few as ten plots have been used in testing predictions obtained on the basis of data from twenty plots (Foody & Cutler 2006).

Visual assessment cannot be recommended as the only accuracy assessment method, but it is a good addition to the quantitative methods. Visual comparison between the predicted map and the original satellite image is useful in analysing and locating errors in relation to the general environmental characteristics visible in a satellite image (I and II) or in maps. In Papers II and III, visual inspection revealed clear misclassifications or mispredictions, which in all cases were quite explicitly related to certain vegetation types.

In the tropics, randomly or systematically sampled validation data are often impossible to collect at reasonable cost. In Río San Juan (IV), field data were collected according to the common procedure in remote sensing studies: locating field plots far from the edges, in homogeneous land cover patches. This procedure avoids transition zones; thus the accuracy of the predictions is commonly overestimated. Transition zones, borders between classes, and subpixel objects such as roads are typically areas of so called mixed pixels, and thus difficult to classify (Smith et al. 2003, Powell et al. 2004,). The problem of mixed pixels appeared in the classification of Yavarí (III) where sampling units located in the forest interior were more accurately classified than units combining two forest types. In Papers I-III, field sampling was conducted along transect lines. In Yasuní (I-II) the transect lines were located within the broad main vegetation types, while in Yavarí (III) the transect lines crossed the main vegetation types. Transect sampling covers vegetation transition zones and thus reduces sampling bias and increases the reliability of accuracy assessment (Guisan & Zimmermann 2000).

Spatial autocorrelation may become a problem in the estimation of prediction errors, specially if field data are sampled along transect lines. The estimated accuracy may be unduly high because of positive spatial autocorrelation in both floristic and satellite image data. Species composition mostly changes gradually, and nearby plots and pixels also tend to be more similar than those further away. Pixel values are also geographically dependent; the reflectance of a pixel is influenced by reflectance from the vegetation of neighbouring pixels and atmospheric scattering. We avoided autocorrelation between plots by running the cross-validation between transect lines rather than individual plots (III) or by applying a minimum distance in selecting the nearest neighbours in error estimation (I and II). In Yasuní we found that 92 % of the selected nearest neighbours in the error estimation of the k-nn predictions were located on another transect line than the one containing the pixel being estimated (II).

Even after careful accuracy assessment, it is not easy to define the reliability of a prediction for operative use and practical applications. Despite some suggestions (Landis & Koch 1977 cited in Monserud & Leemans 1992, Thomlinson et al. 1999), there are no commonly accepted accuracy levels, even in the most common applications such as land-cover classifications. The accuracy of a classification or prediction depends on what is predicted and where; thus no categorical limits for reliability can be given. In poorly known areas, including most rain forest areas, predictions whose accuracy might be considered low in well-surveyed areas can nevertheless provide valuable and previously completely unavailable information. Even a somewhat inaccurate estimate is probably useful, as long as the alternative is a pure guess.

Links of floristic patterns to soil characteristics

The axes of the NMDS ordination of Melastomataceae and pteridophyte species in Yasuní (II) were interpreted in terms of edaphic variables and topography: the floristic gradient of axis 1 was chiefly related to topography, axis 2 was well explained by drainage, and axis 3 was related to soil cation content. The patterns of pteridophyte and Melastomatacaea species richness in the ecological category (poor soil species, rich soil species, floodplain species, swamp forest species) were also studied in Yasuní, and these patterns were visually related mainly to topographical variation (I): rich soil species were more numerous in high hill areas, while the number of poor soil species was highest in areas of low hills.

The results obtained in Yasuní were well in line with earlier studies showing that patterns of Melastomataceae, pteridophyte and tree species composition were related to edaphic patterns (Ter Steege et al. 1993, Clark et al. 1995, 1999, Vormisto et al. 2000, Potts et al. 2002, Phillips et al. 2003, Tuomisto et al. 2003a, 2003b, Paoli et al. 2006, Ruokolainen et al. 2007). In these studies the most important edaphic factors were found to be soil parent material, soil nutrients and drainage.

Topographic patterns and the composition of pteridophyte species correlated in Yavarí (III), although elevational differences in the area were low compared to the spatial resolution (90m) and vertical accuracy (error < 16m) of the used DEM. In the Yavarí area, topography, edaphic characteristics and forest types are related in such a way that forests in the fertile Pebas formation are found further from the river at higher elevations, while terrace forests grow in relatively nutrient-poor river terraces closer to the river (Salovaara et al. 2004). This relationship may have contributed to the correlation found between elevation and species composition. The western Amazonian landscape is a complex mosaic of sediment of different origins and ages, due to tectonic forces and changes in river courses (Salo et al. 1986, Räsänen et al. 1990). In other Amazonian lowland rain-forest areas, topography may thus be correlated with edaphic features positively, negatively or not at all. This may complicate the use of elevation as auxiliary data in large-area mapping. The tropical tree species composition in Costa Rica has also been shown to correlate with topographic position, even when the effect of soil is removed (Clark et al. 1999). However, in our study (IV) in Río San Juan, elevation did not discriminate well between forest types (IV).

Edaphic and topographic factors may naturally be overridden for instance by climate or geographical distance / dispersal limitation at other spatial scales than the landscape scale (Pyke et al. 2001, Phillips et al. 2003). Environmental data alone are thus not a good predictor of global biogeographical patterns. The rain-forest areas of two continents may be similar in environmental characteristics, but share few species or none.

The above results and the relatively high accuracies of the predictions of floristic variation (I, II and III) using understorey indicator species give support to the hypothesis that patterns of species composition are at least partly determined by environmental characteristics, such as edaphic factors and topography. The results also support the use of soil data, and to some extent topographic data as auxiliary data in predicting floristic variation. It is mostly difficult to obtain spatially accurate information on rain-forest soils at an appropriate scale and resolution, but remotely sensed elevation data are commonly available.

Applicability of the results

Accurate predictions and maps showing patterns of floristic variation would be an important tool in conservation planning and forest management, and would provide valuable source data in modelling for instance the distributions of animal taxa (Ferrier et al. 2002). The vegetation classification produced for Yavarí has already been successfully used in predicting abundances of mammal species (Salovaara 2005). The applicability of the predictions depends on their accuracy and the level at which they reflect more general floristic patterns, but also on the time and money needed to apply the procedure in practice. The accuracy of the predictions achieved in the papers of the thesis can be considered high, especially when we take into account the thematic resolution of the forest type discrimination in Papers I-IV. The floristic patterns of trees, Melastomataceae and pteridophytes have been shown to be correlated, as well as to correlate with edaphic and topographical variation (Ter Steege et al. 1993, Clark et al. 1995, Ruokolainen et al. 1997, Clark et al. 1998, Vormisto et al. 2000, Potts et al. 2002, Tuomisto et al. 2003a, 2003b, Jones 2006, Paoli et al. 2006, Ruokolainen et al. 2007). If floristic patterns are environmentally determined, this would mean that the patterns predicted reflect more general floristic variation than just that of the species used in the study.

The economical feasibility of the procedures used in this study in tropical rain forests depends chiefly on the extent of field sampling needed. Field sampling can be made more effective by limiting sampling to floristic indicator taxa (Ruokolainen et al. 1997, Higgins & Ruokolainen 2004). When reliable and rapid field inventory methods are used with remote sensing, the time and effort needed for surveying floristic variation in rain forests can be significantly reduced. In Papers I-III field sampling was limited to one or two indicator plant groups (the pteridophytes and Melastomataceae species). Even if the species were growing in the understorey, they indicated the patterns of canopy trees to a degree allowing floristic variation to be distinguished and mapped by means of remote sensing.

CONCLUSIONS

The main finding of my thesis is that floristic variation in lowland rain forest, expressed in the form of ordination axis scores, vegetation classes or numbers of species in ecological categories, can be predicted relatively accurately on the basis of the spectral values of a medium-resolution Landsat TM or ETM+ image. It was possible to distinguish and map floristic variation within *tierra firme* forest, which has been traditionally seen as a single broad forest type (Pires et al. 1985). We were able to divide *tierra firme* into two forest types, harbouring different sets of species. Moreover, we managed to predict similarity in species composition as a continuous, easily mapped three-dimensional variable (scores of three ordination axes). It is, however, apparent that many of the details of floristic differences observed in the field cannot be distinguished by remote sensing. The species richness of taxonomic groups was not predicted accurately, at least with the data and the methods used.

The k-nn estimation method proved to perform well in predicting continuous floristic variation and in distinguishing between forest types. In the case study of Río San Juan it proved to have higher classification power compared to linear disriminant analysis. Another advantage of the k-nn method is that the effect of the artificial radiometric gradient, which

my be a problem at least in large-area mapping, can be reduced simply by employing a distance limit in searching the nearest neighbours. All four studies were conducted within a relatively small area and using relatively few field plots. If the k-nn method is employed in large-area mapping, the parameters (e.g. number of neighbours, search distance) and the features of satellite image and auxiliary environmental data used, as well as their weighting, need to be optimised. The k-nn method can be considered for practical prediction and mapping of floristic variation if field observations on species are available in feasible densities. The need for a map of floristic patterns in many tropical forest areas is urgent. Landscape-scale vegetation type maps or maps of gradual changes in species composition, expressed for instance in terms of ordination axis scores, would be most welcome even with a certain degree inaccurate predictions, as the alternative is often a pure guess.

The spectral information of the satellite image and remotely derived DEM provide highly valuable information of floristic variation at landscape scale. This information is relatively uniform in quality, spatially continuous and inexpensive - all characteristics that speak for its use. The results of this study have highlighted the importance of feature selection or weighting, the discrimination power of the MIR and NIR bands, and the use of relatively large pixel units (pixel windows or segments) in predicting floristic variation and classifying floristically defined forest types.

The studies provide further evidence for the hypothesis that environmental factors significantly control floristic patterns in tropical rain forests. The hypothesis was put to a strict test by using understorey indicator species to define floristic variation, which was then modelled by satellite image data which cannot directly observe the understorey. The modelling was successful, showing that the distributions of both understorey indicator species and canopy trees must be controlled by common environmental factors. This means that an inventory of one or two indicator taxa combined with remote sensing can successfully increase our knowledge of floristic variation in rain forest, at reduced cost and time compared to traditional methods.

REFERENCES

Anderson, A. B. 1981. White-Sand Vegetation of Brazilian Amazonia. Biotropica 13: 199-210.

Bartholome, E. & Belward, A. S. 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. International Journal of Remote Sensing 26: 1959-1977.

Boyd, D. S., Foody, G. M. & Curran, P. J. 1999. The relationship between the biomass of Cameroonian tropical forests and radiation reflected in middle infrared wavelengths (3.0-5.0 mu m). International Journal of Remote Sensing 20: 1017-1023.

Cayuela, L., Benayas, J. M., Justel, A. & Salas-Rey, J. 2006. Modelling tree diversity in a highly fragmented tropical montane landscape. Global Ecology and Biogeography 15: 602-613.

Clark, D. A., Clark, D. B., Sandoval, R. & Castro, M. V. 1995. Edaphic and human effects on landscape-scale distributions of tropical rain-forest palms. Ecology 76: 2581-2594.

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.

–, Palmer, M. W. & Clark, D. A. 1999. Edaphic factors and the landscape-scale distributions of tropical rain forest trees. Ecology 80: 2662-2675.

Colby, R. 1991. Defining and mapping vegetation types in mega-diverse tropical forests. Photogrammetric Engineering and Remote Sensing 57: 531-537.

Collett, L. J., Goulevitch, B. M. & Danaher, T. J. 1998. SLATS Radiometric Correction: A Semi-Automated, Multi-Stage Process for the Standardisation of Temporal and Spatial Radiometric Differences. Proceedings of Proceedings of the 9th Australasian Remote Sensing and Photogrammetry Conference, July, Sydney, Australia. Available at: http://www.nrw.qld.gov.au/slats/pdf/arspc9_rad_corrections1.pdf. [Cited 10 May 2007].

Condit, R. 1996. Defining and mapping vegetation types in mega-diverse tropical forests. Trends in Ecology & Evolution 11: 4-5.

–, Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., Nunez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H. C., Losos, E. & Hubbell, S. P. 2002. Beta-diversity in tropical forest trees. Science 295: 666-669.

Congalton, R. G. & Green, K. 1999. Assessing the accuracy of remotely sensed data. Boca Raton FL. 137p.

Crippen, R. E. 1989. A simple spatial -filtering routine for the cosmetic removal of scan-line noise from Landsat TM P-tape imagery. Photogrammetric Engineering and Remote Sensing 55: 327-331.

De Grandi, G. F., Mayaux, P., Malingreau, J. P., Rosenqvist, A., 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. 1996. Patterns of tree species richness in rain forests of the middle Caqueta area, Colombia, NW Amazonia. Biotropica 28: 142-158.

–, Svenning, J. & Wright, S. J. 2002. Beta diversity in tropical forests. Science 295: 636-637.

Duque, A., Sanchez, M., Cavelier, J. & Duivenvoorden, J. F. 2002. Different floristic patterns of woody understorey and canopy plants in Colombian Amazonia. Journal of Tropical Ecology 18: 499-525.

Encarnación, F. 1985. Introducción a la flora y vegetación de la Amazonia peruana: estado actual de los estudios, medio natural y ensayo de una clave de determinación de las formaciones vegetales en la llanura amazónica. Candollea 40: 237-252.

Eva, H. D., Belward, A. S., De Miranda, E. E., Di Bella, C. M., Gond, V., Huber, O., Jones, S., Sgrenzaroli, M. & Fritz, S. 2004. A land cover map of South America. Global Change Biology 10: 731-744.

Faith, D. P. & Ferrier, S. 2002. Linking species-compositional dissimilarities and environmental data for biodiversity assessment. Australian Museum Online. Available at: http://www.amonline.net.au/systematics/faith_linking.htm). [Cited 20 Sept 2007].

– & Walker, P. A. 1996a. Environmental diversity: On the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. Biodiversity and Conservation 5: 399-415.

– & Walker, P. A. 1996b. How do indicator groups provide information about the relative biodiversity of different sets of areas? on hotspots, complementarity and pattern-based approaches. Biodiversity Letters 3: 18-25.

Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? Systematic Biology 51: 331-363.

– & Guisan, A. 2006. Spatial modelling of biodiversity at the community level. Journal of Applied Ecology 43: 393-404.

– & Watson, G. 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. Australian Goverment, Department of the Environment and Water Resources. Available at: http://www.deh.gov.au/biodiversity/publications/technical/surrogates/pubs/surrogates.pdf. [Cited 20 Jan 2007].

–, Drielsma, M., Manion, G. & Watson, G. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. Biodiversity and Conservation 11: 2309-2338.

–, Manion, G., Elith, J. & Richardson, K. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Diversity and Distributions 13: 252-264.

–, Watson, G., Pearce, J. & Drielsma, M. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodiversity and Conservation 11: 2275-2307.

Foody, G. M. 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment 80: 185-201.

– & 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.

– & Cutler, M. E. J. 2006. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. Ecological Modelling 195: 37-42.

– & Hill, R. A. 1996. Classification of tropical forest classes from Landsat TM data. International Journal of Remote Sensing 17: 2353-2367.

Franco-Lopez, H., Ek, A. R. & Bauer, M. E. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment 77: 251-274.

Fuller, D. O. 2006. Tropical forest monitoring and remote sensing: A new era of transparency in forest governance? Singapore Journal of Tropical Geography 27: 15-29.

Funk, V. A. & Richardson, K. S. 2002. Systematic data in biodiversity studies: Use it or lose it. Systematic Biology 51: 303-316.

–, Richardson, K. S. & Ferrier, S. 2005. Survey-gap analysis in expeditionary research: where do we go from here? Biological Journal of the Linnean Society 85: 549-567.

Gjertsen, A. K., Tomter, S. & Tomppo, E. 2000. Combined use of NFI sample plots and Landsat TM data to provide forest information on municipality level. Proceedings of Conference on Remote Sensing and Forest Monitoring, June 1-3 1999, Rogow, Poland. p. 167-174.

Griffiths, G. H., Lee, J. & Eversham, B. C. 2000. Landscape pattern and species richness; regional scale analysis from remote sensing. International Journal of Remote Sensing 21: 2685-2704.

Guisan, A. & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

Haapanen, R., Ek, A. R., Bauer, M. E. & Finley, A. O. 2004. Delineation of forest/nonforest land use classes using nearest neighbor methods. Remote Sensing of Environment 89: 265-271.

Hernandez-Stefanoni, J. L., Pineda, J. B. & Valdes-Valadez, G. 2006. Comparing the use of indigenous knowledge with classification and ordination techniques for assessing the species composition and structure of vegetation in a tropical forest. Environmental Management 37: 686-702.

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.

Higgins, M. A. & Ruokolainen, K. 2004. Rapid tropical forest inventory: a comparison of techniques based on inventory data from western Amazonia. Conservation Biology 18: 799-811.

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

– & 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.

Howard, S. M. & Lacasse, J. M. 2004. An evaluation of gap-filled landsat SLC-off imagery for wildland fire burn safety mapping. Photogrammetric Engineering and Remote Sensing 70: 877-880.

Hubbell, S. P. 2001. The Unified neutral theory of biodiversity and biogeography. Princeton University Press. 375 p.

Ingram, J. C., Dawson, T. P. & Whittaker, R. J. 2005. Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. Remote Sensing of Environment 94: 491-507.

Jones, M. M. 2006. Mesoscale patterns in the species composition and distributions of rain forest plants. Annales Universitatis Turkuensis Serie A 2 202: 1-32.

Kalliola, R., Puhakka, M., Salo, J., Tuomisto, H. & Ruokolainen, K. 1991. The dynamics, distribution and classification of swamp vegetation in Peruvian Amazonia. Annales Botanici Fennici 28: 225-239.

Katila, M. & Tomppo, E. 2001. Selecting estimation parameters for the Finnish multisource National Forest Inventory. Remote Sensing of Environment 76: 16-32.

Kaufman, Y. J. & Remer, L. A. 1994. Detection of forests using Mid-Ir Reflectance - an application for aerosol studies. IEEE Transactions on Geoscience and Remote Sensing 32: 672-683.

Kellndorfer, J., Walker, W., Pierce, L., Dobson, C., Fites, J. A., Hunsaker, C., Vona, J. & Clutter, M. 2004. Vegetation height estimation from Shuttle Radar Topography Mission and National Elevation Datasets. Remote Sensing of Environment 93: 339-358.

Kerr, J. T. & Ostrovsky, M. 2003. From space to species: ecological applications for remote sensing. Trends in Ecology & Evolution 18: 299-305.

Kim, H. J. & Tomppo, E. 2006. Model-based prediction error uncertainty estimation for k-nn method. Remote Sensing of Environment 104: 257-263.

Kleinn, C., Corrales, L. & Morales, D. 2002. Forest area in Costa Rica: A comparative study of tropical forest cover estimates over time. Environmental Monitoring and Assessment 73: 17-40.

Kuplich, T. M. 2006. Classifying regenerating forest stages in Amazonia using remotely sensed images and a neural network. Forest Ecology and Management 234: 1-9.

Landis, J. R. & Koch, G. G. 1977. The measurement of observer agreement for categorial data. Biometrics 33: 159-174.

Legendre, P. & Legendre, L. 1998. Numerical Ecology. Elsevier. 853p.

Lillesand, T. M. & Kiefer, R. W. 2000. Remote Sensing and Image Interpretation. John Wiley & Sons. 724p.

Lips, J. M. & Duivenvoorden, J. F. 2001. Caracterización ambiental. In: Duivenvoorden, J.F. (ed.). Evaluación de recursos vegetales no maderables en la Amazonía noroccidental. Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam. p. 19-46.

Lobo, A. 1997. Image segmentation and discriminant analysis for the identification of land cover units in ecology. IEEE Transactions on Geoscience and Remote Sensing 35: 1136-1145.

Lobo, A. & Gullison, R. E. 1998. Mapping the tropical landscape of Beni (Bolivia) from Landsat-TM imagery: beyond the 'forest/non-forest' legend. In: Dallmeier, F. & Comiskey, J.A. (eds.). Forest biodiversity research, monitoring and modeling : conceptual background and old world case studies. UNESCO, Paris. p. 159-181.

Lu, D. 2005. Integration of vegetation inventory data and Landsat TM image for vegetation classification in the western Brazilian Amazon. Forest Ecology and Management 213: 369-383.

Lu, D. S., Mausel, P., Brondizio, E. & Moran, E. 2003. Classification of successional forest stages in the Brazilian Amazon basin. Forest Ecology and Management 181: 301-312.

Lucas, R. M., Honzak, M., Curran, P. J., Foody, G. M. & Nguele, 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 Iquitos, Perú. In: Kalliola, R. & Paitán, S.F. (eds.), Geoecología y desarrollo Amazónico: Estudio integrado en la zona de Iquitos, Perú. University of Turku. p. 35-57.

Margules, C. R. & Pressey, R. L. 2000. Systematic conservation planning. Nature 405: 243-253.

Mayaux, P., Bartholome, E., Fritz, S. & Belward, A. 2004. A new land-cover map of Africa for the year 2000. Journal of Biogeography 31: 861-877.

–, Holmgren, P., Achard, F., Eva, H., Stibig, H. & Branthomme, A. 2005. Tropical forest cover change in the 1990s and options for future monitoring. Philosophical Transactions of the Royal Society B-Biological Sciences 360: 373-384.

–, Richards, T. & Janodet, E. 1999. A vegetation map of Central Africa derived from satellite imagery. Journal of Biogeography 26: 353-366.

McCune, B. & Grace, J. B. 2002. Analysis of ecological communities. MjM Software Design. 300 p.

McRoberts, R. E., Tomppo, E. O., Finley, A. O. & Heikkinen, J. 2007. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sensing of Environment 111: 466-480.

Minnaert, J. L. & Szeicz, G. 1961. The reciprocity principle in lunar photometry. Astrophysics Journal 93: 403-410.

Monserud, R. A. & Leemans, R. 1992. Comparing global vegetation maps with the Kappa statistic. Ecological Modelling 62: 275-293.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. 2000. Biodiversity hotspots for conservation priorities. Nature 403: 853-858.

Nagendra, H. & Gadgil, M. 1999. Satellite imagery as a tool for monitoring species diversity: an assessment. Journal of Applied Ecology 36: 388-397.

Narendra, P. M. & Goldberg, M. 1980. Image segmentation with directed trees. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2: 185-191.

Nelson, B. W. 1994. Natural disturbance and change in the Brazilian Amazon. Remote Sensing Reviews 10: 105-125.

Nilsson, M. 1997. Estimation of forest variables using satellite image data and airborne Lidar. Acta Universitatis Agriculturae Sueciae, Silvestria 17: 1-32.

Oliveira, A. A. d. & Nelson, B. W. 2001. Floristic relationships of terra firme forests in the Brazilian Amazon. Forest Ecology and Management 146: 169-179.

Paoli, G. D., Curran, L. M. & Zak, D. R. 2006. Soil nutrients and beta diversity in the Bornean Dipterocarpaceae: evidence for niche partitioning by tropical rain forest trees. Journal of Ecology 94: 157-170.

Paradella, W. R., Dasilva, M. F. F., Rosa, N. D. & Kushigbor, C. A. 1994. A geobotanical approach to the tropical rain-forest environment of the Carajas mineral province (Amazon region, Brazil), based on digital TM-Landsat and DEM data. International Journal of Remote Sensing 15: 1633-1648.

Pedroni, L. 2003. Improved classification of Landsat Thematic Mapper data using modified prior probabilities in large and complex landscapes. International Journal of Remote Sensing 24,: 91-113.

Pekkarinen, A. 2002. Image segment-based spectral features in the estimation of timber volume. Remote Sensing of Environment 82: 349-359.

Phillips, O. L., Vargas, P. N., Monteagudo, A. L., Cruz, A. P., Zans, M. E. C., Sanchez, W. G., Yli-Halla, M. & Rose, S. 2003. Habitat association among Amazonian tree species: a landscape-scale approach. Journal of Ecology 91: 757-775.

Pires, J. M. & Prance, G. T. 1985. The vegetation types of the Brazilian Amazon. In: Prance, G. T, & Lovejoy, T.E. (eds.), Key Environments: Amazonia. Pergamon Press, Oxford. p. 109-145.

Pitman, N., Terborgh, J., Silman, M. R. & Nuez, P. 1999. Tree species distributions in an upper Amazonian forest. Ecology 80: 2651-2661.

–, Terborgh, J. W., Silman, M. R., Nunez, P., Neill, D. A., Ceron, C. E., Palacios, W. A. & Aulestia, M. 2001. Dominance and distribution of tree species in upper Amazonian terra firme forests. Ecology 82: 2101-2117.

Potts, M. D., Ashton, P. S., Kaufman, L. S. & Plotkin, J. B. 2002. Habitat patterns in tropical rain forests: A comparison of 105 plots in Northwest Borneo. Ecology 83: 2782-2797.

Powell, R. L., Matzke, N., de Souza, C., Clark, M., Numata, I., Hess, L. L., Roberts, D. A., Clark, M., Numata, I., & Roberts, D. A. 2004. Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. Remote Sensing of Environment 90: 221-234.

Pyke, C. R., Condit, R., Aguilar, S. & Lao, S. 2001. Floristic composition across a climatic gradient in a neotropical lowland forest. Journal of Vegetation Science 12: 553-566.

Ramos Bendaña, Z. S. 2004. Estructura y composición de un paisaje boscoso fragmentado: herramienta para el diseño de estrategias de conservación de la biodiversidad. MSc thesis. CATIE. 114 p.

Reese, H., Nilsson, M., Pahlen, T. G., Hagner, O., Joyce, S., Tingelof, U., Egberth, M. & Olsson, H. 2003. Countrywide estimates of forest variables using satellite data and field data from the national forest inventory. Ambio 32: 542-548.

Roberts, D. A., Keller, M. & Soares, J. V. 2003. Studies of land-cover, land-use, and biophysical properties of vegetation in the Large Scale Biosphere Atmosphere experiment in Amazonia. Remote Sensing of Environment 87: 377-388.

Ruokolainen, K. & Tuomisto, H. 1998. Vegetación natural de la zona de Iquitos. In R. Kalliola & S. Flores Paitán (Eds.), Geoecología y desarrollo amazónico estudio integrado en la zona de Iquitos, Perú. University of Turku. p. 253-365.

–, Linna, A. & Tuomisto, H. 1997. Use of Melastomataceae and pteridophytes for revealing phytogeographical patterns in Amazonian rain forests. Journal of Tropical Ecology 13: 243-256.

–, Tuomisto, H., Macía, M. J., Higgins, M. A. & Yli-Halla, M. 2007. Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? Journal of Tropical Ecology 23: 13-25.

Räsänen, M. E., Salo, J. S., Jungnert, H. & Pittman, L. R. 1990. Evolution of the Western Amazon Lowland Relief: impact of Andean foreland dynamics. Terra Nova 2: 320-332.

Sader, S. A., Waide, R. B., Lawrence, W. T. & Joyce, A. T. 1989. Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data. Remote Sensing of Environment 28: 143-148.

Salo, J., Kalliola, R., Häkkinen, I., Mäkinen, Y., Niemelä, P., Puhakka, M. & Coley, P. D. 1986. River dynamics and the diversity of Amazon lowland forest. Nature 322: 254-258.

Salovaara, K. 2005. Habitat heterogeneity and the distribution of large-bodied mammals in Peruvian Amazonia. Reports from the Department of Biology, University of Turku 53: 1- 39.

–, Cardenas, G. G. & Tuomisto, H. 2004. Forest classification in an Amazonian rainforest landscape using pteridophytes as indicator species. Ecography 27: 689-700.

Sanford Jr, R. L., Paaby, P., Luvall, J. C. & Phillips, E. 1994. Climate, geomorphology, and aquatic systems. In: McDade, L.A., Bawa, K.S., Hespenheide, H.A. & Hartshorn, G.S. (eds.). La Selva: ecology and natural history of a neotropical rain forest. University of Chicago Press. p. 19-33.

Schulman, L., Ruokolainen, K., Junikka, L., Saaksjarvi, I. E., Salo, M., Juvonen, S. K., Salo, J. & Higgins, M. 2007. Amazonian biodiversity and protected areas: Do they meet? Biodiversity and Conservation 16: 3011-3051.

Sesnie, S. E. 2007. A geospatial data integration framework for mapping and monitoring landscape diversity in Costa Rica's San Juan - La Selva Biological Corridor. University of Idaho, USA. p. 154. [Online PhD dissertation]. Available at: http://orton.catie.ac.cr/REPDOC/A1315I/A1315I.PDF.

–, Gessler, P., Finegan, B. & Thessler, S. in press. Integrating Landsat TM and SRTM-DEM derived variables for habitat classification and monitoring with decision trees in complex neotropical environments. Remote Sensing of Environment.

Shepard, Jr. G. H., Yu, D. W. & Nelson, B. W. 2004. Ethnobotanical ground-thruthing and forest diversity in the western Amazonia. Advances in Economic Botany 15: 133-171.

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

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.

Stehman, S. V. 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment 62: 77-89.

Stibig, H. -., Belward, A. S., Roy, P. S., Rosalina-Wasrin, U., Agrawal, S., Joshi, P. K., Hildanus, Beuchle, R., Fritz, S., Mubareka, S. & Giri, C. 2007. A land-cover map for South and Southeast Asia derived from SPOT-VEGETATION data. Journal of Biogeography 34: 625-637,

Tchouto, M. G. P., De Boer, W. F., De Wilde, J. J. F. E. & Van der Maesen, L. J. G. 2006. Diversity patterns in the flora of the Campo-Ma'an rain forest, Cameroon: do tree species tell it all? Biodiversity and Conservation 15: 1353-1374.

Ter Steege, H., Jetten, V. G., Polak, A. M. & Werger, M. J. A. 1993. Tropical Rain-Forest Types and Soil Factors in a Watershed Area in Guyana. Journal of Vegetation Science 4: 705-716.

–, Sabatier, D., Castellanos, H., Van Andel, T., Duivenvoorden, J., De Oliveira, A. A., Ek, R., Lilwah, R., Maas, P. & Mori, S. 2000. An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. Journal of Tropical Ecology 16: 801-828.

Thenkabail, P. S., Hall, J., Lin, T., Ashton, M. S., Harris, D. & Enclona, E. A. 2003. Detecting floristic structure and pattern across topographic and moisture gradients in a mixed species Central African forest using IKONOS and Landsat-7 ETM+ images. International Journal of Applied Earth Observation and Geoinformation 4: 255-270.

–, Enclona, E. A., Ashton, M. S., Legg, C. & De Dieu, M. J. 2004. Hyperion, IKONOS, ALI, and ETM plus sensors in the study of African rainforests. Remote Sensing of Environment 90: 23-43.

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.

Toivonen, T., Kalliola, R. & Ruokolainen, K. 2002. Gradient phenomenon in Amazonian Landsat TM images: True or illusion? Proceedings of the Second Workshop of the EARSel Special Interest Group on Remote Sensing for Developing Countries, September 18-20, Bonn, Germany. p. 43-50.

–, Kalliola, R., Ruokolainen, K. & Malik, R. N. 2006. Across-path DN gradient in Landsat TM imagery of Amazonian forests: A challenge for image interpretation and mosaicking. Remote Sensing of Environment 100: 550-562.

Tomppo, E., Gagliano, C., De Natale, F. & Katila, M. 2007. Predicting categorical variables by means of an improved k-NN, Manuscript, Finnish Forest Research Institute.

Tomppo, E. 1991. Satellite image-based national forest inventory of Finland. International Archives of Photogrammetry and Remote Sensing 28: 419-424.

– 1996. Application of Remote Sensing in Finnish National Forest Inventory. Proceedings of International workshop of application of remote sensing in European forest monitoring, 14th-16th October, Vienna, Austria. p. 377-388.

– & Halme, M. 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-N-N estimation: a genetic algorithm approach. Remote Sensing of Environment 92: 1-20.

–, Goulding, C. & Katila, M. 1999. Adapting Finnish multi-source forest inventory techniques to the New Zealand preharvest inventory. Scandinavian Journal of Forest Research 14: 182-192.

–, Korhonen, K. T., Heikkinen, J. & Yli-Kojola, H. 2001. Multi-source inventory of the forests of the Hebei Forestry Bureau, Heilongjiang, China. Silva Fennica 35: 309-328.

–, Olsson, H., Sthål, G., Nilsson, M., Hagner, S. & Katila, M. in press. Creation of forest data bases by combining national forest inventory field plots and remote sensing data. Remote Sensing of Environment.

Trigg, S. N., Curran, L. M. & McDonald, A. K. 2006. Utility of Landsat 7 satellite data for continued monitoring of forest cover change in protected areas in Southeast Asia. Singapore Journal of Tropical Geography 27: 49-66.

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

Tuomisto, H. & Ruokolainen, K. 2006. Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. Ecology 87: 2697-2708.

– & Poulsen, A. D. 1996. Influence of edaphic specialization on pteridophyte distribution in neotropical rain forests. Journal of Biogeography 23: 283-293.

–, 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.

–, Ruokolainen, K. & Yli-Halla, M. 2003. Dispersal, environment, and floristic variation of western Amazonian forests. Science 299: 241-244.

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

–, Poulsen, A. D., Ruokolainen, K., Moran, R. C., Quintana, C., Celi, J. & Canas, G. 2003b. Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. Ecological Applications 13: 352-371.

–, Ruokolainen, K., Poulsen, A. D., Moran, R. C., Quintana, C., Canas, G. & Celi, J. 2002. Distribution and diversity of pteridophytes and Melastomataceae along edaphic gradients in Yasuni National Park, Ecuadorian Amazonia. Biotropica 34: 516-533.

Valencia, R., Foster, R. B., Villa, G., Condit, R., Svenning, J. C., Hernandez, C., Romoleroux, K., Losos, E., Magard, E. & Balslev, H. 2004. Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. Journal of Ecology 92: 214-229.

Vermote, E. F., Tanre, D., Deuze, J. L., Herman, M. & Morcrette, J. J. 1997. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview. IEEE Transactions on Geoscience and Remote Sensing 35: 675-686.

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 Amazonia. Remote Sensing of Environment 87: 470-481.

Villasenor, J. L., Maeda, P., Rosell, J. A. & Ortiz, E. 2007. Plant families as predictors of plant biodiversity in Mexico. Diversity and Distributions 13: 871-876.

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

Walker, W. S., Kellndorfer, J. M. & Pierce, L. E. 2007. Quality assessment of SRTM C- and X-band interferometric data: Implications for the retrieval of vegetation canopy height. Remote Sensing of Environment 106: 428-448.

Whittaker, R. H. 1972. Evolution and measurements of species diversity. Taxon 21: 213-251.

Williams-Linera, G., Palacios-Rios, M. & Hernandez-Gomez, R. 2005. Fern richness, tree species surrogacy, and fragment complementarity in a Mexican tropical montane cloud forest. Biodiversity and Conservation 14: 119-133.

Öhmann, J. L. & Gregory, M. J. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. Canadian Journal of Forest Research 32: 725-741.