Detection and prediction of biodiversity patterns as a rapid assessment tool in the tropical forest of East Usambara, Eastern Arc Mountains, Tanzania

Nina Sengupta

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Wildlife Science

J. D. Fraser, Wildlife Science, Chair
J. B. Campbell, Geography
R.G. Oderwald, Forestry
D. F. Stauffer, Wildlife Science
J.R. Walters, Biology
R.H. Wynne, Forestry

November, 20th 2003
Blacksburg, Virginia

Keywords: rain forest, humid tropical forest, satellite image, remote sensing, rapidly collectable field data, assessment, East Usambara, Eastern Arc, Amani, Nilo

Detection and prediction of biodiversity patterns as a rapid assessment tool in the tropical forest of East Usambara, Eastern Arc Mountains, Tanzania

Nina Sengupta

Abstract

As a strategy to conserve tropical rainforests of the East Usambara block of the Eastern Arc Mountains, Tanzania, I developed a set of models that can identify above-average tree species richness areas within the humid forests. I developed the model based on geo-referenced field data and satellite image-based variables from the Amani Nature Reserve, the largest forest sector in the East Usambara. I then verified the model by applying it to the Nilo Forest Reserve. The field data, part of the Tanzanian National Biodiversity Database, were collected by Frontier-Tanzania between 1999 and 2001, through the East Usambara Conservation Area Management Program, Government of Tanzania. The field data used are rapidly collectible by people with varied backgrounds and education. I gathered spectral reflectance values from pixels in the Landsat Enhanced Thematic Mapper (Landsat ETM) image covering the study area that corresponded to the ground sample points. The spectral information from different bands formed the satellite image-based variables in the dataset. The best satellite image logistic regression and discriminant analysis models were based on a single band, raw Landsat ETM mid-infrared band 7 (RB7). In the Amani forest, the RB7-based model resulted in 65.3% overall accuracy in identifying above average tree species locations. When the logistic and discriminant models were applied to Nilo forest sector, the overall accuracy was 62.3%. Of the rapidly collectible field variables, only tree density (number of trees) was selected in the logistic regression and the discriminant analysis models. Logistic and discriminant models using both RB7 and number of trees recorded 76.3% overall accuracy in Amani, and when applied to Nilo, 76.8% accuracy. It is possible to apply and adapt the current set of models to identify above-average tree species richness areas in East Usambara and other forest blocks of the Eastern Arc Mountains. Potentially, managers and researchers can periodically use the model to rapidly assess, monitor, update, and map the tree species rich areas within the forest. The same or similar models could be applied to check their applicability in other humid tropical forest areas.

Acknowledgments

I thank my advisor, Dr. J.D. Fraser; my committee, Dr. Randy Wynne, Dr. Jeff Walters, Dr. R. Oderwald, Dean Stauffer, and Dr. Jim Campbell; and my department for their continued guidance, patience, encouragement, and support.

I extend special thanks to the following: Dr. B.R. Murphy for his consideration and support as the department head when my other dissertation project prematurely ended. My advisor Jim Fraser, for not driving an SUV; Randy Wynne, for being one of the founder members and patrons of the Community Supported Agriculture in the valley; and Pankaj Gupta for being such an effective GSA president.

My experience with the dissertation project has been nothing less than a roller coaster ride through inspiring insights, incredible bureaucracies, international politics, and unimaginable help, cooperation, and kindness from people I knew and from complete strangers. Here I list a few individuals and institutions (in no particular order) that did not only provide wind to the dissertation sail, but provided the sail itself. I thank the following for funds, resources, data, and networking support: Dr. W.A. Rogers, Dr.Gerry Hertel, Dr. Veli Pojohnen, Kathryn Doody, Albert Ntemi Salu, Alok Mallick, Bruce Watson, Keun P. Kim, Valentin Parvu, Dr. Erkki Tomppo, Hassani Maingo, Godfrey Mathews, Isa Singano (who we lost recently), Mama Mshana, Mr. C. T. Sawe, Dr. Baba Matunda, Mr. C. Dull, the Jane Goodall Foundation and Dr. Jane Goodall, United Nations Development Program/Global Environment Facility-East Africa, United States Department of Agriculture-Forest Service, East Usambara Conservation Area Management Program, Government of Tanzania, Frontier-Tanzania (for their immense help and for collecting the best field data from rainforest that I have ever worked with), CARE-Tanzania, METLA (Forest Research Institute)-Finland, EOSAT-Landsat, World Bank Graduate Scholarship Program, and Virginia Tech. Special thanks to Dr. Randy Wynne and Department of Forestry, for letting me use their laptop during my fieldwork in Tanzania.

Over the years I was inspired, challenged, and entertained by my colleagues, friends, and family here in Blacksburg and across the globe. Here I can list only a few (in no particular order): *Mejomama* (Dr. A.R. Dasgupta), *Chhotomama* (Mr. S.R. Dasgupta), Dr. Ashish Ghosh, *Meshomoshai* (Mr. D.P. Sengupta), Dr. Rauf Ali, Dr.L. Ramakrishnan, Dr. K. Anupama, *Montudadu* (Mr. K. Sen), *Babumama* (Mr.S. Dasgupta), *Chotokaka* (Mr.S. Sengupta), Dr. J-P Pascal, Dr. Tony Whitten, Dr. Terry Lehman, Dr. P. Monga, Mr.J.Ranadey, Mr.Tarun Coomar, Jay Swami, Dr.Piyush Mathur, Dr. Unna Chokkalingam, Dr. Somyadeep Dasgupta, Meeta Mehrotra & Matt Miller, Jackie & Dave Nutter, Cynthia Chapelka, Ruthie & Howard Leeb, Nancy Galloway & Andre Laporte, Peter & Pam Schmitthenner, Inga Williams, Luz Borrero-Yu, Karen Hockett, Gabriella Zilahi-Ballogh, Juan Guzman-Aranda, Julie Delabbio, Joan Marie, Nandini Basuthakur & Sumit Ghosh, Dr. D. Muchoney, Dr. D. Baker, Gayatri Sadgopan, Nayantara Nandakumar, Madhurima Gupta, Prema Gera, Gurvindersingh Bhatia, Geeta Vaidyanathan, Mr.K.C. Chatterjee, Prateep Chatterjee, Gopal, Dr. Raman Kumar, Dr. Eileen Crist, Jane Vance Siegle, Dr. Karen De Pauw, Dr. Jamie Raiser, Dr. K.C. Malhotra, Dr. R.L. Bramhachari, Dr. Tim Boyle, Dr. Herb Saterlee, *Didi* (Mithu Roy), *Bhulu* (Chandan Sengupta), my entire extended family, all the women in my life, friends, my fellow students, faculty members, staff at Virginia Tech, and numerous others. Most of all I thank my parents Dilip and Manjula Sengupta for their unconditional love and support in my ongoing effort to be myself.

I hope my current and future work can contribute towards conserving rainforests and other wilderness areas; and towards maintaining and/or regaining indigenous peoples' rights over their lands and knowledge. Lastly, I thank every tree and the amazing living creatures within rainforests (including the leeches and mosquitoes who are always after me) for allowing me the opportunities to soak my senses in their awesome grandeur from time to time.

This research is about building models (or systems);
thought this will be an appropriate reminder :

Systematize we must,
but even in making and holding the system,
we should always keep firm hold
on this truth
that all systems are
in their nature
transitory and incomplete.

Shri Aurobindo

Table of Contents

List of Figures

List of Tables

1.0 Introduction

Areas under the tropical humid forests that house two-thirds of all terrestrial species (National Academy of Sciences 1980) have been halved within the last fifty years (Myers 1984). Estimates show that between 1990, the Kyoto Protocol base year, and 1997, humid tropical forests were deforested at an annual rate of $5.8 \pm 1.4 \times 10^6$ hectares (Achard et al. 2002) or higher (Food and Agriculture Organization 2001). This deforestation has led to a marked reduction in open and dense natural forests. In the face of such rapid changes, biodiversity conservation in tropical forests requires the following four key skills: to quickly and accurately delineate existing forests, to assess changes in their cover over time, to identify biodiversity patterns within forests, and to understand underlying ecosystem functioning for early detection of ecosystem responses to natural or human-induced alterations.

Improving the ability to precisely differentiate between forested and non-forested areas, using remote sensing data, at various scales is still one of the major research issues in both tropical and temperate regions (e.g. Wolter et al. 1995; Jeanjean and Achard 1997; Mayaux and Lambin 1997; Shimabukuro et al. 1998; Pelkey et al. 2000; Blinn et al. 2002; Sgrenzaroli et al. 2002). To detect changes in forest cover quickly it is necessary to delineate forest cover over time. Only a few studies have estimated forest cover changes; these have had varied, and even disputed, results (Hays and Sader 2001; Stokstad 2001). Jenkins et al. (2003) emphasized the need for a program to regularly and consistently document changes in biodiversity and natural habitats. While research to delineate forests, identify different types of forests, and detect changes in forest covers over time is continuing, scientists and managers are focusing on

detecting and understanding the underlying causes of the spatial patterns of species distribution within forests. This, in turn, will aid precision forestry operation and conservation (e.g. Conduit 1996; Trisurat et al. 2000; Franco-Lopez et al. 2001).

A continental or regional scale study revealed a high congruence of species-rich areas of different taxa (Myers et al. 2000). The authors and their colleagues at Conservation International (CI), a leading Non Government Organization (NGO), defined the term "hotspots" for areas that host exceptionally high concentrations of endemic species (of primarily plants) but have experienced massive habitat loss (Myers 1988; Mittermeier et al. 1998; Myers et al. 2000). Each hotspot contains at least 0.5% or 1,500 of worlds known plant species as endemics (Mittermeier et al. 1998; Myers et al. 2000). The scientists identified 25 such hotspots and recommended that they be conserved on a priority basis (Myers et al. 2000). The majority of these hotspots, including the Eastern Arc Mountains in Tanzania, are tropical rainforests. Together, the hotspots include 44% of all known vascular plants, and 35% of all known species of mammals, birds, reptiles, and amphibians (Myers et al. 2000). High biodiversity areas have been prioritized for conservation by other leading groups, including the World Wide Fund for Nature (WWF) (Olson and Dinerstein 1998), Birdlife International (Bibby 1998), the World Conservation Union (IUCN) (WWF & IUCN 1994 -1997), the World Resources Institute (WRI) (Ayensu et al. 1999), and The Nature Conservancy (TNC) (TNC 1997; Groves et al. 2000). Broadly, the high biodiversity areas identified by these different conservation groups match with each other and with Myer et al.'s (2000) hotspots (da Fonseca 2000; Redford et al. 2003).

However, most conservation decisions are made at finer geographic scales than hotspots or eco-regions (Mace et al. 2000). In addition, there is a need to develop and adopt techniques that allow assessment both at finer and larger scales. Therefore, it is necessary to develop scale-independent techniques that allow researchers and managers to assess biodiversity patterns within hotspots (Mace et al. 2000). Such techniques need to be procedurally simple for easy replication within or outside protected areas, especially where trained personnel and/or recent baseline data are rare.

The assessment of biodiversity patterns in rainforests would be enhanced if, at local scales species-rich areas of a few indicator taxa typically overlapped with areas of rare and endemic species of their own taxonomic group and high biodiversity areas of other taxa. Unfortunately, there typically is little or no such overlap (Burbidge et al. 1992; Prendergast et al. 1993; Lawton et al. 1994; van Jaarsveld et al. 1998). The exception may be in tropical forests acting as refugia, which may show relatively high overlap in species-rich or endemic species-rich areas. Further, both primary and secondary habitats of an indicator taxon together include the rarity/endemic areas of that taxon and high diversity areas of other taxa (Williams et al. 1996; Balmford et al. 1996*a*; Balmford et al. 1996*b*; Balmford and Gaston 1999). However, due to widespread habitat loss, many species currently exist in the peripheries of their original ranges (Lomolino and Channell 1998; Brooks 2000; Channell and Lomolino 2000). These peripheral habitats may actually be secondary habitats to some species. Determining an appropriate indicator taxon (or taxa) and identifying all primary and secondary habitats for the selected taxon (often from existing data collected for very different purposes) may be the only widely available practical means to identify high-biodiversity areas for conservation (Balmford 1998; Howard 1998).

However, mapping current ranges of species is itself difficult (Gaston 1994). Even when possible, determining and mapping primary and secondary habitats of indicator taxa are not quick or easy tasks, especially within rainforests.

In tropical forests, trees provide habitats and/or resources for almost all other forest species (e.g. Hall and Swaine 1976; Huston 1994; Whitmore 1998). Therefore, the development of a simple model that could roughly suggest the locations of tree species-rich areas within rainforests may be a first step toward helping managers and researchers make more informed decisions about conservation of trees and the forest species that closely associate with them. The process of identifying tree species-rich areas can be rapid if the models are able to use satellite image-based variables, possibly integrated with some rapidly collectable field data. Formulating such a model or a set of models is my primary goal in this dissertation.

If satellite images could be used in formulating models, it would substantially reduce time for forest assessment. Research that uses remote sensing or satellite images in tropical forest areas at local scale often classify areas under forest into several habitat classes (Hill 1999; Muchoney et al. 2000; Trisurat et al. 2000; Roy and Joshi 2002). Many also correlate the classified habitat or forest categories to the abundance of certain species, groups of species, or taxa of interest (Raven et al. 1995; Nagendra and Gadgil 1999). In doing so, such studies use single, multi-source (optical or radar), multiple resolution, or multi-temporal remote sensing data in their analyses. Often the spectral values of multiple bands or band ratios (simple ratio, or different types of vegetation indices), used in further analysis, can detect parameters of interest. Supervised classification (Nagendra and Gadgil 1999), neural network analysis (Carpenter et al.

1999; Pax-Lenney et al. 2001), cluster analysis (Lewis 1994), and image segmentation (Shimbukuro et. al. 1998), have been used for model building, mapping and/or classification.

Canopy reflectance in satellite images over tropical forest areas may show little visual variation in reflectance, even though the habitats within the rainforests are patchy (Nee and May 1992; Pacala and Tilman 1994; Tuomisto et al. 1995; Whitmore 1998). Moreover, variations that do appear in canopy reflectance may be suppressed while detecting larger scale patterns or characteristics, like delineating forest cover or forest types at a regional, continental, or global scale. The accuracy of large-scale classification in the humid tropics improves with low-pass filtering or segmentation (Hill and Foody 1994; Hill 1999), which reduces the class variance in images (Hill and Foody 1994). [Both filtering and segmentation are processes for separating images to component regions, thus removing noise or variance]. Therefore, in a classification approach, to some degree, thematic resolution, (i.e. precision of the measurements or categories for a particular theme; Veregin 2001), is lost in the filtering or segmentation process (Rajaniemi et al. *in press*). In addition, the processes that differentiate the habitat classes assume such classes to be internally homogenous; this is rarely the ground reality in tropical rain forests. Even if the image resolution and/or habitat permits differentiation into classes, like biophysical features, species composition, species richness etc., a loss of variance may make such identification difficult at a finer scale.

Studies using a classification approach often try to associate habitat types to the abundance or occurrence of certain taxa or species. They may not be always successful for two reasons; first, the habitat characteristics that determine the distribution or abundance of the concerned species

may not be detectable through low or medium-resolution satellite data. Conversely, the vegetation gradients that satellite image analysis can detect may not influence the abundance or richness of the taxon or particular species of interest. At a finer scale, a species of interest in a location may be associated with the dominant or detectable vegetation gradient of that location; whereas in another location it may not. Therefore, one can study and assess the same species or its habitat using satellite images in one location but not in another (Lavers and Haines-Young 1997). To map a species distribution pattern only over large areas (regional, continental, or global scale), delineating forest types based on satellite image and ancillary data and then associating species to those habitats can be useful tools. However, these techniques are not very useful in detecting variations of vegetation or species distribution at finer scales.

Recently, comparatively simple "classification-free" approaches directly used pixel reflectance from original or derived satellite image (i.e. ratios, standard deviation, texture, and mean of band reflectance over pre-determined kernel sizes). This approach was used to derive biophysical parameters and land covers (Muchoney et al. 2000), and to predict distribution of single species (Lavers and Haines-Young 1997; Scharier 1999) or groups of species (Jakubauskas and Price 1997; Rajaniemi et al. *in press*). Rajaniemi et al. (*in press*) successfully formulated a satellite-based non-parametric classification-free model, using the K-nearest neighbor (KNN) analysis, to predict species number for certain groups of small trees/shrubs (*Melastomataceae*), and ferns (*Pteridophytes*) in the Amazon.

Classification-free approaches, such as the one used in this research, directly use the pixel reflectance without removing any existing noise or variance within satellite images. Such noise

or variance could be indicative of species richness, structure, or composition within forests. Variation in satellite image reflectance could be associated with variation in forest productivity, phenology, and dynamics (Conduit 1996; Duivenvoorden 1998); or they could correlate with plant species composition (Ruokolainen and Tuomisto 1998 referred to in Rajaniemi et al. *in press* and Foody and Cutler 2002). Possibly, variances in the satellite band reflectance over humid forests represent differences in biodiversity, species richness, forest structure, species composition, edaphic/environmental characteristics, species interaction, and other chaotic fluctuations that lead to patchiness within rainforests (Condit 1996). Some of these underlying parameters may have greater influence in some humid forests, in certain areas within such forests, or in certain seasons, than in other areas or time.

Here, I have used a classification-free approach to evaluate the use of remotely- sensed data, data collected on site within rainforests, and several analytical techniques, to identify areas of above average tree species richness, in East Usambara, Tanzania.

2.0 Research Objectives

The aim of this research was to formulate a set of models that will enable managers and scientists to rapidly detect areas with above-average tree species richness within humid tropical forests. The objectives of my research were:

1. To develop a model to detect spatial patterns of above-average tree species richness within the Amani Nature Reserve using satellite image records from pixels corresponding to geo-referenced field plots.
2. To develop models using satellite image-based variables and rapidly collectable ground data from the geo-referenced field plots to determine if the addition of field data can improve models for the Amani Nature Reserve
3. To test the predictive models derived for the Amani Nature Reserve by applying them to the Nilo Forest Reserve.

Ideally, when models from this research are further developed in the future, they can be applied to all forest sectors within East Usambara, and ultimately, to other forest blocks within the Eastern Arc Mountains. The managers and researchers would be able to use the models to assess and monitor forests and should be able to improve them over time by incorporating parameters presently unavailable. In addition, through time series application of the models, researchers or managers could detect changes in forest cover or characteristics, identify early signs of threats to species or groups of species, decide where further interventions and/or research would be necessary, and map to communicate changes to stakeholders.

3.0 Study Area

The humid tropical forests of East Usambara in Tanzania consist of nine to fourteen small forest sectors. Two of these, the Amani Nature Reserve (Amani N.R.), and the Nilo Forest Reserve (Nilo F.R.) are my study areas. Together, these sectors form the East Usambara Forest Block. Eleven such distinct forest blocks, extending from Udzungwa in Tanzania in the south to Taita Hills in Kenya in the north, form the Eastern Arc Mountains (Figure 1). Eastern Arc is one of the 25 global biodiversity hotspots identified by Myer et al. (2000).

3.1 Eastern African tropical forests and Eastern Arc

The tropical moist forests of eastern Africa cover about 10,000 km^2, and are smaller than similar habitats in western Africa. Geologically, the eastern tropical forests fall into three categories: (1) coastal forests on sedimentary rock, which lie between the eastern arc and the coast; (2) forests of more recent origin, which lie atop the volcanic rock; and (3) the oldest of all, the Eastern Arc forests on the ancient crystalline mountains (Wasser and Lovett 1993). Rainforests in East Africa have existed for more than 30 million years, though they receded due to widespread habitat changes during the major extinction spell in the Pleistocene age. Parts of the east-African moist forest persisted due to the stable Indian Ocean currents, which brought rains (Hamilton 1982). The rainforests continued to exist in the Eastern Arc Mountains as discrete blocks of

naturally-fragmented forest refugia located in areas with high rainfall, and surrounded by arid, low rainfall areas (Wasser and Lovett 1993).

Long isolation and fragmentation have resulted in high levels of endemism and diversity in the Eastern Arc forests (Lovett et al. 2000). Both paleo- and neoendemic species thrive in the east African rain forests, with most of the paleoendemics (endemic ancient species with few or no close relatives) occurring in the Eastern Arc. The Eastern Arc covers 5,340 km^2 (Newmark 1998). Though it comprises only 0.005% of the mainland area (940,295 km^2), it hosts18% of all plant species found in mainland Tanzania (Newmark 1999). About one third of the moist forest plant species found there are endemic (Iversen 1991). Similar or greater endemism exists in other taxonomic groups (Hoffman 1993; Howell 1993; Scharff 1993; Newmark 1999).

Human beings have lived in the Eastern Arc Mountains for more than 2,000 years (Schmidt 1989). Anthropogenic activities in the last 200 years have lead to the loss of approximately 77% of the natural forest cover (Newmark 1998). Presently, the estimated forest cover for the Eastern Arc is only 2000 km^2 and, compared to its original cover, it is very fragmented (Myers et al. 2000). The Eastern Arc is the smallest of the 25 biodiversity hotspots identified by Myers et al (2000). The forests still host an exceptionally high endemism with 1500 endemic plants and 121 endemic vertebrates (Myers et al. 2000). Because the Eastern Arc covers a small area, any further habitat loss may cause serious threat to the forest's ability to host its rich biological diversity (Brooks et al. 2001).

3.2 East Usambara Forest

The East Usambara forest block of the Eastern Arc Mountains ($4^o48'$-$5^o13'$S and $38^o32'$-$38^o48'$E) covers 413 km^2 and contains an estimated 221 km^2 of closed forest (Huang et al. 2003; Newmark 1998). Compared to other African rainforests, the Eastern Arc forests, and especially the East Usambara forests, have more species with range size rarity (i.e. species with restricted geographic range) and near-endemic species (i.e. endemic within certain narrow ecological zones). Of the tree species recorded in the East Usambara forests, many (58%) are represented by ≤ 3 individuals (Huang et al. 2003). Amani N.R. and Nilo F.R. are, respectively, the first and the second largest forest sectors within the East Usambara (Figure 2).

Scientific recognition of the uniqueness and high diversity of the humid tropical forests in East Usambara started early. A botanical survey in the late 1890s, an amphibian study in 1928, and a survey on avifauna in1930s are some examples of scientific research in this area (Beharrell et al. 2002). Even though considerable biological information existed from the East Usambara, especially from the Amani area, historically systematic surveys were few and/or were not readily accessible to researchers and managers. In this regard, the East Usambara was like most other humid tropical forests.

Between 1954 and 1978, 50% of East Usambara's mature forests were cleared (Rodgers and Homewood 1982; Rodgers 1993). Much of the remaining forests in the East Usambara continue to be in various stages of use, degradation, or conversion (Newmark 1998; Seddon et al. 1999).

3.3 Amani Nature Reserve and Nilo Forest Reserve

In 1997, the Tanzanian government formed the Amani Nature Reserve (Amani N.R.) out of six former forest reserves (Amani Sigi, Amani East, Amani West, Kwamsambia, Kwamkoro, and Mnyuzi), public land, and land donated by a tea plantation. Together, these covered 83.8 km^2 (5° $1^{'}$ $410^{"}$-5° $04^{'}30^{"}$ S and 38° $30^{'}$ $34^{"}$-38° $40^{'}06^{"}$ E). The lowland and submontane forests of the Amani N.R. occur between 190m and 1130m above the mean sea level (Doody et al. 2001). The Nilo Forest Reserve (Nilo F.R.), registered in 1999, comprises 60.25 km^2 (4° $50^{'}$ - 4° $59^{'}$S and 38° 37' - 38° $41^{'}$ E), and hosts submontane, montane, lowland, and riverine forests. Nilo F.R.'s elevation ranges from 400m to 1,506m above the mean sea level, with the 1,506m being the highest peak in East Usambara (Beharrell et al. 2002). Unlike most of Amani N.R., the general topography of Nilo F.R. is steep, scarped, and dissected (Seddon et al. 1999). Even though, the two forests are physiologically distinct from each other, both record high diversities in several species groups. Of the East Usambara forest sectors, Amani N.R. records the highest diversity in several taxonomic groups, followed by Nilo F.R. (Emberton et al. 1997; Doody et al. 2001; Beharrell et al. 2002).

These forests are also important as water catchment areas (Rodgers 1993; Beharrell et al. 2002). Amani is the source of water for Sigi, the major river in East Usambara, and Nilo F.R. supplies the Hundu, Bombo, and Muzi rivers. These rivers, and the numerous streams within both reserves, serve the forest flora and fauna, human settlements, and agriculture nearby (Doody et al. 2001; Beharrell et al. 2002).

As the habitat, loss continues in East Usambara's forest sectors, urgent conservation intervention is necessary for the Amani N.R., Nilo F.R. and other forest sectors of East Usambara, and the Eastern Arc. The United Nations Development Program/Global Environment Facility (UNDP-GEF) has initiated a conservation, management, and sustainable use effort in the entire Eastern Arc Mountain area (pers. comm. Dr. W.A. Rogers). Current research techniques, if adopted, may help in that conservation effort.

4.0 Methods

I used a "classification-free" approach to create models to identify high tree diversity areas within the tropical humid forests of East Usambara, Tanzania. In this classification-free approach, assessment of above-average or average/below average diversity does not depend upon classifying vegetation into habitat types (Muchoney et al. 2000; Rajaniemi et al. *in press*). I created two sets of models; one set used only satellite image band information as predictor variables. The other set used both satellite data and rapidly collectable field data. Since I could map the satellite image-based models, I ensured those models primarily predicted above average tree species richness areas within forests, in or outside protected areas, rather than within non-forests.

High diversity areas for various taxonomic groups have been defined differently (Thomas and Malorie 1985; Prendergast et al. 1993; Lawton et al. 1994) and arbitrarily (Williams et al. 1996) in literature to suit study objectives and data. In current research, mean tree species count in 50m X 20m (or 0.1 hectare) plots was 15.1 in Amani N.R. and 11.6 tree species in the Nilo F.R. Therefore, I defined high diversity locations in the Amani N.R. as areas that host ≥ 16 tree species / km^2, and in the Nilo F.R. as areas that host ≥ 13 species /km^2 (unless specified otherwise). This allowed the datasets to split in similar sized groups of above-average and average/below average tree species richness groups, which allowed further analysis and comparison. In this document, I have used the terms "high diversity", "above average diversity", or "above average tree species richness" synonymously to indicate areas of above average tree

diversity. I have referred to the areas that did not qualify as high diversity as "not-high diversity" areas or "average/below-average diversity" areas.

4.1 Data Source, Collection, and Preparation

Database

The field data that I used for my research came from a database created by the East Usambara Conservation Area Management Program (EUCAMP), a collaborative effort by the Tanzanian Forestry and Beekeeping Division of the Ministry of Tourism, Natural Resources, and Environment and its various partners (Beharrell et al. 2002). EUCAMP commissioned forest surveys in several East Usambara forest sectors in July 1995 to address the need for a scientifically collected dataset, to initiate an ongoing assessment and monitoring program in the area, and to train local personnel in data collection and assessment procedures. As the lead group, Frontier-Tanzania (a U.K. based conservation NGO), conducted surveys in the East Usambara forests in collaboration with the University of Dar es Salaam, the Tanzania Forestry Research Institute staff, and a wide network of international taxonomists and experts (Beharrell et al. 2002).

In the Amani, Catchment Forest Officers, and villagers from Maramba, Tanga, and Kiswani assisted the Frontier-Tanzania survey team. In the Nilo, the team included EUCAMP forest officers, overseas volunteers, and residents of the nearby settlements of Kwamkole, Kizara,

Maramba, Kuze, Amani, and Tanga. All survey records were entered in a biodiversity database (Microsoft Access 1995) commissioned by the EUCAMP, and linked to the Tanzanian National Biodiversity Database. Soon, some of it will be available on the World Wide Web for scientific research.

Field Data

Field Data Collection

The Frontier-Tanzania staff conducted surveys in the Amani N.R. between January 1999 and March 2000, and in the Nilo F.R. between June 2000 and March 2001. In each reserve, there is at least one known position (latitude/longitude) marked with a boundary stone. The most accessible of the marked boundaries was the starting point for transects. The survey staff drew reserve boundaries on a topographic sheet and/or on a digital map relative to the marked boundary in the field. They marked transects from border to border, East to West, and tagged every 450m on the ground. They then created a grid, in which each cell measured 450m (E-W) X 900m (N-S) (Figure 3). They established one vegetation plot 50m (E-W) X 20m (N-S) at the southeastern corner of each grid cell.

Within each vegetation plot, the team recorded, tagged, and identified trees that were ≥10cm diameter at breast height (dbh). Botanists from the Tanzanian Forestry Research Institute (TAFORI) and, when required, international experts, identified the species. They also visually estimated canopy cover (0-50%, >50%), shrub cover (0-50%, >50%), ground cover (0-50%, >50%), canopy height (≤30m, >30m), and water association (presence of water i.e. river, lake,

steam, lake/pond, marsh/swamp, dry river bed, other vs. no-water). Altitude (m) was measured with an altimeter and slope (degrees) was measured with a clinometer at the southeast corner of each vegetation plot. The survey team established 3m X 3m grids at the center of the vegetation plots, where they counted seedlings, and also visually estimated percent herb cover, percent leaf/vegetative litter cover, and percent of bare soil (pers. comm. Kathryn Doody, Albert Ntemi, Mama Mshana 2001-2003).

In the Nilo F.R. forest, the coordinates of all 69 plots that I used were recorded with Geographical Positioning System (GPS) receivers; only six plots were so recorded in the Amani N.R. In 2001, when I visited Amani N.R., I recorded the locations of 30 additional plots with a GPS unit.

Field Data Preparation

I accessed all vegetation plot-related data for Amani N.R. and Nilo F.R.from the database and deleted any incomplete records.

If GPS records were available, I included them as geographic locations; if not, I calculated the geographic position of each plot relative to two or more plots with known GPS readings and positions on the topographic map. I first tested this procedure by calculating locations for a few plots for which I had the actual GPS records. I compared the GPS and calculated positions to see if they matched well. Then I applied the procedure to plots without GPS records. Next, I converted all GPS and calculated coordinates to WGS 84 (37S) to match with the satellite image projection. The GPS records may have a 10m inherent error. To ensure that the coordinates

represent a location truly within the 50m X 20m plot, I calculated and used the coordinates for the centers of each 50m X 20m plot (Figure 4).

I validated the species records from the database to eliminate errors due to typing, spelling, and the use of different nomenclature for the same species. To minimize such errors, I checked the records using several available floras for the region (Mabberley 1997; Schulman et al. 1998; Flora of Tropical East Africa 1952-1989). I then had the dataset verified by two experts (pers. comm. Dr. Veli Pohjonen, EUCAMP 2002; Dr. D. Porter, Virginia Tech 2002).

Satellite Image Data

Satellite image source

I used a single geocoded Landsat Enhanced Thematic Mapper (Landsat ETM) scene (Path: 167/ Row 63; acquired on December 20th 1999; roughly covering 183km X 170km; projection WGS 84 37S) obtained from United States Department of Agriculture-Forest Service (Figure 5). Landsat ETM is an optical satellite image that has a resolution of 30m X 30m and records seven multi-spectral bands and one panchromatic band. Of the multi-spectral bands, the near-infrared band 4 (0.76-0.9 μm) and mid-infrared bands, 5 (1.55-1.75 μm) and 7 (2.08-2.35 μm) often are used singly or in combination with each other or bands 1, 2, and 3, for vegetation, soil, mineral analysis, and for calculating vegetation or moisture indices (Fiorella and Ripple 1993; Carranza and Hale 2002; Wilson and Sader 2002).

Satellite image data preparation

I used Erdas Imagine Software (ERDAS Imagine 8.5, 2001) for all satellite image analysis.

The satellite image contained some cloud and cloud-shadow areas within the study area. I eliminated the cloud and cloud-shadow areas in two stages. In the first stage, I performed an unsupervised classification of the available "raw" imagery using the ISODATA (Interactive Self-Organizing Data Analysis) technique (Jensen 1996; Erdas Imagine 8.5, 2001) to classify most cloud and cloud-shadow areas into distinct classes. I reclassified those classes to zero, and formulated an image where the reclassified pixels remained zero but all other pixels retained their original band information. I refer to this resultant image as the "original" image.

In the second stage, I tried isolating and eliminating the cloud and cloud-shadow areas that persisted in the original image by using a "region grow" algorithm, in which, I designated a cloud-covered or cloud-shadow pixel as a training or reference pixel. The algorithm designates an area adjoining that training pixel (geographical distance 1 to 5 pixels) spectrally nearest (spectral distance zero to 1) to the training pixel. After ensuring all areas demarcated by the algorithm included only cloud, haze, or cloud-shadow areas I eliminated them from the image using the same method that I used in the first stage. The "final image" that I used for further analysis had little or no cloud shadow areas remaining.

I created a vector layer with the geographic/GPS plot locations, converted it to WGS 84 37S to match with the satellite image projection and then overlaid it to the final image (ArcView GIS 3.2 1992-1999) . From the vector layer and the dataset, I eliminated all plots located within the data gaps caused by the cloud and cloud-shadow removal (Figure 6).

I created six more satellite images using the final image: three "mean" and three "texture" (i.e. variance) images, with window sizes 3 X 3, 5 X 5, and 7 X 7 pixels. I calculated the mean and variance for each pixel by considering the original value of the pixel and the values of its neighboring pixels in the final image. The total number of pixels included in the calculation depended on the kernel or the window sizes. For example, a 3 X 3 window placed the pixel for which the calculation was being done at the center, surrounded by its eight neighboring pixels. For each image, in the initial dataset, I included the brightness values of all bands (band 1, 2, 3, 4, 5, and 7) from pixels corresponding to the ground sampled plots, for each image.

Variable Selection

Because there were many variables, I screened them before initiating model fitting. I followed the variable selection procedure suggested by Hosmer and Lemeshow (2000). Initially, I selected field or raw satellite image-based variables for further analyses if their values differed ($p \leq 0.05$) between the above and average/below average tree diversity groups in the Wilcoxon Signed-Rank Test (for continuous variables) or Chi-Square tests (for categorical variables). I also selected other derivative satellite image-based variables, for example bands derived by calculating mean over a 3X3 pixel window, if their p-values were ≤ 0.0001 (Table 1).

Of the field variables, I eliminated those that were measured by approximation or were categorical (canopy height and percent litter cover). I retained tree density, seedling frequency, and altitude in the final dataset for further analysis.

With the aim to keep the number of variables to a minimum, before further analysis, I created an interim dataset with only the satellite image-based variables, to select a subset for the final dataset. In addition, I calculated two Normalized Difference Moisture Indices [NDMI = (near infrared band – mid infrared band)/ (near infrared band + mid infrared band)] using near infrared band 4 (RB4) and mid-infrared band 7 (RB7), then with RB4 and RB5. I included both NDMIs in the interim dataset (Table 2).

In the interim dataset, if two satellite image-based variables were highly correlated ($r \geq 0.90$), I kept the one with lowest p-value (Table 1). If their p-values were same, I kept the red or infrared band based parameters, since they are often used in vegetation, soil, and moisture analysis (Fiorella and Ripple 1993; Wilson and Sader 2002; Carranza and Hale 2002). Accordingly, I eliminated RB2, RB5, and band 2 of the mean image calculated over 5X5 pixel window (M55B2) (Table 2).

I used logistic regression analysis with stepwise selection, to select the most predictive of the remaining variables. Since the satellite-image based variables are often highly correlated ($r \geq 0.80$), some predictive variable may not be selected in presence of other correlated variables. To find the most predictive of the satellite variables, I used the logistic regression analysis several times, each time eliminating the most predictive variable(s) selected in the previous run. I selected the satellite variables for the final dataset based on three criteria, which I applied sequentially with increasing weight. The first criterion was the percent of above average tree diversity areas in the dataset correctly predicted by the model. The second was the nature of the

variable. If both raw bands and derived (window) variables had similar predictive value, I used the raw band variable. The third and most important criterion was the model's ability to predict above-average species richness areas that actually fall within humid tropical forests (in or outside protected areas) rather than within other land use areas.

Stepwise selection first selected visible blue raw band 1 (RB1) as the best predictor in presence of all other variables (Table 2). Since the raw band was selected as the best predictor, I eliminated the mean bands (calculated over various window sizes) from the subsequent analyses. In subsequent steps, RB1, visible red raw band 3 (RB3), and mid infrared band 7 (RB7) were selected as the best predictors (Table 3). I projected each model on the satellite image to finally test whether they predict areas within humid tropical forest or outside (Figure 7). There are two marked tea-estate areas embedded within the protected area boundaries of Amani N.R.. While these areas have some remnant patches of humid forest that may host some species rich areas, and also have some shade-trees within the plantation area, most of the area is under a mono-culture of tea-bushes; and thereby distinctly different from humid-forests. There are two marked tea-estate areas in Amani N.R. surrounded by the humid forest of the protected area. I calculated the number of above-average diversity pixels predicted by each of the three models within these tea estate areas (Figure 8).

Based on the results, I included the near and the mid-infrared raw bands (RB4 and RB7 respectively) from the interim dataset in the final dataset. Even though some of the mean infrared band variables were highly correlated ($r \geq 0.80$, or their $0.05 \leq$ p-values ≥ 0.001) and were not included in the interim dataset, I included infrared mean bands in the final dataset since I

included the raw bands in the final dataset. This was to test if means calculated over various window sizes, i.e. coarser scale, were more predictive of high tree diversity areas than the raw bands. In absence of moisture data from the field, I included both moisture indices (NDMIs) in the final set.

The final data set for the Amani N.R. included the three field variables and thirteen satellite image-based variables (Table 4). I included the same set of variables, except the number of seedlings, which was incomplete, in the Nilo F.R. data set.

4.2 Data Analysis, Model Generation, and Application

First, I used only satellite image-based variables in model selection process ("satellite model"); later, I added rapidly collectable field variables, along with satellite variables ("field/satellite model") to test if the addition of field variables improved the model. Once I generated the models, I cross-validated them to test their appropriateness for the data used. To evaluate the performance of the model outside Amani N.R., I cross-validated and selected the best models, and then I applied them to the Nilo F.R (Table 5).

I used the SAS software package procedures (SAS System for Windows 8.02) to perform logistic, discriminant, and multiple linear regression analyses. I used a K-nearest neighborhood (KNN) program for the KNN analysis (pers. comm. Dr. E. Tomppo 2001-2003).

For models predicting above-average tree spices richness, I calculated Kappa statistics ($\hat{K}$). It is calculated by comparing actual agreement, i.e. agreement between model results and actual reference data, with chance agreement, i.e. if above or average/below average groups were assigned at random (Congalton et al. 1983; Verbyla 1995). [$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy), where the expected classification accuracy is chance agreement (Verbyla 1995)]. Higher and positive $\hat{K}$-value indicates stronger evidence that the model has better predictability than chance.

Predicting Above Average Tree Species Richness Areas

Logistic Regression Analysis

Based on the variable selection process with interim dataset, I had an apriori knowledge that single raw bands could be predictive of above-average tree diversity. I determined the best model from the final dataset using the Akaike Information Criteria (AIC, given by AIC = -2 ln L + 2k , where L is the maximum likelihood estimate and k is the number of independent parameters) (Burnham and Anderson 1998; Anderson et al. 2000). I first calculated the AIC for models based on each of the satellite image-based parameters. Then, in the second step, I calculated the AIC for models using each field parameter in combination with each of the eleven satellite image-based parameters. I transformed the raw AIC values to AIC weights (w_i), $w_i = \exp(-\frac{1}{2}\Delta AIC_i) / \sum_{r=1}^{R} \exp(-\frac{1}{2}\Delta AIC_r)$, where $\Delta AIC_i = AIC_i$ – minimum(AIC) and "R" is the number of models. The lower the AIC for a model, the lower are the ΔAIC_is, and the higher the AIC weights. For each individual model, the AIC weight acts as the weight of evidence in favor

or against that model; thus the higher the weight, the better the model (Burnham and Anderson 1998).

I also calculated full models, in which I forced all variables in the model, and compared it with the model that used a stepwise selection process, involving forward selection ($p \leq 0.05$) and backward elimination ($p > 0.05$). I verified each model using the "Hosmer and Lemeshow goodness of fit test" (Hosmer and Lemeshow 2000). I selected the best satellite and the best field/satellite model based on the lowest AIC (Anderson et al. 2000) and highest overall percent correct record.

In the future, data from different forest sectors in East Usambara and other areas in the Eastern Arc forests may need to be combined in some manner to generate a model for the entire area. To explore on how this might be done, I combined the two data sets (n = 187; 118 from Amani N.R. and 69 from Nilo F.R.). Here I defined above average tree species richness in two ways. In the "pooled" set, I calculated the mean of the entire dataset and defined plot with ≥ 15 tree species as high diversity. In the "combined" set, I defined as high diversity plots as those plots with $\geq$ the mean for the respective areas (≥ 16 for Amani N.R. and ≥ 13 for Nilo F.R). I performed logistic regression analyses using the pooled and the combined sets, and compared them.

Discriminant Analysis

I used discriminant analysis with the Amani N.R. data to find a model that would identify the high tree diversity areas, first using only the satellite image-based variables, and then using both the satellite and field variables. The aim was primarily to compare results with the logistic

procedure. I selected variables through a stepwise selection process with forward selection ($p \leq 0.05$) and backward elimination ($p > 0.05$). For each type of model, after variable selection, I performed the discriminant analysis and simultaneously a leave-one-out cross validation.

Predicting Species Number

I performed multiple linear regression and non-parametric K-nearest neighbor (KNN) analyses (Tomppo 1991; Franco-Lopez et al. 2001; Katila and Tomppo 2001) using the Amani N.R. and Nilo F.R. data to predict the number of tree species per plot.

Multiple Linear Regression

I fit this multiple linear regression model using the Amani N.R data (Table 2). One model used only satellite variables (satellite model) and the other used both satellite and rapidly collectable field variables (field/satellite model). I generated the models using a stepwise selection process with forward selection ($p \leq 0.05$) and backward elimination ($p > 0.05$). I evaluated the multiple linear regression models based on their significance and their R-square values, where, higher R-square values indicated better models. I also looked at the residual plots for each multiple regression analysis, to see if there were any outliers or high influence data record in each model, and whether the model changed in absence of any such outliers.

K-Nearest Neighbor Analysis

I applied K-nearest neighbor (KNN) analysis to predict species number per pixel (Katila and Tomppo 2001). The KNN analysis assumes that a strong correlation exists between the band

brightness values of the pixels corresponding to the ground sample locations and the biological and physical conditions at those locations. For each unknown pixel in the satellite image, KNN attempts to predict a value for a selected parameter based on known average values for the same parameter from K (user-specified) neighboring pixels. (Tomppo 1991; Franco-Lopez et al. 2001; Katila and Tomppo 2001).

I tried different K values from one to ten (K = 1, 2,..,10) to find the smallest number of pixels to consider in each analysis. Since the number of sample points is limited in the current dataset, I set a large geographic distance (=100km) over which the K neighboring pixels could be obtained. I performed three KNN analyses on combined Amani N.R. and Nilo F.R. data. The first two analyses generated the KNN satellite model (used only satellite image-based variables) and the KNN field/satellite model (used both satellite image-based and field variables) respectively. I used only Landsat ETM band 7 (RB7) as the predictor in the satellite model, and I used RB7 and the number of tree species as predictors in the field/satellite model. I performed a third KNN analysis where I predicted tree density per pixel using all satellite image-based variables in the dataset. Each KNN analysis involved a leave-one-out cross validation test (Katila and Tomppo 2001), where each known record was treated as unknown, and species number or tree density was predicted for later comparison with the corresponding known value.

To compare the cross-validated KNN results with results of other analyses, I grouped the KNN results for each pixel (corresponding to each plot) into high diversity and average/below-average diversity classes based on the known values from corresponding data points. In one set, I considered ≥ 15 trees per plot as high diversity; and in the second set, I treated plots from Amani

N.R. ≥ 16 tree species and that from Nilo F.R. with ≥ 13 species as high diversity areas. For both sets, I calculated the correlation between the actual and predicted species numbers for high diversity and average/below average diversity groups.

5.0 Results

5.1 Predicting Above-Average Species Richness

Satellite Image-Based Models

Logistic Regression Analysis

The model using the satellite image-based variable RB7 gave the lowest Akaike Information Criteria (AIC) value and highest AIC weight (w_i) (Table 6). Based on the highest percent accuracy (65.3%) and the lowest AIC number, the stepwise selection procedure selected the same model (Table 7). The model (Table 8) correctly identified 72.4% of the high tree diversity areas in the Amani N.R (Table 9). This prediction is 31% ($\hat{K}$) better than just random assignment of plots to above-average richness classes. Of the predicted high diversity areas identified by the model, 62.6% were correct. I projected the model on the Landsat ETM image of the area, where each pixel represents the logistic model, and the areas with high probability (i.e. $\theta = \geq 0.5$) of hosting above-average tree species richness are highlighted (Figure 9).

The model for Amani N.R. produced 62.3% overall accuracy ($\hat{K} = 0.24$). when applied to the Nilo F.R. (high diversity: ≥ 13 tree species per plot) (Table 10). Model application in the Nilo F.R. correctly identified 60.7% of the high tree diversity areas and, 53.1% of the classified high diversity areas were correct. When I defined the above average tree species richness for the Nilo F.R. based on the Amani N.R. data (i.e. ≥ 16 tree species per plot), the overall percent of

correctly classified area was 58% and $\hat{K}$ reduced to 0.14. (Table 11). The model correctly identified 56.5% of the high tree diversity areas, and 40.6% of the predicted high diversity locations actually hosted ≥ 16 tree species.

The logistic regression models with the Amani-Nilo pooled dataset (high biodiversity areas: ≥ 15 tree species/plot) (Table 12, Table 13) and the combined dataset (high diversity areas: ≥ 16 tree species/plot for records from the Amani N.R. and ≥ 13 tree species/plot for records from the Nilo F.R. records), generated similar predictability (Table 14, Table 15). Analysis with both sets, selected Landsat ETM satellite band 7 (RB7) as the best predictor variable in the stepwise selection process.

Discriminant Analysis Models

Like the logistic regression analysis (Table 9), the stepwise selection process in the discriminant analysis selected the brightness values of Landsat ETM satellite band 7 (RB7) as the best predictor variable (Table 16 and Table 17). When I applied the discriminant function (Table 16) to the Nilo F.R. area, the ability of the model to correctly identify high tree diversity areas of Nilo F.R. (Table 18) were identical to Amani N.R. logistic model application to the Nilo F.R. (Table 10).

Rapidly Collectible Field Data and Satellite Image-Based Models

Logistic Regression Analysis

The lowest AIC and highest AIC weight (w_i) indicate that the model based on the number of trees ≥ 10cm dbh (Ntree) and the RB7 is the best (Table 6). The same model was chosen by the stepwise selection procedure in the logistic regression analysis, based on highest percent

accuracy (76.3%) and lowest AIC number (134.28) (Table 7). The logistic model (Table 19) using RB7 and number of trees, could predict 81% of the high tree diversity areas in the Amani N.R. (Table 20). Of the predicted high diversity areas, 73.4% were correct. This prediction is 53% ($\hat{K}$) better than just random assignment of plots to above-average richness classes. Thus, addition of field data greatly improved the model's ability to correctly isolate the average/below average diversity areas; and marginally improved its ability to identify the high diversity areas.

I applied the Amani N.R. field/satellite model to the Nilo F.R. The overall accuracy of the Nilo F.R. field/satellite model (76.8% Table 21), was very similar to that from the Amani N.R.(76.3%, Table 20). The $\hat{K}$ also improved from 24% in the satellite model application to 49% in the field/satellite application. However, the Amani N.R. model application to the Nilo F.R., less accurately (53.5%) identified high tree diversity areas and more accurately identified the average/below average diversity areas (93%, Table 21). This improved the users' accuracy for high diversity areas (i.e. percent of predicted high diversity areas that actually were high diversity) to 83% in the field/satellite model application. Compared to the satellite image-based model (Table 18), the field/satellite model application in the Nilo F.R. (Table 21), correctly identified a lower percent of high diversity areas, and, in contrast, recorded higher percent accuracies in all other categories. The improved accuracies were due to the model's improved ability to identify average/below average diversity areas with the inclusion of the field-based predictor.

I used stepwise selection procedures in the logistic field/satellite models for the two Amani-Nilo datasets: pooled set (high diversity: ≥ 15 species/plot) and the combined set (high diversity: ≥ 16

species/plot for Amani N.R. records, and ≥ 13 species/plot for Nilo F.R. records). The stepwise procedure selected Landsat ETM satellite bands 7 (RB7), and number of trees (Ntree) as the best predictors for both sets (Table 22 and Table 24). Compared to the satellite image-based models for the pooled (Table 13) and the combined set (Table 15), the field/satellite models both improved with the inclusion of the field variable (Table 23 and Table 25).

Discriminant Analysis Models:

Initially, the variable selection process in the discriminant analysis selected the number of seedlings (Nseed) and the number of trees (Ntree) as appropriate field-based predictor variables, along with Landsat ETM band 7. However, the predictabilities of the discriminant function analysis, using only the tree density (Ntree) and the satellite image band (76.3%, Table 28), improved more than when using the number of seedlings (Nseed) also (71.1%, Table 26). The cross-validated discriminant analysis results for Amani N.R. using satellite image-based variable and the tree density (Table 28) resulted in exactly the same overall accuracy (76.3%) as the logistic regression field/satellite model (Table 20). The discriminant analysis could include 84.4% (Table 28) of the high diversity areas in the model, while the logistic could include 81% (Table 20). The logistic result, on the other hand, could ensure that 73.4% the high diversity areas designated by the model actually were high diversity, as compared to 72% by the discriminant model.

I applied the discriminant function (Table 27) from the Amani N.R. field/satellite model, which used RB7 and tree density, to the Nilo F.R. The result (Table 29) showed an improved predictability over the satellite model application to the Nilo F.R. (Table 18), primarily in its ability to identify the average/below average diversity areas. The predictabilities of the

discriminant model (Table 29) were similar to the logistic field/satellite model application to the Nilo F.R. (Table 21).

5.2 Predicting Species Number

Multiple Linear Regression Models

The stepwise selection process in multiple linear regression with only satellite image-based variables selected RB7 as the most predictive variable (Figure 10). However, the regression (p 0.001) resulted in an R^2 value of only 0.11 (Table 30). I plotted the predicted tree species number against the residual (Figure 11). It seemed one data point could be acting as an outlier exerting influence on the result (also visible in Figure 10). Amani N.R. dataset had few other similar records but they were eliminated from the field data set due to incomplete record or other reasons. However, just to test how much influence a single data point is exerting, I removed the record from the dataset and ran the multiple linear regression analysis once again. The stepwise selection process in multiple linear regression with only satellite image-based variables still selected RB7 as the most predictive variable. The regression ($p \leq 0.001$) result improved only slightly from in an R^2 value of only 0.11 in presence of the outlier to 0.14. The residual plot for the second analysis did not show any difference exc ept for absence of removed record (Figure 12).

Just to be able to compare with the results from the logistic regression (Table 9), I compared the actual tree species count per plot with the predicted number of tree species per plot generated by the multiple linear regression model using all records. I grouped into high diversity class if the

actual species numbers were ≥ 16 per plot. The overall predictability of the multiple linear regression model was 61.8% (Table 31) as opposed to 65.3% in the logistic regression model for Amani N.R.(Table 9).

While using both satellite image-based and field variables from Amani N.R., the stepwise selection procedure of the multiple linear regression chose brightness values of RB7 and the number of trees per plot as the best predictors (Figure 13). Compared to the satellite model, the results ($p \leq 0.05$) improved, to an R^2 value of 0.35 (Table 32). I plotted the predicted tree species number against the residual (Figure 14). The same data point, identified in the satellite model, seems to be a potential outlier in the field/satellite model also. To test how much influence that particular record is exerting, I removed the record from the dataset and ran the multiple linear regression analysis once again. The stepwise selection process in multiple linear regression with field and satellite image-based variables still selected RB7 and number of trees per plot as the most predictive set of variables. The regression ($p \leq 0.001$) result did not improve; the R^2 value remained at 0.35 in absence of the record. The residual plot for the second analysis also did not show any difference except for absence of removed record (Figure 15).

Once again, I compared the actual tree species numbers by plot with the predicted numbers generated by the multiple linear regression analysis using all records, to compare the result with the logistic model (Table 20), I summarized the results considering areas with ≥ 16 tree species per plot as high diversity (Table 33). The overall predictability and ability to identify high diversity areas for the multiple linear regression model was 72.8% and 74.1% respectively (Table 33) compared to 76.3% and 81% for the logistic regression model (Table 20).

K-nearest neighbor Analysis

The cross-validated K-nearest neighbor (KNN) analysis result, using RB7 and tree density as predictors, yielded a 68% overall correlation (r = 0.68) between the actual and predicted species numbers calculated by the analysis. The calculated mean of 13.875 was also close to the actual mean (Table 34). However, the primary objective of this exercise was to be able to identify or predict species numbers for areas that host above average tree species richness. The correlation between actual and predicted species number of average/below average tree diversity areas was above 60% (r = 0.69 and 0.68) for both pooled (high diversity: $\geq$15 tree species/plot) and combined (high diversity: Amani plots with $\geq$16 tree species and Nilo plots with $\geq$13 tree species) datasets. This relation was much lower for above average richness areas (Table 35). To compare with the logistic results for the pooled (Table 23) and the combined set (Table 25), I grouped the actual tree species number and their corresponding predicted numbers from the KNN analysis into above-average and average/below average diversity groups. Then from the group, I obtained the pooled-KNN (high diversity: $\geq$15 tree species/plot) and the combined-KNN (high diversity: Amani plots with $\geq$16 tree species and Nilo plots with $\geq$13 tree species) results. The overall percent of correctly classified plots in pooled-KNN is 69.5% (Table 36), which is comparable to the logistic pooled model result of 72.1% (Table 23). The pooled-KNN results showed a slight improvement in its ability to detect high tree diversity areas (74.7%), and to ensure that 74.4% of the high diversity areas identified by the model are above-average species diversity. The overall predictability of the combined-KNN model is 72.7%, (Table 37) as opposed to 75.9% by the combined logistic model (Table 25). Compared to the combined logistic model results, the KNN combined set model showed lower ability to identify the high

diversity areas and its ability of the predict high tree diversity areas (69%), and could ensure that of the predicted high diversity areas only 69.8% were correct. The two other satellite variable-based KNN, one predicting tree species numbers and the other predicting tree densities, were deemed unsuccessful as they only yielded overall correlation (r) between 0.1 to 0.3.

5.3 Models Compared

In all types of models, addition of the field variable, tree density, improved the predictability of the models compared to the satellite image (RB7) based models. However, within model classes (satellite only, field/satellite) the different statistical techniques (Logistic regression, Discriminant analysis, Multiple linear regression, and K-nearest neighbor analysis) generated similar predictions (e.g. 61.8 – 65.3 total % correct for the Amani N.R. satellite models, Table 38).

6.0 Discussion

The "classification-free" approach using Landsat ETM mid-infrared band 7 predicted above average tree species richness locations within the Amani N.R. forest with reasonable accuracy. The results of the model improved further when the number of trees, a rapidly collectable field variable, was added to the model. The results suggest that above average diversity areas can be relatively rapidly assessed and to some extent mapped, within the humid tropical forest. The models held well when applied to the Nilo F.R., another reserve within the East Usambara, suggesting the models could be applicable to other forest sectors in the East Usambara, within or outside the protected areas.

No other the classification-based or classification-free approach has reported that a single infrared (Landsat ETM) band-based model could locate above average tree species richness areas within a tropical humid forest habitat. Compared to any known available models, the current ones are simple, and require a minimum manipulation of satellite imagery. The simplicity of the model will allow managers to adopt it for rapid assessment, especially where recent baseline data are not available. Even when such data are available, the model will enable periodic updating of the data, assessing, and monitoring.

It remains to be determined if similar results can be obtained from other forest blocks within the Eastern Arc Mountain or humid tropical forests in other regions or continents. Although Landsat ETM band 7 is predictive of high tree diversity areas within forests, initially other visible bands (band 1 and then band 3) were selected by the regression analysis as predictive variables. The

underlying physical factors that drive the relationship between band and ground conditions are unknown. However, from the areas designated as above-average tree diversity in the images by different band-based models, it seems the underlying conditions that were influencing visible blue band (RB1) were very different from those influencing visible red band (RB3) or the mid-infrared band (RB7). Plant pigments primarily absorb visible lights; chlorophyll pigments are known to absorb violet-blue and red lights. There absorption changes over season and plant species (Verbyla 1995). These factors and others may have contributed to the differences in the images for various band-based models. Understanding these underlying factors influencing various bands may allow comparison of the ecology and management of areas.

Tuomisto et al. (1995) and Nagendra and Gadgil (1999) believe that edaphic conditions greatly influence the plant species composition and vegetation cover. Both the soil and vegetation characteristics, in turn, have a strong influence on variances in satellite image reflectance. Landsat ETM band 7 (RB7) also is indicative of soil moisture conditions (Verbyla 1995; Campbell 1996). While mineral imaging for hydrothermally altered rocks (like clay and iron oxide) with Landsat ETM image over heavily vegetated areas in Philippines, Carranza and Hale (2002) found band 7 (and band 5) to be indicative of clay mineral areas. Sandy-clay or clay with Ferralsols (soils enriched with oxides of iron and aluminum) characterize the soils in East Usambara forests (Hamilton 1989). Carranza and Hale (2002) also reported that the vegetation and hydrothermally altered minerals are similar in their reflectance in Landsat ETM band 7 (and band 5), posing difficulty in their use for mineral study.

Mid-infrared wavelengths also may indicate moisture conditions in general, or leaf moisture (Wilson and Sader 2002); and hydrologic conditions influence the plant cover (Silvertown et al. 1999). My informal observations on habitat and tree characteristics within the Amani N.R. suggest a strong influence of moisture condition. The undisturbed, and previously disturbed but well-recovering high diversity plots, to me, seemed distinctly more humid, moist, and had bodies of water within or adjacent to the plot. They also were typically associated with low reflectance in mid-infrared bands.

To investigate whether moisture influenced the mid-infrared values, I calculated the normalized difference moisture index (NDMI) and included it in the analysis to see if the model selection process chose it as a predictor variable. If the moisture index-variable had been selected and showed a high predictability for above-average diversity areas, it would have suggested a strong influence of the moisture regime in identifying high tree species rich areas. The model did not select NDMI. However, NDMI is perhaps not the ideal indicator of moisture regime. NDMI is called “moisture” or “wetness” index in lieu of a better term (Wilson and Sader 2002). It is sensitive to leaf water content, shadowing, and some other factors yet unidentified (Horler and Ahern 1986) and/or it represents the combined effect of differences in forest structure and water content (Cohen et al. 1995). When NDMI for a forest area was calculated and recorded over time and compared with results from other indices, like the Normalized Difference Vegetation Index (NDVI); NDMI showed an improved ability to detect changes in forest characteristics, especially changes due to partial cuts or thinning within forests (Wilson and Sader 2002). NDMI is also known to differentiate between old growth and mature forests better than other procedures (Fiorella and Ripple 1993). This research applies NDMI slightly differently. It compares the

index between plots, which is a much smaller scale of application than the studies that have used NDMI in the past. Perhaps NDMI was not selected as one of the best predictors because it was suited for larger, regional scale analysis than plot-wise comparison. Thus, it remains unclear if moisture gradient is one of the main biophysical characters that somehow influenced the number of species in the study areas. Contrasting gradients in biophysical parameters, like moisture, (i.e. presence, abundance, absence of water/moisture etc), may have a strong localized influence on species richness, composition, and tree density in East Usambara. This can be tested if humidity is recorded from within the forest and/or perennial and non-perennial streams within forests are mapped. However, moisture alone may not be able to explain the species richness or composition in all areas, and in some areas, other gradients (like soil nutrition, slope orientation, and so on) may be more important (Condit 1996).

To a certain degree, the current research was successful in identifying above-average high tree diversity areas within the humid forests. However, multiple influencing factors, including moisture condition, may determine the specific species richness and composition within the plots. The absence of data on all factors influencing the actual species distribution in the East Usambara could be one explanation why the multiple linear regressions and non-parametric KNN method could not predict number of tree species per plot. Logistic regression was successful because it only differentiated between above average tree species rich areas and others. However, all analytical techniques were moderately successful in identifying above average tree species richness. Multiple linear, logistic regression, and discriminant analyses all identified the same set of parameters (Landsat ETM band 7 and number of trees) to be the most predictive of all available variables. Using KNN, Rajaniemi *et al.* (in press) estimated species

richness for *Pteridophytes* (ferns) and *Melastomataceae* (small trees and shrubs) within lowland humid forests in the Amazonian Ecuador. Their species definition included both morphospecies and taxonomic species and could estimate fern species fairly accurately (root mean square errors between 1.9 and 5) (Rajaniemi *et al.* in press). Even though the plot-based correlation between actual and predicted data is not available in their paper, the study clearly indicates promising results and an avenue for further research to improve estimating species richness of different rainforest plant groups. Although KNN did not produce good result in the current research, the analysis, especially for predicting tree density per plot, can potentially be improved. When successful, the pixel-wise results of the KNN analysis allow researchers to map the parameter of interest on the satellite image. Thus, improved pixel-wise tree density estimates will allow mapping the logistic and/or KNN models that use both satellite image-based variable and tree densities. Perhaps setting a limit for the geographic and spectral distance for the K nearest neighbor used for calculation, or incorporating *Mahalanobis* distance in predicting parameters and/or application of the new weighted KNN (pers. comm. E.Tomppo 2003) can improve the predictions.

Spectral signatures of pixels in a satellite image may be determined by one or more factors. Thus, the same spectral signature could be generated from different habitat types and it will not be always possible to predict a field or habitat characteristic from the spectral signature. Similarly , when predicting species richness, application of a model beyond the area in which the model was constructed may produce spurious results. This phenomenon is known as the "signature extension" problem (Jensen 1996). The signature extension problem also may exist within habitat types. My results suggest that Landsat ETM band 7 reflectance corresponds to

"local peaks" in tree species richness or a combination of factors that co-vary with tree species richness, instead of a specific species number. This could yet be another reason why the multiple linear regression models and KNN analyses did not reliably predict species numbers. I could have tested this local peak concept by applying the Amani N.R.-based model to similar plot-based data from areas outside the natural forest near Amani, in species poor areas. If the Amani N.R. models predicted species numbers, then extrapolation to areas outside the natural forest would not identify any high diversity areas. However, if the model represents/corresponds to local peaks in species number, then when extrapolated, the model would distinguish more speciose areas from the poorer areas. In the absence of such data, I tried to test this pattern in another way. I applied the model to Nilo F.R., first, maintaining high tree diversity definition from the Amani N.R. ($\geq$16species/plot); then, defining high diversity based on local peaks in Nilo F.R. records ($\geq$13 tree species). The latter model showed considerably better predictability.

Assuming the signature extension problem exists for the East Usambara forests, and the model detections correspond to a "local peak" in tree species, exporting the model far beyond the area in which it was fit may produce poor results. Possibly, a model should be used within a certain optimal area. The size of such an optimum area may differ from one humid tropical forest to another. This could be evaluated by using data from a large forest area, such as, the Udzungwa, the largest forest block of the Eastern Arc. A model could be constructed locally with the high diversity areas defined with local data. The model could then be tested in concentric areas of increasing size. The point where an addition in area results in decreasing accuracy may indicate the optimum size over which the definition of local-peak in species richness holds true. The size of the optimum area will differ from place to place; the greater the habitat and environmental

variability over space, the smaller will be the optimal area. An indication of the existence of such an optimal size for a particular definition of above-average diversity is perhaps present in the current research. After developing the model for the Amani N.R. and applying it to the Nilo F.R., I combined the data from both forest sectors and fitted a logistic regression model. I did this twice, first taking the above average tree species richness of the pooled data set as the definition of high diversity area ($\geq$15species/plot). I then maintained the definition of high diversity area in each forest sector ($\geq$16species/plot for Amani N.R. records and $\geq$13species/plot for Nilo F.R.) in the combined set. The latter always showed slightly better results in its ability to identify a higher percent of actual species rich areas in the forest sectors (i.e. higher producer's accuracy). Since the models do not reliably predict species numbers, this sensitivity to the local peak rather than peak species number is actually beneficial for the model application. Because of this, the model can be much more flexible and applicable to different forest areas than if it had predicted a specific species number.

Both classification-based and classification-free analyses for within-forest variations in biota or biogeophysical characteristics in the humid tropical forests have benefited when remote sensing/ satellite image-based data were used in combination with spatial ground and/or other ancillary data (Gastellu-Etchegorry et al. 1993; Riaza et al. 1998; Muchoney and Strahler 2002). Accurate and up-to-date field or geographic data are rare for tropical forest areas. Remote sensing with supplementary data, which can sometimes be inaccurate, and/or out of date, often is used for estimating forest related variables over a small area (Rao et al. 1998). Collecting field data is time consuming, expert-intensive, and thereby expensive. To formulate a rapid assessment and monitoring program within a tropical rainforest, the field data incorporated in a model should

able to be rapidly collected. It should also be such that a large number of people with varied expertise could be easily trained on how to collect them if needed. From the available Amani N.R. and Nilo F.R. dataset, only a few variables fit this description, and the model selection process picked only one: tree density. The number of trees, along with Landsat ETM band7, improved the prediction for above average tree species richness in the study area.

The addition of tree density usually increased the models' predictability, with marked improvement in the detection of average/below average diversity areas. Huang et al. (2003) found tree density to be higher in intact forests compared to those formally logged, burned, or cultivated. Above-average tree diversity areas in the current research may include intact forests as well as some that previously burned or were disturbed but are presently well recovering (pers. comm. Albert Ntemi and Kathrine Doody 2001). The average/below average diversity plots include those that may be presently disturbed, previously burned, disturbed, or cultivated, and sites that naturally hosts a low number of species. My findings suggest that most above-average diversity plots host high tree density also. However, misclassifications of average/below average diversity areas as high diversity areas after adding tree density data in the model suggest that on occasion average/below average tree diversity areas may have high tree numbers. Upon adding tree density data in the model, especially in Nilo forests, the model could detect fewer above-average diversity areas compared to the satellite model; indicating not all high tree diversity areas have high tree densities, and low tree density might also occur in high diversity plots. However, in general by adding tree density data, the predictability of the model improved mostly due to their improved abilities to isolate the average/below average areas; suggesting that the average/ below average diversity plots are more closely associated with low tree density.

Huang et al.(2003)'s study aimed at understanding the impact of human disturbance in different forest sectors of East Usambara. They found no difference in number of tree species between intact forests and those formally logged, burned, or cultivated. The authors also pointed out that disturbance regime can be an important, but not necessarily the only factor that determines species diversity within humid tropical forests (Huang et al. 2003). The above-average species diversity areas in the current research may include intact, naturally species-rich areas, previously disturbed but well recovering patches, etc. The distribution pattern of above-average richness may be due to a number of underlying ecological processes, including lack of disturbance (intact plot), intermediate disturbance (Molino and Sabatier 2001), ecotonal or edge effect, limited dispersal (Chave and Leigh 2002), species aggregation (Condit et al. 2000), patchiness of resources (Nee and May 1992; Pacala and Tilman 1994), random variation (Gentry 1988), or effects of other, unmeasured environmental variations.

In general, in the Amani and Nilo forests, areas that have low tree-species numbers host fewer trees; in contrast, species-rich locations have high tree density. The high frequency of trees in high species locations is perhaps due to the presence of common and rare species. The latter tend to aggregate more than common species (Condit et al. 2000). In addition, some of the high species sites, being relatively undisturbed, host few, but large, canopy trees. Great variability exists in understory light condition in undisturbed or old growth forests (Nicotra et al. 1999). Understory species are more diverse than canopy species (Croat 1978). The light conditions, in turn, could encourage the presence of both light and shade tolerant varieties, and species of

different plant functional types (Huston 1994). All these factors may contribute to high tree numbers and high tree species in some plots. However, departure from this pattern is observed in areas where high tree counts are associated with a low number of species and vice versa. In the East Usambara forests, 0.22% of the most common species account for 90.2% of the total number of trees (Huang et al. 2003). Some plots may host a higher number of trees, but most of them are common species, and therefore report low species richness. In addition, species richness decreased with increased frequency of some commonly occurring tree species (Huang et al. 2003). This may account for some of the misclassifications in the field/satellite model.

On the other hand, undisturbed, old growth forest patches that are stable and dominated by a few canopy species, may show high species number but not a high number of trees. This is evident from the results of the model application to the Nilo F.R. area, where, with the addition of tree density data, a smaller percentage of actual species-rich sites could be detected. At the same time, the ability to detect average/below average diversity locations increased greatly. This supports the observation that, while some high species sites host low tree numbers, most of the low tree density sites are associated with low species count in the Nilo forest. Nilo F.R. is distinct among the East Usambara forest blocks because of its high altitude and steep gradient (Seddon et al. 1999). In contrast, Amani N.R. has, in general, gentler slopes and thereby greater accessibility. Consequently, chances of anthropogenic disturbance probably are greater in Amani N.R. It is possible that high stem density is more closely associated with relatively undisturbed or well-recovering high species richness plots in the Amani N.R. than in the Nilo F.R. In Nilo, forest structure, species composition, and similar factors, rather than number of trees, perhaps could differentiate areas of above-average tree species richness more easily. It is

important to investigate these underlying relationships and origins of species richness, or endemism, further. If identified, the underlying factors can be appropriately incorporated in management and conservation actions; for example, to maintain a certain disturbance level if intermediate disturbance explains the richness, to avoid management practices in areas where endemism can be explained by long term adoption to stability, and so on (Lovett et al. 2000).

As demand for forest resources and land currently under forest continues to grow, it is important to be able to quickly assess and monitor high diversity areas within existing patches. Current research shows that the models for the Amani N.R. perform well in Amani and can reliably detect the high tree diversity areas within the Nilo F.R. Given that much of the remaining tropical humid forestland is rapidly degrading (WCMC 1992; Nepstad et al. 1999; Achard 2002), the current research responds to the urgent need for a set of simple models that can roughly suggest the locations of tree species-rich areas within rainforests. This, itself, is the first step toward helping managers and researchers gain the ability to measure, assess, quickly identify threats and thereby make more informed decisions about conservation of trees and the species that associate with trees.

Most developing countries have no system of monitoring biodiversity, even though countries like Tanzania, as signatory to the Convention of Biological Diversity (CBD) treaty, are obligated to assess and monitor diversity (Danielsen et al. 2000). Sheil (2002) suggested that, with low financial and human resources, technical data collection is not cost effective for the tropical forest regions. He also criticized research practices by pointing that, while scientists are obsessed with precision measurements, managers would rather identify and address threats

quickly in the tropical forests (Sheil 2002). The simple rapid assessment and monitoring model derived by this research addresses the needs of both managers and scientists. It provides them with rapid, usable, and updatable information. The increased capabilities of hand-held GPS to record coordinates, and the ability of personal computers to perform memory-intensive but simple satellite image analyses, make these models easy for managers and researchers in Tanzania to adopt and use in the field. Moreover, these models could be the first step towards understanding the pattern and the underlying distribution of species in East Usambara and in the Eastern Arc forests. This may enable managers and researchers to address the real mechanisms or factors that are affected by a threat, as well as the threat (biological or otherwise) itself.

In the current research, all satellite or field/satellite models developed by the different analytical techniques generated similar predictabilities. In finer comparison, the choice was between a model's ability to correctly detect percent of above-average richness areas (i.e.% actual correct) and its ability to ensure that a high percent of the areas predicted as above-average richness by the model actually contained high number of tree species (i.e. % predicted correct). Logistic and discriminant analysis showed better predictabilities than other models. However, discriminant analysis assumes that data are normally distributed, while this was the case in this study, it may not be true in all studies. If data are non-normally distributed, logistic regression should generate better results (Press and Wilson 1978). Moreover, logistic models are simpler to apply when producing maps from satellite images. If discriminant analysis models generated distinctly better results compared to the logistic regression model in identifying above-average richness areas, despite the complexity, it would have been better to adopt the discriminant model for further and

future analysis. However, given the results, logistic regression appears to be the preferred analytical technique now.

The level of prediction achieved by the current models can and should be improved through further research to discover the underlying factors influencing the satellite image reflections. If, in the near future, KNN analysis can be improved or other recent techniques can be adopted to reliably detect stem density (Brandtberg et al. 2003; Gray 2003; Thenkabail et al. 2003) the field/satellite model could be mapped, without extensive fieldwork. However, researchers and managers need not wait for such further improvements to apply the current model. The current models by themselves will improve their abilities to assess and monitor the within forest diversity and in turn can contribute towards further improvement of the same or similar models in the future.

If in the future, it becomes difficult to obtain Landsat ETM images, or if Landsat or other images do not record bands equivalent to the current Landsat ETM+ band 7 (2.09-2.35 µm), perhaps similar models can be developed based on other mid-infrared bands. For example, Landsat ETM+ band 5, 1.55-1.75 µm, which is highly correlated with Landsat ETM band 7 (RB7), or similar bands from other satellite sources, but have lower predictive capabilities than RB7, may still be used for similar model generation. During the initial analyses, Landsat ETM visible red band, RB3 (0.63-0.69 µm)-based model showed better results than other visible bands. Perhaps, in absence of Landsat mid-infrared band 7 or equivalent ones, the visible red band by itself could be used for prediction, or perhaps the predictability of the model to detect above-average diversity could be improved by combining it with other available mid-infrared bands.

7.0 Literature Cited

Achard, F., H.D. Eva, H. Stibig, P. Mayaux, J. Gallego, T. Richards, and J. Malingreau. 2002. Determination of deforestation rates of the world's humid tropical forests. Science **297:** 999-1002.

Anderson, D.R., K.P. Burnham, W.L. Thompson, 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management **64(4):** 912-923.

ArcView GIS 3.2 software. 1992-1999. Environmental System Research Institute, Inc.

Ayensu, E., D. van R. Classen, M. Collins, A. Dearing, L. Fresco, M. Gadgil, H. Gitay, G. Glaser, C. Juma, J. Krebs, R. Lenton, J. Lubchenco, J. A. McNeely, H. A. Mooney, P. Pinstrup-Andersen, M.Ramos, P. Raven, W. V. Reid, C. Samper, J. Sarukhan, P. Schei, J.G. Tundisi, R.T. Watson, X. Guanhua, A.H. Zakri. 1999. International ecosystem assessment. Science **286:** 685-686.

Balmford, A., M.J.B.Green, and M.G.Murray. 1996_a. Using higher-taxon richness as surrogate for species richness: I. Regional tests. Proceedings of Royal Society of London **263:** 1269-1274.

Balmford, A., A.H.M. Jayasuriya, and M.J.B.Green. 1996_b. Using higher-taxon richness as surrogate for species richness: II. Local applications. Proceedings of Royal Society of London **263:** 1571-1575.

Balmford, A. 1998. On hotspots and the use of indicators for reserve selection. Trends in Ecology and Evolution **13(10):** 409

Balmford, A. and K.J. Gaston 1999. Why biodiversity surveys are good value. Nature **398:** 204-205.

Beharrell, N.K., E. Fanning, and K.M. Howell (editors). 2002. Technical paper 53 Nilo Forest Reserve: a biodiversity survey. EUCAMP, Ministry of Natural Resources and Tourism, Tanzania, Forestry and Beekeeping Division, Department of International

Bibby, C.J. 1998. Selecting areas of conservation. Pages 176-201 in W. J. Sutherland, editors. Conservation science and action. Blackwell Science, Oxford, MA, USA.

Blinn, C.E., R.H. Wynne, and J.A. Scrivani, 2002. Factorial analysis of IGSCR parameters using a hypothesis testing approach. Technical Papers, 68th Annual Meeting of the American Society for Photogrammetry and Remote Sensing, Washington, DC, USA.

Brooks, T. 2000. Living on the edge. Nature **403:** 26-29.

Brooks, T.M., R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, A.B. Rylands, W.R. Konstant, P.Flick, J. Pilgrim, S.Oldfield, G. Magin, and C. Hilton-Taylor. 2001. Habitat loss and extinction in the hotspots of biodiversity. Conservation Biology **16(4):** 909-923.

Brandtberg, T., T.A. Warner, R.E. Landenberger, J.B. McGraw. 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from eastern deciduous forest in North America. Remote Sensing of Environment. **85:** 290-303

Burbidge, A.H., K. Leicester, S. McDavitt, and J.D. Majer.1992. Ants as indicators of disturbance at Yanchep National Park, Western Australia. *Journal of the Royal Society of Western Australia* **75:** 89-95.

Burnham, K.P. and D.R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag New York Inc., New York, USA.

Campbell, J.B. 1996. Introduction to Remote Sensing (2nd Edition). The Guilford Press, New York USA. Pp 622.

Carpenter, G., Gopal, S., Martens, S., and Woodcock, C. 1999. A Neural Network Method for Mixture Estimation for Vegetation Mapping, Remote sensing of the Environment. **70 (2):** 138-152.

Carranza, E. J.M. and M. Hale. 2002. Mineral imaging with Landsat Thematic Mapper data for hydrothermal alteration mapping in heavily vegetated terrain. International Journal of Remote Sensing. **23(22):** 4827-4852

Channell, R. and M. V. Lomolino. 2000. Dynamic biogeography and conservation of endangered species. Nature **403:** 84-86.

Chave, J., and E. G. Leigh. 2002. A spatially explicit neutral model of -diversity in tropical forests. Theoretical Population Biology **62:** 153-168

Cohen, W.B., T.A. Spies, M. Fiorella. 1995. Estimating the age and structure of forests in multi-ownership landscape of western Oregon, USA. International Journal of Remote Sensing **16:** 721-746.

Condit, R. 1996. Defining and mapping vegetation types in mega-diverse tropical forests. Trends in Ecology and Evolution **11(1):** 4-5

Condit, R., P.S. Aston, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S. P. Hubbell, R. B. Foster, A. Itoh, J. V. LaFrankie, H. S. Lee, E. Losos, N. Manokaran, R. Sukumar, T. Yamakura. 2000. Spatial patterns in the distribution of tropical tree species. Science **288:** 1414-1418

Congalton, R., R. G. Oderwald, R. A. Mead. 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. Photogrammetric Engineering and Remote Sensing **49 (12):** 1671-1678.

Croat, T. 1978. Floor of Barro Colorado Island. Stanford University Press, Stanford, CA, USA.

da Fonseca, G.A.B. 2000. Following Africa's lead in setting priorities. Nature **405:** 393-394.

Danielsen, F., D. S. Balete, M.K. Poulsen, M. Enghoff, C.M. Nozawa, and A.E. Jensen. 2000. A simple system for monitoring biodiversity in protected areas of a developing country. Biodiversity and Conservation **9:** 1671-1705.

Doody, K.Z., K.M. Howell, and E. Fanning (editors). 2001. Technical paper 52 Amani Nature Reserve: a biodiversity survey. EUCAMP, Ministry of Natural Resources and Tourism, Tanzania, Forestry and Beekeeping Division, Department of International Development Co-operation, Finland; Frontier-Tanzania; University of Dar es Salaam; Society for Environmental Exploaration, Tanga, Tanzania.

Duivenvoorden, J. F. and J.M. Lips. 1998. Mesoscale patterns of tree species diversity in Colombian Amazonia. Forest biodiversity in North, Central and South America, and the Caribbean: research and monitoring (eds. F. Dallmeier and J.A. Comiskey). Pp. 535-549. Man and the Biosphere Series 21. UNESCO, Paris and The Parthenon Publishing Group, New York, USA.

Emberton, K.C., T.A. Pearce, P.F. Kasigwa, P. Tattersfield, and Z. Habibu. 1997. High biodiversity and regional endemism in land snails of eastern Tanzania. Biodiversity Conservation **6:** 1123-1136

ERDAS Imagine 8.5 software. 2001. Geographic imaging made simple, ERDAS, Inc.

Fiorella, M. and W.J. Ripple. 1993. Determining successional stage of temperate conifer forests with Landsat satellite data. Photogrammetric Engineering and Remote Sensing **59:** 239-246

Flora of Tropical East Africa Series. 1952-1989. Published under the authority of the Secretary of State for the Colonies by the Crown Agents for Oversea Governments and Administrations, London, or Balkema, Neatherlands on behalf of the East African Governments. Prepared in assistance of the East African Herbarium at the Royal Kew Botanical Gardens, UK.

Food and Agriculture Organization (FAO). 2001. Global forest resources assessment 2000: main report. FAO, Rome.

Foody, G.M. and Cutler, M.E. 2000. Remote sensing of biodiversity: using neural network to estimate the diversity and composition of a Bornean Tropical Rainforest from Landsat TM data. 497-499. IGARSS 2002 : remote sensing, integrating our view of the planet : 2002 IEEE International Geoscience and Remote Sensing Symposium : 24th Canadian Symposium on Remote Sensing : proceedings : Westin Harbour Castle, Toronto, Canada, June 24-28

Franco-Lopez, H., A. R. Ek, M. E. Bauer. 2001. Estimation and mapping of forest stand density, volume, and cover type using k-nearest neighbors method. Remote Sensing of Environment **77 (3):** 251-274.

Gastellu-Etchegorry, J.P. and V. Trichon. 1998. A modeling approach of PAR environment in a tropical rain forest in Sumatra: application to remote sensing. Ecological Modelling. **108:** 237-264

Gastellu-Etchegorry, J.P. Estreguil, J.P., Mougin, C., and Y. Laumonier. 1993. A GIS based methodology for small scale monitoring of tropical trees- a case study in Sumatra. International Journal of Remote Sensing **14(12):** 2349-2368

Gaston, K.J. 1994. Measuring geographic range sizes. Ecography **17(2):** 198-205

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

Groves, C., L. Valutis, D. Vosick, B. Neely, K. Wheaton, J. Touval, B. Runnels. 2000. Designing a geography of hope: a practioner's handbook to ecoregional conservation planning. Vol.I & II. The Nature Conservancy, Washington, D.C., USA.

Gray, A. 2003. Monitoring stand structure in mature costal Douglas-fir forests: effect of plot size. Forest Ecology and Management **175:** 1-16.

Hall, J. B. and M.D. Swaine. 1976. Classification and ecology of closed-canopy forest in Ghana. Journal of Ecology **64:** 913-951.

Hamilton, A.C. 1982. Environmental history of Africa: A study of the quaternary. Academic press, New York, USA, pp 328.

Hamilton, A.C. and R. Bensted-Smith. 1989. Forest conservation in the East Usambara Mountains, Tanzania. IUCN, Gland and Forestry Division, Ministry of Lands, Natural Resources and Tourism, Tanzania.

Hayes, D. and S.A. Sader. 2001. Change detection techniques for monitoring forest clearing and regrowth in a tropical moist forest. Photogrammetric Engineering and Remote Sensing **67(9):** 1067-1075

Hill, R.A. and G.M. Foody. 1994. Separability of rain-forest types in Tambopata-Candamo reserve zone, Peru. International Journal of Remote Sensing **15:** 2687-2693

Hill, R.A. 1999. Image segmentation for humid tropical forest classification in Landsat T.M. data. International Journal of Remote Sensing **20(5):** 1039-1044

Hoffman, R.L. 1993. Biogeography of East African montane forest millipedes, pp 103-114 *in* Biogeography and Ecology of the Rain Forests of Eastern Africa with J.C. Lovett and S.K. Wasser as editors, Cambridge University Press.

Horler, D.N. H. and Ahern, F.J. 1986. Forestry information content of Thematic Mapper data. International Journal of Remote Sensing **7:** 405-428.

Hosmer, D.W. and S. Lemeshow. 2000. Applied logistic regression (2nd Edition). John Wiley and Sons, Inc. USA, pp

Howard, P.C., P.Viskanic, R. T.R.B. Davenport, F.W. Kigenyi, M.Baltzer, C.J. Dickinson, J. S. Lwanga, R.A. Matthews, and A.Balmford. 1998. Complementarity and use of indicator groups for reserve selection in Uganda. Nature **394:** 472-475

Howell, K.M. 1993. Herpetofauna of the eastern African forests. pp 173-202 *in* Biogeography and Ecology of the Rain Forests of Eastern Africa with J.C. Lovett and S.K. Wasser as editors, Cambridge University Press, UK.

Huang, W., V. Pohjonen, S. Johansson, M. Nashanda, M.I.L. Katigula, and O. Luukkanen. 2003. Species diversity, forest structure and species composition in Tanzanian tropical forests. Forest Ecology and Management **173:** 11-24.

Huston, M.A. 1994. Biological diversity: the coexistence of species in changing landscapes. Cambridge university press, Cambridge, U.K., pp 681.

Iversen, S.T. 1991.The Usambara mountains, NE Tanzania: Phytogeography of the vascular plant flora. *Symbolae Botanicae Upsaliensis* 24: 1-234.

Jakubauskas, M.E. and K. P. Price. 1997. Empirical relationships between structural and spectral factors of Yellowstone Lodgepole Pine forests. Photogrammetric Engineering and Remote Sensing **63(12):**1375-1381.

Jeanjean, H. and F. Achard. 1997. A new approach for tropical forest area monitoring using multiple spatial resolution satellite sensor imagery. International Journal of Remote Sensing **18(11):** 2455-2461

Jenkins, M. R.E. Green, and J. Madden. 2003. The challenge of measuring global change in wild nature: are things getting better or worse? Conservation Biology **17(1):** 20-23

Jensen, J.R. 1996. Introductory digital image processing. Prentice Hall, New Jersey, USA, pp 318.

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

Lawton, J.H., J.R. Prendergast, and B.C. Eversham. 1994. The numbers and spatial distributions of species: analysis of British data. pp 177-195 *in* Systematics and conservation evaluation with P.L. Forey, C.J. Humpries, and R.I. Vane-Wright as editors. Clarendon Press, Oxford, UK.

Lavers, C. and R. Haines-Young. 1997. The use of satellite imagery to estimate Dunlin *Calidris alpina* abundance in Caithness and Sutherland and in the Shetland Islands. Bird Study **44(2):** 220-226

Lewis, M.M. 1994. Species composition related to spectral classification in an Australian spinifex hummock grassland. International Journal of Remote Sensing **40:** 125-136

Lomolino, M.V. and R. Channell. 1998. Range collapse, re-introductions, and biogeographic guidelines for conservation. Conservation Biology **12(2):** 481-484

Lovett, J.C., S. Rudd, J. Taplin, and C. Fridmodt-Moller. 2000. Patterns of biodiversity in Africa south of the Sahara and their implications for conservation management. Biodiversity and Conservation **9:** 37-46

Mabberley, D.J. 1997. The Plant-Book: A portable dictionary of the vascular plants (2nd Ed). Cambridge University Press, UK. 858pp

Mace, G.M., A. Balmford, L. Boitani, G. Cowlishaw, A.P. Dobson, D.P. Faith, K.J. Gaston, C.J. Humphries, R.I. Vane-Wright, P.H. Williams. 2000. It's time to work together and stop duplicating conservation efforts. Nature **405:** 393.

Mayaux, P. and E. F. Lambin. 1997. Tropical forest area measured from global land-cover classifications: inverse calibration models based on spatial textures. Remote Sensing of Environment **59:** 29-43.

Mittermeier, R.A., N. Myers, J. B. Thomsen, G.A.B. da Fonseca, S. Olivieri. 1998. Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. Conservation Biology **12(3):** 516-520

Molino, J-F. and Sabatier, D. 2001. Tree diversity in tropical rain forests: a validation of the intermediate disturbance hypothesis. Nature **294:**1702-1704

Muchoney, D., J. Borak, H.Chi, M. Friedl, S.Gopal, J. Hodges, N.Morrow, A. Strahler. 2000. Application of the MODIS global supervised classification model to vegetation and land cover mapping of Central America. International Journal of Remote Sensing **21(6):** 1115- 1138.

Muchoney, D. and Strahler, A. 2002. Regional vegetation mapping and direct land surface parameterization from remotely sensed and site data. International Journal of Remote Sensing. **23(6):** 1125-1142

Myers, N. 1984. The primary source: tropical forests and our future (1st edition). Norton W.W. & Company, Inc., New York, USA, pp 399.

Myers, N. 1988. Threatened biotas: 'hotspots' in tropical forests. Environmentalist **8;** 187-208

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

National Academy of Sciences.1980. Research Priorities in Tropical Biology. National Academy of Sciences, Washington DC, USA

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

Nee, S. and R. M. May. 1992. Dynamics of metapopulations: habitat destruction and competitive coexistence. Journal of Animal Ecology **61:** 37-40

Nepstad, D.C., A. Verssimo, A. Alencar, C. Nobre, E. Lima, P. Lefebvre, P. Schlesinger, C. Potter, P. Moutinho, E. Mendoza, M. Cochrane, V. Brooks. 1999. Can we defy nature's end? Nature **398:** 505-508.

Newmark, W.D. 1998. Forest area, fragmentation, and loss in the Eastern Arc Mountains: implications for the conservation of biological diversity. Journal of East African Natural History **87:** 1-8

Newmark, W.D. 1999. Ecological monitoring: its importance for the conservation of biological diversity in the Eastern Arc Forests. Bulletin of East African Natural History **29:** 4-6.

Nicotra, A.B., R.L. Chazdon, and S.V. Iriarte. 1999. Spatial heterogeneity of light and woody seedling regeneration in tropical wet forests. Ecology **80:** 1908-1926.

Olson, D.M. and E. Dinerstein. 1998. The Global 2000: A representation approach to conserving the earth's most biologically valuable ecoregions. **12(3):** 502-515.

Pacala, S.W. and D. Tilman. 1994. Limiting similarly in mechanistic and spatial models of plant competition in heterogeneous environments. American Naturalist **143(2):** 222-257

Pax-Lenny, M., C.E. Woodcock., S.A. Macomber, S.Gopal, and C.Song. 2001. Forest mapping with generalized classifier and Landsat TM data. Remote Sensing of Environment **77:** 241-250.

Pelkey, N.W., C.J. Stoner, and T.M. Caro. 2000. Vegetation in Tanzania: assessing long term trends and effects of protection using satellite imagery. Biological Conservation **94:** 297-309.

Prendergast, J.R., R.M. Quinn, J.M. Lawton, B.C. Eversham, and D.W. Giboons. 1993. Rare species, the coincidence of diversity hotspots and conservation strategies. Nature **365:** 335-337.

Press, S.J. and S. Wilson. 1978. Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association **73(364):** 699-705

Rajaniemi, S., E. Tomppo, K. Ruokolainen, H. Tuomisto. *In press.* Estimating pteridophyte and Melastomataceae species richness by means of satellite images and field data in Western Amazonian rain forest.

Rao, J. N. K. 1998. Small area estimation. In : S. Kotz, C.B. Read, and D.L. Banks (Eds), Encycopedia of statistical sciences, updated volume 2 (pp. 621-628). Wiley, New York.

Ravan, S.A., P.S. Roy, and C.M. Sharma. 1995. Space remote sensing for spatial vegetation characterization. Journal of Bioscience **20:** 427-438.

Redford, K.H., P. Coppolillo, E.W. Sanderson, G.A.B. da Fonseca, E.Dinerstein, C.Groves, G. Mace, S. Maginnis, R.A. Mittermeier, R. Noss, D. Olson, J. G. Robinson, A. Vedder, and M. Wright. 2003. Mapping the conservation landscape. Conservation Biology **17(1):** 116-131

Riaza, A., M.L. Martinez-Torres, R. Ramon-Lluch, J. Alonso, P. Heras. 1998. Evolution of equatorial vegetation communities mapped using Thematic Mapper images through a geographical information system (Guinea, Equatorial Africa). International Jurnal of Remote Sensing **19(1):** 43-54

Rodgers, W.A., and K.M. Homewood. 1982. Species richness and endemism in the Usambara Mountain forests, Tanzania. Biological Journal of the Linnean Society **18:** 197-242

Rodgers, W.A, 1993. The conservation of the forest resources of eastern Africa: past influences, present practices and future needs, pp 283-332 *in* Biogeography and Ecology of the Rain Forests of Eastern Africa with J.C. Lovett and S.K. Wasser as editors, Cambridge University Press.

Roy, P.S. and P.K. Joshi. 2002. Forest cover assessment in north-east India – the potential of temporal wide swath satellite sensor data (IRS-1C WiFs). International Journal of Remote Sensing. **23(22):** 4881-4896

SAS 8.02 software. 1999-2001. SAS Institute Inc, Cary, NC. USA.

Seddon, N., J.M.M. Ekstrom, D.R. Capper, I.S. Isherwood, R. Muna, R.G. Pople, E. Tarimo, J. Timothy. 1999. The importance of the Nilo and Nguu North Forest Reserves for the conservation of montane forest birds Tanzania. Biological Conservation **87:** 59-72

Schairer, G. 1999. Landscape level evaluation of Northern Bobwhite habitats in eastern Virginia using Landsat TM imagery. Masters Thesis. Virginia Tech.

Scharff, N. 1993. The Linyphiid spider fauna (Araneae: Linyphiidae) of mountain forests in Eastern Arc Mountains. pp 173-202 *in* Biogeography and Ecology of the Rain Forests of Eastern Africa with J.C. Lovett and S.K. Wasser as editors, Cambridge University Press.

Schmidt, P.R. 1989. Early exploitation and settlement in the Usambara Mountain. Pp. 75-78 *in* Forest Conservation in the East Usambara Mountains, Tanzania with A.C. Hamilton and R. Bensted Smith as editors, IUCN, Gland, Switzerland.

Schulman, L., L. Junikka, A. Mndolwa, I. Rajabu. 1998. Trees of Amani Nature Reserve NE Tanzania. The Ministry of Natural Resources and Tourism, Tanzania. Helsinki University Printing House. 335pp.

Sgrenzaroli, M., G.F. DeGrandi, H. Eva, and F.Achard. 2002. Tropical forest cover monitoring: estimates from GRFM JERS-1 radar mosaics using wavelet zooming techniques and validation. International Journal of Remote Sensing **23(7):** 1329-1355

Sheil, D. 2002. Why doesn't biodiversity monitoring support conservation priorities in the tropics?. Unasylva. International Journal of Forestry and Forest Industries **53(2):** http://www.fao.org/DOCREP/004/Y3582E/y3582e11.htm#k

Shimabukuro, Y.E., G.T. Batista, E.M.K. Mello, J.C. Moreira, V. Duarte. 1998. Using shade fraction image segmentation to evaluate deforestation in Landsat Thematic Mapper images of the Amazon region. International Journal of Remote Sensing **19(3):** 535-541.

Silvertown, J., M.E. Dodd, D.J.G. Gowing, and J.O. Mountford. 1999. Hydrologically defined niches reveal a basis for species richness in plan communities. Nature **400:** 61-63.

Stokstad, E. 2001. U.N. Report suggests slowed forest losses. Science **291:** 2294

The Nature Conservancy. 1997. Designing a geography of hope: guidelines for ecoregion-based conservation in The Nature Conservancy. The Nature Conservancy, Arlington, VA, USA.

Thenkabail, P.S., J. Hall, T. Lin, M.S. Aston, D. Harris, E.A. Enclona. 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

Thomas, C.D. and Mallorie, H.C. 1985. Rarity, species richness and conservation: butterflies of the Atlas Mountains in Morocco. Biological Conservation **33**:95-117

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

Trisurat, Y., A. Eiumnoh, S. Murai, M.Z. Hussain, and R. P. Shrestha. 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(10):** 2031-2042.

Tuomisto, H. K. Ruokolainen, R. Kalliola, A. Linna, W. Danjoy, and Z. Rodriguez. 1995. Dissecting Amazonian biodiversity. Science **269:** 63-66.

Van Jarsveld, A.S. S. Freitag, S. L. Chown, C. Muler, S. Koch, H. Hull, C. Bellamy, M.Kruger, S. Endrody-Younga, M.W. Mansell, and C.H. Scholtz. 1998. Biodiversity assessment and conservation strategies. Science **279 (5359) :** 2106-2108.

Verbyla, D.L. 1995. Satellite Remote Sensing of Natural Resources. Lewis Publishers, New York, USA. Pp 198.

Veregin, H. 2001. Data Quality Measurement and Assessment http://www.olemiss.edu/depts/geology/courses/ge470/gistop_11.htm. Department of Geography, The University of Mississippi. Accessed in November 2003.

Wasser, S.K. and Lovett, J.C., 1993. Introduction to the biogeography and ecology of the rain forests of eastern Africa, pp 3-8 *in* Biogeography and Ecology of the Rain Forests of Eastern Africa with J.C. Lovett and S.K. Wasser as editors, Cambridge University Press.

Whittmore, T.C. 1998. An introduction to tropical rain forests. 2nd edition. Oxford university press, New York, USA.

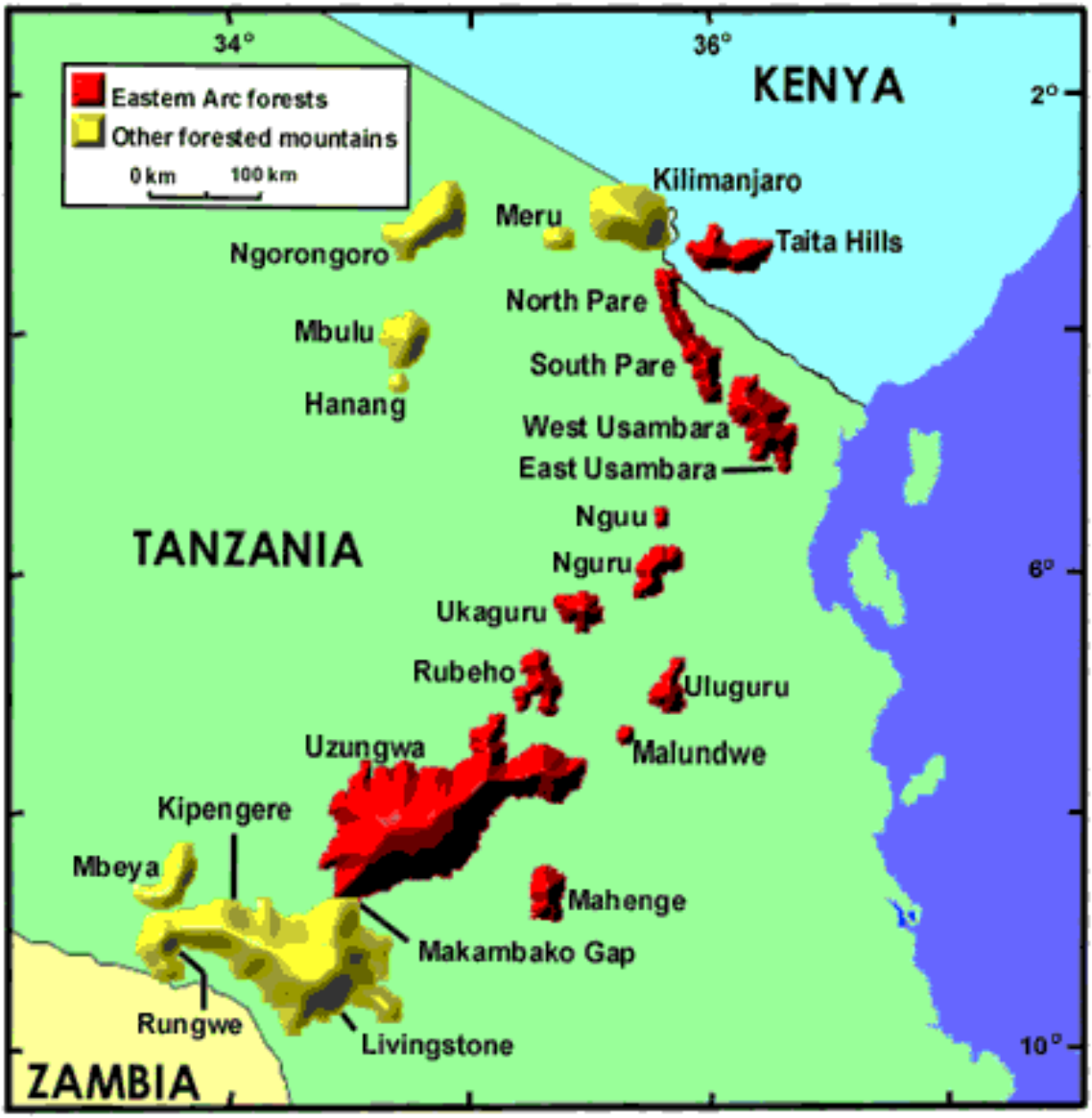
Williams P., D.Giboons, C.Margules, A.Rebelo, C.Humphries, and R.Pressy. 1996. A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British Birds. Conservation Biology **10(1):** 155-174.

Wilson, E. H. and Sader, S.A. 2002. Detection of forest harvest type using multiple dates of Landsat TM imagery. Remote Sensing of Environment **80:** 385-396.

Wolter, P.T., D. J. Mladenoff, G.E. Host, and T.R. Crow. 1995. Improved forest classification in the Northern Lake States using multi-temporal Landsat imagery. Photogrammetric Engineering and Remote Sensing **61:** 1129-1143. TA593A2P5

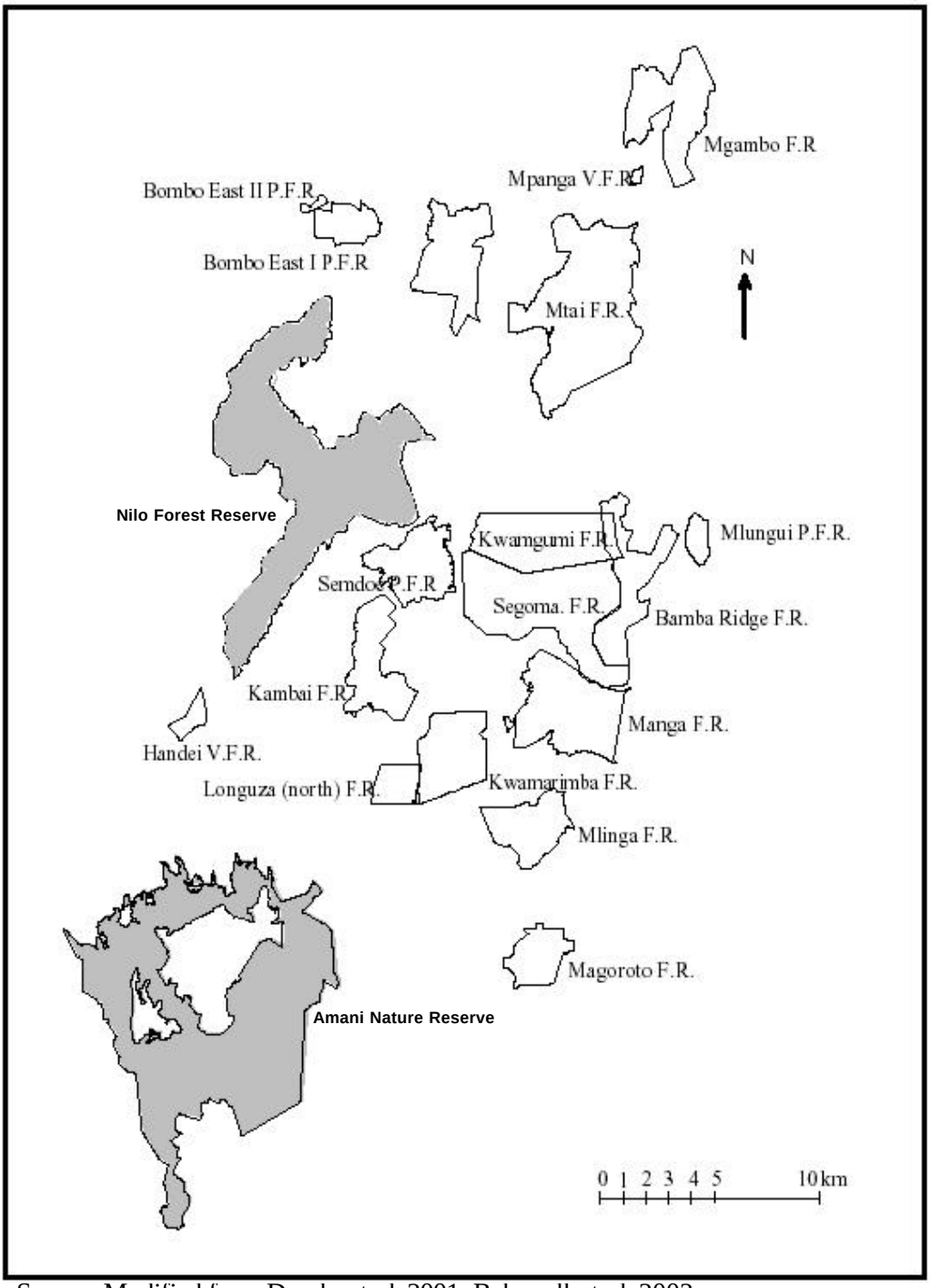
World Conservation Monitoring Centre. 1992. Global biodiversity: status of earth's living resources. Chapman and Hall, London, UK.

WWF and IUCN. 1994-1997. edited by S.D. Davis, V.H. Heywood and A.C. Hamilton. Centers of Plant Diversity: A guide and strategy for their conservation 3 Vols, IUCN Publications, Cambridge, UK



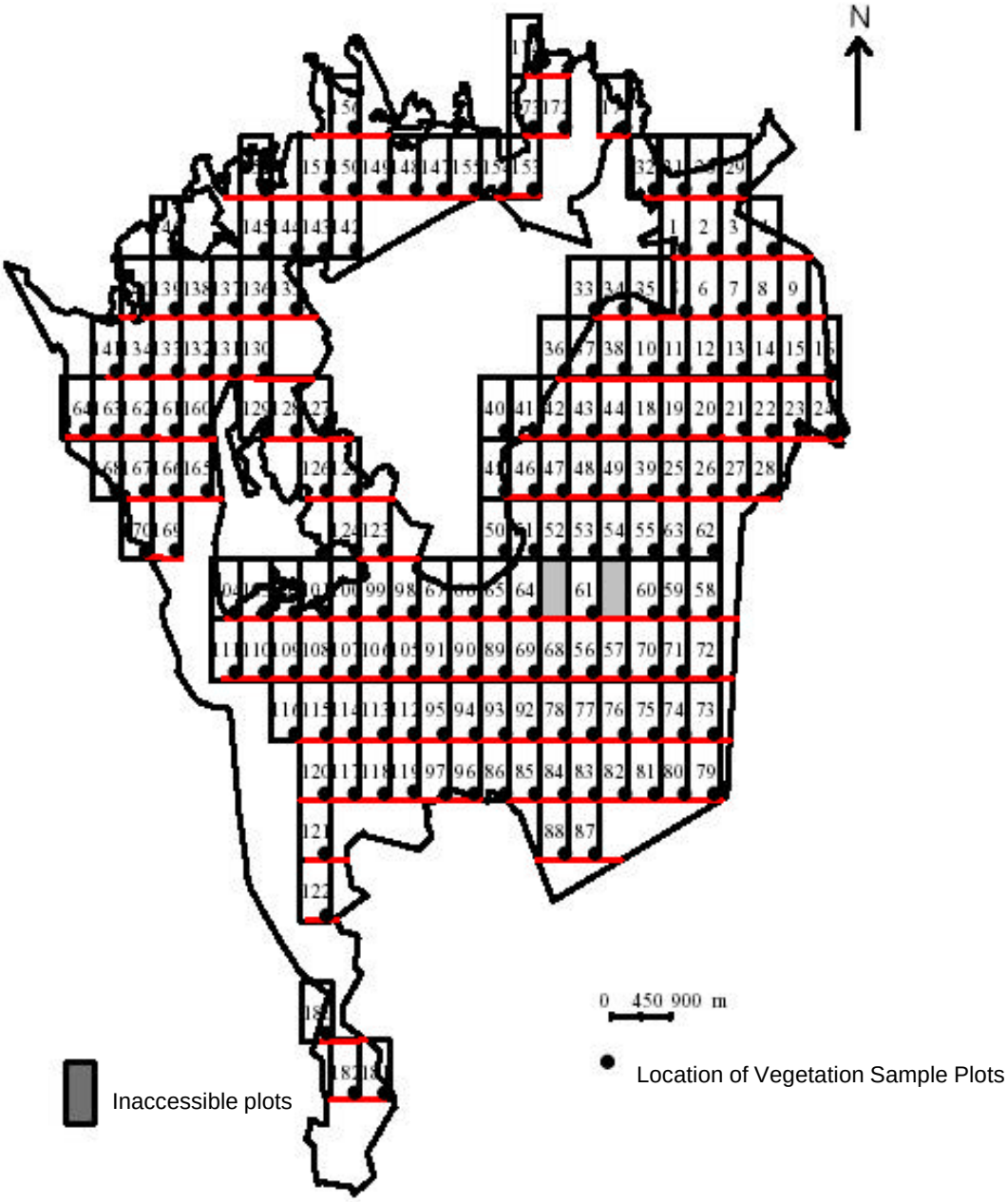
Source: Eastern Arc. 2002 The Bugwood Network – University of Georgia, College of Agricultural and Environmental Sciences and Warnell School of Forest Resources. August 8th 2001.http://www.easternarc.org/html/map.html

Figure 1: The Eastern Arc Mountain Forest in Tanzania and Kenya.



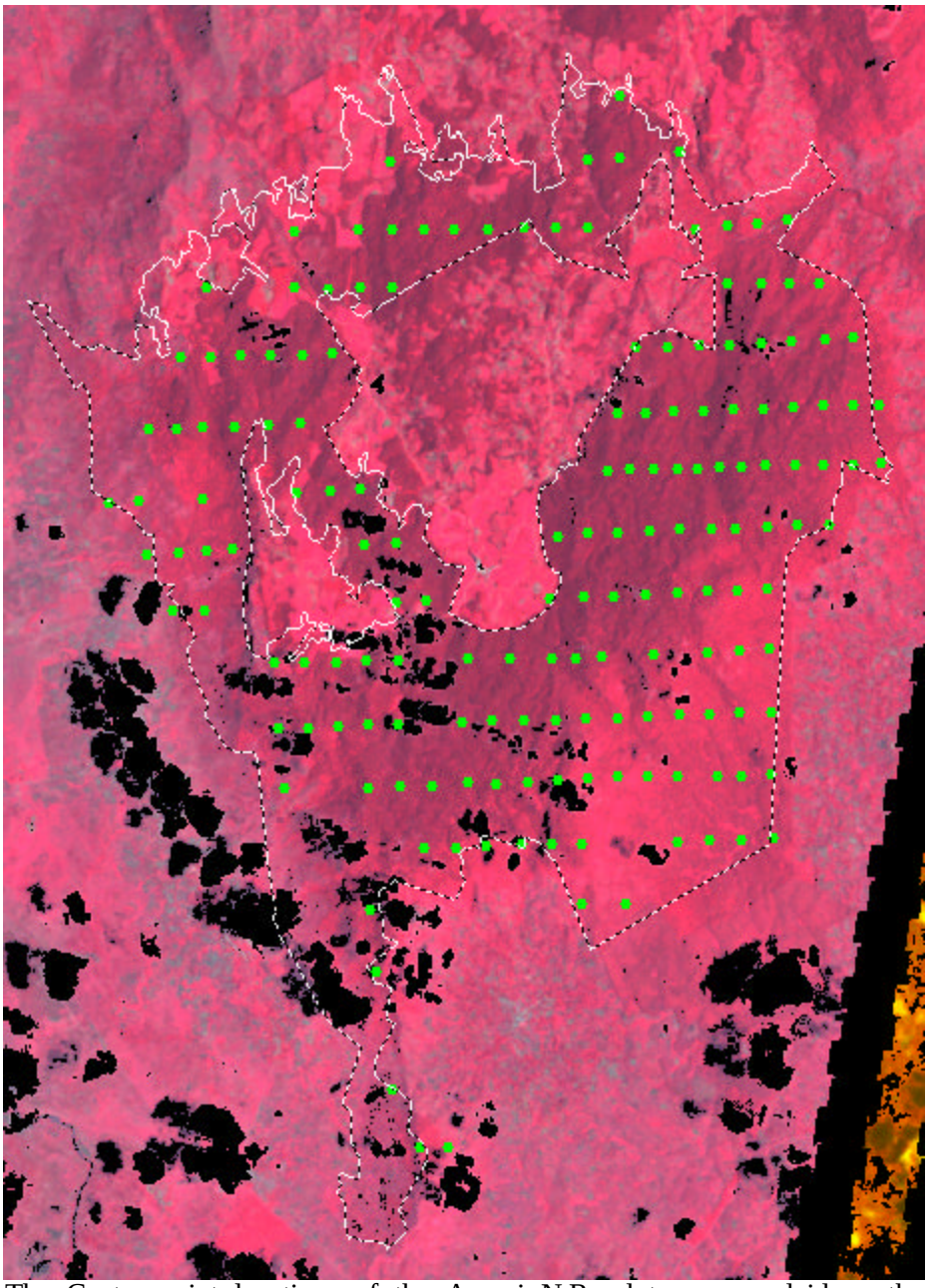
Source: Modified from Doody et al. 2001; Beharrell et al. 2002

Figure 2: East Usambara forest sectors in Tanzania, with the Amani and the Nilo reserves highlighted**.**



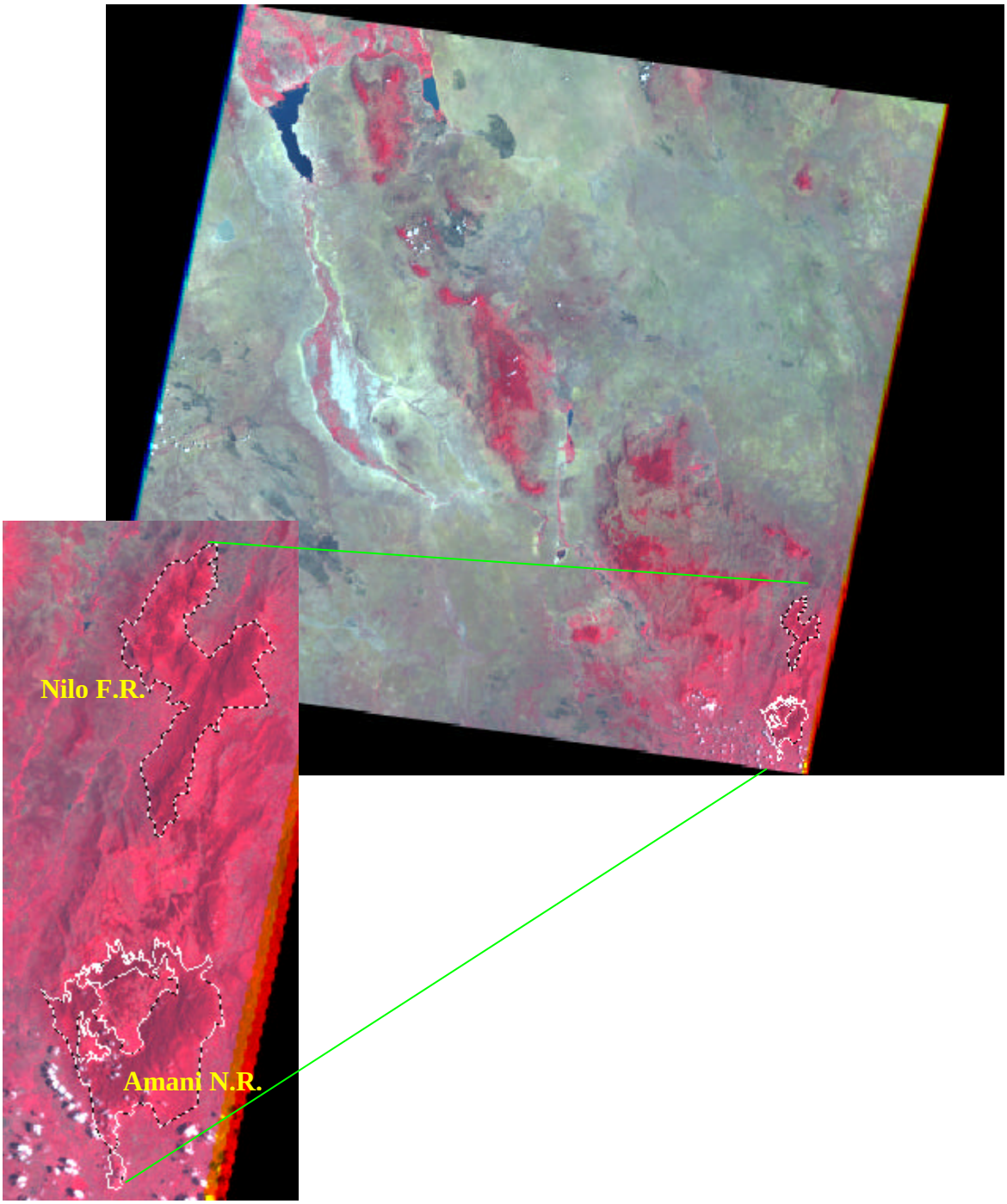
Source: Adopted from Doody et al. 2001.

Figure 3: Schematic Diagram of the sampling design followed in the Amani Nature Reserve, East Usambara, Tanzania.



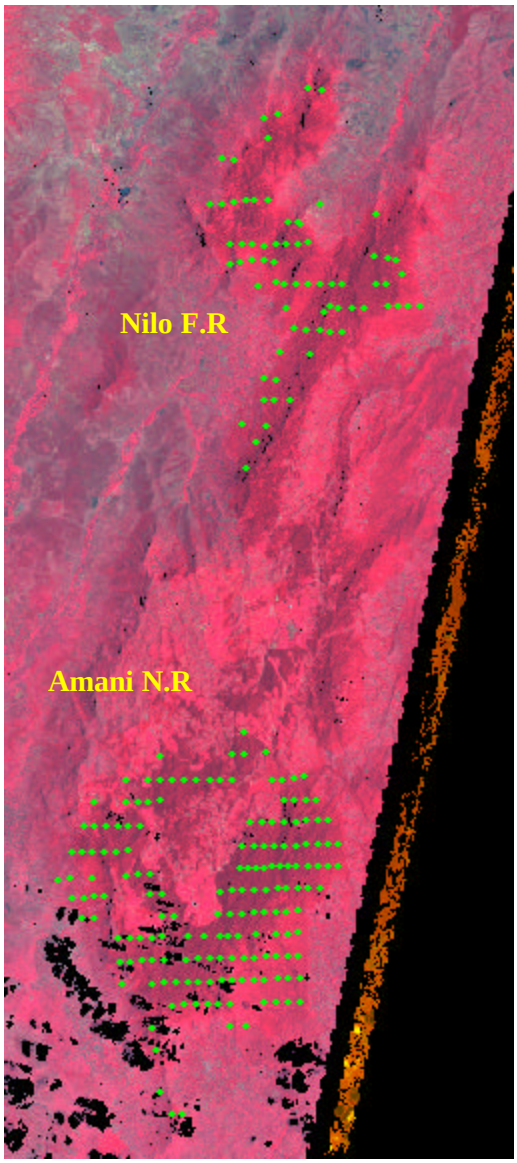
The Center point locations of the Amani N.R. plots are overlaid on the Landsat ETM image for the area. The darker clumps represent the previous cloud or cloud-shadow areas. A fine dotted line demarcates the approximate boundary for the Amani Nature Reserve.

Figure 4: Locations of vegetation sample plots within the Amani Nature Reserve, East Usambara, Tanzania.



Landsat ETM raw image representing bands 4 (near-infrared band in red), 3(in green) and 2 (in blue)

Figure 5: Locations of the Amani N.R. and the Nilo F.R. forests, East Usambara, Tanzania, on the Landsat ETM satellite image used in the current research.



Center points of the sampled plots in Amani N.R. and Nilo F.R., marked (green dots) and overlaid on Landsat ETM satellite image for the area

Figure 6: Sample plot locations of plots included in the analysis from the Amani N.R. and the Nilo F.R. forests, East Usambara, Tanzania.

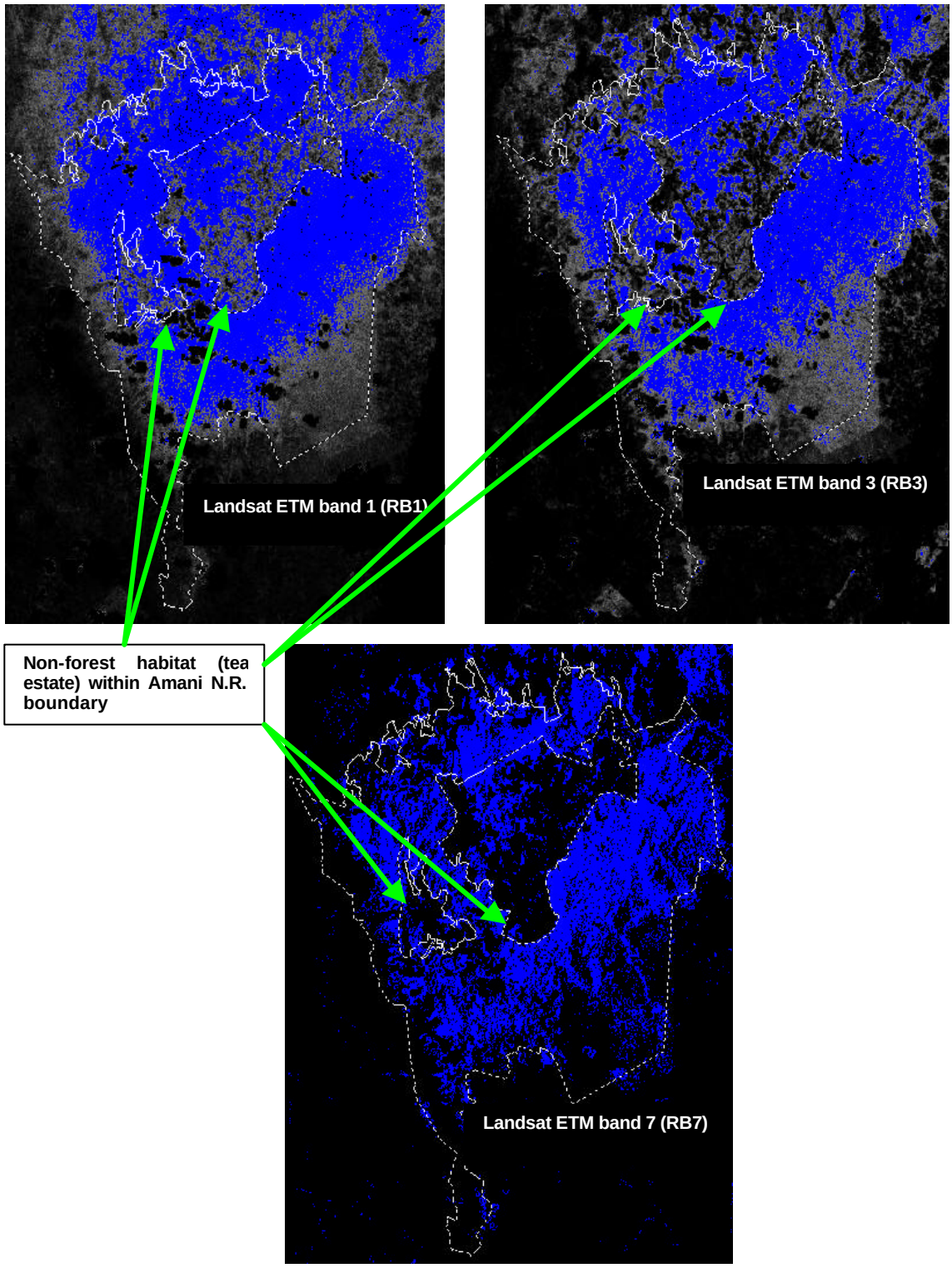


Figure 7: Logistic regression models for tree species richness areas projected on Landsat ETM satellite image for Amani N.R., East Usambara, Tanzania, where each pixel highlighted in blue reflects a probability ($\theta \geq 0.5$) of hosting above-average tree species richness based on three different bands, visible band 1, visible band 3, and mid-infrared band 7.

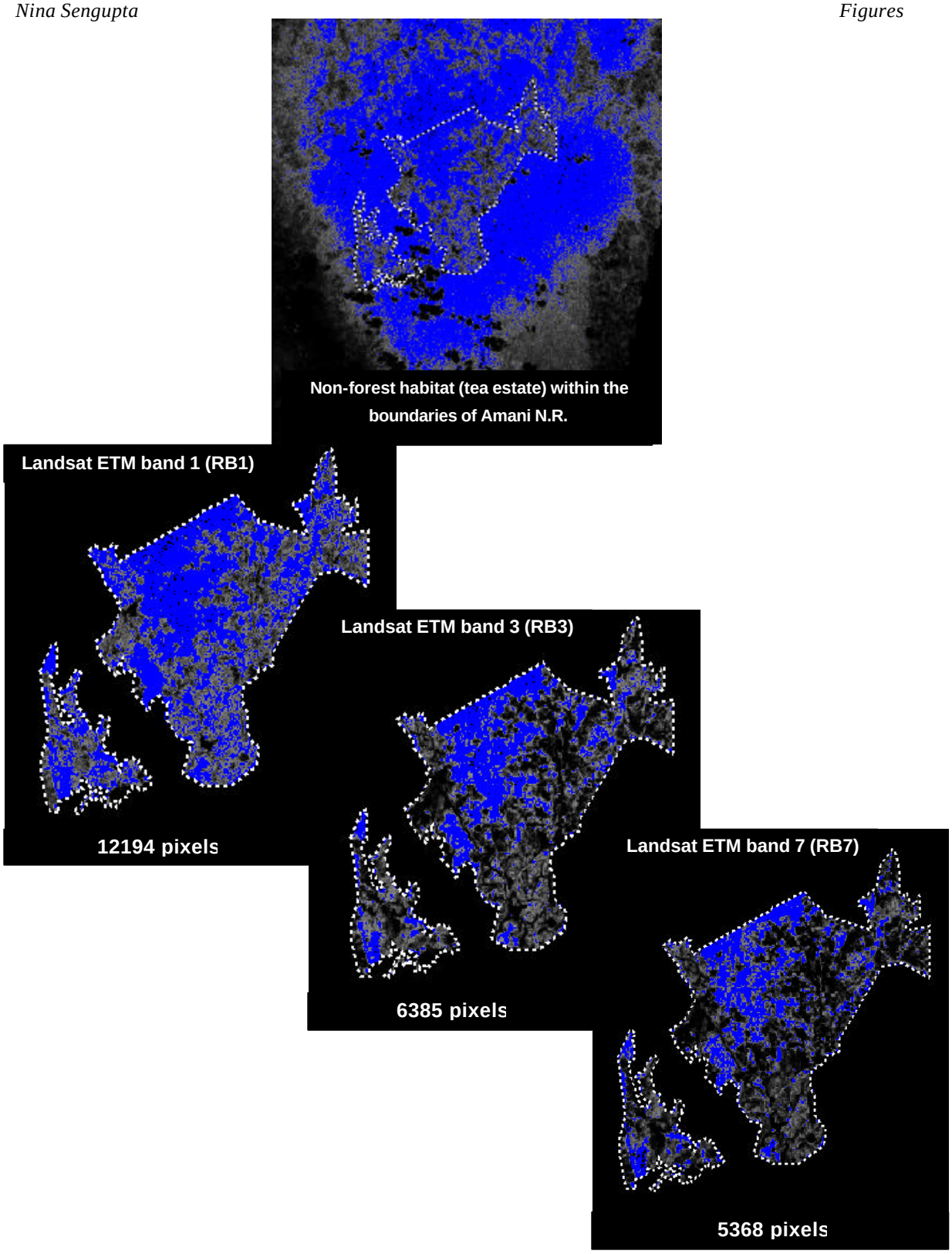


Figure 8: Number of pixels indicating above-average tree species richness in the non-forest habitat (tea-estate) within the boundaries of the Amani N.R., East Usambara, Tanzania, as predicted by logistic regression model based on visible band 1, visible band 3, and mid-infrared band 7.

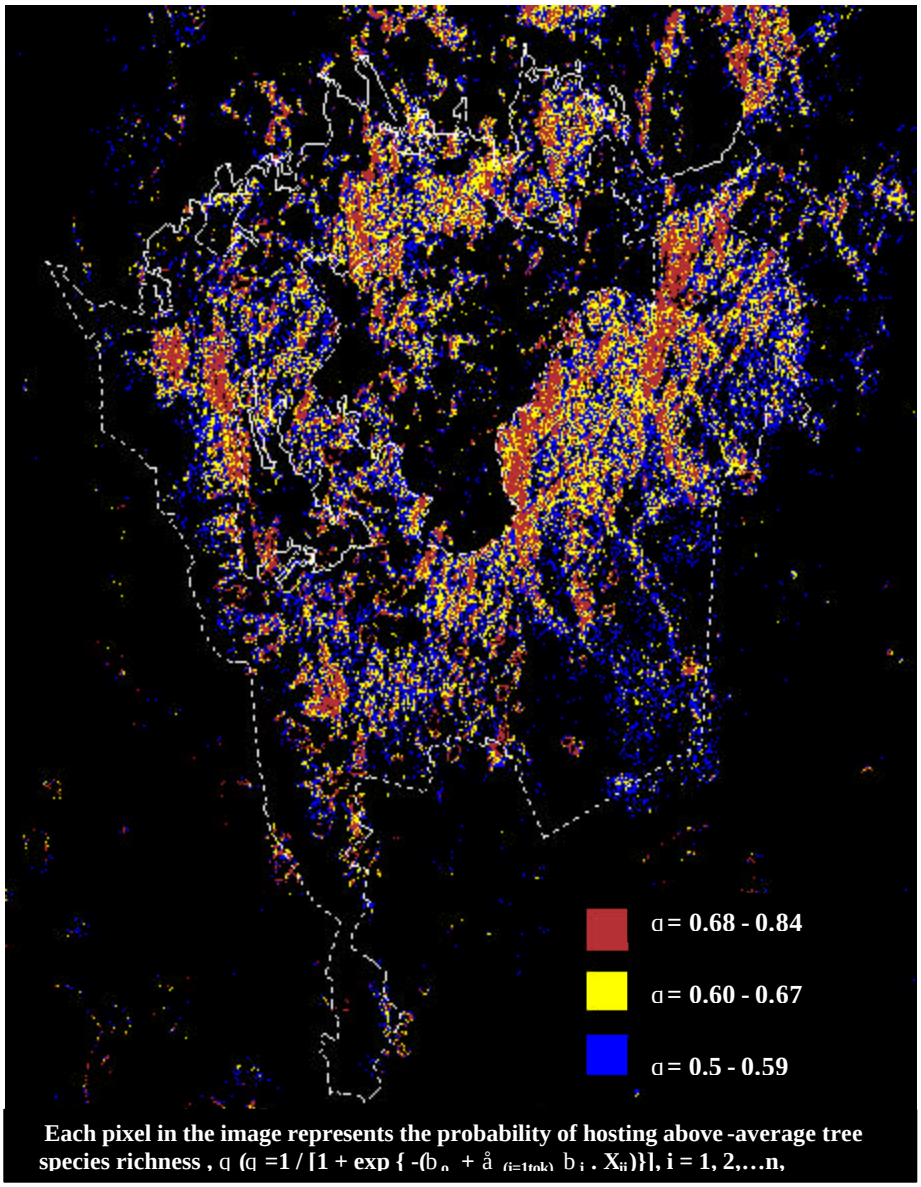


Figure 9: Above average tree species richness areas within the Amani N.R. forest, East Usambara, Tanzania, are highlighted. Each highlighted pixel reflects a probability ($\theta \geq 0.5$) of hosting above-average tree species richness identified based on a logistic regression model using Landsat ETM mid-infrared band 7.

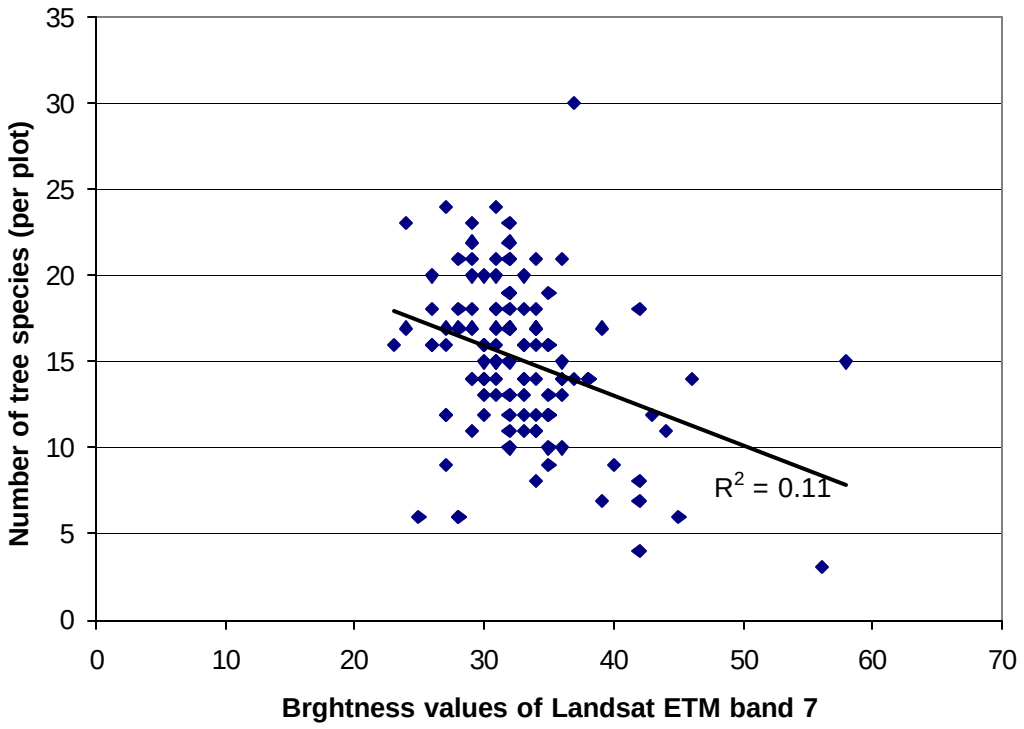


Figure 10: Chart showing the distribution of tree species number per 0.1 ha plot in the Amani N.R., East Usambara Tanzania, against the brightness value of the Landsat ETM band 7 (RB7). The trend line showing the correlation (R^2).

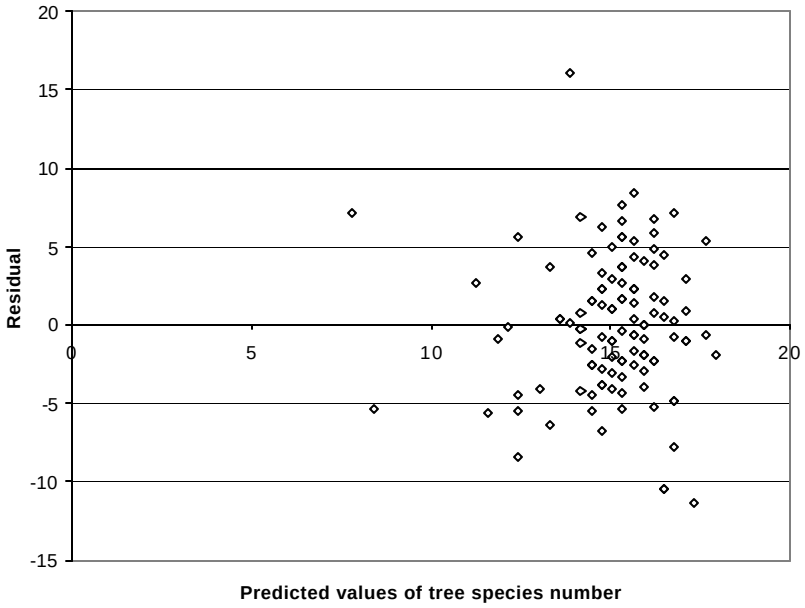


Figure 11: Chart showing the distribution of predicted tree species number per 0.1 ha plot in the Amani N.R., East Usambara Tanzania, against the residuals generated from the Landsat ETM band 7 (RB7) – based model.

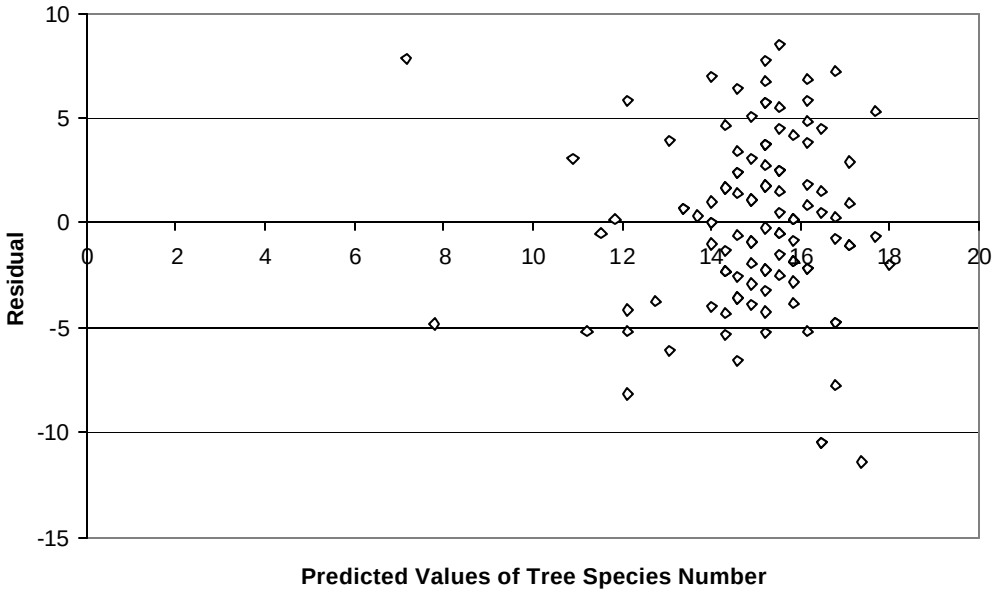


Figure 12: Chart showing the distribution of predicted tree species number per 0.1 ha plot against the residuals generated from the Landsat ETM band 7 (RB7) – based model for the Amani N.R. forest, East Usambara Tanzania, after records from one plot, a possible outlier, was removed.

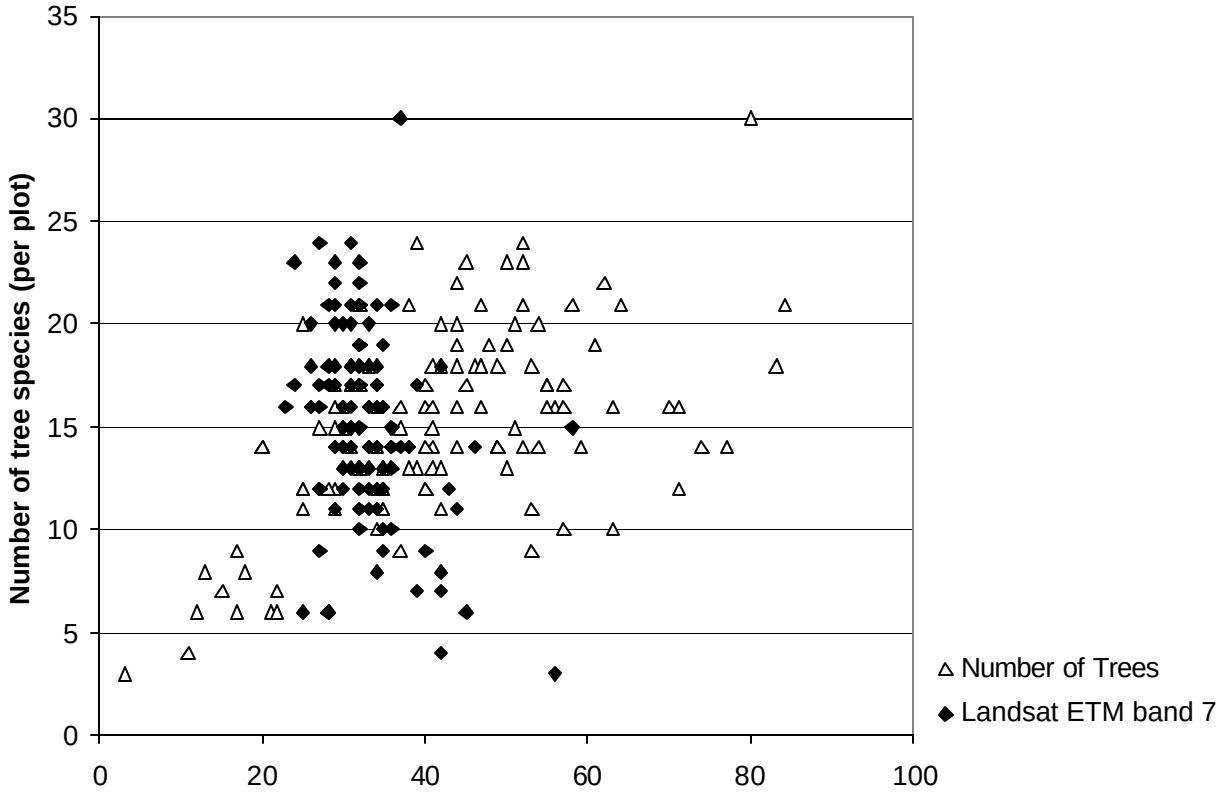


Figure 13 : Chart showing the distribution of tree species number per 0.1 ha plot in the Amani N.R., East Usambara Tanzania, against the brightness value of the Landsat ETM band 7 (RB7) and Number of Trees (per 0.1 ha plot).

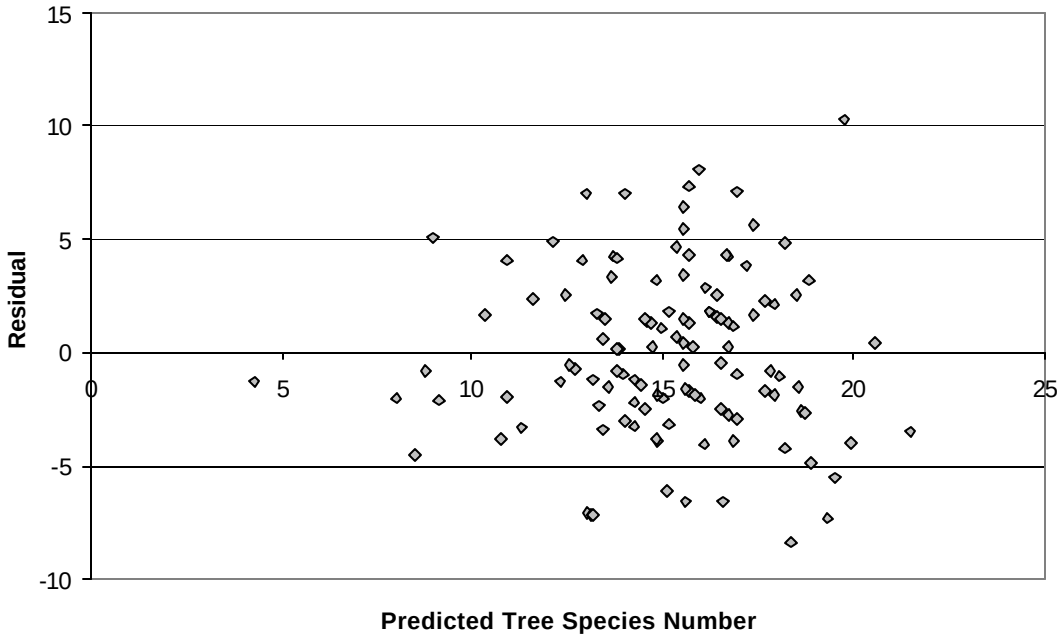


Figure 14: Chart showing the distribution of predicted tree species number per 0.1 ha plot in the Amani N.R., East Usambara Tanzania, against the residuals generated from the tree frequency per plot and Landsat ETM band 7 (RB7) - based model.

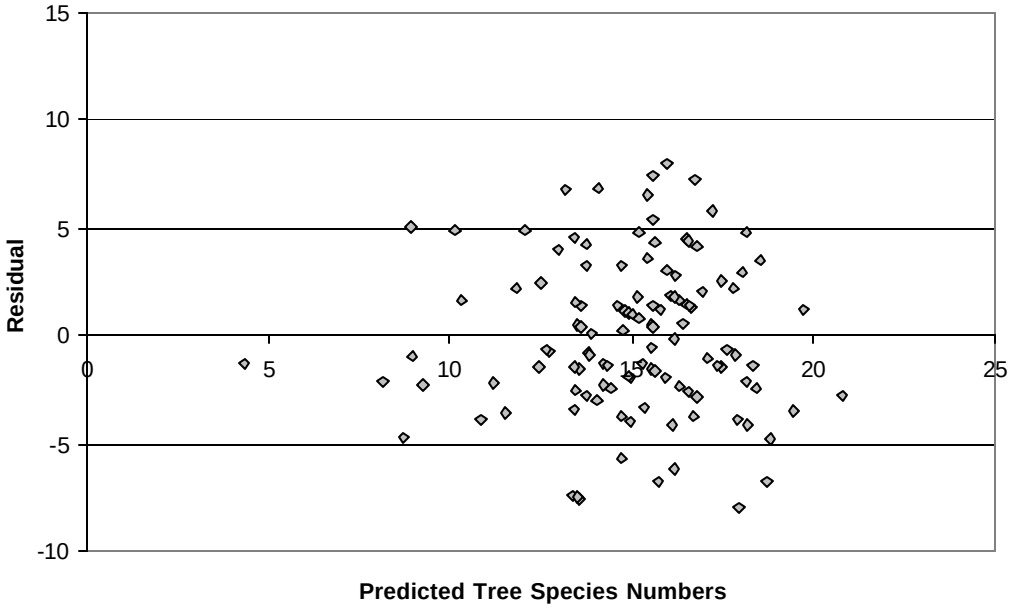


Figure 15: Chart showing the distribution of predicted tree species number per 0.1 ha plot against the residuals generated from the tree frequency per plot and Landsat ETM band 7 (RB7) – based model for the Amani N.R. forest, East Usambara Tanzania, after records from one plot, a possible outlier, was removed.

Table 1. Field and satellite-image-based variable selection from the Amani N.R. data set, East Usambara, Tanzania, for further analysis.

Variables:	Mean or Percent		W Scores or c^2 Value	P-value
	High-div area(58)	Not-high div area(60)		
Field Data				
Number of trees	73.4	46	4259	<0.0001
Number of seedlings	65.4	53.7	3797	0.0295
Altitude	67.9	51.3	3940	0.0043
Percent litter cover	67.2	51.9	3901	0.0065
Slope	63.3	55.7	3676	0.1136
Percent herb cover	55.2	63.5	3207	0.0923
Percent bare soil	62.5	56.5	3625	0.1584
Canopy height (%)	15.25	84.75	6.9637	0.0083
Percent ground cover (%)	18.64	81.36	1.7696	0.1853
Percent shrub cover (%)	32.2	67.8	3.3987	0.0652
Percent tree canopy cover (%)	55.08	44.92	0.1235	0.7253
Water association(%)	33.9	66.1	1.0715	0.3006
Satellite Image-based Data				
Raw band 1	45.6	72.8	2650	<0.0001
Raw band 2	47.1	71.4	2732	0.0001
Raw band 3	46.3	72.1	2691	<0.0001
Raw band 4	51.9	66.8	3010	0.0178
Raw band 5	50.1	68.5	2906	0.0034
Raw band 7	49.5	69.1	2873	0.0018
Mean(3X3) band 1	46.7	71.8	2709	<0.0001
Mean(3X3) band 2	47	71.5	2728	0.0001
Mean(3X3) band 3	46.8	71.7	2715	<0.0001
Mean(3X3) band 4	52.2	66.5	3030	0.0238
Mean(3X3) band 5	50	68.5	2905	0.0033
Mean(3X3) band 7	49	69.6	2843	0.0011
Mean(5X5) band 1	47.1	71.4	2733	0.0001
Mean(5X5) band 2	46.8	71.7	2714	<0.0001
Mean(5X5) band 3	47	71.5	2729	0.0001
Mean(5X5) band 4	51.9	66.7	3014	0.0189
Mean(5X5) band 5	49	69.5	2845	0.0011
Mean(5X5) band 7	47.7	70.8	2770	0.0002
Mean(7X7) band 1	47.1	71.4	2732	0.0001
Mean(7X7) band 2	46.5	72	2701	<0.0001
Mean(7X7) band 3	47.2	71.3	2738	0.0001
Mean(7X7) band 4	51.3	67.4	2976	0.0106
Mean(7X7) band 5	48.1	70.4	2792	0.0004
Mean(7X7) band 7	47.2	71.3	2742	0.0001
Texture(3X3) band 1	53.3	65.4	3097	0.0569
Texture(3X3) band 2	51.5	67.1	2991	0.0149
Texture(3X3) band 3	53.5	65.3	3103	0.0613
Texture(3X3) band 4	57.5	61.3	3339	0.5484
Texture(3X3) band 5	52.6	66.1	3054	0.0324
Texture(3X3) band 7	53.9	64.8	3131	0.0881
Texture (5X5) band 1	52.7	66	3058	0.0346
Texture (5X5) band 2	51	67.6	2962	0.0086
Texture (5X5) band 3	50.4	68.2	2925	0.0047
Texture (5X5) band 4	59.7	59.2	3465	0.9422
Texture (5X5) band 5	54.3	64.4	3151	0.1075

Continued

Variables:	**Mean or Percent**		**W Scores or c^2 Value**	**P-value**
	High-div area(58)	**Not-high div area(60)**		
Texture (5X5) band 7	53.9	64.9	3127	0.0816
Texture (7X7) band 1	53.4	65.3	3102	0.0607
Texture (7X7) band 2	53.2	65.5	3086	0.0497
Texture (7X7) band 3	52.8	65.9	3062	0.0367
Texture (7X7) band 4	59.6	59.4	3457	0.9764
Texture (7X7) band 5	55.6	63.1	3230	0.2352
Texture (7X7) band 7	54.7	64.1	3174	0.1393

Table 2. List of satellite image-based variables that were included in the interim data set for variable selection process, before initiating model fitting to detect above average tree species richness areas within the humid tropical forest of Amani N.R., East Usambara, Eastern Arc, Tanzania.

	Source	Variables
Satellite image-based variables	-Original Landsat ETM image (without cloud cover)	bands 1, 2, 3, 4, 5, 7
	- Mean image calculated from original image where mean calculated within a 3 X 3 window	bands 1, 3
	- Mean image calculated from original image where mean calculated within a 5 X 5 window	bands 2
	- Mean image calculated from original image where mean calculated within a 7 X 7 window	bands 2
	- Normalized Difference Moisture Index (NDMI)	NDMI1, NDMI2

Normalized Difference Moisture Index or NDMI = (near infrared band – mid infrared band)/ (near infrared band + mid infrared band): NDMI1 calculated with band4 and band 7, NDMI2 with band 4 and band 5.

Table 3. Satellite image-based variable selection based on predictions for above average tree species richness or high tree diversity areas in Amani N.R, East Usambara, Tanzania using logistic regression equation.

	Actual[1]		Total	Correct
Predicted[2]	High tree diversity	Not high diversity	n	%
Landsat ETM band 1 (RB1)-based Model				
High tree diversity area	47	28	75	62.6
Not high diversity area	11	32	43	74.4
Total (n)	58	60	118	
Correct (%)	**81.0**	53.3		66.9
Landsat ETM band 3 (RB3)-based Model				
High tree diversity area	45	26	71	63.3
Not high diversity area	13	34	47	72.3
Total (n)	58	60	118	
Correct (%)	**77.5**	56.6		66.9
Landsat ETM band 7 (RB7)-based Model				
High tree diversity area	42	25	67	62.6
Not high diversity area	16	35	51	68.6
Total (n)	58	60	118	
Correct (%)	**72.4**	58.3		65.3
Landsat ETM band 4 (RB4) and NDMI1 -based Model				
High tree diversity area	22	14	36	61.1
Not high diversity area	36	46	82	56.0
Total (n)	58	60	118	
Correct (%)	**37.9**	76.6		57.6

Number of high tree diversity locations = 58, Not high tree diversity locations = 60 (Above average species tree richness or high tree diversity locations for Amani N.R. are those with $\geq$16 trees per plot).

[1]Actual column records number of above average (high) or average/below average (not-high) diversity areas included in the dataset used to create the model. Actual Correct % for high tree diversity (highlighted) refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

NDMI1 (Normalized Difference Moisture Index) = (band 4 – band7)/ (band 4 + band 7)

Table 4. List of satellite image-based variables that were included in the final data set before initiating model fitting to detect above average tree species richness areas within the humid tropical forest of Amani N.R., East Usambara, Eastern Arc, Tanzania.

	Source	Variables
Field based variables	East Usambara M.S. Access data base (maintained by EUCAMP)	number of trees, number of seedlings, and altitude
Satellite image-based variables	-Original Landsat ETM image (without cloud cover)	bands 4, 7
	- Mean image calculated from original image where mean calculated within a 3 X 3 window	bands 4, 5, 7
	- Mean image calculated from original image where mean calculated within a 5 X 5 window	bands 4, 5, 7
	- Mean image calculated from original image where mean calculated within a 7 X 7 window	bands 4, 5, 7
	- Normalized Difference Moisture Index (NDMI)	NDMI1, NDMI2

All fourteen variables were present for each record in Amani N.R. data set. Records for number of seedlings and altitude were missing from some records of Nilo F.R. data. All Nilo F.R. records included in the analysis had satellite image-based variables and number of tree records.

Normalized Difference Moisture Index or NDMI = (near infrared band – mid infrared band)/ (near infrared band + mid infrared band): NDMI1 calculated with band4 and band 7, NDMI2 with band 4 and band 5.

Table 5. A list of analyses performed to predict areas of high tree species richness or species number in Amani N.R. and Nilo F.R. forests of East Usambara, Tanzania.

Data used:	Amani N.R.	Nilo F.R.	Amani and Nilo combined
Analysis			
Logistic Regression			
	- Satellite model		
		-Extrapolated satellite model (with Nilo F.R. high diversity are defined as $\geq$13 sp per plot)	
		- Extrapolated satellite model (with Nilo F.R. high diversity are defined as $\geq$16 sp p er plot)	
	- Field/satellite model		
		- Extrapolated field/satellite model (with Nilo F.R. high diversity are defined as $\geq$13 sp per plot)	
			- Pooled model (where high diversity area defined as $\geq$15sp per plot
			- Combined model (where high diversity area for Amani defined as $\geq$16 sp and for Nilo $\geq$13sp
	- Satellite model (using moisture index as a variable)		
	- Satellite model (using moisture index as a variable)		
Discriminant Analysis			
	- Satellite model		
		- Extrapolated satellite model (with Nilo F.R. high diversity are defined as $\geq$13 sp per plot)	
	- Field/satellite model (with Nseed as one of the field variables)		
	- Field/satellite model		
		- Extrapolated field/satellite model (with Nilo F.R. high diversity are defined as $\geq$13 sp per plot)	
Linear regression Analysis			
	- Satellite model		
	- Field/satellite model		
K-nearest neighbor Analysis			
	-Satellite model		
	- Field/satellite model		
	- Field/satellite model (with number of tree as dependent and all other variables except number of species as dependent variable		

”Satellite model” refer to all models using one or more of the eleven satellite image-based variables, and “Field/satellite model” refer to models using satellite and field based variables. All field/satellite models except otherwise mentioned used Landsat ETM band 7 and number of trees in the model.

Table 6. Model selection in logistic analysis for predicting above average tree species richness areas within the tropical forest of Amani N.R., East Usambara, Tanzania, using Akaike's Information Criteria (AIC).

Model		**Band**	**AIC**	D_i **AIC**	w_i
Satellite Image-based models		RB7	150.73	0	0.879572
		M55B7	157.954	7.224	0.023747
		M77B7	158.039	7.309	0.022758
		M33B7	158.428	7.698	0.018736
		M33B5	158.790	8.060	0.015634
		M55B5	159.171	8.441	0.012922
		M33B5	159.710	8.980	0.009869
		RB4	159.833	9.103	0.009281
		M77B4	161.951	11.221	0.003219
		M55B4	162.701	11.971	0.002212
		M33B4	162.853	12.123	0.002050
	Field data Used				
Satellite Image	Number of trees	RB7	134.276	0	0.928034
And field data-	-do-	RB4	142.781	8.505	0.013205
based models	-do-	M33B7	143.023	8.747	0.011700
	-do-	M55B7	143.373	9.097	0.009821
	-do-	M55B5	143.749	9.473	0.008138
	-do-	M33B5	144.02	9.744	0.007107
	-do-	M77B5	144.083	9.807	0.006887
	-do-	M77B7	144.198	9.922	0.006502
	-do-	M77B4	145.735	11.459	0.003015
	-do-	M55B4	146.09	11.814	0.002525
	-do-	M33B4	146.283	12.007	0.002292
	Number of seedlings	RB7	149.474	15.198	0.000465
	Altitude	RB7	151.685	17.409	0.000154
	Number of seedlings	RB4	155.017	20.741	0.000029
	-do-	M55B7	155.945	21.669	0.000018
	-do-	M77B7	156.013	21.737	0.000017
	-do-	M77B5	156.124	21.848	0.000016
	-do-	M33B7	156.344	22.068	0.000015
	-do-	M55B5	156.486	22.21	0.000014
	-do-	M33B5	156.904	22.628	0.000013
	-do-	M77B4	157.889	22.613	0.0000069
	-do-	M55B4	158.633	24.357	0.0000047
	-do-	M33B4	158.667	24.391	0.0000046
	Altitude	M55B7	159.376	25.1	0.0000032
	-do-	M77B7	159.387	25.111	0.0000032
	-do-	M33B7	159.953	25.677	0.0000024
	-do-	M77B5	160.023	25.747	0.0000023
	-do-	M55B5	160.481	26.205	0.0000018
	-do-	RB4	160.942	26.666	0.0000015
	-do-	M33B5	161.098	26.822	0.0000013
	-do-	M77B4	163.222	28.946	0.0000004
	-do-	M55B4	164.098	29.822	0.0000003
	-do-	M33B4	164.324	30.048	0.0000002

Δ_i AIC or Delta-AIC = AIC_i - minimum AIC. That is, the difference between AIC values of each i model and the smallest AIC value in the models compared. w_i or Akaike weights = $\exp(-1/2\Delta AIC_i) / \sum_{r=1}^{R} \exp(-1/2\Delta AIC_r)$, where R is the number of models.

Table 7. Logistic regression models for predicting above average tree species richness areas within the tropical forest of Amani N.R., East Usambara, Tanzania.

Logistic Model Variables[2]	**Model Selection** AIC	Overall % Correct[1]	**Model verification** Hosmer & Lemeshow Goodness of Fit Test
Satellite-image-based models			
Variables used (11): RB4, RB7, M33B4, M33B5, M33B7, M55B4, M55B5, M55B7, M77B4, M77B5, M77B7	164.37	54.2	0.41
Variable used (1)[3]: RB7	150.73	65.3	0.89
Satellite-image and field based models			
Variables used (14): RB4, RB7, M33B4, M33B5, M33B7, M55B4, M55B5, M55B7, M77B4, M77B5, M77B7, Ntree, Nseed, Alti	147.38	69.5	0.38
Variable used (2)[3]: RB7, Ntree	134.28	76.3	0.29

[1] Overall percent correct is measured at θ i.e. the probability of high diversity area = 0.5. [The logistic equation: $\theta = 1 / [1 + \exp\{-(\beta_o + \sum_{(j=1 to k)} \beta_j . X_{ij})\}]$, i = 1, 2,…n, where β_o is the beta value of the intercept, β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the independent variables present or selected by the regression process].

[2] Variables: All satellite image-based variables are brightness values recorded for the pixels corresponding to the ground sampled locations for different bands: RB4 and RB7 refers to (raw) values of bands 4 and 7 respectively. Mean values of certain pixels surrounding the sampled pixels were included. If the mean is taken over 3X3 window then '33' appears in the variable name, and M33B4, M33B5, M33B7 refer to mean brightness values over 3X3 pixels taken for bands 4, 5, and 7. Three window sizes are used 3X3, 5X5, and 7X7. Accordingly, M55B4, M55B5, M55B7 and M77B4, M77B5, and M77B7 refer to window sizes 5X5 and 7X7 for bands 4, 5, and 7 respectively. Field Variables: Ntree = number of trees ≥ 10cm dbh, Nseed = number of seedlings, Alti = altitude measured in meters within each plot.

[3] The best model using selected variables was chosen by comparing Akaike Information Criteria (AIC) of different models. The same variables were also selected in the model when logistic stepwise selection procedure was used to select the most appropriate variables.

Table 8. Logistic regression* parameter estimates for predicting high tree diversity areas within the tropical forest of Amani N.R., Tanzania using only satellite image-based variables.

Variables	Parameter Estimates			
	β	SE	χ^2	P-value
Intercept	5.7770	1.6802	11.8215	0.0006
RB 7	-0.1790	0.0519	11.8741	0.0006

*The logistic equation: $\theta = 1 / [1 + \exp\{-(\beta_o + \sum_{(j=1 to k)} \beta_j . X_{ij})\}]$, i = 1, 2,…n, where θ is the probability of high diversity area, β_o is the beta value of the intercept, β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the independent variables selected by the stepwise regression process.

Table 9. Predictions for above average tree species richness or high tree diversity areas in Amani N.R, East Usambara, Tanzania based on logistic equation using Landsat ETM satellite image-based variable, band 7 (RB7).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**42**	25	67	62.6	
Not high diversity area	16	**35**	51	68.6	
Total (n)	58	60	118		
Correct (%)	72.4	58.3		**65.3**	0.31

Number of high tree diversity locations = 58, Not high tree diversity locations = 60 (Above average species tree richness or high tree diversity locations for Amani N.R. are those with $\geq$16 trees per plot).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 10. Results of applying logistic regression model (with satellite image-based variable Landsat ETM band 7 (RB7) as predictor) generated in Amani N.R. to Nilo F.R. area to predict high tree diversity or the above average tree species richness areas in Nilo F.R., East Usambara Tanzania. High tree diversity area in Nilo F.R. has been defined based on the Nilo forest record ($\geq$13 tree species per plot).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**17**	15	32	53.1	
Not high diversity area	11	**26**	37	70.2	
Total (n)	28	41	69		
Correct (%)	60.7	63.4		**62.3**	0.24

Number of high tree diversity locations in Nilo = 28, Not high tree diversity locations = 41 (Above average species tree richness or high tree diversity locations for Nilo F.R. are those with $\geq$13 trees per plot).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 11. Results from applying logistic regression model generated in Amani N.R. to Nilo F.R. area to predict high tree diversity or the above average tree species richness areas in Nilo F.R., East Usambara Tanzania. High tree diversity area in Nilo F.R. has been defined based on the Nilo forest record (≥16 tree species per plot).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**13**	19	32	40.6	
Not high diversity area	10	**27**	37	73.0	
Total (n)	23	46	69		
Correct (%)	56.5	58.6		**58.0**	0.14

Number of high tree diversity locations in Nilo = 23, Not high tree diversity locations = 46 (Above average species tree richness or high tree diversity locations for Nilo F.R. are those with ≥16 trees per plot).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 12. Logistic regression[1] parameter estimates for predicting above average tree species richness areas within the tropical forest in Tanzania from the pooled[2] Amani N.R and Nilo F.R. forests using only satellite image-based variables.

Variables	Parameter Estimates			
	β	SE	χ^2	P-value
Intercept	3.1066	0.9977	9.6962	0.0018
RB 7	-0.0982	0.0302	10.5564	0.0012

1. The logistic equation: $\theta = 1 / [1 + \exp\{ -(\beta_o + \sum_{(j=1 to k)} \beta_j . X_{ij})\}]$, i = 1, 2,…n, where θ is the probability of high diversity area, β_o is the beta value of the intercept, β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the independent variables selected by the stepwise regression process.

2. In the pooled data set, above average species tree richness or high tree diversity locations were those with 15 tree species per plot. Number of high tree diversity locations = 87, Not high tree diversity locations = 100

Table 13. Predictions based on logistic regression for above average tree species richness or high tree diversity areas in the pooled dataset* that include records from Amani N.R, Nilo F.R., East Usambara, Tanzania using Landsat ETM satellite image-based variable, band 7 (RB7).

	Actual[1]		Total correct		$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**47**	31	78	60.0	
Not high diversity area	40	**69**	108	63.8	
Total (n)	87	100	187		
Correct (%)	54.0	69.0		**62.0**	0.23

* In the pooled data set, above average species tree richness or high tree diversity locations were those with $\geq$15 tree species per plot. Number of high tree diversity locations = 87, Not high tree diversity locations = 100

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 14. Logistic regression[1] parameter estimates for predicting above average tree species richness areas within the tropical forest in Tanzania from the combined[2] Amani N.R and Nilo F.R. forests using only satellite image-based variables.

Variables	Parameter Estimates			
	β	SE	χ^2	P-value
Intercept	4.4724	1.1392	15.4130	<0.0001
RB 7	-0.1410	0.0349	16.3327	<0.0001

1. The logistic equation: $\theta = 1 / [1 + \exp\{ -(\beta_o + \sum_{(j=1 to k)} \beta_j . X_{ij})\}]$, i = 1, 2,…n, where θ is the probability of high diversity area, β_o is the beta value of the intercept, β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the independent variables selected by the stepwise regression process.

2. In the combined data set, above average species tree richness or high tree diversity locations were defined based on the individual forest the records came from. Plots from Amani N.R. with $\geq$16 tree species per plot and from Nilo F.R. with $\geq$13 tree species per plot were defined as high diversity areas or areas with above average species richness. Number of high tree diversity locations = 86, Not high tree diversity locations = 101.

Table 15. Predictions based on logistic regression for above average tree species richness or high tree diversity areas in the combined dataset* that include records from Amani N.R, Nilo F.R., East Usambara, Tanzania using Landsat ETM satellite image-based variable, band 7 (RB7).

	Actual[1]		Total correct		$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**47**	31	78	60.2	
Not high diversity area	39	**70**	109	64.2	
Total (n)	86	101	187		
Correct (%)	54.6	69.3		**62.6**	0.24

*In the combined data set, above average species tree richness or high tree diversity locations were defined based on the individual forest the records came from. Plots from Amani N.R. with $\geq$16 tree species per plot and from Nilo F.R. with $\geq$13 tree species per plot were defined as high diversity areas or areas with above average species richness. Number of high tree diversity locations = 86, Not high tree diversity locations = 101.

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 16 Linear discriminant function* parameters to predict high tree diversity areas within the tropical forest of Amani N.R., Tanzania using only satellite image-based variables.

Variable	Parameter Estimates	
	High tree diversity (=1)	Not high diversity (=0)
Constant	-18.07899	-22.73494
RB7	1.16964	1.31163

***The linear discriminant function can be written as: = constant + $\beta_1 X_1 + \beta_2 X_2 \ldots + \beta_j X_j$, where β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the discriminating variable.

Table 17. Predictions for above average tree species richness or high tree diversity areas in Amani N.R., East Usambara, Tanzania based on discriminant analysis using Landsat ETM satellite image-based parameter band 7 (RB7).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**42** (61.5%)	25	67	62.6	
Not high diversity area	16	**35** (65.2%)	51	68.6	
Total (n)	58	60	118		
Correct (%)	72.4	58.3		**65.3**	0.31

Number of high tree diversity locations in Amani N.R. = 58, Not high tree diversity locations = (Above average species tree richness or high tree diversity locations for Amani N.R. are those with $\geq$16 trees per plot). Numbers in parenthesis indicate posterior probabilities of observations of belonging to the group that were classified.

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 18. Results from applying discriminant function generated in Amani N.R. to Nilo F.R. area to predict high tree diversity or the above average tree species richness areas in Nilo F.R., East Usambara Tanzania using only satellite image-based variable (Landsat ETM band 7, RB7). High tree diversity area in Nilo F.R. has been defined based on the Nilo forest record ($\geq$13 tree species per plot).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**17**	15	32	53.1	
Not high diversity area	11	**26**	37	70.2	
Total (n)	28	41	69		
Correct (%)	60.7	63.4		**62.3**	0.24

Number of high tree diversity locations in Nilo = 28, Not high tree diversity locations = 41 (Above average species tree richness or high tree diversity locations for Nilo F.R. are those with $\geq$13 trees per plot).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 19. Logistic regression* parameter estimates for predicting high tree diversity areas within the tropical forest of Amani N.R., East Usambara, Tanzania using satellite image and field-based variables.

Variables	Parameter Estimates			
	β	SE	χ^2	P-value
Intercept	3.9542	1.9634	4.0562	0.0440
RB 7	-0.2100	0.0630	11.0961	0.0009
Ntree	0.0647	0.0170	14.4259	0.0001

*The logistic equation: $\theta = 1 / [1 + \exp\{-(\beta_o + \sum_{(j=1 to k)} \beta_j . X_{ij})\}]$, i = 1, 2,…n, where θ is the probability of high diversity area, β_o is the beta value of the intercept, β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the independent variables selected by the stepwise regression process.

Table 20. Predictions for above average tree species richness or high tree diversity areas in Amani N.R, East Usambara, Tanzania based on logistic equation using Landsat ETM satellite image-based variable, band 7 (RB7) and rapidly collectable field parameter, number of trees (≥10cm dbh).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**47**	17	64	73.4	
Not high diversity area	11	**43**	54	79.6	
Total (n)	58	60	118		
Correct (%)	81.0	71.6		**76.3**	0.53

Number of high tree diversity locations = 58, Not high tree diversity locations = 60 (Above average species tree richness or high tree diversity locations for Amani N.R. are those with ≥16 trees per plot).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 21. Results of applying logistic regression model (with satellite image-based variable Landsat ETM band 7 (RB7) and field variable, number of trees (Ntree) as predictors) generated in Amani N.R. to Nilo F.R. area to predict high tree diversity or the above average tree species richness areas in Nilo F.R., East Usambara Tanzania. High tree diversity area in Nilo F.R. has been defined based on the Nilo forest record ($\geq$13 tree species per plot).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
Model					
High tree diversity area	**15**	03	18	83.4	
Not high diversity area	13	**38**	51	74.5	
Total (n)	28	41	69		
Correct (%)	53.5	92.6		**76.8**	0.49

Number of high tree diversity locations in Nilo = 28, Not high tree diversity locations = 41 (Above average species tree richness or high tree diversity locations for Nilo F.R. are those with $\geq$13 trees per plot).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 22. Logistic regression[1] parameter estimates for predicting above average tree species richness areas within the tropical forest in Tanzania from the pooled[2] Amani N.R and Nilo F.R. forests using both satellite image-based and field variables.

Variables	Parameter Estimates			
	β	SE	χ^2	P-value
Intercept	0.0283	1.2777	0.0005	0.9823
RB 7	-0.0885	0.0367	5.8133	0.0159
Ntree	0.0710	0.0128	30.9334	<0.0001

1. The logistic equation: $\theta = 1 / [1 + \exp\{ -(\beta_o + \sum_{(j=1tok)} \beta_j . X_{ij})\}]$, i = 1, 2,…n, where θ is the probability of high diversity area, β_o is the beta value of the intercept, β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the independent variables selected by the stepwise regression process.

2. In the pooled data set, above average species tree richness or high tree diversity locations were those with $\geq$15 tree species per plot. Number of high tree diversity locations = 87, Not high tree diversity locations = 100

Table 23. Predictions based on logistic regression for above average tree species richness or high tree diversity areas in the pooled dataset* that include records from Amani N.R, Nilo F.R., East Usambara, Tanzania using Landsat ETM satellite image-based (band 7, RB7) and field variables (number of trees, Ntree).

	Actual[1]		Total correct		$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
Model					
High tree diversity area	**61**	26	87	70.0	
Not high diversity area	26	**74**	100	74.0	
Total (n)	87	100	187		
Correct (%)	70.1	74.0		**72.1**	0.44

* In the pooled data set, above average species tree richness or high tree diversity locations were those with $\geq$15 tree species per plot. Number of high tree diversity locations = 87, Not high tree diversity locations = 100

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 24. Logistic regression[1] parameter estimates for predicting above average tree species richness areas within the tropical forest in Tanzania from the combined[2] Amani N.R and Nilo F.R. forests using both satellite image-based and field variables.

Variables	Parameter Estimates			
	β	SE	χ^2	P-value
Intercept	2.0262	1.4082	2.0703	0.1502
RB 7	-0.1533	0.0433	12.5570	<0.0004
Ntree	0.0722	0.0131	30.4966	<0.0001

1. The logistic equation: $\theta = 1 / [1 + \exp\{ -(\beta_o + \sum_{(j=1 to k)} \beta_j . X_{ij})\}]$, i = 1, 2,…n, where θ is the probability of high diversity area, β_o is the beta value of the intercept, β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the independent variables selected by the stepwise regression process.

2. In the combined data set, above average species tree richness or high tree diversity locations were defined based on the individual forest the records came from. Plots from Amani N.R. with 16 tree species per plot and from Nilo F.R. with $\geq$13 tree species per plot were defined as high diversity areas or areas with above average species richness. Number of high tree diversity locations = 86, Not high tree diversity locations = 101.

Table 25. Predictions based on logistic regression for above average tree species richness or high tree diversity areas in the combined dataset* that include records from Amani N.R, Nilo F.R., East Usambara, Tanzania using Landsat ETM satellite image-based (band 7 ,RB7) and field variable (number of trees, Ntree).

	Actual[1]		Total correct		$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**65**	24	89	73.0	
Not high diversity area	21	**77**	98	78.5	
Total (n)	86	101	187		
Correct (%)	75.5	76.2		**75.9**	0.52

*In the combined data set, above average species tree richness or high tree diversity locations were defined based on the individual forest the records came from. Plots from Amani N.R. with 16 tree species per plot and from Nilo F.R. with $\geq$13 tree species per plot were defined as high diversity areas or areas with above average species richness. Number of high tree diversity locations = 86, Not high tree diversity locations = 101.

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 26. Discriminant analysis results of locating high tree diversity areas in Amani, N.R, East Usambara Tanzania based on satellite image (raw band 7) and field-based predictors (number of trees and number of seedlings).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**46** (68.7%)	22	68	67.6	
Not high diversity area	12	**38** (73.6%)	50	76.0	
Total (n)	58	60	118		
Correct (%)	79.3	63.3		**71.1**	0.43

Number of high tree diversity locations = 58, Not high tree diversity locations = 60 (Number of species per plot $\geq$16 considered as high tree diversity location for Amani N.R). Percent figures in parenthesis indicate posterior probability of the observations belonging to that correctly classified group. Number if parentheses indicate posterior probability of observations belonging to the group that it has been classified.

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 27. Linear discriminant function* for Nspdiv, where Nspdiv=1 are the high tree diversity areas (and Nspdiv=0 are not) in Amani, N.R, East Usambara Tanzania based on satellite image variable (raw band 7: RB7) and field based variable (number of trees: Ntree).

Variable	Nspdiv	
	High tree diversity (=1)	Not high diversity (=0)
Constant	-24.48385	-26.59101
RB7	1.19750	1.33325
Ntree	0.24685	0.19155

*The linear discriminant function can be written as: = constant + $\beta_1 X_1 + \beta_2 X_2 \ldots + \beta_j X_j$, where β_j is the beta value of the *j* independent variables, and X_{ij} are data value of the discriminating variables.

Table 28. Discriminant analysis results of locating high tree diversity areas in Amani, N.R, East Usambara Tanzania based on satellite image (raw band 7) and field-based predictor (number of trees).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**49** (67.3%)	19	68	72.0	
Not high diversity area	9	**41** (71%)	50	82.0	
Total (n)	58	60	118		
Correct (%)	84.4	68.3		**76.3**	0.53

Number of high tree diversity locations = 58, Not high tree diversity locations = 60 (Number of species per plot $\geq$16 considered as high tree diversity location for Amani N.R). Percent figures in parenthesis indicate posterior probability of the observations belonging to that correctly classified group. Number if parentheses indicate posterior probability of observations belonging to the group that it has been classified.

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 29. The results of extrapolating field/satellite discriminant analysis model (based on band7 and Number of trees) from Amani N.R. to Nilo F.R. area to predict high tree diversity areas in Nilo F.R., East Usambara, Tanzania. High tree diversity area in Nilo F.R. based on Nilo forest records ($\geq$13 trees species per plot).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**16**	03	19	84.2	
Not high diversity area	12	**38**	50	76.0	
Total (n)	28	41	69		
Correct (%)	57.1	92.6		**78.3**	0.53

Number of high tree diversity locations in Nilo = 28, Not high tree diversity locations = 41 (Number of species per plot $\geq$13 considered as high tree diversity location for Nilo).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 30. Multiple linear regression analysis results for predicting tree species number per plot in Amani, N.R, East Usambara Tanzania based on satellite image-based variable (Landsat ETM band 7, RB&).

R-Square = 0.11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	291.6849	291.684	15.41	0.0001
Error	116	2195.8828	18.9300		
Corrected Total	117	2487.5678			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	24.61084	2.45324	1905.13195	100.64	<.0001
RB7	-0.28946	0.07374	291.68496	15.41	0.0001

Table 31 Predictions for above average tree species richness or high tree diversity areas in Amani N.R, East Usambara, Tanzania based the predicted tree species numbers generated by the multiple linear regression model using Landsat ETM satellite image-based variable, band 7 (RB7).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**32**	19	51	62.7	
Not high diversity area	26	**41**	67	61.1	
Total (n)	58	60	118		
Correct (%)	55.1	68.3		**61.8**	0.24

Number of high tree diversity locations = 58, Not high tree diversity locations = 60 (Above average species tree richness or high tree diversity locations for Amani N.R. are those with $\geq$16 trees per plot).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 32 Multiple linear regression analysis results for predicting tree species number per plot in Amani, N.R, East Usambara Tanzania based on satellite image-based variable (Landsat ETM band 7, RB7) and field data (number of trees per plot, Ntree).

R-Square = 0.35

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	892.43902	446.21951	32.17	<.0001
Error	115	1595.12877	13.87068		
Corrected Total	117	2487.56780			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	16.08336	2.46756	589.27416	42.48	<.0001
RB7	-0.21866	0.06403	161.75535	11.66	0.0009
Ntree	0.14676	0.02230	600.75407	43.31	<.0001

Table 33 Predictions for above average tree species richness or high tree diversity areas in Amani N.R, East Usambara, Tanzania based the predicted tree species numbers generated by the multiple linear regression model using Landsat ETM satellite image-based variable, band 7 (RB7) and rapidly collectable field parameter, number of trees (≥10cm dbh).

	Actual[1]		Total	Correct	$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
High tree diversity area	**43**	17	60	71.6	
Not high diversity area	15	**43**	58	74.1	
Total (n)	58	60	118		
Correct (%)	74.1	71.6		**72.8**	0.34

Number of high tree diversity locations = 58, Not high tree diversity locations = 60 (Above average species tree richness or high tree diversity locations for Amani N.R. are those with ≥16 trees per plot).

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 34. Results of K-nearest neighbor analysis with combined Amani N.R. and Nilo F.R. data set using Landsat ETM band 7 (RB7) and number of trees per plot (Ntree) as predictors, where K nearest neighbor used to calculate species per plot is six (K=6).

Number of Nearest Neighbor used	Actual Mean	Calculated Mean	RMSE	Bias	Std.Error of Bias	r
6	13.8395	13.875	3.657	0.035	0.267	0.68

Table 35. Results of correlation between actual tree species numbers, and numbers predicted by the K-nearest neighbor analysis with combined Amani N.R. and Nilo F.R. data set using Landsat ETM band 7 (RB7) and number of trees per plot (Ntree) as predictors, where K nearest neighbor used to calculate species per plot is six (K=6).

	High diversity plots	Not-high diversity plots
Correlation (r) between actual and predicted tree species numbers corresponding to high tree diversity areas (high diversity: $\geq$15 tree species per plot)	0.23	0.69
Correlation (r) between actual and predicted tree species numbers corresponding to high tree diversity areas (high diversity: $\geq$16 tree species per plot for Amani N.R. records and $\geq$13 tree species per plot for Nilo F.R. records)	0.20	0.68

Table 36. Predictions based on actual tree species numbers from Amani N.R. and Nilo F.R., East Usambara, Tanzania, grouped as the pooled dataset* then compared to the corresponding group based on predicted tree species numbers generated by the K-nearest neighbor analysis using Landsat ETM satellite image-based (band 7, RB7) and field variables (number of trees, Ntree).

	Actual[1]		Total correct		$\hat{K}$
Predicted[2]	High tree diversity	Not high diversity	n	%	
Model					
High tree diversity area	**65**	35	100	74.7	
Not high diversity area	22	**65**	87	65.0	
Total (n)	87	100	187		
Correct (%)	74.7	65.0		**69.5**	0.39

* In the pooled data set, above average species tree richness or high tree diversity locations were those with $\geq$15 tree species per plot. Number of high tree diversity locations = 87, Not high tree diversity locations = 100

[1] The actual column contains the number of above average (high) or average/below average (not-high) diversity areas in the dataset used to create the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2]Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 37. Predictions based on actual tree species numbers from Amani N.R. and Nilo F.R., East Usambara, Tanzania, grouped as the combined dataset* then compared to the corresponding group based on predicted tree species numbers generated by the K-nearest neighbor analysis using Landsat ETM satellite image-based (band 7, RB7) and field variables (number of trees, Ntree).

	Actual		Total correct		$\hat{K}$
Predicted	High tree diversity	Not high diversity	n	%	
Model					
High tree diversity area	**58**	25	83	69.8	
Not high diversity area	26	**78**	104	75.0	
Total (n)	84	103	187		
Correct (%)	69.0	75.7		**72.7**	0.45

* In the combined data set, above average species tree richness or high tree diversity locations were defined based on the individual forest the records came from. Plots from Amani N.R. with ≥ 16 tree species per plot and from Nilo F.R. with ≥ 13 tree species per plot were defined as high diversity areas or areas with above average species richness. Number of high tree diversity locations = 86, Not high tree diversity locations = 101.

Actual column records actual number of above average (high) or average/below average (not-high) diversity areas included in the model. Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are included in the model.

Predicted row records the number of above (high) or average/below average (not-high) diversity areas that are predicted by the model. Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Table 38. Comparison of different models and analytical methods to predict above average tree species richness in Amani and Nilo forests, East Usambara, Tanzania, using either only Landsat ETM band 7 (RB7) or using both RB7 and rapidly collectible field variable, number of trees as predictors.

	Total % correct	Actual (% correct)[1]		Predicted (% correct)[2]		$\hat{K}$
		High diversity	Not-high diversity	High diversity	Not-high diversity	
Satellite Image-based Model						
Amani N.R. Logistic	65.3	72.4	58.3	62.6	68.6	0.31
Amani N.R. Discriminant	65.3	72.4	58.3	62.6	68.6	0.31
Amani N.R. Multiple-Linear	61.8	55.1	68.3	62.7	61.1	0.24
Nilo F.R. Logistic	62.3	60.7	63.4	53.1	70.2	0.24
Nilo F.R. Discriminant	62.3	60.7	63.4	53.1	70.2	0.24
Amani-Nilo pooled[3] Logistic	62.0	54.0	69.0	60.0	63.8	0.23
Amani-Nilo combined[4] Logistic	62.6	54.6	69.3	60.2	64.2	0.24
Field/Satellite Image-based Model						
Amani N.R. Logistic	76.3	81.0	71.6	73.4	79.6	0.53
Amani N.R. Discriminant	76.3	84.4	68.3	72.0	82.0	0.53
Amani N.R. Multiple-Linear	72.8	74.1	71.6	71.6	74.1	0.34
Nilo F.R. Logistic	76.8	53.4	92.6	83.4	74.5	0.49
Nilo F.R. Discriminant	78.3	57.1	92.6	84.2	76.0	0.53
Amani-Nilo pooled[3] Logistic	72.1	70.1	74.0	70.0	74.0	0.44
Amani-Nilo combined[4] Logistic	75.9	76.2	75.5	73.0	78.5	0.52
Amani-Nilo pooled[3] KNN	69.5	74.7	65.0	74.7	65.0	0.39
Amani-Nilo combined[4] KNN	72.7	69.0	75.7	69.8	75.0	0.45

[1] Actual Correct % for high tree diversity refers to the percent of known above average tree species richness areas that are predicted by the model to contain high /above-average tree species richness.

[2] Predicted Correct % refers to the percent of above or average/below average tree diversity areas predicted by the model are actually correct.

[3] In the pooled data set, above average species tree richness or high tree diversity locations were those with ≥ 15 tree species per plot. Number of high tree diversity locations = 87, Not high tree diversity locations = 100

[4] In the combined data set, above average species tree richness or high tree diversity locations were defined based on the individual forest the records came from. Plots from Amani N.R. with ≥ 16 tree species per plot and from Nilo F.R. with ≥ 13 tree species per plot were defined as high diversity areas or areas with above average species richness. Number of high tree diversity locations = 86, Not high tree diversity locations = 101.

$\hat{K}$ = (overall classification accuracy – expected classification accuracy)/ (1- expected classification accuracy)

Curriculum Vitae

NINA SENGUPTA
911 University City Boulevard. Apt A-11
Blacksburg. VA 24060. USA.
FAX 540-231-7580 Phone 540-961-0615 E-mail nsengupt@vt.edu

SUMMARY
Training and work experience in: tropical biodiversity conservation, management, and planning issues including rapid biodiversity assessment (using satellite and ground data), spatial patterns of biodiversity, watershed conservation, integrating conservation with sustainable development goals in participation with local people, recommending improvements in Environmental Impact Assessments for biodiversity.

WORK EXPERIENCE

1.Teaching Assistant, and Research Associate at Virginia Tech (Since August 1996): for courses like Conservation of Biodiversity, Human dimensions in Wildlife Management, Population Ecology, and so on. (Served as a student member of the University Council, a prime decision-making body of the university). Currently a research associate in NASA-funded "Carbon from community" project in Mali.

2. Short-term Consultant (biodiversity assessment)

- The World Bank, Washington, D.C. (2002): Reviewed selected projects from south Asian countries to assess their conformity with the existing biodiversity/natural habitat related Bank policies and recommended some necessary changes.
- United Nations Development Program/GEF,India (1999): Participated in a project development (PDF) activity for spatial biodiversity information based decision-making system for the tropical forests of the Andaman Islands. Base work including satellite image interpretation for the area was initiated and carried out since 1996 in collaboration with Virginia Tech and Society for Environment and Development, a local NGO in India.
- The Rijksherbarium/Hortus Botanicus, Netherlands (1997): Reviewed environmental assessment reports of projects funded by the World Bank and International Finance Corporation in limestone or karst areas of south and Southeast Asian countries. The project results were discussed at an international conference in Hanoi, Vietnam in 1999 and were included in a World Bank publication called Biodiversity and Cultural Property in the Management of Limestone Resources, Lessons from East Asia'.
- The World Bank, Washington, D.C. (1995): Reviewed environmental assessment reports from almost all South and Southeast Asian countries to assess the extent to which biodiversity issues were addressed and suggested options for improvement. Suggestions were adapted as guidelines in the Bank for future projects involving biodiversity issues. Both this and the assessment done in 1997 were later incorporated in the World Bank's 'Biodiversity and Environmental Assessment Toolkit' (http://wbln0018.worldbank.org/essd/essd.nsf/Biodiversity/Front+Page)

3. Ecologist - Development Alternatives – DA. (1994) a large multi-disciplinary NGO based in India. *Selected projects undertaken are listed later.*

4. Research Student -Dept. of Ecology, French Institute (1990) – vegetation cover analysis with visual interpretation of SPOT satellite image and ground data.

Voluntary work/Community Involvement - Speaker and film coordinator for International Club, a student organization at Virginia Tech, interested in global issues. Volunteer in a local community-based, not-for-profit movie theatre and whenever possible in other community events. Also, involved with

Association of India's Development, a group that monitor small grass-root level developmental projects in India for funding.

ACADEMIC RECORDS

1. **Doctoral candidate in Wildlife Science at Virginia Tech,** USA (**completed** 2003). Research involved integrating field and satellite data from within rainforest of East Usambara, Tanzania to identify and test high biodiversity areas. Results will help rapid assessment, monitoring, and proactive management in the data-poor rainforest areas. Project collaborators: Virginia Tech, UNDP-East Africa, EUCAMP-Tz, Frontier-Tz, and CARE.
2. **Master of Science in Ecology at Salim Ali School of Ecology,** Pondicherry University, India. Thesis was on behavior and conservation of the endangered Gaur (*Bos gaurus gaurus)*, the largest of the bovines. Worked on a rare high-altitude population in Western Ghat, India.
3. **Bachelor of Science at Calcutta University,** Calcutta, India. Geography, Economics, and Political Science.

RECENT HONORS & SELECTED WORKSHOP/TRAINING

1. Part of winning team in the **World Bank Development Marketplace 2003** for "C-neutral bio-diesel project" (#1551) where natural resource monitoring and conservation are integral (www.developmentmarketplace.org/).
2. **2003 Women's Leadership Award** from American Association of University Women and YMCA.
3. **K-nearest neighbor** (non-parametric) **method** developed/adopted and used to assess forestry parameters in temperate regions. The method has potential for use in natural tropical forests. Received training on invitation from Prof E.Tomppo, who developed the application at METLA, Finnish Forest Research Institute at Helsinki in June-July 2001.
4. Invited participant to the international workshop at Edinburgh, U.K. in 1997 on the prospects of **Rapid biodiversity assessment of tropical rainforests**.
5. Scholarship from **The World Bank Graduate Scholarship Program (WBGSP)** 1994-1996.
6. Wildlife management and its application in Tanzania (November-December 1991). Organized by the **Wildlife Conservation and Management Training Program**, Smithsonian Institution, USA.

SELECTED PROJECTS UNDERTAKEN (Over the years)

Integrating Conservation Objectives with Development/Management Options:
(Developing conservation/management modules and employment generation options for local people.)

- ***Nature Reserves***
  (1) *Jaldapara Asian Rhino Sanctuary, India*. Involved rapid survey of the sanctuary and participatory assessment in adjoining villages, critical analysis of all information. Project funded by IUCN.
  (2) *Utilization of Wilderness in Wandoor, a village in the Andaman Islands, India*. Involved assessment of forest and wilderness resources in and around the village, need and wilderness resource use patterns by villagers. Wandoor village adjoining the Wandoor Marine National Park was not included within the Park management. Recommendations to Forest Dept. lead to its inclusion later. Project funded by Development Alternatives in collaboration with ANET, a local NGO.
- ***Micro Enterprise***
  (3) *Biodiversity & Micro Enterprise:* Worked in a team on an initial report on biodiversity, development and inter-linked social aspects in order to identify a few biomass-based micro-enterprises that may be conducive to conservation (and gender equity) in Madhya Pradesh, India.

- **Other Sectors**

  (4) *Impact of Mining on Tribal Community in Rajmahal, Bihar*. Developed a realistic compensation and training package for the local 'santhal' tribal community about to be relocated due to a coalmine expansion. Options were sensitive to their economic, socio-cultural needs and environmental concerns. Project funded by a Canadian mining firm working with the Eastern Coalfields Ltd., India.

* *Publications on these projects are available with Development Alternatives / funding a gencies*

LANGUAGES KNOWN

Bengali, English, Hindi.

COUNTRY EXPERIENCE

India (extensive), Tanzania (dissertation work + was trained on wildlife management issues). Worked through consultancies on biodiversity and natural resource conservation issues of: China, South Korea, Philippines, Malaysia, Indonesia, Vietnam, Thailand, Bhutan, Bangladesh, and Sri Lanka. Recently received training on some new forest analysis method in METLA, Finland.