Do we know enough?

The impacts of data biases, insufficient sampling and degradation on biodiversity estimates in Tanzanian forests – implications for conservation planning

ANTJE AHRENDS
Ph.D.

THE UNIVERSITY of York

Environment Department

February 2010

Abstract

In the face of continued loss of biodiversity and limited conservation resources, conservation needs to be as efficient as possible. This relies on an unbiased knowledge of the distribution of biodiversity and the threats to it. While we cannot afford to delay conservation action until better data are available, it is essential to understand the limitations of and mitigate biases in the data that we base critical conservation decisions on. This thesis analyses the implications of data biases, insufficient data and degradation and formulates recommendations for conservation planning. While it focuses on vascular plants in the coastal forests and Eastern Arc Mountains in Tanzania, an area recognised internationally for its high biodiversity, the concepts presented are generic with application outside the study area and across other taxonomic groups.

The thesis' findings indicate that the plant data used to underpin conservation decisions in the study area are severely biased and that some of the data are insufficiently representative to capture the true patterns. Funding, botanist and sampling intensity biases, which are circularly linked, may partly explain why despite decades of research surprisingly little consensus has been reached on patterns in biodiversity. In the study area, these effects account for a substantial proportion in the variation in perceived plant diversity patterns with alarming consequences for the reliability of conservation priority assessments. This and the evidence for quickly spreading degradation highlight the urgent need for more efficient surveys and systematic conservation planning.

The recommendations include focusing surveys on under-researched locations, distributing botanical expertise more equally across the study area, and using pilot studies prior to surveys to determine the minimum needed sample size. Most importantly, coordinated efforts are needed to develop conservation goals and strategically plan surveys in order to mitigate data biases, and avoid duplication and ad-hoc donor-driven conservation planning. These recommendations may be applicable to other regions and even to the global scale, where partly competing conservation prioritisation schemes determine where billions of dollars in conservation investments are spent, each based on their own data and analyses. Problems associated with these schemes are seldom communicated lest public support may be compromised; however, understanding and acknowledging the gaps in our knowledge on the distribution of biodiversity and threats to it are essential to conserve the world's biodiversity more effectively.

Contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendix

List of tables

Chapter 1

No tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Acknowledgements

This study would not have been possible without the help and support of numerous individuals and organisations. Funding was provided by a Marie Curie Actions grant to my supervisor Dr. Rob Marchant (contract MEXT-CT-2004-517098). I would like to thank him for his support, advice and guidance. I am further grateful to Prof. Chris Thomas and Prof. Jon C. Lovett for additional supervision. Special thanks go to Prof. Carsten Rahbek and Prof. Neil D. Burgess for inspiration, guidance and fruitful collaboration on the papers.

This thesis is based on data that has been collected over many years by numerous individuals and organisations. I am indebted to Frontier Tanzania, Missouri Botanical Garden, Museo Tridentino di Scienze Naturali, the Forest and Beekeeping Division of Tanzania, the Tanzanian Forest Conservation Group, the Wildlife Conservation Society of Tanzania, WWF Tanzania, and Jon Lovett, Andrew Marshall, Michele Menegon, John Hall, James Taplin, Phil Clarke, Victoria Wilkins Kindemba, Nisha Owen, Nike Doggart, Claire Bracebridge, Kathryn Doody, Gerry Hertel, Trevor Jones, Kerstin Pfliegner and Francesco Rovero amongst many others for sharing their field data with me. Roy Gereau produced a fully revised list of accepted names and endemism status for the Eastern Arc Mountain plant taxa, which constituted a major advance for plant research in the region and was key to this thesis. I am further grateful to Phil Platts and Jon Green for assisting with the cleaning of spatial coordinates, and to the Valuing the Arc project (notably, Ruth Swetnam, Phil Platts, Boniface Mbilyini, Neil Burgess and Andrew Balmford) for providing GIS layers.

Throughout the phases of my own fieldwork, I received support from numerous organisations and individuals. I am grateful to Boniface Mhoro, Alex Ngovi, Janes Mdanga, ‘Mkude’, Phil Platts, Omar Majani, Mhiua Mapunda, Paul Kihara and Jemma Finch, who provided invaluable assistance to the fieldwork, and Frank Mbago, Suleiman Omari, Jon Lovett, Kaj Vollesen, Sally Bidgood and Henk Beentje who assisted me with the specimen identification in the herbarium. Administrative support was provided by the Institute of Resource Assessment of the University of Dar es Salaam and by the WWF Tanzania Project Office, notably Peter Sumbi to whom I owe special thanks. I would further like to thank Simon Milledge, TRAFFIC East and Southern Africa, for lending us field equipment. I am grateful to the Tanzanian Commission for Science and Technology (COSTECH) and the Director of the Tanzanian Forest and Beekeeping Division for granting me permission to carry out the field work (permit number 2006-277-NA-2006-93). Field work for Chapter 5 was already carried out during my M.Sc. (COSTECH permit number 2004-272-MSF-2004-50), and I am extremely grateful to Neil Burgess, Boniface Mhoro, Simon Milledge, Haji Mkuagura, Agostino John, Peter Sumbi, Charles Meshak, Nike Doggart, Stéphanie Duvail, Olivier Hamerlynck, Kathryn Doody, Graham Anderson, Jason Rubens amongst many others for facilitating this work.

I am indebted to the co-authors of the papers for the inspiration they provided, their advice on analysis, presentation and discussion and for patiently correcting my Denglish, most notably Roy Gereau, Phil Platts, Mark Bulling and Simon Lewis. This work would not have been possible without them. Dave Hay and Alan Sneath have been stars sorting out computing resources for me. My sincere gratitude also to Pete Hollingsworth, Toby Pennington, Chris Ellis, James Richardson and other colleagues from the Royal Botanic Garden Edinburgh for immensely helpful discussions and their support in finishing this work; and to my examiners Andrew Balmford, Shonil Bhagwat and Dave Raffaelli for their very helpful comments on the thesis.

Special thanks go to my KITE colleagues and friends Phil Platts, Alistair Jump, Cassian Mumbi, Stephen Rucina, Andy Marshall and especially Jemma Finch, who were simply the best that could have happened to me. Finally, thanks to my family and Mark Bulling, whose encouragement, support and love were invaluable.

Author's declaration

This thesis is the result of my own work, except where due acknowledgement is made, and it has not been previously included in a thesis, dissertation or report submitted to this university or any other institution for a degree, diploma or other academic qualification. Data underlying Chapter 5 have already been collected during my M.Sc., but the analysis presented in Chapter 5 has been part of this thesis. Numerous colleagues have kindly contributed guidance, advice, ideas and data, and they have been included as co-authors on each chapter. However, I have played the dominant role in all data analysis, interpretation and writing.

Signed

Antje Ahrends

Chapter 1 – Introduction

Parts of this chapter in prep. for:

Ahrends, A., Jump, A.S., Hollingsworth, P.M., Lovett, J.C., Burgess, N.D., Gereau, R.E., Bulling, M.T. & Marchant, R. ‘Major gaps in our knowledge of plant biodiversity in the Eastern Arc Mountains: implications for conservation and future research’ in prep. for *Journal of the East African Natural History Society*.

Background

Conservation is at a critical stage. Since the first United Nations conference on the environment in Stockholm in 1972 international conservation goals have progressively been formalised in a series of Multilateral Environmental Agreements, and numerous new non-governmental conservation organisations formed (Niekisch 2000). Ambitious goals have been set, such as the '2010 target' of *"achieving by 2010 a significant reduction of the current rate of biodiversity loss at the global, regional and national levels as a contribution to poverty alleviation and to the benefit of all life on Earth"* (Convention on Biological Diversity Decision VI/26; World Summit on Sustainable Development UN Doc.A/CONF.199.20). The availability of primary biodiversity data for effective conservation planning has also dramatically improved through the digitisation and consolidation in global data portals of herbarium and museum data (Bisby 2000; Graham *et al.* 2004; Lughadha & Miller 2009). Yet, it is unlikely that the unprecedented pace of biodiversity loss that characterises the 20th and 21st centuries has declined significantly (Millenium Ecocosystem Assessment 2005; Mooney 2010). On the contrary - the rate at which we are losing biodiversity is projected to increase in the face of global environmental change (Brook *et al.* 2008; Stork *et al.* 2009; Stork 2010). This and the limited resources available (James *et al.* 1999) force us to focus conservation efforts on the most important areas in greatest need (Margules & Pressey 2000) and make it imperative to conserve biodiversity as effectively as possible

Effective conservation depends on reliable information on the distribution of biodiversity (Wilson 1985; Pimm & Lawton 1998) and threats (Wilson *et al.* 2005). At least a dozen schemes have been proposed to prioritise areas based on their conservation importance (Davis *et al.* 1997; Bryant *et al.* 1998; Stattersfield *et al.* 1998; Fishpool & Evans 2001; Mittermeier *et al.* 2002; Olson & Dinerstein 2002; Sanderson *et al.* 2002; Eken *et al.* 2004; Plantlife International 2004; Hoekstra *et al.* 2005; Mittermeier *et al.* 2005; Ricketts *et al.* 2005). A number of these schemes (Stattersfield *et al.* 1998; Olson & Dinerstein 2002; Mittermeier *et al.* 2005) have become very influential in guiding global conservation action and funds; for example, biodiversity hotspots (Mittermeier *et al.* 2005) alone have raised over US$ 750 million of funding since 1988 (Myers & Mittermeier 2003). Two components that are critical to most schemes are measures of irreplaceability and vulnerability. Most schemes prioritise high levels of irreplaceability (frequently levels of endemism), with vascular plants playing a particularly important role (Brooks *et al.* 2006) as they are amongst the better known taxonomic groups (Lughadha 2004; Barthlott *et al.* 2005; Kier *et al.* 2005), and their use as an indicator of diversity in other groups of terrestrial organism is being tested for systematic conservation planning (Bladt *et al.* 2008; Larsen *et al.* 2009). Some schemes also include a measure of vulnerability, with high vulnerability either contributing positively (reactive schemes) or negatively (proactive schemes) to the conservation importance of an area (Brooks *et al.* 2006). This thesis focuses on the measurement of these aspects, and therefore they are described in more detail below.

Measuring irreplaceability

With few exceptions (Stattersfield *et al.* 1998) global analyses of irreplaceability are restricted to estimates by specialists. However, particularly at the regional scale, such analyses are now increasingly supplemented by data-driven approaches (for example Fishpool & Evans 2001; Eken *et al.* 2004). This has been facilitated by the increased availability of high-resolution distribution data for the better studied taxonomic groups (Orme *et al.* 2005; Lamoreux *et al.* 2006) and sophistication of conservation prioritisation techniques (Margules & Pressey 2000; Ferrier *et al.* 2004). However, knowledge of species distributions is far from being adequate yet (Whittaker *et al.* 2001; Lomolino 2004), and particularly at the regional scale, where most conservation decisions take place (Mace *et al.* 2000; Ferrier 2002; Ferrier *et al.* 2004), collection intensity may be insufficient to allow for effective conservation planning (Da Fonseca *et al.* 2000). Several methods have been proposed to fill distribution knowledge gaps in the short term, including environmentally driven predictive modelling (Da Fonseca *et al.* 2000; Ferrier 2002; Ferrier *et al.* 2004; Küper *et al.* 2006), use of spatio-environmental surrogates such as landscape heterogeneity (Pressey *et al.* 2000; Ferrier 2002; Rouget *et al.* 2003), and combinations of data-driven and expert opinion based approaches (Margules & Pressey 2000). However, models and the use of surrogates will perform poorly if there are strong environmental and/or spatial biases in the data they are based upon (Platts *et al.*2010). Even expert estimates will to some degree be limited and biased by available knowledge and therefore, contrary to Brooks *et al.* (2006) high congruence between published species accounts and expert estimates (Krupnick & Kress 2003) may not sufficiently demonstrate the validity of existing conservation priorities. Many studies have shown that perceived levels of biodiversity at local and regional scales are a direct function of collection intensity (Kress *et al.* 1998; Reddy & Davalos 2003; Kier *et al.* 2005; Tobler *et al.* 2007; Soria-Auza & Kessler 2008), and have cautioned against biological generalisations or conservation recommendations based on such data, particularly in notoriously under-collected systems such as tropical forests (Prance *et al.* 2000). Conservation may not be effective if the biases affecting perceived irreplaceability are not understood and accounted for in conservation planning (Williams *et al.* 2002; Grand *et al.* 2007).

Given the need for unbiased biodiversity data, biodiversity surveys continue to be an important aspect of conservation assessments (Balmford & Gaston 1999). Constant shortages of funds for biodiversity surveys (Da Fonseca *et al.* 2000) and basic taxonomic work (e.g. Disney 1998; Jaspars 1998) make it necessary that biodiversity surveys are as efficient as possible. This is hampered by a persistent disagreement over the most efficient sampling protocols. Several studies have compared the efficiency of numerous methods used for plant species richness assessments in tropical forests (Hall 1991; Stern 1998; Wessels *et al.* 1998; Campbell *et al.* 2002; Phillips *et al.* 2003; Gordon & Newton 2006a; Archaux *et al.* 2007); however, the conclusions of these studies lack consistency. A consequence of this lack of clarity over which protocol is best to use is that a wide range of different methods are employed. As such, many studies might be rendered incompatible with each other, and there is a general uncertainty over data reliability as the employed sampling strategies may not only be inefficient but also insufficiently representative (Gordon & Newton 2006b). There is an urgent need to understand the factors that influence sampling efficiency and to establish protocols for the minimum needed sample size upon which conservation decisions and ecological inferences can reliably be based.

Measuring vulnerability

In Africa, areas that are important in terms of biodiversity conservation have been shown to be congruent with human population density (Balmford *et al.* 2001), and population growth rates in biodiversity hotspots may be higher than the average growth rate in developing countries (Cincotta *et al.* 2000). Areas that are already degraded or pre-disposed to degradation, for example due to their location near infrastructure and/or centres of demand, or unstable governance and high levels of corruption (O'Connor *et al.* 2003; Smith *et al.* 2003), may be more costly to conserve (Balmford *et al.* 2003). Furthermore, vulnerability of areas may, in part, stem from an intense reliance of the adjacent population on ecosystem services. It is therefore essential that vulnerability, socio-economic considerations and ecosystem service provision are incorporated into conservation planning (Naidoo *et al.* 2008). This should take place at all planning stages – from the initial identification and involvement of relevant stakeholders, through the formulation of goals and targets, to the implementation of conservation action (Wilson *et al.* 2005). An assessment of relative irreplaceability *versus* vulnerability can provide guidance in deciding whether a particular vulnerable area should receive conservation priority (as it may contain unique biological features not represented by other areas) or whether the costs of maintaining the area outweigh the potential gains (Margules & Pressey 2000). However, current measures of vulnerability (summarised in Wilson *et al.* 2005) are frequently retrospective and not predictive (Brooks *et al.* 2006) due to the intensity and projected impact of threats being particularly difficult to forecast (Wilson *et al.* 2005). Therefore the reliable prediction of levels of exposure, intensity and impact of degradation in areas of conservation interest would be of significant importance in conservation planning and protected area management. Furthermore, these predictions would aid the implementation of schemes such as 'Reducing Emissions from Deforestation and Degradation' (REDD) (e.g. Miles & Kapos 2008), currently negotiated by parties of the United Nations Framework Convention on Climate Change. Due to difficulties associated with complete-coverage measurement of degradation such schemes tend to centre policy around measures of deforestation (Achard *et al.* 2007; DeFries *et al.* 2007; Ramankutty *et al.* 2007).

Overview of the thesis

The obstacles to effective conservation planning detailed above (first, biases in the biological data that underpin conservation decisions; second, lack of clarity over the most efficient survey design; and third lack of spatially explicit and predictive data on the vulnerability of areas) are characteristic at many high-biodiversity areas in the tropics. This thesis analyses the extent and impact of these challenges and provides recommendations for future biodiversity surveys and conservation planning. It focuses on vascular plants in the Eastern Arc Mountains (EAM) and coastal forests (CF) of Tanzania, which provide a good model system because their highly diverse flora is relatively well documented (Beentje & Smith 2001) and extensive degradation data are available (Appendix). Yet, as elsewhere, there is a lack of clarity over whether these data suffice for systematic conservation planning, and, if not, how to most efficiently collect further data. The overall goal of the thesis is to strengthen the knowledge base for effective conservation planning in the study area, but the concepts presented are generic and it is hoped that they will find application outside the region and across other taxonomic groups.

Specifically, the thesis aims are:

1. To identify biases in the available vascular plant data for the study area and to assess the degree to which these may influence regional conservation planning.
2. To analyse which factors influence the efficiency of tree species richness assessments in tropical forests and to identify the most efficient sampling protocol. Furthermore, to assess whether insufficient sampling may have been a source of error in the deduction of plant species richness patterns in the study area.
3. To establish whether forest degradation is predictable and thereby amenable to strategic conservation planning.
4. To provide guidance for future vegetation and degradation surveys and conservation planning in the area.

Below I give an overview of the study area, outlining our present knowledge, addressing its formation, current climatic conditions, the origin of its high biodiversity, the history of human impact on the area and the conservation problems it faces. This is followed by a justification for focussing on the vascular plants in this area and details of the data utilised in the study, including their collation, problems encountered during the process and the key characteristics of the database structure. Finally, I outline the structure of each chapter.

Study area

The EAM and CF of Tanzania (Figs. 1.1 and 1.2) are amongst the most important sites for conservation globally (Stattersfield *et al.* 1998; Olson & Dinerstein 2002; Mittermeier *et al.* 2005), and by comparison to many other tropical hotspots they have been relatively well studied rendering them an ideally suited model system for this study. Botanical exploration in the EAM goes back over 130 years and in the CF ~110 years. The EAM are a chain of 13 ancient crystalline mountain blocs composed of over 570 million years old heavily metamorphosed Precambrian basement rock and estimated to have been uplifted in the Miocene 30 million years ago (Schlüter 1997). The tropical mountains stretch from south-east Kenya to south-central Tanzania (3° - 9° south; 34° - 39° east) and are under the direct climatic influence of the Indian Ocean. Today, they support ~3,300 – 5,100 km² of tropical forest, which may be less than 30% of the estimated original forested area in prehistoric times (~ 2000 years ago) (Newmark 1998; Myers *et al.* 2000; Newmark 2002; Burgess *et al.* 2007; Platts *et al.* 2010). The lowland dry CF have long been overshadowed by the apparently more diverse moist mountain forests, but conservation interest in this area increased when biodiversity surveys revealed levels of localised endemism comparable to those of the EAM (Brenan 1978; Burgess *et al.* 1998; Clarke 1998; Burgess & Clarke 2000), and found a large number of shared near-endemic species (*n* = 234) between the EAM and CF (Roy Gereau, unpublished data). The lowland CF are thought to have once formed a belt along the East African coast ranging from southern Somalia to northern Mozambique (-1° - 25° south; 34° - 44° east). Approximately 3,170 km² of this habitat are remaining (693 km² in Tanzania) (Burgess & Clarke 2000), confined to around 250 small and highly fragmented patches, most of which are less than 5 km² in size (Rodgers 2000). Their bedrock is highly varied and mostly of sedimentary origin (Burgess & Clarke 2000), accumulated since the Triassic, and more pronounced during the Jurassic era when Africa as a whole broke away from the remainder of Gondwanaland (~120 million BP) (Schlüter 1997).The limits of the CF area are set by rainfall (decreasing to the north), seasonality (increasing to the south), and altitude (increasing to the west) (Burgess *et al.* 1998).

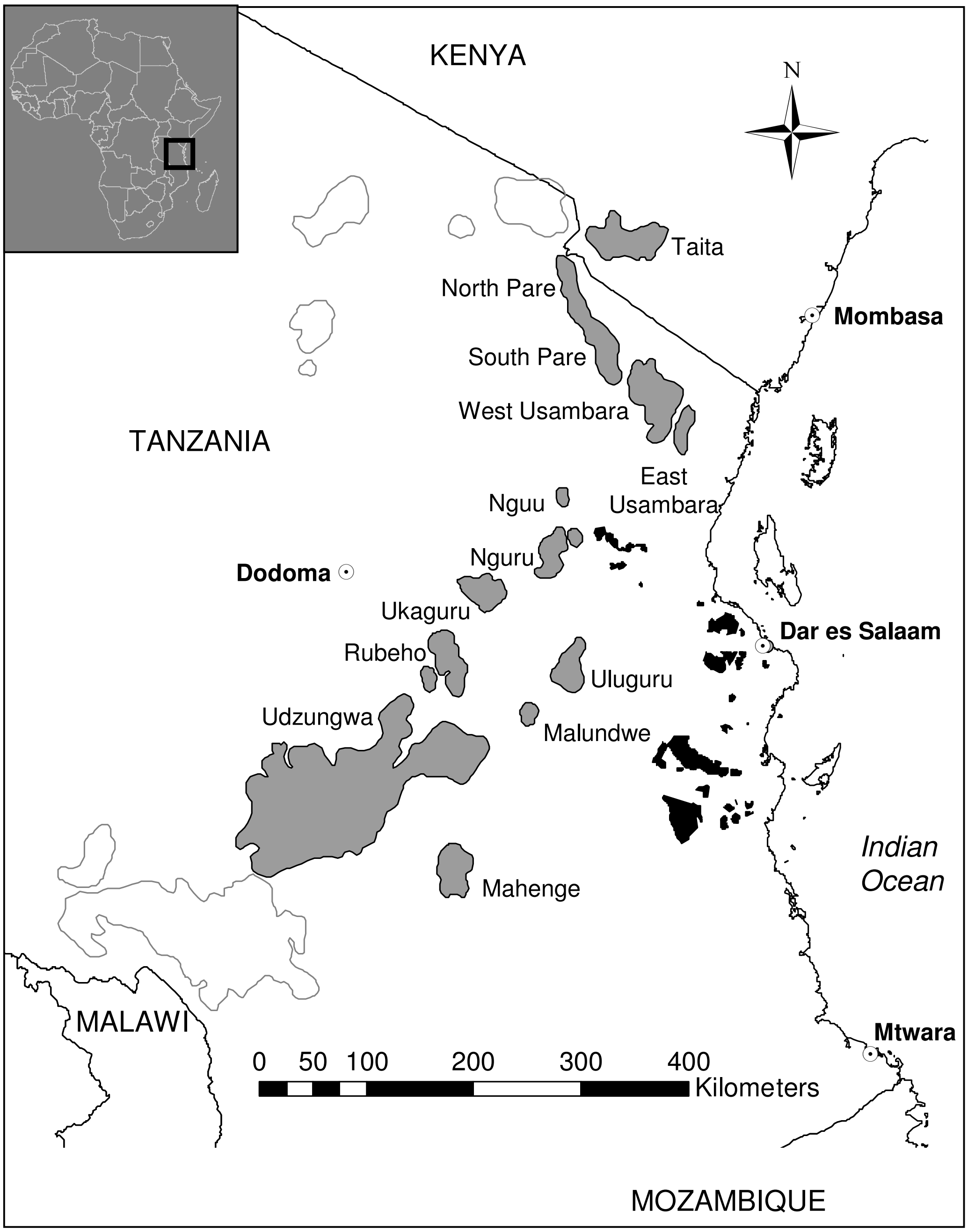


Figure 1.1. Thesis study area. The Eastern Arc Mountains (EAM) are grey and the coastal forests (of the Tanzanian Coast (Pwani) and Dar es Salaam Regions) black polygons. Polygons outlined in grey are other mountains that do not form part of the EAM.

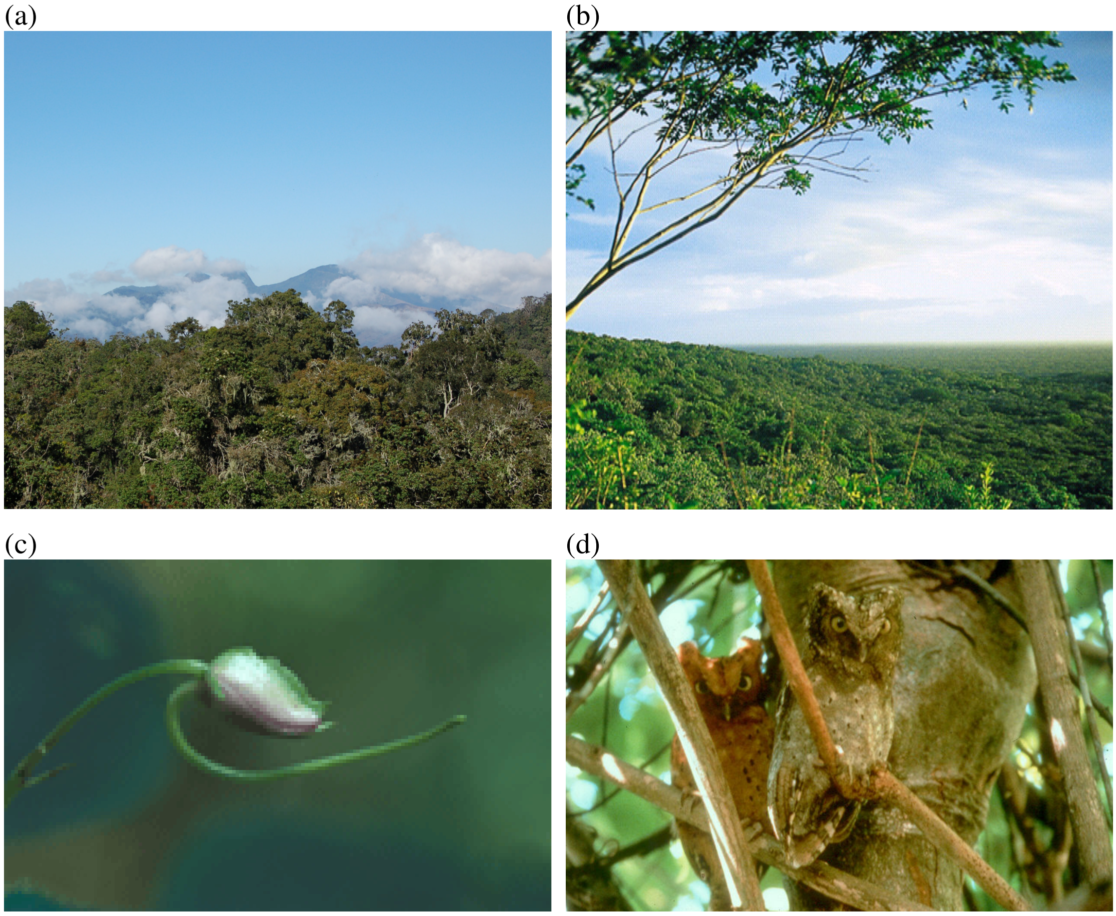


Figure 1.2. Photographs of Eastern Arc Mountain and coastal forests. (a) West Kilombero Scarp Forest Reserve in the Udzungwas. (b) Arabuko-Sokoke Forest Reserve in Kenya – the largest remaining fragment of coastal forest vegetation in East Africa. (c) *Impatiens* (Balsaminaceae) – a genus within which remarkable recent local radiations have occurred in the Eastern Arc Mountains (neo-endemism). Photo: Michele Menegon. (d) Sokoke Scops-Owl (*Otus ireneae* Ripley) – a putative palaeo-endemic confined to Arabuko-Sokoke Forest and the lower slopes of the East Usambara Mountains (Stevenson & Fanshawe 2004). Photo: ICBP.

Archaeological and high-resolution palaeoecological data suggest that anthropogenic influences on the EAM vegetation date back at least 2,000 years (Finch and Marchant 'A 1300-yr record of forest history and human impacts from the Eastern Arc Mountains of Tanzania' in prep); however, an intensified impact is only apparent at the turn of the 20th century when much selective logging occurred under colonial rule. While the first forest reserves were already established under German colonial rule, a total ban on all timber harvesting was implemented as late as 1984 (Finch and Marchant, in prep). This intensive logging over decades may be responsible for locally low densities of high-value timber species such as *Ocotea usambarensis* Engl.. Today the EAM forests are conserved in over 150 forest reserves, 4 nature reserves and 2 national parks (Burgess *et al.*, 2007). Outside these protected areas most forest has been cleared (Burgess *et al.*, 2007), with an estimated total loss in forest area of 25% since 1955 (Hall *et al.*, 2009) and strong degradation pressures continue. For example, stump data indicate that an average of 7% and up to 50% of trees have been cut in gazetted forest reserves over the last ~10 years (Ahrends, unpublished data). Degradation pressures in the CF are even stronger due to their greater accessibility and close vicinity to Dar es Salaam, and they were the target of massive-scale illegal timber logging mainly for the Chinese market between 2000 and 2005 (Milledge *et al.* 2007). These forests also constitute one of the main sources of charcoal for expanding urban centres such as Dar es Salaam with up to 70% of trees cut in the forest reserves close to that city (Chapter 5; Fig. 1.3). Effective conservation is thus vital to conserve these biologically important areas.

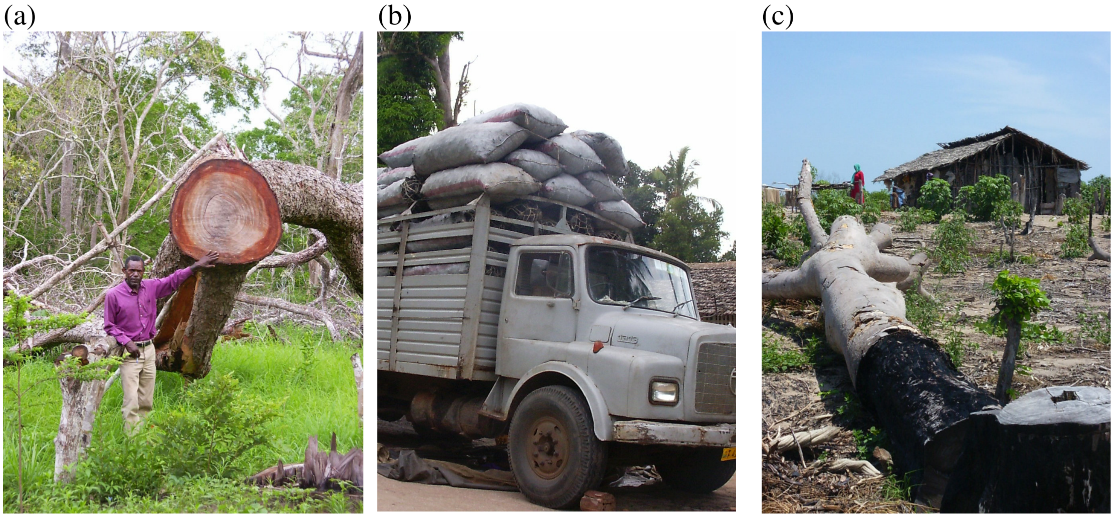

(a) (b) (c)

Figure 1.3. Degradation in the coastal forests. (a) Illegal timber logging in Ruvu South Forest Reserve. (b) Transport of charcoal from Kibiti to Dar es Salaam. (c) In unprotected areas logging and charcoal burning are frequently followed by agricultural encroachment (Kisiju).

The EAM and CF originally formed a joint hotspot (Mittermeier *et al.* 1998) but have since been split (Mittermeier *et al.* 2005). Since 1980, when intensified scientific and conservation interest in the region emerged, over US$ 200 million have been invested in its conservation and biodiversity exploration (Chapter 2). As elsewhere, decisions about priority areas for conservation have typically been based on a combination of species level data and expert advice with numbers of endemic species ranking amongst the most important criteria (e.g. CEPF 2003; FBD 2005). As such, 80% of the total EAM funding invested since 1980 has focused on the Usambara and Udzungwa Mountains, which are perceived to have the highest levels of endemism (Chapter 2). The same data have recently been used to count single site endemics for the identification of core areas as part of the nomination of the EAM as a World Heritage Site (WHC 2010). However, potential collection biases, survey needs and individual site vulnerability to degradation have not been systematically addressed in the development of these species based conservation plans.

Why plants?

The thesis focuses on the vascular plants in these areas for several reasons. Firstly, patterns in plant biodiversity can be indicative of patterns in other taxonomic groups (Kier *et al.* 2005) and vascular plant diversity is an important criterion in many global and regional conservation prioritisation schemes (Davis *et al.* 1997; Mittermeier *et al.* 1997; Olson & Dinerstein 2002; Eken *et al.* 2004; Plantlife International 2004; Mittermeier *et al.* 2005). Secondly, in the EAM plant record density is comparatively high with explorations going back over 130 years and totalling almost 88,000 records (Appendix). Thirdly, vascular plants are generally amongst the better known taxonomic groups with a list of all accepted plant names being in development for the Global Strategy of Plant Conservation (Paton *et al.* 2008), which means that plant diversity estimates for the EAM may be more reliable and can be set into a broader continental/global context. Fourthly, dynamics over time can be inferred from palaeo records, rendering plants particularly suitable for the study of historical and future vegetation dynamics. Thus, focussing on vascular plants for this study offers several key advantages and can reflect generic patterns and issues that affect other groups. But it must be stressed that future work will need to extend such work across the other taxonomic groups if we are to build a sound understanding of the area and develop reliable conservation priorities.

Data collation for the thesis

Numerous individuals and organisations have collected plant data in the EAM and CF of Tanzania, with the earliest listed plant record from 1877. However, with the exception of the Missouri Botanical Garden TROPICOS catalogue (http://www.tropicos.org/) no effort has yet been made to collate all these data in a central database. The several thousand plant records were held in highly disparate sources (partly not digitised, i.e. only in field books, or in chaotically organised spreadsheets), with individual collectors frequently mixing different taxonomic classifications, and substantial parts of the data were thought to be entirely lost (Frontier Tanzania, personal communication). Efforts to analyse plant diversity and to identify endemic species were largely unconnected, and widely differing conclusions were made. For example, published estimates for the level of endemism in the area have ranged from ~800 to over 2,000 species (Lovett 1989; Chapman & Chapman 1996; Myers *et al.* 2000). Differences in these estimates will be due to genuine knowledge gaps but problems have been exacerbated in some important cases by misquotations permeating the literature (Roy Gereau, personal communication). The situation was further complicated by the widespread use of frequently unclarified synonyms. Paradoxically, identifying and correcting synonyms is generally more challenging than identifying taxonomic novelties due to there being a rigid publishing system for the description of novel plant species, but new synonyms are only occasionally and sporadically documented (Lughadha 2004). The only available comprehensive nomenclator (list of accepted species names for an area; Lughadha 2004) was the African Flowering Plants Database (2008). However this does not list all of the used synonyms, nor the more recently discovered species from the area. The database is also somewhat unpopular amongst Tanzanian botanists as its family classification diverges in many ways from the Flora of Tropical East Africa (Turrill *et al.* 1952-), the primary literature source used for species identifications in the region. The consequence of the taxonomic uncertainty was that almost all published identification sources and check lists use a slightly different taxonomic system (compare e.g. Burtt *et al.* 1940; Turrill *et al.* 1952-; Beentje 1994; Kokwaro 1994; Lovett *et al.* 2006), which meant that species compilations for conservation assessments were chaotic and could only be interpreted by a few specialists. This problem is exacerbated by the paucity of genetic level data for the region as divergent sister species from different mountains might be lumped into single poorly defined morphospecies, leading to underestimates of the degree of endemism. Further complications are presented by taxonomically uncertain species such as *Macaranga capensis sensu lato* (Baill.) Sim, which is split by some into two varieties, or even the sister species *M. capensis* (Baill.) Sim and *M. kilimandscharica* Pax (Polhill 1988; Lovett *et al.* 2006); or *Saintpaulia*, where the number of estimated species varies greatly (from ~8 to 17) depending on the species concept applied (compare Möller & Cronk 1997; Lindqvist & Albert 1999, 2001).

The first two years of this Ph.D. were spent collating all available vegetation assessments from the EAM in a relational database and Geographic Information System (GIS), cleaning the data from errors, highlighting all potential misidentifications, and standardising the species names to accepted names in the African Flowering Plants Database (2008). The taxonomy was further updated according to a fully revised list of accepted names (with reference to taxonomic revisions and monographs) produced by Roy Gereau. The data were then analysed in terms of geographic biases and two surveys carried out (July – October 2006: Uluguru and Udzungwa, 462 records; August – October 2007: Nguu and North Pare, 389 records) in order to fill two of the most pressing survey gaps: only four plant records and no vegetation plot data existed for the remote Nguu Mountains and no vegetation plot data for the North Pare Mountains, with no collections being anticipated from these areas in the near future. Furthermore, 26,253 candidate EAM records were obtained from the

Missouri Botanical Garden TROPICOS database (a download of all data collected in EAM districts) and manually assigned to one of the EAM blocs based on the locality description of the collector. Matching the records according to their coordinates in a GIS would have resulted in an incorrect assignment of over 30% of the records (n = 8,235) because many of the collections were conducted in pre-GPS times and their coordinates have subsequently been assigned, often using broad locality categories. The final database comprises 70,081 records (57,900 of which are identified to species) from 2,216 vegetation plots, in addition to 17,731 records from TROPICOS (n total = 87,812). All potentially threatened species have been identified (n = 698) according to the Eastern Africa Plant Red List Authority (2008), and all endemic plants on the basis of an analysis by Roy Gereau. In total, 3,475 vascular plant species (4,236 taxa) have been recorded (3,986 species without a strict altitudinal limit of ≥ 500 m), of which 409 (500 taxa) are thought to be endemic (464 without a strict lower altitudinal limit) (Roy Gereau, unpublished data). These figures differ greatly from the previously most widely cited estimates of ~1,500 endemic plants over a total of ~3,000 species for the EAM and CF combined (Myers *et al.* 2000). Wherever possible, identifier and herbarium were documented but for the majority of the plot-based records verification of primary herbarium material will not be possible as such collections contain many sterile specimens which are discarded following identification - a common problem with data that is systematically or randomly collected. Therefore, the database lists the original specimen numbers assigned in the field as some collectors will keep their specimens privately.

The database also includes information from degradation transects for both the EAM and CF (n transects = 1,057; n records = 273,564). The vast majority of these records have been collected in the EAM (n = 261,546) and do not contain any information on species identity or size. These assessments were based on a rapid approach (Doody *et al.* 2001) whereby along transects of 10 m width all standing, naturally dead and cut poles (≥ 2 m straight length; ≥ 50 to < 150 mm diameter at reference height (drh; 1.3 m up the stem or above buttresses) and trees (≥ 3 m straight length; ≥ 150 mm drh) are counted in sections of 50 m length. Disturbance records for the CF (n = 12,018), collected by the PhD candidate for her M.Sc., detail species, drh and height class (Ahrends 2005). These records are also currently being added to the vegetation data, augmenting to total number of available plant records for the study area to n = 99,830.

The database provides metadata with data provider, contacts, conditions for the use of the data and potential problems associated with each dataset, reflecting the aim of constructing this database, i.e. to standardise the data and to make them easily available for conservation assessments. As such, it has already been provided to the ‘Valuing the Arc’ project

(Mwakalila *et al.* 2009) and the EAM World Heritage Site nomination (WHC 2010). At present, memoranda of understanding are required with each data provider for third party data users. Negotiations with the data contributors are under-way to publish the data as a paper and thereby make them more easily accessible. The Appendix gives details of the database structure and the fields contained therein.

Outline of the analytical chapters

Chapter 2 - Funding begets biodiversity

Effective conservation relies on an unbiased knowledge of the distribution of biodiversity (Da Fonseca *et al.* 2000; Margules & Pressey 2000). Species richness, endemism and threatened species are emphasised by many regional and global conservation prioritisation schemes (Brooks *et al.* 2006). Once an area is regarded as important in terms of biodiversity it typically attracts funding (e.g. Dalton 2000). This chapter analyses whether there is a circularity between inventory funding and perceived biodiversity, whereby the conservation status of areas already perceived as important is strengthened through further survey investment while other areas with less initial funding remain overlooked.

Chapter 3 - Conservation and the botanist effect

Another potential bias in conservation prioritisations is the spatial distribution of taxonomic knowledge (Mace 2004). This chapter tests whether plant recording in the EAM is biased by observer skill and available resources, and to this end analyses reported species richness and numbers of threatened and endemic species based on 24,024 records collected by 13 botanists.

Chapter 4 – Optimising sampling protocols for plant species richness assessments in tropical forests

The previous two chapters highlight the great need for efficient biodiversity surveys to reduce biases and fill the gaps in data used for conservation planning. This is particularly important in the notoriously understudied tropical forests (Prance *et al.* 2000). However, a persisting disagreement over the most efficient vegetation sampling protocols (e.g. Stern 1998; Phillips *et al.* 2003; Gordon & Newton 2006a; Archaux *et al.* 2007) means that a large range of potentially incomparable assessment methods are being employed (e.g. Gentry 1982; Hall 1991; Stohlgren *et al.* 1995; Campbell *et al.* 2002; Sheil *et al.* 2003). While surveys should ideally be rapid and cost-effective, it is also important that they are sufficiently intensive to capture the true diversity patterns (Gordon & Newton 2006b). Chapter 4 takes a simulation approach to analyse the factors that may influence performance of tree sampling in tropical forests and to develop guidelines for choosing an appropriate sampling protocol.

Chapter 5 - Predictable waves of forest degradation spreading from an African city.

The first three chapters focus on assessments of irreplaceability; however, another important consideration in conservation planning is the vulnerability of areas as this is likely to be associated with conservation urgency, costs and effectiveness (Balmford *et al.* 2003; Wilson *et al.* 2005). While impressive progress has been made with the quantification of deforestation from remotely sensed data (Achard *et al.* 2007; DeFries *et al.* 2007; Ramankutty *et al.* 2007; DeFries *et al.* 2010), forest degradation remains much more difficult to measure (Tang *et al.* in press). Given that degradation can have severe impacts on biodiversity (Bawa & Seidler 1998) and carbon storage (Nepstad *et al.* 1999; Asner *et al.* 2005), understanding the expansion rates and patterns of tropical forest degradation is key to managing these habitats sustainably for biodiversity conservation and climate change mitigation. Predicting tropical forest degradation is also one of the major challenges to REDD. Putative patterns have been proposed (e.g. Gentry & Vasquez 1988; Vasquez & Gentry 1989), however, not tested systematically. This chapter tests economic resource use theory (von Thünen 1966; Angelsen 2007) applied to tropical forest degradation by analysing ground-based data on forest degradation at different distances from a demand centre (Dar es Salaam, Tanzania) over a decade.

References

Achard F., DeFries R., Eva H., Hansen M., Mayaux P. & Stibig H.J. (2007). Pan-tropical monitoring of deforestation. *Environmental Research Letters*, 2.

African Flowering Plants Database (2008). Conservatoire et Jardin botaniques de la Villle de Genève and South African National Biodiversity Institute Pretoria. http://www.ville-ge.ch/musinfo/bd/cjb/africa/.

Ahrends A. (2005). Patterns of degradation in lowland coastal forests in Coast Region, Tanzania. University of Greifswald, Greifswald.

Angelsen A. (2007). Forest Cover Change in Space and Time: Combining the von Thünen and Forest Transition Theories. *World Bank Policy Research Working Paper*, 4117.

Archaux F., Berges L. & Chevalier R. (2007). Are plant censuses carried out on small quadrats more reliable than on larger ones? *Plant Ecology*, 188, 179-190.

Asner G.P., Knapp D.E., Broadbent E.N., Oliveira P.J.C., Keller M. & Silva J.N. (2005). Selective logging in the Brazilian Amazon. *Science*, 310, 480-482.

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

Balmford A., Gaston K.J., Blyth S., James A. & Kapos V. (2003). Global variation in terrestrial conservation costs, conservation benefits, and unmet conservation needs. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 1046-1050.

Balmford A., Moore J.L., Brooks T., Burgess N., Hansen L.A., Williams P. & Rahbek C. (2001). Conservation conflicts across Africa. *Science*, 291, 2616-2619.

Barthlott W., Mutke J., Rafiqpoor D., Kier G. & Kreft H. (2005). Global Centers of vascular plant diversity. *From Plant Taxonomy to Evolutionary Biology*, 92, 61-83.

Bawa K.S. & Seidler R. (1998). Natural forest management and conservation of biodiversity in tropical forests. *Conservation Biology*, 12, 46-55.

Beentje H. (1994). *Kenya trees, shrubs and lianas*. National Museums of Kenya, Nairobi.

Beentje H. & Smith S. (2001). FTEA and after. *Systematics and Geography of Plants*, 71, 265-290.

Bisby F.A. (2000). The quiet revolution: Biodiversity informatics and the internet. *Science*, 289, 2309-2312.

Bladt J., Larsen F.W. & Rahbek C. (2008). Does taxonomic diversity in indicator groups influence their effectiveness in identifying priority areas for species conservation? *Animal Conservation*, 11, 546-554.

Brenan J. (1978). Some aspects of the phytogeography of tropical Africa. *Annals of the Missouri Botanical Garden*, 65, 437-478.

Brook B.W., Sodhi N.S. & Bradshaw C.J.A. (2008). Synergies among extinction drivers under global change. *Trends in Ecology & Evolution*, 23, 453-460.

Brooks T.M., Mittermeier R.A., da Fonseca G.A.B., Gerlach J., Hoffmann M., Lamoreux J.F., Mittermeier C.G., Pilgrim J.D. & Rodrigues A.S.L. (2006). Global biodiversity conservation priorities. *Science*, 313, 58-61.

Bryant D., Nielsen D. & Tangley L. (1998). The last frontier forests. *Issues in Science and Technology*, 14, 85-87.

Burgess N.D., Butynski T.M., Cordeiro N.J., Doggart N.H., Fjeldså J., Howell K.M., Kilahama F.B., Loader S.P., Lovett J.C., Mbilinyi B., Menegon M., Moyer D.C., Nashanda E., Perkin A., Rovero F., Stanley W.T. & Stuart S.N. (2007). The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, 134, 209-231.

Burgess N.D. & Clarke G.P. (eds.) (2000). *Coastal Forests of Eastern Africa*. IUCN Publication Services Unit, Cambridge.

Burgess N.D., Clarke G.P. & Rodgers W.A. (1998). Coastal forests of eastern Africa: status, endemism patterns and their potential causes. *Biological Journal of the Linnean Society*, 64, 337-367.

Burtt B.D., Hora F.B. & Greenway P.J. (1940). *Check-Lists of the Forest Trees and Shrubs of the British Empire - Tanganyika Territory Part I*. Imperial Forestry Institute, Oxford.

Campbell P., Comiskey J., Alonso A., Dallmeier F., Nunez P., Beltran H., Baldeon S., Nauray W., De La Colina R., Acurio L. & Udvardy S. (2002). Modified whittaker plots as an assessment and monitoring tool for vegetation in a lowland tropical rainforest. *Environmental Monitoring and Assessment*, 76, 19-41.

CEPF (Critical Ecosystem Partnership Fund) (2003). Eastern Arc Mountains and coastal forest hotspot Ecosystem Profile. CEPF Dar es Salaam, Nairobi. http://www.cepf.net/Documents/final.easternarc.ep.pdf

Chapman C.A. & Chapman L.J. (1996). Mid-elevation Forests: A history of disturbance and regeneration. In: *East African Ecosystems and their Conservation* (ed. Young TP & McClanahan TR). Oxford University Press, New York.

Cincotta R.P., Wisnewski J. & Engelman R. (2000). Human population in the biodiversity hotspots. *Nature*, 404, 990-992.

Clarke G.P. (1998). A new regional centre of endemism in Africa. In: *Chronology, taxonomy and ecology of floras of Africa and Madagascar* (eds. Huxely CR, Lock JM & Cutler DE). Royal Botanic Gardens, Kew, London.

Da Fonseca G.A.B., Balmford A., Bibby C., Boitani L., Corsi F., Brooks T., Gascon C., Olivieri S., Mittermeier R.A., Burgess N., Dinerstein E., Olson D., Hannah L., Lovett

J., Moyer D., Rahbek C., Stuart S. & Williams P. (2000). It's time to work together and stop duplicating conservation efforts ... following Africa's lead in setting priorities. *Nature*, 405, 393-394.

Dalton R. (2000). Biodiversity cash aimed at hotspots. *Nature*, 406, 818-818.

Davis S.D., Heywood V.H., Herrera-MacBryde O., Villa-Lobos J. & Hamilton A.C. (1997). Centres of plant diversity. A guide and strategy for their conservation. (ed. WWF/IUCN) Gland.

DeFries R., Achard F., Brown S., Herold M., Murdiyarso D., Schlamadinger B. & de Souza C. (2007). Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environmental Science & Policy*, 10, 385-394.

DeFries R.S., Rudel T., Uriarte M. & Hansen M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*. doi:10.1038/ngeo756

Disney H. (1998). Rescue plan needed for taxonomy. *Nature*, 394, 120-120.

Doody K.Z., Howell K.M. & Fanning E. (2001). Udzungwa Mountains Biodiversity Surveys – Methods Manual. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Eastern Africa Plant Red List Authority (2008). List of Potentially Threatened Plant Taxa in the Eastern Arc Mountains of Tanzania. Unpublished report presented to the Forestry and Beekeeping Division, Tanzania Ministry of Natural Resources and Tourism. (ed. Africa Plant Red List Authority) Dar es Salaam.

Eken G., Bennun L., Brooks T.M., Darwall W., Fishpool L.D.C., Foster M., Knox D., Langhammer P., Matiku P., Radford E., Salaman P., Sechrest W., Smith M.L., Spector S. & Tordoff A. (2004). Key biodiversity areas as site conservation targets. *Bioscience*, 54, 1110-1118.

FBD (Forest and Beek Keepign Division of the Ministry of Natural Resources and Tourism Tanzania) (2005). Biodiversity Baseline for the Eastern Arc Mountains of Tanzania and Kenya. Conservation and Management of the Eastern Arc Mountain Forests Project, Dar es Salaam.

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

Ferrier S., Powell G.V.N., Richardson K.S., Manion G., Overton J.M., Allnutt T.F., Cameron S.E., Mantle K., Burgess N.D., Faith D.P., Lamoreux J.F., Kier G., Hijmans R.J., Funk V.A., Cassis G.A., Fisher B.L., Flemons P., Lees D., Lovett J.C. & Van Rompaey R. (2004). Mapping more of terrestrial biodiversity for global conservation assessment. *Bioscience*, 54, 1101-1109.

Fishpool L.D.C. & Evans M.I. (2001). *Important Bird Areas in Africa and associated islands*. BirdLife International, Cambridge.

Gentry A.H. (1982). Patterns of Neotropical Plant-Species Diversity. *Evolutionary Biology*, 15, 1-85.

Gentry A.H. & Vasquez R. (1988). Where have all the ceibas gone - a case-history of mismanagement of a tropical forest resource. *Forest Ecology and Management*, 23, 73-76.

Gordon J.E. & Newton A.C. (2006a). Efficient floristic inventory for the assessment of tropical tree diversity: A comparative test of four alternative approaches. *Forest Ecology and Management*, 237, 564-573.

Gordon J.E. & Newton A.C. (2006b). The potential misapplication of rapid plant diversity assessment in tropical conservation. *Journal for Nature Conservation*, 14, 117-126.

Graham C.H., Ferrier S., Huettman F., Moritz C. & Peterson A.T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, 19, 497-503.

Grand J., Cummings M.P., Rebelo T.G., Ricketts T.H. & Neel M.C. (2007). Biased data reduce efficiency and effectiveness of conservation reserve networks. *Ecology Letters*, 10, 364-374.

Hall J.B. (1991). Multiple-nearest-tree sampling in an ecological survey of Afromontane catchment forest. *Forest Ecology and Management*, 42, 245-299.

Hoekstra J.M., Boucher T.M., Ricketts T.H. & Roberts C. (2005). Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters*, 8, 23-29.

James A.N., Gaston K.J. & Balmford A. (1999). Balancing the Earth's accounts. *Nature*, 401, 323-324.

Jaspars M. (1998). Tough time for taxonomy. *Nature*, 394, 413-413.

Kier G., Mutke J., Dinerstein E., Ricketts T.H., Kuper W., Kreft H. & Barthlott W. (2005). Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, 32, 1107-1116.

Kokwaro J.O. (1994). *Flowering Plant Families of East Africa*. East African Educational Publishers, Nairobi.

Kress W.J., Heyer W.R., Acevedo P., Coddington J., Cole D., Erwin T.L., Meggers B.J., Pogue M., Thorington R.W., Vari R.P., Weitzman M.J. & Weitzman S.H. (1998). Amazonian biodiversity: assessing conservation priorities with taxonomic data. *Biodiversity and Conservation*, 7, 1577-1587.

Krupnick G.A. & Kress W.J. (2003). Hotspots and ecoregions: a test of conservation priorities using taxonomic data. *Biodiversity and Conservation*, 12, 2237-2253.

Küper W., Sommer J.H., Lovett J.C. & Barthlott W. (2006). Deficiency in African plant distribution data - missing pieces of the puzzle. *Botanical Journal of the Linnean Society*, 150, 355-368.

Lamoreux J.F., Morrison J.C., Ricketts T.H., Olson D.M., Dinerstein E., McKnight M.W. & Shugart H.H. (2006). Global tests of biodiversity concordance and the importance of endemism. *Nature*, 440, 212-214.

Larsen F.W., Bladt J. & Rahbek C. (2009). Indicator taxa revisited: useful for conservation planning? *Diversity and Distributions*, 15, 70-79.

Lindqvist C. & Albert V.A. (1999). Phylogeny and conservation of African violets (Saintpaulia: Gesneriaceae): New findings based on nuclear ribosomal 5S non-transcribed spacer sequences. *Kew Bulletin*, 54, 363-377.

Lindqvist C. & Albert V.A. (2001). A high elevation ancestry for the Usambara Mountains and lowland populations of African violets (Saintpaulia, Gesneriaceae). *Systematics and Geography of Plants*, 71, 37-44.

Lomolino M.V. (2004). Conservation biogeography. In: *Frontiers of Biogeography: new directions in the geography of nature* (eds. Lomolino MV & Heaney LR). Sinauer Associates Sunderland, Massachusetts.

Lovett J.C. (1989). Tanzania. In: *Floristic Inventory of Tropical Countries: The Status of Plant Systematics, Collections, and Vegetation, Plus Recommendations for the Future* (eds. Campbell DG & Hammond HD). New York Botanical Garden, New York.

Lovett J.C., Ruffo C.K., Gereau R.E. & Taplin J.R.D. (2006). *Field Guide to the Moist Forest Trees of Tanzania.* The Society for Environmental Exploration, London and Dar es Salaam.

Lughadha E.N. (2004). Towards a working list of all known plant species. *Philosophical Transactions of the Royal Society B-Biologial Sciences*, 359, 681-7.

Lughadha E.N. & Miller C. (2009). Accelerating global access to plant diversity information. *Trends in Plant Science*, 14, 622-628.

Mace G.M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society B-Biologial Sciences*, 359, 711-719.

Mace G.M., Balmford A., Boitani L., Cowlishaw G., Dobson A.P., Faith D.P., Gaston K.J., Humphries C.J., Vane-Wright R.I., Williams P.H., Lawton J.H., Margules C.R., May R.M., Nicholls A.O., Possingham H.P., Rahbek C. & van Jaarsveld A.S. (2000). It's time to work together and stop duplicating conservation efforts. *Nature*, 405, 393-393.

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

Miles L. & Kapos V. (2008). Reducing greenhouse gas emissions from deforestation and forest degradation: Global land-use implications. *Science*, 320, 1454-1455.

Milledge S.A.H., Gelvas I.K. & Ahrends A. (2007). Forestry, Governance and National Development: Lessons Learned from a Logging Boom in Southern Tanzania. TRAFFIC East/Southern Africa, Tanzania Development Partners Group, Tanzania Ministry of Natural Resources and Tourism, Dar es Salaam.

Millenium Ecocosystem Assessment (2005). Current state and trends. In: *Millenium Ecosystem Assessment* (eds. Hassan R, Scholes R & Ash N) Washington DC.

Mittermeier R.A., Mittermeier C.G., Gil P.R., Pilgrim J., Fonseca G., Konstant W.R. & Brooks T. (2002). *Wilderness: Earth's Last Wild Places*. Cemex, Mexico City.

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

Mittermeier R.A., Robles-Gil P., Hoffmann M., Pilgrim J.D., Brooks T.M., Mittermeier G.C., Lamoreux J.L. & Fonseca, G. (2005). *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Ecoregions*. 2 edn, Cemex, Mexico City.

Mittermeier R.A., Robles Gil P. & Mittermeier C.G. (1997). *Megadiversity*. Cemex, Mexico City.

Möller M. & Cronk Q.C.B. (1997). Phylogeny and disjunct distribution: Evolution of Saintpaulia (Gesneriaceae). *Proceedings of the Royal Society of London Series B-Biological Sciences*, 264, 1827-1836.

Mooney H.A. (2010). The ecosystem-service chain and the biological diversity crisis. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 365, 31-39.

Mwakalila S., Burgess N., Ricketts T., Olwero N., Swetnam R., Mbilinyi B., Marchant R., Mtalo F., White S., Munishi P., Malimbwi R., Smith C., Jambiya G., Marshall A., Madoffe S., Fisher B., Kajembe G., Morse-Jones S., Kulindwa K., Green R., Turner R.K., Green J. & Balmford A. (2009). Valuing the Arc: Linking Science with Stakeholders to Sustain Natural Capital. *The Arc Journal*, 23, 25-30.

Myers N. & Mittermeier R.A. (2003). Impact and acceptance of the hotspots strategy: Response to Ovadia and to Brummitt and Lughadha. *Conservation Biology*, 17, 1449-1450.

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

Naidoo R., Balmford A., Costanza R., Fisher B., Green R.E., Lehner B., Malcolm T.R. & Ricketts T.H. (2008). Global mapping of ecosystem services and conservation priorities. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 9495-9500.

Nepstad D.C., Verissimo A., Alencar A., Nobre C., Lima E., Lefebvre P., Schlesinger P., Potter C., Moutinho P., Mendoza E., Cochrane M. & Brooks V. (1999). Large-scale impoverishment of Amazonian forests by logging and fire. *Nature*, 398, 505-508.

Newmark W. (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, 29-36.

Newmark W.D. (2002). Ecological Studies. Conserving biodiversity in East African forests: A study of the Eastern Arc Mountains. In: *Ecological Studies. Conserving biodiversity in East African forests: A study of the Eastern Arc Mountains*. Springer-Verlag, New York.

Niekisch M. (2000). Internationaler Naturschutz. In: *Arten-, Biotop- und Landschaftsschutz* (eds. Dahl HJ, Niekisch M, Riedl U & Scherfose V). Economica Verlag, Bonn.

O'Connor C., Marvier M. & Kareiva P. (2003). Biological vs. social, economic and political priority-setting in conservation. *Ecology Letters*, 6, 706-711.

Olson D.M. & Dinerstein E. (2002). The Global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden*, 89, 199-224.

Orme C.D.L., Davies R.G., Burgess M., Eigenbrod F., Pickup N., Olson V.A., Webster A.J., Ding T.S., Rasmussen P.C., Ridgely R.S., Stattersfield A.J., Bennett P.M., Blackburn T.M., Gaston K.J. & Owens I.P.F. (2005). Global hotspots of species richness are not congruent with endemism or threat. *Nature*, 436, 1016-1019.

Paton A.J., Brummitt N., Govaerts R., Harman K., Hinchcliffe S., Allkin B. & Lughadha E.N. (2008). Towards Target 1 of the Global Strategy for Plant Conservation: a working list of all known plant species - progress and prospects. *Taxon*, 57, 602-611.

Phillips O.L., Vasquez Martinez R., Nunez Vargas P., Lorenzo Monteagudo A., Chuspe Zans M.E., Galiano Sanchez W., Pena Cruz A., Timana M., Yli-Halla M. & Rose S. (2003). Efficient plot-based floristic assessment of tropical forests. *Journal of Tropical Ecology*, 19, 629-645.

Pimm S.L. & Lawton J.H. (1998). Ecology - Planning for biodiversity. *Science*, 279, 2068-2069.

Plantlife International (2004). Identifying and protecting the world's most important plant areas. Plantlife International, Salisbury.

Platts, P.J., Ahrends, A., Gereau, R.E., McClean, C., Lovett, J.C., Marshall, A., Pellikka, P., Mulligan, M., Fanning, E., Marchant, R. (2010). Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. *Diversity and Distributions* 16, 628-642.

Polhill R.M. (ed.) (1988). *Flora of Tropical East Africa: Euphorbiaceae 2*. A.A. Balkema, Rotterdam.

Prance G.T., Beentje H., Dransfield J. & Johns R. (2000). The tropical flora remains undercollected. *Annals of the Missouri Botanical Garden*, 87, 67-71.

Pressey R.L., Hager T.C., Ryan K.M., Schwarz J., Wall S., Ferrier S. & Creaser P.M. (2000). Using abiotic data for conservation assessments over extensive regions: quantitative methods applied across New South Wales, Australia. *Biological Conservation*, 96, 55-82.

Ramankutty N., Gibbs H.K., Achard F., Defriess R., Foley J.A. & Houghton R.A. (2007). Challenges to estimating carbon emissions from tropical deforestation. *Global Change Biology*, 13, 51-66.

Reddy S. & Davalos L.M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, 30, 1719-1727.

Ricketts T.H., Dinerstein E., Boucher T., Brooks T.M., Butchart S.H.M., Hoffmann M., Lamoreux J.F., Morrison J., Parr M., Pilgrim J.D., Rodrigues A.S.L., Sechrest W., Wallace G.E., Berlin K., Bielby J., Burgess N.D., Church D.R., Cox N., Knox D., Loucks C., Luck G.W., Master L.L., Moore R., Naidoo R., Ridgely R., Schatz G.E., Shire G., Strand H., Wettengel W. & Wikramanayake E. (2005). Pinpointing and preventing imminent extinctions. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 18497-18501.

Rodgers A. (2000). Why a book on coastal forests. In: *The Coastal Forests of Eastern Africa* (eds. Burgess ND & Clarke GP). IUCN Publication Services Unit, Cambridge.

Rouget M., Cowling R.M., Pressey R.L. & Richardson D.M. (2003). Identifying spatial components of ecological and evolutionary processes for regional conservation planning in the Cape Floristic Region, South Africa. *Diversity and Distributions*, 9, 191-210.

Sanderson E.W., Jaiteh M., Levy M.A., Redford K.H., Wannebo A.V. & Woolmer G. (2002). The human footprint and the last of the wild. *Bioscience*, 52, 891-904.

Schlüter T. (1997). *Geology of East Africa*. Borntraeger, Berlin.

Sheil D., Ducey M.J., Sidiyasa K. & Samsoedin I. (2003). A new type of sample unit for the efficient assessment of diverse tree communities in complex forest landscapes. *Journal of Tropical Forest Science*, 15, 117-135.

Smith R.J., Muir R.D.J., Walpole M.J., Balmford A. & Leader-Williams N. (2003). Governance and the loss of biodiversity. *Nature*, 426, 67-70.

Soria-Auza R.W. & Kessler M. (2008). The influence of sampling intensity on the perception of the spatial distribution of tropical diversity and endemism: a case study of ferns from Bolivia. *Diversity and Distributions*, 14, 123-130.

Stattersfield A.J., Crosby M.J., Long A.J. & Wege D.C. (1998). *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation*. BirdLife International, Cambridge.

Stern M.J. (1998). Field comparisons of two rapid vegetation assessment techniques with permanent plot inventory data in Amazonian Peru. In: *Forest Biodiversity Research, Monitoring and Modeling: Conceptual Background and Old World Case Studies.* (eds. Dallmeier F & Comiskey JA). UNESCO/Parthenon Publishing Group, New York.

Stevenson T. & Fanshawe J. (2004). *Birds of East Africa.* A & C Black Publishers Ltd, London.

Stohlgren T.J., Falkner M.B. & Schell L.D. (1995). A Modified-Whittaker Nested Vegetation Sampling Method. *Vegetatio*, 117, 113-121.

Stork N.E. (2010). Re-assessing current extinction rates. *Biodiversity and Conservation*, 19, 357-371.

Stork N.E., Coddington J.A., Colwell R.K., Chazdon R.L., Dick C.W., Peres C.A., Sloan S. & Willis K. (2009). Vulnerability and Resilience of Tropical Forest Species to Land-Use Change. *Conservation Biology*, 23, 1438-1447.

Tang L., Shao G., Piao Z., Dai L., Jenkins M.A., Wang S., Wu G., Wu J. & Zhao J. (in press). Forest degradation deepens around and within protected areas in East Asia. *Biological Conservation.*

Tobler M., Honorio E., Janovec J. & Reynel C. (2007). Implications of collection patterns of botanical specimens on their usefulness for conservation planning: an example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. *Biodiversity and Conservation*, 16, 659-677.

Turrill W.B., Milne-Redhead C.E., Hubbard C.E., M P.R., J B.H. & al. e. (1952-). *The Flora of Tropical East Africa.* A.A. Balkema, Rotterdam.

Vasquez R. & Gentry A.H. (1989). Use and misuse of forest-harvested fruits in the Iquitos area. *Conservation Biology*, 3, 350-361.

von Thünen J.H. (1966). Der Isolierte Staat in Beziehung der Landwirtschaft und Nationalökonomie. In: *Von Thünen's Isolated State* (ed. Hall P). Pergamon Press, Oxford.

Wessels K.J., Van Jaarsveld A.S., Grimbeek J.D. & Van der Linde M.J. (1998). An evaluation of the gradsect biological survey method. *Biodiversity and Conservation*, 7, 1093-1121.

WHC (2010). Eastern Arc Mountains World Heritage Nomination. Paris. http://rapidshare.com/files/342874331/E_Arc_Mountains_World_Heritage_Nomination_100127_FINAL.pdf

Whittaker R.J., Willis K.J. & Field R. (2001). Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, 28, 453-470.

Williams P.H., Margules C.R. & Hilbert D.W. (2002). Data requirements and data sources for biodiversity priority area selection. *Journal of Biosciences*, 27, 327-338.

Wilson E.O. (1985). Time to Revive Systematics. *Science*, 230, 1227-1227.

Wilson K., Pressey R.L., Newton A., Burgman M., Possingham H. & Weston C. (2005). Measuring and incorporating vulnerability into conservation planning. *Environmental Management*, 35, 527-543.

Chapter 2 – Funding begets biodiversity

Submitted to *Diversity and Distributions* (Biodiversity Research and Reviews)

Funding begets biodiversity

Antje Ahrends[1,2], Neil D. Burgess[3,4,5], Roy E. Gereau[6], Rob Marchant[1], Mark T. Bulling[7], Jon C. Lovett[8], Philip J. Platts[1], Victoria Wilkins Kindemba[9], Nisha Owen[9], Eibleis Fanning[9] and Carsten Rahbek[5]

[1]Environment Department, University of York, YO105DD, UK; [2] Botanic Garden Edinburgh, 20A Inverleith Row, EH3 5LR, UK; [3] Zoology Department, University of Cambridge, CB2 3EJ, UK; [4] WWF US, 1250 24th St NW, Washington DC, USA; [5] Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Universitetsparken 15, DK- 2100 Copenhagen Ø, Denmark; [6] Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA; [7] Oceanlab, University of Aberdeen, Newburgh, Aberdeenshire, AB416AA, UK; [8] Twente Centre for Studies in Technology and Sustainable Development, University of Twente, P.O. Box 217, 7500 Enschede, The Netherlands; [9] The Society for Environmental Exploration (Frontier), 50-52 Rivington Street, London, EC2A 3QP, UK

Abstract

Conservation priority assessments are typically based on levels of species richness, endemism and threat. Areas identified as important receive the majority of conservation investments, often facilitating further research that results in more species discoveries. Thus, the conservation status of such areas may be circularly strengthened while other areas with less initial funding may remain overlooked. We analysed time series data (1980 - 2007) of funding and plant species records (n = 75,631) in the Eastern Arc Mountains, Tanzania. Funding explained 65% of variation in perceived biodiversity, and was driving conservation priorities and not *vice versa*. These results suggest that biodiversity assessments may be biased by investment, which might also help to explain why little consensus has been reached on biodiversity patterns. Continued biodiversity loss commands urgent conservation action even if our knowledge is incomplete; however, by concentrating inventory funds in areas already identified as important, we risk losing others of underestimated or unknown value.

Introduction

Biodiversity values are widely ranked according to species richness and the prevalence of threatened and endemic taxa (Brooks *et al.* 2006). In conjunction with parameters such as levels of threat, these diversity metrics underpin several schemes that identify global (Stattersfield *et al.* 1998; Olson & Dinerstein 2002; Mittermeier *et al.* 2005) and regional (Fishpool & Evans 2001; Eken *et al.* 2004; Plantlife International 2004) conservation priorities. Once an area is considered to be of high conservation priority, it typically attracts funding for further research and conservation. Priority locations such as Biodiversity Hotspots, Global 200 and Key Biodiversity Areas receive the majority of investments made available by global conservation funds and organisations (e.g. Dalton 2000). Since intensification of research in an area is likely to result in the discovery of more species (Nelson *et al.* 1990; Reddy & Davalos 2003; Kier *et al.* 2005; Soria-Auza & Kessler 2008), including threatened and endemic species, the priority status of that area may be strengthened in a circular fashion. Meanwhile, areas that have received little or no initial funding may remain perpetually overlooked. Thus, funding may beget perceived biodiversity importance and bias our understanding of conservation priorities.

Plant species richness is broadly related to environmental conditions, including levels of both anthropogenic and natural disturbance (O'Brien *et al.* 2000; Hawkins *et al.* 2003). Predictors for levels of endemism and the number of threatened species are less clear; however, both of these variables tend to increase with species richness (Jetz *et al.* 2004). High levels of endemism have also been attributed to past climate configurations (Taplin & Lovett 2003; Jetz *et al.* 2004). We might therefore expect that environmental factors, including levels of disturbance, would be better predictors for species diversity patterns than funding or survey effort. We tested this hypothesis in the Eastern Arc Mountains (EAM) of Tanzania (Fig. 2.3a), an area of outstanding biodiversity value. This range of ancient tropical mountains was a suitable test candidate as it is one of the most important sites for conservation globally (Stattersfield *et al.* 1998; Olson & Dinerstein 2002; Mittermeier *et al.* 2005) and is one of the better studied global conservation priority areas, with botanical exploration going back over 130 years.

Materials and methods

Study Area

The EAM are a chain of 13 ancient crystalline mountain blocs composed of heavily metamorphosed Precambrian basement rock and estimated to have been uplifted in the Miocene 30 million years ago (Schlüter 1997). The mountains stretch from south-east Kenya to south-central Tanzania and are under the direct climatic influence of the Indian Ocean (Fig. 2.3a). Today, they support 3,300 – 5,100 km² of tropical forest, which may be less than

30% of the estimated original forested area in prehistoric times (~ 2000 years ago) (Newmark 1998; Myers *et al.* 2000; Newmark 2002; Burgess *et al.* 2007; Platts *et al.* 2010).

Data

The species data were derived from an extensive dataset totalling 75,631 records from the Missouri Botanical Garden's TROPICOS database (http://www.tropicos.org/) and from 2,216 vegetation plot assessments. Plant species records, representing 3,986 vascular plant species, were taxonomically standardised by reference to the African Flowering Plants Database (African Flowering Plants Database 2008) and further updated by reference to taxonomic revisions and monographs (i.e. there were no changes in species numbers due to changes in the species concept). We also identified all potentially threatened plants according to Eastern Africa Plant Red List Authority (2008) and all endemic plants on the basis of an analysis by R.E.G. of the plant records from the TROPICOS database.

Funding data were derived from a comprehensive collation of all inventory, research and conservation projects that have taken place in the EAM since 1980 (n = 134), which is when explicit research and conservation interest targeted at the area emerged. All funding data were standardised to US$ in the year 2007 (US$2007) with a GDP deflator (www.measuringworth.com). The deflation calculation was made separately for the expenses in every project year.

Environmental data included climate, topography and disturbance. Climatic predictors were derived according to Platts *et al.* (2008) with climate surfaces obtained from the Centre for Resource and Environmental Studies (http://fennerschool.anu.edu.au/). These were then summarised for each mountain bloc. Topography, forest cover data and estimates of the population density were based on Burgess *et al.* (2007). Disturbance per mountain bloc was calculated as the percentage of trees and poles cut in 949 transects, which totalled 536 km in length and were distributed evenly across mountain blocs. Humans are the single major source of disturbance in the EAM vegetation, as natural disturbances by cyclones, earthquakes, or volcanic eruptions are extremely rare. For a list of all environmental predictor variables see Supporting information for Chapter 2 Table S2.1.

Analysis

Statistical models were developed as follows: Because many of the 24 climatic and the five topographic and forest cover predictors were correlated, a situation that may lead to inflated standard errors (Zuur *et al.* 2009a) and alter the significance levels of predictors and interaction terms (Sithisarankul *et al.* 1997), Principal Component Analysis (PCA) was used to replace the two variable sets with their uncorrelated components (for the PCA rotated component matrixes see Tables S2.2 and S2.3). Hierarchical partitioning (Chevan &

Sutherland 1991) allowed us to estimate the independent and conjoint contributions of all predictors. As a starting point, we fitted a linear regression. Where validation procedures, following Zuur *et al.* (2007), indicated problems associated with heterogeneity of variance we used linear regression with generalised least squares (GLS) (Pinheiro *et al.* 2008; Zuur *et al.* 2009b) estimation procedure. GLS was preferred over a Poisson general linear model, as the latter assumes a particular residual distribution that in our case was not matched, and the ranges of our dependent variables were large (i.e. close to continuous). To define the best random structure for the GLS, we first compared models including all starting predictors with different variance function structures (without variance covariate, power of variance covariate, exponential of variance covariate, and constant plus power of a variance covariate) estimated with restricted maximum likelihood, and chose the model with the lowest AIC (Akaike Information Criterion; Sakamoto & Ishiguro 1986) and the most even spread of residuals. To find the minimum adequate model, we used a backward stepwise selection on the basis of the partial *F*-statistic for regressions, and the likelihood ratio test obtained by maximum likelihood for GLS. Where model validation revealed a Cook's distance greater than one for one or several of the data points, the analysis was undertaken both with and without extreme observations.

The consistency of the choice of independent variables and the backward stepwise selection procedure model was checked by using two further selection methods where variables had not been replaced by their principal components:

1. Hierarchical partitioning with the full set of predictive variables. Because the hierarchical partitioning function implemented in the R library hier.part (Walsh & Mac Nally 2008) currently only allows for the simultaneous analysis of 12 predictors, we randomly selected 12 predictors for the hierarchical partitioning and averaged the results for each predictor over 100 repetitions. Candidate predictors that had a significantly higher contribution score than the rest of the variables were chosen. Modeling procedures were as above.
2. Stepwise exclusion of predictors based on univariate models. The total set of candidate predictors was reduced to the strongest uncorrelated set (Pearson's $r < 0.7$) according to the predictive power of variables in univariate tests (Quinn & Keough 2002). This was followed by hierarchical partitioning as above.

The respective contribution of each variable towards explaining the variation in overall perceived plant species richness was established by decomposing the variance in a partial regression (Zuur *et al.* 2007), whereby for each variable the percentage drop in model fit (r^2) is measured when that variable is omitted from the model. This technique allowed us to establish the contribution of all remaining predictor variables, separately and jointly, towards the level of explained variance. In order to evaluate the trend over time, the above modelling procedure followed by partitioning of variance was also performed for perceived species richness at decade intervals (1989 and 1999) (with the funding predictor being calculated for the same decade intervals).

Cross-correlations (Chatfield 2003) were calculated between the amount of funding invested in an area in every year between 1980 and 2007 and the number of new species records for that area, with time lags ranging from five years before to five years after.

We take into consideration that the number of data points used in the analysis was relatively small. However, the dataset contributing to each of these points was extensive. This, in conjunction with the strongly emerging pattern and the consistency established with the model validation procedures, increases our confidence in the reliability of the analysis.

The PCA was calculated in SPSS 11.5; all other statistical analyses were performed in the "R" statistical and programming environment version 2.8.1 (R Development Core Team 2008) and its libraries hier.part (Walsh & Mac Nally 2008) and nlme (Pinheiro *et al.* 2009).

Results

In the EAM, as in many other parts of the world, funding for biodiversity inventories is scarce. Between 1980 and 2007 investments in conservation and research in the region totalled US$2007 117 million. Of this amount only 3% has been invested in botanical inventories. Documented vascular plant richness in the EAM, to date totalling 3,986 species, has increased in three distinct phases following early explorations pre-1980 (Fig. 2.1): during the first intensified exploration in the 1980s; between 1995 and 2000; and from 2004 onwards. The recorded number of threatened and/or endemic species shows a pattern very similar to that of perceived species richness. The funding pattern also shows phases: initial, relatively low levels of funding in the 1980s; a very rapid increase from 1994 to 2001; followed by a second rapid increase from 2004 to present.

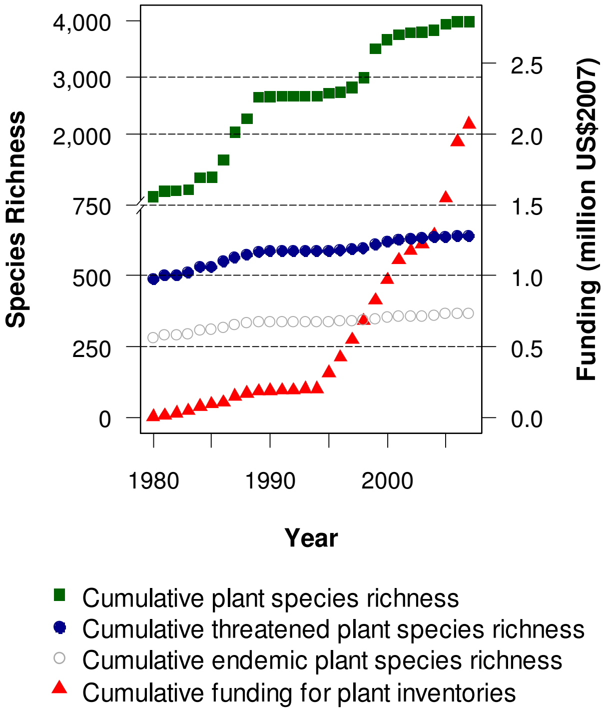


Figure 2.1. Cumulative perceived species richness and funding in the Eastern Arc Mountains between 1980 and 2007. Funding is for plant inventories and has been standardised to US$ in the year 2007. Note the axis break on the first Y axis.

Funding (cumulative investments per mountain bloc for plant inventories between 1980 and 2007) emerged as the best predictor for total perceived plant species richness within mountain blocs. In total, 36 predictors and interactions between these were tested: 34 environmental predictors (climate, topography and disturbance); a survey intensity predictor (number of records per mountain bloc); and the funding predictor (Table S2.1). Partial regressions showed that 65% of the variation in perceived plant species richness is explained by the funding for botanical inventories alone, whereas only 11% is explained by environmental characteristics and disturbance (Fig. 2.2; Table S2.4). The recorded number of threatened and endemic plant species was closely related to overall perceived plant species richness. Both were best predicted by survey intensity in combination with environmental characteristics (Table S2.4). All model results were consistent across all predictor selection procedures.

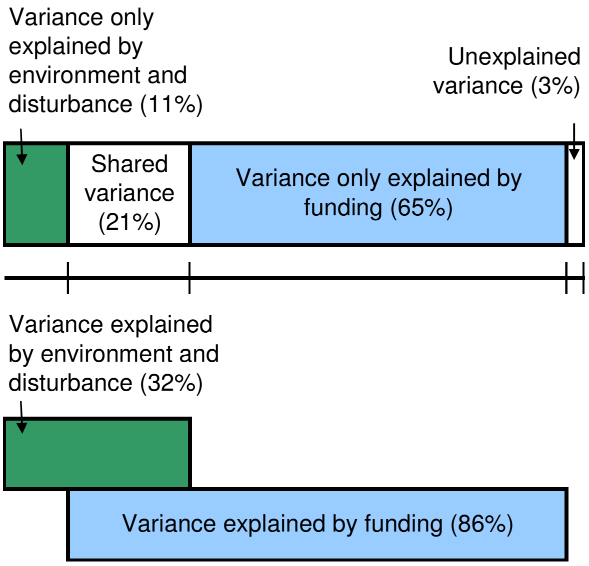


Figure 2.2. Explained variance in perceived plant species richness for the linear regression model, partitioned between its predictors. Funding is the cumulative investment in plant inventories 1980 - 2007. The environmental and disturbance predictors are a principal component representative of maximum potential evapotranspiration (PET), PET range and maximum suitable temperature days (i.e. water-energy, heterogeneity and optimal growth conditions), and the percentage of trees cut. Independent data points *n* = 11 (mountain blocs).

Cross-correlations showed distinctive patterns at the bloc resolution level, revealing their different exploration histories. Correlations between funding and perceived species richness with a negative time lag suggest that perceived species richness is driving funding; conversely, a positive time lag suggests that funding is driving perceived species richness. A significant correlation for a time lag of zero also supports the hypothesis that funding is driving perceived species richness levels. This is due to the high probability that species will be found relatively quickly once funding has been allocated (for *n* = 34 projects we have data on funding start year and year that the majority of the fieldwork took place: for 27 projects time lag = 0 yrs; for 7 projects time lag = 1 yr; average time lag = 0.21 yrs ± 0.07 SE), compared to the much lower probability that an increase in richness would trigger the writing of a proposal with funding being allocated for implementation within the same year. The latter is difficult to show as an increase in perceived species richness cannot be assigned to a certain year; however, Fig. S2.1 shows for example for the Udzungwa Mountains that there was a time lag of c. 5 years between the description of many species from the area (late 1980s) and the allocation of further funding (mid 1990s). For each mountain bloc, cross-correlations are highest at time lags of zero or greater (Fig. 2.3b) and indicate overall that funding is a stronger driver of perceived species richness than *vice versa*. Discoveries of threatened and endemic species, which typically require more intensive study, are also likely to be indirectly driven by funding, as both spatial and temporal patterns in threatened, endemic and overall perceived species richness are very similar (Fig. S2.1a, b) and survey intensity is partly determined by available funding (Pearson [funding, number of records] = 0.7).

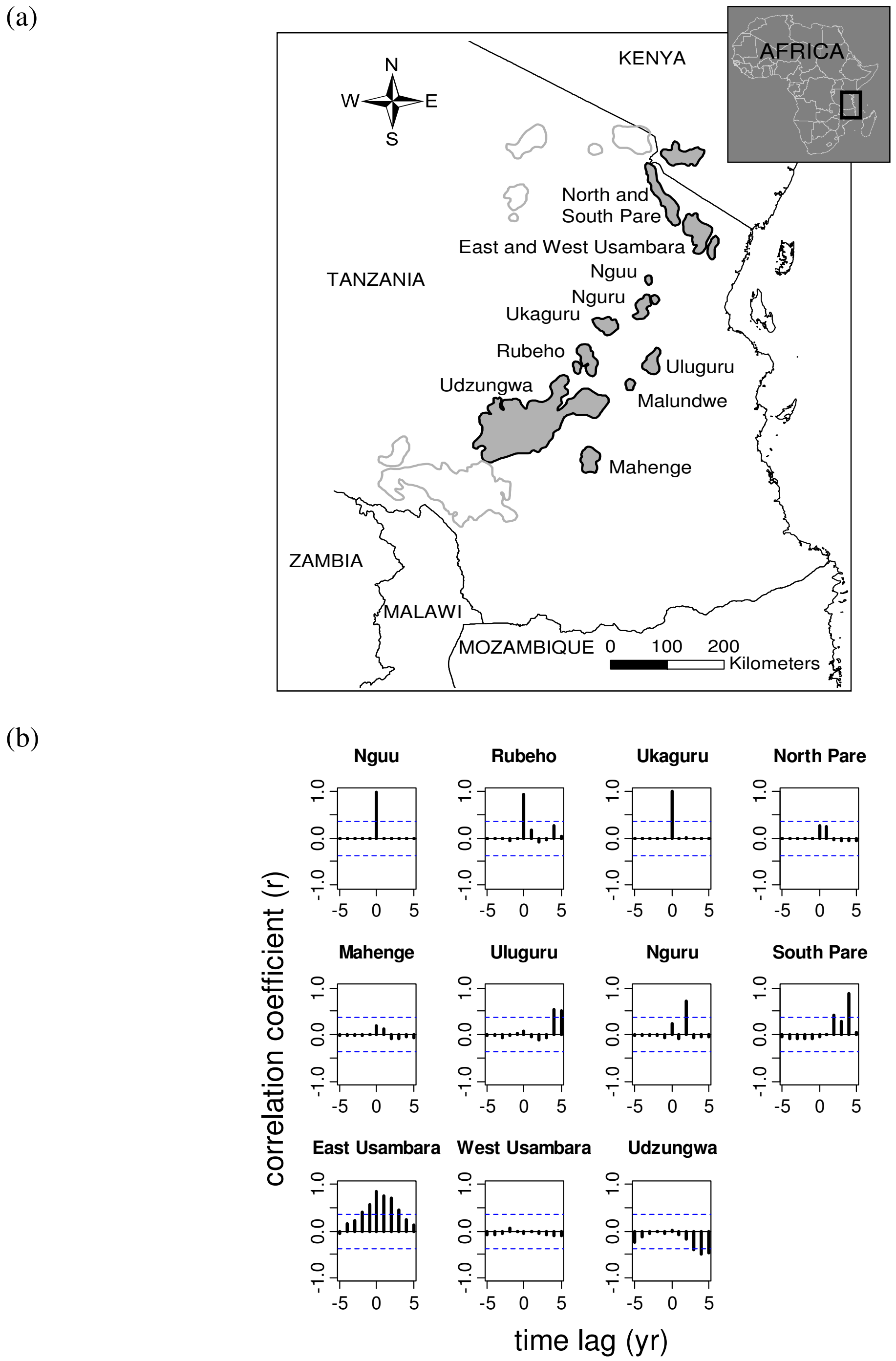


Figure 2.3. Investments in plant inventories and recorded biodiversity in the Eastern Arc Mountains. (a) Map of the Eastern Arc Mountains. (b) Cross-correlations between investment in plant inventories (1980 - 2007) and new plant species recorded. Horizontal axis represents time lag between investment and species discovery, and dotted lines the 95% upper and lower confidence bands. For most mountain blocs correlations are significant for zero or positive time lags suggesting that funding is driving perceived species richness. Correlations are negligible or negative for mountains already well known before the study period where investments did not result in the discovery of further species (West Usambara, Udzungwa) (Fig. S2.2).

The influence of funding on the levels of perceived species richness was strong in the 1980s (Fig. 2.4), which was the starting period of intensified botanical exploration of the EAM. During this time, a new species for the area would be found with, on average, an investment of less than US$2007 100, and funding levels explained a remarkable 99% of variation in the number of new species found (Table S2.4). The influence of funding slightly weakened in the 1990s, a phase of highly intense botanical exploration. On average, a new species record was found with every investment of US$2007 250 and funding was no longer the sole significant explanatory variable. From 2000 onwards, research in the EAM began to target the lesser researched mountain blocs, resulting in the discovery of 477 new species for the region, 29 of which are potentially threatened and/or endemic. On average, a new species for the area was found with every investment of US$2007 500. However, an asymptote in species richness has not yet been approached and the explanatory power of environmental variables is just above 10%, suggesting that many further botanical inventories will be needed in order to get a reliable view of species richness patterns in this area. Similarly, patterns in the observed distribution of threatened and endemic species, which were solely driven by survey intensity in the 1980s, were increasingly related to environmental factors in the 1990s and 2007, with the model coefficient for the survey intensity predictor steadily decreasing (Table S2.4).

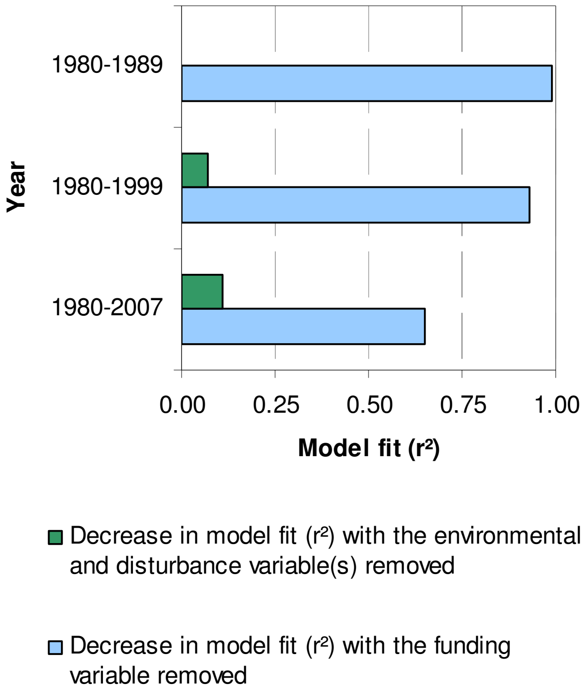


Figure 2.4. Relative importance of the predictors in explaining plant perceived species richness over time. Funding is the cumulative investment in plant inventories between 1980 and the end of the investigated decade. The environmental and disturbance predictors are a principal component representative of maximum potential evapotranspiration, evapotranspiration range and maximum suitable temperature days and the percentage of trees cut. Independent data points $n = 11$ (mountain blocs).

Funds for conservation, research and inventory are highly correlated with each other (Table S2.5), showing that the same mountain blocs are targeted for these different purposes. The close tie between the funds for inventory and those for conservation indicates that where initial biodiversity inventories result in the discovery of new species, further funding is then attracted for conservation and research. This partly finances further inventories and results in the discovery of more species for that area, strengthening its value for conservation and research in a spiral fashion. At the same time, other areas receive relatively little initial funding for inventories; hence the number of species discovered in these areas remains small and their conservation and research status low. The selection of initial areas for investment is often related to accessibility, interest in the area, size, (historical/colonial) land ownership and political considerations (Kadmon *et al.* 2004; Halpern *et al.* 2006). The positive association between perceived species richness and the percentage of trees cut in the EAM, for example, may be a reflection of non-linear effects of disturbance on species richness, but it may also be because easily accessible areas close to roads and markets are targeted by both logging companies and botanists, as historically remote areas would have been extremely difficult if not impossible to reach.

Discussion

This study has three main implications. Firstly, we may have a much distorted view of species diversity patterns, and may not have sufficient data to identify conservation priority areas with certainty, particularly at the site scale (Da Fonseca *et al.* 2000). In the EAM, as elsewhere, research and conservation investment is biased towards the areas that we think are important, and our understanding of the relative conservation importance of areas within different biomes across the world may be biased towards those that have received the most funding for biodiversity inventories. This may also partially explain why surprisingly little consensus has been achieved on the distributional pattern and drivers of species richness (Rahbek *et al.* 2007; Gotelli *et al.* 2009), and possibly also patterns of endemism, which have, for example, been attributed to evolutionary effects and historical climate configurations (Taplin & Lovett 2003; Jetz *et al.* 2004; Buckley & Jetz 2007; Carnaval & Moritz 2008). Secondly, promising progress has been made with species distribution modelling techniques (Elith *et al.* 2006), which can be used to establish the probability of an area's conservation importance for one or several species (Da Fonseca *et al.* 2000; Graham *et al.* 2004). Climatically based fine-scale species distribution models for the area (Platts *et al.* 2010), in line with individual-based rarefaction curves (see Fig. S2.3) for relatively well sampled mountain blocs with > 3000 records (n = 6), suggest that for example the conservation importance of the Rubeho and Nguru Mountains may be underestimated.

However, the predictions from these models for new areas and future scenarios will only be as good as the data that underpin them (Rondinini *et al.* 2006). We recommend that funding and associated sampling intensity biases be considered in the development of these models to achieve more accurate predictions, or at least to inform relevant measures of uncertainty. Thirdly, inventories from potentially important but lesser known areas will be necessary to assess whether their conservation importance has been underestimated. We are not advocating a highly cost-intensive global standardisation of survey intensity for all areas, and we agree that in the face of rapid global biodiversity loss it is important to dedicate efforts to conservation even if our knowledge of patterns in species richness is incomplete (Meir *et al.* 2004; McDonald-Madden *et al.* 2008; Grantham *et al.* 2009). However, a lack of high-quality biodiversity surveys may prove economically disadvantageous (Balmford & Gaston 1999), and by dedicating funds nearly exclusively towards areas that have already been identified to be of high importance, we risk losing other areas of equal importance with greatly underestimated or unknown conservation status. Increasing the funding for biodiversity inventories in potentially important but under-researched areas would reduce this risk and provide us with a more balanced assessment of global diversity patterns, allowing effective conservation of more of the world's biodiversity.

Acknowledgements

We thank A. Balmford, S.A. Bhagwat, P.M. Hollingsworth, R.T. Pennington and D. Raffaelli for critical discussions. Funding for A.A. and P.J.P. was provided by a Marie Curie Actions grant to R.M., and for M.T.B. by NERC. The authors gratefully acknowledge contributions of plant, disturbance and funding data: Frontier Tanzania, the Tanzania Forest Conservation Group, the Tanzanian Forestry and Beekeeping Division, A.R. Marshall, and J.B. Hall provided plant and disturbance field data. BirdLife International (G. Eshiamwata, P. Ndang'ang'a, and P. Nnyiti,), the Critical Ecosystem Partnership Fund (G. Love), DANIDA (T. Blomley), T. Evans, FINNIDA (M. Määttä, M. Nummelin, J. Saramäki, and V. Turpeien), GTZ (Steffen Rötcher), G. Hertel, Missouri Botanical Garden (Peter Phillipson and Porter P. Lowry), Museo Tridentino di Scienze Naturali (M. Menegon, F. Rovero), B. Newmark, T. Struhsaker, The Field Museum of Natural History Chicago (B. Stanley), TFCG (N. Doggart), the UNDP/GEF Cross Border Biodiversity Programme (L. Persha), University of Copenhagen (J. Fjeldså and L. Hansen), University of York (A.R. Marshall), and WWF (P.J. Stephenson and P. Sumbi) provided funding data.

References

African Flowering Plants Database (2008). Conservatoire et Jardin botaniques de la Villle de Genève and South African National Biodiversity Institute Pretoria. http://www.ville-ge.ch/musinfo/bd/cjb/africa/.

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

Brooks T.M., Mittermeier R.A., da Fonseca G.A.B., Gerlach J., Hoffmann M., Lamoreux J.F., Mittermeier C.G., Pilgrim J.D. & Rodrigues A.S.L. (2006). Global biodiversity conservation priorities. *Science*, 313, 58-61.

Buckley L.B. & Jetz W. (2007). Environmental and historical constraints on global patterns of amphibian richness. *Proceedings of the Royal Society B-Biological Sciences*, 274, 1167-1173.

Burgess N.D., Butynski T.M., Cordeiro N.J., Doggart N.H., Fjeldså J., Howell K.M., Kilahama F.B., Loader S.P., Lovett J.C., Mbilinyi B., Menegon M., Moyer D.C., Nashanda E., Perkin A., Rovero F., Stanley W.T. & Stuart S.N. (2007). The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, 134, 209-231.

Carnaval A.C. & Moritz C. (2008). Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, 35, 1187-1201.

Chatfield C. (2003). *The analysis of time series: an introduction*. 6 edn. Chapman and Hall, Ltd, London.

Chevan A. & Sutherland M. (1991). Hierarchical Partitioning. *American Statistician*, 45, 90-96.

Da Fonseca G.A.B., Balmford A., Bibby C., Boitani L., Corsi F., Brooks T., Gascon C., Olivieri S., Mittermeier R.A., Burgess N., Dinerstein E., Olson D., Hannah L., Lovett J., Moyer D., Rahbek C., Stuart S. & Williams P. (2000). It's time to work together and stop duplicating conservation efforts ... following Africa's lead in setting priorities. *Nature*, 405, 393-394.

Dalton R. (2000). Ecologists back blueprint to save biodiversity hotspots. *Nature*, 406, 926-926.

Eastern Africa Plant Red List Authority (2008). List of Potentially Threatened Plant Taxa in the Eastern Arc Mountains of Tanzania. Unpublished report presented to the Forestry and Beekeeping Division, Tanzania Ministry of Natural Resources and Tourism. (ed. Africa Plant Red List Authority) Dar es Salaam.

Eken G., Bennun L., Brooks T.M., Darwall W., Fishpool L.D.C., Foster M., Knox D., Langhammer P., Matiku P., Radford E., Salaman P., Sechrest W., Smith M.L., Spector

S. & Tordoff A. (2004). Key biodiversity areas as site conservation targets. *Bioscience*, 54, 1110-1118.

Elith J., Graham C.H., Anderson R.P., Dudik M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.M., Peterson A.T., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberon J., Williams S., Wisz M.S. & Zimmermann N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129-151.

Fishpool L.D.C. & Evans M.I. (2001). *Important Bird Areas in Africa and associated islands*. BirdLife International, Cambridge.

Gotelli N.J., Anderson M.J., Arita H.T., Chao A., Colwell R.K., Connolly S.R., Currie D.J., Dunn R.R., Graves G.R., Green J.L., Grytnes J.A., Jiang Y.H., Jetz W., Lyons S.K., McCain C.M., Magurran A.E., Rahbek C., Rangel T., Soberon J., Webb C.O. & Willig M.R. (2009). Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, 12, 873-886.

Graham C.H., Ferrier S., Huettman F., Moritz C. & Peterson A.T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, 19, 497-503.

Grantham H.S., Wilson K.A., Moilanen A., Rebelo T. & Possingham H.P. (2009). Delaying conservation actions for improved knowledge: how long should we wait? *Ecology Letters*, 12, 293-301.

Halpern B.S., Pyke C.R., Fox H.E., Haney J.C., Schlaepfer M.A. & Zaradic P. (2006). Gaps and mismatches between global conservation priorities and spending. *Conservation Biology*, 20, 56-64.

Hawkins B.A., Field R., Cornell H.V., Currie D.J., Guegan J.F., Kaufman D.M., Kerr J.T., Mittelbach G.G., Oberdorff T., O'Brien E.M., Porter E.E. & Turner J.R.G. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84, 3105-3117.

Jetz W., Rahbek C. & Colwell R.K. (2004). The coincidence of rarity and richness and the potential signature of history in centres of endemism. *Ecology Letters*, 7, 1180-1191.

Kadmon R., Farber O. & Danin A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, 14, 401-413.

Kier G., Mutke J., Dinerstein E., Ricketts T.H., Kuper W., Kreft H. & Barthlott W. (2005). Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, 32, 1107-1116.

McDonald-Madden E., Bode M., Game E.T., Grantham H. & Possingham H.P. (2008). The need for speed: informed land acquisitions for conservation in a dynamic property market. *Ecology Letters*, 11, 1169-1177.

Meir E., Andelman S. & Possingham H.P. (2004). Does conservation planning matter in a dynamic and uncertain world? *Ecology Letters*, 7, 615-622.

Mittermeier R.A., Robles-Gil P., Hoffmann M., Pilgrim J.D., Brooks T.M., Mittermeier G.C., Lamoreux J.L. & Fonseca, G. (2005). *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Ecoregions*. 2 edn, Cemex, Mexico City.

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

Nelson B.W., Ferreira C.A.C., Dasilva M.F. & Kawasaki M.L. (1990). Endemism Centers, Refugia and Botanical Collection Density in Brazilian Amazonia. *Nature*, 345, 714-716.

Newmark W. (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, 29-36.

Newmark W.D. (2002). Ecological Studies. Conserving biodiversity in East African forests: A study of the Eastern Arc Mountains. In: *Ecological Studies. Conserving biodiversity in East African forests: A study of the Eastern Arc Mountains*. Springer-Verlag, New York.

O'Brien E.M., Field R. & Whittaker R.J. (2000). Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography. *Oikos*, 89, 588-600.

Olson D.M. & Dinerstein E. (2002). The Global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden*, 89, 199-224.

Pinheiro J., Bates D., DebRoy S., Sarkar D. & the R Core team (2009). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-96. Available at: http://CRAN.R-project.org/package=nlme.

Plantlife International (2004). Identifying and protecting the world's most important plant areas. Plantlife International, Salisbury.

Platts, P.J., Ahrends, A., Gereau, R.E., McClean, C., Lovett, J.C., Marshall, A., Pellikka, P., Mulligan, M., Fanning, E., Marchant, R. (2010). Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. *Diversity and Distributions* 16, 628-642.

Platts P.J., McClean C.J., Lovett J.C. & Marchant R. (2008). Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty. *Ecological Modelling*, 218, 121-134.

R Development Core Team (2008). *R: A language and environment for statistical computing*, Vienna. Available at: www.R-project.org.

Rahbek C., Gotelli N.J., Colwell R.K., Entsminger G.L., Rangel T. & Graves G.R. (2007). Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proceedings of the Royal Society B-Biological Sciences*, 274, 165-174.

Reddy S. & Davalos L.M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, 30, 1719-1727.

Rondinini C., Wilson K.A., Boitani L., Grantham H. & Possingham H.P. (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, 9, 1136-1145.

Sakamoto Y. & Ishiguro M.K.G. (1986). *Akaike Information Criterion Statistics*. D. Reidel Publishing Company, Dordrecht/Boston/Lancaster/Tokyo.

Schlüter T. (1997). *Geology of East Africa*. Borntraeger, Berlin.

Sithisarankul P., Weaver V.M., Diener-West M. & Strickland P.T. (1997) Multicollinearity may lead to artificial interaction: An example from a cross sectional study of bio-markers. Southeast Asian Journal of Tropical Medicine and Public Health 28, 404–409.

Soria-Auza R.W. & Kessler M. (2008). The influence of sampling intensity on the perception of the spatial distribution of tropical diversity and endemism: a case study of ferns from Bolivia. *Diversity and Distributions*, 14, 123-130.

Stattersfield A.J., Crosby M.J., Long A.J. & Wege D.C. (1998). *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation*. BirdLife International, Cambridge.

Taplin J.R.D. & Lovett J.C. (2003). Can we predict centres of plant species richness and rarity from environmental variables in sub-Saharan Africa? *Botanical Journal of the Linnean Society*, 142, 187-197.

Walsh C. & Mac Nally R. (2008). hier.part: Hierarchical Partitioning. R package version 1.0-3. Available at: http://CRAN.R-project.org/package=hier.part.

Zuur A.F., Ieno E.N. & Elphick C.S. (2009a). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 3–14.

Zuur A.F., Ieno E.N. & Smith G.M. (2007). *Analysing Ecological Data*. Springer, New York.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. (2009b). *Mixed Effects Models and Extension in Ecology with R*. Springer, New York.

Supporting information

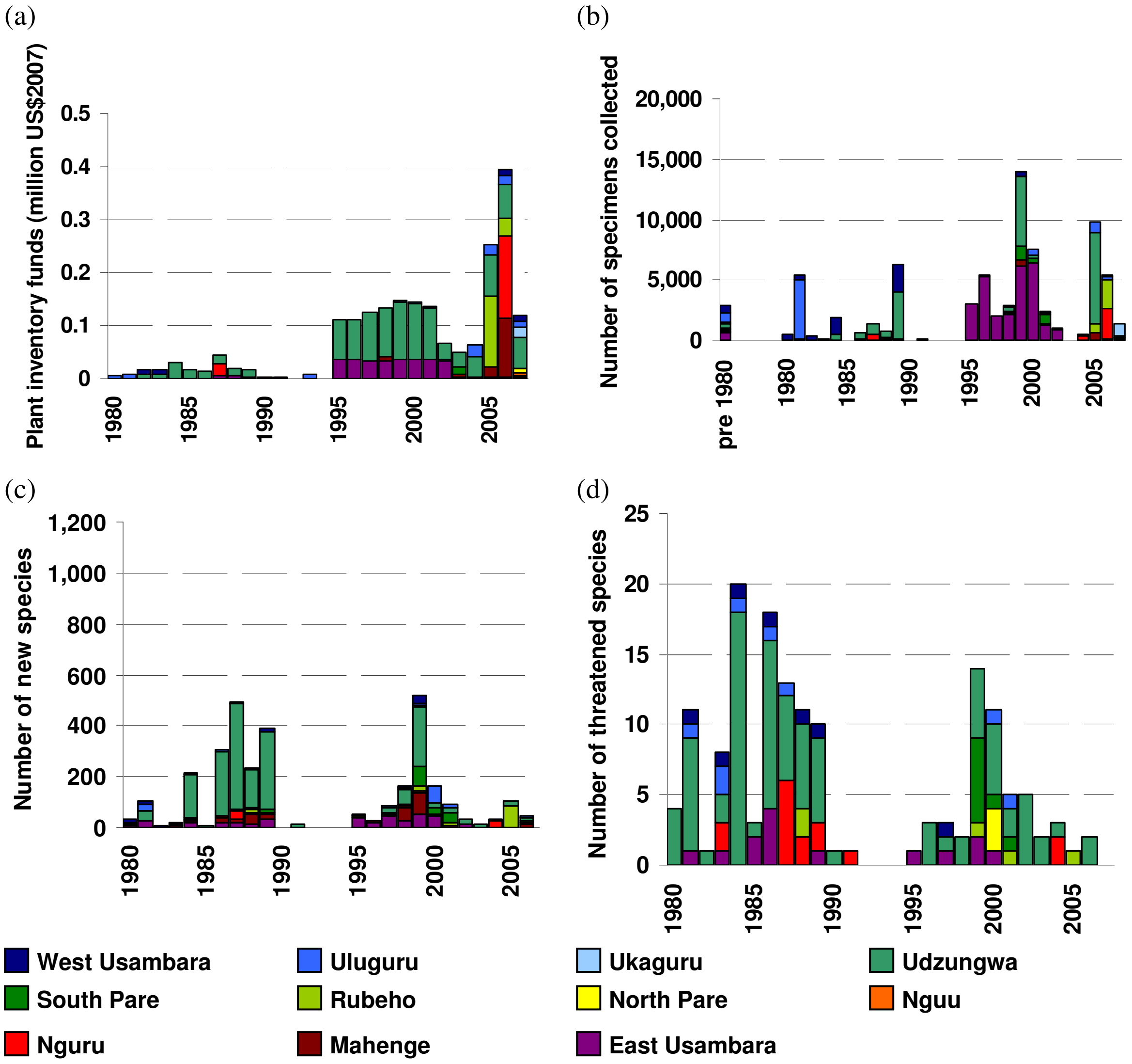


Figure S2.1. Levels of funding, number of specimens and discoveries of new and threatened species over time in individual mountain blocs. In the 1980s and 1990s funding and specimen collections focused on the Udzungwa Mountains, and consequently many new and threatened species were discovered there. Intensive collections in the already well explored East Usambara Mountains (late 1990s) resulted in the discovery of fewer new records. From 2005 onwards funding and specimen collections increasingly targeted at the lesser known mountains, with the consequence that new and threatened species were discovered there, particularly in the Rubeho and Nguru Mountains.

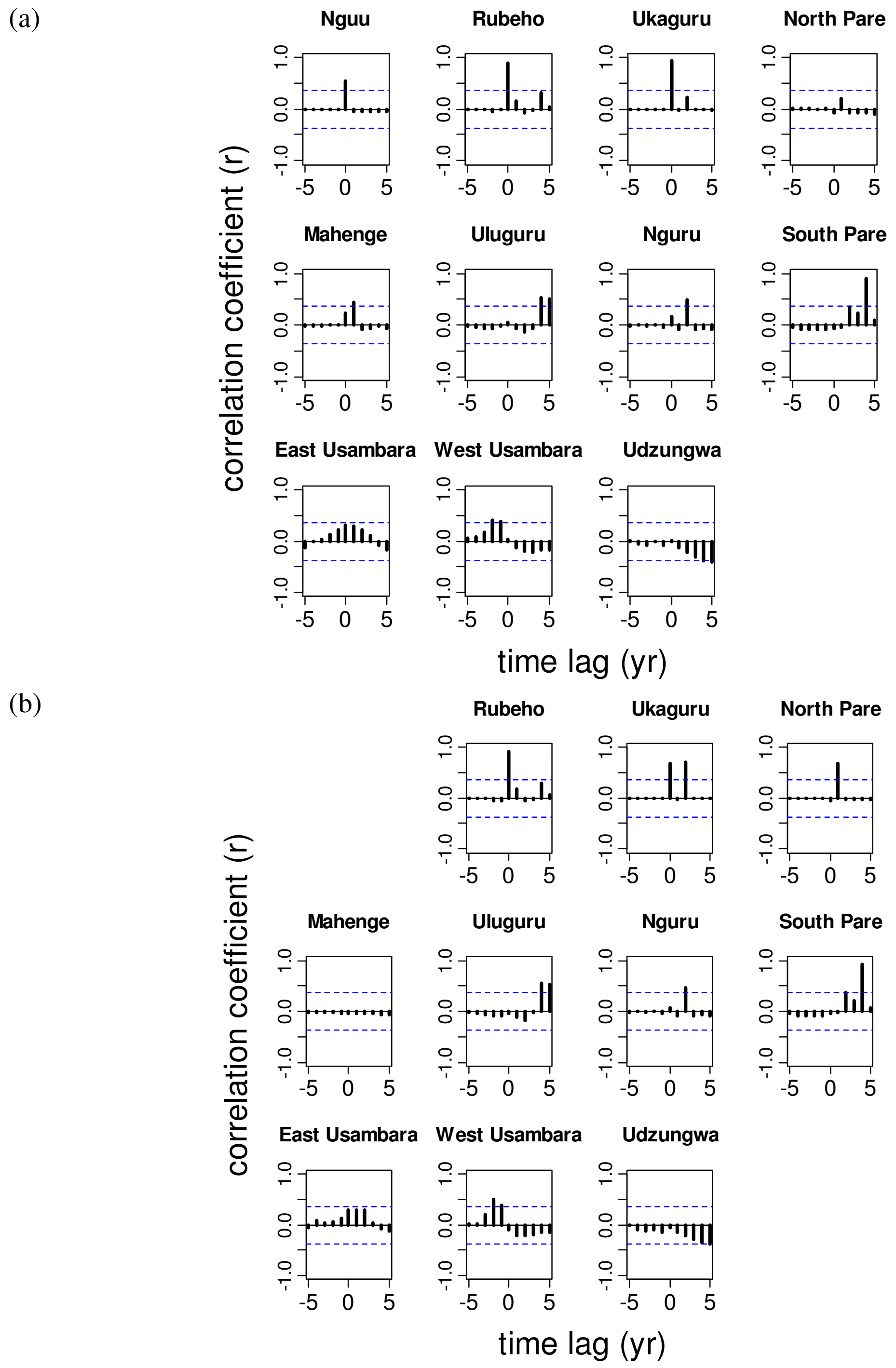


Figure S2.2. Cross-correlations between investment in plant inventories (1980 - 2007) and new threatened (a) and endemic (b) plant species recorded. Horizontal axis represents time lag between investment and species discovery, and dotted lines the 95% upper and lower confidence bands. As for species richness, correlations are negligible or negative for mountains already well known before the study period, indicating that endemic and threatened are not necessary discovered last. Many of the region's endemic and/or threatened species are conspicuous (e.g. *Allanblackia ulugurensis* Engl. has very showy flowers), are locally dominant (e.g. *Schefflera lukwangulensis* (Tennant) Bernardi on the Lukwangule Plateau, Uluguru Mountains), or well known (e.g. *Saintpaulia* spp.).

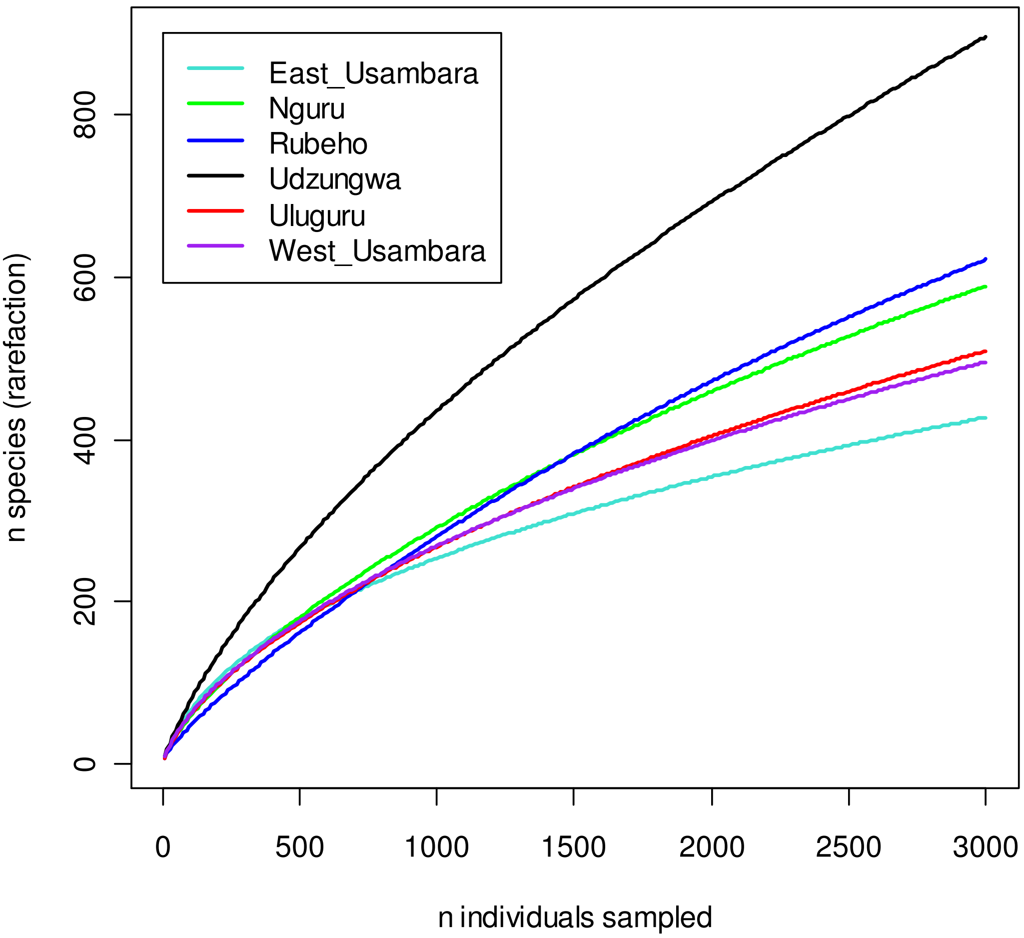


Figure S2.3. Individual-based rarefaction curves for relatively well sampled mountain blocs with > 3000 records. The rarefaction curves have been computed using the R library *vegan* (Oksanen *et al*. 2010).

Predictor	**Calculation**
Temperature	
Mean annual temperature (°C)	
Minimum temperature, coldest month (°C)	
Maximum temperature, warmest month (°C)	
Minimum of minimum temperature, coldest month (°C)	
Maximum of maximum temperature, warmest month (°C)	
Temperature range (°C)	
Precipitation	
Mean annual precipitation (mm)	
Precipitation, driest month (mm)	
Precipitation, wettest month (mm)	
Minimum precipitation, driest month (mm)	
Maximum precipitation, wettest month (mm)	
Precipitation range (mm)	
Potential evapotranspiration (PET)	Thornthwaite, C. W. 1948 An approach towards a rational classification of climate. Geogr. Rev. 38, 55-94.
Mean PET (cm yr-1)	
Minimum PET (cm yr-1)	
Maximum PET (cm yr-1)	
PET range (cm yr-1)	
Annual moisture index (AMI)	Mean annual precipitation/potential evapotranspiration (Platts *et al.* 2008)
Mean AMI (cm yr-1)	
Minimum AMI (cm yr-1)	
Maximum AMI (cm yr-1)	
AMI range (cm yr-1)	
Temperature-days	Suitable growth conditions bounded by maximum and mini-mum temperatures of the warmest and coldest months, respectively, across which a species has presences. Averaged over 40 tree species (Platts *et al.* 2008).
Mean temperature-days (°-days yr-1)	
Minimum temperature-days (°-days yr-1)	
Maximum temperature-days (°-days yr-1)	
Temperature-days range (°-days yr-1)	
Potentially indicative of past climate stability	Distance from the mountain bloc centre to the nearest point along the coast.
Distance from the Indian Ocean (km)	
Topography	
Minimum forested elevation (m a.s.l.)	
Maximum forested elevation (m a.s.l.)	
Forested elevation range (m a.s.l.)	
Mountain bloc size (km^2)	
Area of forest cover (km^2)	
Disturbance	
Population density in the mountain bloc area	
Percentage of poles cut	
Percentage of trees cut	
Percentage of total cuts	

Table S2.1. Starting environmental and disturbance predictors used in the modelling process. All climate predictor were derived at a spatial resolution of three arc-minutes (0.05 decimal degrees).

	Component		
	1	**2**	**3**
Mean annual precipitation	0.352	0.929	0.080
Mean temperature	0.850	0.039	-0.172
Precipitation, wettest month	0.225	0.954	0.046
Precipitation, driest month	0.859	0.025	0.377
Maximum temperature, warmest month	0.796	-0.432	-0.212
Minimum temperature, coldest month	0.898	0.095	0.018
Minimum of minimum temperature,	0.931	0.056	-0.293
Maximum of maximum temperature,	-0.036	-0.686	0.570
Temperature range	-0.892	-0.173	0.379
Minimum precipitation, driest month	0.916	0.006	0.206
Maximum precipitation, wettest month	0.060	0.935	0.273
Precipitation range	-0.097	0.942	0.240
Mean potential evapotranspiration	0.965	-0.105	0.097
Minimum potential evapotranspiration	0.929	0.126	-0.267
Maximum potential evapotranspiration	0.086	0.081	0.953
Potential evapotranspiration range	-0.460	-0.010	0.871
Mean annual moisture index	-0.233	0.946	0.135
Minimum annual moisture index	0.070	0.856	-0.217
Maximum annual moisture index	-0.393	0.693	0.537
Annual moisture index range	-0.488	0.306	0.721
Mean temperature-days	0.962	0.040	-0.102
Minimum temperature-days	0.889	0.130	-0.378
Maximum temperature-days	0.087	0.512	0.727
Temperature-days range	-0.791	0.040	0.574

Table S2.2. Rotated component matrix (rotation method: varimax with Kaiser normalisation; rotation converged in five iterations) for the climate principal components used in the analysis.

	Component	
	1	**2**
Minimum forested elevation	-0.358	0.897
Maximum forested elevation	0.740	0.590
Forested elevation range	0.859	-0.237
Mountain bloc size	0.843	-0.193
Area of forest cover	0.919	-0.166

Table S2.3. Rotated component matrix (rotation method: varimax with Kaiser normalisation; rotation converged in three iterations) for the topography principal components used in the analysis.

Dependent variable	**Fit type**	**Model**	**F(OLS) / logLik (GLS)**	***p***	**df**	**r^2**	**Adj. r^2**
Perceived plant species richness 1989	OLS	y = 82.05 + 0.02 × plant_funds ($p \leq 0.001$)	59.11	≤ 0.001	9	0.99	0.99
Perceived plant species richness 1999	OLS	y = 68.11 + 30.14 × tree_cut ($p \leq 0.001$) + 0.003792 × plant_funds ($p \leq 0.001$)	111	≤ 0.001	8	0.97	0.96
Perceived plant species richness 2007	OLS	y = 200.2 + 126.9 × clim ($p \leq 0.05$) + 27.13 × tree_cut ($p \leq 0.05$) + 0.002 × plant_funds ($p \leq 0.001$)	149.1	≤ 0.001	7	0.97	0.95
Recorded number of threatened species 1989	OLS	y = 36.89 + 0.03 × n_records ($p \leq 0.001$)	27.2	≤ 0.001	9	0.75	0.72
Recorded number of threatened species 1999	GLS	y = 36.75 + 0.02 × n_records ($p \leq 0.05$) + 38.79 × clim ($p \leq 0.001$) - 29.26 × bloc ($p \leq 0.01$)	-44.84		7		
Recorded number of threatened species 2007	GLS	y = 72.27+ 0.01 × n_records ($p \leq 0.05$) + 42.03 × clim ($p \leq 0.05$)	-53.83		8		
Recorded number of endemic species 1989	OLS	y = 18.13 + 0.02 × n_records ($p \leq 0.001$)	34.72	≤ 0.001	9	0.79	0.77
Recorded number of endemic species 1999	GLS	y = 22.37 + 0.01 × n_records ($p \leq 0.1$) + 27.74 × clim ($p \leq 0.001$) - 23.74 × bloc ($p \leq 0.001$)	-41.71		7		
Recorded number of endemic species 2007	GLS	y = 12.65+ 0.003 × n_records ($p \leq 0.01$) + 14.46 × bloc ($p \leq 0.001$)	-46.72		8		

Table S2.4. Minimum adequate models for perceived plant species richness, and recorded numbers of threatened and endemic plant species richness. “Tree_cut“ is the percentage of cut trees per mountain bloc; “clim” is the climate principal component representative of maximum potential evapotranspiration, potential evapotranspiration range and maximum temperature days; “plant_funds” is the plant inventory funding between 1980 and 1989, 1999, and 2007 respectively; “n_records” is the number of plant records.

	Inventory funds	Plant inventory funds	Other research funds
Conservation funds	0.921	0.856	0.64
Inventory funds		0.984	0.579
Plant inventory funds			0.561

Table S2.5. Correlations (Pearson) between total funding (1980 - 2007) for inventories, conservation and other research in the Eastern Arc Mountains.

Link from Chapter 2 to Chapter 3

Results of the previous chapter indicated that there may be a circularity between funding for plant inventories and perceived biodiversity, and that we may have a highly distorted view of regional biodiversity priorities, biased by investment. Funding was highly collinear with the number of records collected, but it explained more variation in the number of species recorded than sampling intensity. It is possible that funding for biodiversity inventories does not only influence the quantity of work undertaken but also its quality, e.g. by facilitating the employment of highly skilled observers and their provision with resources such as access to one of the worldwide leading herbaria for the collection area and identification literature. The second results chapter therefore concentrates on analysing whether projects with higher levels of funding indeed employ better trained observers, and whether there is a relationship between the level of training and resources of a botanist and the number of species they record. An understanding of this relationship is essential as declining resources for descriptive taxonomy and biodiversity inventories may mean that the quality of the data collected decreases, and is biased, whereby more species are reported from areas where better trained surveyors have been employed.

The Udzungwa Mountains (here Ndundulu Forest) harbour the largest fragments of remaining Eastern Arc Mountain forest.

Chapter 3 – Conservation and the botanist effect

In revision in *Biological Conservation* (Full length articles)

Conservation and the botanist effect

Antje Ahrends[1,2], *Carsten Rahbek*[3], *Mark T. Bulling*[4], *Neil D. Burgess*[3,5,6], *Jon C. Lovett*[7], *Victoria Wilkins Kindemba*[8], *Nisha Owen*[8], *Albert Ntemi Sallu*[9], *Andrew R. Marshall*[10], *Boniface E. Mhoro*[11], *Eibleis Fanning*[8] and *Rob Marchant*[2]

[1] Royal Botanic Garden Edinburgh, 20A Inverleith Row, EH3 5LR, UK; [2]Environment Department, University of York, YO105DD, UK; [3] Centre for Macroecology, Institute of Biology, University of Copenhagen, Universitetsparken 15, DK- 2100 Copenhagen Ø, Denmark; [4] Oceanlab, University of Aberdeen, Main Street, Newburgh, Aberdeenshire, AB416AA, UK; [5] Zoology Department, University of Cambridge, CB2 3EJ, UK; [6]WWF US, 1250 24th St NW, Washington DC, USA; [7] Twente Centre for Studies in Technology and Sustainable Development, University of Twente, P.O. Box 217, 7500 Enschede, Netherlands; [8] The Society for Environmental Exploration (Frontier), 50-52 Rivington Street, London, EC2A 3QP, UK; [9]Community Volunteer Services-Tanzania, P.O. Box 303, Tanga, Tanzania; [10]Flamingo Land Ltd., Kirby Misperton, Malton, North Yorkshire, YO17 6UX, UK; [11] University of Dar es Salaam, P. O Box 35091, Dar es Salaam, Tanzania; [12] TRAFFIC, P.O. Box 106060, Dar es Salaam, Tanzania

Abstract

Over the last few decades, resources for descriptive taxonomy and biodiversity inventories have substantially declined. This could mean that the quality of collected biodiversity data decreases overall, and non-uniformly from area to area, depending on the resources available for species identification. We tested this hypothesis with tropical tree records (n = 24,024) collected from the Eastern Arc Mountains, Tanzania, between 1980 and 2007. We focussed on the backgrounds of 13 botanists, whose collections represent 80% of the total plant records for this region. Our findings suggest that botanists with better practical training in tropical plant identification record more species and more species of conservation concern. Combined with the number of person-days, training explained 96% of the variation in the numbers of species the botanists found, while predicted species richness in the survey area (estimated using regional-scale climatically-driven distribution models) was not significant. Training was almost as important for explaining numbers of threatened and endemic species as the relative richness of these species in the survey area. Botanists with better resources and training (access to herbarium, PhD) also provided higher quality data (fewer misidentifications, less misspellings). Our analysis suggests that it may be necessary to account for recorder training when comparing diversity across sites, in particular numbers of rare and endemic species, and for global data portals to provide such information. We also

suggest that greater investment in the training of botanists would pay dividends in terms of increased recording efficiency and reliability of data, and hence improve conservation planning and implementation on the ground.

Introduction

Species losses are occurring at unprecedented levels (Wilson 2000; Novacek & Cleland 2001) and anthropogenic pressures have been identified as the major cause (Vitousek *et al.* 1997). The rate at which we are losing biodiversity is projected to increase in the face of global environmental change (Brook *et al.* 2008; Stork 2010). In order to conserve species and ecosystems effectively we need reliable information on the distribution of biodiversity (Pimm & Lawton 1998), particularly because limited resources (James *et al.* 1999) force us to focus conservation efforts on the most important areas in greatest need (Margules & Pressey 2000).

At the same time, resources for descriptive taxonomy, collections and biodiversity inventories are declining (Disney 1989; Ehrenfeld 1989; Whitehead 1990; Gaston & May 1992; Gee 1992; Wheeler *et al.* 2004). Higher education institutions often do not replace retiring taxonomists (Feldmann & Manning 1992), and while university ecology and conservation curricula increasingly emphasise statistics and the use of Geographical Information Systems, the number of courses offered in systematic biology has been widely reduced (Noss 1996). These trends and the fact that measures for academic performance such as the citation index do not favour basic taxonomic work (Valdecasas *et al.* 2000; Samyn & Massin 2002) reduce the incentive for students to enter a career in systematic biology. Today, natural history is often thought of as a hobby (Rivas 1997) and there is an increasing reliance on amateur taxonomists (Hopkins & Freckleton 2002) and volunteer labour (Darwall & Dulvy 1996; Brandon *et al.* 2003; Haag 2005; Brightsmith 2008; Lovell *et al.* 2009; Schmeller *et al.* 2009). Lessening support for basic biodiversity inventories hits the tropics particularly hard because their biodiversity remains severely understudied (Prance *et al.* 2000). With funds for biodiversity assessments and the training of botanists and zoologists being chronically inadequate (Da Fonseca *et al.* 2000), the collection of tropical data is increasingly dominated by institutions that employ untrained volunteers (Cousins 2007; Brightsmith 2008) and 'parataxonomists' (Basset *et al.* 2004).

Declining resources for taxonomy and training may mean that the quality of collected biodiversity data decreases, and also varies from area to area depending on the available resources for taxonomic identification. Almost two thirds of a sample of 80 recent ecological papers did not state how correct identifications were verified, suggesting that neither expert taxonomists' knowledge nor identification literature were used (Bortolus 2008). Random observer effects, introducing noise in reported species richness and numbers of species of conservation concern, are widely acknowledged (Archaux 2009; Leps & Hadincova 1992). Systematic effects have been documented less frequently; however, understanding and accounting for such effects is extremely important as they may introduce directional biases in census estimates. It is conceivable that a field botanist with less training and fewer resources may be more prone to misidentifications (Scott & Hallam 2003), remain with a larger number of unidentified specimens, and identify fewer species and rarities. Such an effect would severely hamper our ability to pinpoint areas of conservation priority because we would be unsure whether the data collected were a reliable reflection of the actual species pool or strongly biased due to the limited taxonomic resources.

In this paper we collated an extensive database of plant records from the Eastern Arc Mountains (EAM), a series of mountain ranges within the Eastern Afromontane biodiversity hotspot (Mittermeier *et al.* 2005), and examined the potential effect of botanists' training and resources. Tanzania provides a good case study area because there, as in many other tropical countries, professional botanists are becoming rare, the herbaria are under-funded and under–staffed, and yet Tanzania has a relatively well-documented flora (Beentje & Smith 2001) and probably the largest number of vascular plant species of any country in tropical Africa (Roy E. Gereau, Missouri Botanical Garden, personal communication).

In our analysis we focused on three questions:

(1) Are the training of botanists and the resources available to them better predictors of the documented numbers of species and threatened or endemic species than (a) actual differences in plant diversity, (b) sampling intensity, and (c) the number of sample locations?

(2) Is the quality of the botanists' data related to their training and resources, e.g. do botanists with less training more frequently misidentify specimens?

(3) Are the perceived plant diversity patterns across the EAM (and associated conservation priorities) biased by a 'botanist effect' (spatial distribution of training of botanists and resources available to them)?

Addressing these questions is an important first step towards understanding the influence of biodiversity inventory training and resources on conservation planning, which is essential in the face of rapid biodiversity loss.

Materials and methods

Study area

The EAM are a chain of 13 ancient crystalline mountain blocs composed of heavily metamorphosed Precambrian basement rock and estimated to have been uplifted in the Miocene 30 million years ago (Schlüter 1997). The mountains stretch from south-east Kenya to south-central Tanzania and are under the direct climatic influence of the Indian Ocean. Today, they support ~3,300 km² - 5,100 km^2 of tropical forest (Newmark 1998; Newmark 2002; Burgess *et al.* 2007; Platts *et al.* 2010), which may be less than 30% of the original forested area (Burgess *et al.* 2007).

Data

Species records. We collated all available vegetation plot assessments for the region (n = 2,247), totalling 70,081 records (62,224 identified to species) collected by 14 (leading) field botanists. All records were taxonomically standardised to the African Flowering Plants Database (2008), and further updated by reference to taxonomic revisions and monographs by Roy E. Gereau. In the analysis only trees with a diameter at reference height (1.3 m along the stem or above buttresses; drh) greater than or equal to 200 mm (n = 33,512 identified to species, in 2,213 plots) were considered, the minimum drh that had been sampled by all botanists. Because the number of trees assessed by the botanists differed (371 – 4,594), we randomly sampled 250 individuals out of all the trees assessed by each respective botanist, and recorded the number of species found, also noting the numbers of threatened (Eastern Africa Plant Red List Authority 2008) and endemic species (Roy E. Gereau, unpublished data) reported. The results were averaged over 1,000 repetitions.

Botanist data. The botanists' training and resources were scored in eight categories (Table 3.1). We derived the scoring system through discussion with three active field botanists on factors that are important for accurate botanical work and objectively measurable. The scores were kept as general as possible, typically only differentiating three categories, in order to minimise errors resulting from subjective decisions near boundary placements. Unable to score all categories with a high level of confidence for one botanist, we retained 13 botanists for the final analysis (24,024 trees in 1,863 plots). The quality of the data collected by the botanists was measured in six categories: percentage of (1) unidentified species, (2) species with uncertain identification, (3) almost certainly misidentified species (species recorded way outside their recognised distribution area (different continent or part of Africa) and that are not known to have been introduced), (4) misspelt species, (5) species with uncertain identification due to spelling errors and (6) unrecognisable species due to spelling errors (see Supporting information for Chapter 3 Table S3.1).

Category	Score	Explanation
Regional experience	0	Less than five years (sum of actual time spent in the field and in the herbarium) of experience in identifying plants from tropical East Africa
	0.5	Five years or more of experience in identifying plants from tropical East Africa but less than 5 years of experience in the Eastern Arc Mountains.
	1	Five years or more of experience in identifying plants from the Eastern Arc Mountains.
MSc	0	No botany related MSc.
	0.5	Partly botany related MSc.
	1	Botany related MSc.
PhD	0	No botany related PhD.
	0.5	Partly botany related PhD.
	1	Botany related PhD.
Training	0	No formal training in tropical plant identification.
	0.5	Less than six months of formal training in tropical plant identification.
	1	Six months or more of formal training in tropical plant identification.
Herbarium access	0	No access to a worldwide leading herbarium for East Africa with good facilities and extensive collections (East African Herbarium, Kew, Missouri) for specimen identification.
	1	Access to a worldwide leading herbarium for East Africa for specimen identification.
Herbarium staff	0	Never worked as herbarium staff.
	0.5	Worked as herbarium staff for part of the career.
	1	Career as herbarium staff.
Access to identification literature	0	No access to identification literature.
	0.5	Occasional access to the complete Flora of Tropical East Africa and other identification literature (e.g. upon visiting a herbarium).
	1	Full access to the complete Flora of Tropical East Africa and other identification literature.
Collaboration with taxonomic experts	0	Never collaborated with expert taxonomists.
	0.5	Occasional collaboration with expert taxonomists.
	1	Regular collaboration with expert taxonomists.

Table 3.1. The scoring system used to measure the training and resources available to the botanists, based on the expert opinion of three leading botanists.

Other predictor variables. In addition to the botanists' training and resources we considered nine other candidate predictors for modelling the number of species found by the botanists, which would be expected to drive the dependent variables in the absence of a botanist effect: minimum altitude sampled, altitude range sampled, number of vegetation plots sampled, number of mountain blocs sampled, number of assessed trees greater than 200 mm drh, number of days spent in the field, and number of days spent in the field multiplied by number of field staff on those days (person-days) and predictive estimates of species richness in the sampled mountain blocs. For threatened and endemic species we also included their relative richness (ratio of these species to all species richness in the sampled mountain blocs). Predicted numbers of species in the sampling areas were based on regional-scale climatically-driven species distribution models (Platts *et al.* 2010). Uncertainties associated with these variables are noted; for example, the botanists may have sampled only a small area of the entire bloc and the model predictions themselves are prone to biases; however, they are a best possible approximation. Another potential problem is that some of the data collected by the botanists were used to develop the models, i.e. there is a risk of circularity which may increase the probability of committing a Type II error (increased chance of accepting the null hypothesis that there is no botanist affect).

Analysis

Species richness and number of threatened/endemic species found by the botanists. In order to establish significant predictors for species richness and numbers of threatened and endemic species found by the botanists we employed a linear regression approach. Firstly, we tested for collinearity between predictor variables (Pearson ≥ 0.7), and reduced the total set of candidate predictors to the strongest uncorrelated predictors (measured by their predictive power when fitted against the dependent variable on their own) (Quinn & Keough 2002) to avoid inflated standard errors of the variables (Zuur *et al.* 2007). There is a risk of this procedure resulting in the exclusion of driving variables, and we therefore present all collinear variables in Table S3.2. The elimination procedure left us with over 10 candidate predictors in all three analyses. Because this set was still impractically large, in each case we used hierarchical partitioning (Chevan & Sutherland 1991) to identify a small subset of the predictors most likely to play a critical role in determining the value of the dependent variable. We then fitted a multiple linear regression model. Validation procedures, following Zuur *et al.* (2009), indicated no problems associated with assumptions of normality and heterogeneity of variance. To find the minimum adequate model in each case, we then applied a backward stepwise selection based on the partial *F*-statistic. Where model validation revealed a Cook's distance greater than one for one or several of the data points, the analysis was undertaken both with and without these observations in order to assess if they had any significant impact on the structure of the minimum adequate model. The respective contribution of each variable towards explaining the variation in reported species richness and numbers of threatened and endemic species was established by decomposing the variance in a partial regression (Zuur *et al.* 2007).

Quality of the data supplied by the botanists. We employed a multivariate approach to establish whether particular aspects of the botanists' training and resources were significantly associated with the quality of the data they provided (for measures see Table S3.1). The analysis consisted of two steps – ordination and vector fitting. The use of Non-metric Multidimensional Scaling (NMDS), one of the most robust ordination methods (Minchin 1987), allowed us to account for non-normality in our data. We based dissimilarity on Gower distance (Gower 1971), and used 20 iterations with random starts to seek the most stable solution (minimum stress). Subsequent to the ordination we fitted predictors representative of the botanists' training and resources (Table 3.1) as vectors onto the ordination. We calculated squared correlation coefficients (R^2) for each predictor and established their significance in 1,000 random permutations.

Relationship between project funding, botanist training and perceived conservation importance. A previous study in the EAM, modelling plant species richness and the number of threatened and endemic plant species recorded to date for each mountain bloc with a starting set of 36 variables (climate, topography, disturbance, number of plant records per mountain bloc and funding for biodiversity surveys; see Supporting information for Chapter 2 Table S2.1) showed that documented plant diversity patterns in the EAM are likely driven by funding (Chapter 2). We tested whether this may partly be due to funding influencing the level of skill of employed botanists, which may in turn increase the number of species found. First, we analysed correlations between funding and botanic training, and subsequently tested for a botanist effect on perceived biodiversity by modelling plant recording efficiency on botanist training and number of field days. We also included an interaction term between these two variables because recording efficiency is likely to vary with the number of available field days (a minimum number of field days are needed to collect efficiently). A botanist training score for each individual mountain bloc was derived as follows

$$\sum_{b=1}^{B} \left(\frac{r_b}{R} t_b\right)$$

where b is an individual botanists, r_b is the number of records made by them, t_b is their training score, and B and R are the total number of botanists and records, respectively. In total, we developed three models (for species richness, and numbers of threatened and endemic species). Recording efficiency was measured as the number of recorded species divided by the logarithm of the number of records. This type of transformation was chosen based on Mosteller and Turkey's bulging rule from the Box-Cox family of transformations (Zuur *et al*. 2007).

The plant species data were based on a recent compilation of all available plant records for the area, a dataset totalling 75,631 records of specimen label data from the Missouri Botanical Garden's TROPICOS database (http://www.tropicos.org/), with specimen collections for the EAM from a wide range of herbarium and literature sources, and from 2,216 vegetation plot assessments (including those used in this paper). (For further details see Chapter 2 and the Appendix). These data were recorded by over 500 collectors; detailed information was available for only 13 of these individuals. The botanist variable is, however, representative in our view because the 13 botanists participating in this study have made major contributions to the reconnaissance of the regional flora: they have collected 80% (n = 60,193) of the currently available plant records for the EAM, and over 90% in four of the 12 mountain blocs (East Usambara, Nguru, Nguu, Ukaguru). The botanists also spanned a range of training and other taxonomic resource levels. Model selection, validation, search for the minimum adequate model and procedures for dealing with extreme observations were as above.

All statistical analyses were performed in the "R" statistical and programming environment version 2.9.2 (R Development Core Team 2009) and its libraries hier.part (Walsh & Mac Nally 2008), nlme (Pinheiro *et al.* 2009) and vegan (Oksanen 2007).

Results

Species richness and number of threatened and endemic species found

The botanists' training in tropical plant identification was a highly significant predictor for reported species richness, with training and the number of person-days combined explaining 96% of the variation in the number of species documented (Table 3.2). Other candidate predictors, such as overall species richness in the collection areas and predictors representative of the heterogeneity of the sampled locations (altitude range, number of plots, and number of mountains) were not found to be significant. The selection of significant predictors was consistent across the analyses with and without a single observation with a Cook's distance greater than one. The coefficients show that, on average, a trained botanist found ~20 more species for every 250 individuals recorded than an untrained botanist, whereby it did not make a difference whether the botanist had received more or less than six months of training on tropical plant identification. The overall model fit dropped by 14% when the botanists' training was removed as an explanatory variable, and by 33% when person-days were removed (Fig. 3.1), suggesting that survey intensity in the field is more important than training in determining species richness.

Models for reported numbers of threatened and endemic species also explained over 90% of the total variance (Table 3.2). They showed that trained botanists found more threatened and endemic species, but contrary to species richness, training of more than six months had an important impact. Training was almost as important as the relative richness of these species of conservation concern in the respective survey areas, whilst the number of trees sampled was of relatively little importance (Table 3.2; Fig. 3.1).

Variable	**Predictor**	**Coef.**	***p***	***F***	**df**	**r^2 (adj.)**
Species richness	General model		≤ 0.001	57.38	8	0.96 (0.94)
	intercept	9.59	≤ 0.05			
	training < 6 months	22.28	≤ 0.01			
	training > 6 months	19.34	≤ 0.01			
	person-days	0.12	≤ 0.001			
Number of threatened species	General model		≤ 0.001	32.71	8	0.94 (0.91)
	intercept	-4.37	≤ 0.001			
	training < 6 months	0.13	> 0.05			
	training > 6 months	1.21	≤ 0.01			
	number trees sampled	2.43×10^{-4}	≤ 0.05			
	relative richness threatened species	0.84	≤ 0.01			
Number of endemic species	General model		≤ 0.001	40.99	9	0.93 (0.91)
	intercept	-6.58	≤ 0.001			
	training < 6 months	0.43	> 0.05			
	training > 6 months	1.48	≤ 0.01			
	relative richness endemic species	0.56	≤ 0.001			

Table 3.2. Model results for species richness and numbers of threatened and endemic species found by the botanists.

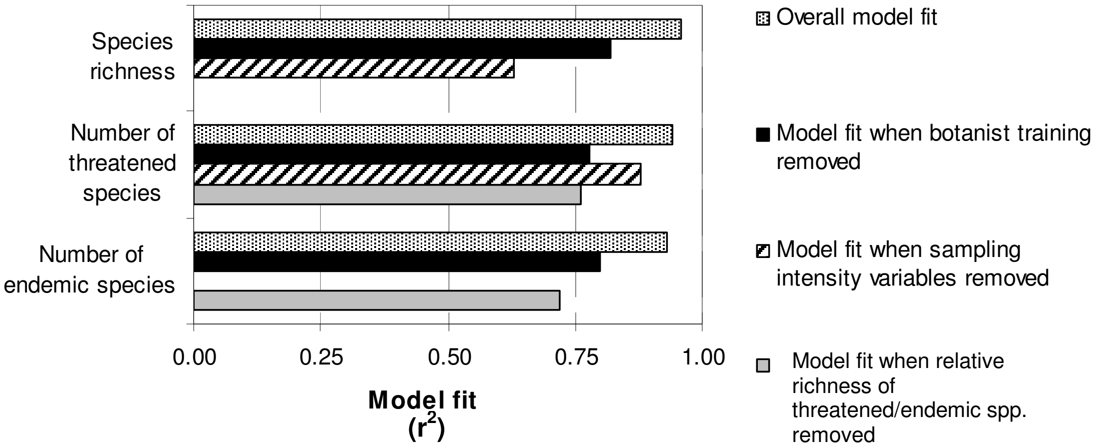
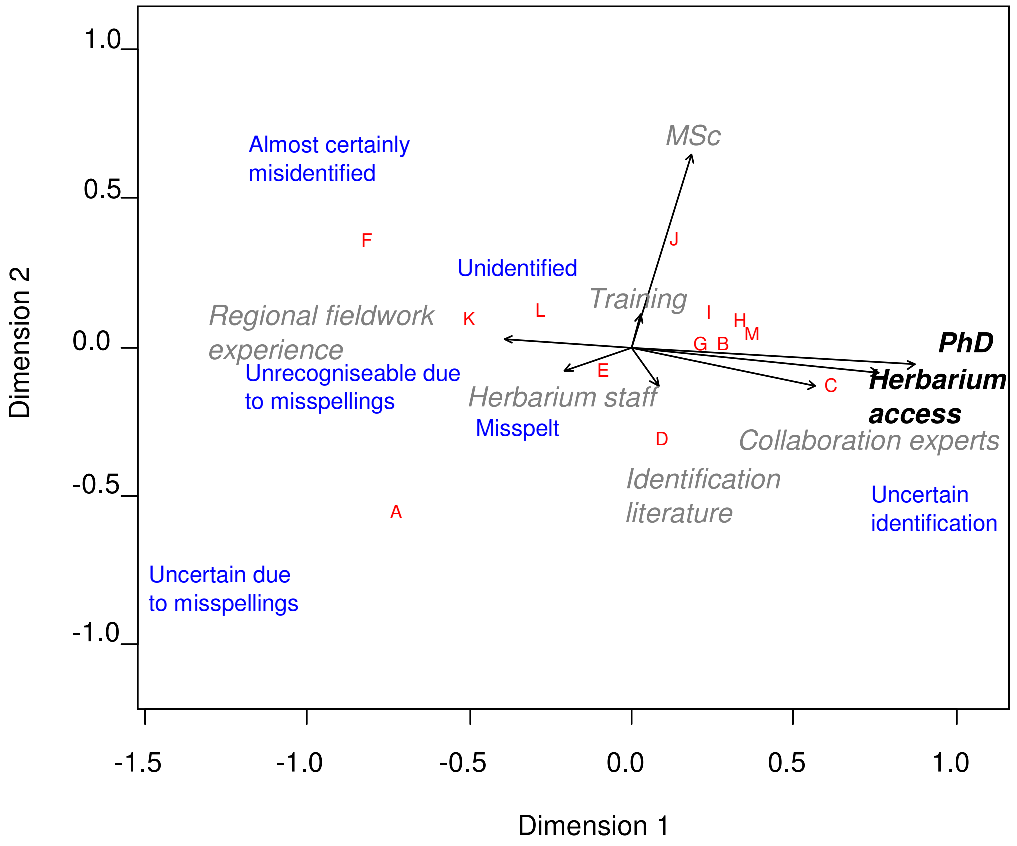


Figure 3.1. Relative importance of the predictors in explaining the number of species, the number of threatened species and the number of endemic species found by the botanists. Note that the gap in the graph for number of endemic species is due to the fact that none of the sampling intensity variables (person days in the case of species richness and numbers of trees sampled in the case of threatened species) was a significant predictor in the model. (For more explanation on the variables see Table 3.1, and for the associated model coefficients see Table 3.2).

Quality of the data supplied by the botanists

Out of 70,081 records, 7,857 (11%) were unidentified, 36 (< 1%) almost certainly misidentified, 4,158 (6%) misspelt and 7% of the misspelt records were entirely unrecognisable. Had the original data been used without checking and corrections being made, overall species richness would have appeared nearly twice as high (1,806 species instead of 925) due to spelling errors, use of synonyms and misidentifications. Multivariate analysis showed that the most significant predictors for data quality were academic training, and access to herbaria (Fig. 3.2). Botanists with a PhD in botany and/or easy access to one of the worldwide leading herbaria for East Africa tended to provide more thoroughly spelt and checked data. Their records were seldom unrecognisable due to spelling mistakes and contained fewer unidentified specimens. They also more frequently marked identifications as uncertain. Overall, data quality was highest in the 1980s and since then has declined while collection rates have slightly increased (Fig. S3.1). This is likely due to an increasing dominance of botanists with less than six months training, who accounted for 14% of the records in the 1980s and 25% since 1990.



	NMDS axis 1	NMDS axis 2	r^2	*p*
Regional fieldwork experience	1.00	-0.07	0.11	>0.1
MSc	-0.28	-0.96	0.33	>0.1
PhD	-1.00	0.07	0.56	<0.01
Training	-0.26	-0.97	0.01	>0.1
Herbarium access	-0.99	0.11	0.43	<0.05
Herbarium staff	0.93	0.36	0.04	>0.1
Access to identification literature	-0.54	0.84	0.02	>0.1
Collaboration with taxonomic experts	-0.97	0.23	0.25	>0.1

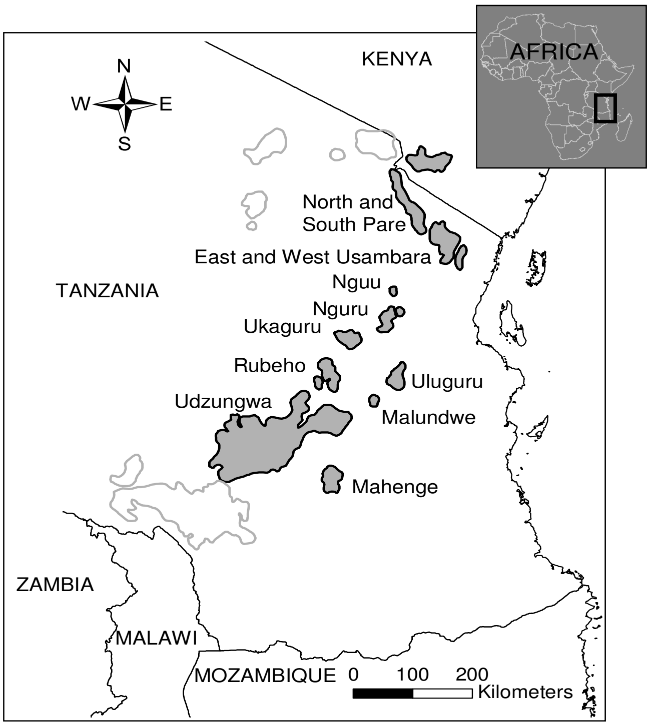
Figure 3.2. Ordination (NMDS) graph of the quality of the collected data and predictors. The predictors are fitted as vectors, pointing in the direction of the most rapid change in the particular predictor. The length of a vector is proportional to the predictor's correlation with the ordination. All predictor names are in italics; significant predictors are highlighted in bold; all others are in grey. All data quality aspects are in blue and the botanists in red, coded as letters from A-M.

Relationship between project funding, botanist training and perceived conservation importance

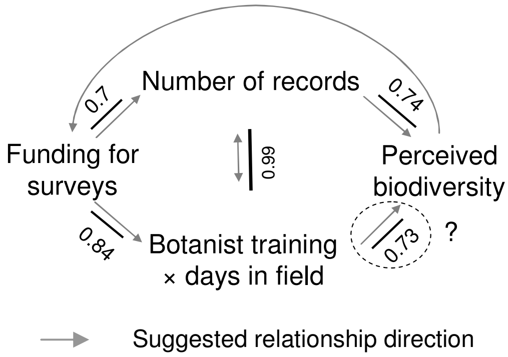
There were significant positive correlations between the available funds for surveys, the number of records collected, the training score of the employed botanists multiplied by their time in the field, and the perceived plant diversity of the respective mountain blocs (Fig. 3.3b). These correlations suggest that funding influences both the number of records sampled and botanist quality, which in turn influence perceived biodiversity. There is also likely circularity between perceived biodiversity and funding (Chapter 2). In order to test for a botanist effect independent of the number of records that have been collected we modelled recording efficiency (number of species found over number of records collected) on the standardised botanist training score and the number of field days. The interaction between these two explanatory variables was significant for all three models (Fig. 3.3c). Number of field days tended to be more significant than the botanist training score (i.e. sufficient time in the field is likely an important efficiency factor), but the botanist training consistently positively influenced recording efficiency, suggesting that independent of the number of records sampled and time there is a botanist effect on the perceived plant diversity patterns in the EAM.

In terms of cost per recorded species, botanists with little training were least, and botanists with an intermediate level of training (six months or less) were most efficient: the average cost per reported species in projects that employed botanists trained to an intermediate level (mean training score of 0.5) were US$ 538, US$ 3,843 and US$ 4,798 per reported species, endemic and threatened species respectively; projects that employed intensively trained botanists (average training score > 0.5) invested US$ 1,559, US$ 11,959 and US$ 18,764; and projects that employed poorly trained botanists (average training score < 0.5) US$ 762, US$ 9,311, and US$ 10,119 (all US$ values standardised to the year 2007 using a GDP deflator; www.measuringworth.com). This is before accounting for the financial cost of correcting identification and entry mistakes. However, botanists with a training of six months or more made a greater overall contribution to reported floristic diversity in the EAM: they documented 296 species (46 threatened and 57 endemic) not reported by botanists with less training; the latter reported only 57 species (7 threatened and 8 endemic) that had not been collected by more intensively trained botanists, although botanists with a training of less than six months provided almost 50% of the plant records for the study area.

(a)



(b)



		Pearson
No. records	Recorded threatened spp.	0.75
Bot. Training x field days		0.75
No. records	Recorded endemic spp.	0.59
Bot. Training x field days		0.61

(c)

Variable	Predictor	Coef.	*p*	*F*	df	r^2 (adj.)
Species richness recording efficiency	General model		≤ 0.05	7.61	7	0.77 (0.66)
	intercept	-39.7	> 0.1			
	training score	110.09	≤ 0.1			
	field days	2.48	≤ 0.05			
	training score : field days	-2.21	≤ 0.05			
Threatened species recording efficiency	General model		≤ 0.001	20.28	7	0.9 (0.85)
	intercept	-5.74	> 0.1			
	training score	14.74	≤ 0.05			
	field days	0.41	≤ 0.01			
	training score : field days	-0.37	≤ 0.01			
Endemic species recording efficiency	General model		≤ 0.05	6.23	7	0.73 (0.61)
	intercept	-3.17	> 0.1			
	training score	6.54	> 0.1			
	field days	0.25	≤ 0.05			
	training score : field days	-0.23	≤ 0.05			

Figure 3.3. Relationships between funding for projects, numbers of records sampled, botanist training and perceived diversity patterns in the Eastern Arc Mountains. (a) Map of the Eastern Arc Mountains. (b) Relationships between funding, numbers of records sampled, botanist training and perceived biodiversity patterns in the Eastern Arc Mountains. Numbers represent Pearson correlation coefficients (for species richness; correlations for threatened and endemic species are given in the table directly below). The arrows show the suggested direction of the relationship. While it has previously been suggested that funding and the number of records influence perceived biodiversity which in turn influences funding (Chapter 2), this paper's analysis focuses on the encircled relationship between botanist resources and perceived biodiversity. (c) Models for recording efficiency. Irrespective of the number of number of records made, better trained botanists and/or those that have more time in the field find more species.

Discussion

Concerned about the decline of support for taxonomy and field biology, we analysed whether this may mean that the quality of collected biodiversity data decreases overall, and non-uniformly from area to area, depending on the resources available for species identification. Our analysis was speculative in nature, due to the inherent subjectivity involved in designing a scoring system, and because we had data from only 13 botanists, who sampled in different areas for which true species richness and levels of endemism and threat were unknown. However, the information collected for each of these botanists and their records were detailed and extensive, and the emerging pattern was strong and consistent across all analyses, increasing our confidence in the results.

Our findings suggest that better trained botanists record more species and more species of conservation concern. Combined with the number of person-days, training almost entirely explained spatial variations in the numbers of species the botanists found. Training (on its own) was almost as important in explaining recorded numbers threatened and endemic species as the relative richness of these species in the survey area. It is necessary to exclude the possibility that better trained botanists simply visited more diverse areas. Correlations between the level of training and the modelled total plant diversity of the areas visited by the botanists are insignificant (Pearson [training, total species richness] = 0.136, $p > 0.1$; Pearson [training, total number of threatened species] = 0.039, $p > 0.1$; Pearson [training, total number of endemic species] = -0.122, $p > 0.1$).

Perceived plant diversity patterns across the EAM (and associated conservation priorities) are largely driven by funding invested for biodiversity inventories (Chapter 2). This may be because better funded projects facilitate more time in the field and employ better trained botanists, which according to our results increases recording efficiency. Further research into these biases in other areas/taxonomic groups and at different scales is necessary, however the finding that funding, botanist and survey intensity effects almost entirely explain the variation in the perceived plant diversity patterns in the EAM is alarming, and stresses the importance of greater transparency of data that underpin conservation decisions. We recommend that schemes using survey data to prioritise areas for conservation collect comprehensive metadata about the origins of those data, and test and potentially account for biases. Furthermore, potential data quality issues should be documented by data portals such as GBIF (http://www.gbif.org/).

The importance of practical training in recording accuracy has previously been documented for invertebrates (Lovell *et al.* 2009), low density populations of an invasive insect (Fitzpatrick *et al.* 2009), coral reef fish (Darwall & Dulvy 1996), and lichens (McCune *et al.* 1997). Vascular plant studies may be affected by recorder skill effects for a number of reasons. Firstly, a less skilled botanist may collect insufficient voucher material and/or field notes. Secondly, they may mistake a new species in the field for a species already collected. Thirdly, the identification of sterile specimens is often not possible with conventional keys and instead requires a high level of familiarity with the regional flora. Fourthly, a less well trained botanist may be more hesitant to identify a specimen as a new species, regional record or rarity. This requires a high level of taxonomic expertise and associated confidence. Finally, declining resources for herbaria mean that the time spent by professional taxonomists helping with the identifications is scarce. Frequently, many years pass between the collection of a specimen and its final verification, and unless databases are stringently updated the records will not be corrected. The botanists' resources and training also had a strong effect on the quality of the data they supplied, which can lead to cascades of errors in ecological research (Bortolus 2008). Analyses of the unchecked dataset would have operated with nearly as many imaginary species as actual species, i.e. would be entirely flawed. Alarmingly, the quality of the plant data collected in the EAM has severely declined since the 1980s (Fig. S3.1).

Good field data, particularly for the tropics, are rare. Regionally focussed distribution models can provide surrogates for full-coverage biodiversity inventories; however such estimates remain biased by the underlying species data (e.g. Platts *et al*. 2010) In this respect, observers with minimal training can make extremely useful contributions by increasing data volumes and mitigating geographic biases (Hopkins & Freckleton 2002; Basset *et al.* 2004; Abadie *et al.* 2008). In the EAM and also elsewhere (Lovell *et al.* 2009; Schmeller *et al.* 2009), such observers frequently contribute the highest volumes of data and cover the largest number of sites, because employment costs are lower and professionals frequently collect for herbaria whose collection volumes are limited by available cupboard space and 'interest', meaning that common and widespread species tend not to be collected, and, often, not even recorded. For example, only 12 out of the 3,450 species recorded in the Missouri Botanical Garden's database TROPICOS for the EAM have 50 or more records, the number generally considered necessary for deriving species distribution models (Coudun & Gegout 2006). It is largely thanks to the contributions of large-scale volunteer or semi-professional data collections (e.g. Frontier Tanzania 2007) that such statistical tools can be used in conservation planning, and that strong ease of access biases are increasingly mitigated. The increasing number of ecotourism holidays (Cousins 2007) can thus make valuable contributions to research and conservation of particular sites (Haag 2005). Our analyses suggest that the most efficient strategy is to employ botanists with an intermediate level of training as they find more species and species of conservation concern per funding and time unit. Expert assessments are more funding and time intensive due to greater thoroughness employed. However, the results also show that rare species are most reliably assessed with a high level of training (more than six months), and this is where collaborations between volunteer-based/semi-professional and professional collections may be particularly fruitful: while volunteer or semi-professional collectors could be tasked with rapid assessments that aim to increase the data volume on readily identifiable species, experts could focus on assessments used for conservation planning.

Finally, the increasing pressure to define species' ranges accurately and to predict their future distribution in the face of rapid global environmental change (Parmesan & Yohe 2003) calls for thorough and unbiased biodiversity inventories. Museums and botanical gardens have a major role to play in this endeavour (Primack & Miller-Rushing 2009) and in the training of field biologists. Greater investment in the training of botanists, we think, would pay dividends due to increased recording efficiency and reliability of data for conservation assessments, and reduced time for data cleaning. Reducing support for taxonomy and field biology means that we risk losing species simply because our data are not good enough.

Acknowledgements

We thank A. Balmford, S.A. Bhagwat, C. Ellis, M. Gibby, D. Harris, P.M. Hollingsworth, R.T. Pennington, R. Polhill, D. Raffaelli, J. Ratter for critical discussions. Funding for A.A. was provided by a Marie Curie Actions grant to R.M., for M.T.B. by NERC and for A.R.M. by NERC, NGS and Margot Marsh Biodiversity Foundation. The authors gratefully acknowledge contributions of plant data by Frontier Tanzania.

References

Abadie J.C., Andrade C., Machon N. & Porcher E. (2008). On the use of parataxonomy in biodiversity monitoring: a case study on wild flora. *Biodiversity and Conservation*, 17, 3485-3500.

African Flowering Plants Database (2008). Conservatoire et Jardin botaniques de la Villle de Genève and South African National Biodiversity Institute Pretoria. http://www.ville-ge.ch/musinfo/bd/cjb/africa/.

Basset Y., Novotny V., Miller S.E., Weiblen G.D., Missa O. & Stewart A.J.A. (2004). Conservation and biological monitoring of tropical forests: the role of parataxonomists. *Journal of Applied Ecology*, 41, 163-174.

Beentje H. & Smith S. (2001). FTEA and after. *Systematics and Geography of Plants*, 71, 265-290.

Bortolus A. (2008). Error cascades in the biological sciences: The unwanted consequences of using bad taxonomy in ecology. *Ambio*, 37, 114-118.

Brandon A., Spyreas G., Molano-Flores B., Carroll C. & Ellis J. (2003). Can volunteers provide reliable data for forest vegetation surveys? *Natural Areas Journal*, 23, 254-261.

Brightsmith D.J. (2008). Rainforest expeditions and Earthwatch as funding partners for macaw (Ara spp.) research in southeastern Peru. *Ornitologica Neotropical*, 19, 173-181.

Brook B.W., Sodhi N.S. & Bradshaw C.J.A. (2008). Synergies among extinction drivers under global change. *Trends in Ecology & Evolution*, 23, 453-460.

Burgess N.D., Butynski T.M., Cordeiro N.J., Doggart N.H., Fjeldså J., Howell K.M., Kilahama F.B., Loader S.P., Lovett J.C., Mbilinyi B., Menegon M., Moyer D.C., Nashanda E., Perkin A., Rovero F., Stanley W.T. & Stuart S.N. (2007). The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, 134, 209-231.

Chevan A. & Sutherland M. (1991). Hierarchical Partitioning. *American Statistician*, 45, 90-96.

Coudun C. & Gegout J.C. (2006). The derivation of species response curves with Gaussian logistic regression is sensitive to sampling intensity and curve characteristics. *Ecological Modelling*, 199, 164-175.

Cousins J.A. (2007). The role of UK-based conservation tourism operators. *Tourism and Management*, 28, 1020-1030.

Da Fonseca G.A.B., Balmford A., Bibby C., Boitani L., Corsi F., Brooks T., Gascon C., Olivieri S., Mittermeier R.A., Burgess N., Dinerstein E., Olson D., Hannah L., Lovett

J., Moyer D., Rahbek C., Stuart S. & Williams P. (2000). It's time to work together and stop duplicating conservation efforts ... following Africa's lead in setting priorities. *Nature*, 405, 393-394.

Darwall W.R.T. & Dulvy N.K. (1996). An evaluation of the suitability of non-specialist volunteer researchers for coral reef fish surveys. Mafia Island, Tanzania - A case study. *Biological Conservation*, 78, 223-231.

Disney R.H.L. (1989). Does Anyone Care. *Conservation Biology*, 3, 414-414.

Eastern Africa Plant Red List Authority (2008). List of Potentially Threatened Plant Taxa in the Eastern Arc Mountains of Tanzania. Unpublished report presented to the Forestry and Beekeeping Division, Tanzania Ministry of Natural Resources and Tourism. (ed. Africa Plant Red List Authority) Dar es Salaam.

Ehrenfeld D. (1989). Is Anyone Listening. *Conservation Biology*, 3, 415-415.

Feldmann R.M. & Manning R.B. (1992). Crisis in Systematic Biology in the Age of Biodiversity. *Journal of Paleontology*, 66, 157-158.

Fitzpatrick M.C., Preisser E.L., Ellison A.M. & Elkinton J.S. (2009). Observer bias and the detection of low-density populations. *Ecological Applications*, 19, 1673-1679.

Frontier Tanzania (2007). Biodiversity Research and Awareness in the lesser-known Eastern Arc Mountains (BREAM): Mahenge Mountains Report, Ulanga District. (eds. Owen N, Wilkins V, Fanning E & Howell KM). Society for Environmental Exploration, University of Dar es Salaam, Critical Ecosystems Partnership Fund.

Gaston K.J. & May R.M. (1992). Taxonomy of Taxonomists. *Nature*, 356, 281-282.

Gee H. (1992). UK Systematic Biology - Peers slam peer-review. *Nature*, 355, 488-488.

Gower J.C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27 857-871.

Haag A. (2005). A trip of a lifetime. *Nature*, 435, 1018-1020.

Hopkins G.W. & Freckleton R.P. (2002). Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Animal Conservation*, 5, 245-249.

James A.N., Gaston K.J. & Balmford A. (1999). Balancing the Earth's accounts. *Nature*, 401, 323-324.

Lovell S., Hamer M., Slotow R. & Herbert D. (2009). An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. *Biodiversity and Conservation*, 18, 3295-3307.

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

McCune B., Dey J.P., Peck J.E., Cassell D., Heiman K., WillWolf S. & Neitlich P.N. (1997). Repeatability of community data: Species richness versus gradient scores in large-scale lichen studies. *Bryologist*, 100, 40-46.

Minchin P.R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, 69, 89-107.

Mittermeier R.A., Robles-Gil P., Hoffmann M., Pilgrim J.D., Brooks T.M., Mittermeier G.C., Lamoreux J.L. & Fonseca, G. (2005). *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Ecoregions*. 2 edn, Cemex, Mexico City.

Newmark W. (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, 29-36.

Newmark W.D. (2002). Ecological Studies. Conserving biodiversity in East African forests: A study of the Eastern Arc Mountains. In: *Ecological Studies. Conserving biodiversity in East African forests: A study of the Eastern Arc Mountains*. Springer-Verlag, New York.

Noss R.F. (1996). The naturalists are dying off. *Conservation Biology*, 10, 1-3.

Novacek M.J. & Cleland E.E. (2001). The current biodiversity extinction event: Scenarios for mitigation and recovery. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 5466-5470.

Oksanen.J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2010). vegan: Community Ecology Package. R package version 1.17-2. Available at: http://CRAN.R-project.org/package=vegan.

Parmesan C. & Yohe G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37-42.

Pimm S.L. & Lawton J.H. (1998). Ecology - Planning for biodiversity. *Science*, 279, 2068-2069.

Pinheiro J., Bates D., DebRoy S., Sarkar D. & the R Core team (2009). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-96. Available at: http://CRAN.R-project.org/package=nlme.

Platts, P.J., Ahrends, A., Gereau, R.E., McClean, C., Lovett, J.C., Marshall, A., Pellikka, P., Mulligan, M., Fanning, E., Marchant, R. (2010). Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. *Diversity and Distributions* 16, 628-642.

Prance G.T., Beentje H., Dransfield J. & Johns R. (2000). The tropical flora remains undercollected. *Annals of the Missouri Botanical Garden*, 87, 67-71.

Primack R.B. & Miller-Rushing A.J. (2009). The role of botanical gardens in climate change research. *New Phytologist*, 182, 303-313.

Quinn Q.P. & Keough M.J. (2002). *Experimental design and data analysis for biologists*. Cambridge University Press, Cambridge.

R Development Core Team (2009). *R: A language and environment for statistical computing*, Vienna. Available at: www.R-project.org.

Rivas J.A. (1997). Natural history: Hobby or science? *Conservation Biology*, 11, 811-812.

Samyn Y. & Massin C. (2002). Taxonomists' requiem? *Science*, 295, 276-277.

Schlüter T. (1997). *Geology of East Africa*. Borntraeger, Berlin.

Schmeller D.S., Henry P.Y., Julliard R., Gruber B., Clobert J., Dziock F., Lengyel S., Nowicki P., Deri E., Budrys E., Kull T., Tali K., Bauch B., Settele J., Van Swaay C., Kobler A., Babij V., Papastergiadou E. & Henle K. (2009). Advantages of Volunteer-Based Biodiversity Monitoring in Europe. *Conservation Biology*, 23, 307-316.

Scott W.A. & Hallam C.J. (2003). Assessing species misidentification rates through quality assurance of vegetation monitoring. *Plant Ecology*, 165, 101-115.

Stork N.E. (2010). Re-assessing current extinction rates. *Biodiversity and Conservation*, 19, 357-371.

Valdecasas A.G., Castroviejo S. & Marcus L.F. (2000). Reliance on the citation index undermines the study of biodiversity. *Nature*, 403, 698-698.

Vitousek P.M., Mooney H.A., Lubchenco J. & Melillo J.M. (1997). Human domination of Earth's ecosystems. *Science*, 277, 494-499.

Walsh C. & Mac Nally R. (2008). hier.part: Hierarchical Partitioning. R package version 1.0-3. Available at: http://CRAN.R-project.org/package=hier.part.

Wheeler Q.D., Raven P.H. & Wilson E.O. (2004). Taxonomy: Impediment or expedient? *Science*, 303, 285-285.

Whitehead P. (1990). Systematics - an Endangered Species. *Systematic Zoology*, 39, 179-184.

Wilson E.O. (2000). The future of conservation biology (vol. 14, pg. 1, 2000). *Conservation Biology*, 14, 920-920.

Zuur A.F., Ieno E.N. & Smith G.M. (2007). *Analysing Ecological Data*. Springer, New York.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. (2009). *Mixed Effects Models and Extension in Ecology with R*. Springer, New York.

Supporting information

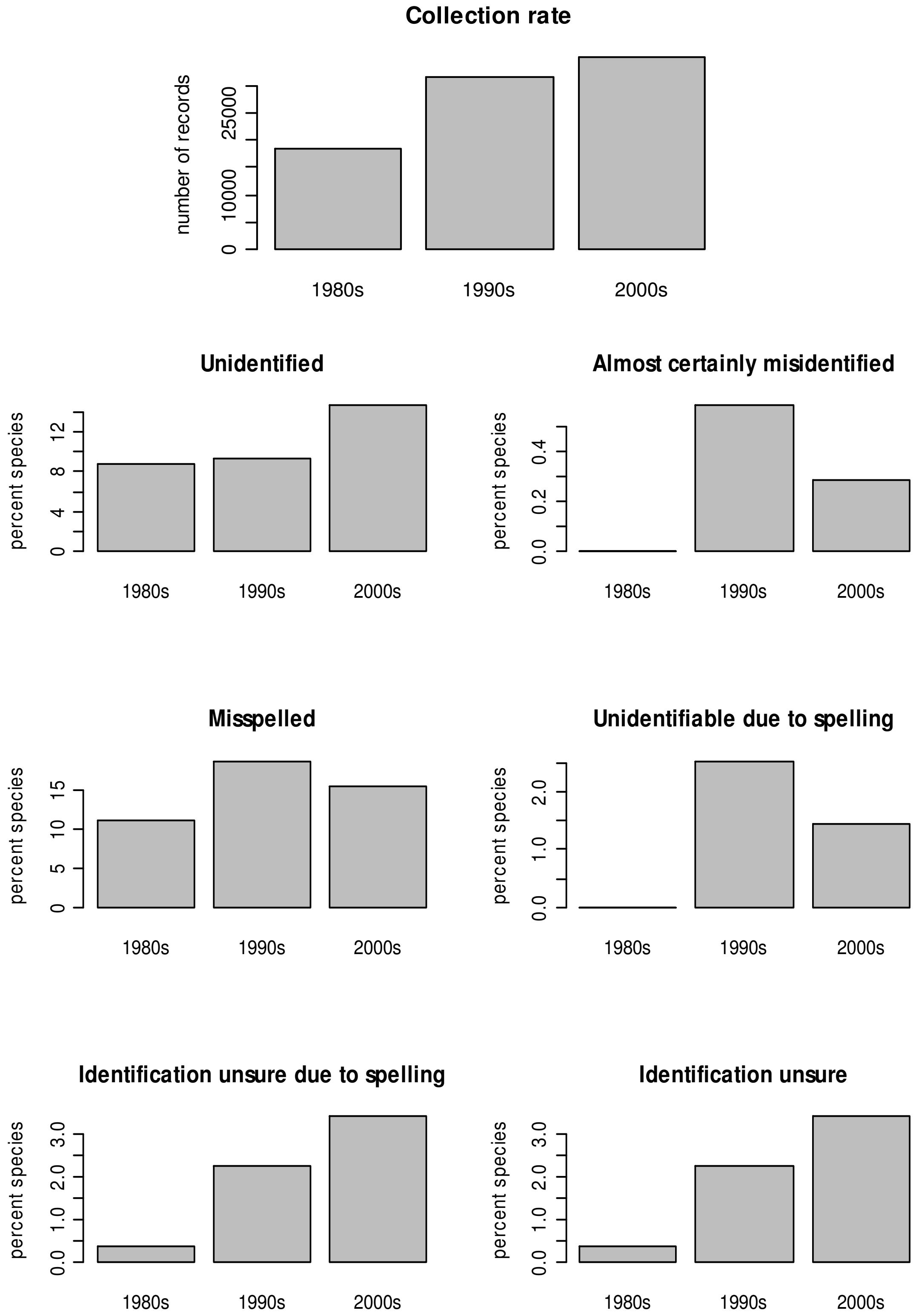


Figure S3.1. Quality of the plant record data collected in the Eastern Arc Mountains over time (based on the 13 botanists in the study).

Variable	**Measure**
Percentage of unidentified species	Unidentified species are only identified to genus, family or not at all. Misidentifications have also been categorised as unidentified.
Percentage of species with uncertain identification	Species with an uncertain identification were marked as uncertain by botanist him-or herself.
Percentage of almost certainly misidentified species	Almost certainly misidentified species were recorded way out of their recognised distribution range (different continent, or West or South Africa) and have not been introduced to East Africa.
Percentage of misspelt species	Misspellings also include sloppiness such as misplaced hyphens and spaces, misspelled capitals, use of abbreviations, etc.
Percentage species with an uncertain identification due to spelling mistakes	Percentage of species that have been so severely misspelt that we were uncertain whether we had interpreted correctly what the botanist had meant to enter.
Percentage of unrecognisable species due to spelling mistakes and sloppiness	Percentage of species that have been so severely misspelt that the name was entirely unrecognisable. This category also includes species names that have been confused (e.g. a specific epithet entered from a different, similarly sounding genus).

Table S3.1. Measurements for the quality of the collected data.

		Pearson
Altitude range	Herbarium staff	0.70
Herbarium access	**PhD**	1.00
Number of days in field	Minimum altitude	-0.71
Number of person-days	Minimum altitude	-0.75
Number of person-days	Number of days in field	0.91
Number of plots	**Number of records**	0.78

Table S3.2. Predictors with levels of correlation ≥ 0.7. Predictors chosen for further analysis (based on their predictive power when fitted on their own) are in bold.

Link from Chapter 3 to Chapter 4

The previous result chapters have shown that there are severe biases in the plant data that underpin conservation decisions in the study area. There consequently is an urgent need for rapid and efficient assessments to mitigate these biases. This is not only true in the Eastern Arc Mountains but in the entire tropical forest biome, given the rapid loss of tropical forest biodiversity and its relatively poor collection. However, despite decades of research field ecologists still agonise over the most efficient sampling protocol for plant species richness assessments. In the Eastern Arc Mountains and coastal forests of Tanzania at least a dozen different sampling protocols have been used, and there is a lack of clarity whether the data are comparable and whether some of the less intensive methods have been sufficiently representative to capture the true pattern. There has been a range of studies in tropical forests that have compared the efficiency of different field protocols, however they are partly contradictory. Some of this confusion may have arisen as such studies typically use field data, where individual abundance distribution, density and patchiness are highly variable. Chapter 4 develops a simulation approach to identify the most efficient sampling protocol for a variety of field conditions, whereby potentially confounding factors are controlled for and their respective influence assessed.

The Eastern Arc Mountains range up 2,635 m in altitude at the Lukwangule Plateau in the Uluguru Mountains (here upper montane forest at 2,600 m in Uluguru South).

Chapter 4 - Optimising sampling protocols for plant species richness assessments in tropical forests

To be submitted to *Ecology Letters* (Letter)

Optimising sampling protocols for plant species richness assessments in tropical forests

Antje Ahrends[1,2], Mark T. Bulling[3], Philip J. Platts[2], Rob Marchant[2], Neil D. Burgess[4,5,6], Jon C. Lovett[7] and Carsten Rahbek[6]

[1] Botanic Garden Edinburgh, 20A Inverleith Row, EH3 5LR, UK; [2] KITE, Environment Department, University of York, YO105DD, UK; [3] Oceanlab, University of Aberdeen, Newburgh, Aberdeenshire, AB416AA, UK; [4] Zoology Department, University of Cambridge, CB2 3EJ, UK; [5] WWF US, 1250 24th St NW, Washington DC, USA; [6] Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Universitetsparken 15, DK- 2100 Copenhagen Ø, Denmark; [7] Twente Centre for Studies in Technology and Sustainable Development, University of Twente, P.O. Box 217, 7500 Enschede, The Netherlands

Abstract

Plant species richness is at the heart of conservation prioritisation and fundamental ecological studies. Despite decades of research, there is little consensus on the drivers of plant species richness patterns and optimal sampling strategies, particularly in tropical forests. By taking a simulation approach, we analysed the factors that may influence sampling performance. We generated a simple pattern whereby species richness peaks in the middle of an environmental gradient and simulated different plot based assessment strategies, with parameters reflecting typical tropical forest tree values. We show that sample size is of extreme importance, with other factors such as patchiness and plot design only mattering at low sample sizes. Even in the most optimistic scenarios a minimum sample size of 1,000 to 2,000 m^2 was needed to capture the species richness pattern. This is not always achieved and insufficient sampling may partly explain persistent disagreements over species richness patterns along environmental gradients such as altitude.

Introduction

Measuring plant species richness accurately is of fundamental importance for biodiversity conservation and ecology. For example, floristic diversity is an important criterion for many global and regional schemes that identify areas of conservation priority (Davis *et al.* 1997; Brooks *et al.* 2006). This is because vascular plants are amongst the better known taxonomic groups, and their use as an indicator of diversity in other groups of terrestrial organism is being tested for systematic conservation planning (Bladt *et al.* 2008; Larsen *et al.* 2009).

Plant species richness is also at the heart of studies on general ecological principles, such as on the much debated drivers of the latitudinal and altitudinal gradient in species richness (Rahbek 1995; O'Brien *et al.* 2000; Whittaker *et al.* 2001; Lavers & Field 2006). Despite decades of research, no consensus has yet been reached on the historic and current environmental factors that are driving species richness patterns (Rahbek *et al.* 2007; Gotelli *et al.* 2009). Understanding these fundamental biogeographic patterns and the biases that may influence our perception of them is key to systematic conservation planning, which aims to adequately represent biodiversity in protected areas and to conserve the processes that generated and sustain it (Margules & Pressey 2000).

It is therefore of concern that there is little consensus on how plant species richness should be assessed in the field, particularly in the notoriously understudied tropical forests (Prance *et al.* 2000). Vegetation plots are probably the most commonly used assessment approach. Numerous different plot methods have been proposed (Gentry 1982; Shmida 1984; Austin & Heyligers 1989; Hall 1991; Dallmeier 1992; Stohlgren *et al.* 1995; Campbell *et al.* 2002; Sheil *et al.* 2003), and many studies have compared them in terms of their efficiency and ability to capture the true plant species richness pattern (Hall 1991; Stern 1998; Wessels *et al.* 1998; Campbell *et al.* 2002; Phillips *et al.* 2003; Gordon & Newton 2006a; Archaux *et al.* 2007). However, the conclusions of these studies lack consistency. The use of diversity indices introduces an additional variable that affects the efficiency of different assessment methods (Condit *et al.* 1996; Gimaret-Carpentier *et al.* 1998). Over a decade ago, Stohlgren *et al.* (1995) remarked that despite a lot of research into the subject, "*field ecologists still agonise over selecting the appropriate sampling protocol*". We believe that, albeit progress has been made, this is still true.

The consequence of the lack of clarity over the most efficient sampling protocol is that a wide range of different methods are employed. Thus many studies are rendered incompatible with each other, and there is a general increase in uncertainty over data reliability as the employed sampling strategies may not only be inefficient, but insufficiently representative to capture the true species richness pattern (Gordon & Newton 2006b). As with the effect of different scales being employed in different studies (Rahbek 2005; Nogues-Bravo *et al.* 2008), differences in sampling protocol may well be a cause for the wide range of proposed species richness patterns.

With very few exceptions (e.g. Austin & Adomeit 1991), comparative studies of the efficiency of sampling protocols have been based on real field data. This may partly explain why a consensus regarding the most efficient sampling protocol has not been reached as, unless the area has been sampled exhaustively, the true richness patterns remain unknown (Austin et al. 2006). Further, results from studies based on field data are likely to be confounded by differing stem densities (Gotelli & Colwell 2001), the degree of patchiness in the spatial distribution of individuals, the strength of the gradient (= turnover in species richness between the localities sampled), noise, and the underlying species abundance distribution. Higgins & Ruokolainen (2004) showed that sampling strategies are rated differently when the delineation of the sample community is changed while sampling protocols are kept constant. In comparison, ecologically meaningful simulated datasets allow for the controlled manipulation of confounding factors and thus a systematic investigation of sampling strategy efficiency and potential biases.

The aims of this paper are (1) to examine the relative importance of factors that influence sampling efficiency and (2) to offer guidelines to ecologists for choosing the most appropriate sampling protocol and intensity for their study. Creating an artificial landscape and species richness patterns, we assess the impact of species aggregation, the strength of the richness gradient, noise and species abundance distribution on the performance of different sampling strategies, and establish the minimum sample size needed to capture the true species richness pattern. The parameterisation of our simulation focuses on tropical forests because in these systems efficient sampling design is particularly important: tropical forests are relatively challenging to inventory; their levels of biodiversity are high; they are often poorly conserved; and systematic conservation planning is hampered by a poor understanding of biodiversity patterns (Phillips *et al.* 2003; Gordon & Newton 2006a). In our simulations, species richness is driven by a single environmental factor. If sampling schemes fail to detect the true gradient in such simplified data, then they are unlikely to be sufficient for real data, where there are likely to be multiple interacting drivers, i.e. more complex gradients need to be captured.

Materials and methods

Overview

We simulated an artificial landscape where an environmental gradient causes species richness to have different values in different areas. This, for example, could be a mountain ranging from 0 - 2,500 m above sea level (a.s.l.) with ten different elevation bands (each 250 m). Each elevation band contains 10,000 trees in an area of 25 ha. In order to keep our conclusions general, we did not vary stem density or area size with altitude. Species richness was set to peak in the two middle elevation bands (between 1,000 and 1,500 m a.s.l.) and to gradually decline above and below that altitude. This pattern in species richness is often attributed to the so called ‘Mid-Domain-effect’ that arguably arises as a consequence of geometric constraints and two hard boundaries (Colwell *et al.* 2004). The choice of model did not reflect favouritism for any of the four main diversity models on altitudinal variation in species richness (Nogues-Bravo *et al.* 2008). Instead one of the simplest models was selected in order to produce an estimate for the absolute minimum sample size needed to pick up the pattern. We sampled each elevation band, trying to find the species richness pattern under a total of 24 different field conditions (Fig. 4.1). The parameterisation of individual density, species richness and the field conditions reflected typical values for tropical forest trees. Using a variety of different sampling techniques, we sampled the gradient 576 times (2,880 times with repetitions), testing all combinations of field conditions and sampling types. Each time, we measured the efficiency of the sample protocol in terms of how closely it captured the true species richness pattern, the proportion of species found, and expended effort.

Simulated field conditions

Species aggregation. We simulated two different spatial scenarios:

1. Complete spatial randomness (CSR). Individuals of each tree species were distributed across space completely randomly.
2. Aggregated. The individuals of each tree species were positioned relative to each other according to an aggregated distribution. This pattern corresponds most closely to realistic field conditions, particularly in tropical forests where nearly all tree species display an aggregated pattern throughout all size classes (Condit *et al.* 2000).

Each elevation band in the simulated landscape measured 500 × 500 m. The chosen tree density (400 trees ha^{-1}) is the lower quartile of tree density values ($\geq$ 10 cm diameter at reference height (drh; 1.3 m up the stem or above buttresses) that we extracted from studies across Africa, Asia and South and Central America in a total of 45 sites (Leigh 1975; Emmons & Gentry 1983; Condit *et al.* 2000; Ahrends & Marchant 2006-2008; La Torre-Cuadros *et al.* 2007; Wittmann *et al.* 2008; Top *et al.* 2009) (see Supporting information for Chapter 4 Fig. S4.1). Choosing a typical but not overly high value for tree density allowed a cautious estimation of the minimum adequate sample size. The spatial distribution of trees (point pattern) was programmed using the following procedure: For each species, a number of ‘parent’ trees were randomly placed within the area. These produced ‘offspring’ which were placed at random locations within a certain radius from the parent, causing an aggregated pattern. The radius around the parent plant for placing offspring was set to 50 m, which is the radius within which aggregation tends to be highest in tropical forests (Condit *et* *al.* 2000). The degree of aggregation was manipulated by varying the probability of an individual being a parent tree. For the complete spatial random pattern this was set to 1, and for the aggregated pattern to 0.2. If the offspring of a parent plant located close to the border fell outside the forest area, we repeatedly randomly positioned the offspring within the parent radius until it was within in the forest area boundary. This resulted in a slightly stronger aggregation of individuals close to the border in the aggregated pattern - a bias that was however smaller than biases due to other methods of edge correction: not allowing parent plants to be located within a distance of less than 50 m from the area edge, ensuring no offspring fall outside the study area would have meant that tree density at the border would have been consistently lower for the CSR pattern; if the area would have been envisioned as a sphere with offspring falling outside being assigned to a location at the opposite end of the area, the aggregation of the tree species would have been distorted. Another option would have been not to correct for the edge effect (as it is a natural property of these systems) and instead disregard offspring outside the area, however, this would have compromised comparability between the simulations as the total number of individuals would have been inconsistent, with potentially strong biases occurring with rarer species.

Strength of gradient. Tree species richness can vary enormously across different tropical forest types and floristic realms (Gentry 1988; Condit *et al.* 2000); for example, in 50 ha plots from 63 to > 800 species ≥ 10 cm drh (Condit *et al.* 2000). In our simulations, we set tree species richness and the strength of the gradient to realistic but not extreme values, allowing us to establish the absolute minimum sampling requirements. We simulated two different gradient strengths:

1. Strong. Species numbers ranged from 100 species at the lowest and highest elevation band to 300 species at mid elevation, and there was a constant turnover of 50 species in species richness between adjacent elevation bands.
2. Weak. Species numbers ranged from 100 species at the lowest and highest elevation band to 140 species at mid elevation with a constant turnover of 10 species in species richness between adjacent elevation bands.

Noise. We simulated two different noise scenarios:

1. No noise.
2. Noise. Noise was generated by randomly sampling from a normal distribution ($\mu = 0$ and $\sigma^2 = 0.5 \times$ species turnover between elevation bands). Following conversion to the nearest integer, this number was added to the number of species present at the particular elevation. This produced a reasonable level of noise in the system scaled to the particular

species turnover rate: random reversals in the biodiversity values at different elevations occurred but the fundamental pattern remained recognisable.

Species abundance distribution. Three of the most common species abundance models for ecological communities were used:

1. Geometric.
2. Log normal.
3. MacArthur's broken stick.

These represent a progression from a model where species are as evenly distributed as realistically still possible in natural communities (MacArthur's broken stick model), through the log normal series to a model where a few species dominate (geometric series) (Magurran 1988).

All the field conditions were selected to cover large ranges in realistic values, which provided information on the potential range in influence that each field condition may have on the efficiency of a sampling regime. This allowed us to establish the minimum acceptable sampling intensity required to counter-balance the negative effects of each aspect of the field conditions on sampling efficiency.

We conducted two further simulation analyses (Table S4.1): (1) sampling in square instead of rectangular plots with all the other simulation parameters as above; (2) sampling all individuals ≥ 2.5 cm drh (instead of ≥ 10 cm drh), whereby all the simulations were as above with the exception of the individual density which was adjusted to reflect a typical density for this size group (2,000 individuals ha^{-1}) (Ahrends, unpublished data).

Sampling

Trees were sampled in plots, i.e. the sampling was restricted to a few locations at each elevation. We simulated four commonly used sampling techniques (= plot dimensions) and six different sampling intensities (= number of plots) (Table 4.1). The plots were aligned north-south, randomly located with the constraint that they were not allowed to overlap. Five repeat simulations of sampling the complete altitudinal gradient were conducted under each permutation of field conditions, sampling technique and sampling intensity. Sampling performance was defined as the ability of the sampling protocol to capture the true species richness pattern, measured using Spearman rank correlations between the sampled pattern and the true pattern. In addition, we assessed the proportion of species found by the sampling protocol. Expended effort was measured as the total plot area sampled, allowing us to calculate two metrics of sampling efficiency: (1) sampling performance / expended effort, (2) proportion of species captured / expended effort.

To investigate the influence of the field conditions on sampling performance, we fitted Boosted Regression Trees (BRT) (Elith *et al.* 2008) to the dependent variables (sampling performance and proportion of species found) and to the candidate predictors (field conditions and sampling protocols). BRT combine the advantages of tree-based methods (automatic fitting of interactions and curvilinear relationships, insensitivity to outliers, and accommodation of diverse variable types) with the increased predictive performance of boosting – a technique that optimises model accuracy by iteratively fitting and combining a large number of decision trees (typically > 1,000; Elith *et al.* 2008). In order to establish optimal model settings (tree complexity (= number of possible interactions), learning rate (= degree of shrinkage of the contribution of each tree), level of stochasticity (= proportion of data of the full training dataset used at each iterative step) and number of trees), we systematically altered these settings and evaluated model accuracy using (1) 10-fold cross-validation and (2) evaluation of the models' predictive performance by splitting the data into a training and an evaluation dataset and calculating the deviance between predicted and actual values (Elith et al. 2008).

All simulations and analyses were conducted using R 2.8.1 (R Development Core Team 2008), the spatstat (Baddeley & Turner 2005) and gbm (Ridgeway 2007) libraries, and the BRT function source code provided by Elith *et al.* (2008).

	Plot dimensions	**Total number of plots**						**Plot dimensions originated in**
Rectangular plots	2×50 m	10	20	50	100	200	500	(Gentry 1982)
	10×50 m	2	4	10	20	40	100	
	20×50 m	1	2	5	10	20	50	(Stohlgren *et al.* 1995)
	100×100 m	NA	NA	NA	1	2	5	(Dallmeier 1992)
Additional square dimensions tested	10×10 m	10	20	50	100	200	500	
	~22×22 m	2	4	10	20	40	100	
	~32×32 m	1	2	5	10	20	50	
Total area covered by plots (m^2)		1,000	2,000	5,000	10,000	20,000	50,000	

Table 4.1. Summary of all plot sampling protocols tested.

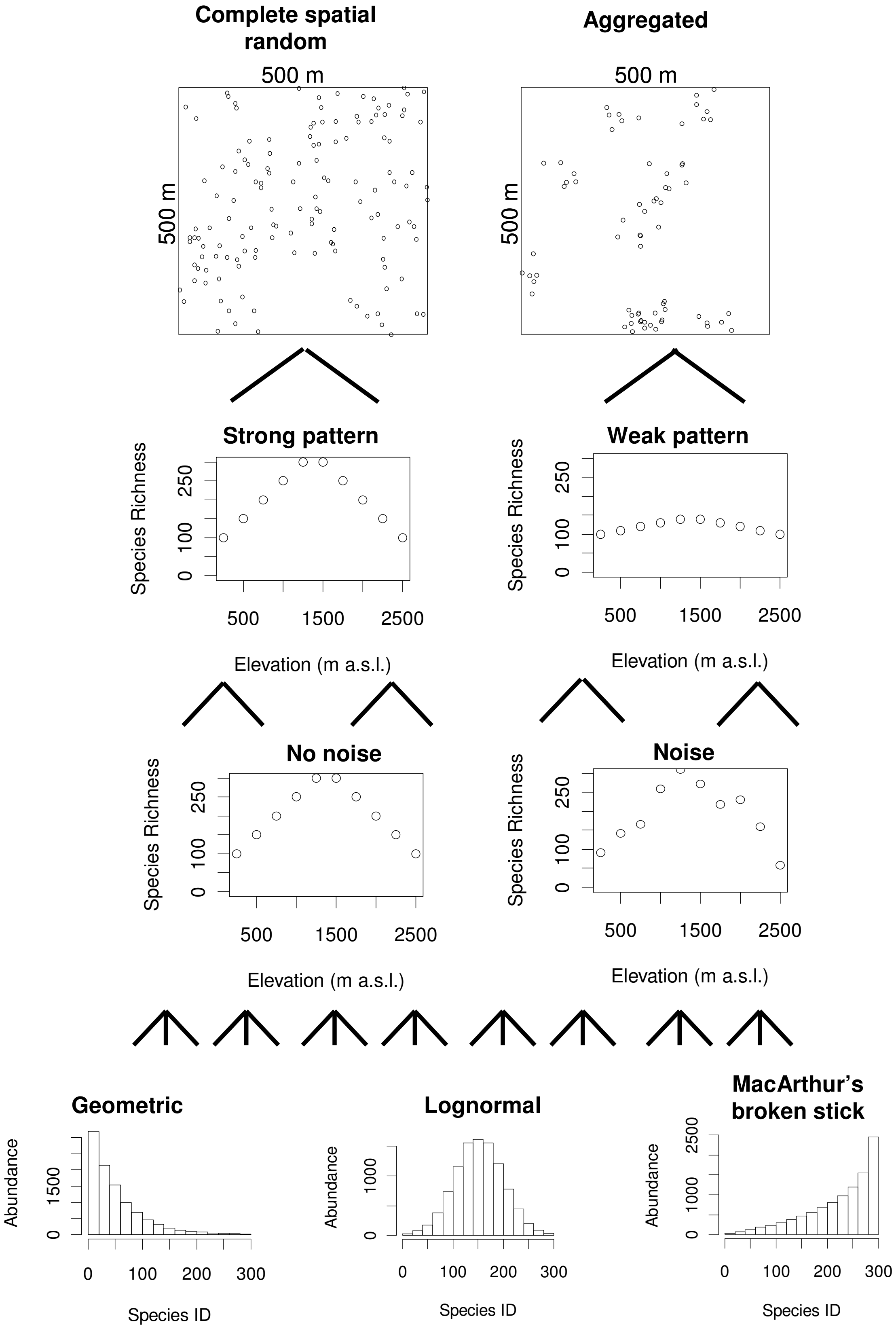
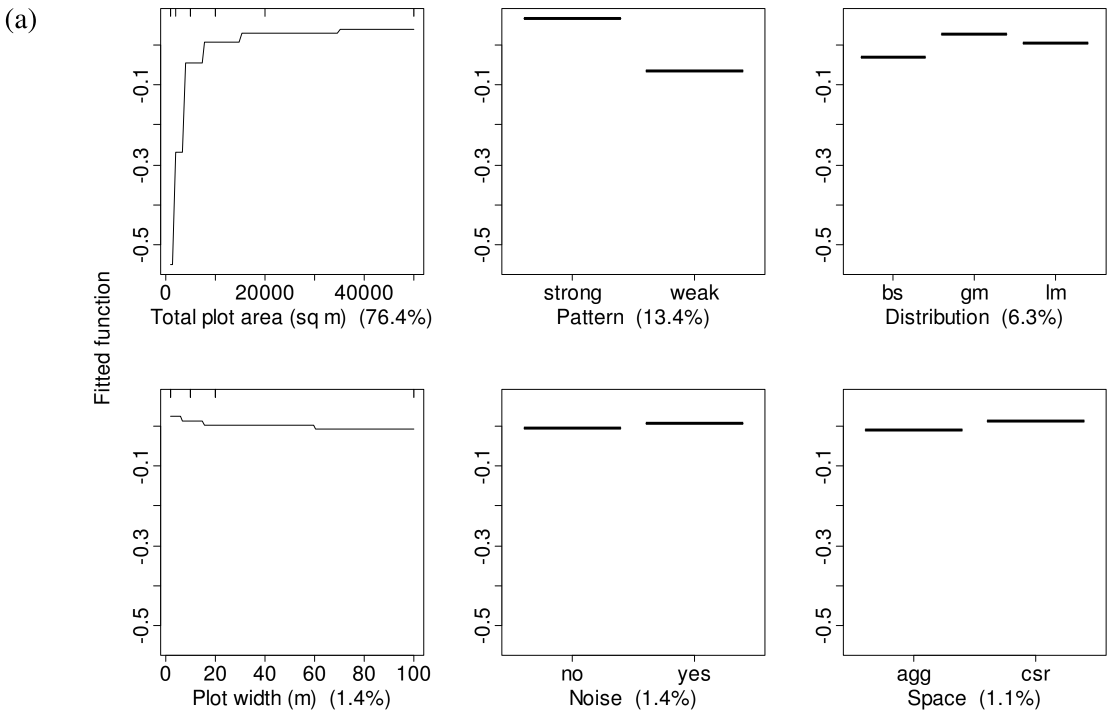


Figure 4.1. Overview of the 24 field scenarios.

Results

Identifying the true species richness pattern

A sufficient total sampling area was critical for the true species richness pattern to be determined (Fig. 4.2). Total plot area was by far the most important predictor, explaining 76% of the variation in the correlation between the sampled and the true species richness pattern (Fig. 4.2a). Sampling performance rose sharply with increasing total plot area, levelling out once the total sampled area was greater than 5,000 m^2. Figure 4.3 illustrates the effect of sampling size; the estimated species richness did not change with altitude at the smallest sampling size and the strength of the pattern only became apparent when at least one ha was sampled. Another important predictor was the strength of the true richness pattern, the strong pattern being detected much more readily than the weak one. Each of the remaining predictors (species abundance distribution, plot shape, noise and spatial aggregation of individuals, in decreasing order of influence) explained less than 10% of the variation. The more even the species abundance distribution, the weaker the correlation between the estimated and real richness pattern; and the larger the plots (and the smaller their number), the poorer the performance of the sampling. Spatial aggregation of individuals within species decreased the efficiency of the sampling, whilst the presence of noise in the richness pattern had no effect. There were significant interactions between total plot area and all other predictors, and between the strength of the true gradient and the species abundance distribution, the spatial pattern and the plot shape (Fig. 4.2b). However, once sample size was sufficiently large the effect of other confounding factors on the sampling protocol performance was minimal. For instance, a more efficient plot design (smaller plot sizes) minimally increased the fit between the sampled and the true pattern when the overall area sampled was low, but was almost negligible when the total area sampled was large (Fig. 4.2c). Overall, the model performance was good with a mean cross-validation coefficient of CV = 0.86 (S.E. = 0.01).



(b)

	Distribution	Pattern	Space	Noise	Plot width	Total plot area
Distribution	0	0.01	0	0	0	0.39
Pattern	0	0	0.01	0	0.05	1.36
Space	0	0	0	0	0	0.03
Noise	0	0	0	0	0	0.06
Plot width	0	0	0	0	0	0.02
Total plot area	0	0	0	0	0	0

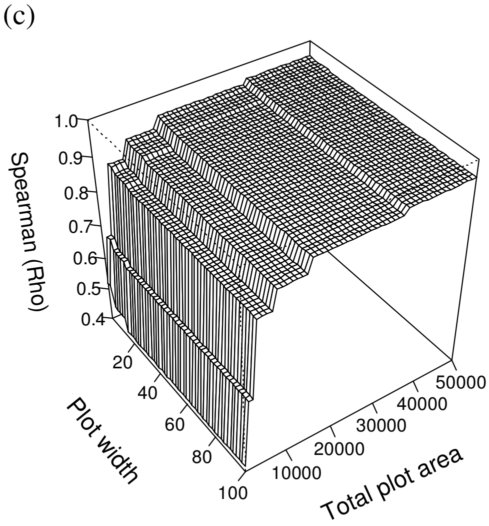


Figure 4.2. Boosted Regression Tree results for the sampling performance at capturing the species richness pattern. (a) Partial effect of each of the tested field conditions on the sampling performance. All other predictors are held constant at their mean. Bracketed values on the x axis detail the relative influence of the predictor variable, which is calculated as the number of times the variable is selected at nodes in the decision tree weighted by the resulting model improvement, averaged over all trees (Elith et al. 2008). Y axes are scaled to have zero mean. Labels for plots corresponding to distribution are bs (broken stick), gm (geometric) and lm (log normal); and for plots corresponding to space csr (complete spatial random) and agg (aggregated). (b) Summary of all pairwise interactions in the model, which are calculated as the residual variance in a linear model that models temporary predictions for combinations of two predictors on the two predictors in question, fitted as factors. Zero residual variance indicates that no interactions are fitted (Elith *et al.* 2008). (c) Interaction between the partial effects of plot design and total area sampled on sampling performance.

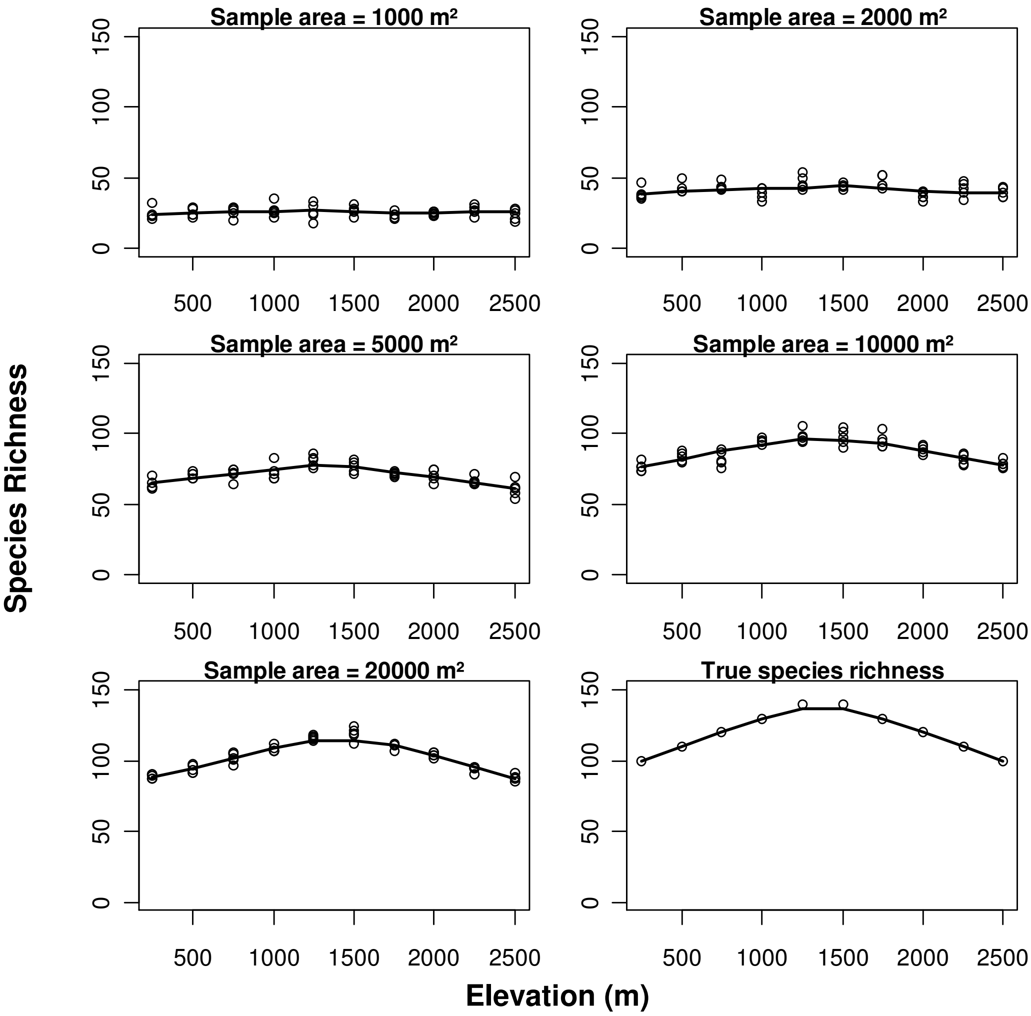
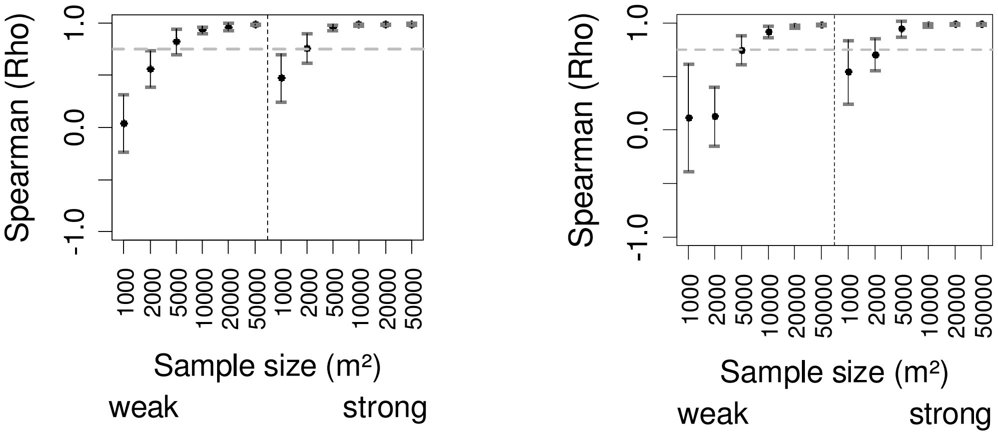


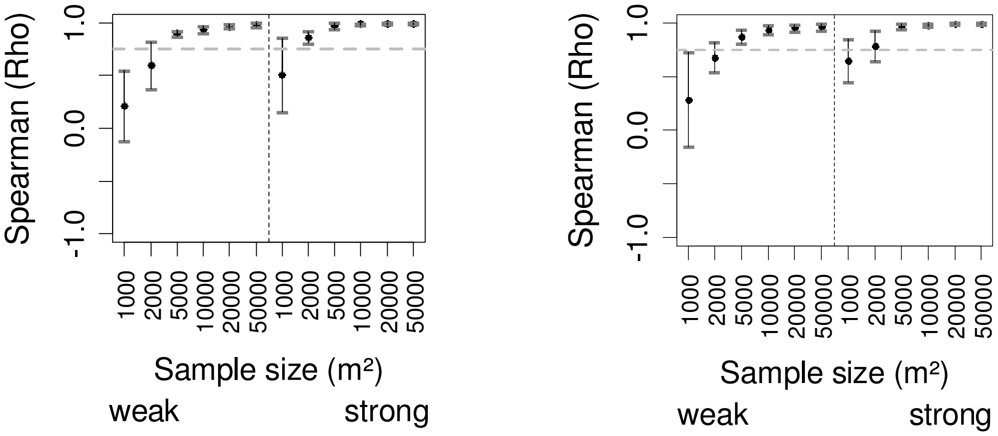
Figure 4.3. Sampled species richness pattern over different sample sizes and true species richness pattern with elevation. The spatial distribution is aggregated, species abundance distribution broken stick and the strength of the gradient weak.

The plot design 2 × 50 m nearly always performed better than larger plots irrespective of the field conditions, although this improvement was sometimes marginal (Fig. S4.2). Using this plot design, we identified the minimum sample size required under the different field conditions to obtain a strong correlation between the estimated and the true species richness gradient (Fig. 4.4). When the strength of the gradient was weak, the minimum needed sample size that allowed a reliable pattern reconstruction under all conditions was 5,000 m² (mean Spearman correlation ≥ 0.75). With a sample size of 2,000 m² pattern estimates were reliable only under good conditions, e.g. geometric species abundance distribution (upper standard deviation (S.D.) Spearman correlation ≥ 0.75). When the gradient was strong, a total sample area of 1,000 m² produced reasonable estimates under good conditions (upper S.D. ≥ 0.75) while 2,000 m² was a more reliable sample size (mean ≥ 0.75). These sample sizes apply for the case where all individuals ≥ 10 cm drh are sampled. When all individuals

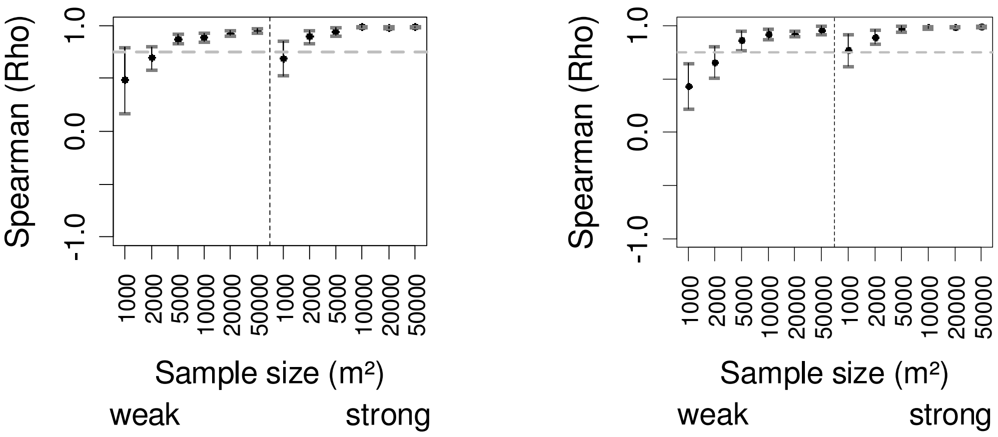
≥ 2.5 cm drh were sampled a sample size of 1,000 m² sufficed under all conditions (Fig. S4.6).



Broken stick CSR **Broken stick aggregated**



Lognormal CSR **Lognormal aggregated**



Geometric CSR **Geometric aggregated**

Figure 4.4. Minimum needed sample sizes for rectangular (2 × 50 m) plots that sample all individuals ≥ 10 cm drh under different field conditions to capture the true species richness pattern (Spearman ≥ 0.75). Black dots are the mean and grey bars the upper and lower standard deviation of the correlation between the sample and the true species richness pattern.

Estimation of the number of species present

The sample size also largely determined the performance of the sampling regimes in terms of capturing the species present. This variable explained 88% of the variation in the proportion of species found. The other predictors – in order of decreasing importance: species abundance distribution, strength of gradient, spatial distribution and plot size – each explained less than 10% of the variation, and noise made no contribution to the model. Even species abundance distributions, weak richness patterns (= less species present), non-aggregated spatial patterns of individuals and higher numbers of repetitions of small plots all contributed to greater proportions of species being found (Figs. S4.1b and S4.2b). The largest interaction existed between plot width and the degree of spatial aggregation of the individuals. When individuals within a species were aggregated, small plots with a large number of repetitions performed considerably better at estimating the number of species than large plots with few repetitions (Fig. 4.2b). However, the effect of plot size was negligible at larger sample sizes (see Fig. S4.3c). Overall, model performance was excellent with an average cross-correlation coefficient of CV = 0.99 (S.E. = 0.00). None of the tested sample protocols captured all species present (Fig. S4.4). With the largest tested sample size of 50,000 m², 84% of the species were found on average, whilst with a sample size of 5,000 m² less than 50% of species were found.

Sampling performance and efficiency

In general, smaller plots performed better than larger plots, and rectangular plots better than square plots at capturing the true species richness pattern (Fig. 4.5). The average efficiency loss (over all field conditions) when using the most inefficient sample method (100 × 100 m plots) compared to the most efficient method (2 × 50 m plots) was 4% (Fig. 4.5). A BRT model tree showed pattern capture efficiency was mostly determined by the plot size (50%), with plot shape accounting for only 1% of the variance (Table S4.2). Plot size played an even greater role in explaining the sampling efficiency in terms of capturing a maximum proportion of species present (64%) while in this case plot shape was close to irrelevant (Table S4.2). The efficiency lost by using the 100 × 100 m plot method instead of sampling in plots of 2 × 50 m was 10% (Fig. 4.5).

The minimum required field efforts were therefore relatively similar between rectangular and square plot methods (compare Figs. 4.4 and S4.5; Table S4.1), and two tailed Student's *t*-tests between the number of species (E(μ)_rectangular = 84.6; E(μ)_square = 84.56) and the overall number of individuals recorded (E(μ)_rectangular = 628.11; E(μ)_square = 627.47) in square and rectangular plots were insignificant ($p > 0.05$).

A BRT model showed that the number of individuals captured did not differ significantly between plot designs (Table S4.2), indicating that this was not the cause of smaller plots performing better than larger plots. We find that what is most important is the way in which the area is covered by plots, i.e. how many different locations are sampled, rather than total number of individuals captured.

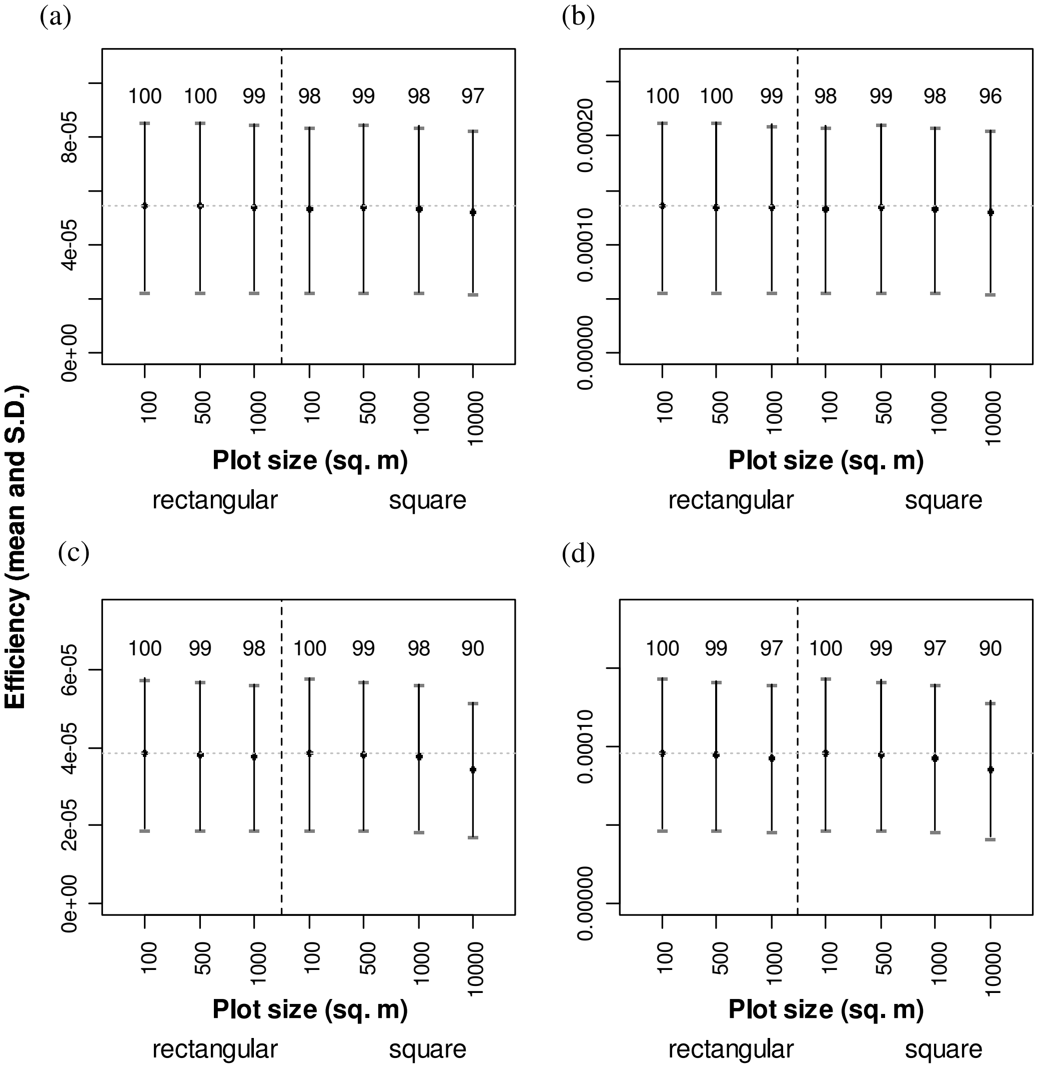


Figure 4.5. Sampling efficiency of the different plot designs tested (mean and standard deviation (S.D.)). (a) Pattern capture (Spearman correlation between true and sampled pattern) / total area sampled. (b) Pattern capture / sum of individuals sampled. (c) Proportion of species captured / total area sampled. (d) Proportion of species captured / sum of individuals sampled. The numbers above the error bars represent the relative plot design efficiency in percent (as percent of the most efficient sample design). Note that only simulation results for 1 - 5 ha total sampled area are included in order to avoid a bias due to the fact that for lower sample sizes no values are recorded for the 100 × 100 m plot method.

Discussion

Although the quest for most efficient sample design has occupied researchers for decades (Bormann 1953), there is little consensus on optimal sampling strategies for biodiversity studies in tropical forests (Gimaret-Carpentier *et al.* 1998; Phillips *et al.* 2003; Gordon & Newton 2006a; Gordon & Newton 2006b). By taking a simulation approach, we can analyse the factors that influence sampling performance and develop guidelines for choosing an appropriate sampling protocol for fieldwork in tropical forests (code for determining the most efficient sampling protocol available on request). The analysis has shown that sample size is of extreme importance, and that other factors such as the individual abundance, distribution, patchiness and the plot design only matter at low sample sizes. Even in the most optimistic scenarios a minimum sample size of 1,000 to 2,000 m^2 was needed to capture the species richness pattern, which is not always achieved and may partly explain disagreements over species richness patterns along environmental gradients such as altitude (Rahbek *et al.* 2007).

When sample sizes are large, it is possible to be confident about the sampled patterns irrespective of confounding field conditions or plot design. However, if sample sizes are small, these field conditions can significantly affect the ability to find the true species richness pattern. Most notably, the strength of the underlying gradient and the species abundance distribution can have severe effects when combined (weak gradients and even species abundance distribution), greatly reducing the ability of the sampling effort to derive the correct richness pattern. In comparison, the spatial arrangement of individuals and noise in the true gradient pattern are less influential. The simulations indicate that under difficult field conditions (weak gradient strength, aggregated spatial pattern and even species abundance distribution), a sample size of five times larger than the widely used Gentry approach (10 repetitions of randomly located 2 × 50 m plots; Gentry 1982) is needed to capture the pattern. However, it should be noted that in the traditional Gentry method all stems greater 2.5 cm are sampled, and if this is followed the method is sufficient to pick up the pattern.

The simulations showed that the smaller the plots the more efficient the sampling. This counters conclusions drawn by Kenkel & Podani (1991) who conclude that researchers should use the largest plot size possible. However, in their example the authors do not correct for overall sample area (larger plots resulted in a larger overall sample size). The assumption that the use of smaller plots may decrease estimation efficiency as they sample small-scale variation (Kenkel & Podani 1991) cannot be backed by our results. Rectangular plots were slightly more efficient than square plots, which confirms Bormann (1953) and studies cited therein. Al Gentry 2 × 50 m plot types were the most reliable (and the smallest) in the simulations. The plot size effect was more prominent with respect to species capture efficiency than to pattern capture efficiency. However, the simulations also indicated that the total area sampled is by far the strongest determinant of sampling performance, with plot size and shape having very little effect in themselves. This suggests that a sub-optimal sample size can only to a very small degree be counter-balanced by a strategically optimal plot design. The comparative importance of sample size to that of sampling design has also been shown for sampling strategies in studies providing data for habitat suitability modelling (Hirzel & Guisan 2002).

Care has to be taken when the aim of field surveys is not only to determine changes of species richness along a gradient but also to sample a maximum amount of the species present. The simulations showed differences in the way that the proportion of species found and the accuracy of the estimated richness pattern were affected by the different field conditions under the various sampling protocols. While even species abundance distributions generally resulted in larger proportions of species being found, they were also associated with greater difficulty in obtaining the true species richness pattern. Spatial aggregation of individuals on the other hand only slightly decreased the performance of sampling protocols in terms of capturing the richness pattern, but had a strong influence on the proportion of species found, especially when a large plot approach with few sample points was taken. Figure 4.6 presents a flowchart of how one may go about selecting the appropriate sampling protocol under different field conditions and resources constraints.

Many different species richness patterns have been reported in the literature. Such differences may be genuine or, at least in part, due to a whole range of different sampling intensities and designs being employed. These simulation results suggest that studies that did not utilise a sufficient sample size are most likely to have shown incorrect estimates of species richness patterns. We simulated a study area measuring only 500 × 500 m at each altitude, an area that is small compared to the extent of altitudinal band areas of most mountains. If an area as small as this requires a total sample area of 2,000 to 5,000 m^2 (0.8% to 2% of the simulated area) at realistic field conditions (a sampling intensity which is not always achieved), a much larger area will likely require an even higher sampling intensity. The simulations suggest that tree diversity studies in tropical forests, with an assessment threshold of 10 cm drh or more, covering less than these minimal sampling areas should be scrutinised carefully for sampling intensity artefacts.

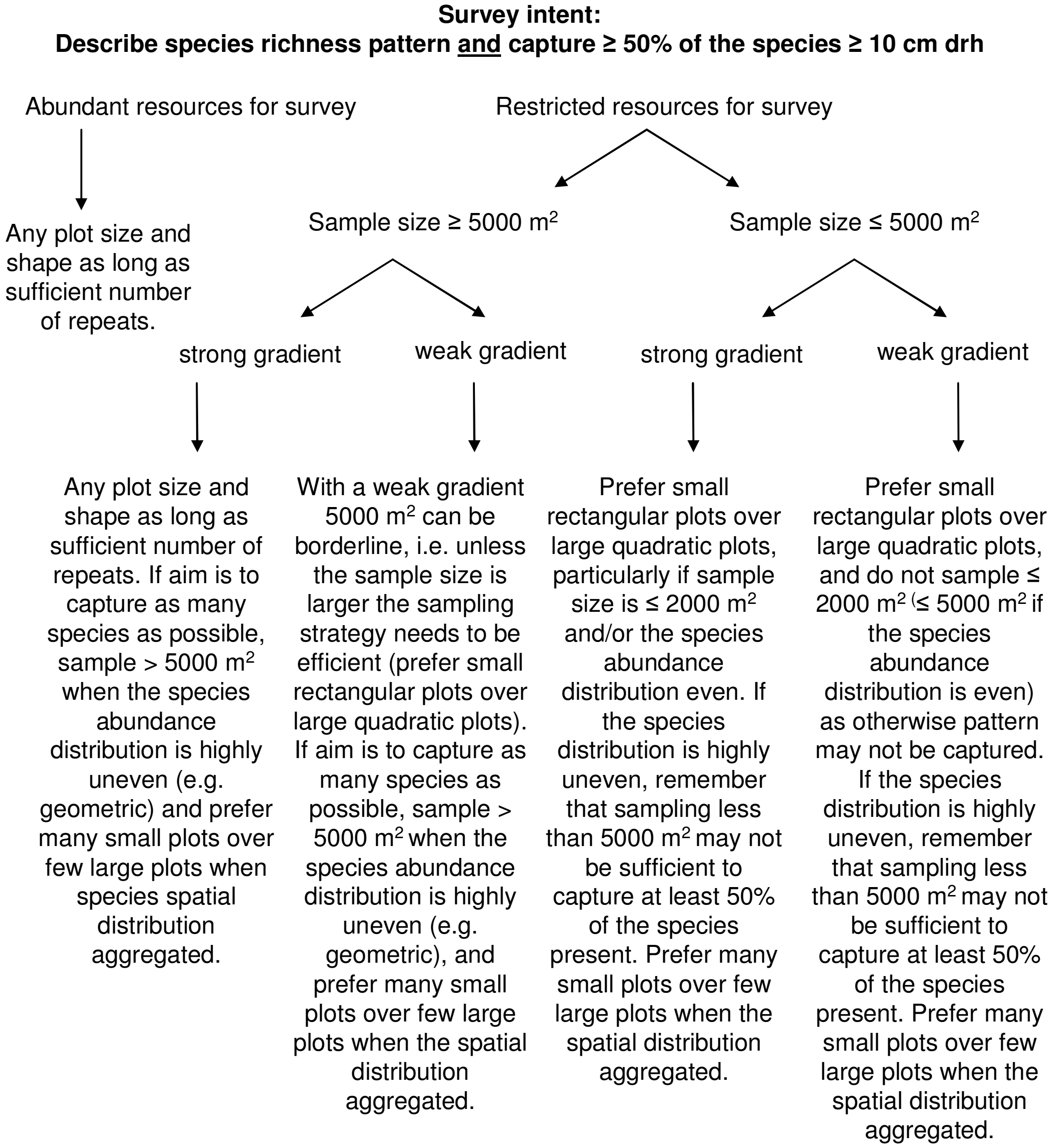


Figure 4.6. Flowchart of how one might want to go about selecting an appropriate sampling protocol under different field conditions and resource constraints. The field conditions (individual density, numbers of species present, strength of gradient and species abundance and spatial distributions) are as in the simulations.

Our study has three main implications: (1) Ecologists do not need to agonise over the best sampling strategy as long as their sample size is large. However, if sampling intensity is low then smaller plots with a larger number of repetitions are preferable to few but big plots and rectangular plots are preferable to square plots. In studies with small sample sizes it is also imperative to consider the level of spatial aggregation within species, the underlying species abundance distribution and the expected strength of the gradient as these variables may significantly influence the accuracy of the estimated richness pattern. (2) Funding for biodiversity inventories is scarce, and highly intensive assessments are becoming increasingly rare. There is a trend towards using data that are already available in large scale meta-analyses. Our analysis suggests that it is valid to compare studies that employ different plot designs in terms of the shape of the species richness gradient (not the strength) as long as sampling intensity is high enough. However, such analyses will be compromised if component studies have insufficient sampling intensities, and care should be taken to eliminate these from the datasets. (3) Intensive sampling of areas is time-consuming and expensive, particularly in remote locations. In the face of global environmental change and the urgency of assessing biodiversity for conservation prioritisation, rapid assessment approaches are increasingly being employed (Gordon & Newton 2006b). We agree that sampling should be efficient; however, we suggest that the importance of suitable sample sizes should not be underestimated. Studies with insufficient sampling will be much more likely to draw incorrect conclusions about species richness patterns, and may lead to flawed assessments and management of biodiversity and conservation priorities.

Acknowledgements

We thank A. Balmford, S.A. Bhagwat, J-A. Grytnes, D. Raffaelli and T. S. Romdal for critical discussions, and D. Hay for facilitating the running of the simulations. Funding for A.A. and P.J.P. was provided by a Marie Curie Actions grant to R.M., and for M.T.B. by NERC.

References

Ahrends A. & Marchant R. (2006-2008). Vegetation data analyses for Mselezi, Sali, Kanga, Nguru South, Pala Ulanga and Ukwiva Forest Reserves in the Eastern Arc Mountains. In: *Biodiversity Research and Awareness in the Lesser-Known Eastern Arc Mountains* (eds. Owen N, Wilkins Kindemba V, Fanning E & Howell KM). Frontier Tanzania BREAM Project, Dar es Salaam.

Archaux F., Berges L. & Chevalier R. (2007). Are plant censuses carried out on small quadrats more reliable than on larger ones? *Plant Ecology*, 188, 179-190.

Austin M.P. & Adomeit E.M. (1991). Sampling strategies costed by simulation. In: *Nature Conservation: Cost Effective Biological Surveys and Data Analysis* (eds. Margules CR & Austin MP). CSIRO Australia.

Austin M.P. & Heyligers P.C. (1989). Vegetation survey design for conservation - Gradsect sampling of forests in northeastern New-South-Wales. *Biological Conservation*, 50, 13-32.

Baddeley A. & Turner R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, 12, 1-42.

Bladt J., Larsen F.W. & Rahbek C. (2008). Does taxonomic diversity in indicator groups influence their effectiveness in identifying priority areas for species conservation? *Animal Conservation*, 11, 546-554.

Bormann F.H. (1953). The statistical efficiency of sample plot size and shape in forest ecology. *Ecology*, 34, 474-487.

Brooks T.M., Mittermeier R.A., da Fonseca G.A.B., Gerlach J., Hoffmann M., Lamoreux J.F., Mittermeier C.G., Pilgrim J.D. & Rodrigues A.S.L. (2006). Global biodiversity conservation priorities. *Science*, 313, 58-61.

Campbell P., Comiskey J., Alonso A., Dallmeier F., Nunez P., Beltran H., Baldeon S., Nauray W., De La Colina R., Acurio L. & Udvardy S. (2002). Modified whittaker plots as an assessment and monitoring tool for vegetation in a lowland tropical rainforest. *Environmental Monitoring and Assessment*, 76, 19-41.

Colwell R.K., Rahbek C. & Gotelli N.J. (2004). The mid-domain effect and species richness patterns: What have we learned so far? *American Naturalist*, 163, E1-E23.

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

Condit R., Hubbell S.P., Lafrankie J.V., Sukumar R., Manokaran N., Foster R.B. & Ashton P.S. (1996). Species-area and species-individual relationships for tropical trees: A comparison of three 50-ha plots. *Journal of Ecology*, 84, 549-562.

Dallmeier F. (1992). Long-term monitoring of biological diversity in tropical forest areas: Methods for establishment and inventory of permanent plots. In: *MAB Digest 11*. UNESCO, Paris.

Davis S.D., Heywood V.H., Herrera-MacBryde O., Villa-Lobos J. & Hamilton A.C. (1997). Centres of plant diversity. A guide and strategy for their conservation. (ed. WWF/IUCN) Gland.

Elith J., Leathwick J.R. & Hastie T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802-813.

Emmons L.H. & Gentry A.H. (1983). Tropical forest structure and the distribution of gliding and prehensile-tailed vertebrates. *American Naturalist*, 121, 513-524.

Gentry A.H. (1982). Patterns of Neotropical Plant-Species Diversity. *Evolutionary Biology*, 15, 1-85.

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

Gimaret-Carpentier C., Pelissier R., Pascal J.P. & Houllier F. (1998). Sampling strategies for the assessment of tree species diversity. *Journal of Vegetation Science*, 9, 161-172.

Gordon J.E. & Newton A.C. (2006a). Efficient floristic inventory for the assessment of tropical tree diversity: A comparative test of four alternative approaches. *Forest Ecology and Management*, 237, 564-573.

Gordon J.E. & Newton A.C. (2006b). The potential misapplication of rapid plant diversity assessment in tropical conservation. *Journal for Nature Conservation*, 14, 117-126.

Gotelli N.J., Anderson M.J., Arita H.T., Chao A., Colwell R.K., Connolly S.R., Currie D.J., Dunn R.R., Graves G.R., Green J.L., Grytnes J.A., Jiang Y.H., Jetz W., Lyons S.K., McCain C.M., Magurran A.E., Rahbek C., Rangel T., Soberon J., Webb C.O. & Willig M.R. (2009). Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, 12, 873-886.

Gotelli N.J. & Colwell R.K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4, 379-391.

Hall J.B. (1991). Multiple-nearest-tree sampling in an ecological survey of Afromontane catchment forest. *Forest Ecology and Management*, 42, 245-299.

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

Hirzel A. & Guisan A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, 157, 331-341.

Kenkel N.C. & Podani J. (1991). Plot size and estimation efficiency in plant community studies. *Journal of Vegetation Science*, 2, 539-544.

La Torre-Cuadros M., Herrando-Perez S. & Young K.R. (2007). Diversity and structural patterns for tropical montane and premontane forests of central Peru, with an assessment of the use of higher-taxon surrogacy. *Biodiversity and Conservation*, 16, 2965-2988.

Larsen F.W., Bladt J. & Rahbek C. (2009). Indicator taxa revisited: useful for conservation planning? *Diversity and Distributions*, 15, 70-79.

Lavers C. & Field R. (2006). A resource-based conceptual model of plant diversity that reassesses causality in the productivity-diversity relationship. *Global Ecology and Biogeography*, 15, 213-224.

Leigh E.G. (1975). Structure and climate in tropical rain-forest. *Annual Review of Ecology and Systematics*, 6, 67-86.

Magurran A.E. (1988). *Ecological Diversity and its Measurement*. Princeton University Press, New Jersey.

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

Nogues-Bravo D., Araujo M.B., Romdal T. & Rahbek C. (2008). Scale effects and human impact on the elevational species richness gradients. *Nature*, 453, 216-U8.

O'Brien E.M., Field R. & Whittaker R.J. (2000). Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography. *Oikos*, 89, 588-600.

Phillips O.L., Vasquez Martinez R., Nunez Vargas P., Lorenzo Monteagudo A., Chuspe Zans M.E., Galiano Sanchez W., Pena Cruz A., Timana M., Yli-Halla M. & Rose S. (2003). Efficient plot-based floristic assessment of tropical forests. *Journal of Tropical Ecology*, 19, 629-645.

Prance G.T., Beentje H., Dransfield J. & Johns R. (2000). The tropical flora remains undercollected. *Annals of the Missouri Botanical Garden*, 87, 67-71.

R Development Core Team (2008). *R: A language and environment for statistical computing*, Vienna. Available at: www.R-project.org.

Rahbek C. (1995). The elevational gradient of species richness - a uniform pattern. *Ecography*, 18, 200-205.

Rahbek C. (2005). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, 8, 224-239.

Rahbek C., Gotelli N.J., Colwell R.K., Entsminger G.L., Rangel T. & Graves G.R. (2007). Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proceedings of the Royal Society B-Biological Sciences*, 274, 165-174.

Ridgeway G. (2007). gbm: Generalized Boosted Regression Models. R package version 1.6-3. Available at: http://CRAN.R-project.org/package=gbm.

Sheil D., Ducey M.J., Sidiyasa K. & Samsoedin I. (2003). A new type of sample unit for the efficient assessment of diverse tree communities in complex forest landscapes. *Journal of Tropical Forest Science*, 15, 117-135.

Shmida A. (1984). Whittaker plant diversity sampling method. *Israel Journal of Botany*, 33, 41-46.

Stern M.J. (1998). Field comparisons of two rapid vegetation assessment techniques with permanent plot inventory data in Amazonian Peru. In: *Forest Biodiversity Research, Monitoring and Modeling: Conceptual Background and Old World Case Studies.* (eds. Dallmeier F & Comiskey JA). UNESCO/Parthenon Publishing Group, New York.

Stohlgren T.J., Falkner M.B. & Schell L.D. (1995). A Modified-Whittaker Nested Vegetation Sampling Method. *Vegetatio*, 117, 113-121.

Top N., Mizoue N., Ito S., Kai S., Nakao T. & Ty S. (2009). Effects of population density on forest structure and species richness and diversity of trees in Kampong Thom Province, Cambodia. *Biodiversity and Conservation*, 18, 717-738.

Wessels K.J., Van Jaarsveld A.S., Grimbeek J.D. & Van der Linde M.J. (1998). An evaluation of the gradsect biological survey method. *Biodiversity and Conservation*, 7, 1093-1121.

Whittaker R.J., Willis K.J. & Field R. (2001). Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, 28, 453-470.

Wittmann F., Zorzi B.T., Tizianel F.A.T., Urquiza M.V.S., Faria R.R., Machado e Sousa N., Modena E.D., Gamarra R.M. & Rosa A.L.M. (2008). Tree Species Composition, Structure, and Aboveground Wood Biomass of a Riparian Forest of the Lower Miranda River, Southern Pantanal, Brazil. *Folia Geobotanica*, 43, 397-411.

Supporting information

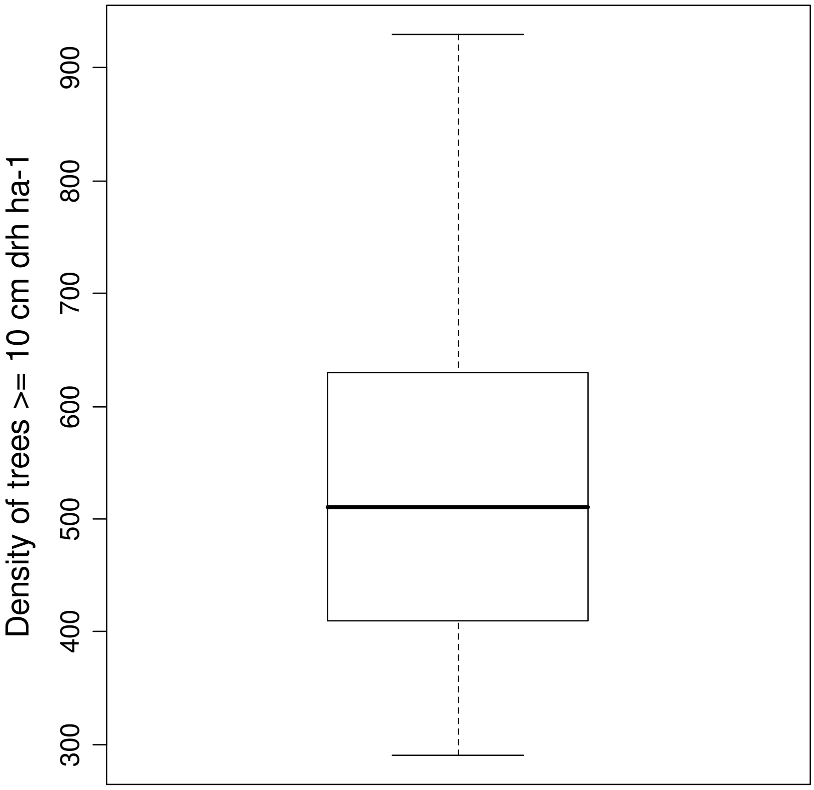


Figure S4.1. Tree density values in tropical forests assembled from the literature (45 sites, across Africa, Asia, and Central and South America).

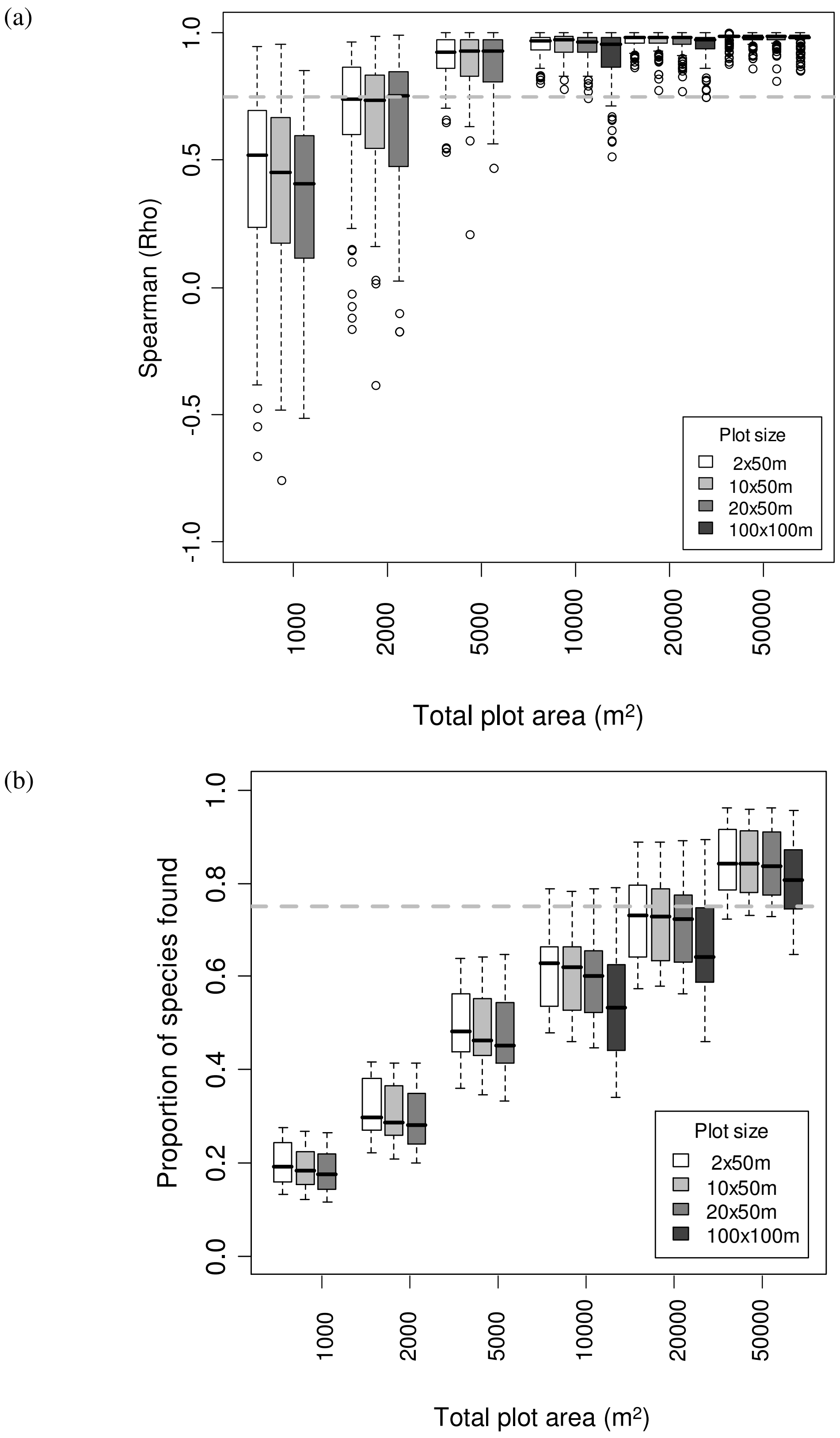
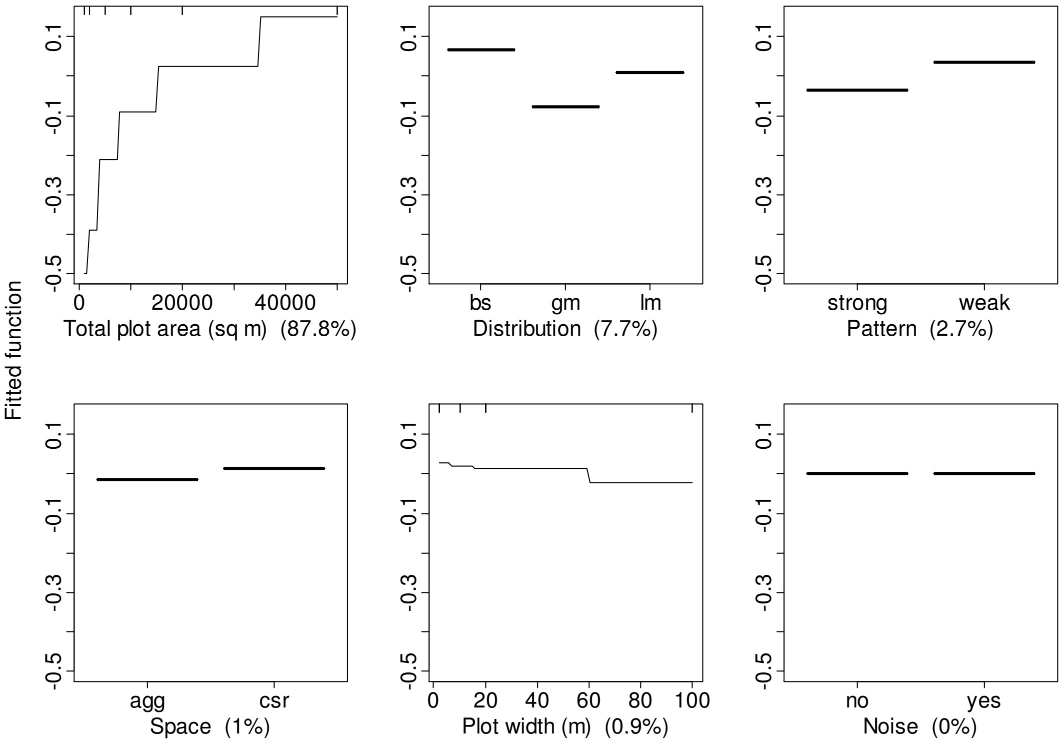


Figure S4.2. Performance of different plot designs in capturing the true species richness pattern (a) and species present (b).

(a)



(b)

	Distribution	Pattern	Space	Noise	Plot width	Total plot area
Distribution	0	0.08	0.01	0	0.04	0.33
Pattern	0	0	0	0	0.04	0.23
Space	0	0	0	0	1.1	0.05
Noise	0	0	0	0	0	0
Plot width	0	0	0	0	0	0.3
Total plot area	0	0	0	0	0	0

(c)

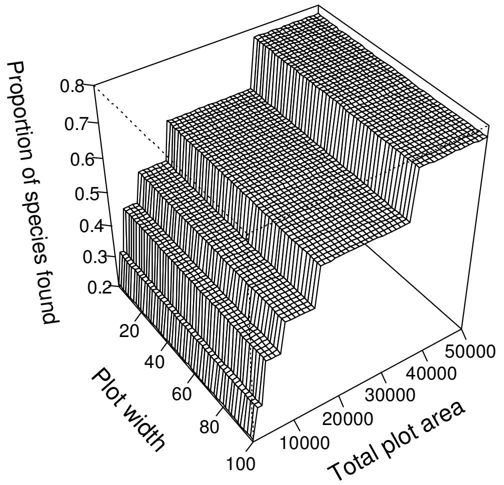
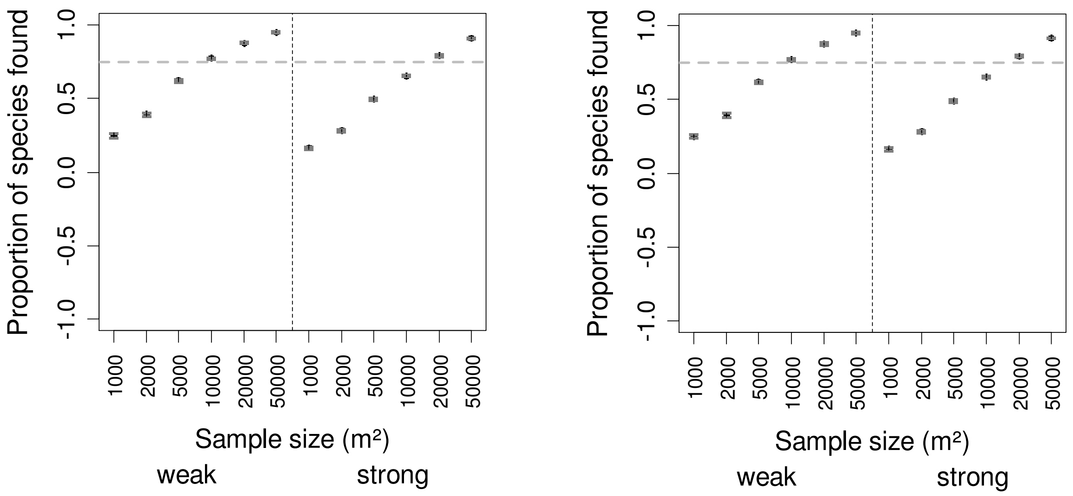
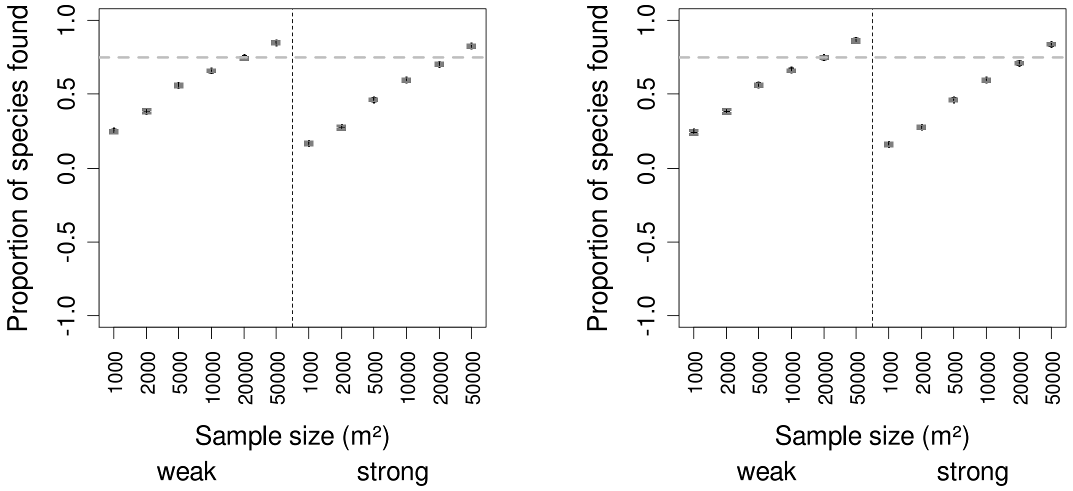


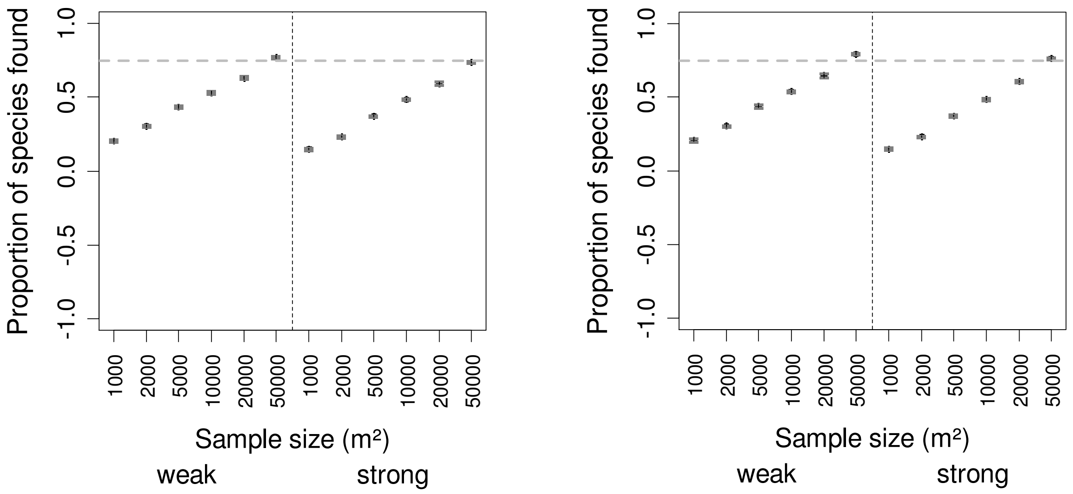
Figure S4.3. Boosted Regression Tree results for proportion of species found. (a) Partial effect of each of the tested field conditions on the proportion of species found in different sampling regimes. All other predictors are held constant at their mean. Bracketed values on the x axis detail the relative influence of the predictor variable, which is calculated as the number of times the variable is selected at nodes in the decision tree weighted by the resulting model improvement, averaged over all trees (Elith et al. 2008). Y axes are scaled to have zero mean. Labels for plots corresponding to distribution are bs (broken stick), gm (geometric) and lm (log normal); and for plots corresponding to space csr (complete spatial random) and agg (aggregated). (b) Summary of all interactions in the model. For interpretation of the values see legend Fig. 4.2. (c) Interaction between the partial effects of plot design and total area sampled on sampling performance.



Broken stick CSR **Broken stick aggregated**

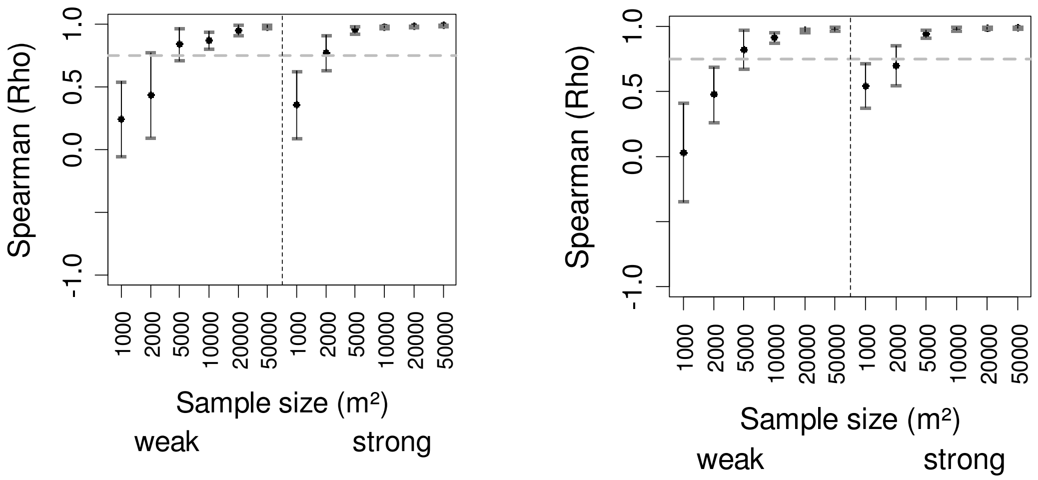


Lognormal CSR **Lognormal aggregated**



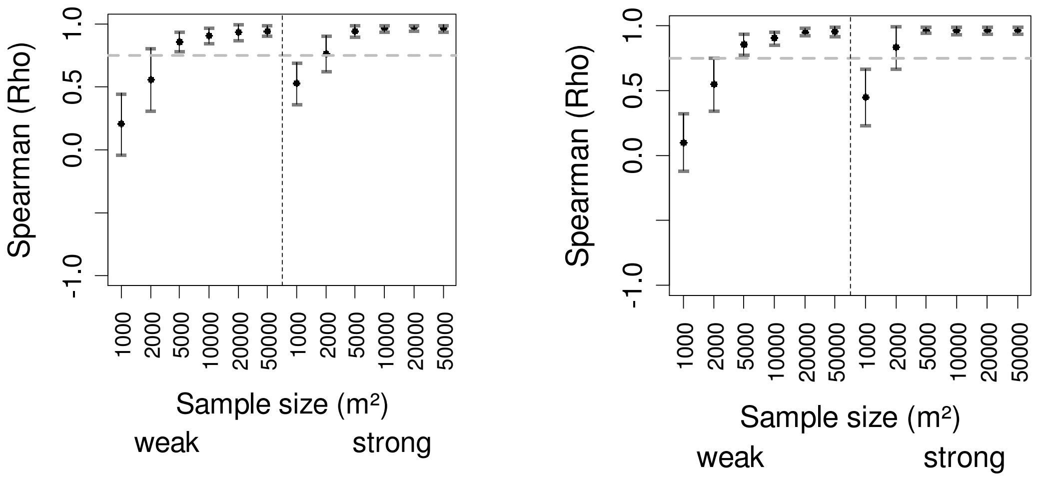
Geometric CSR **Geometric aggregated**

Figure S4.4. Minimum needed sample sizes for rectangular (2 × 50 m) plots sampling individuals ≥ 10 cm stem diameter to capture at least 75% of the species present. Black dots are the mean and grey bars the upper and lower standard deviation of the proportion of species found.



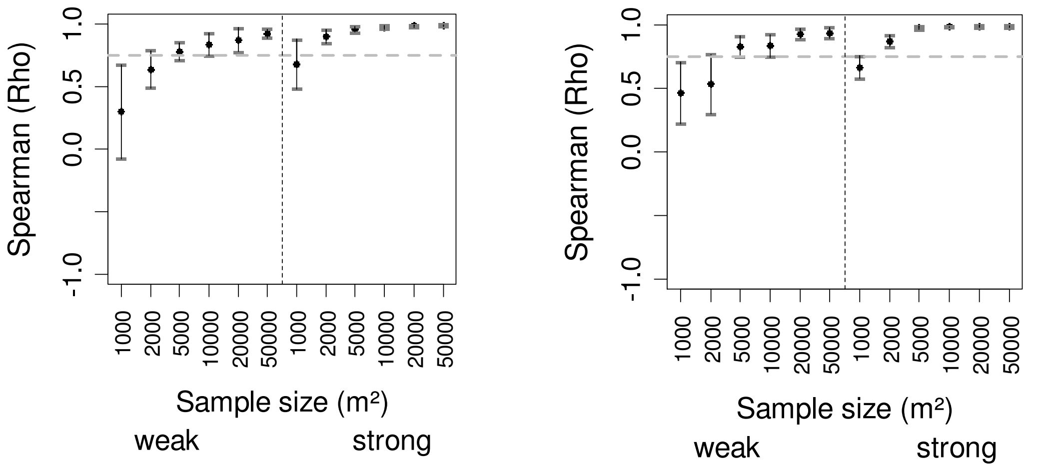
Broken stick CSR

Broken stick aggregated



Lognormal CSR

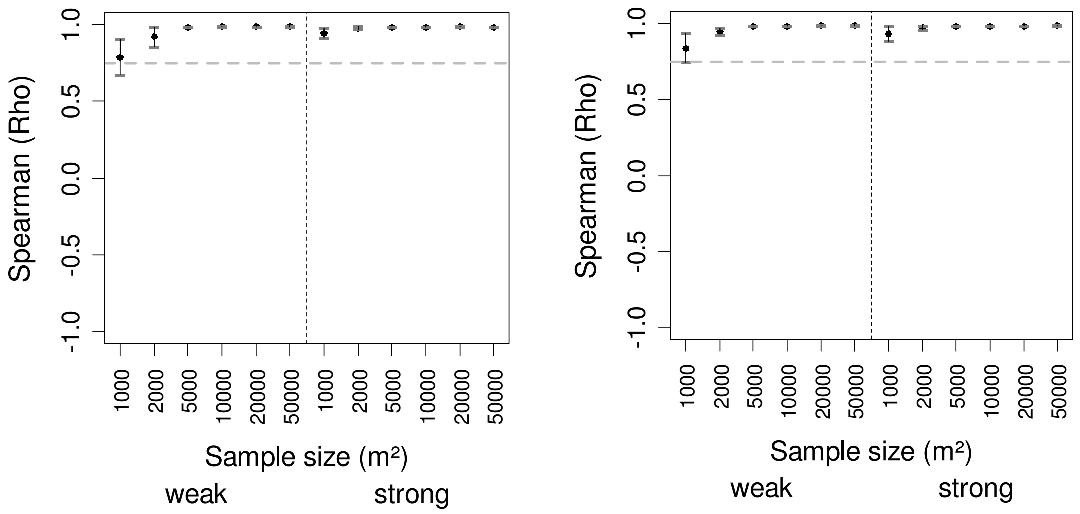
Lognormal aggregated



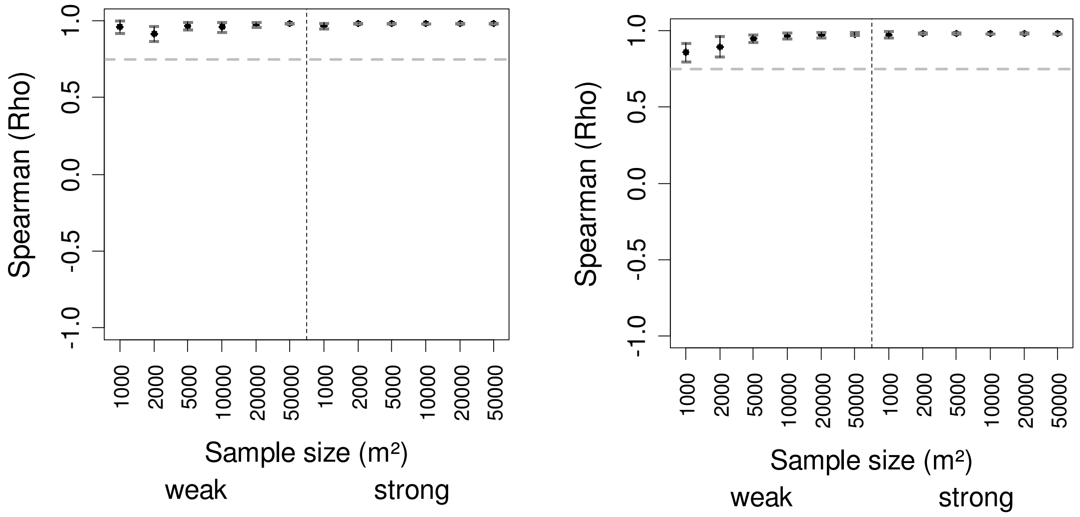
Geometric CSR

Geometric aggregated

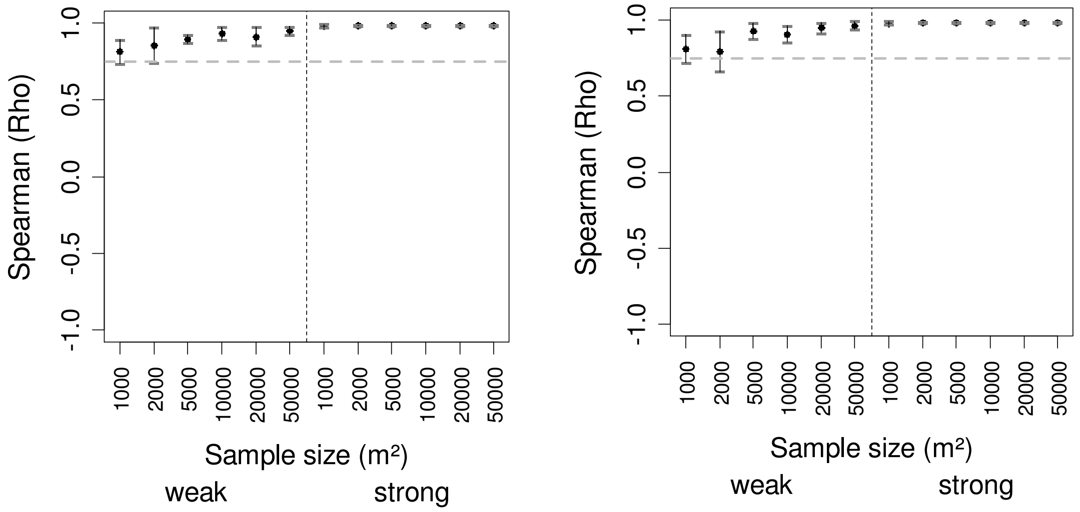
Figure S4.5. Minimum needed sample sizes for square (10 × 10 m) plots sampling individuals ≥ 10 cm stem diameter to capture the true species richness pattern (Spearman ≥ 0.75). Black dots are the mean and grey bars the upper and lower standard deviation of the correlation between the sample and the true species richness pattern.



Broken stick CSR **Broken stick aggregated**



Lognormal CSR **Lognormal aggregated**



Geometric CSR **Geometric aggregated**

Figure S4.6. Minimum needed sample sizes for rectangular (2 × 50 m) plots sampling individuals ≥ 2.5 cm stem diameter to capture the true species richness pattern (Spearman $\geq$ 0.75). Black dots are the mean and grey bars the upper and lower standard deviation of the correlation between the sample and the true species richness pattern.

No Individuals (density/ha)	**Plot sizes**	**Most efficient plot size(s)**	**Minimum needed sample size** to capture richness pattern out of tested sample sizes: mean Spearman ≥ 0.75; (upper S.D. ≥ 0.75)
10,000 (400)	2×50, 10×50, 20×50, 100×100	2×50	weak gradient: 5,000 m^2 (2,000 m^2) strong gradient: 2,000 m^2 (1,000 m^2)
10,000 (400)	10×10, ~22×22, ~32×32, 100×100	10×10, ~22×22	weak gradient: 5,000 m^2 (2,000 m^2) strong gradient: 2,000 m^2 (1,000 m^2)
50,000 (2,000)	2×50, 10×50, 20×50, 100×100	2×50	weak gradient: 1,000 m^2 strong gradient: 1,000 m^2

Table S4.1. Summary of all simulations. In all simulations the area at each sampling site (e.g. altitudinal band) was set to 25 ha (500 × 500 m), and three species abundance distributions (broken stick, lognormal and geometric), two strength of gradient (weak, strong), two spatial aggregations (csr and aggregated), two levels of noise (no noise, noise of half the strength of the gradient), and six total area sample sizes (1,000; 2,000; 5,000; 10,000; 20,000; 50,000 m^2) were tested.

Dependent	**Predictors (% contribution)**	**Number of trees**	**Interactions**	**Cross validation correlation (S.E.)**
Pattern capture	Distribution (6%), pattern (13%), space (1%), noise (1%), plot size (1%), total plot area (76%)	1500	8 (total plot area with distribution (0.39), pattern (1.36), space (0.03), noise (0.06), plot size (0.02); pattern with distribution (0.01), space (0.01), plot size (0.05))	0.86 (0.01)
Species capture	Distribution (8%), pattern (3%), space (1%), noise (0%), plot size (1%), total plot area (88%)	2100	9 (total plot area with distribution (0.33), pattern (0.23), space (0.04), plot size (0.29); plot size with distribution (0.04), pattern (0.04), space (1.1); distribution with space (0.01), pattern (0.08))	1 (0.00)
Pattern capture efficiency	Distribution (11%), pattern (36%), space (2%), noise (1%), plot size (49%), plot shape (1%)	500	0	0.31 (0.01)
Species capture efficiency	Distribution (13%), pattern (22%), space (2%), noise (0%), plot size (64%), plot shape (0%)	550	0	0.38 (0.01)
Sum of assessed individuals	Distribution (0%), pattern (0%), space (0%), noise (0%), plot size (0%), total plot area (100%)	50	0	1 (0.00)

Table S4.2. Summary of all Boosted Regression Tree models. The following settings were applied to all models: error distribution: Gaussian; maximum permitted tree complexity = 5; learning rate = 0.005; bag fraction = 0.75.

Link from Chapter 4 to Chapter 5

The previous three result chapters have focused on assessments of irreplaceability, however, quantifying vulnerability is an equally important cornerstone of efficient conservation planning. While deforestation is relatively straight forward to infer from remotely-sensed data, large-scale measurement and predictions of forest degradation still pose a severe challenge to conservation planners and initiatives such as ‘Reducing Emissions from Deforestation and Degradation’ (REDD). Patterns and drivers of degradation and its impact on biodiversity and carbon storage are only understood in outline. Based on field degradation data collected at increasing distance from the centre of demand Dar es Salaam, Chapter 5 tests whether economic theory can be used to predict the spread of degradation across landscapes, and analyses degradation impacts on timber values, carbon storage and biodiversity.

Degradation can have severe impacts but these are difficult to predict (here Pugu Forest Reserve close to Dar es Salaam).

Chapter 5 – Predictable waves of forest degradation spreading from an African city

Online early in *Proceedings of the National Academy of Science, USA*
Social Sciences/Sustainability Science (doi: 10.1073/pnas.0914471107)

Predictable waves of forest degradation spreading from an African city

Antje Ahrends[1], Neil D. Burgess[2], Simon L. Lewis[3], Mark T. Bulling[4], G. Philip Clarke[5], Boniface E. Mhoro[6], and Simon A. H. Milledge[7]

[1] Environment Department, University of York, YO10 5DD, UK; Royal Botanic Garden Edinburgh, 20A Inverleith Row, EH3 5LR, UK; [2] Zoology Department, University of Cambridge, CB2 3EJ, UK; WWF US, 1250 24th St NW, Washington DC, USA; Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Universitetsparken 15, DK- 2100 Copenhagen Ø, Denmark; [3] Earth & Biosphere Institute, School of Geography, University of Leeds, West Yorkshire, LS2 9JT, UK; [4] Oceanlab, University of Aberdeen, Newburgh, Aberdeenshire, AB41 6AA, UK; [5] The Hermitage, Crewkerne, Somerset, TA18 8HA, UK; [6] University of Dar es Salaam, P. O Box 35091, Dar es Salaam, Tanzania; [7] TRAFFIC, P.O. Box 106060, Dar es Salaam, Tanzania

Abstract

Tropical forest degradation emits ~0.5 Pg carbon yr^{-1}, reduces biodiversity, and facilitates forest clearance. Understanding degradation drivers and patterns is therefore crucial to manage forests to mitigate climate change and reduce biodiversity loss. Putative patterns of degradation have variously been described, but these have not been assessed quantitatively, nor tested systematically. Economic theory predicts that forest degradation should entail the systematic removal of high-value forest products around demand centres, and that exploitation should expand as concentric waves, with each new wave targeting lower-value products. We tested this theory by monitoring forest patches 10 - 220 km from Dar es Salaam, Tanzania in 1991 and 2005. Our predictions were confirmed: high-value logging expanded 9 $km \cdot y^{-1}$, and an inner wave of lower value charcoal production 2 $km \cdot y^{-1}$. Both carbon storage and species richness significantly increased with each kilometre from Dar es Salaam (0.2 Mg carbon ha^{-1}; 0.1 species ha^{-1}). This suggests that tropical forest degradation can be modelled and predicted, which might enable better targeting of policies to reduce carbon and biodiversity loss from forest systems in developing countries.

Introduction

Approximately one-third of remaining tropical forest has been degraded through selective logging (Johns 1997), a practice that adds ~0.5 Pg carbon yr^{-1} to the atmosphere (Putz *et al.*

2008). Forest degradation is broadly defined as the long-term reduction of the overall potential supply of goods and services, including carbon storage, wood production and biodiversity conservation. The impacts of individual forms of tropical forest degradation are understood in outline, for example, industrial logging reduces carbon stocks (Nepstad *et al.* 1999; Asner *et al.* 2005; Berry *et al.* in press) and changes biodiversity, often reducing it (Bawa & Seidler 1998; Willott 1999; Gardner *et al.* 2009; Berry *et al.* in press). However, our understanding of the drivers of forest degradation and how it proceeds across landscapes over time is based on a small number of observational studies (Gentry & Vasquez 1988; Vasquez & Gentry 1989; Tang *et al.* in press). While remote-sensing technology has facilitated progress in spatial modelling of deforestation (Chomitz & Gray 1996; DeFries *et al.* 2010), putative patterns in forest degradation dynamics have not been tested systematically, preventing the development of broadly applicable predictions. Yet this knowledge is critical if optimal policies are to be implemented to manage forests to mitigate climate change and biodiversity loss, in particular the currently negotiated instrument ‘Reducing Emissions from Deforestation and Forest Degradation’ (REDD) within the UN Framework Convention on Climate Change (Miles & Kapos 2008; Putz *et al.* 2008).

Economic resource use theory provides a general model to predict patterns of forest degradation. This asserts that land is allocated to the activity that provides the maximum **net** value (‘rent’ in economic terms), which in turn is the largest gain from the land minus the costs involved to obtain that gain (von Thünen 1966; Chomitz & Gray 1996; Angelsen 2007). This translates to a prediction that waves of forest degradation will emanate from major demand centres and expand into nearby forested areas, and will target resources in sequence, starting from those of highest value (Chomitz & Gray 1996). Such a sequence of demand, linked to resource utilisation, has been demonstrated for unmanaged fisheries, where it is termed 'fishing down the food web' (Pauly *et al.* 1998; Berkes *et al.* 2006; Scales *et al.* 2006), but has not been shown for the exploitation of differently-valued tropical forest products, and has also not been linked to impacts on forest degradation. Here we test economic resource-use theory predictions against ground observations of forest degradation in 10 forests over 14 years (1991 - 2005) from the demand centre of Dar es Salaam (DES), Tanzania.

Materials and methods

Study Area

DES is a rapidly expanding city of some 3 - 4 million people on the Indian Ocean coast of East Africa. Forest product demand is increasing sharply to meet expanding markets overseas (particularly China), as well as rising domestic demand for building materials and cooking fuel (Liu & Diamond 2005; UN-HABITAT 2008). This pattern of increasing consumption, combined with weak resource management practices is typical of other tropical regions (Geist & Lambin 2002; Ehrlich & Pringle 2008; Laurance 2008), making the forests around DES a potentially valuable model system for testing forest degradation theory. The East African coastal forests - thought to once have formed a belt along the East African coast from southern Somalia to northern Mozambique - now remain as a series of highly fragmented forest patches, covering less than 10% of their climatically suitable habitat (Burgess & Clarke 2000). Due to their exceptionally high levels of relict endemism, multiple conservation priority setting schemes have identified them as one of the most important areas for biodiversity conservation worldwide (Stattersfield *et al.* 1998; Olson & Dinerstein 2002; Mittermeier *et al.* 2005).

Field data collection

In 1990 - 1994 (median = January 1st 1991) 11 forests were sampled in coastal Tanzania, noting the type of extractive activities that were occurring in each forest (Clarke & Dickinson 1995). In each of the forests one to three plots of different size from 0.025 to 0.25 ha were located at random in areas stratified according to the type of forest vegetation (n total = 45 plots; area sampled 4 ha), with all trees $\geq$ 100 mm diameter at reference height (1.3 m along the stem or above buttresses; drh) measured and identified to species. In 2004 - 2005 (median = January 7th 2005) eight of these forests were re-sampled and two additional forests surveyed. The forests were chosen to span 10 - 220 km distance from DES and to have similar climate, topography, soils and socio-economic conditions (see Supporting information for Chapter 5 Fig. S5.1; Table S5.1). Within each forest we randomly located transects (10 × 500 - 1,500 m length) to sample 0.1% of the area of each of the ten forests. Within each transect all trees and stumps $\geq$ 50 mm drh were recorded as alive, naturally dead or cut, measured and identified to species (where possible) (n = 12,018). Again we noted the types of extractive activities occurring, including quantifying the number of charcoal production pits.

Calculation of variables for the analysis (2005 data)

Timber value: Trees qualifying as timber were defined as all trees with straight stems at least 3 m in length and $\geq$ 150 mm drh. We categorised each stem as ‘high-value’, ‘medium-value’, ‘low-value’ timber or ‘suitable only for charcoal production’ based on published use-data (Bryce 2003; The United Republic of Tanzania 2004), and calculated the density of suitable timber trees and the average drh and basal area sum of all trees.

Value of extracted timber: Calculations included average stump diameter and average forests' 'stump value', which was based on the mean royalties that the Tanzanian Government collects for felling the respective species (US$ value in the year 2005).

Carbon stocks: Above-ground carbon stocks were computed using the Dry Forest allometric equation of Chave *et al.* (2005). This computes the carbon stored in individual trees utilising drh and the wood specific gravity of each stem. Wood specific gravity was taken from a global database (Chave *et al.* 2009). When species-specific wood specific gravity data was not available, genus-level values were used (Lewis *et al.* 2009).

Species diversity: Three area-standardised species richness estimates were calculated (Mao Tau, Chao 1, Jaccard 2) for all trees $\geq$ 150 mm drh, each computed over a sub-sample of eight plots (corresponding to 0.4 ha) with 50 iterations, using EstimateS (Colwell 2006). Species richness standardised by area is sometimes referred to as 'species density' (Gotelli & Colwell 2001) as opposed to species richness standardised by number of individuals.

Forest use changes between 1991 and 2005

An estimate of total carbon loss in Dar es Salaam and Pwani Regions in Tanzania was calculated as follows: the dominant extractive activity (charcoal burning, medium-value timber logging and high-value timber logging (= degradation stages)) was established for all forests (n = 33) in 1991 and 2005. The degradation stage was assigned on the basis of survey data where available (for n = 12 forests; see Fig. 5.1 and Table 5.1), and for forests for which no data were available we estimated the degradation stage based on the degradation wave predictions established in this study verified against expert opinion. The loss of carbon associated with the transition from one degradation stage to another (e.g. from medium-value timber logging to charcoal burning) was calculated as the average difference in carbon stored between forests in these different degradation stages in 2005. The rationale for basing the rates of carbon loss on spatial data in 2005, instead of on temporal changes since 1991, is the greater data availability for 2005. Total carbon stored in the area in 1991 and 2005 was computed as the sum of the products of each forest's size (ha) and the average amount of carbon stored ha^{-1} in a forest of that particular degradation stage. Given that forests may have remained in the same degradation stage but still lost considerable amounts of carbon, our estimate is likely to be conservative.

Modelling spatial degradation patterns in 2005

We fitted linear regression models for each of eight dependent variables (average stump value (US$ 2005), average stump diameter (mm), standing timber (n ha^{-1}), basal area (m^2 ha^{-1}), standing carbon (Mg ha^{-1}), and Mao Tau, Chao 1 and Jaccard 2 indices) with four predictor variables: (1) distance from DES (km), (2) distance from the main DES road (km)

(Fig. S5.1), (3) forest truck accessibility, and (4) population density in wards within 2 km distance of the focal forest reserve. Distance from DES and distance from the main DES road were measured using a car odometer, thus, they represent road distances. Forest Truck accessibility was measured as the sum of scores for three attributes: (1) main road to forest graded, (2) roads within forest graded, (3) terrain easily accessible, i.e. not hilly or mountainous. If an attribute was fulfilled the score 1 was assigned, otherwise 0. Figures for average population density by wards were based on the National Bureau of Statistics Tanzania (2002).

If the correlation between independent variables was ≥ 0.7 (Pearson), the variable that was least correlated with the particular dependent variable was excluded from the process. Linear regression models were fitted including all terms and the interactions. Model validation procedures followed Zuur *et al.* (2009) and indicated no heterogeneity of variance, and the presence of normality in the residuals. To find the minimal adequate model, we applied a backward stepwise selection using the partial *F*-statistic. Model validation on the final model was then carried out once more. It is noted that the number of data points available for the analysis was small. However, the dataset contributing to each of these points was extensive. This, in conjunction with the strongly emerging pattern and its consistency across all tested variables increases our confidence in the reliability of the analysis.

All statistical analyses were performed in the “R” statistical and programming environment version 2.9.2. (R Development Core Team 2009) and its library ‘boot’ (Canty & Ripley 2009). Standard errors are based on 1,000 nonparametric bootstrap replicates.

Results

Changes between 1991 and 2005

We found three distinct degradation waves emanating from DES in accordance with economic and fisheries theory (Fig. 5.1; Table 5.1; Table S5.2). In 1991, the innermost degradation wave, which comprised the extraction of low-value wood for charcoal production, extended up to 50 km from DES and was dominant within a 20-km distance from the city. This largely provides cooking fuel for DES residents. A second degradation wave extracted low and medium-value timber for local and DES consumption in construction, and export. This middle wave extended 20 – 100 km from DES and was dominant at distances of 20 – 50 km from DES where high-value timber species were almost exhausted. Beyond 50 km an outer wave of forest use consisted of high-value timber logging, for DES consumption in construction, and export.

The order of concentric degradation waves remained the same in 2005 (Fig. 5.1; Table 5.1; Table S5.1), but had expanded significantly. Charcoal production had become the dominant use up to 50 km from DES but extended to 170 km from DES, with the outer boundary of this wave having moved 120 km since 1991, and the outer boundary of the area where charcoal production is the dominant use having moved 30 km (2 km·y^{-1}). Charcoal production sites (pits or earth mound kilns, on average 8 × 8 m in size) covered ~8% of forest area within 20 km distance of DES. At 170 km distance they covered 0.3%, and beyond 210 km distance no charcoal production was found (Fig. 5.2a). The sharp drop in charcoal production sites at increasing distance from DES is likely driven by the cost of transporting the charcoal to DES, making its production an increasingly marginal activity at greater distance from this city. However, as nearby forests are exhausted charcoal prices increase, and charcoal production further from DES becomes more attractive, as shown by DES charcoal prices which increased from US$ 0.18 kg^{-1} to US$ 0.27 kg^{-1} (US$ 2009, calculated using a GDP deflator; www.measuringworth.com) from 1997 to 2007 (Hofstad 1997; Luoga *et al.* 2000).

Medium-value timber logging (round wood export value up to US$ 250 per m^3 in 2005 (Milledge *et al.* 2007)) dominated at 50 – 170 km from DES, and the outer bound of this second wave had moved 110 km, reaching the forests south of the Rufiji River (Fig. 5.1). The high-value timber logging wave (round wood export value ~US$ 330 per m^3 in 2005 (Milledge *et al.* 2007)) started at 170 km distance from DES (the inner boundary of high-value timber logging wave having moved ∽9 km·y^{-1}), was dominant starting from 210 km and continued to at least 220 km (Fig. 5.1). Timber logging removed tree species in sequence. Two high-value timber species (*Milicia excelsa* (Welw.) C.C.Berg and *Brachylaena huillensis* O. Hoffm.) have been entirely depleted, and stocks of two others (*Pterocarpus angolensis* DC. and *Khaya anthotheca* (Welw.) C.DC.) have been almost exhausted (Table S5.1). Mean timber value per tree stump increased US$ 0.1 km^{-1} distance from DES (US$ value in the year 2005), illustrating the economic consequences of the degradation wave (Fig. 5.2b).

Thus, in line with resource use theory, forest degradation waves are expanding rapidly from DES: the charcoal burning and medium-value timber ‘wave fronts’ have, respectively, expanded 120 km and 110 km between 1991 and 2005. The inner boundary of the high-value timber logging wave has expanded 120 km, and is now almost beyond the edge of the study area. Multiple regression models showed that out of the four tested factors that may predict levels of degradation in the ten forests studied in 2005, distance from DES - with one exception - was the sole significant predictor and explained between 60 and 80% of variation (Table S5.2). Only the average forests' stump value was better explained by the forests' lorry accessibility, which may indicate accessibility is a stronger factor in high value timber logging than distance; however, the two variables were also strongly correlated (Pearson = -0.72). Distance from the main road was also marginally correlated with distance from Dar es Salaam (Pearson = 0.65) (Table S5.2) and may be another important driving factor but in order to avoid inflated standard errors of the variable coefficients due to high collinearity (Zuur *et al.* 2009) it was excluded from the model.

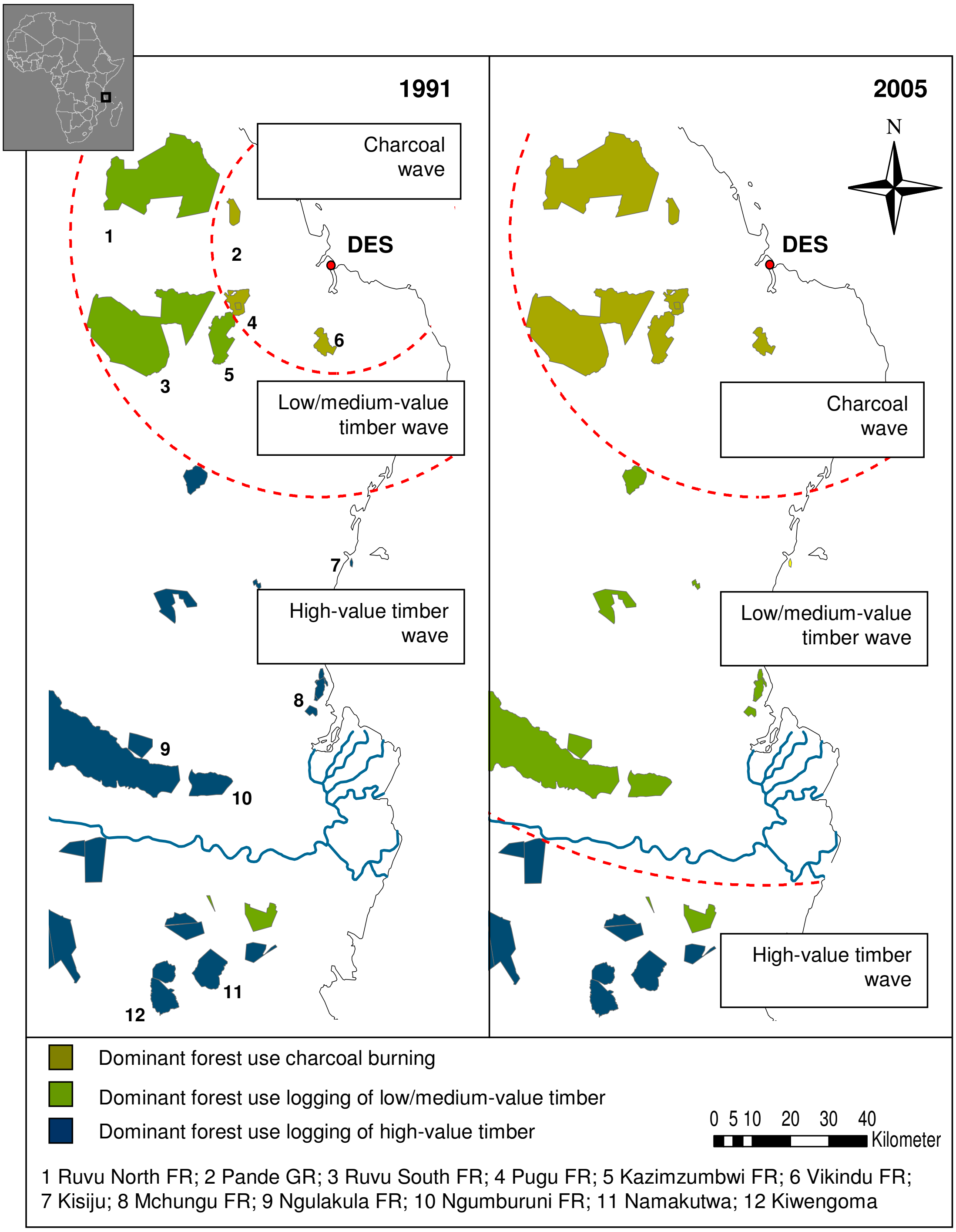


Figure 5.1. Map of the degradation waves of dominant forest use in the study area in 1991 and 2005. Charcoal burning has moved c. 30 km road distance from DES in this time period and medium-value timber logging 160 km. The outer boundary of high-value timber logging was outside the study area already in 1991. Muhoro and Nyamwagne, the two forests south of Rufiji River that do not follow the general degradation pattern, are not natural forests. Note that the forest boundaries are also likely to have changed between 1991 and 2005 as there has been some forest clearance and agricultural encroachment at the boundaries. This is not reflected in the figure.

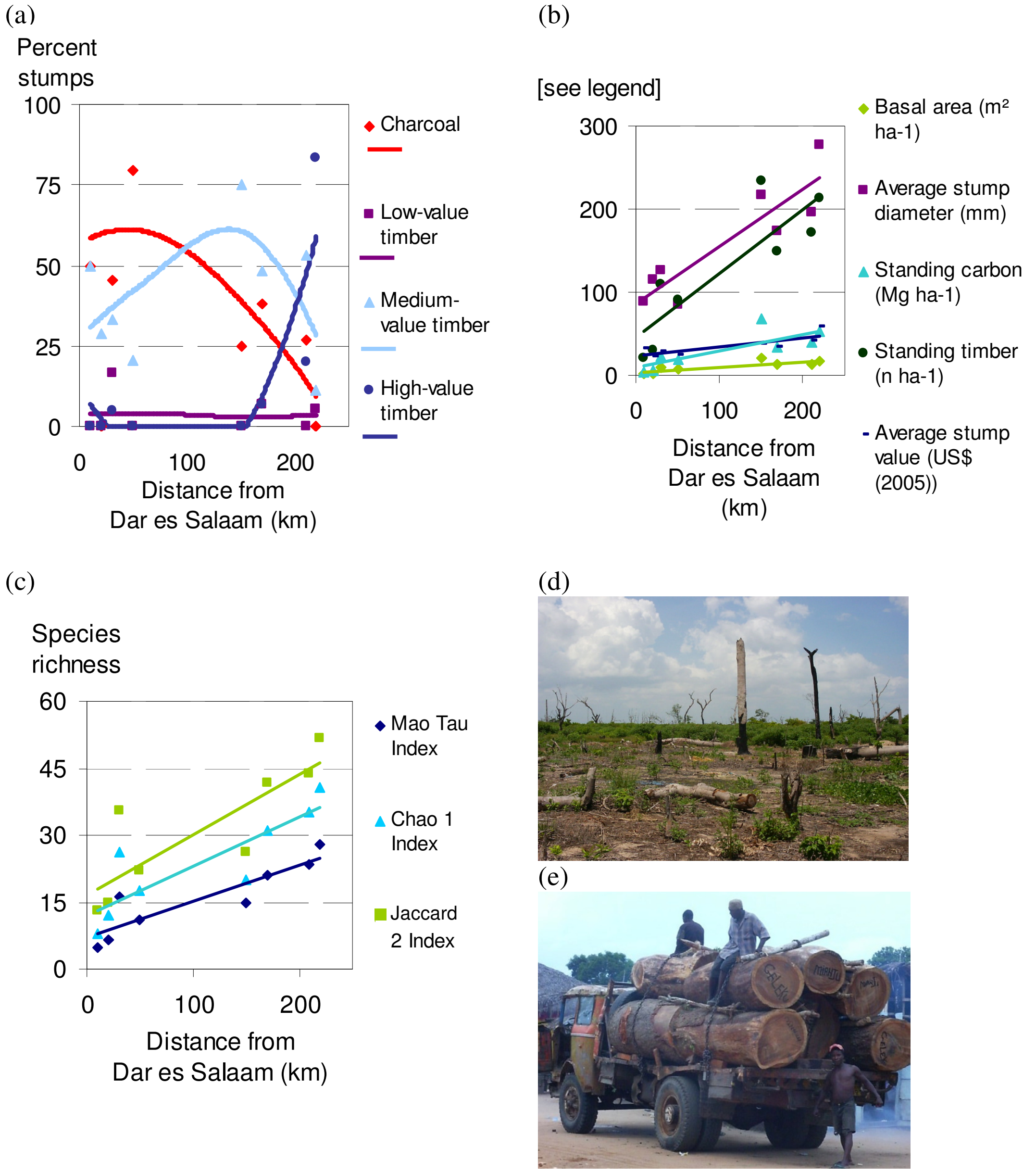


Figure 5.2. Patterns in forest use and condition at increasing distance from Dar es Salaam (DES). (a) Forest use at increasing distance from DES, quantified as numbers of stumps for trees used for charcoal burning (r^2 = 0.73) or as low-, medium- or high-value timber (r^2 = 0.01, 0.41 and 0.74, respectively), presented as interpolated lines using a loess function (span = 2). The low-value timber pattern is much less clear than the other extraction waves, which is presumably due to the fact that low value timber is used both commercially and for subsistence locally, i.e. the pattern is influenced by different factors and may therefore be blurred. (b) Forest condition across distance from DES measured using forest structure (basal area, average stump diameter, standing timber, standing carbon), and economic (average stump value) variables. (c) Estimated tree species richness at differing distances from DES using observed species richness (Mao Tau) and total species richness (Chao 1 and Jaccard 2) estimators, each randomised 50 times over 8 samples (0.4 ha). (d) Kisiju forest at 90 km distance from DES has been almost removed since 1994 and all woody resources converted to charcoal. (e) Illegally logged high-value timber harvested at 200 km distance from DES. Number of independent data points for the production of (a), (b) and (c) were *n* = 8 (forests), respectively.

Distance DES (km)	Forest	Selective high-value logging		Medium and low-value logging		Charcoal burning	
		1991	2005	1991	2005	1991	2005
10	Vikindu	NO-	NO-	NO-	NO-	YES+	YES
20	Pande	NO-	NO-	YES	NO-	YES+	YES
30	Pugu	NO-	NO-	YES+	NO-	YES+	YES+
32	Kazimzumbwi	NO-		YES		YES+	
40	Ruvu North	NO-		YES		YES+	
50	Ruvu South	YES	NO-	YES+	YES+	YES+	YES+
90	Kisiju[1]	YES	NO-	YES+	NO-	YES+	NO-
150	Mchungu	YES	NO-	NO+	YES+	NO+	NO+
170	Ngumburuni		NO-		YES+		YES+
210	Namakutwa	YES+	YES	NO+	YES+	NO+	NO+
220	Kiwengoma	YES+	YES	NO+	NO+	NO+	NO+

Table 5.1. Degradation status in 11 study forests in 1991 and 2005, at increasing distance from Dar es Salaam. The sequence would be: NO+: abundant resources, not yet exploited; YES+: exploited with abundant resources; YES: exploited with some resources remaining; NO -: not exploited because resources are exhausted.

Consequences of forest degradation

The systematic depletion of forest resources resulted in reduced timber stocks, carbon storage and biodiversity (Fig. 5.2b, c and d). In 2005, within a 20 km radius of DES forests had 25 trees ha^{-1} (bootstrapped S.E. = ±3.46), compared to 99 trees ha^{-1} (bootstrapped S.E. = ±6.35) within a radius of > 20 and ≤ 50 km distance, and 193 trees ha^{-1} (bootstrapped S.E. = ±15.11) ≥ 200 km (Fig. 5.2b). Thus, at ≤ 50 km from DES forest canopies are no longer closed, and at ≤ 20 km distance the forests are practically removed. Between 1991 and 2005 tree density (trees ha^{-1}) in a given forest significantly declined ($p \leq 0.05$; paired t-test; $n = 6$), while there were also marginally significant reductions in aboveground carbon (Mg ha^{-1}) and mean tree diameter ($p \leq 0.1$; paired t-test; $n = 4$), further illustrating how the forest condition has deteriorated (Figs. S5.2 and S5.3).

Total species richness increased from 8 - 13 tree species (depending upon the index used) per sample unit (0.4 ha) in forests closest to DES to 41 - 52 tree species ≥ 220 km from DES

---

[1] Note that the unprotected Kisiju forest (90 km from DES) was completely destroyed over the study period (Fig. 5.2e).

(Fig. 5.2c). Similarly above-ground carbon storage increased from 4 Mg carbon ha^{-1} (bootstrapped S.E. = ±2.84) nearest to DES to 52 Mg carbon ha^{-1} (bootstrapped S.E. = ±4.99) furthest from DES (Fig. 5.2b). A first-order estimate of the above-ground carbon lost from study area (258,000 ha) between 1991 and 2005 is ~0.21 Tg carbon yr^{-1} (total loss of above-ground carbon over the 33 closed-canopy forests in the study area (= 2.97 Tg) / 14 yrs), equivalent to over a quarter of the annual emissions of carbon from fossil fuel use in Tanzania over the same period (Boden *et al.* 2009)[2].

Discussion

The progressive decline in value of harvested woody resources at a given distance from DES over the past decade and increasing distance of transport for equivalent-value products over time, suggests an unsustainable 'logging down the profit margin' scenario akin to the sequential 'fishing down the food web' resource utilisation patterns seen in unmanaged marine habitats (Pauly *et al.* 1998). At current levels of demand and the outward expansion of the exploitation waves, we predict that there will be no high-value timber species remaining in the Tanzanian coastal forests up to 220 km from DES in early 2010 (Fig. 5.2e), and up to the southern Tanzanian border within 37 years[3]. A recently opened bridge across the Ruvuma River at the southern Tanzanian border will likely facilitate further encroachment of the degradation wave into Mozambique. In contrast, charcoal burning is predicted to continue to expand in line with urban demand and a lack of affordable alternatives (Kirilenko & Sedjo 2007), but may never reach the Tanzanian border because transport costs may make alternatives less expensive. However, new waves of charcoal extraction may expand from growing urban centres in the south of Tanzania.

Governance in the study forest reserves has been extremely poor: in 2005 alone the Tanzanian government lost estimated revenue of US$ 58 million due to illegal timber logging, with China importing tenfold more timber from Tanzania than declared **total** exports from this country (Milledge *et al.* 2007). The reserves in this study were mostly protective forest reserves (Table S5.1) and this relatively high protection status did not prevent the spread of degradation; however, degradation pressures were highest outside gazetted reserves (Kisiju, Fig. 5.2e).

---

[2] The estimated annual emission of carbon from fossil fuel use in Tanzania (Boden *et al.* 2009) presumably do not include emissions from charcoal burning.

[3] Inner wave boundary was at 170 km in 2005 and moves 9 km a^{-1}. Road distance between DES and the southern Tanzanian border is ~500 km.

Our Tanzania study provides three insights of importance to the debate on utilising payments to ensure that the decisions of economic actors in maximising the value from a given piece of land favours carbon storage and high biodiversity rather than an activity that releases carbon to the atmosphere and reduces biodiversity. First, deforestation and degradation waves should be seen in state-of-the art models of the ‘opportunity costs’ of avoiding deforestation and degradation, i.e. the foregone economic benefits from alternative land uses, if these models are capturing the complex real-world spatio-temporal patterns of carbon losses from forests. This may allow discrimination among models which give a three-fold difference in the costs of reducing deforestation by 10% by 2030 (Kaimowitz & Angelsen 1998), and would help reduce the uncertainty associated with the costs of post-Kyoto REDD schemes, proposed as an important mechanism within the United Nations Framework Convention on Climate Change to mitigate carbon dioxide emissions through conserving the tropical forests. Second, carbon fluxes from degradation are significant, suggesting that mitigating degradation, rather than merely avoiding deforestation should feature more strongly in ecosystem-service payment schemes such as proposed REDD schemes. Third, the use of models of degradation dynamics may allow the identification of areas to focus REDD policies and incentives so that they are tailored to specific areas where carbon is vulnerable to being lost, and are relevant to the type of degradation activity occurring (e.g. low-value timber *vs*. high-value timber). Achieving carbon dioxide emissions reductions in this way is also time-limited and economically irreversible, because once degradation has occurred it cannot then be avoided in the future.

Acknowledgements

The authors thank N. Doggart, C. Meshack, and P. Sumbi for assistance with fieldwork arrangements. A. Balmford, S. A. Bhagwat, P. M. Hollingsworth, J. C. Lovett, R. Marchant, A. R. Marshall, R. T. Pennington, D. Raffaelli and two anonymous reviewers provided helpful comments on the manuscript. Funding was provided by CEPF, WWF and DAAD (A.A.), the Royal Society (S.L.L.) and NERC (M.T.B.); data from the 1990/94 survey were provided by Frontier Tanzania.

References

Angelsen A. (2007). Forest Cover Change in Space and Time: Combining the von Thünen and Forest Transition Theories. *World Bank Policy Research Working Paper*, 4117.

Asner G.P., Knapp D.E., Broadbent E.N., Oliveira P.J.C., Keller M. & Silva J.N. (2005). Selective logging in the Brazilian Amazon. *Science*, 310, 480-482.

Bawa K.S. & Seidler R. (1998). Natural forest management and conservation of biodiversity in tropical forests. *Conservation Biology*, 12, 46-55.

Berkes F., Hughes T.P., Steneck R.S., Wilson J.A., Bellwood D.R., Crona B., Folke C., Gunderson L.H., Leslie H.M., Norberg J., Nystrom M., Olsson P., Osterblom H., Scheffer M. & Worm B. (2006). Ecology - Globalization, roving bandits, and marine resources. *Science*, 311, 1557-1558.

Berry N.J., Phillips O.L., Lewis S.L., Hill J.K., Edwards D.P., Tawatao N.B., Ahmad N., Magintan D., Khen C.V., Maryati M., Ong R.C. & Hamer K.C. (in press). The value of logged tropical forests: lessons from northern Borneo. *Biodiverstiy and Conservation*.

Boden T.A., Marland G. & Andres R.J. (2009). Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center.

Bryce J.M. (2003). *The commercial timber of Tanzania.* 3 edn. Tanzania Forestry Research Institute, Morogoro.

Burgess N.D. & Clarke G.P. (eds.) (2000). *Coastal Forests of Eastern Africa*. IUCN Publication Services Unit, Cambridge.

Canty A. & Ripley B. (2009). boot: Bootstrap R (S-Plus) Functions. R package version 1.2-38. Available at: http://CRAN.R-project.org/package=boot.

Chave J., Andalo C., Brown S., Cairns M.A., Chambers J.Q., Eamus D., Folster H., Fromard F., Higuchi N., Kira T., Lescure J.P., Nelson B.W., Ogawa H., Puig H., Riera B. & Yamakura T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145, 87-99.

Chave J., Coomes D., Jansen S., Lewis S.L., Swenson N.G. & Zanne A.E. (2009). Towards a worldwide wood economics spectrum. *Ecology Letters*, 12, 351-366.

Chomitz K.M. & Gray D.A. (1996). Roads, land use, and deforestation: A spatial model applied to belize. *World Bank Economic Review*, 10, 487-512.

Clarke G.P. & Dickinson A. (1995). Status Reports for 11 Coastal Forests in Coast Region, Tanzania. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Colwell R.K. (2006). EstimateS: Statistical estimation of species richness and shared species from samples. http://viceroy.eeb.uconn.edu/estimates

DeFries R.S., Rudel T., Uriarte M. & Hansen M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*. doi:10.1038/ngeo756

Ehrlich P.R. & Pringle R.M. (2008). Where does biodiversity go from here? A grim business-as-usual forecast and a hopeful portfolio of partial solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 11579-11586.

Gardner T.A., Barlow J., Chazdon R., Ewers R.M., Harvey C.A., Peres C.A. & Sodhi N.S. (2009). Prospects for tropical forest biodiversity in a human-modified world. *Ecology Letters*, 12, 561-582.

Geist H.J. & Lambin E.F. (2002). Proximate causes and underlying driving forces of tropical deforestation. *Bioscience*, 52, 143-150.

Gentry A.H. & Vasquez R. (1988). Where have all the ceibas gone - a case-history of mismanagement of a tropical forest resource. *Forest Ecology and Management*, 23, 73-76.

Gotelli N.J. & Colwell R.K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4, 379-391.

Hofstad O. (1997). Woodland deforestation by charcoal supply to Dar es Salaam. *Journal of Environmental Economics and Management*, 33, 17-32.

Johns A.G. (1997). *Timber production and biodiversity conservation in tropical rain forests*. Cambridge University Press, Cambridge.

Kaimowitz D. & Angelsen A. (1998). *Economic models of tropical deforestation: a review*. Center for International Forestry Research, Bangor.

Kirilenko A.P. & Sedjo R.A. (2007). Climate change impacts on forestry. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 19697-19702.

Laurance W.F. (2008). The need to cut China's illegal timber imports. *Science*, 319, 1184-1184.

Lewis S.L., Lopez-Gonzalez G., Sonke B., Affum-Baffoe K., Baker T.R., Ojo L.O., Phillips O.L., Reitsma J.M., White L., Comiskey J.A., Djuikouo M.N., Ewango C.E.N., Feldpausch T.R., Hamilton A.C., Gloor M., Hart T., Hladik A., Lloyd J., Lovett J.C., Makana J.R., Malhi Y., Mbago F.M., Ndangalasi H.J., Peacock J., Peh K.S.H., Sheil D., Sunderland T., Swaine M.D., Taplin J., Taylor D., Thomas S.C., Votere R. & Woll H. (2009). Increasing carbon storage in intact African tropical forests. *Nature*, 457, 1003-1007.

Liu J.G. & Diamond J. (2005). China's environment in a globalizing world. *Nature*, 435, 1179-1186.

Luoga E.J., Witkowski E.T.F. & Balkwill K. (2000). Economics of charcoal production in miombo woodlands of eastern Tanzania: some hidden costs associated with commercialization of the resources. *Ecological Economics*, 35, 243-257.

Miles L. & Kapos V. (2008). Reducing greenhouse gas emissions from deforestation and forest degradation: Global land-use implications. *Science*, 320, 1454-1455.

Milledge S.A.H., Gelvas I.K. & Ahrends A. (2007). Forestry, Governance and National Development: Lessons Learned from a Logging Boom in Southern Tanzania. TRAFFIC East/Southern Africa, Tanzania Development Partners Group, Tanzania Ministry of Natural Resources and Tourism, Dar es Salaam.

Mittermeier R.A., Robles-Gil P., Hoffmann M., Pilgrim J.D., Brooks T.M., Mittermeier C.G., Lamoreux J.L. & Fonseca G. (2005). *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Ecoregions*. 2 edn, Cemex, Mexico City.

National Bureau of Statistics Tanzania (2002). Population Census of 2002. Dar es Salaam.

Nepstad D.C., Verissimo A., Alencar A., Nobre C., Lima E., Lefebvre P., Schlesinger P., Potter C., Moutinho P., Mendoza E., Cochrane M. & Brooks V. (1999). Large-scale impoverishment of Amazonian forests by logging and fire. *Nature*, 398, 505-508.

Olson D.M. & Dinerstein E. (2002). The Global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden*, 89, 199-224.

Pauly D., Christensen V., Dalsgaard J., Froese R. & Torres F. (1998). Fishing down marine food webs. *Science*, 279, 860-863.

Putz F.E., Zuidema P.A., Pinard M.A., Boot R.G.A., Sayer J.A., Sheil D., Sist P., Elias & Vanclay J.K. (2008). Improved tropical forest management for carbon retention. *Plos Biology*, 6, 1368-1369.

R Development Core Team (2009). *R: A language and environment for statistical computing*, Vienna. Available at: www.R-project.org.

Scales H., Balmford A., Liu M., Sadovy Y. & Manica A. (2006). Keeping bandits at bay? *Science*, 313, 612-613.

Stattersfield A.J., Crosby M.J., Long A.J. & Wege D.C. (1998). *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation*. BirdLife International, Cambridge.

Tang L., Shao G., Piao Z., Dai L., Jenkins M.A., Wang S., Wu G., Wu J. & Zhao J. (in press). Forest degradation deepens around and within protected areas in East Asia. *Biological Conservation*.

The United Republic of Tanzania (2004). Subsidiary Legislation. Includes: Notice of the Commencement Date and Regulation of the Forest (Act, No. 14 of 2002). Government Printer, Dar es Salaam.

UN-HABITAT (2008). The State of African Cities 2008. UN-HABITAT, Nairobi.

Vasquez R. & Gentry A.H. (1989). Use and misuse of forest-harvested fruits in the Iquitos area. *Conservation Biology*, 3, 350-361.

von Thünen J.H. (1966). Der Isolierte Staat in Beziehung der Landwirtschaft und Nationalökonomie. In: *Von Thünen's Isolated State* (ed. Hall P). Pergamon Press, Oxford.

Willott S.J. (1999). The effects of selective logging on the distribution of moths in a Bornean rainforest. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 354, 1783-1790.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. (2009). *Mixed Effects Models and Extension in Ecology with R*. Springer, New York.

Supporting information

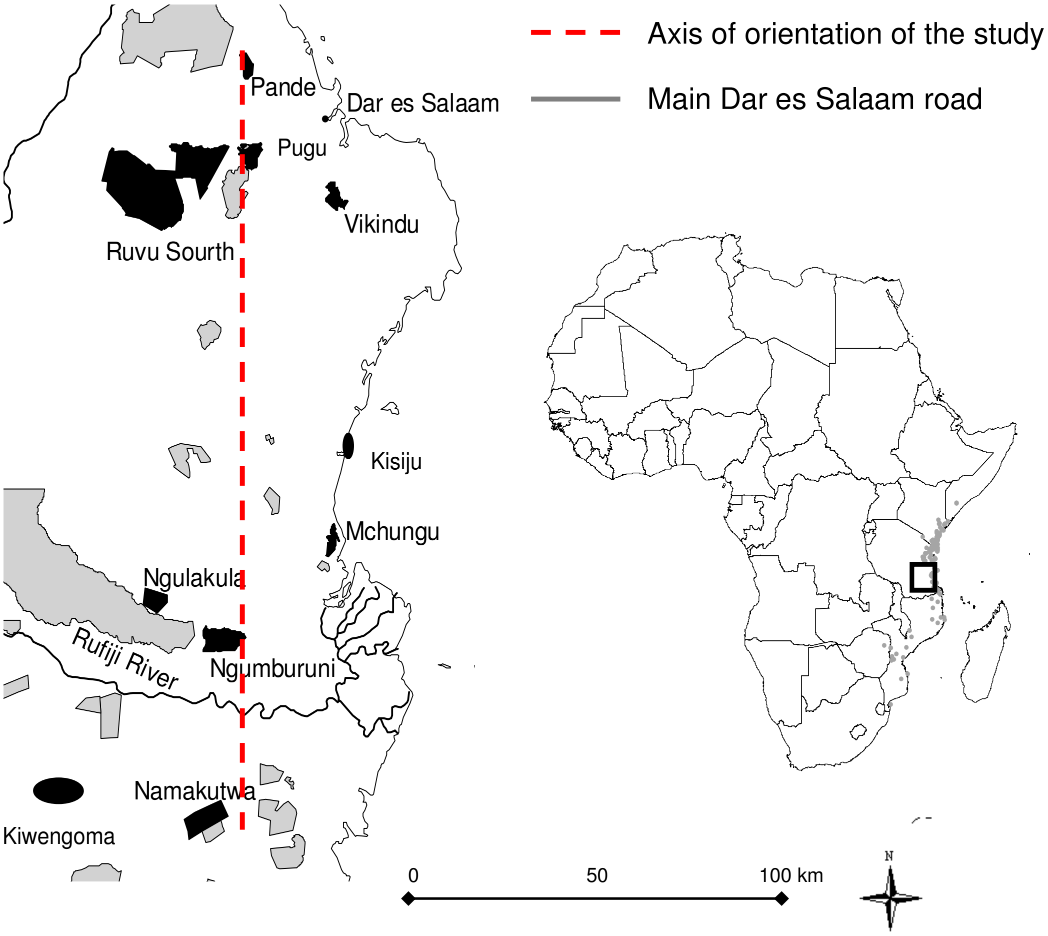


Figure S5.1. Location of the study area in Tanzania. The study area stretches from DES to south of the Rufiji River. Solid black polygons are sampled forests and grey polygons are unsampled forest. Grey dots on the Africa map show the location of remaining related East African coastal forests.

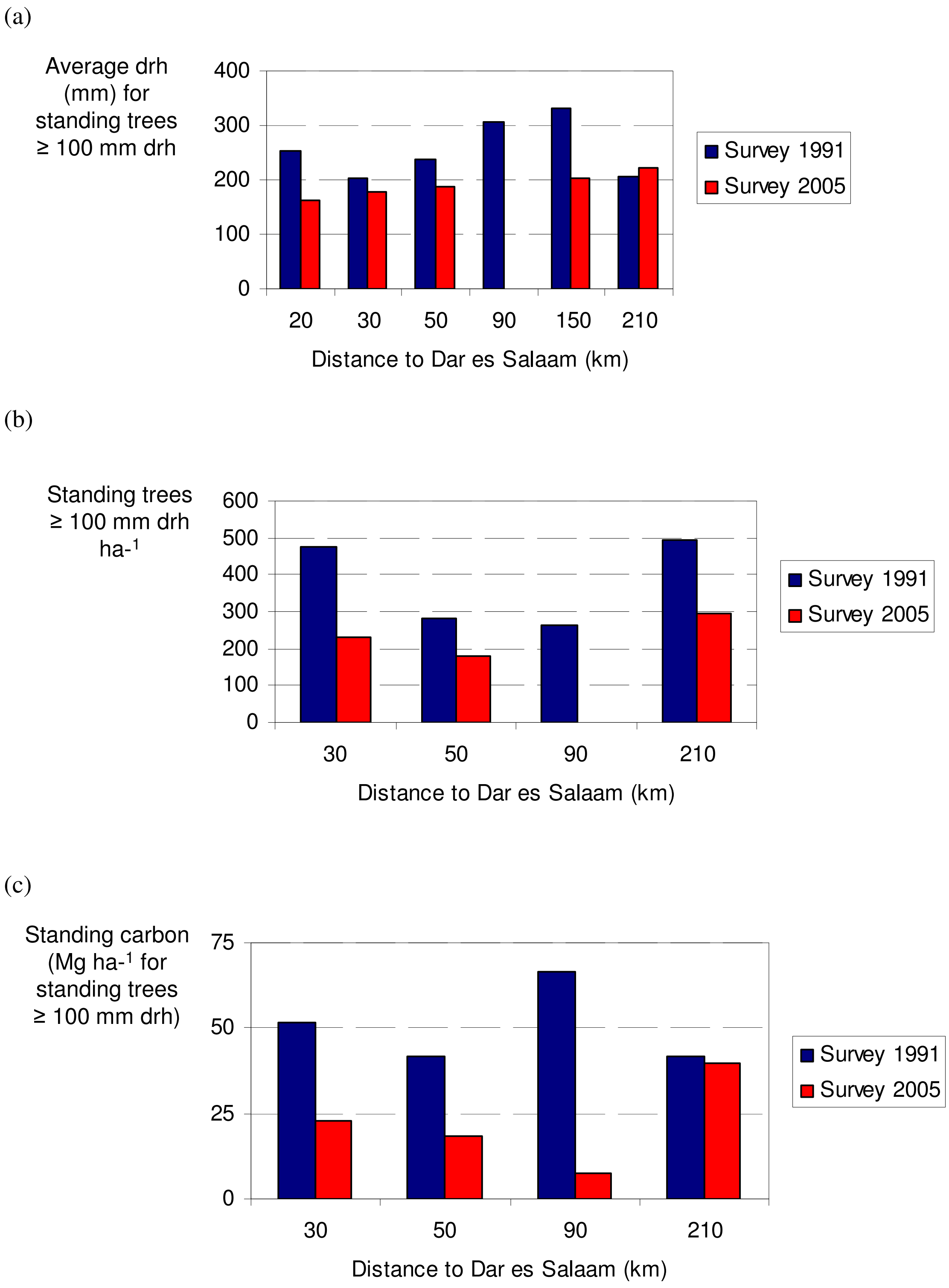


Figure S5.2. Changes in stem diameter (drh) (a), stem density ha^{-1} (b) and standing carbon ha^{-1} (c) for trees ≥ 100 mm drh between 1991 and 2005 in coastal forests located between Dar es Salaam (DES) and the Rufiji river with matched datasets. In 1991, drh, standing trees and carbon differ between forests due to differences in vegetation type or forest history. In 2005, patterns in these forest structure variables are driven by distance from DES. The forest at 90 km distance (Kisiju) has been entirely cleared for agriculture in 2005, and Namakutwa at 210 km distance has been cleared for farming c. 100 years ago (Clarke 1992) and is regenerating.

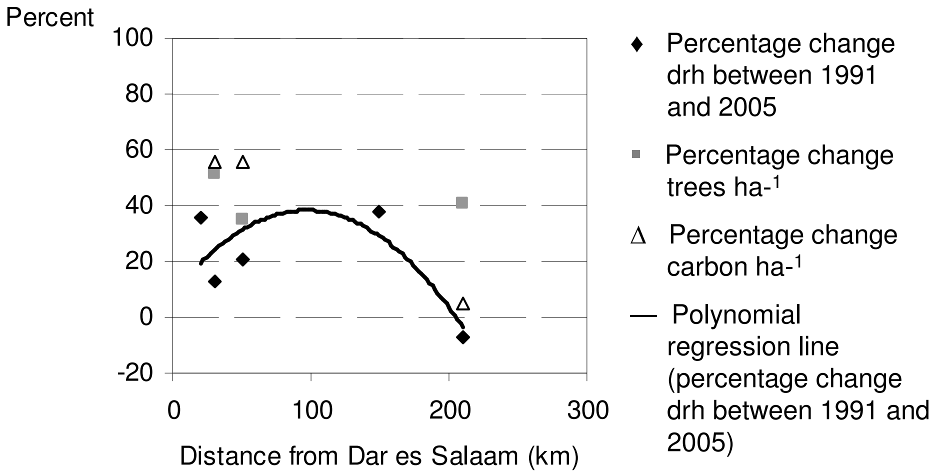


Figure S5.3. Percentage change in stem diameter (drh), standing trees ha^{-1} and standing carbon ha^{-1} for trees ≥ 100 mm drh between 1991 and 2005 in forests with matched datasets with a fitted polynomial regression line for percentage change in drh (r^2 = 0.56). Greatest percentage in drh occurred at ~100 km distance from DES. Closer to DES the percentage change is lower as the forests had already been degraded in 1991, and at ≥ 100 km distance from DES the percentage change is lower as there degradation has not yet resulted in significant changes in the forest structure. Note that Kisiju at 90 km distance is an outlier and has not been included in this analysis. The percentage change in standing trees and carbon ha^{-1} could not be fitted due the low number of available data points.

	Forests									
	Vikindu	Pande	Pugu	Ruvu South	Kisiju	Mchungu	Ngulakula	Ngumburuni	Namakutwa	Kiwengoma
District	Kisarawe	Kinondoni	Kisarawe	Kisarawe/ Kibaha	Mkulanga	Rufiji	Rufiji	Rufiji	Rufiji	Rufiji
Size in km²	17.96	12.26	21.79	350.00	2.00	10.35	22.00	40.00	46.34	20.25
Legal status[4]	Prod. FR	GR	Prot. FR	Prot. FR	No gaz.	Prot. FR	Prot. FR	Prot. FR	Prot. FR	Prot. FR
Number of 10 × 50 m sections	8	63	55	69	14	20	3	59	36	47
Percent of total area sampled[5,6]	0.02[7]	0.22	0.13	0.15	0.35	0.10	0.01	0.15	0.15	0.12

Table S5.1. General information and sampling intensity across the 10 study forests in coastal Tanzania sampled in 2005.

---

[4] Prod. = Productive; Prot. = Protective; FR = Forest Reserve; GR = Game Reserve; No gaz = No gazettment (i.e. general land).

[5] In forests with an area larger than 40 km², sampling was confined to randomly chosen regions within the forest, and these were sampled with the above intensity.

[6] Kisiju and Ngulakua FR were excluded from the analysis. Kisiju is not a protected area and has almost entirely been converted into agricultural land. Sampling intensity in Ngulakula was too low.

[7] Sampling intensity low, but representative as large parts of Vikindu FR were converted into agricultural land and the actual area covered by forest remnants is small.

Category	Dependent variable	Model	*p*	*F*	df	Adj. r^2
Forest use	Average stump value (US$ (2005))	y = 59.091 - 17.273 * Access1 – 26.031 * Access2 – 31.136 * Access3	≤ 0.05	7.22	4	0.727
	Average stump diameter (mm)	y = 86.267 + 0.685 * DisDES	≤ 0.01	25.03	6	0.774
Remaining woody resources	Standing timber (n ha^{-1})	y = 44.4 + 0.772 * DisDES	≤ 0.01	17.97	6	0.708
	Basal area/ha (m^2 ha^{-1})	y = 3.744 + 0.064 * DisDES	≤ 0.01	15.92	6	0.681
Carbon stock	Standing carbon (Mg ha^{-1})	y = 9.069 + 0.203 * DisDES	≤ 0.05	10.93	6	0.587
Remaining biodiversity	Mao Tau Index	y = 7.061 + 0.081 * DisDES	≤ 0.01	25.19	6	0.776
	Chao 1 Index	y = 11.942 + 0.111 * DisDES	≤ 0.01	16.9	6	0.694
	Jaccard 2 Index	y = 16.665 + 0.135 * DisDES	≤ 0.01	14.32	6	0.656

Table S5.2. Multiple regression model results for spatial degradation patterns in 2005.

Table S5.3

Forest	Past reports	Past findings on timber values	This study's findings on timber values	Past extraction activities	Present extraction activities
Vikindu	(Hawthorne 1984; Clarke & Dickinson 1995)	Hardly any timber species; few individuals of *Afzelia quanzensis* (II) and *Khaya anthotheca* (I).	No timber species.	Clearing and afforestion with exotic species. Pole cutting. Charcoal burning.	Pole cutting and charcoal burning.
Pande	(Mwasumbi et al. 1994; Clarke & Dickinson 1995)	Few timber species in low densities; few individuals of *Afzelia quanzensis* (II), *Brachylaena huillensis* (I) and *Bombax rhodognaphalon* (IV) and very few individuals of *Milicia excelsa* (I).	Hardly any timber species; only very few individuals of *Bombax rhodognaphalon* (IV)	Logging and charcoal burning.	Pole cutting and charcoal burning (albeit not recent).
Pugu	(Howell 1981; Clarke & Dickinson 1995)	Most timber species had already been removed during colonial era (including e.g. *Khaya anthotheca* (I) and *Milicia*	Low-value timber species such as *Antiaris toxicaria* (V) left in the vicinity of the District Forest	Logging and charcoal burning.	Charcoal burning.

Table S5.3

Forest	**Past reports**	**Past findings on timber values**	**This study's findings on timber values**	**Past extraction activities**	**Present extraction activities**
		excelsa (I)); low-value timber species such as *Antiaris toxicaria* (V) left.	Office in an area that regularly patrolled. In other parts of the reserve most timber trees removed.		
Ruvu South	(Clarke & Dickinson 1995)	Only few timber species, e.g. logging of *Brachylaena huillensis* (I) occurred and suitable trees were already scarce.	No individuals of *Brachylaena huillensis* (I) were found; a number of individuals of *Julbernardia globiflora* (II) had been harvested relatively recently.	Logging. Charcoal burning not yet perceived as a threat.	Logging and charcoal burning.
Kisiju	(Hawthorne 1984; Stubblefield 1992)	Timber species such as *Afzelia quanzensis* (II), *Baphia kirkii* (II) and *Hymenaea verrucosa* (V) were abundant.	A single forest patch of 0.005 km^2 remained. Extremely poor and landless squatters have logged, burned to charcoal and cultivated the entire forest area.	Agricultural encroachment and pole cutting.	Clear-felling, charcoal burning and pole cutting followed by agricultural encroachment.
Mchungu	(Waters & Burgess 1994; Clarke & Dickinson 1995)	Timber species such as *Afzelia quanzensis* (II), *Baphia kirkii* (II) and *Hymenaea verrucosa* (V) were abundant.	Timber species such as *Afzelia quanzensis* (II), *Baphia kirkii* (II) and *Hymenaea verrucosa* (V) were abundant.	With the exception of few trees felled for local use, none reported.	Logging (recent).
Ngumburuni	No assessment.	No assessment.	Two stumps of high-value species were found: *Pterocarpus angolensis* (I) and *Diospyros mespiliformis* (I). Of these species, no sizeable standing trees were found. Stumps indicated that during the last 10 years harvesting had focused on *Afzelia quanzensis* (II), *Baphia kirkii* (II), and *Pteleopsis myrtifolia* (IV).	No assessment.	Logging and charcoal burning.
Namakutwa	(Clarke 1992; Clarke & Dickinson	Past extraction of timber species such as *Milicia excelsa* (I) and *Pterocarpus angolensis*	No individuals of *Milicia excelsa* (I) and *Pterocarpus angolensis* (I) were	Logging.	Logging.

Table S5.3

Forest	**Past reports**	**Past findings on timber values**	**This study's findings on timber values**	**Past extraction activities**	**Present extraction activities**
	1995)	(I) had reduced the economic value of the forest, however, a few *Milicia excelsa* individuals (I) remained, and *Bombax rhodognaphalon* (IV) and *Hymenaea verrucosa* (V) were abundant.	found. Stumps indicated that during the last 10 years harvesting had focused on *Afzelia quanzensis* (II), *Albizia versicolor* (II), *Julbernardia globiflora* (II), *Millettia stuhlmannii* (II) and *Pterocarpus angolensis* (I). Of these, hardly any sizeable trees remained in the forest.		
Kiwengoma	(Sheil & Burgess 1990; Waters & Burgess 1994; Clarke & Dickinson 1995)	Moist forest with a high proportion of valuable timber species, notably *Khaya anthotheca* (I), *Pterocarpus tinctorius* (I) and *Milicia excelsa* (I).	*Milicia excelsa* (I) was not found. Recent logging of large individuals of *Khaya anthotheca* (I) and *Pterocarpus tinctorius* (I) in moist forest, and *Pterocarpus angolensis* (I) and *Albizia versicolor* (II) in woodlands had reduced the number of large and high-value timber trees. Timber species with less value such as *Afzelia quanzensis* (II) and *Hymenaea verrucosa* (V) were found in large quantities.	Logging of high-value species such as *Khaya anthotheca* (I) and *Milicia excelsa* (I).	Logging (recent) of high and medium-value species such as *Khaya anthotheca* (I), *Pterocarpus tinctorius* (I) and *Albizia versicolor* (II).

Table S5.3. Changes in forest composition since surveys in the 1990s and the repeat survey of this paper in 2005. Timber classes are in brackets after the species' names. Throughout the study area, high-value Class I timber species such as *Milicia excelsa* (Welw.) C.C.Berg and *Brachylaena huillensis* O. Hoffm., which were present in the 1990s, have been logged to exhaustion and are no longer recorded; and other high-value species such as *Pterocarpus angolensis* DC. and *Khaya anthotheca* (Welw.) C.DC. are now rare. Class II timber species, such as *Afzelia quanzensis* Welw., have also been locally depleted. The majority of recently harvested timber species were *Baphia kirkii* Bak. and *Hymenaea verrucosa* Gaertn., which are Class II and V timber species respectively. We also found changes in the sizes of harvested trees: There was an average decrease of 200 mm drh between old and recent stumps of Class I and II timber species. These dramatic shifts in harvested species and harvest areas have also been documented with trade data (Milledge & Kaale 2003; Milledge *et al.* 2007).

Supporting references

Clarke G.P. (1992). Site description and conservation evaluation: Namakutwa-Nyamuete Forest Reserve, Rufiji District, Tanzania. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Clarke G.P. & Dickinson A. (1995). Status Reports for 11 Coastal Forests in Coast Region, Tanzania. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Hawthorne W.D. (1984). Ecological and Biogeographical Patterns in the Coastal Forests of East Africa. University of Oxford, Oxford.

Howell K.M. (1981). Pugu Forest Reserve - Biological Values and Development. *African Journal of Ecology*, 19, 73-81.

Milledge S.A.H., Gelvas I.K. & Ahrends A. (2007). Forestry, Governance and National Development: Lessons Learned from a Logging Boom in Southern Tanzania. TRAFFIC East/Southern Africa, Tanzania Development Partners Group, Tanzania Ministry of Natural Resources and Tourism, Dar es Salaam.

Milledge S.A.H. & Kaale B.K. (2003). Bridging the Gap: Linking Timber Trade with Infrastructure Development and Poverty Eradication Efforts in Southern Tanzania. TRAFFIC East/Southern Africa, Dar es Salaam.

Mwasumbi L.B., Burgess N.D. & Clarke G.P. (1994). Vegetation of Pande-Kiono Coastal Forests, Tanzania. *Vegetatio*, 113, 71-81.

Sheil R.D. & Burgess N. (1990). Preliminary results of biological surveys in Zaraninge (Kiono) and Kierengoma (Matumbi Hills) Coastal Forests, Tanzania: January to March 1990. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Stubblefield L.K. (1992). Preliminary Results of a Biological Value Survey of Kisiju – Dendene Forest, Tanzania: November to December 1992. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Waters T. & Burgess N. (1994). Preliminary results of biological surveys of Mchungu and Kiwegnoma (Matumbi) forests, and short visits to seven other forested sites in coastal Tanzania: July to September 1990. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Chapter 6 – Conclusions

Parts of this chapter in prep. for:

Ahrends, A., Jump, A.S., Hollingsworth, P.M., Lovett, J.C., Burgess, N.D., Gereau, R.E., Bulling, M.T. & Marchant, R. ‘Major gaps in our knowledge of plant biodiversity in the Eastern Arc Mountains: implications for conservation and future research’ in prep. for *Journal of the East African Natural History Society*.

This thesis analysed the impact of three fundamental challenges to conservation planning: biases in the biological data that underpin conservation decisions, lack of clarity over the most efficient survey design which may lead to insufficient field data collection, and lack of spatially explicit and predictive data on degradation. Below I discuss each of them in turn in light of the findings from the individual chapters, and then summarise the implications for future research and conservation.

Data biases

This work found a number of biases in the data available for the study area. Firstly, plant survey effort is heavily spatially biased (Fig. 6.1), and a significant relationship between survey effort and species endemism has been demonstrated for both animals (Doggart *et al.* 2006) and plants (Chapter 2).

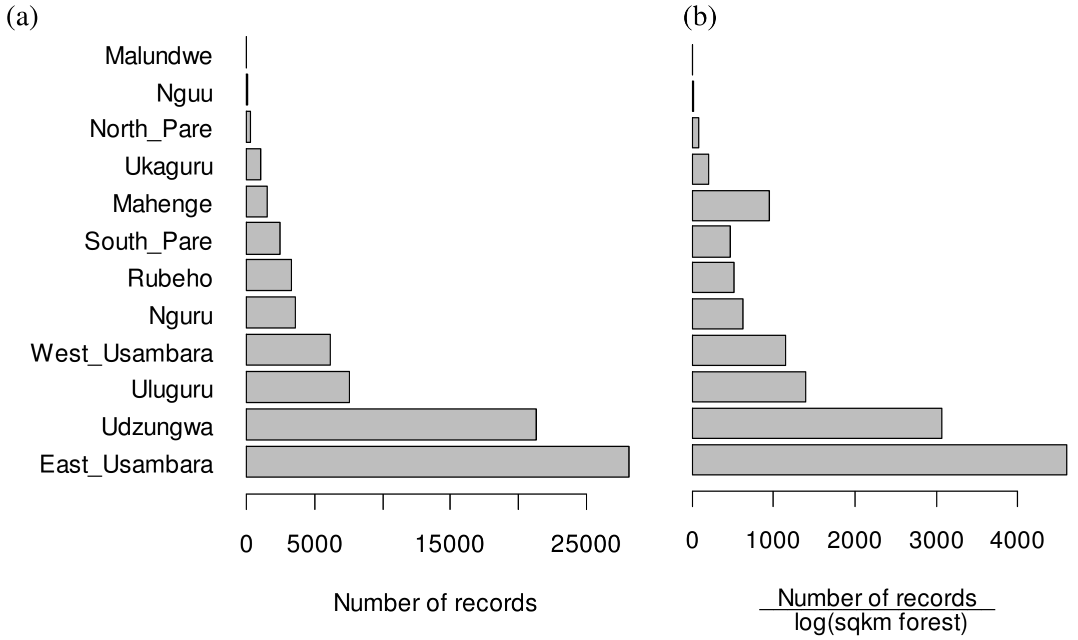


Figure 6.1. Plant record collection intensity across the Eastern Arc Mountain blocs. (a) Total number of records. (b) Number of records standardised by log (area).

There is also evidence that some of these spatial biases originate from circularities between funding and perceived biodiversity, whereby surveys tend to focus on areas already known to be species rich, resulting in the discovery of more species from these areas and reinstating their conservation status, while other areas with little or no initial exploration may remain overlooked (Fig. 6.2; Chapter 2). Furthermore, areas perceived as important for plant biodiversity may have become ‘magnets’ for expert botanists as these are the areas where the majority of investment is made and employment found, and because these areas are perceived as botanically interesting (Fig. 6.2). Analyses in Chapter 3 suggest that the resulting unequal spatial distribution of taxonomic knowledge may in turn lead to biases in biodiversity inventories. Sampling in under-researched areas was dominated by volunteers and semi-professionals. In the North Pare and Rubeho Mountains botanists with a training of less than 6 months collected over 75% of all specimens, and in the Nguu and Ukaguru Mountains almost 100%, i.e. a third of the Eastern Arc Mountain (EAM) blocs have scarcely been visited by experts. Consequently, funding and associated survey intensity and botanist effects explain over 60% of the variation in plant diversity patterns on their own, and environmental variables just over 10% (Chapters 2 and 3), with time series analyses suggesting that funding is driving perceived biodiversity and not *vice versa* (Chapter 2).

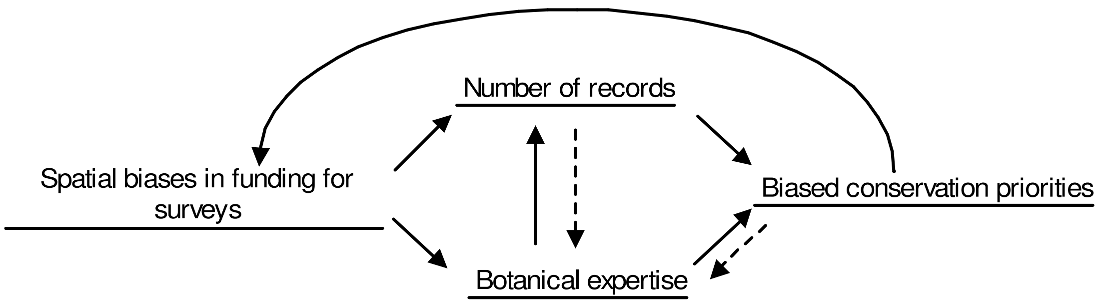


Figure 6.2. Suggested relationships between funding, number of records, botanic expertise and conservation priorities.

Conservation planning based on such data may therefore be flawed, in particular because investments are influenced by factors beyond perceived biodiversity priorities, e.g. by accessibility, ease of running a project, historical relationships, political interest and stability, geographical focus of organisations, donor wishes, or simple opportunity (Halpern *et al.* 2006). Thus, some of today's conservation priorities may originally have been the focus of investment due to reasons other than biodiversity value, which led to spiralling increases in perceived conservation interest of particular areas, potentially over-shadowing other equally or more important areas. This is very likely the case in the EAM where conservation priorities have tended to change as new data from more difficult to access mountain blocs have become available (e.g. Doggart *et al.* 2006; Menegon *et al.* 2009). Regional scale conservation priorities are potentially more strongly affected by such biases than those at the global scale as regional assessments are increasingly based on actual survey data rather than expert opinion (Brooks *et al.* 2006). However, expert opinion will also be influenced by the current state of knowledge, and experts may make inaccurate estimates. For example, levels of plant endemism in the study area have previously been estimated to be 50% (1,500 endemic species and a total of 3,000 species), resulting in the area's designation as a global hotspot (Myers *et al.* 2000); however, at the time floristic inventories were concentrated on the 'hotspots' within the EAM (Chapter 2). Since then many putative

EAM endemics have been discovered in other biogeographic areas (Hemp 2006), and a comprehensive review of the endemism status of ~4,000 EAM species indicates that endemism levels are 11.8% (409 species of 3,475 species recorded in the EAM with a strict lower altitudinal limit of ≥ 500 m; Roy Gereau, unpublished data). These spatial biases may also extend beyond the scale of the study area: within sub-Saharan Africa, for example, Angola, the Central African Republic, the Democratic Republic of Congo, Somalia and Sudan are vastly under-collected due to difficult research conditions, and the conservation value of many of these areas is unknown (Küper *et al.* 2004; Küper *et al.* 2006). It may not be surprising that with the exception of the Horn of Africa no area in these countries has been designated as a biodiversity hotspot (Mittermeier *et al.* 2005). Thus the biases in biodiversity estimates found in the study due to investment biases and circularities may also affect global conservation priorities.

Such biases will further compromise ecological studies, and may partly explain why despite decades of research there is surprisingly little consensus on species richness and endemism patterns and causes (Rahbek 2005; Gotelli *et al.* 2009). In the EAM, for example, it seems that levels of endemism are highest in the Usambara Mountains (Lovett 1993; Scholes *et al.* 2006; Platts *et al*. 2010), which are situated close to the Indian Ocean. This pattern has been attributed to the putative stabilising influence of the Indian Ocean monsoon system on the region's coastal climate (Hamilton 1981; Fjeldså *et al.* 1997; Fjeldså & Lovett 1997). Uninterrupted moisture availability is hypothesised to have enabled continuous forest cover even during glacial periods. Furthermore, the complex topography of the EAM may have allowed altitudinal movement of taxa, minimising climatically linked extinctions (Lovett *et al.* 2005). However, much of the supporting evidence for these hypothesises is largely circumstantial (Burgess *et al.* 2007). At present, only two palaeoecological records that date past the last glacial maximum exist for these mountains: from Dama Swamp in the Udzungwa Mountains (Mumbi *et al.* 2008), covering ~24,000 years; and from the Lukwangule Plateau in the Uluguru Mountains (Finch *et al.* 2009), covering ~48,000 years. Both indicate relatively stable forest composition, and therefore do not provide an explanation for vastly differing levels of endemism and species richness across the individual EAM blocs. Thus, there is not yet sufficient evidence to reject the hypothesis that mountain blocs close to the Indian Ocean are perceived as more biodiverse simply because proximity to the Indian Ocean also means proximity to Dar es Salaam, infrastructure and less strenuous fieldwork conditions. It is important to note that this does not bring into question the undoubted conservation importance of such areas nor the merits of the stability hypothesis, but it does advocate for a transparent presentation of potential data problems when heavily biased data are used to derive ecological theories and the associated conservation implications (e.g. Lovett *et al.* 2000).

Data biases may also occur due to ‘taxonomic inflation’ (Isaac *et al.* 2004), whereby some groups (Alroy 2003) and areas have received more attention from phylogeographic studies, and the associated description of ‘new’ species using the phylogeographic species concept “*makes hotspots appear even hotter*” (Isaac *et al.* 2004). While the description of new plant species is still heavily based on morphological characteristics (McDade 1995; Knapp *et al.* 2005), i.e. the taxonomic inflation is less severe than in charismatic large vertebrate groups (Mallet *et al.* 2005), biases are still present. In the EAM, for example, plant phylogenetic research has been concentrated in the East Usambaras and Ulugurus to differentiate between morphologically poorly resolved species of *Saintpaulia* (Möller & Cronk 1997b; Möller & Cronk 1997a; Lindqvist & Albert 1999, 2001). Such spatial and taxonomic biases in the applied species concept will severely compromise comparisons of species diversity (Mace 2004).

In the absence of sufficient data for conservation planning, climatic species distribution models are appealing tools to help guide conservation planning (Da Fonseca *et al.* 2000; Ferrier *et al.* 2004; Graham *et al.* 2004; Rondinini *et al.* 2006). However, such models are biased by their input data (Platts *et al*. 2010), i.e. in a situation of severe inequalities in the spatial distribution of survey effort they are unlikely to capture the true distributional patterns. Figures 6.3 and 6.4 show that while data collections in the EAM have sampled altitude and mean temperature relatively well, there are severe biases with respect to annual precipitation and moisture index whereby collections have predominantly focused on areas that receive more moisture. It may be possible to account for biases, e.g. by adding a variable to the model that reflects the bias (offset), or by using a mixed model whereby the source of bias is fitted as a random effect, allowing the model intercept and/or slope to vary according to e.g. observer skill or level of investment (Zuur *et al.* 2009). However, such an approach relies on consistent and thorough documentation of potential biases in biodiversity survey data, particularly for supra-regional or global data collations such as the Global Biodiversity Information Facility (GBIF, http://www.gbif.org/), where there is no direct contact any more between the data collector and the data analyst.

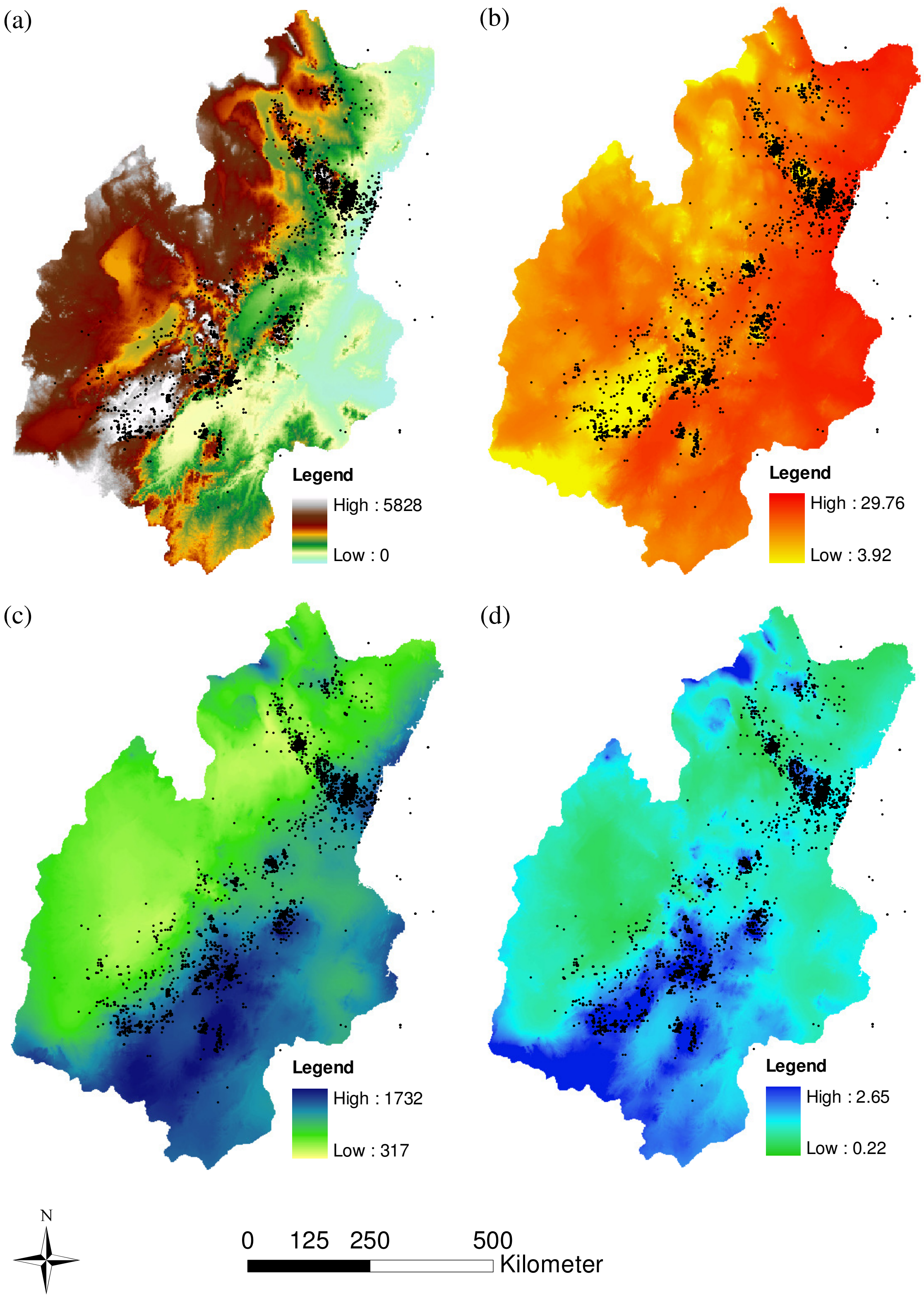


Figure 6.3. Plant collections in the study area (black dots) across elevation (m) (a), mean temperature (˚C) (b), annual rainfall (mm) (c) and annual moisture index (d). Elevation data is based on a high-resolution (~90 m) digital elevation model (SRTM DEM version 4, http://srtm.csi.cgiar.org/), and climate on surfaces supplied by the Centre for Resource and Environmental Studies (CRES), Australian National University at a resolution of 3 arc minutes (~5.5 km) (http://fennerschool.anu.edu.au/), downscaled by Phil Platts, University of York, to a resolution of 1 km. Calculation of the annual moisture index follows Platts *et al.* (2008).

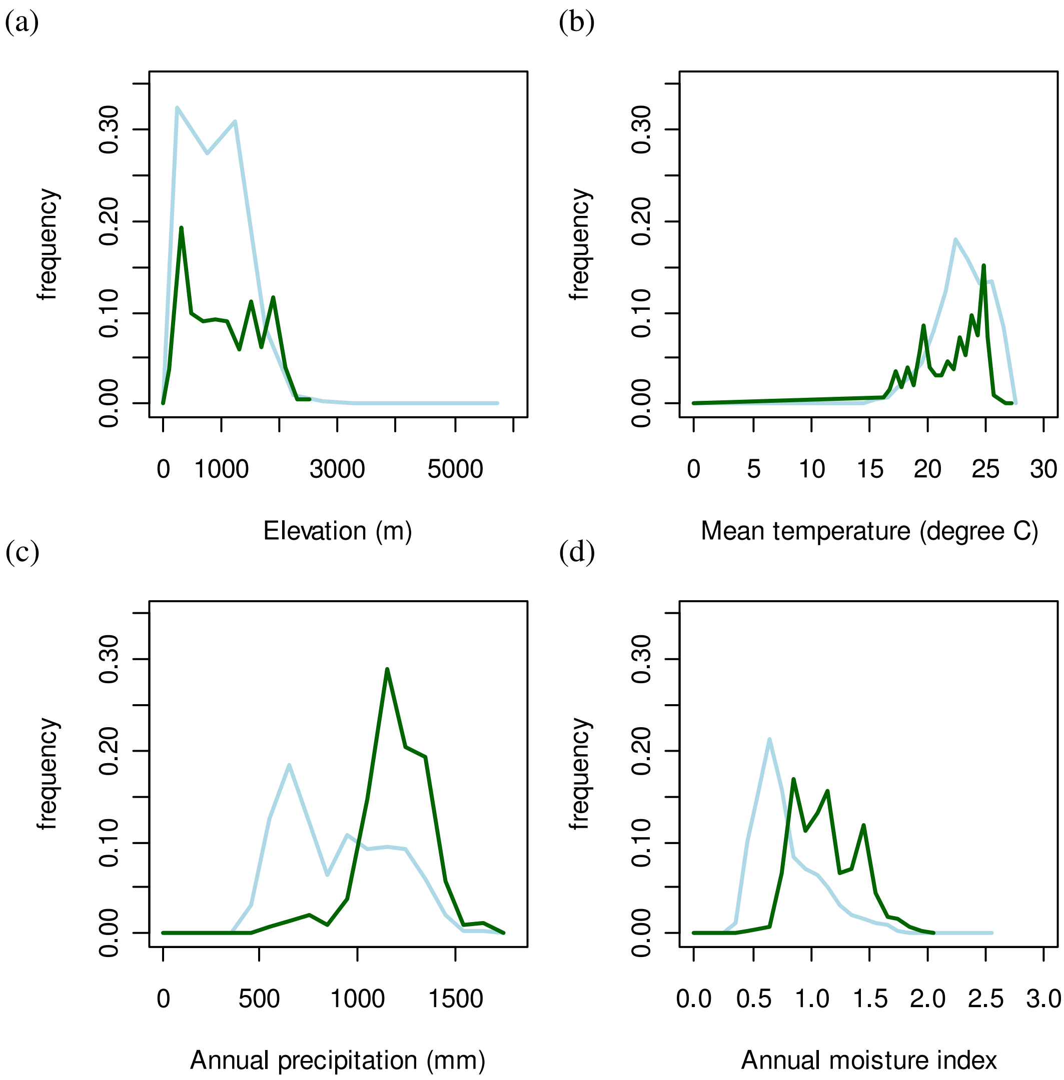


Figure 6.4. Frequency proportion plots of environmental characteristics over a 2 km × 2 km resolution grid covering the study area. Light blue lines represent all cells within the study area and dark green lines all cells that have been sampled. The area underneath each curve sums to one.

Future research collections offer the opportunity to address current data biases through the strategic collection of new field data. On average, between 1980 and 2007, a botanical survey in the EAM cost US$ 1,482 (all US$ values standardised to year 2007) per day; with the average cost per day decreasing for longer surveys (e.g. > 30 days US$ 1,250) and increasing again for very long surveys (e.g. > 100 days US$ 1,640). This may seem expensive; however, these figures are total costs, including international flights, salaries, overheads and herbarium identification. Investments for plant biodiversity surveys account for less than 2% (and overall biodiversity surveys for 4%) of the total investments (> US$ 117 million) in conservation and research projects in the EAM between 1980 and 2007. Of these 2%, almost 75% have been invested in the heavily researched mountain blocs (Usambara, Uluguru and Udzungwa Mountains). In the face of continued and rapid global biodiversity loss we cannot afford to delay conservation action until better survey data become available (Fuller *et al.* 2007; McDonald-Madden *et al.* 2008; Grantham *et al.* 2009), but a more balanced distribution of inventory funds between areas may significantly improve the efficiency and efficacy of conservation by providing more reliable data on which conservation decisions can be based. For example, an investment of US$ 490,000 (< 1% of total investments between 1980 and 2007) for biodiversity inventories in the six least researched mountain blocs (Mahenge, Nguru, Nguu, North Pare, Rubeho, and Ukaguru) between 2000 and 2007 resulted in the discovery of numerous new species across all surveyed taxa (Menegon *et al.* 2004; Channing *et al.* 2005; Menegon *et al.* 2008; Rovero *et al.* 2008) (including 477 new species of plant for the region, 29 of which were potentially threatened and/or endemic), and altered existing conservation priorities (Doggart *et al.* 2006; Menegon *et al.* 2009). This shows that in the EAM and potentially elsewhere the knowledge base for conservation planning could be significantly improved by investing just a fraction of the available survey funds into research of under-explored areas - without any extra expenses occurred or conservation action delayed.

Lack of clarity on efficient protocols and insufficient data collection

There is a "*need for speed*" in biodiversity assessments and conservation decisions (McDonald-Madden *et al.* 2008) but results in Chapter 4 indicate that, irrespective of biases, failure to fulfil a certain minimum sampling intensity may lead to wrong conclusions about the nature of the species richness gradient and flawed assessments of biodiversity and conservation priorities. Sacrificing accuracy for speed is therefore likely to be counter-productive (Gordon & Newton 2006b; McDonald-Madden *et al.* 2008). Studies on altitudinal species richness gradients in the EAM, for example, largely based on variable area plots measuring the 20 nearest trees of ≥ 200 mm diameter at reference height (1.3 m along the stem or above buttresses; drh) (area sample size on average c. 1,000 m^2), concluded that there is no change in species richness with altitude (Lovett 1996, 1999; Lovett *et al.* 2006) – a gradient that would be new to science (Nogues-Bravo *et al.* 2008). However, simulations in Chapter 4 show that this apparent pattern could be an artefact of insufficient sampling (Fig. 4.3). Even in the most optimistic scenarios implemented in the simulation, a minimum sample size of 1,000 to 2,000 m^2 was needed to capture the true species richness pattern, and this is often not achieved in the field. Consequently, differences in sampling protocol may well be a cause for the wide range of proposed species richness patterns.

A lack of consensus on optimal sampling strategies for biodiversity studies in tropical forests (Gimaret-Carpentier *et al.* 1998; Phillips *et al.* 2003; Gordon & Newton 2006a; Gordon & Newton 2006b) means that a wide range of different methods are employed, with it being unclear whether the data are comparable and sufficiently representative to capture the true pattern (Gordon & Newton 2006b). Chapter 4 shows that as long as sampling intensity is large enough, sample design and other confounding factors have minimal influence on the outcome, i.e. patterns (not species numbers) established with different sample designs and in different areas are comparable. However, if sample size is small, the sample protocol should be given careful consideration. Furthermore, Chapter 4 suggests that previous studies may have differed in their findings on the most efficient sampling methodology due to a lack of standardisation for potentially confounding factors as they were mostly based on field data. The simulations in comparison allowed the controlled and precise adjustment of these potentially confounding factors and resulted in consistently identifying small rectangular plots as the most efficient design.

A further question is whether or not species richness patterns, once identified, are influential in conservation decisions. Species richness is never used as the single criterion for conservation planning (Brooks *et al.* 2006) as levels of concordance between species richness and endemism differ between taxonomic groups, region and scale (compare for example Prendergast *et al.* 1993; Ricketts 2001; Jetz *et al.* 2004; Orme *et al.* 2005; Lamoreux *et al.* 2006; Qian & Ricklefs 2008). Concordance has generally been found to be low at the fine scale where conservation decisions take place, i.e. conservation of species richness and high levels of endemism are not necessarily overlapping goals. However, species richness remains important as it is relatively easy to measure, whereas establishing the endemism status of species requires a detailed knowledge of their ranges. Comprehensive lists of endemic species for an area are often not available, particularly in the tropics. In the EAM such a list was only achieved in 2009 after over 30 years of intensive botanical exploration. Furthermore, in the simulations species richness was more difficult to capture the more even the species abundance distribution was (Chapter 4), i.e. assessing the tail of the distribution (rare species) was most difficult. This suggests that sampling protocols that fail to detect the true species richness gradient are even less likely to adequately capture the levels of rarity and thus miss an important component of conservation planning.

Measuring degradation

The benefits of including measures of degradation into conservation planning were discussed in the introduction. However, data on degradation are frequently even more scarce than data on biodiversity. Our understanding of the drivers of forest degradation and how it proceeds across landscapes over time is based on a small number of observational studies (e.g. Gentry & Vasquez 1988; Vasquez & Gentry 1989). Furthermore, the impacts of different forms of forest degradation are only understood in outline, and so far research has largely focused on deforestation (Achard *et al.* 2007; DeFries *et al.* 2007; Gibbs *et al.* 2007; Ramankutty *et al.* 2007; DeFries *et al.* 2010), with much less of a focus on the impacts of degradation on carbon storage (Putz *et al.* 2008), or biodiversity (Bawa & Seidler 1998; Ehrlich & Pringle 2008).

This gap in the data may be partly bridged if patterns of degradation could be reliably predicted. Chapter 5 analyses whether economic models (von Thünen 1966; Kaimowitz & Angelsen 1998; Angelsen 2007) can be used to predict the spread of degradation. The results suggest that it is indeed possible to develop broadly applicable predictions of degradation. In coastal Tanzania degradation has spread remarkably closely to the patterns suggested by the economic theory - in concentric waves that systematically remove high-value forest products around a demand centre (Dar es Salaam), and expand, with each new wave targeting lower-value products. Distance from Dar es Salaam explained between 60 and 80% of the spatial variation in degradation.

The severity of the impacts of degradation in this area and elsewhere (e.g. Koenig 2008; Tang *et al.* in press) reinforce the need for reliable degradation metrics to be incorporated in the formulation of conservation strategies in forest regions (Miles & Kapos 2008; Putz *et al.* 2008). The most disturbed forests had three times lower species richness and over ten times lower carbon storage than the least disturbed forests. A first-order estimate of the above-ground carbon lost from the Tanzanian Coast Region between 1991 and 2005 is ~0.21 Tg carbon yr^{-1}, equivalent to over a quarter of the annual emissions of carbon from fossil fuel use in Tanzania over the same period (Boden *et al.* 2009).

The case study in Chapter 5 suggests that it is possible to base degradation predictions on distance from centers of demand. Further, important variables were the distance from the main road and the accessibility of the forest reserve to trucks. In other areas the patterns may not always be as clear as in the coastal forests (CF) of Tanzania where immense governance shortfalls have allowed an almost unhindered and rapid spread of illegal commercial logging between 2000 and 2005, resulting in an estimated loss of revenue of US$ 58 million in 2005 alone (Milledge *et al.* 2007). Where stricter controls on illegal timber extraction are implemented, patterns will be more diffuse and driven more strongly by opportunity. However, degradation is still likely to be greater in areas close to concentrations of demand and this assumption can serve as a valuable first assessment of risk when no data on degradation are available for conservation planning.

Implications for future surveys

There is a chronic scarcity of funds for biodiversity inventories and (non-molecular) taxonomy e.g. (Disney 1989; Ehrenfeld 1989; Whitehead 1990; Gaston & May 1992; Gee 1992; Wheeler *et al.* 2004) and public expenditure for institutions such as botanic gardens are likely to decline further since the most recent ‘credit crunch’. Consequently, biodiversity surveys need to be as efficient as possible. There are various ways in which the efficiency of biodiversity inventories could be improved, including (1) focusing them on locations and taxonomic groups where a maximum gain in new biodiversity knowledge is likely to be achieved, (2) balancing employment costs for the surveyor according to the survey purpose, (3) surveying using the most efficient methodology, and (4) collaboration between different organisations and a central collation of data. Below, each of these is examined in more detail with reference to the study area and in the light of the results of the thesis.

Focusing surveys on areas where a maximum gain in new biodiversity knowledge is expected. As highlighted above, focusing just a fraction of survey funds on under-researched areas can result in large knowledge gains and mitigate some of the biases in the data. Basic niche theory suggests that areas with different climatic envelopes are likely to contain different groups of species (Chase & Leibold 2003). Therefore, a low cost and potentially effective way to plan new surveys may be to use a combination of fine-scale climatically-driven species distribution models and/or rarefaction analysis, in combination with maps of under-collected environmental space, to identify the locations that are likely to be most fruitful for future surveys. Climatically based fine-scale species distribution models for the area (Platts *et al.* 2010), in line with individual-based rarefaction curves (see Fig. S2.3) suggest that for example the conservation importance of the Rubeho and Nguru Mountains may be underestimated. In terms of under-researched environmental space, a comparison of the environmental space in the study area and that covered by the areas that have been studied (Figure 6.5) clearly shows that the drier environments towards the west of the study area have received far less research attention than the moist forests further to the east. This difference is likely to be due to the expectation that these areas will harbour fewer species and are therefore less attractive for study, but also due to their distance from Dar es Salaam and other infrastructure. The Nguu, North Pare and Ukaguru mountains harbour a particularly high fraction of under-collected environmental space (Fig. 6.5 and Table 6.1) and species distribution models suggest that they may be home to a much greater diversity than currently recorded (Platts *et al.* 2010). Furthermore, much of the CF environmental space is highly under-researched (Fig. 6.5 and Table 6.1). These areas should consequently be preferred locations for future surveys and at least one visit by an expert botanist should occur before an attempt is made to firmly establish their conservation status. Table 6.1 details the gaps between environmental space of individual reserves within the study area and how much such space has been sampled in surveys to date, thus highlighting potential areas to be targeted for future surveys.

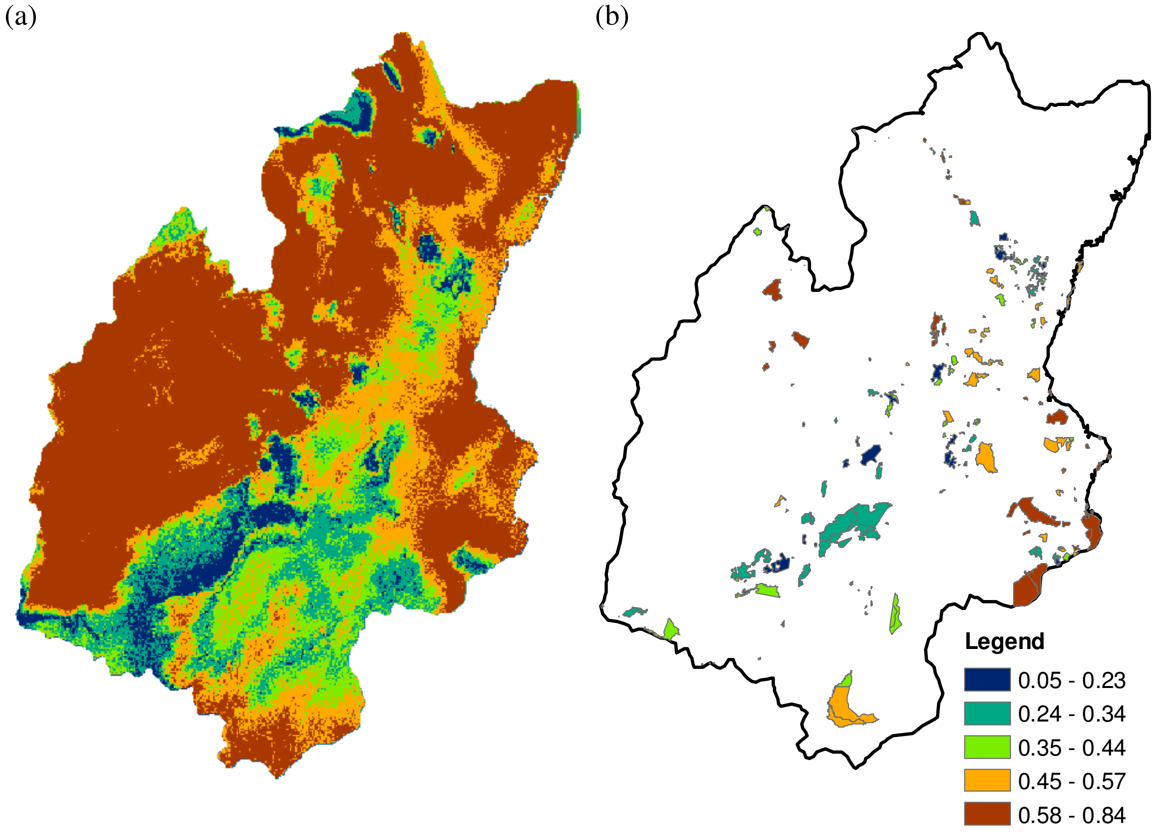


Figure 6.5. Survey gaps in environmental space. (a) Raster of survey gaps in environmental space calculated for each cell (2 × 2 km) as follows: proportion of cells with this value for an environmental variable (variables detailed below) (+/- 0.5 standard deviation) in the entire study area minus proportion of cells that have been sampled with this environment value (+/- 0.5 standard deviation). Resulting differences are standardised to range between 0 and 1. The survey gap map is the mean of survey gaps across the following environmental categories: altitude, slope, aspect (eastness), mean, minimum and maximum annual temperature, annual rainfall, rainfall during driest and wettest month, potential evapotranspiration and annual moisture index (calculations follow Platts *et al.* 2008). (b) Survey gaps across forest reserves (colour filled shapes) and Eastern Arc Mountain blocs (hollow shapes), calculated as the average of pixel values in (a) for each forest reserve.

Forest reserve	Gap	Area	Forest reserve	Gap	Area	Forest reserve	Gap	Area
Mamboya	0.73	Ukaguru	Kisima Gonja	0.45	West Usambara	Uponera	0.28	Ukaguru
Mkongo	0.71	Nguu	South Gendagenda	0.45	coastal	Ruvu	0.28	Uluguru
Mamboto	0.70	Ukaguru	Pugu	0.45	coastal	Kisinga-Rugaro	0.28	Udzungwa
Nyumburuni	0.68	coastal	Myoe	0.44	Mahenge	Chome	0.28	South Pare
Mramba	0.67	North Pare	Nambiga	0.43	Mahenge	Dindili	0.28	Uluguru
Ngulakula	0.67	coastal	Kiranga-Hengae	0.43	South Pare	Ihanga	0.27	Udzungwa
Ruhai River	0.65	coastal	Korogwe Fuel	0.42	coastal	Bombo West	0.27	East Usamba
Mkuli	0.65	Nguu	Kanga	0.42	Nguru	Image	0.27	Udzungwa
Vumari	0.65	South Pare	Mindu	0.41	Uluguru	Longuza	0.26	East Usamba
Mohoro	0.65	coastal	Kihuhwi	0.41	East Usamba	Tawi Village	0.26	coastal
Pumula	0.65	Nguu	Nguru ya Ndege	0.40	Uluguru	Wota	0.26	Rubeho
Chambogo	0.65	South Pare	Chamanyani	0.39	Uluguru	Pangawe East	0.26	Uluguru
Ruvu North Fuel	0.64	coastal	Namakutwa Nyamulete	0.39	coastal	Bombo East 2	0.26	East Usamba
Mohoro River	0.64	coastal	Iringa	0.38	Udzungwa	Mkungwe	0.25	Uluguru
North Nguru	0.64	Nguu	Kasanga	0.38	Uluguru	Kichi Hill	0.25	coastal
Mtanza Msona	0.63	coastal	Kitulanghalo	0.38	Uluguru	Sao Hill	0.24	Udzungwa
Vikindu	0.63	coastal	Mkindo (or Mikindo)	0.38	Nguru	Mvuha	0.23	Uluguru
Kiverenge	0.63	North Pare	Ndelemai	0.38	West Usambara	Kiwengoma	0.23	coastal
Kilindi	0.62	Nguu	Pangawe West	0.38	Uluguru	Mtai	0.23	East Usamba
Kikoka	0.62	coastal	Balangai West	0.37	West Usambara	Ukwiva	0.23	Rubeho
Lungonya	0.61	coastal	Luhombero Luwegu	0.37	Mahenge/coastal	Mkussu	0.22	West Usambara
Mchungu	0.60	coastal	Mamiwa Kisara	0.37	Ukaguru	Kikongoloi	0.22	West Usambara
Rau	0.60	North Pare	Ngongwa-Busangi	0.36	Ukaguru	Bamba Ridge	0.21	East Usamba
Masanganya	0.59	coastal	Manga	0.36	East Usamba	Kimboza	0.21	Uluguru
Kiono Zaraninge	0.58	coastal	Baga I (Mzinga)	0.36	West Usambara	Semdoe	0.21	East Usamba
Minja	0.57	North Pare	Mufindi Scarp	0.35	Udzungwa	Uluguru North	0.20	Uluguru
Kazimzumbwi	0.56	coastal	Handei Village	0.35	East Usamba	Kindoroko	0.20	North Pare
Rudewa	0.55	Nguu	Magombera	0.35	Udzungwa	Vugiri	0.20	West Usambara
Kwizu	0.55	South Pare	Tongwe	0.34	East Usamba	Kitara Ridge	0.20	West Usambara
Ruvu South	0.55	coastal	Kilengwe	0.32	Uluguru	Milindo	0.20	Ukaguru
Morogoro Fuel	0.54	Uluguru	Nilo	0.32	East Usamba	New Dabaga-Ulongambi	0.20	Udzungwa
Msumbugwe	0.54	coastal	Kwamarimba	0.32	East Usamba			
Mafi Hill	0.54	West Usambara	Uzungwa Scarp	0.32	Udzungwa	Shagayu	0.19	West Usambara
Kataplimwa	0.53	Udzungwa	Mahenge Scarp	0.32	Mahenge	Mlinga	0.17	East Usamba
Bondo	0.52	Nguru	Kihuhwi Sigi	0.32	East Usamba	Bombo East 1	0.17	East Usamba
Mtita	0.52	coastal	Pala Mountain	0.31	Rubeho	Shume Magamba	0.17	West Usambara
Mbwara Village	0.51	coastal	Iwande	0.30	Udzungwa	Idewa	0.17	Udzungwa
Pagale	0.49	Nguru	Nyanganje	0.30	Udzungwa	Uluguru South	0.16	Uluguru
Nyandira	0.48	Uluguru	Mfundia	0.30	East Usamba	Mufindi Tea Est.	0.16	Udzungwa
Pongwe	0.48	coastal	Mgambo	0.30	East Usamba	Ikwamba	0.16	Ukaguru
Uzigua	0.48	Nguru	Amani	0.30	East Usamba	Nguru South	0.15	Nguru
Mselezi	0.47	Mahenge	Nambunju Village	0.30	coastal	Kwamgumi	0.14	East Usamba
Muhulu	0.47	Mahenge	Kilombero	0.29	Udzungwa	Mafwomero	0.14	Rubeho
Kwani	0.47	coastal	Kihiliri	0.29	Rubeho/Ukaguru	Mang'alisa	0.13	Rubeho
Talagwe	0.47	Nguru/Ukaguru	Segoma	0.29	East Usamba	Kambai	0.08	East Usamba
Konga	0.46	Uluguru	West Kilombero Scarp	0.29	Udzungwa	Kigogo	0.08	Udzungwa
Ndolwa	0.45	West Usambara	Mlali	0.28	Ukaguru	Lulanda	0.05	Udzungwa

Table 6.1. Forest reserves and their environmental space survey gap value (descending order), calculated as the mean research gap for all 2 × 2 km cells within each reserve. The higher the figure the less environmental space of that reserve has been sampled (values run from 0 to 1). For further details see Fig. 6.5.

Balancing employment costs for the surveyor according to the survey purpose. Botanical assessments by experts are expensive and time intensive but as shown in Chapter 3, the resulting species lists are likely to be more comprehensive and reliable. This is particularly true for species of conservation concern, which frequently include taxonomic novelties or single mountain bloc endemics which may be closely related to (and thus morphologically similar to) more widespread species. Consequently, expert botanists should be employed in assessments for conservation planning such as determining the conservation priority of different areas. Sampling by experts should also be spatially dispersed across the study area in order to generate a more accurate representation of biodiversity patterns and minimise potential biases due to some areas acting as ’botanist magnets’. However, the analysis in Chapter 3 indicated that botanists with intermediate levels of training in tropical plant identification (more than zero and less than six months) were more efficient than experts (more species records per unit time and cost), a likely effect of them more readily leaving a specimen unidentified or even misidentified. In the EAM, volunteer or semi-professionals have collected almost 50% of the available plant records. Therefore they play an extremely important role in the reconnaissance of flora, providing rapid assessments that aim to increase the data volume on readily identifiable species, and/or generate data for ecological analyses in random or systematically designed assessments (e.g. vegetation plots). Such data collections are the prerequisite to many statistical tests and techniques such as species distribution modelling, which require a minimum number of 10 - 50 unbiased and independent data points per species. This is rarely provided by professional botanists, who in most cases collect for botanic gardens and herbaria where restricted cupboard space means that rarely more than ten specimens per species are collected. The species abundance distribution in herbaria is often biased towards taxonomically difficult groups or species that are otherwise of interest, e.g. because they are rare. Furthermore, herbarium data almost always show road and other accessibility biases (Rondinini *et al.* 2006). Results from Chapter 3 indicated that botanists with no training in tropical plant identification were least efficient, and should probably not be employed for botanical assessments unless closely supervised by more experienced colleagues.

Degradation data collections on the other hand require no particular expertise and can be carried out rapidly, relatively inexpensively and with the participation of local communities. Doody *et al.* (2001) detail a protocol that has been widely used in the study area (Appendix). In contrast to the field data collection methods used in Chapter 5, these rapid assessments simply record standing trees and stumps, not differentiating between species or stem sizes. A potential shortfall of such an approach is that it is not possible to make a distinction between the different forms of degradation. Patterns, sources and flows for commercial timber logging and subsistence logging may vary widely, and rapidly collected degradation data may consequently not allow the identification of agents causing degradation, and patterns may be blurred. The combining of rapid degradation assessment approaches with more detailed assessments that establish species, timber class, size and potential destination of all extracted trees would therefore allow the collection of relative large coarse-grained datasets on degradation, whilst also providing more scarce but more detailed targeted data at greater resolutions.

Surveying in the most efficient way. Surveys should be as rapid as possible; however, increased speed should stem from advances in techniques and increased efficiency (e.g. Gardner *et al.* 2008) instead of sacrifices in the reliability of the collected data. Surveys that do not adhere to the minimum needed sample size may fail to reach their goal (Chapter 4), and thereby constitute a risky expenditure of time and money. The collected data will still be extremely useful for species recording purposes; however, the intended study can only be reliably conclusive if further data are collected. Given that travel costs to and from the research location frequently constitute the greatest expense, it may be more efficient to determine the minimum needed sample size prior to the main field work by collecting a few pilot samples and utilising the guidelines presented in Chapter 4. The code for running the simulations required to do this is being made available on request in the hope that this may help plant ecologists to plan future surveys more efficiently.

Collaboration and central collation of data. Collaboration amongst data collecting institutions and data users (e.g. conservation organisations) and central collation of all available data is a prerequisite to efficient field data collection. Pooling of data from different sources will considerably increase our ability to identify possible data biases, strategically plan new surveys, and establish confidence levels when the data are used in conservation decisions. Furthermore, the data are likely to have been more thoroughly scrutinised and be in a format which is suitable for wider use in supra-regional or global analyses of conservation priorities. Studies such as this thesis can be used for some of this work. However, the permanent maintenance and consequent updating of a central data collation is likely to require a dedicated institution and the direct allocation of staff time. Given the costs of surveys this seems a highly reasonable and valuable investment to ensure the long-term use and impact of the data.

A large amount of further data could be gained through the digitisation of museum and herbarium data: thus far, depending on the institution, only between 10 and 15% of the specimens have been digitised; Graham *et al.* 2004. There has been an intensive debate on whether this is a worthwhile investment or would risk the diversion of much needed funds for biodiversity surveys (Brooke 2000a, b; Graves 2000; Scoble 2000; Wirtz 2000). Digitised museum and herbarium data undoubtedly constitutes an enormously useful reference source and allows for a wide range of interesting research questions to be addressed, in particular relating trends over time such as range shifts (Graham *et al.* 2004; Tingley & Beissinger 2009). Furthermore, digitisation costs can be as low as US$ 1 per specimen (Brooke 2000b), whereas the average cost per record in the EAM 1980 – 2007 was US$ 12.51. However, in the light of the findings in Chapters 2 and 3 and the problems associated with herbarium data (biases, potential misidentifications and spatial uncertainties for records collected in pre-GPS times; Graham *et al.* 2004) the large-scale digitisation of herbarium specimens cannot replace systematic data collections for conservation planning. In the EAM, vast quantities of vegetation plot data on the more common and easy to identify species, collected by semi-professionals, have been combined with taxonomically more stringently assessed herbarium collections, and this allowed for a wide range of ecological analyses (e.g. Platts *et al.* 2008; Platts *et al.* 2010). Random or systematic collections also have their shortfalls; in particular they suffer from containing many sterile specimens, which are frequently discarded subsequent to their identification, denying any future verification process. A potential solution may be to photo-document the relevant characters of all recorded specimens and/or to store the voucher specimens with a designated institution. All digital photo material and information on the storage location of specimens should form part of the central data collation. This would allow for subsequent verifications of doubtful material, which are usually dealt with by simply omitting the record from any analysis.

In conclusion, there is much scope for making surveys more efficient in the EAM whereby only small changes to the standard collecting procedures could considerably improve the usability of field data and its reliability in conservation planning, i.e. increase the return value for biodiversity surveys. This extends to maximizing the realisation of the potential of these data by establishing a well maintained and documented accessible database for the area which would allow the use of the data in multiple studies and meta-analyses.

Implications for conservation planning

In the EAM and CF systematic conservation planning is within reach: following intensive biological exploration many of the major taxonomic groups have been relatively well recorded, an increasing number of genetic studies are becoming available (Couvreur *et al.* 2008; Fjeldså & Bowie 2008; Blackburn & Measey 2009), ecosystem services are being quantified (Mwakalila *et al.* 2009) and work is underway to extend degradation predictions (Chapter 5) to the entire study area based on disturbance data collated during this thesis (Appendix). Systematic conservation planning has already been attempted, but on the basis of counts of species of conservation concern with little or no consideration of potential sampling biases and/or vulnerability (CEPF 2003, 2007; WHC 2010). These gaps could easily be filled: with respect to vascular plants, an investment of just ~US$ 120,000 for expert-led inventories in the four most under-researched mountain blocs (Nguu, North Pare, Rubeho, Ukaguru; Figs. 6.1 and 6.5) (US$ 1,500 cost per survey day × 20 field days × 4 sites) could vastly improve the knowledge base for such conservation decisions. If the survey needs for other taxonomic groups are formulated in a similar fashion, and there is increased collaboration amongst institutions that collect primary data as well as transparency in and understanding of the limitations and potential errors in the data, the efficiency of both surveys and conservation planning is likely to improve significantly. It has been argued that in the face of uncertainty and on-going degradation it may be more efficient to use simple decision rules rather than to develop cost and labour intensive comprehensive conservation plans for near-optimal locations of protected areas (Meir *et al.* 2004). While this may be true, severe biases (Chapters 2 and 3) will likely compromise even simple decisions.

Current conservation planning within the study area can also be relatively ad-hoc, reactive to donor wishes and based on whichever data are available at the time. There has been little coordinated effort to identify joint conservation goals *a priori*. A joint conservation strategy that explicitly formulates how irreplaceability and vulnerability should be measured (which indicators), how these aspects should be weighted against each other, and how much biodiversity and other ecosystem services should be conserved may be a better way forward. Another vital aspect to conservation planning in the area, given the high population density in areas rich in species of conservation concern (Balmford *et al.* 2001) and the fact that humans rely on the ecosystem services provided by these areas, is the integration of conservation projects with development projects. Combined conservation and development efforts that ensure that local communities benefit from and are involved in the management of natural resources (e.g. Arabuko Sokoke Forest Management Team 2002) may significantly enhance the effectiveness of conservation (e.g. Blomley *et al.* 2008), and improve governance which has been shown to be a major factor in successful forest conservation (O'Connor *et al.* 2003; Smith *et al.* 2003; Milledge *et al.* 2007).

The needs for identifying and mitigating data biases, and for integrated and systematic conservation planning may well be true for other areas and even at the global scale, where there is redundancy and duplication amongst the major conservation prioritisation schemes, which each collate their own data, formulate their own (often competing) goals and raise their own funds (Mace *et al.* 2000). Worse still, some of these schemes that pool US$ billions of conservation resources have taken a rather defensive approach towards critical analyses (Myers & Mittermeier 2003; Whittaker *et al.* 2005). Based on counts of species and rarities, these schemes have an appearance of objectivity and scientific rigour, which is part of their appeal to donors and the public. The EAM are a prime-example for how inaccurate estimates can be misleading, with the original estimate of 50% plant endemism (Myers *et al.* 2000) qualifying the area as a hotspot, but with a likely more accurate estimate of 11.8% (Roy Gereau, unpublished data) based on the extensive data collated being below the required threshold of 1,500 endemic species[1]. The communication of problems associated with global conservation priority schemes is understandably problematic as these schemes are part of the branding and fundraising strategies of many non-governmental organisations. Furthermore, the sudden cutting of funds, resulting from revised estimates not fulfilling certain benchmarks, to areas such as the EAM that are on their way to becoming model examples for conservation planning and that provide vitally important ecosystem services to the entire nation (Mwakalila *et al.* 2009), is likely to be counter-productive. However, critical analysis and incorporation of new data and advances in biogeographic research (Whittaker *et al.* 2005) does not necessarily lower the profile of such prioritisation schemes, and donors and the general public may have an understanding for the need of schemes such as hotspots to be reasonably flexible in their assessments and recommendations. On the contrary, in the face of climate change, which has already led to shifts in species distributions (Parmesan & Yohe 2003), it is particularly important that the generally static conservation prioritisation schemes have the capacity to adapt (e.g. Pressey *et al.* 2007; Hannah *et al.* 2008) without losing their public support.

Given the urgent need for both identifying and implementing optimal conservation strategies across the globe, increased collaboration between scientists and conservationists is essential. However, these sectors tend to use different communication platforms (even languages) and operate on different timescales. The research community tends to be driven by the need for the rapid publication of articles in high impact journals (often not accessible to the general public or people in developing countries) as well as short-term grant funding cycles, and rapidly changes and adapts to the arrival of new theories and modification of current theory. In contrast, the conservation community tends to work by implementing long-term strategies and gains most of its funding from donor and public support, communicating via its own subset of journals or more public-centred media. Due to concern over compromising support and funding, conservation organisations may be hesitant to accept scientific criticism on popular schemes (Myers & Mittermeier 2003), but generally there is an openness towards using newer findings, techniques and data as long as these are communicated in an accessible way (see for example several interviews with prominent conservationists at http://www.biogeography.org/html/fb.html). Scientists on the other hand frequently take

---

[1] The Eastern Arc Mountains are now part of the Eastern Afromontane hotspot (Mittermeier *et al.* 2005).

little initiative to ensure that their results reach application in conservation - a problem that has been termed “knowing but not doing” (Knight *et al.* 2008). The prerequisites for a more fruitful dialogue are, amongst others, that scientific publications appear in locally accessible media as well as journals, and that there is a change in attitude amongst scientists regarding measures of their performance. In its appealing simplicity and apparent objectivity but inflexibility and narrow scope, the citation index paradoxically parallels exactly the attributes of some of the global conservation prioritisation schemes. A possible route to greater collaboration between the scientific and conservation communities may be for funding organisations and donors to explicitly target some funds for the organising and running of workshops and visits to organisations and institutes involving both scientists and conservationists. Encouraging progress has been made with the implementation of a model comparable to the Intergovernmental Panel on Climate Change for biodiversity and ecosystem service conservation (http://www.ipbes.net/en/index.asp). Furthermore, university curricula for conservation related studies should include practical conservation experience for the students to gain contacts and familiarity with relevant organisations and to learn about the implementation of research (e.g. Knight *et al.* 2008).

In summary, this thesis critically examined the extent and state of our present knowledge of biodiversity patterns and degradation in the EAM and CF of Tanzania. This area is clearly of high conservation importance. However, substantial gaps in our knowledge and their consequences were identified, which are currently hampering our ability to effectively conserve the biodiversity in these areas. Recommendations for addressing these shortfalls were made leading to a general recognition that there is a great need for a coordinated conservation planning framework that aims to draw together interlinked components of the system such as biodiversity, evolutionary history and potential, vulnerability and ecosystem services, and which can be utilised for optimising future conservation and survey strategies. Importantly, many gaps between our current knowledge and the requirements for such a framework could be addressed relatively inexpensively. Given the ongoing degradation and large scale environmental change in this area, this needs to be done rapidly.

References

Achard F., DeFries R., Eva H., Hansen M., Mayaux P. & Stibig H.J. (2007). Pan-tropical monitoring of deforestation. *Environmental Research Letters*, 2.

Alroy J. (2003). Taxonomic inflation and body mass distributions in North American fossil mammals. *Journal of Mammalogy*, 84, 431-443.

Angelsen A. (2007). Forest Cover Change in Space and Time: Combining the von Thünen and Forest Transition Theories. *World Bank Policy Research Working Paper*, 4117.

Arabuko Sokoke Forest Management Team (2002). Strategic Forest Management Plan 2002 - 2027. Gedi, Kenya.

Balmford A., Moore J.L., Brooks T., Burgess N., Hansen L.A., Williams P. & Rahbek C. (2001). Conservation conflicts across Africa. *Science*, 291, 2616-2619.

Bawa K.S. & Seidler R. (1998). Natural forest management and conservation of biodiversity in tropical forests. *Conservation Biology*, 12, 46-55.

Blackburn D.C. & Measey G.J. (2009). Dispersal to or from an African biodiversity hotspot? *Molecular Ecology*, 18, 1904-1915.

Blomley T., Pfliegner K., Isango J., Zahabu E., Ahrends A. & Burgess N. (2008). Seeing the wood for the trees: an assessment of the impact of participatory forest management on forest condition in Tanzania. *Oryx*, 42, 380-391.

Boden T.A., Marland G. & Andres R.J. (2009). Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center.

Brooke M.D. (2000a). Costs and benefits of Web access to museum data - Reply. *Trends in Ecology & Evolution*, 15, 375-375.

Brooke M.D. (2000b). Why museums matter. *Trends in Ecology & Evolution*, 15, 136-137.

Brooks T.M., Mittermeier R.A., da Fonseca G.A.B., Gerlach J., Hoffmann M., Lamoreux J.F., Mittermeier C.G., Pilgrim J.D. & Rodrigues A.S.L. (2006). Global biodiversity conservation priorities. *Science*, 313, 58-61.

Burgess N.D., Butynski T.M., Cordeiro N.J., Doggart N.H., Fjeldså J., Howell K.M., Kilahama F.B., Loader S.P., Lovett J.C., Mbilinyi B., Menegon M., Moyer D.C., Nashanda E., Perkin A., Rovero F., Stanley W.T. & Stuart S.N. (2007). The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, 134, 209-231.

CEPF (Critical Ecosystem Partnership Fund) (2003). Eastern Arc Mountains and coastal forest hotspot Ecosystem Profile. CEPF Dar es Salaam, Nairobi. http://www.cepf.net/Documents/final.easternarc.ep.pdf

CEPF (Critical Ecosystem Partnership Fund) (2007). CEPF's US $ 7 million conservation pledge. *Arc Journal*, 20.

Channing A., Menegon M., Salvidio S. & Akker S. (2005). A new forest toad from the Ukaguru mountains, Tanzania (Bufonidae : Nectophrynoides). *African Journal of Herpetology*, 54, 149-157.

Chase J.M. & Leibold M.A. (2003). Ecological Niches: Linking Classical and Contemporary Approaches. Chicago University Press, Chicago.

Couvreur T.L.P., Chatrou L.W., Sosef M.S.M. & Richardson J.E. (2008). Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *Bmc Biology*, 6, 10.

Da Fonseca G.A.B., Balmford A., Bibby C., Boitani L., Corsi F., Brooks T., Gascon C., Olivieri S., Mittermeier R.A., Burgess N., Dinerstein E., Olson D., Hannah L., Lovett J., Moyer D., Rahbek C., Stuart S. & Williams P. (2000). It's time to work together and stop duplicating conservation efforts ... following Africa's lead in setting priorities. *Nature*, 405, 393-394.

DeFries R., Achard F., Brown S., Herold M., Murdiyarso D., Schlamadinger B. & de Souza C. (2007). Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environmental Science & Policy*, 10, 385-394.

DeFries R.S., Rudel T., Uriarte M. & Hansen M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*. doi:10.1038/ngeo756

Disney R.H.L. (1989). Does Anyone Care. *Conservation Biology*, 3, 414-414.

Doggart N., Perkin A., Kiure J., Fjeldsa J., Poynton J. & Burgess N. (2006). Changing places: How the results of new field work in the Rubeho Mountains influence conservation priorities in the Eastern Arc Mountains of Tanzania. *African Journal of Ecology*, 44, 134-144.

Doody K.Z., Howell K.M. & Fanning E. (2001). Udzungwa Mountains Biodiversity Surveys – Methods Manual. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Ehrenfeld D. (1989). Is Anyone Listening. *Conservation Biology*, 3, 415-415.

Ehrlich P.R. & Pringle R.M. (2008). Where does biodiversity go from here? A grim business-as-usual forecast and a hopeful portfolio of partial solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 11579-11586.

Ferrier S., Powell G.V.N., Richardson K.S., Manion G., Overton J.M., Allnutt T.F., Cameron S.E., Mantle K., Burgess N.D., Faith D.P., Lamoreux J.F., Kier G., Hijmans R.J., Funk V.A., Cassis G.A., Fisher B.L., Flemons P., Lees D., Lovett J.C. & Van Rompaey R. (2004). Mapping more of terrestrial biodiversity for global conservation assessment. *Bioscience*, 54, 1101-1109.

Finch J., Leng M.J. & Marchant R. (2009). Late Quaternary vegetation dynamics in a biodiversity hotspot, the Uluguru Mountains of Tanzania. *Quaternary Research*, 72, 111-122.

Fjeldså J. & Bowie R.C.K. (2008). New perspectives on the origin and diversification of Africa's forest avifauna. *African Journal of Ecology*, 46, 235-247.

Fjeldså J., Ehrlich D., Lambin E. & Prins E. (1997). Are biodiversity 'hotspots' correlated with current ecoclimatic stability? A pilot study using the NOAA-AVHRR remote sensing data. *Biodiversity and Conservation*, 6, 401-422.

Fjeldså J. & Lovett J.C. (1997). Biodiversity and environmental stability. *Biodiversity and Conservation*, 6, 315-323.

Fuller T., Sanchez-Cordero V., Illoldi-Rangel P., Linaje M. & Sarkar S. (2007). The cost of postponing biodiversity conservation in Mexico. *Biological Conservation*, 134, 593-600.

Gardner T.A., Barlow J., Araujo I.S., Avila-Pires T.C., Bonaldo A.B., Costa J.E., Esposito M.C., Ferreira L.V., Hawes J., Hernandez M.I.M., Hoogmoed M.S., Leite R.N., Lo-Man-Hung N.F., Malcolm J.R., Martins M.B., Mestre L.A.M., Miranda-Santos R., Overal W.L., Parry L., Peters S.L., Ribeiro M.A., da Silva M.N.F., Motta C.D.S. & Peres C.A. (2008). The cost-effectiveness of biodiversity surveys in tropical forests. *Ecology Letters*, 11, 139-150.

Gaston K.J. & May R.M. (1992). Taxonomy of Taxonomists. *Nature*, 356, 281-282.

Gee H. (1992). UK Systematic Biology - Peers slam peer-review. *Nature*, 355, 488-488.

Gentry A.H. & Vasquez R. (1988). Where have all the ceibas gone - a case-history of mismanagement of a tropical forest resource. *Forest Ecology and Management*, 23, 73-76.

Gibbs H.K., Brown S., Niles J.O. & Foley J.A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2.

Gimaret-Carpentier C., Pelissier R., Pascal J.P. & Houllier F. (1998). Sampling strategies for the assessment of tree species diversity. *Journal of Vegetation Science*, 9, 161-172.

Gordon J.E. & Newton A.C. (2006a). Efficient floristic inventory for the assessment of tropical tree diversity: A comparative test of four alternative approaches. *Forest Ecology and Management*, 237, 564-573.

Gordon J.E. & Newton A.C. (2006b). The potential misapplication of rapid plant diversity assessment in tropical conservation. *Journal for Nature Conservation*, 14, 117-126.

Gotelli N.J., Anderson M.J., Arita H.T., Chao A., Colwell R.K., Connolly S.R., Currie D.J., Dunn R.R., Graves G.R., Green J.L., Grytnes J.A., Jiang Y.H., Jetz W., Lyons S.K., McCain C.M., Magurran A.E., Rahbek C., Rangel T., Soberon J., Webb C.O. & Willig

M.R. (2009). Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, 12, 873-886.

Graham C.H., Ferrier S., Huettman F., Moritz C. & Peterson A.T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, 19, 497-503.

Grantham H.S., Wilson K.A., Moilanen A., Rebelo T. & Possingham H.P. (2009). Delaying conservation actions for improved knowledge: how long should we wait? *Ecology Letters*, 12, 293-301.

Graves G.R. (2000). Costs and benefits of Web access to museum data. *Trends in Ecology & Evolution*, 15, 374-374.

Halpern B.S., Pyke C.R., Fox H.E., Haney J.C., Schlaepfer M.A. & Zaradic P. (2006). Gaps and mismatches between global conservation priorities and spending. *Conservation Biology*, 20, 56-64.

Hamilton A.C. (1981). The quaternary history of African forests - its relevance to conservation. *African Journal of Ecology*, 19, 1-6.

Hannah L., Dave R., Lowry P.P., Andelman S., Andrianarisata M., Andriamaro L., Cameron A., Hijmans R., Kremen C., MacKinnon J., Randrianasolo H.H., Andriambololonera S., Razafimpahanana A., Randriamahazo H., Randrianarisoa J., Razafinjatovo P., Raxworthy C., Schatz G.E., Tadross M. & Wilmee L. (2008). Climate change adaptation for conservation in Madagascar. *Biology Letters*, 4, 590-594.

Hemp A. (2006). Vegetation of Kilimanjaro: hidden endemics and missing bamboo. *African Journal of Ecology*, 44, 305-328.

Isaac N.J.B., Mallet J. & Mace G.M. (2004). Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution*, 19, 464-469.

Jetz W., Rahbek C. & Colwell R.K. (2004). The coincidence of rarity and richness and the potential signature of history in centres of endemism. *Ecology Letters*, 7, 1180-1191.

Kaimowitz D. & Angelsen A. (1998). *Economic models of tropical deforestation: a review*. Center for International Forestry Research, Bangor.

Knapp S., Lughadha E.N. & Paton A. (2005). Taxonomic inflation, species concepts and global species lists. *Trends in Ecology & Evolution*, 20, 7-8.

Knight A.T., Cowling R.M., Rouget M., Balmford A., Lombard A.T. & Campbell B.M. (2008). Knowing but not doing: Selecting priority conservation areas and the research-implementation gap. *Conservation Biology*, 22, 610-617.

Koenig R. (2008). Critical Time for African Rainforests. *Science*, 320, 1439-1441.

Küper W., Sommer J.H., Lovett J.C. & Barthlott W. (2006). Deficiency in African plant distribution data - missing pieces of the puzzle. *Botanical Journal of the Linnean Society*, 150, 355-368.

Küper W., Sommer J.H., Lovett J.C., Mutke J., Linder H.P., Beentje H.J., Van Rompaey R., Chatelain C., Sosef M. & Barthlott W. (2004). Africa's hotspots of biodiversity redefined. *Annals of the Missouri Botanical Garden*, 91, 525-535.

Lamoreux J.F., Morrison J.C., Ricketts T.H., Olson D.M., Dinerstein E., McKnight M.W. & Shugart H.H. (2006). Global tests of biodiversity concordance and the importance of endemism. *Nature*, 440, 212-214.

Lindqvist C. & Albert V.A. (1999). Phylogeny and conservation of African violets (Saintpaulia: Gesneriaceae): New findings based on nuclear ribosomal 5S non-transcribed spacer sequences. *Kew Bulletin*, 54, 363-377.

Lindqvist C. & Albert V.A. (2001). A high elevation ancestry for the Usambara Mountains and lowland populations of African violets (Saintpaulia, Gesneriaceae). *Systematics and Geography of Plants*, 71, 37-44.

Lovett J.C. (1993). Eastern Arc moist forest flora. In: *Biogeography and Ecology of the Rain Forests of Eastern Africa* (eds. Lovett JC & Wasser). Cambridge University Press, Cambridge.

Lovett J.C. (1996). Elevational and latitudinal changes in tree associations and diversity in the Eastern Arc mountains of Tanzania. *Journal of Tropical Ecology*, 12, 629-650.

Lovett J.C. (1999). Tanzanian forest tree plot diversity and elevation. *Journal of Tropical Ecology*, 15, 689-694.

Lovett J.C., Marchant R., Taplin J. & Kuper W. (2005). The oldest rainforests in Africa: stability or resilience for survival and diversity? *Phylogeny and Conservation*, 8, 198-229.

Lovett J.C., Marshall A.R. & Carr J. (2006). Changes in tropical forest vegetation along an altitudinal gradient in the Udzungwa Mountains National Park, Tanzania. *African Journal of Ecology*, 44, 478-490.

Lovett J.C., Rudd S., Taplin J. & Frimodt-Moller C. (2000). Patterns of plant diversity in Africa south of the Sahara and their implications for conservation management. *Biodiversity and Conservation*, 9, 37-46.

Mace G.M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 359, 711-719.

Mace G.M., Balmford A., Boitani L., Cowlishaw G., Dobson A.P., Faith D.P., Gaston K.J., Humphries C.J., Vane-Wright R.I., Williams P.H., Lawton J.H., Margules C.R., May R.M., Nicholls A.O., Possingham H.P., Rahbek C. & van Jaarsveld A.S. (2000). It's time to work together and stop duplicating conservation efforts. *Nature*, 405, 393-393.

Mallet J., Isaac N.J.B. & Mace G.M. (2005). Response to Harris and Froufe, and Knapp et al: Taxonomic inflation. *Trends in Ecology & Evolution*, 20, 8-9.

McDade L.A. (1995). Species concepts and problems in practice – insights from botanical monographs. *Systematic Botany*, 20, 606-622.

McDonald-Madden E., Bode M., Game E.T., Grantham H. & Possingham H.P. (2008). The need for speed: informed land acquisitions for conservation in a dynamic property market. *Ecology Letters*, 11, 1169-1177.

Meir E., Andelman S. & Possingham H.P. (2004). Does conservation planning matter in a dynamic and uncertain world? *Ecology Letters*, 7, 615-622.

Menegon M., Burgess N., Doggart N., Loader S. & Owen N. (2009). The South Nguru Mountains-a new jewel in the Eastern Arc. *Oryx*, 43, 174-175.

Menegon M., Doggart N. & Owen N. (2008). The Nguru mountains of Tanzania, an outstanding hotspot of herpetofaunal diversity. *Acta Herpetologica*, 3, 107-127.

Menegon M., Salvidio S. & Loader S.P. (2004). Five new species of Nectophrynoides Noble 1926 (Amphibia Anura Bufonidae) from the Eastern Arc Mountains, Tanzania. *Tropical Zoology*, 17, 97-121.

Miles L. & Kapos V. (2008). Reducing greenhouse gas emissions from deforestation and forest degradation: Global land-use implications. *Science*, 320, 1454-1455.

Milledge S.A.H., Gelvas I.K. & Ahrends A. (2007). Forestry, Governance and National Development: Lessons Learned from a Logging Boom in Southern Tanzania. TRAFFIC East/Southern Africa, Tanzania Development Partners Group, Tanzania Ministry of Natural Resources and Tourism, Dar es Salaam.

Mittermeier R.A., Robles-Gil P., Hoffmann M., Pilgrim J.D., Brooks T.M., Mittermeier C.G., Lamoreux J.L. & Fonseca G. (2005). *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Ecoregions*. 2 edn, Cemex, Mexico City.

Möller M. & Cronk Q.C.B. (1997a). Origin and relationships of Saintpaulia (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. *American Journal of Botany*, 84, 956-965.

Möller M. & Cronk Q.C.B. (1997b). Phylogeny and disjunct distribution: Evolution of Saintpaulia (Gesneriaceae). *Proceedings of the Royal Society of London Series B Biological Sciences*, 264, 1827-1836.

Mumbi C.T., Marchant R., Hooghiemstra H. & Wooller M.J. (2008). Late Quaternary vegetation reconstruction from the Eastern Arc Mountains, Tanzania. *Quaternary Research*, 69, 326-341.

Mwakalila S., Burgess N., Ricketts T., Olwero N., Swetnam R., Mbilinyi B., Marchant R., Mtalo F., White S., Munishi P., Malimbwi R., Smith C., Jambiya G., Marshall A., Madoffe S., Fisher B., Kajembe G., Morse-Jones S., Kulindwa K., Green R., Turner R.K., Green J. & Balmford A. (2009). Valuing the Arc: Linking Science with Stakeholders to Sustain Natural Capital. *The Arc Journal*, 23, 25-30.

Myers N. & Mittermeier R.A. (2003). Impact and acceptance of the hotspots strategy: Response to Ovadia and to Brummitt and Lughadha. *Conservation Biology*, 17, 1449-1450.

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

Nogues-Bravo D., Araujo M.B., Romdal T. & Rahbek C. (2008). Scale effects and human impact on the elevational species richness gradients. *Nature*, 453, 216-U8.

O'Connor C., Marvier M. & Kareiva P. (2003). Biological vs. social, economic and political priority-setting in conservation. *Ecology Letters*, 6, 706-711.

Orme C.D.L., Davies R.G., Burgess M., Eigenbrod F., Pickup N., Olson V.A., Webster A.J., Ding T.S., Rasmussen P.C., Ridgely R.S., Stattersfield A.J., Bennett P.M., Blackburn T.M., Gaston K.J. & Owens I.P.F. (2005). Global hotspots of species richness are not congruent with endemism or threat. *Nature*, 436, 1016-1019.

Parmesan C. & Yohe G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37-42.

Phillips O.L., Vasquez Martinez R., Nunez Vargas P., Lorenzo Monteagudo A., Chuspe Zans M.E., Galiano Sanchez W., Pena Cruz A., Timana M., Yli-Halla M. & Rose S. (2003). Efficient plot-based floristic assessment of tropical forests. *Journal of Tropical Ecology*, 19, 629-645.

Platts, P.J., Ahrends, A., Gereau, R.E., McClean, C., Lovett, J.C., Marshall, A., Pellikka, P., Mulligan, M., Fanning, E., Marchant, R. (2010). Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya. *Diversity and Distributions* 16, 628-642.

Platts P.J., McClean C.J., Lovett J.C. & Marchant R. (2008). Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty. *Ecological Modelling*, 218, 121-134.

Prendergast J.R., Quinn R.M., Lawton J.H., Eversham B.C. & Gibbons D.W. (1993). Rare Species, the Coincidence of Diversity Hotspots and Conservation Strategies. *Nature*, 365, 335-337.

Pressey R.L., Cabeza M., Watts M.E., Cowling R.M. & Wilson K.A. (2007). Conservation planning in a changing world. *Trends in Ecology & Evolution*, 22, 583-592.

Putz F.E., Zuidema P.A., Pinard M.A., Boot R.G.A., Sayer J.A., Sheil D., Sist P., Elias & Vanclay J.K. (2008). Improved tropical forest management for carbon retention. *Plos Biology*, 6, 1368-1369.

Qian H. & Ricklefs R.E. (2008). Global concordance in diversity patterns of vascular plants and terrestrial vertebrates. *Ecology Letters*, 11, 547-553.

Rahbek C. (2005). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, 8, 224-239.

Ramankutty N., Gibbs H.K., Achard F., Defriess R., Foley J.A. & Houghton R.A. (2007). Challenges to estimating carbon emissions from tropical deforestation. *Global Change Biology*, 13, 51-66.

Ricketts T.H. (2001). Aligning conservation goals: Are patterns of species richness and endemism concordant at regional scales? *Animal Biodiversity and Conservation*, 24, 91-99.

Rondinini C., Wilson K.A., Boitani L., Grantham H. & Possingham H.P. (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, 9, 1136-1145.

Rovero F., Menegon M., Leonard C., Perkin A., Doggart N., Mbilinyi M. & Mlawila L. (2008). A previously unsurveyed forest in the Rubeho Mountains of Tanzania reveals new species and range records. *Oryx*, 42, 16-17.

Scholes R.J., Küper W., Biggs R., Mwangi E., Raharimampionona R., Lowry P., Sene E., Ashton P.M.S., Blake S. & Justice C.O. (2006). Biodiversity. In: *Africa Environment Outlook 2 - Our Environment, Our Wealth* (eds. Mohamed-Katere JC & Sabet M). Divison of Early Warning and Assessment (DEWA), United Nations Environment Programme, Nairobi.

Scoble M.J. (2000). Costs and benefits of Web access to museum data. *Trends in Ecology & Evolution*, 15, 374-374.

Smith R.J., Muir R.D.J., Walpole M.J., Balmford A. & Leader-Williams N. (2003). Governance and the loss of biodiversity. *Nature*, 426, 67-70.

Tang L., Shao G., Piao Z., Dai L., Jenkins M.A., Wang S., Wu G., Wu J. & Zhao J. (in press). Forest degradation deepens around and within protected areas in East Asia. *Biological Conservation*.

Tingley M. & Beissinger S. (2009). Detecting range shifts from historical species occurences: new perspectives on old data. *Trends in Ecology & Evolution*, 24, 625-633.

Vasquez R. & Gentry A.H. (1989). Use and misuse of forest-harvested fruits in the Iquitos area. *Conservation Biology*, 3, 350-361.

von Thünen J.H. (1966). Der Isolierte Staat in Beziehung der Landwirtschaft und Nationalökonomie. In: *Von Thünen's Isolated State* (ed. Hall P). Pergamon Press, Oxford.

WHC (2010). Eastern Arc Mountains World Heritage Nomination. Paris. http://rapidshare.com/files/342874331/E_Arc_Mountains_World_Heritage_Nomination_100127_FINAL.pdf

Wheeler Q.D., Raven P.H. & Wilson E.O. (2004). Taxonomy: Impediment or expedient? *Science*, 303, 285-285.

Whitehead P. (1990). Systematics - an Endangered Species. *Systematic Zoology*, 39, 179-184.

Whittaker R.J., Araujo M.B., Paul J., Ladle R.J., Watson J.E.M. & Willis K.J. (2005). Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, 11, 3-23.

Wirtz P. (2000). Costs and benefits of Web access to museum data. *Trends in Ecology & Evolution*, 15, 374-375.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. (2009). *Mixed Effects Models and Extension in Ecology with R*. Springer, New York.

Appendix – Thesis database

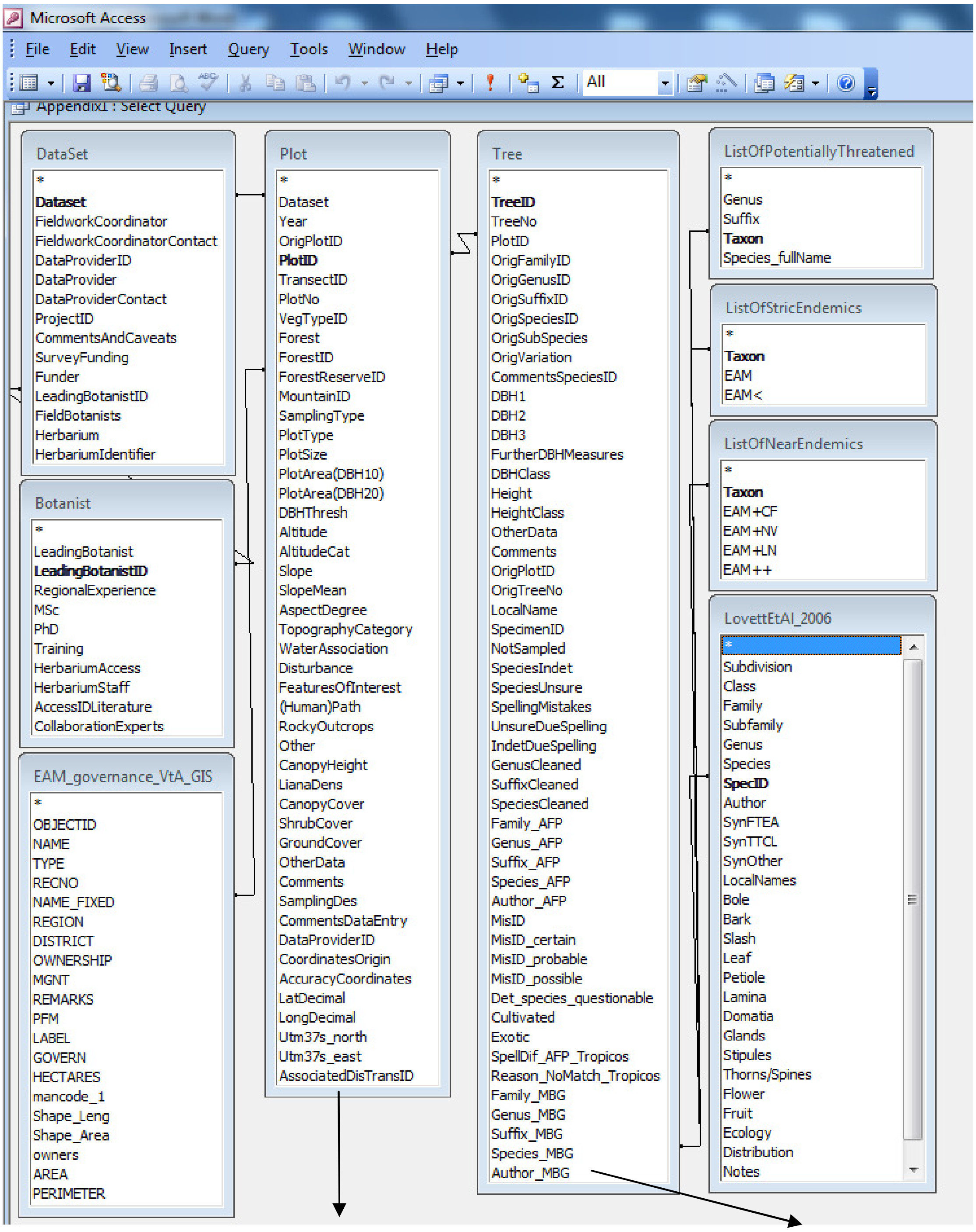
The thesis database comprises two integrated sub-databases – vegetation and disturbance. They are described separately below.

Vegetation database

The vegetation database contains 70,081 Eastern Arc Mountain (EAM) plant records (57,900 of which are identified to species) from 2,216 vegetation plots and 474 species (mostly trees), and can be linked to further 26,253 (17,731 from EAM blocs) records from the Missouri Botanic Garden TROPICOS database (*n* total records EAM = 87,812; *n* total species = 3,986). A further 12,018 records from the coastal forests (CF) are currently being added. The database comprises nine related tables, providing information on individual tree measurements, tree species characteristics (Lovett *et al.* 2006), species range size (endemism) (Roy Gereau, unpublished data), species threat status (Eastern Africa Plant Red List Authority 2008), vegetation plot characteristics, forest reserve attributes (Mwakalila *et al.* 2009), and metadata including contacts, potential caveats, survey funding and levels of training of the field botanists (Fig. App.1). The taxonomic system is standardised to the African Flowering Plants Database (2008), and a fully revised list produced by Roy Gereau on the basis of taxonomic references and revisions. Various different sampling protocols have been used to generate the plot data with plot sizes ranging from 625 – 2,500 m^2, randomly or systematically located, while the TROPICOS data has generally been collected opportunistically.

The major strength of the plot data is the comparatively high data volume. While structural measurements will be relatively accurate, the reliability of the species identification is more doubtful: much of the data has been collected by students, volunteers and/or semi-professional botanists, and many (sterile) voucher specimens will not have been mounted for herbaria, though the specimens may have been stored by the data collector. Furthermore, the vegetation plot records are generally only identified to species (in contrast, the TROPICOS records are frequently taxonomically resolved to sub-species or variety). When using the vegetation plot data it is therefore advisable to account for training of the leading field botanist, contact them over doubtful records, and to generally treat identifications with care.

The plot data has been collected by Frontier Tanzania (Victoria Wilkins Kindemba, Nisha Owen, Nike Doggart, Claire Bracebridge, Kathryn Doody amongst many others) (*n* = 43,811), Antje Ahrends (*n* = 12,869), Jon Lovett (*n* = 11,232), Andrew Marshall (*n* = 7,345), UNDP/GEF Cross Border Biodiversity Programme (*n* = 4,153), John Hall (*n* = 2,589) and James Taplin (*n* = 100). Species identifications of the two largest data contributions (collected recently by Frontier Tanzania and Antje Ahrends) are potentially the least accurate as for a large number of specimens final identification is still pending. Phil Platts and Jon Green assisted with the cleaning of spatial coordinates.



Integrates with the disturbance data (430 Eastern Arc vegetation plots are situated right next to a disturbance transect; all coastal forest data comprise both vegetation and disturbance assessments).

Can be linked to the TROPICOS database which contains 26,253 further records.

Figure App.1. Screenshot of the vegetation plot database. (Some fields have been omitted to ease visualisation). Note that not all rows contain data for all fields, e.g. not every data provider has assessed canopy cover or height of the trees.

Disturbance database

The disturbance database contains 1,057 transects (826 with spatial coordinates) and 273,564 records, of which 261,546 have been collected in the EAM (covering all mountain blocs with the exception of Malundwe), and 12,018 in the CF (these contain both disturbance and vegetation data, i.e. overlap with records in the vegetation database). The basic collection method is illustrated in Fig. App.2. With the exception of assessments in the CF, no information on species identity or tree/pole size has been collected. For a screenshot of the database see Fig. App.3.

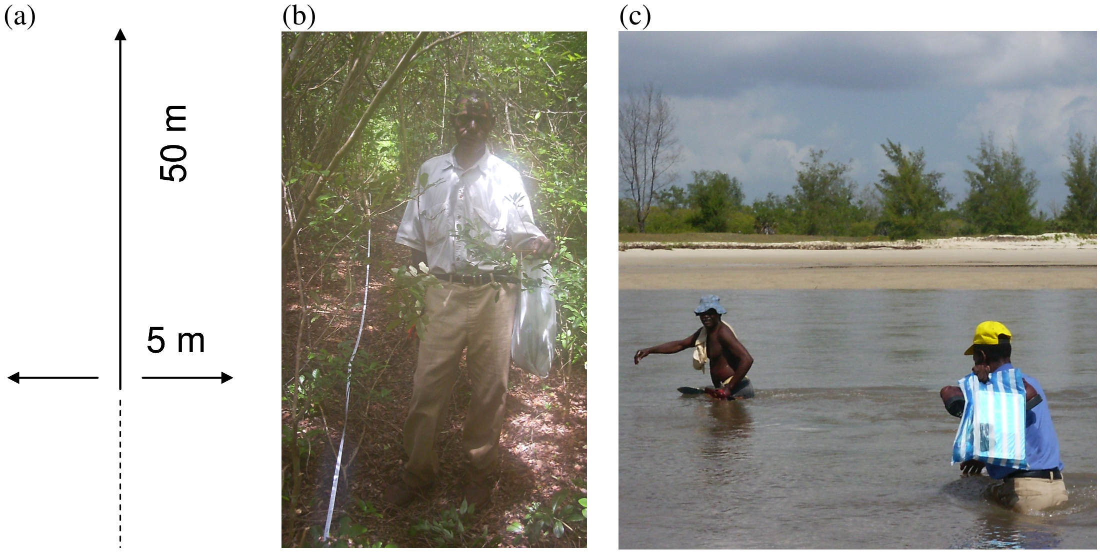
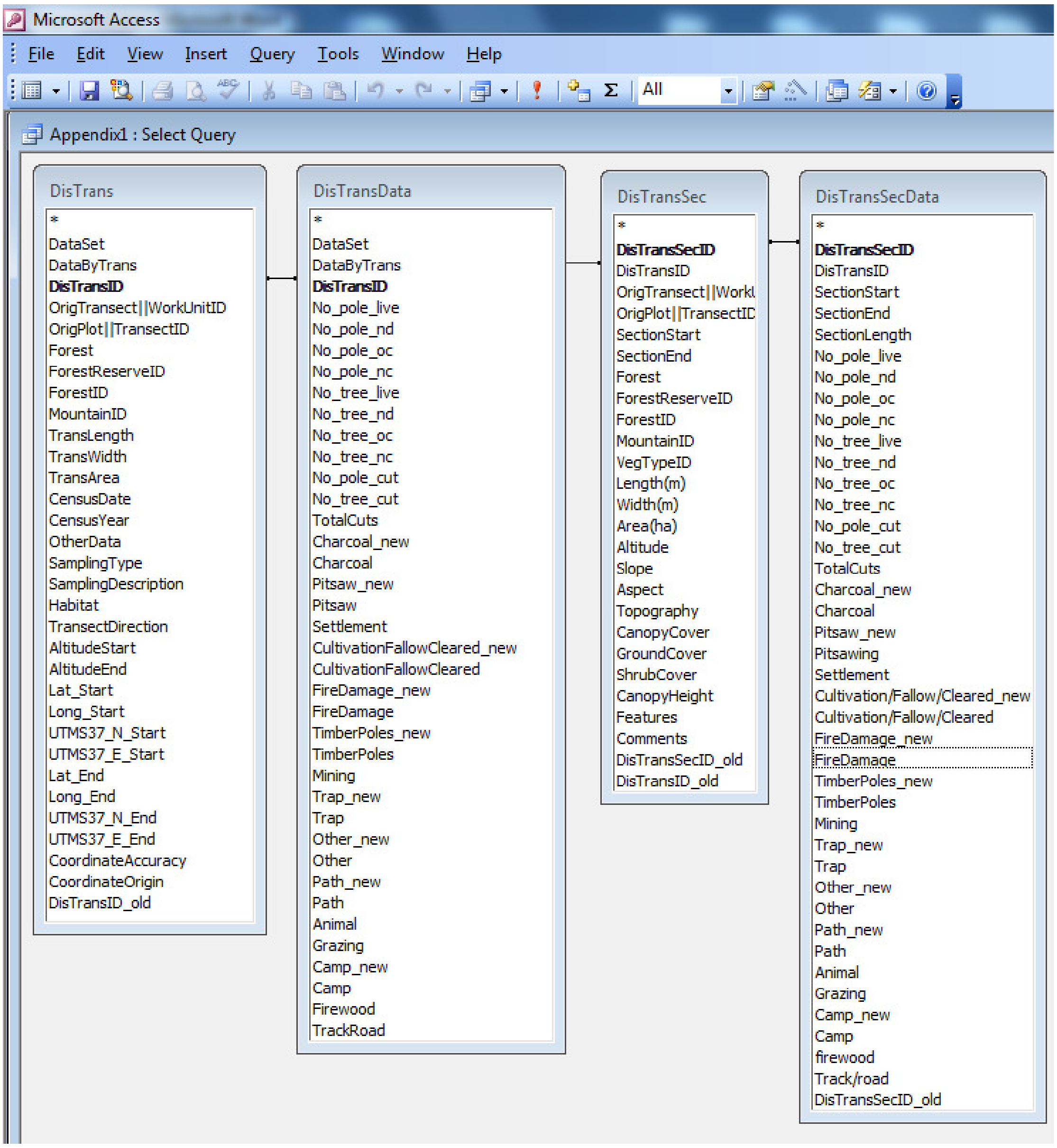


Figure App.2. Disturbance transect assessments. (a) The basic methodology (Doody *et al.* 2001): along transects of 10 m width all standing, naturally dead and cut poles (≥ 2 m straight length; ≥ 50 to < 150 mm drh) and trees (≥ 3 m straight length; ≥ 150 mm drh) are counted in sections of 50 m length. (b) Optionally, species and size of the poles and trees can be assessed – here done by Boniface Mhoro in Pande Game Reserve. (c) The transects are (stratified) randomly or systematically located.

Of the total disturbance transect length of ~575 km, Frontier Tanzania contributed 306.47 km, the Forest and Beekeeping Division of Tanzania 104.35, Museo Tridentino di Scienze Naturali (Francesco Rovero) 66 km, the Tanzanian Forest Conservation Group 42.75, the Wildlife Conservation Society Tanzania 35.8 km and Antje Ahrends 18.7 km; the latter dataset includes species identification and size measurements.

At present, use of both the vegetation plot and disturbance data requires memoranda of understanding with each data provider. Negotiations with the data contributors are underway to publish the data as a paper and therewith make it more easily accessible.



DisTrans
*
DataSet
DataByTrans
DisTransID
OrigTransect||WorkUnitID
OrigPlot||TransectID
Forest
ForestReserveID
ForestID
MountainID
TransLength
TransWidth
TransArea
CensusDate
CensusYear
OtherData
SamplingType
SamplingDescription
Habitat
TransectDirection
AltitudeStart
AltitudeEnd
Lat_Start
Long_Start
UTMS37_N_Start
UTMS37_E_Start
Lat_End
Long_End
UTMS37_N_End
UTMS37_E_End
CoordinateAccuracy
CoordinateOrigin
DisTransID_old

DisTransData
*
DataSet
DataByTrans
DisTransID
No_pole_live
No_pole_nd
No_pole_oc
No_pole_nc
No_tree_live
No_tree_nd
No_tree_oc
No_tree_nc
No_pole_cut
No_tree_cut
TotalCuts
Charcoal_new
Charcoal
Pitsaw_new
Pitsaw
Settlement
CultivationFallowCleared_new
CultivationFallowCleared
FireDamage_new
FireDamage
TimberPoles_new
TimberPoles
Mining
Trap_new
Trap
Other_new
Other
Path_new
Path
Animal
Grazing
Camp_new
Camp
Firewood
TrackRoad

DisTransSec
*
DisTransSecID
DisTransID
OrigTransect||Work
OrigPlot||TransectI
SectionStart
SectionEnd
Forest
ForestReserveID
ForestID
MountainID
VegTypeID
Length(m)
Width(m)
Area(ha)
Altitude
Slope
Aspect
Topography
CanopyCover
GroundCover
ShrubCover
CanopyHeight
Features
Comments
DisTransSecID_old
DisTransID_old

DisTransSecData
*
DisTransSecID
DisTransID
SectionStart
SectionEnd
SectionLength
No_pole_live
No_pole_nd
No_pole_oc
No_pole_nc
No_tree_live
No_tree_nd
No_tree_oc
No_tree_nc
No_pole_cut
No_tree_cut
TotalCuts
Charcoal_new
Charcoal
Pitsaw_new
Pitsawing
Settlement
Cultivation/Fallow/Cleared_new
Cultivation/Fallow/Cleared
FireDamage_new
FireDamage
TimberPoles_new
TimberPoles
Mining
Trap_new
Trap
Other_new
Other
Path_new
Path
Animal
Grazing
Camp_new
Camp
firewood
Track/road
DisTransSecID_old

Figure App.3. Screenshot of the disturbance transect database. (Some fields have been omitted to ease visualisation). Note that not all rows contain data for all fields. Metadata tables similar to the ones in the vegetation plot database are in production.
]

References

African Flowering Plants Database (2008). In. Conservatoire et Jardin botaniques de la Villle de Genève and South African National Biodiversity Institute Pretoria. http://www.ville-ge.ch/musinfo/bd/cjb/africa/.

Doody K.Z., Howell K.M. & Fanning E. (2001). Udzungwa Mountains Biodiversity Surveys – Methods Manual. In: *Frontier Tanzania Technical Reports*. The Society for Environmental Exploration and The University of Dar es Salaam, Dar es Salaam.

Eastern Africa Plant Red List Authority (2008). List of Potentially Threatened Plant Taxa in the Eastern Arc Mountains of Tanzania. Unpublished report presented to the Forestry and Beekeeping Division, Tanzania Ministry of Natural Resources and Tourism. (ed. Africa Plant Red List Authority) Dar es Salaam.

Lovett J.C., Ruffo C.K., Gereau R.E. & Taplin J.R.D. (2006). *Field Guide to the Moist Forest Trees of Tanzania.* The Society for Environmental Exploration, London and Dar es Salaam.

Mwakalila S., Burgess N., Ricketts T., Olwero N., Swetnam R., Mbilinyi B., Marchant R., Mtalo F., White S., Munishi P., Malimbwi R., Smith C., Jambiya G., Marshall A., Madoffe S., Fisher B., Kajembe G., Morse-Jones S., Kulindwa K., Green R., Turner R.K., Green J. & Balmford A. (2009). Valuing the Arc: Linking Science with Stakeholders to Sustain Natural Capital. *The Arc Journal*, 23, 25-30.