Spatial Modelling, Phytogeography and Conservation in the Eastern Arc Mountains of Tanzania and Kenya

Philip John Platts

PhD

THE UNIVERSITY OF YORK

York Institute for Tropical Ecosystem Dynamics

Environment Department

March 2012

This thesis is dedicated to my father
Philip Graham Platts

27th November 1937 – 11th February 2012

Thesis Abstract

Forests in the Eastern Arc Mountains are amongst the oldest and most biodiverse on Earth. They are a global priority for conservation and provide ecosystem services to millions of people. This thesis explores how spatial modelling can provide direction for conservation and botanical survey, and contribute to understanding of phytogeographical relationships. The ecoregion is rigorously defined by terrain complexity, vegetation distribution and established geoclimatic divisions, providing a coherent platform upon which to collate and monitor biological and socioeconomic information. Accordingly, 570 vascular plant taxa (species, subspecies and varieties) are found to be strictly endemic. The human population exceeds two million, with median density more than double the Tanzania average. Population pressure (accrued across the landscape) is shown to be greatest adjacent to the most floristically unique forests. Current knowledge on species distributions is subject to sampling bias, but could be systematically improved by iterative application of the bioclimatic models presented here, combined with targeted fieldwork. Tree data account for 80% of botanical records, but only 18% of endemic plant species; since conservation priorities differ by plant growth form, future fieldwork should aim to redress the balance. Concentrations of rare species correlate most strongly with moisture availability, whilst overall richness is better predicted by temperature gradients. Climate change impacts are projected to be highly variable, both across space and between species. Concordant with the theory that past climatic stability facilitated the accumulation of rare species, contemporary climates at sites of known endemic richness are least likely to be lost from dispersal-limiting mountain blocs during the 21st century. Faced with rapid population growth and the uncertainty of climate change, priorities for governance are to facilitate sustainable forest use and to maintain/restore habitat connectivity wherever possible. Overall, the thesis demonstrates that decision makers concerned with biodiversity conservation, particularly in mountain and coastal regions, should be wary of inferring local patterns of change from broad-scale models. The current study is a step toward spatially refined understanding of conservation priorities in the Eastern Arc Mountains, whilst novel methodologies have wider application in the fields of species distribution modelling and mountain analysis.

Contents

List of Tables

List of Figures

Acknowledgements

This research was funded by the Marie-Curie programme of the European 6th Framework, under contract MEXT-CT-2004-517098 to Dr. Rob Marchant. Additional funding for meetings and workshops was provided by the Leverhulme Trust, the Royal Society and Diversitas. I am grateful to the Tanzanian Commission for Science and Technology (COSTECH) and to the Forestry and Beekeeping Division of Tanzania for permission to conduct fieldwork. For logistical support, I thank WWF-Tanzania, the British Institute in East Africa and the University of Dar es Salaam. Many thanks to Dr. Rob Marchant for his supervision, support and unflappable calm, and to Dr. Colin McClean for additional supervision and modelling advice. Thanks also to Prof. Jon Lovett and Prof. Neil Burgess for sharing their knowledge of the Eastern Arc Mountains, and to Kathryn Addison and Dave Hay for help with computing and administration.

The analytical chapters of this thesis would not have been possible without field data contributed by Antje Ahrends, Roy Gereau, Jon Hall, Jon Lovett, Andy Marshall, Frontier-Tanzania and the Missouri Botanical Garden. Special thanks go to Antje and Roy for their heroic efforts collating and cleaning tens of thousands of plant records – their botanist brains were invaluable throughout. I thank Mark Mulligan, Stefan Hagemann and Daniela Jacob for providing climate surfaces, Petri Pellikka and Valuing the Arc for land cover data, the Cartography Laboratory and Map Office in Dar es Salaam, and Sue White and Biniam Ashagre for advice on hydrological modelling.

I owe thanks to Alex Ngowi, the priests and villagers at Kilomeni Mission in North Pare, staff at CEFA Hostel in Dar es Salaam, the field assistance of Mhiua Mapunda, Paul Kihara and Jafed in Nguu, and to Omar Majani for his extraordinary tree climbing skills. During the final stages of write-up, I was lucky to have the support of Prof. Andrew Balmford, Dr. Ruth Swetnam and Dr. Julian Bayliss in Cambridge. Thanks to KITE friends, Alistair Jump, Antje Ahrends, Jemma Finch, Andy Marshall, Cassian Mumbi and Stephen Rucina, and also to Julia Latham, Liz Curmi and Marion Pfeifer. I am grateful to Dr. Steve Willis and Prof. Dave Raffaelli for examining the thesis.

Thank you to my family, especially Mum and Dad, and to Loz, Sophie and Edie for looking after us in Cambridge. Most of all, I thank Hayley, for patience and encouragement, and for giving us our beautiful son Leonard.

Author's Declaration

This thesis represents my own work and has not been submitted previously to the University of York or any other institution for a degree, diploma or other qualification. Colleagues who contributed data or supervision are included as co-authors on the respective chapters, wherein the nature of their contributions are detailed. However, I played the dominant role in all model design, data analysis, interpretation and writing.

Signed

Philip John Platts
March 2012

“*Nature favours those organisms which leave the environment*
in better shape for their progeny to survive”

James Lovelock

Chapter 1 – General Introduction

Overview

Plants are the physiological foundation of all terrestrial ecosystems, harnessing the sun's energy to convert carbon dioxide and water into glucose and oxygen. Animals breathe the oxygen, consume the sugars and utilise vegetative architecture and associated soil formations to build nests, dens and burrows. Plants regulate nutrient cycling, hydrological flows and local climate, and impact regional and global climate *via* albedo, transpiration and carbon fixation. For the perpetuation of Earth's most unique and threatened ecosystems in the face of human pressures and environmental change, estimates of current and future distributions of plant species are crucial (Fig. 1.1).

One sixth of the world's vascular flora is endemic to the African continent, of which many species are confined to tropical mountain forests in Tanzania, Kenya, Uganda, Cameroon, Guinea and Côte d'Ivoire (Scholes *et al.*, 2006). Tropical forests are the most biologically diverse ecosystems on the planet (Gentry, 1992), housing over half of the world's biological diversity in just 7% of the land area (Bartholome and Belward, 2005). They contain a quarter of the carbon in the terrestrial biosphere and sequester large amounts annually, so affecting global circulation patterns (Bonan, 2008; Lewis *et al.*, 2009). The 20th century saw rates of tropical deforestation and degradation accelerate to critical levels, but also an increased awareness of the consequences, particularly regarding biodiversity loss, climate regulation and impacts on human well-being (Millennium Ecosystem Assessment, 2005). In recent decades, a number of scientific panels and political commitments have been established in an effort to mitigate long-term damage (e.g., Intergovernmental Panel on Climate Change, http://www.ipcc.ch/; Intergovernmental Platform on Biodiversity and Ecosystem Services, http://www.ipbes.net/; Convention on Biological Diversity, http://www.cbd.int/; Reduced Emissions from Deforestation and Forest Degradation, http://www.un-redd.org/). Meanwhile, advances in computational power and analysis tools, data collation and dissemination, have facilitated new scientific methodologies. Remotely sensed data, climate simulations, dynamic vegetation models and species distribution models allow us to monitor and predict how natural systems respond to climatic and socio-economic change (Asner *et al.*, 2010; Doherty *et al.*, 2010; Franklin, 2010; Swetnam *et al.*, 2011; Pfeifer *et al.*, 2012).

This thesis first introduces some background information on threats to forest habitats, issues regarding their conservation and the role of spatial modelling. The focus then centres on the Eastern Arc Mountains of Tanzania and Kenya, one of the most important and challenging sites for conservation globally (Myers, 1990; Brooks *et al.*, 2002). The Eastern Arc and its unique forest habitats provide the context for the analytical thesis chapters that follow.

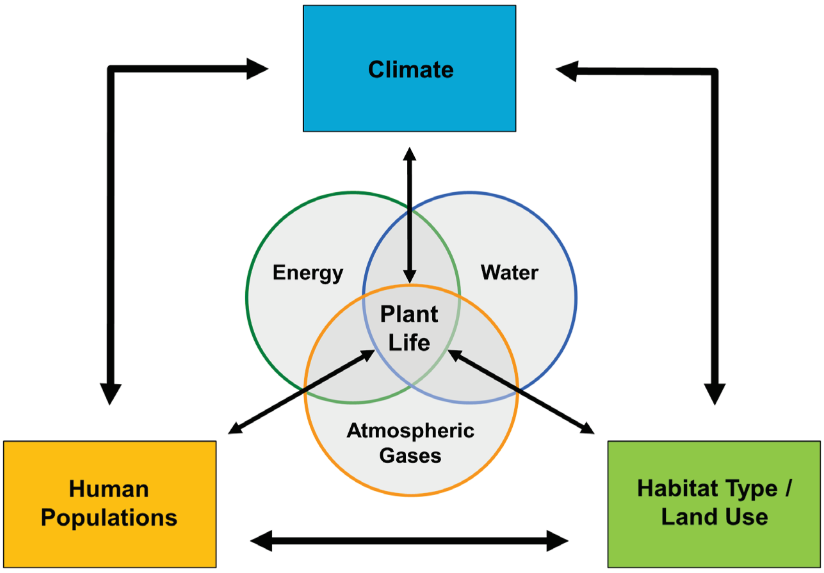


Figure 1.1. Key components of the Earth system, with plant life at the centre.

Forest degradation and loss

In a synthesis of satellite imagery and published reports, Asner *et al.* (2009) found that around half of the humid tropical forest biome has less than 50% tree cover remaining, and that between the years 2000 and 2005 approximately one fifth was subject to timber harvesting. According to the FAO, 131 million ha (20%) of African forest was affected by fire each year between 1998 and 2002. Fire is a naturally occurring ecosystem process, but the frequency of burning has increased due to agriculture and charcoal production, to the detriment of carbon stocks and biodiversity (Tansey *et al.*, 2004).

One of the earliest records of forest clearance by humans dates back some 9000 yBP (Ghab Valley, Syria; Yasuda *et al.*, 2000). It is, however, more recently that technological advancements, international trade, increased permanence of settlement and population growth have resulted in the large-scale removal of forests for timber, wood-based fuel, and to make way for arable and grazing lands (DeFries *et al.*, 2010). The industrial revolution of the 18th and 19th centuries greatly accelerated deforestation rates, especially in Europe, North

America and parts of Asia. Today, clearance of temperate forests has mostly abated: existing farmlands, plantations and imports are sufficient to support human populations, and there has been a general increase in the proportion of these forests conserved for their biological diversity (FAO, 2006). Forest cover in Europe and Asia is actually increasing, either because of expansion into abandoned agricultural lands or because of afforestation programs, particularly in China (4 million ha.y^{-1} 2000-2005; FAO, 2006). These secondary forests help to offset carbon stocks lost elsewhere, but are typically less biodiverse compared with primary forests (e.g., due to simpler vertical structure).

Whilst accrued wealth and low population growth rates in economically developed countries have eased anthropogenic pressures on the temperate forest biome, forest habitats in the tropical zone are under severe threat from disturbance and loss, with catastrophic consequences for biodiversity, climate regulation and sustainable resource provision. Driven by rapidly expanding human populations and associated pressure on natural resources, Africa alone suffered a net forest loss of over 4 million ha annually (0.62%.y^{-1}) between 2000-2005, equivalent to 110 km^2 per day (FAO, 2006). Deforestation rates in South America were similar (0.5%.y^{-1}; 2000-2005), but in Brazilian Amazonia have slowed in recent years, thanks to dramatic changes in political, economic and legal frameworks (Hecht, 2012). In the neotropics and south-east Asia, export markets and urban growth are the primary drivers of change (DeFries *et al.*, 2010). In Africa, local demand for small-scale agricultural plots and for wood-based fuels, such as charcoal, remain most important (Ahrends *et al.*, 2010; Fisher, 2010). In Tanzania, 90% of energy consumed is wood fuel originating from forest habitats (Milledge *et al.*, 2007).

The FAO define deforestation to be the depletion of crown cover to below 10%. This implies a somewhat generous definition of ‘forest’, with the consequence that heavily degraded habitats are omitted from estimates of loss. Rates of forest degradation (as opposed to outright clearance) resulting from, for example, shifting agriculture, selective logging and unsustainable extraction of non-timber forest products (NTFPs), are very difficult to assess remotely, but can have significant long-term impacts on forest composition and function.

Conservation prioritisation

Limited funds for tackling unprecedented rates of biodiversity loss dictate targeted, science-led approaches to conservation planning (Margules and Pressey, 2000). Without setting priorities, resources would be spread too thinly to maximise species conservation (Pimm *et al.*, 2001; Myers, 2003). Global priority metrics focus on the numbers of rare and endemic taxa, especially amongst flowering plants, and on the degree of threat (Olson and Dinerstein, 1998; Myers *et al.*, 2000; Brooks *et al.*, 2006; Wilson *et al.*, 2006). Biologically rich areas with high irreplaceability, which are also vulnerable to degradation and loss, are given the highest rank. At more local scales, patterns of biodiversity and threat can vary considerably, as can the costs of conservation (Naidoo and Ricketts, 2006), so further analyses are necessary to determine action on the ground (Mace *et al.*, 2000; Ferrier, 2002).

Priority metrics are unavoidably sensitive to the information available at the time of assessment. Conservation decisions are spatially and taxonomically biased by previous levels of investment, site accessibility and historical patterns of exploration (Reddy and Davalos, 2003; also Ahrends *et al.*, 2011a in Appendix I). Consequently, many areas remain under-valued due to a lack of knowledge (Küper *et al.*, 2006; Grand *et al.*, 2007). Fine-scale data on the distribution of species are very time consuming to collect and, particularly for rare species, require high levels of taxonomic expertise (Ahrends *et al.*, 2011b in Appendix I). In tropical forests, exhaustive sampling is logistically and financially prohibitive. Whilst further exploration and taxonomy are undoubtedly required, many unique habitats are being lost or degraded faster than we can document the species they contain. Modelling taxonomic effort over time, Joppa *et al.* (2011) estimate that *c.* 15% of flowering plant species have yet to be discovered, and that Tanzania in particular will become relatively much richer as the missing species are accounted for. Another study, which identifies predictable patterns in the allocation of species to higher taxonomic ranks, suggests that 86% of all species, 28% of plants, await description (Mora *et al.*, 2011). In prioritising sites for conservation, a key trade-off is between ‘counting to extinction’ and acting upon proximate estimates, whereby gaps in knowledge are addressed using expert opinion and, more recently, spatial modelling of biophysical indicators (Pressey *et al.*, 2000).

Species distribution modelling

Species distribution modelling (SDM) – essentially, the spatial interpolation of known species occurrence – has become a popular tool for guiding conservation effort. The rapid progression in SDM over recent decades coincides with advances in geographical information systems (Fig. 1.2), mapping tools that have greatly facilitated its implementation. Free internet access to large databases of herbaria specimens and other species information (Graham *et al.*, 2004), digitised environmental data and a latent demand for predictive spatial models, have combined to accelerate the popularity of SDM in scientific research.

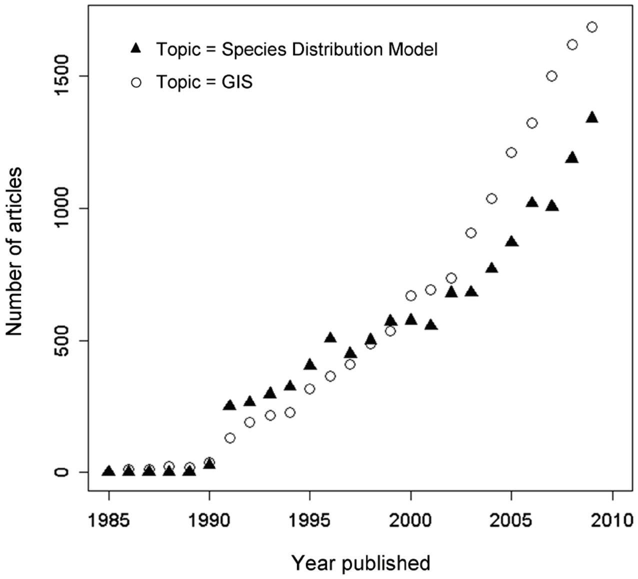


Figure 1.2. Searching the ISI Web Science for the keywords "species distribution model" and "GIS" (geographical information system) shows an explosion in the number of articles making reference to these fields since 1990. The correlation between the two trends is highly significant (Pearson's $r = 0.96$; $p = 5.97 \times 10^{-14}$).

Whilst modelling approaches are not immune to bias in the data (Lozier *et al.*, 2009; Chapman, 2010), careful application of statistical, rule-based or expert-derived models can provide planners with valuable insights into predictable controls on species' realised or potential niches (Loiselle *et al.*, 2003; Kearney *et al.*, 2010). Unvisited sites estimated to be viable for species of conservation concern may be targeted for future research (Guisan *et al.*, 2006), making the most of dwindling funds for inventory work (Lawton *et al.*, 1998) and potentially uncovering areas of high conservation importance (e.g., Raxworthy *et al.*, 2003; Menon *et al.*, 2010). Spatial predictions summed over species and/or taxonomic groups are useful for biodiversity assessment, or else richness can be modelled directly as a function of environmental factors. Such applications involve assumptions regarding biotic interactions, non-equilibrium states and past colonisation events, but nonetheless can yield useful information when considered at the appropriate spatial scales (Araújo and Peterson, 2012; Godsoe and Harmon, 2012).

Austin (2002) describes three components to the SDM framework: the ecological model, the data model and the statistical model. The ecological model is an *a priori* understanding or hypothesis of which factors might limit a species distribution. The data model consists of georeferenced observations for the response (species occurrence, diversity or density) and regional maps of environmental gradients (predictor variables). The statistical model is a mathematical basis or rule set for parameterising the species-environment relationship and for assessing uncertainty in results. Applications of SDM in ecology and conservation are increasingly far-reaching: biodiversity assessment; range estimation and improved sampling for rare and endangered species; risk assessment for invasive species/pathogens; reserve design; habitat management and restoration; historical climate/vegetation reconstruction; and impact assessment for climate and land use change (Elith *et al.*, 2006; Franklin, 2010).

Describing the response

Central to any SDM application is a form of ecological gradient analysis, which seeks to describe how species patterns correlate with environment factors across a landscape (Whittaker, 1967). Simple predictive models can be obtained by constructing bioclimatic or 'multidimensional niche' envelopes, defined variously according to the environmental limits of a documented species distribution (Busby, 1991; Walker and Cocks, 1991). Envelope models can yield a summary of environmental conditions across sample sites, and predict species presence or absence elsewhere (Fig. 3a). More complex methods, such as logistic regression, offer insight into the functional forms of species response (Fig. 3b-d).

Early gradient analysis posited that species' responses to environmental gradients should follow symmetric unimodal distributions (Whittaker, 1956), approximated by Gaussian functions in generalised linear models (GLMs). Later work demonstrated that observed response shapes are often skewed, bimodal or even more complex (Austin, 2002). Because distribution data necessarily document the realised niche, modelled relationships can deviate from the underlying functional response, distorted by factors beyond the scope of the model such as competition, disturbance and barriers to dispersal – anything that restricts a species from establishing in otherwise suitable areas (Fig. 1.3). A bimodal fit is perfectly plausible given competitive interactions, for example, even if the underlying response is bell-shaped (Austin and Smith, 1989).

Non-parametric methods, such as generalised additive models (GAMs), are now widely used in plant ecology. GAMs fit piecewise curves constructed from smoothing splines, each focussed on a distinct portion of parameter space (Fig. 1.3c). Thus, the functional form of response is driven by the data, rather than *a priori* by the analyst (Yee and Mitchell, 1991). In comparative studies, GAM usually outperforms strictly parametric methods in species distribution modelling (Meynard and Quinn, 2007).

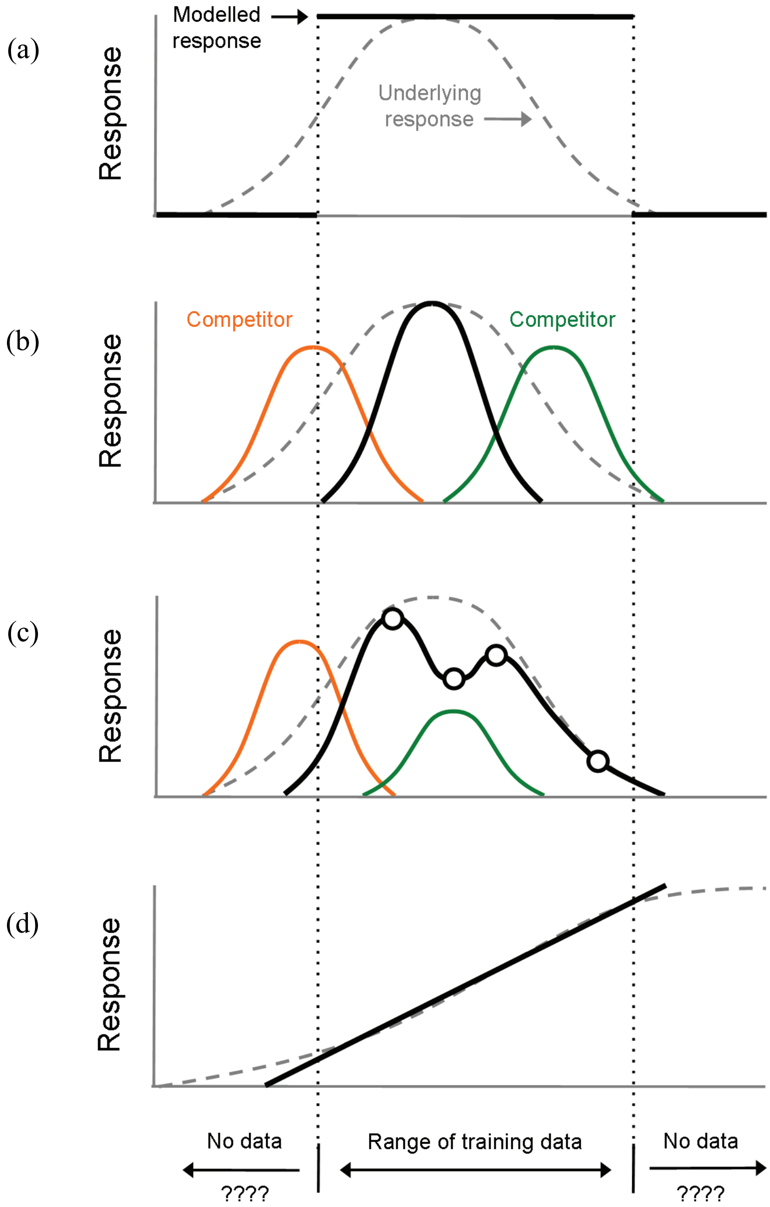


Figure 1.3. Modelled response of a hypothetical species distribution (y-axis) to an environmental predictor (x-axis) where the underlying response shape is Gaussian. (a) Simple envelope model predicts a dichotomy of species presence and absence. (b) Linear regression fits a response narrower than the fundamental niche due to competition near the edges of the species' environmental range. (c) As in b, but competitive interactions are unequal and biased towards the lower end of the gradient; here, non-parametric methods such as GAM (diagram shows splines joined at knots) are used to model the realised niche. (d) Straight line response fitted to a wider-ranging species, because only the lower end of its range has been sampled. In all cases, prediction uncertainty is highest beyond the limits of training data.

Prediction

Gradient analysis is essential to ensure an ecologically meaningful model, and can itself provide useful information for conservation. However, the principle application of SDM in conservation planning is spatial prediction. That is, extrapolation of the modelled relationships to predict species occurrence, abundance or diversity across a landscape. Assuming that the target species' relationship with its environment remains unchanged over the time span in question (niche conservatism), models can also be extrapolated for historical reconstruction or for exploring scenarios of future change. In this respect, space and time are interchangeable: both are novel parameter space for extrapolation.

In spatial prediction, intra-specific responses are assumed similar across a landscape, unless explicitly parameterised otherwise using, for example, geographically weighted regression. The strength of this assumption depends on many factors, including levels of genetic variation, especially between disjunct populations, and differences in community composition across the study region. Extrapolating models to different points in time raises similar issues, and in addition the fact that species are constantly evolving and adapting to changes in their biotic and abiotic surroundings (Dobrowski *et al.*, 2011). Thomas *et al.* (2001) report examples of rapid niche adaption in Britain over a 20 year period: butterfly species have colonised new habitat types, and species of bush cricket have dispersed to previously unreachable locations, facilitated by new longer-winged phenotypes. Conversely, a study of birds, mammals and butterflies in Mexico provides evidence for niche conservatism spanning millions of years (Peterson *et al.*, 1999). Thus, the strength of temporal SDM prediction is dependent on the time span, species and region in question (Peterson, 2011). Dominant species tend to exhibit less variation in their realised niches than inferior competitors (Pearman *et al.*, 2008). Long-lived sessile organisms with low dispersal capacity, such as large tree species, may also conserve their niches for longer, due to the reduced opportunity for selection and adaption to new habitats (Holt, 1996).

Envelope models make the transparent assumption that the entire bioclimatic range of the target species has been sampled: conditions outside the observed niche (or some percentile thereof) are deemed strictly and uniformly unsuitable; conditions within are considered uniformly viable for the species (Fig. 1.3a). SDM methods that explicitly model the shape of the response, such as GLM and GAM, do not impose such a dichotomy and therefore permit non-zero predictions beyond the environmental limits of the training data (Fig. 3b-d). Since the species/background data may or may not be sufficient to represent the entire study region, either now or in the future, such extrapolations are subject to high uncertainty. This aspect of model validation is particularly pertinent given current concern over climate change. Governments, conservation organisations and scientists are increasingly faced with the challenge of predicting ecosystem response to novel climates (Williams *et al.*, 2007), yet model uncertainty beyond the environmental limits of training data is often overlooked (Thuiller *et al.*, 2004; Araújo and Guisan, 2006; Pearson *et al.*, 2006).

Thesis aims and objectives

The analytical chapters of this thesis (Chapters 2 through 5) explore some of the issues raised above in the context of the Eastern Arc Mountains in Tanzania and Kenya (EAMs). This is the smallest hotspot of biodiversity identified by Myers (1990), characterised by complex topography and steep climatic gradients, as well as regional and seasonal fluctuations in temperature and rainfall, precluding meaningful inference from global- or continental-scale modelling.

Due to the breadth of conservation and livelihood challenges in the EAMs, the methods and findings presented here should have wider relevance, especially to other Afromontane systems (Fig. 1.4). In particular, the thesis provides methods for defining mountain endemism, for assessing uncertainty in niche-based model predictions, for downscaling climate forecasts, and for linking future climate predictions with hypotheses of past ecoclimatic stability. Specifically, the thesis aims to:

1. Establish a quantitative and repeatable method for delineating the EAMs
2. Advance SDM methodologies and identify key predictors of EAM plant distributions
3. Predict fine-scale patterns of plant richness and rarity
4. Highlight the potential biological importance of lesser-known mountain blocs
5. Provide high-resolution, regionally-forced climate forecasts for the EAMs, and so investigate potential climate change impacts on some of the Earth's rarest species

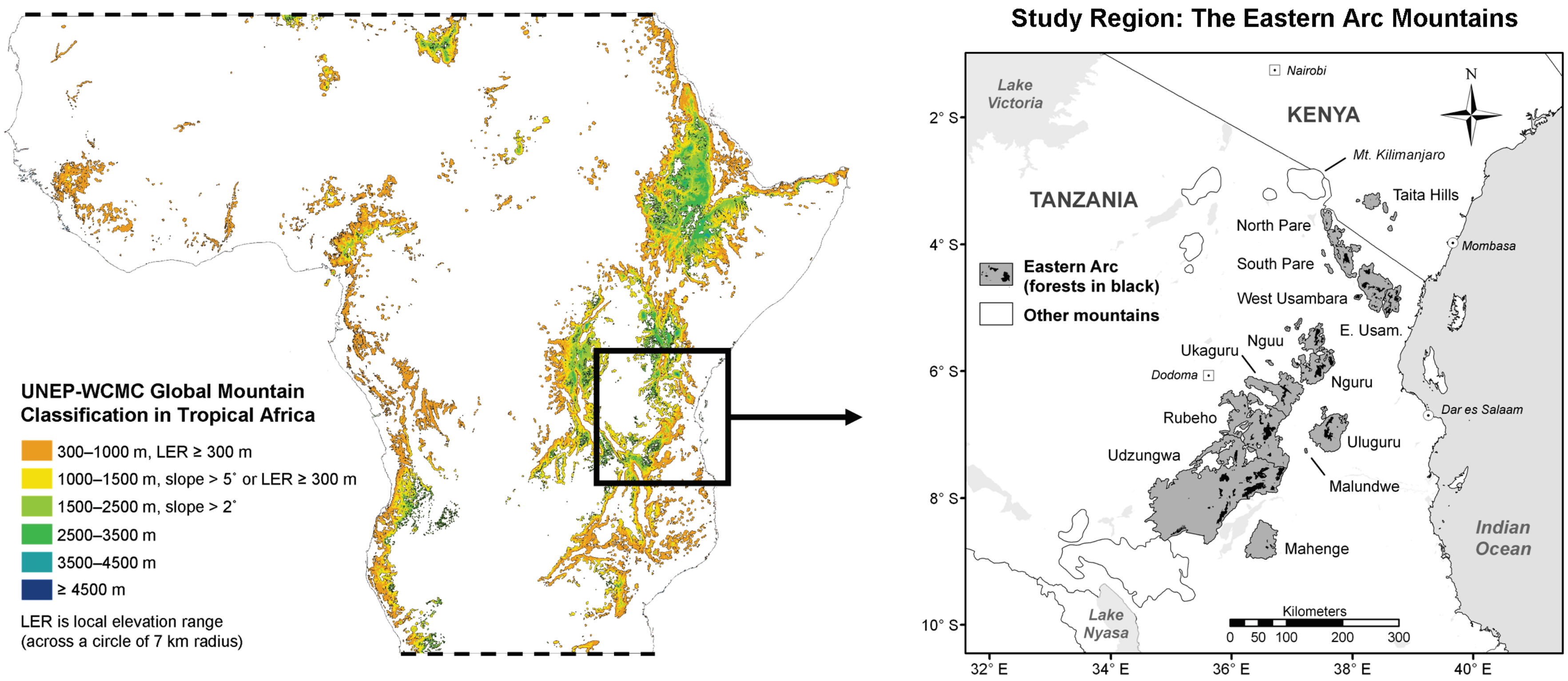


Figure 1.4. Left: mountain habitats in tropical Africa (Kapos *et al.*, 2000); at regional to local scales, there is scope to refine mountain delineation using higher resolution elevation and vegetation data, as shown here for the Eastern Arc (derivation in Chapter 2). Right: crystalline mountain blocs (grey) that comprise the Eastern Arc chain; adjacent highlands have different substrate (e.g., Mt. Kilimanjaro) or have climatic regimes that are influenced more by the Great African Lakes than by the Indian Ocean (e.g., Lake Nyasa Highlands).

Study region

Geology and climate

The EAMs consist of 13 disjunct ranges ('blocs'), stretching from the Taita Hills in southeast Kenya to the Udzungwa Mountains in south-central Tanzania (Fig. 1.4). The substrate is Precambrian crystalline basement rock, uplifted from the African plateau during the Miocene *c.* 30 MY BP (Schlüter, 1997), although initial faulting may have begun over 100 MY earlier (Griffiths, 1993). Elevations range from 121-2636 m above mean sea level (the summit is Kimhandu Peak in the Uluguru range).

Mean annual temperatures across the study region range from 12.4-24.1 ˚C (mean, 20.7 ˚C; WorldClim interpolated climatology; Hijmans *et al.*, 2005). The warmest months are November through March, with mean daily maxima exceeding 34 ˚C on lower slopes. The coolest months are June through August, when mean daily minima drop below 5 ˚C at high altitudes (frosts are not uncommon in the upper-montane zone). Slopes close to the Indian Ocean are several degrees cooler than equivalent altitudes elsewhere. Air temperatures in the understory vary according to distance from the forest edge and the degree of solar radiation penetrating the canopy (Newmark, 2001; Newmark, 2005).

Annual rainfall is typically in the range 500-2000 mm.y^{-1}, but in some blocs exceeds 3000 mm.y^{-1} (precipitation-radar measurements, 1997-2006; Mulligan, 2006). Windward eastern aspects are wettest due to orographic rainfall and mist driven by Indian Ocean currents (Marchant *et al.*, 2007) and sustain continuous broadleaved forest cover where land use allows. Drier western flanks tend to support deciduous woodland rather than moist forest assemblages (Newmark, 1998), whilst grassland and heathland are common on the uppermost montane plateaus (Finch and Marchant, 2011). The north of the study region is generally recognised to have two distinct peaks in rainfall: short rains from March to May and heavier rains from October to December. In the south, one rainy season prevails from December to April. Analysis of recent precipitation-radar measurements suggests a rather more complex and variable distribution of seasonal rains. A consistent pattern is that East Usambara and Uluguru experience the most benign dry seasons, with some eastern aspects perhumid (> 100 mm.m^{-1}; Pócs, 1976). Drier spells are most prolonged and intense on western margins of Ukaguru, Rubeho and Udzungwa, which receive little rain for three or more consecutive months.

Biological importance

The EAMs support around 3300-5700 km^2 of moist tropical forest (Newmark, 2002; Burgess *et al.*, 2007a), including some of the world's most important and fragile sites for biodiversity conservation (Brooks *et al.*, 2002; Mittermeier *et al.*, 2004). The mountains contain over 14% of the vascular plant species indigenous to tropical Africa in less than 0.25% of the land area[1]. Of these species, 471 are strictly endemic (Chapter 6), including some now common in western households (Fig. 1.5).

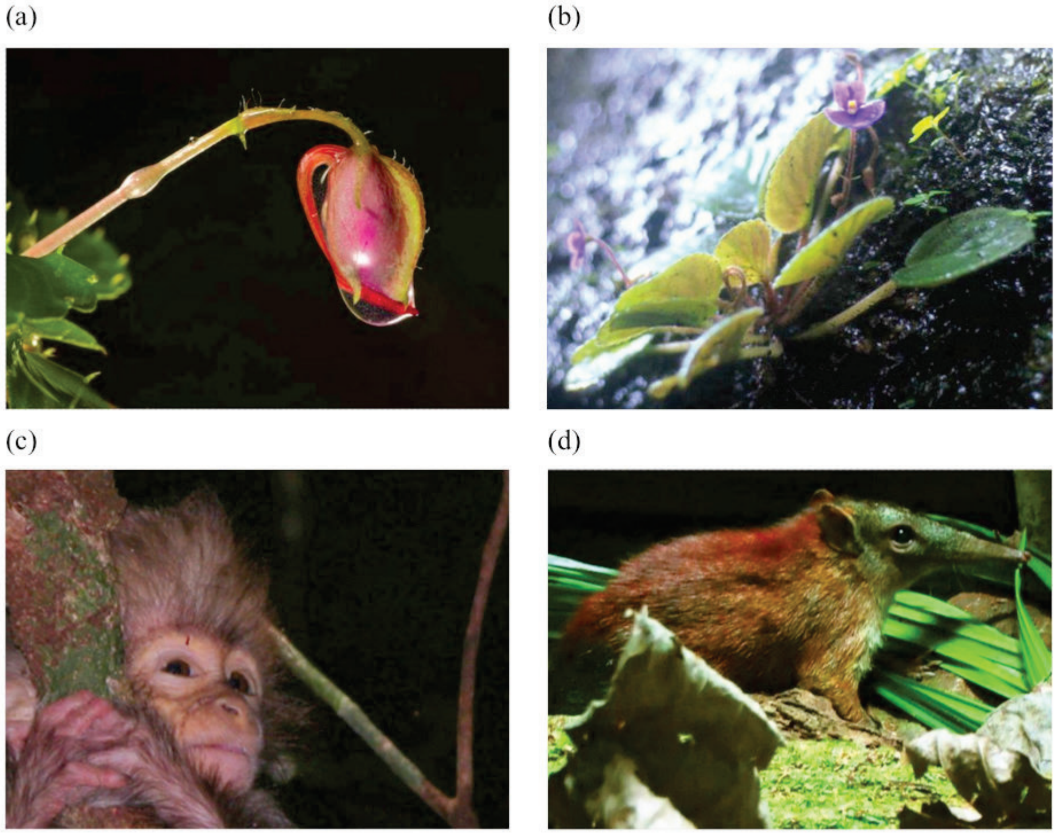


Figure 1.5. Flora and fauna of the EAMs. Well known houseplants: (a) Busy Lizzie (*Impatiens* sp.), (b) African Violet (*Saintpaulia* sp.). Recently discovered: (c) critically endangered *Kipunji* genus, (d) grey-faced sengi / 'elephant shrew' (*Rhyncocyon udzungwensis*) – both in the 'top ten' new species of the decade (Conservation International and BBC Natural History Unit, 2010). Photo credits: a, Michele Menegon; b, Nobby Cordeiro; c, Trevor Jones; d, Francesco Rovero.

---

[1] Excluding Madagascar and inland water bodies, tropical Africa covers *c*. 22 million km^2 and contains 26,848 indigenous vascular plant species (AFPD, 2010, corrected to remove introduced species; R.E. Gereau and H. Beentje, pers. comm.). The EAMs, as here defined in Chapter 2, cover *c*. 52,000 km^2 (including plateaus) and contain 3845 indigenous vascular plant species (Gereau *et al*., in prep.).

A number of forest-dependent birds and mammals are endemic, such as the Usambara eagle owl (*Bubo vosseleri*) and the recently discovered grey-faced sengi (*Rhyncocyon udzungwensis*; Fig. 1.5). Concentrations of range-restricted herpetofauna and invertebrates are especially impressive – single site endemism within the latter is as high as 80% (ICBP, 1992; Lovett and Wasser, 1993; Stattersfield *et al.*, 1998; Burgess *et al.*, 2007a; Poynton *et al.*, 2007). Faced with ongoing habitat loss and degradation, the EAMs are considered a 'hyper hot' priority for conservation (Myers *et al.*, 2000; Brooks *et al.*, 2002).

Delineation and classification

The term 'Eastern Arc' was first established in the 1980s, when botanical explorations in the region were documenting large numbers of forest species of highly restricted distribution (Lovett, 1985). Some mountains in the chain were already known for their biological importance, notably in the Usambara and Uluguru ranges (Polhill, 1968; Pocs, 1976). Then in the late 1970's, research in the Udzungwa bloc revealed that a number of the Usambara endemics also occurred farther south, indicating that the centre of endemism may be more extensive. As scientific exploration continued, rapid and widespread exploitation was taking hold, particularly mechanised logging in the Usambaras and eastern Udzungwa (Bjorndalen, 1992). In the absence of detailed scientific surveys for many of the blocs, it became imperative to formulate a predictive definition for the area of endemism (J.C. Lovett, pers. comm.). In 1988, a geological and climatological description of the Eastern Arc was established, defining the forests as "those occurring on crystalline mountains in south-east Kenya and eastern Tanzania under the direct climatic influence of the Indian Ocean" (AETFAT conference, Hamburg; see Lovett, 1990). The term Eastern Arc was subsequently adopted by the Tanzania Forestry Action Plan to delimit an area of high conservation importance (Bensted-Smith and Msangi, 1989). Soon after the mountains were recognised globally for their biodiversity value (Myers, 1990).

According to White's (1983) classification of African vegetation, the Tanzanian flora consist of Afromontane, Zanzibar-Inhambane and Lake Victoria phytogeographical types. Lovett's (1990) refinement of these spatial divisions identifies the EAMs according to their soil type and main climatic influence – proxies for long-term isolation and climatic stability, factors hypothesised to be central to the exceptional concentrations of endemism (Lovett and Wasser, 1993; Fjeldså and Lovett, 1997). Forests influenced predominantly by the climatic regimes of the Great African Lakes are not considered part of the Eastern Arc; Lake Victoria was considerably smaller during the last glacial maximum, disrupting rainfall patterns and associated forest vegetation (Hamilton, 1982). Forests on volcanic soils are also omitted, these mountains being geologically younger (e.g., Mt. Kilimanjaro, 1-2 MY old; Schlüter, 1997) and containing fewer species of restricted distribution.

Mittermeier *et al.* (1998) originally combined Tanzania's coastal forests and those of the EAMs within a single biodiversity hotspot. This classification was later revised, with the mountains now belonging to the Eastern Afromontane Biodiversity Hotspot (Mittermeier *et al.*, 2004). From a taxonomic standpoint, the split is controversial due to submontane and coastal habitats in Tanzania sharing a large number of range-restricted taxa (279 vascular plant species; Gereau *et al.*, in prep.). From a conservation perspective, the grouping of Afromontane habitats is arguably appropriate to address the management challenges specific to mountain regions, their ecology and people (Kreutzmann, 2001).

Scientific interest in the EAMs has grown in recent years, resulting in numerous assessments of biological and natural capital importance (e.g., CEPF, 2003; Burgess *et al.*, 2007a; Mwakalila *et al.*, 2009). Whilst the predictive, landscape-scale definition of Lovett (1990) has been widely applied, prior to the current work (Chapter 2) precise spatial limits for the area of endemism have been lacking – a result of the inherent subjectivity in defining mountain extent (Gerrard, 1990) and of broader uncertainties in delimiting areas of endemism (Anderson, 1994; Harold and Mooi, 1994). Thus, fundamental biogeographical questions such as "what is the area of the EAM?" and "how many (endemic) species occur there?" have been troublesome to address with rigour or consistency until now.

Forest use and protection

Humans have an exceptionally long history of forest use in East Africa (Marchant *et al.*, 2010). Some of the earliest hominoid remains were discovered close to the study region, in Olduvai Gorge, northern Tanzania. Occupation of the EAMs by hunter gatherers and subsistence farmers at least pre-dates the Common Era (Lovett and Wasser, 1993). Historically, subsistence use by small local communities and nomadic pastoralists, such as the collection of poles, firewood, food and medicinal plants, had minimal long-term impact on forest structure. Over the past 200 years, however, population growth and associated demand for agricultural land, wood-based fuels and other forest products, have impeded forest regeneration (Newmark, 1998; Finch and Marchant, 2011). Commercial practices such as mining, clearance for plantations and extraction of valuable timber species such as *Khaya anthotheca* (African mahogany), *Milicia excelsa*, *Pterocarpus* and *Olea* spp. have greatly impacted forest extent and condition, whilst returning little in the way of benefits to local communities (Bjorndalen, 1992; Burgess *et al.*, 2005). In 1984, commercial timber harvesting in catchment forest reserves was banned, but illegal pit-sawing continues in many areas. Present-day forests represent less than 30% of their preclearance extent (Newmark, 2002; Burgess *et al.*, 2007a; Hall *et al.*, 2009). There are around 200 fragments across the 13 mountain blocs (Fig. 1.4), with a mean patch size of 21 km^2 (median, 1.5 km^2; range, 0.01-526 km^2). In Taita, just 2% of the original forest area remains.

Figure 1.6. Examples of forest conversion and use. (a) Amani Nature Reserve in East Usambara: submontane forest on the left, tea estates on the right. (b) Forests capture fog inputs and regulate seasonal water flows. (c) Firewood collection day in Udzungwa National Park – the primary fuel for rural communities. (d) Terraced slopes in Chome Forest Reserve, South Pare. Photo credits: a-b, Philip Platts; c, Julia Latham; d, Jemma Finch.

Effective management is important across a range of spatial and thematic scales (Mwakalila *et al.*, 2009): globally, for biodiversity conservation and climate change mitigation; nationally, for water supplies, energy production, climate regulation, soil conservation and nutrient cycling; and locally, for communities directly reliant on forest products (Fig. 1.6).

Three quarters of the EAM forest area is gazetted, of which over half is designated for catchment protection or multi-resource use (Burgess *et al.*, 2007b). Levels of degradation within these forest reserves vary greatly from place to place, depending on levels of staffing, proximity to roads and settlements, and cultural traditions (Green *et al.*, 2012). Some are under Joint Forest Management, whereby local communities enter agreements with government regarding forest use. In addition, local communities manage forests on village land. Both participatory management schemes have potential to improve forest condition (Blomley *et al.*, 2008), but ambiguous tenure rights and often long-winded official ratification risk reducing incentives for sustainable use. More stringent, state regulation applies to nature reserves (gazetted for forest conservation) and national parks, which combined cover nearly a third of the total forest area. Strong management effectively reduces degradation within these sites, but might arguably lead to leakage (displacement of resource depletion to surrounding areas) or poverty if resource needs are not otherwise fulfilled (Pfeifer *et al.*, 2012 in Appendix I).

Data collection

The compilation of environmental data involved sourcing spatial estimates of topography, climate, soil type and land cover. Some datasets were freely available for use (e.g., elevation); others were unpublished or incomplete and obtained *via* collaboration with the data providers (e.g., land cover and cloud data). Additional map data, such as historical atlases and maps of the field sites were collected during field visits with kind assistance from the Cartography Laboratory and Map Office in Dar es Salaam. Botanical data for the thesis were sourced from two large databases. The first collates *c*. 70,000 plant records from over 2000 vegetation plot assessments, conducted by Antje Ahrends (Edinburgh Botanical Garden), Jon Hall (University of Wales, UK), Jon Lovett (University of Twente, Netherlands), Andrew Marshall (University of York, UK) and researchers at Frontier-Tanzania (for details see Ahrends, 2010). The second database, TROPICOS, provides collection localities for *c*. 18,000 herbarium specimens (http://tropicos.org/). Subsets of data used for modelling were cleaned in collaboration with field researchers and database managers.

In July and August 2007, I assisted in vegetation plot assessments targeting two of the most under-researched mountain blocs. The first site was in Nguru North Forest Reserve (Nguu), the second in Kindoroko Forest Reserve (North Pare). At each location, three 20×50 m plots were established, from which a total of 170 botanical specimens were collected. In Nguru North, moist submontane forest was found between 1400-1500 m, with canopy heights of 20-40 m. Species such as *Entandrophragma excelsum*, *Myrianthus holstii*, *Newtonia buchananii* and *Trilepisium madagascariense* were most common in the plots. The forest was in good condition with little evidence of disturbance. A number of fallen trees were observed with new growth in their wake, indicative of a mature forest dynamic.

Kindoroko plots were situated near the highest ridge in North Pare. Montane forest (1600-1800 m; canopy height, 20 m) was in generally good health, but with some evidence of pole cutting (for construction) and bark removal (medicinal use) near paths. Non-native *Eucalyptus* and agricultural plots were common around the forest edge. Upper montane forest (1800-2100 m; canopy height, 10-20 m) was not recently disturbed, but near the central ridge regenerated over heathland, perhaps following fire events. The most frequently collected specimens were *Albizia gummifera*, *Aphloia theiformis*, *Macaranga capensis*, *N. buchananii*, *Rapanea melanophloeos*, *Syzygium guineense*, *Tabernaemontana pachysiphon* and *Xymalos monospora* (Ahrends & Platts, unpublished data 2007).

Outline of analytical chapters

The core of the thesis is presented in the style of scientific articles, each accessible to a wide scientific audience *via* peer-reviewed journal publications. For the articles to be coherent bodies of work independent of the thesis *in toto*, there exists some necessary overlap between introductory sections. Together, Chapters 2 through 5 address the thesis aims and objectives described above. Chapter 6 is a summary discussion, which synthesises key findings and their impacts to date, and suggests potential directions for future research.

Chapter 2 – Mountain Limits

The importance of high-resolution mapping in the management of mountain systems and the services they provide means that precise definitions of spatial extent are increasingly required. Further, given the global importance of the EAMs as an area of endemism, it is essential that the meaning of “area” in this context be defined explicitly. Anderson (1994) wrote that “To speak of a taxon as endemic […] without specifying the area is meaningless”. Numerous reports, publications and websites consulted at the time of writing this thesis presented conflicting estimates of where the EAM boundaries lay, and thus of area, biodiversity and endemism within. Chapter 2 proposes an exact placement for EAM limits, derived according to quantitative rules for landform classification, the distribution of mountain vegetation and established phytogeographical divisions.

Chapter 3 – Predicting Tree Distributions

Despite a relatively long history of botanical exploration in the region, this chapter represents one of the first published accounts of regionally focussed plant distribution modelling in the EAMs. The study applies regression techniques to parameterise the responses of 40 large tree taxa (38 species) to a suite of climatic and topographic predictor variables. The results demonstrate that 'best-model' predictions can vary according to the method used for variable selection, and that a model average can yield a superior trade-off between precision and generality. The effects of weighting data (to standardise sample prevalence) and of incorporating autocovariate terms (to address spatial autocorrelation) are also investigated. In addition, a novel graphical tool is proposed for visualising extrapolation uncertainty in predictive models.

Chapter 4 – Distribution Models and Conservation Priority

Equipped with more extensive botanical inventory data and the methodological findings of Chapter 3, this study investigates the potential of distribution models to inform management of which areas should be prioritised for conservation. The environmental responses of 452 vascular plant taxa are modelled with respect to climate, topography and soil properties, and then extrapolated to the wider EAM region. This chapter provides spatially complete information for conservation planning in well-researched areas, and predictive estimates conservation importance elsewhere. Large disparities between levels of richness confirmed *vs*. predicted at the mountain bloc resolution provide direction for future fieldwork. Differences between patterns predicted for the different growth forms of plant are discussed.

Chapter 5 – Climate Change

Output from Global Circulation Models (GCMs) is impractically coarse for regional-scale analysis of mountain species, with horizontal resolutions typically in the range 200-600 km. Here, a regional climate model is used to dynamically downscale GCM forecasts to 55 km. Anomalies for the 21st century are then spatially interpolated and added to higher-resolution baseline grids, yielding detailed scenarios of change across the EAMs. Potential impacts on the region's endemic flora are investigated using the definitions, tools and findings described in Chapters 2-4. In line with the theory that climatic stability facilitates the accumulation of rare species, this chapter finds significant correlations between endemic plant richness and the persistence of 21st century climates within dispersal-limiting mountain blocs. Species' responses across the altitudinal gradient are mediated as much by changes in seasonality and moisture, as by the monotonic relation between altitude and mean annual temperature, questioning the common application of temperature lapse-rates alone to forecast climate change impacts in mountain regions.

References

Ahrends, A. (2010) *Do we know enough? The impacts of data biases, insufficient sampling and degradation on biodiversity estimates in Tanzanian forests – implications for conservation planning*, PhD Thesis, University of York, UK.

Ahrends, A., Burgess, N. D., Gereau, R. E., Marchant, R., Bulling, M. T., Lovett, J. C., Platts, P. J., Kindemba, V. W., Owen, N., Fanning, E. and Rahbek, C. (2011a) Funding begets biodiversity. *Diversity and Distributions*, **17**, 191-200.

Ahrends, A., Burgess, N. D., Milledge, S. A. H., Bulling, M. T., Fisher, B., Smart, J. C. R., Clarke, G. P., Mhoro, B. E. and Lewis, S. L. (2010) Predictable waves of sequential forest degradation and biodiversity loss spreading from an African city. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 14556-14561.

Ahrends, A., Rahbek, C., Bulling, M., Burgess, N. D., Platts, P. J., Lovett, J. C., Kindemba, V. W., Owen, N., Sallu, A. N., Marshall, A. R., Mhoro, B. E., Fanning, E. and Marchant, R. (2011b) Conservation and the botanist effect. *Biological Conservation*, **144**, 131-140.

Anderson, S. (1994) Area and endemism. *Quarterly Review of Biology*, **69**, 451-471.

Araújo, M. B. and Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688.

Araújo, M. B. and Peterson, T. (2012) Uses and misuses of bioclimatic envelope modelling. *Ecology*, in press.

Asner, G. P., Loarie, S. R. and Heyder, U. (2010) Combined effects of climate and land-use change on the future of humid tropical forests. *Conservation Letters*, **3**, 395-403.

Asner, G. P., Rudel, T. K., Aide, T. M., Defries, R. and Emerson, R. (2009) A Contemporary Assessment of Change in Humid Tropical Forests. *Conservation Biology*, **23**, 1386-1395.

Austin, M. P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.

Austin, M. P. and Smith, T. M. (1989) A new model for the continuum concept. *Vegetatio*, **83**, 35-47.

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

Bensted-Smith, R. and Msangi, T. H. (1989) *Report on the conservation of ecosystems and genetic resources*, Tanzania Forestry Action Plan, The United Republic of Tanzania.

Bjorndalen, J. E. (1992) Tanzania's vanishing rain-forests: assessment of nature conservation values, biodiversity and importance for water catchment. *Agriculture, Ecosystems and Environment*, **40**, 313-334.

Blomley, T., Pfliegner, K., Isango, J., Zahabu, E., Ahrends, A. and 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.

Bonan, G. B. (2008) Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science*, **320**, 1444-1449.

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

Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Rylands, A. B., Konstant, W. R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G. and Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909-923.

Burgess, N., Kilahama, F., Nderumaki, M., Kahemela, A. and Sawe, C. (2005) Mining in the Eastern Arc Mountains: the situation in early October 2004. *Oryx*, **39**, 13-14.

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. and Stuart, S. N. (2007a) The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, **134**, 209-231.

Burgess, N. D., Loucks, C., Stolton, S. and Dudley, N. (2007b) The potential of forest reserves for augmenting the protected area network in Africa. *Oryx*, **41**, 151-159.

Busby, J. R. (1991) BIOCLIM a bioclimatic analysis and prediction system. *Plant Protection Quarterly*, **6**, 8-9.

CEPF (2003) *Ecosystem profile for the Eastern Arc Mountains & Coastal Forests of Tanzania and Kenya*. Final version 31 July 2003, updated March 2005, Critical Ecosystem Partnership Fund (http://www.cepf.net/).

Chapman, D. S. (2010) Weak climatic associations among British plant distributions. *Global Ecology and Biogeography*, **19**, 831-841.

DeFries, R. S., Rudel, T., Uriarte, M. and Hansen, M. (2010) Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*, **3**, 178-181.

Dobrowski, S. Z., Thorne, J. H., Greenberg, J. A., Safford, H. D., Mynsberge, A. R., Crimmins, S. M. and Swanson, A. K. (2011) Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. *Ecological Monographs*, **81**, 241-257.

Doherty, R. M., Sitch, S., Smith, B., Lewis, S. L. and Thornton, P. K. (2010) Implications of future climate and atmospheric CO2 content for regional biogeochemistry, biogeography and ecosystem services across East Africa. *Global Change Biology*, **16**, 617-640.

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. and Zimmermann, N. E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

FAO (2006) *Global Forest Resources Assessment 2005*, Food and Agriculture Organization of the United Nations, Rome, Italy.

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

Finch, J. and Marchant, R. (2011) A palaeoecological investigation into the role of fire and human activity in the development of montane grasslands in East Africa. *Vegetation History and Archaeobotany*, **20**, 109-124.

Fisher, B. (2010) African exception to drivers of deforestation. *Nature Geoscience*, **3**, 375-376.

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

Franklin, J. (2010) *Mapping Species Distributions*, Cambridge University Press, Cambridge, UK.

Gentry, A. H. (1992) Tropical forest biodiversity - distributional patterns and their conservation significance. *Oikos*, **63**, 19-28.

Gereau, R. E., Hall, J., Hemp, A., Ahrends, A. and Platts, P. J. (in prep.) Distribution and endemism of vascular plants in the Eastern Arc Mountains.

Gerrard, A. J. (1990) *Mountain Environments: an examination of the Physical Geography of Mountains*, Belhaven Press, London, UK.

Godsoe, W. and Harmon, L. J. (2012) How do species interactions affect species distribution models? *Ecography*, in press.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C. and 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. and Neel, M. C. (2007) Biased data reduce efficiency and effectiveness of conservation reserve networks. *Ecology Letters*, **10**, 364-374.

Green, J. M. H., Burgess, N. D., Green, R. E., Madoffe, S. S., Munishi, P. K. T., Turner, R. K. and Balmford, A. (2012) Estimating management costs of protected areas: a novel approach from the Eastern Arc Mountains, Tanzania. *Biological Conservation*, in press.

Griffiths, C. J. (1993) The geological evolution of East Africa. In *Biogeography and ecology of the rain forests of Eastern Africa* (Eds, Lovett, J. C. and Wasser, S.) Cambridge University Press, Cambridge, UK, pp. 9-21.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A. and Zimmermann, N. E. (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**, 501-511.

Hall, J., Burgess, N. D., Lovett, J., Mbilinyi, B. and Gereau, R. E. (2009) Conservation implications of deforestation across an elevational gradient in the Eastern Arc Mountains, Tanzania. *Biological Conservation*, **142**, 2510-2521.

Hamilton, A. C. (1982) *Environmental History of East Africa. A study of the Quarternary*, Academic Press, London, UK.

Harold, A. S. and Mooi, R. D. (1994) Areas of endemism - definition and recognition criteria. *Systematic Biology*, **43**, 261-266.

Hecht, S. B. (2012) From eco-catastrophe to zero deforestation? Interdisciplinarities, politics, environmentalisms and reduced clearing in Amazonia. *Environmental Conservation*, **39**, 4-19.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Holt, R. D. (1996) Demographic constraints in evolution: Towards unifying the evolutionary theories of senescence and niche conservatism. *Evolutionary Ecology*, **10**, 1-11.

ICBP (1992) *Putting Biodiversity on the Map: Priority Areas for Global Conservation*, International Council for Bird Preservation, Cambridge, UK.

Joppa, L. N., Roberts, D. L., Myers, N. and Pimm, S. L. (2011) Biodiversity hotspots house most undiscovered plant species. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 13171-13176.

Kapos, V., Rhind, J., Edwards, M., Price, M. F. and Ravilious, C. (2000) Developing a map of the world's mountain forests. In *Forests in sustainable mountain development: a state of knowledge report for 2000* (Eds, Price, M. F. and Butt, N.) CAB International, Wallingford, UK, pp 4-9.

Kearney, M. R., Wintle, B. A. and Porter, W. P. (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, **3**, 203-213.

Kreutzmann, H. (2001) Development indicators for mountain regions. *Mountain Research and Development*, **21**, 132-139.

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

Lawton, J. H., Bignell, D. E., Bolton, B., Bloemers, G. F., Eggleton, P., Hammond, P. M., Hodda, M., Holt, R. D., Larsen, T. B., Mawdsley, N. A., Stork, N. E., Srivastava, D. S. and Watt, A. D. (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, **391**, 72-76.

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. and Woll, H. (2009) Increasing carbon storage in intact African tropical forests. *Nature*, **457**, 1003-U1003.

Loiselle, B. A., Howell, C. A., Graham, C. H., Goerck, J. M., Brooks, T., Smith, K. G. and Williams, P. H. (2003) Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, **17**, 1591-1600.

Lovett, J. C. (1985) Moist forests of eastern Tanzania. *Swara*, **8**, 8-9.

Lovett, J. C. (1990) Classification and status of the moist forests of Tanzania. *Notifications from the Institute for General Botany Hamburg*, **23a**, 287–300.

Lovett, J. C. and Wasser, S. K. (1993) *Biogeography and Ecology of the Rainforests of Eastern Africa*, Cambridge University Press, Cambridge, UK.

Lozier, J. D., Aniello, P. and Hickerson, M. J. (2009) Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *Journal of Biogeography*, **36**, 1623-1627.

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. and van Jaarsveld, A. S. (2000) It's time to work together and stop duplicating conservation efforts. *Nature*, **405**, 393-393.

Marchant, R., Finch, J., Kinyanjui, R., Muiruri, V., Mumbi, C., Platts, P. J. and Rucina, S. (2010) Palaeoenvironmental perspectives for sustainable development in East Africa. *Climate of the Past Discussions*, **6**, 963-1007.

Marchant, R., Mumbi, C., Behera, S. and Yamagata, T. (2007) The Indian Ocean dipole - the unsung driver of climatic variability in East Africa. *African Journal of Ecology*, **45**, 4-16.

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

Menon, S., Choudhury, B. I., Khan, M. L. and Peterson, A. T. (2010) Ecological niche modeling and local knowledge predict new populations of Gymnocladus assamicus a critically endangered tree species. *Endangered Species Research*, **11**, 175-181.

Meynard, C. N. and Quinn, J. F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**, 1455-1469.

Milledge, S. A. H., Gelvas, I. K. and 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 / Ministry of Natural Resources of Tourism, Dar es Salaam, The United Republic of Tanzania.

Millennium-Ecosystem-Assessment (2005) *Ecosystems and Human Well-being: Synthesis*, Island Press, Washington DC, USA.

Mittermeier, R. A., Myers, N., Thomsen, J. B., da Fonseca, G. A. B. and 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., Brooks, T., Mittermeier, C. G., Lamoreux, J. and da Fonseca, G. A. B. (2004) *Hotspots Revisited*, CEMEX, Mexico.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. and Worm, B. (2011) How many species are there on Earth and in the ocean? *PLoS Biology*, **9**, e1001127.

Mulligan, M. (2006) Global gridded 1km TRMM rainfall climatology and derivatives. Version 1 (http://www.ambiotek.com/1kmrainfall).

Mwakalila, S., Burgess, N. D., Ricketts, T., Olwero, N., Swetnam, R., Mbilinyi, B., Marchant, R., Mtalo, F., White, S., Munishi, P., Marshall, A., Malimbwi, R., Smith, C., Jambiya, G., Madoffe, S., Fisher, B., Kajembe, G., Morse-Jones, S., Kulindwa, K., Green, J. and Balmford, A. (2009) Linking Science with Stakeholders to sustain Natural Capital. *The Arc Journal*, **23**, 22-27.

Myers, N. (1990) The biodiversity challenge: expanded hot-spots analysis. *Environmentalist*, **10**, 243-256.

Myers, N. (2003) Biodiversity hotspots revisited. *Bioscience*, **53**, 916-917.

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

Naidoo, R. and Ricketts, T. H. (2006) Mapping the economic costs and benefits of conservation. *Plos Biology*, **4**, 2153-2164.

Newmark, W. (2005) Diel variation in the difference in air temperature between the forest edge and interior in the Usambara Mountains, Tanzania. *African Journal of Ecology*, **43**, 177-180.

Newmark, W. D. (1998) Forest area, fragmentation, and loss in the Eastern Arc Mountains: implications for the conservation of biological diversity. *Journal of the East African Natural History Society*, **87**, 29–36.

Newmark, W. D. (2001) Tanzanian forest edge microclimatic gradients: Dynamic patterns. *Biotropica*, **33**, 2-11.

Newmark, W. D. (2002) *Conserving Biodiversity in East African Forests: a Study of the Eastern Arc Mountains*, Springer, Berlin, Germany.

Olson, D. M. and Dinerstein, E. (1998) The global 200: A representation approach to conserving the Earth's most biologically valuable ecoregions. *Conservation Biology*, **12**, 502-515.

Pearman, P. B., Guisan, A., Broennimann, O. and Randin, C. F. (2008) Niche dynamics in space and time. *Trends in Ecology & Evolution*, **23**, 149-158.

Pearson, R. G., Thuiller, W., Araujo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P. and Lees, D. C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704-1711.

Peterson, A. T. (2011) Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography*, **38**, 817-827.

Peterson, A. T., Soberon, J. and Sanchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265-1267.

Pfeifer, M., Disney, M., Quaife, T. and Marchant, R. (2012) Terrestrial ecosystems from space: a review of earth observation products for macroecology applications. *Global Ecology and Biogeography*, in press.

Pfeifer, M., Burgess, N. D., Swetnam, R. D., Platts, P. J., Willcock, S. and Marchant, R. (2012) Protected areas: mixed success in conserving East Africa's evergreen forests. *PloS one*, **7**, e39337.

Pimm, S. L., Ayres, M., Balmford, A., Branch, G., Brandon, K., Brooks, T., Bustamante, R., Costanza, R., Cowling, R., Curran, L. M., Dobson, A., Farber, S., da Fonseca, G. A. B., Gascon, C., Kitching, R., McNeely, J., Lovejoy, T., Mittermeier, R. A., Myers, N., Patz, J. A., Raffle, B., Rapport, D., Raven, P., Roberts, C., Rodriguez, J. P., Rylands, A. B., Tucker, C., Safina, C., Samper, C., Stiassny, M. L. J., Supriatna, J., Hall, D. H. and Wilcove, D. (2001) Environment - Can we defy nature's end? *Science*, **293**, 2207-2208.

Pocs, T. (1976) Vegetation mapping in the Uluguru mountains (Tanzania, East Africa). *Boissiera*, **24**, 477-498.

Polhill, R. M. (1968) Tanzania. In *Conservation of vegetation in Africa south of the Sahara* (Eds, Hedberg, I. and Hedberg, O.) Almqvist and Wiksells Boktryckeri AB, Uppsala, Sweden.

Poynton, J. C., Loader, S. P., Sherratt, E. and Clarke, B. T. (2007) Amphibian diversity in east african biodiversity hotspots: Altitudinal and latitudinal patterns. *Biodiversity and Conservation*, **16**, 1103-1118.

Pressey, R. L., Hager, T. C., Ryan, K. M., Schwarz, J., Wall, S., Ferrier, S. and 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.

Pócs, T. (1976) Bioclimatic Studies in the Uluguru Mountains (Tanzania, East Africa) I. *Acta Botanica Academiae Scientarium, Hungariae*, **20**, 115-135.

Raxworthy, C. J., Martinez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A. and Peterson, A. T. (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837-841.

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

Schlüter, T. (1997) Geology of East Africa. In *Contributions to the regional geography of the Earth* (translated from German), Borntraeger, Berlin, Germany.

Scholes, R. J., Küper, W., Biggs, R., Mwangi, E., Raharimampionona, R., Lowry, P., Sene, E., Ashton, P., Blake, S. and Justice, C. O. (2006) Biodiversity. In *Africa Environment Outlook 2 – Our Environment, Our Wealth* (Eds, Mohamed-Katerere, J. C. and Sabet, M.) Division of Early Warning and Assessment (DEWA) United Nations Environment Programme, Nairobi, Kenya, pp. 226-261.

Stattersfield, A. J., Crosby, M. J., Long, A. J. and Wege., D. C. (1998) *Endemic bird areas of the world: priorities for biodiversity conservation*, BirdLife International, Cambridge, UK.

Swetnam, R. D., Fisher, B., Mbilinyi, B. P., Munishi, P. K. T., Willcock, S., Ricketts, T., Mwakalila, S., Balmford, A., Burgess, N. D., Marshall, A. R. and Lewis, S. L. (2011) Mapping socio-economic scenarios of land cover change: A GIS method to enable ecosystem service modelling. *Journal of Environmental Management*, **92**, 563-574

Tansey, K., Gregoire, J. M., Binaghi, E., Boschetti, L., Brivio, P. A., Ershov, D., Flasse, S., Fraser, R., Graetz, D., Maggi, M., Peduzzi, P., Pereira, J., Silva, J., Sousa, A. and Stroppiana, D. (2004) A global inventory of burned areas at 1km resolution for the year 2000 derived from SPOT VEGETATION data. *Climatic Change*, **67**, 345-377.

Thomas, C. D., Bodsworth, E. J., Wilson, R. J., Simmons, A. D., Davies, Z. G., Musche, M. and Conradt, L. (2001) Ecological and evolutionary processes at expanding range margins. *Nature*, **411**, 577-581.

Thuiller, W., Brotons, L., Araújo, M. B. and Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165-172.

Walker, P. A. and Cocks, K. D. (1991) HABITAT - A procedure for modeling a disjoint environmental envelope for a plant or animal species. *Global Ecology and Biogeography Letters*, **1**, 108-118.

White, F. (1983) *The vegetation of Africa. A descriptive memoir to accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa*, UNESCO, Paris, France.

Whittaker, R. H. (1956) Vegetation of the Great Smoky Mountains. *Ecological Monographs*, **26**, 1-69.

Whittaker, R. H. (1967) Gradient Analysis of Vegetation. *Biological Reviews of the Cambridge Philosophical Society*, **42**, 207-&.

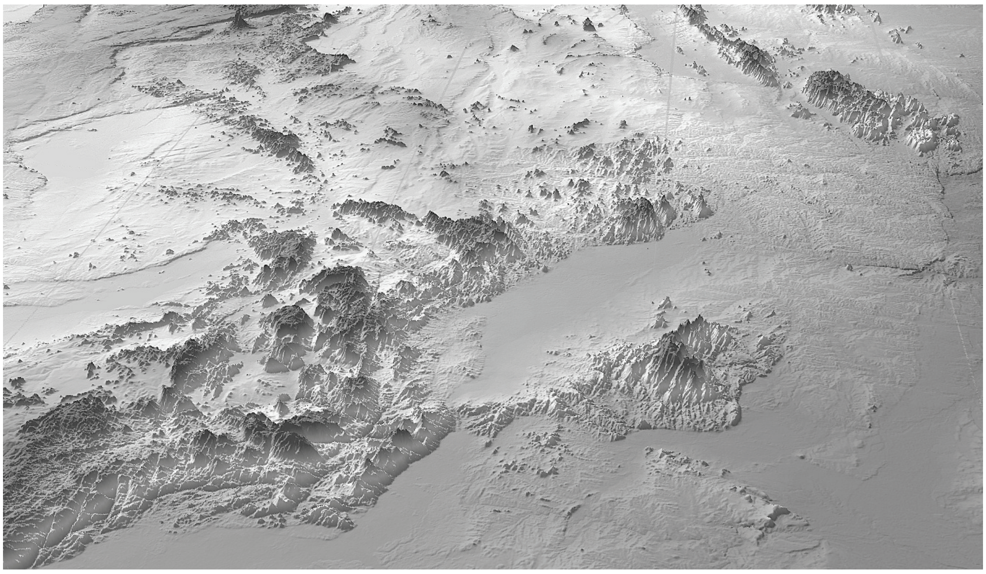
Williams, J. W., Jackson, S. T. and Kutzbacht, J. E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5738-5742.

Wilson, K. A., McBride, M. F., Bode, M. and Possingham, H. P. (2006) Prioritizing global conservation efforts. *Nature*, **440**, 337-340.

Yasuda, Y., Kitagawa, H. and Nakagawa, T. (2000) The earliest record of major anthropogenic deforestation in the Ghab Valley, northwest Syria: a palynological study. *Quaternary International*, **73-4**, 127-136.

Yee, T. W. and Mitchell, N. D. (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science*, **2**, 587-602.

Chapter 2 – Mountain Limits



Published in *Environmental Conservation* (Sep. 2011), Vol. 38, Issue 3, Pages 312-324

Delimiting tropical mountain ecoregions for conservation

Philip J. Platts [a,g,*], Neil D. Burgess [b], R. E. Gereau [c], Jon C. Lovett [d], Andrew R. Marshall [a,e], Colin J. McClean [a], Petri K. E. Pellikka [f], Ruth D. Swetnam [g], Rob Marchant [a]

[a] The York Institute for Tropical Ecosystem Dynamics (KITE), Environment Department, University of York, Heslington, York YO10 5DD, UK. [b] Centre for Macroecology, Evolution and Climate, Department of Biology, Copenhagen University, Universitetsparken 15, DK-2100, Copenhagen, Denmark. WWF-US Conservation Science Program, 1259 24th Street NW, Washington DC, USA. [c] Missouri Botanical Garden, P.O. Box 299, Saint Louis, MO 63166-0299, USA. [d] Twente Centre for Studies in Technology and Sustainable Development (CSTM), University of Twente, Postbus 217, 7500 AE, Enschede, Netherlands. [e] Flamingo Land Ltd, Kirby Misperton, Malton, North Yorkshire YO17 6UX, UK. [f] Department of Geography, University of Helsinki, P.O. Box 64, 00014 Helsinki, Finland. [g] Conservation Science Group, Zoology Department, University of Cambridge, Cambridge CB2 3EJ, UK.

** Corresponding author. Email: philip.platts@gmail.com*

Abstract

Ecological regions aggregate habitats with similar biophysical characteristics within well-defined boundaries, providing spatially consistent platforms for monitoring, managing and forecasting the health of interrelated ecosystems. A major obstacle to the implementation of this approach is imprecise and inconsistent boundary placement. For globally important mountain regions such as the Eastern Arc (Tanzania and Kenya), where qualitative definitions of biophysical affinity are well established, rule-based methods for landform classification provide a straightforward solution to ambiguities in region extent. The method we present encompasses the majority of both contemporary and estimated preclearance forest cover within strict topographical limits. Many of the species here tentatively considered 'near-endemic' could be reclassified as strictly endemic according to the derived boundaries. LandScan and census data show population density inside the ecoregion to be higher than in rural lowlands, and lowland settlement to be most probable within 30 km. This definition should help to align landscape scale conservation strategies in the Eastern Arc Mountains and promote new research in areas of undocumented biological importance. Similar methods could work well in other regions where mountain extent is poorly resolved.

Keywords: boundary placement; conservation priority; Eastern Arc Mountains; ecological regions; ecosystem services; mapping; plant endemism; population pressure; topography.

Introduction

Ecological regions (ecoregions) are widely employed as units for conservation priority setting (Olson and Dinerstein, 2002; Burgess *et al.*, 2006). By aggregating ecosystems with similar biophysical characteristics, interdependences and spheres of influence within ecoregions, assessment and management can target the system as a whole and avoid any disconnect between science and policy. This holistic approach is more faithful to environment-ecosystem interconnections, including the role of human populations, but by definition imposes strict and static divisions on complex and ever changing landscapes. Careful delineation is therefore crucial. Landform, geology, climate, vegetation and evolutionary history, as well as cultural and political considerations, are all important (McMahon *et al.*, 2004). In practice, different criteria favour different objectives and no single solution is optimal for all applications in all regions (Olson *et al.*, 2001).

Some biogeographical units such as mountains, wetlands and islands appear by their nature to be clearly defined in space. In the case of mountains, however, there is no universally accepted method for marking the transition to lowlands, and thus no consistent way of defining precisely the geographical limits of a mountainous region (Gerrard, 1990). The essence of the problem, as in any ecoregion, is that environmental gradients are continuous (from sea level to mountain top) and so any spatial dichotomy is necessarily subjective. Freely available digital elevation data, together with improvements in desktop mapping software, have brought advances in the development of a systematic process by which to define and study mountains. Practitioners can now experiment with different ways of bounding their region of interest, so that qualitative definitions of biological affinity may where appropriate progress to spatially explicit rule-based algorithms, increasing the prominence of such regions and their specific management challenges on the political stage (Browne *et al.*, 2004; McMahon *et al.*, 2004).

As assessment mechanisms for biodiversity and ecosystem services move towards implementation within international frameworks (Larigauderie and Mooney, 2010), there is a requirement for rule-based definitions that can resolve ambiguities in the placement of ecoregion boundaries. In the present study, using the Eastern Arc Mountains (EAMs) of

Tanzania and Kenya as an example, we discuss some of the challenges and uncertainties in defining tropical mountain ecoregions. This chain of 13 block-faulted massifs harbours one of the world's most important concentrations of biodiversity across a series of fragile sites (Brooks *et al.*, 2002; Mittermeier *et al.*, 2004). The "Eastern Arc Mountains [sic] Forests" have recently been proposed for UNESCO World Heritage status (http://whc.unesco.org/), and will soon be subject to increased international attention as the United Nations REDD Programme is piloted in Tanzania (Burgess *et al.*, 2010). We propose for the first time a rigorous set of topographical limits for these mountain habitats, using as a starting point the global mountain topology developed by the UNEP World Conservation Monitoring Centre (UNEP-WCMC; Kapos *et al.*, 2000). Increasing the spatial resolution from 1 km to 90 m, we rescale terrain parameters using empirical and remotely sensed data on the distribution of endemic plant species and mountain forests respectively, before bounding the ecoregion according to topographic prominence and established biogeographical distributions.

A qualitative definition for the Eastern Arc

The EAMs were first described as distinct from surrounding Afromontane habitats in the 1980s due to the exceptional proportion of rare species (Lovett, 1985). The distinction is qualitatively explained according to key environmental characteristics thought to underlie the high levels of endemism (Fig. 2.1a): first, the great age of the Precambrian crystalline and metamorphic substrate compared to geologically more recent volcanoes such as Kilimanjaro and Meru (> 30 million *vs*. < 2 million years old; Schlüter, 1997); second, the relative consistency of rain-bearing winds from the Indian Ocean compared to more variable climatic regimes within the great lake catchments of Nyasa/Malawi, Tanganyika and Victoria. Accordingly, the chain extends roughly 750 km from the Taita Hills in south-east Kenya to the Udzungwa Mountains in south-central Tanzania; volcanoes to the north are excluded (different substrate), as are highlands south of the Makambako Gap (geologically younger and rainfall patterns not directly influenced by the Indian Ocean).

In Fig. 2.1b we identify the 13 most commonly cited ranges (blocs). Sometimes omitted are: the Taita Hills (the only bloc north of the Kenya-Tanzania border, e.g. Burgess *et al.*, 2009); the Nguu (sometimes combined with Nguru, e.g. Newmark, 2002); and Malundwe Mountain due to its small extent and position within a chain of low elevation hills (e.g. Stanley and Olson, 2005). An occasional addition is Image, but this mountain is more usually included in Udzungwa (Mbilinyi *et al.*, 2006).

Mountain limits

The lack of a systematic basis for boundary placement in the EAMs has meant that the precise spatial extents depicted have varied from study to study, hindering information exchange and multilateral action in response to new data or changes on the ground. Mountain limits have been defined variously by hand, using elevational contours, forest or reserve boundaries, or a combination of these factors (e.g. Doggart *et al.*, 2006). Many depictions can be traced back to Lovett (1992): drawn from contours on a Shell Road Map, these boundaries were intended as diagrammatic and heuristic; however a number of derivatives are now in common usage, conspicuous in their repetition of shortfalls in the original. A 1000 m cut-off for the Taita Hills, for instance, depicts bloc-extent far to the west of the mountains proper but excludes important sites for conservation in the east.

Elevation offers a simple, intuitive, but often insufficient method for delimiting mountain regions (Messerli and Ives, 1997). With respect to plant endemism in the Eastern Arc, one convention is to impose a uniform lower limit of 500 m (e.g., government reports, World Heritage application), intended to distinguish montane habitats from the nearby coastal forests (Lovett *et al.*, 2000). We find this definition to be too strict in the east, excluding patches of forest from the lower slopes of six mountain blocs. Conversely, because the East African interior is characterised by a highland plateau, almost all non-mountainous land on the western side exceeds 500 m. Similarly for herpetofauna, species turnover is highest at around 400 m in the north-east but closer to 800 m in the south (Poynton *et al.*, 2007).

A global mountain typology

Effective mountain delineation across all aspects requires consideration of not just elevation, but also steepness of slope and terrain roughness (Gerrard, 1990). Implementing these criteria on a global scale, UNEP-WCMC developed a map of the world's mountains and mountain forests by classifying 1 km land parcels according to elevation, slope and local relief (Kapos *et al.*, 2000; Blyth *et al.*, 2002). The classification consists of six elevational bands, with terrain constraints strictest at low elevations (Fig. 2.1). Whilst this typology effectively represents the EAMs on a world map, it is too broadly defined to identify all features important at the site level or to distinguish between adjacent ranges (e.g., South Pare *vs*. West Usambara; Fig. 2.1b). Kapos *et al.* (2000) suggest that for quantitative applications at sub-national scales their map could be refined using high-resolution vegetation data together with an appropriate measure of relative relief – criteria we now test for the EAMs.

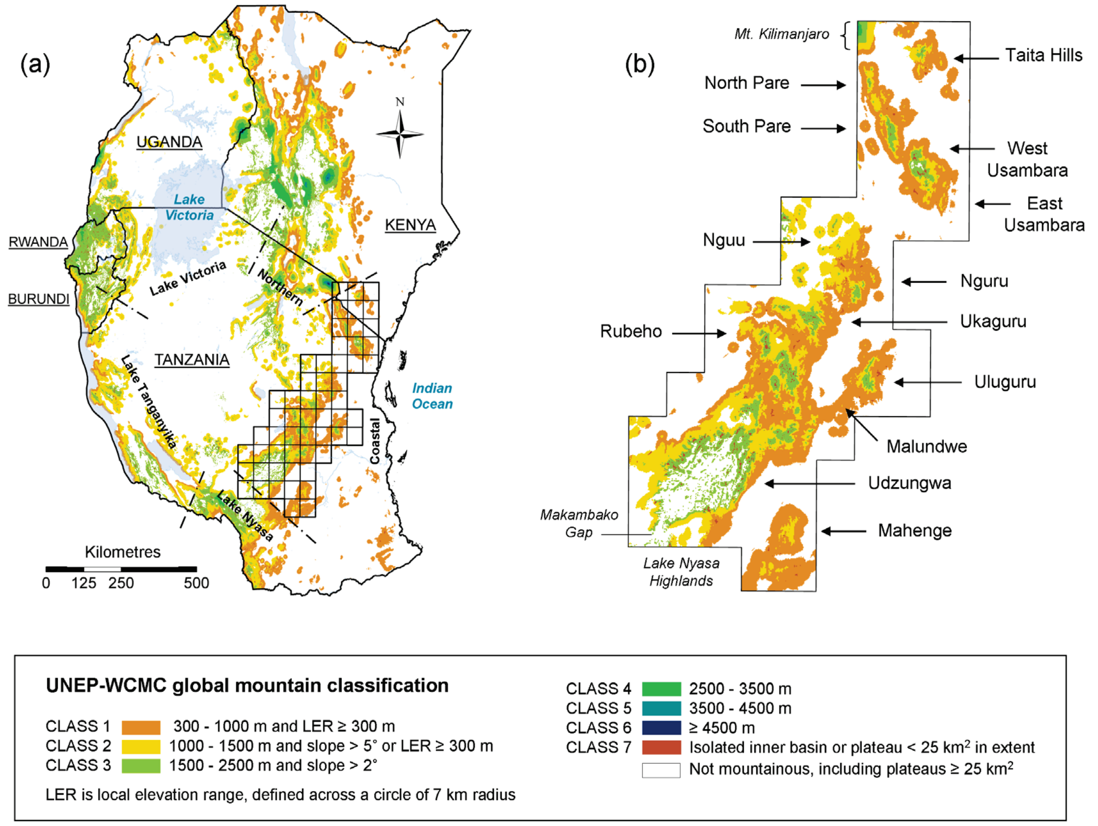


Figure 2.1. Global mountain typology defined by UNEP-WCMC at 1 km resolution. (a) East Africa, showing divisions (dashed lines) in Tanzanian forest on the basis of geology and climate (from Lovett, 1990): Coastal, Eastern Arc and Northern forests are all under the direct climatic influence of the Indian Ocean (rather than the Great Lakes), but only the Eastern Arc forests are on Precambrian crystalline bedrock. (b) Zoomed perspective of quarter degree grid squares that intersect the 13 Eastern Arc ranges (blocs).

Methods

Regional refinements to the global typology

Elevational zonation

We adhered mainly to the elevational classes as defined globally (Fig. 2.1), with the exception of the lowest mountain category, which was applied without the 300 m lower limit. As in other regions where mountains occur along a coastline (Nordregio, 2004), marked topography and associated forest vegetation extend almost to sea level. For higher elevations, the hypsographic curve for eastern Tanzania and south-east Kenya (cumulative height *vs.* relative area) exhibits similarities with the breakpoints defined by UNEP-WCMC: frequencies are highest for land at elevations of 1000-1500 m, with a further change in trend at elevations *c.* 2500 m. The unqualified inclusion of all land above 2500 m is perhaps inappropriate for regions with large highland plateaus (*cf.* Meybeck *et al.*, 2001), but is of little consequence here where only mountain summits exceed this elevation.

Terrain parameters

The spatial resolution of the entire typology was increased from 1 km to 90 m. CGIAR-SRTM elevation data (Jarvis *et al.*, 2008) were extracted for the 56 half-degree grid squares that intersect with the EAMs (Fig. 2.1) and resampled from a three arc-second geographic projection to a 90 m equal-area conic projection (Africa Albers). Slope constraints were increased from 5° to 10° (class 2) and from 2° to 5° (class 3) to correct for the greater topographic heterogeneity captured per unit area at this finer resolution.

At the global scale, the local elevation range parameter (hereafter, $\text{LER}_{\text{RADIUS}}$) is intended to capture “older mountains of regional significance” (Kapos *et al.*, 2000). The EAM blocs consist entirely of such old weathered slopes, but in some cases are relatively small and, particularly in the east, rise abruptly from surrounding plains. A parameterisation of the LER better suited to regional-scale analysis was therefore sought. With a view to maximising the inclusion of forest habitats and endemic plant records within a minimal mountain area, we tested LER radii in the range 250 m to 2 km and thresholds in the range 5% to 30% (of the radius). In exploratory analyses, larger radii led to excessive buffering around the base of the mountains or, with stricter thresholds, the omittance of outlying peaks; smaller radii returned patterns akin to the slope parameter, which is roughly equivalent to LER_{90m}.

The final step in the derivation of each variant of this regional typology was to apply a spatial filter, such that grid squares adopted the majority mountain class within a 500 m radius. This resolved fine-scale anomalies in raster grids, aiding the transition to vector format.

Matching to mountain vegetation

The decision of which parameterisation was best suited to the EAMs was guided by overlaying high-resolution (1 ha) vegetation data. Indigenous broadleaved forests in the Taita bloc were identified from SPOT multi-spectral satellite images (Clark and Pellikka, 2009). Tanzanian forests were based on MNRT (1997; updated with later imagery from 2000 onwards by the Remote Sensing and GIS Laboratory, Sokoine University of Agriculture). We extracted all forest classified as submontane, montane or upper-montane, and additionally considered any lowland forest contiguous with a submontane patch, thus removing the elevational bias of vegetation classification (Pócs, 1976). These estimates were then corrected according to Mbilinyi *et al.* (2006), Marshall *et al.* (2010 in Appendix I), Conservation International (2008), and local knowledge (see for example http://celp.org.uk/projects /tzforeco/). The total forest area identified was 4388 km^2.

Given that around 70% of forest cover has been lost since the turn of the 20th century and that forests on some eastern aspects might regenerate farther downslope if it were not for land use barriers and frequent burning, it was appropriate to also consider estimates of 'palaeoecological' extent (Hall *et al.*, 2009; Tanzania only). Hereafter, these estimates are termed 'preclearance' rather than palaeoecological. Forest would have extended beyond the base of the mountains during interglacial periods (e.g., Holocene and Eemian) and consisted of different vegetation assemblages in cooler, potentially drier environments of lower CO_2 concentration associated with the last glacial period (c. 10 000-114 000 yrs BP; Elenga *et al.*, 2000; Vincens *et al.*, 2007).

Lastly, we overlaid point distributions for 378 species of vascular plants strictly endemic to the EAMs (according to a 500 m lower limit). Over 2000 spatially referenced herbarium specimen records were available (http://www.tropicos.org/), representing 77% of all strict-endemics and spanning all plant growth forms. To investigate the efficacy of the 500 m lower limit for endemism, and to avoid bias resulting from such a convention, we further considered the distributions of 53 'near-endemic' plant species (455 records, some below 500 m) that are documented only within the half-degree cells in Fig. 2.1b and whose known distributions do not extend to coastal forests, Neogene Volcanoes or the Lake Nyasa Highlands. Based on locality information supplied with these records, we assumed a spatial accuracy of one arc-min (< 2 km).

Bounding the chosen regional typology

Amalgamating classified features

First, adjacent mountain classes were dissolved such that each spatially distinct feature was represented by a single polygon. With the exception of small, isolated fragments (< 1 km^2

and > 1 km from nearest neighbour), these features were buffered by 1 km and then simplified and smoothed using bend reduction and polynomial approximation tools in ArcGIS Desktop 9.3.1 (exponential kernel, tolerance = 1 km). This method ensured the preservation of all mountainous features on a 1 km grid (a popular format for regional mapping) and yielded topologically simpler boundaries that additionally enclosed unclassified habitats benefiting from close proximity to complex relief. Following Blyth *et al.* (2002), isolated inner basins and plateaus were filled if less than 25 km^2.

Mountain selection based on relative relief

The distinction between a mountain and a hill is largely semantic, with equivalent relative relief considered mountainous in one region whilst merely hilly in another (OED, 1989). It was however appropriate to make such a distinction in the EAMs so that minor relief not commonly perceived as mountainous could be systematically removed. Thompson (1964) suggests a topographic prominence of 2000 ft to be a good rule of thumb, which we rounded down to the metric equivalent of 600 m. A cluster of features, bounded as above, was considered ‘mountainous’ if attaining an altitude of at least 600 m relative to adjacent elevations; ‘hilly’ clusters (< 600 m prominence) were included in the ecoregion only if known to support natural forest vegetation and/or strictly endemic plant species.

Results

Sensitivity to local elevation range

The biggest changes in overall mountain area resulted from varying the LER threshold according to fixed percentages of its defining radius (Table 2.1). The lowest threshold (5%) was too permissive, merging adjacent range-boundaries and including minor undulations throughout the lowlands (Fig. 2.2). Higher thresholds were more effective in terms of forest and endemic plant records captured per unit area, but if too high (30%) then lower slopes and mountain foothills were curtailed. A 20% threshold provided the best compromise between forest inclusion and range distinction. With the threshold percentage held constant, changing the LER radius affected polygon complexity but had little impact on overall mountain area. Small radii tracked closely fine-scale changes in relief; large radii better represented broad-scale trends in the landscape. Both LER_{500m} and LER_{1km} worked well, but we preferred the former because slightly more forest and endemism records were included, and because the latter excluded lower slopes in south-west Mahenge (inspection of 3D elevation).

At this level ($\text{LER}_{500\text{m}}$, 20% threshold), 94% of the estimated forest area (both present-day and preclearance) was classified as mountainous. The remaining 6% was largely within 1 km of a classified feature and so would be captured by the ecoregion boundary (Table 2.1). One exception was open-canopy forest in the lowest parts of Matundu (south-east Udzungwa), which was omitted by even the most permissive LER threshold. Respectively, 99% and 93% of records for strictly and nearly endemic plant species originate from areas classified as mountainous by the chosen typology.

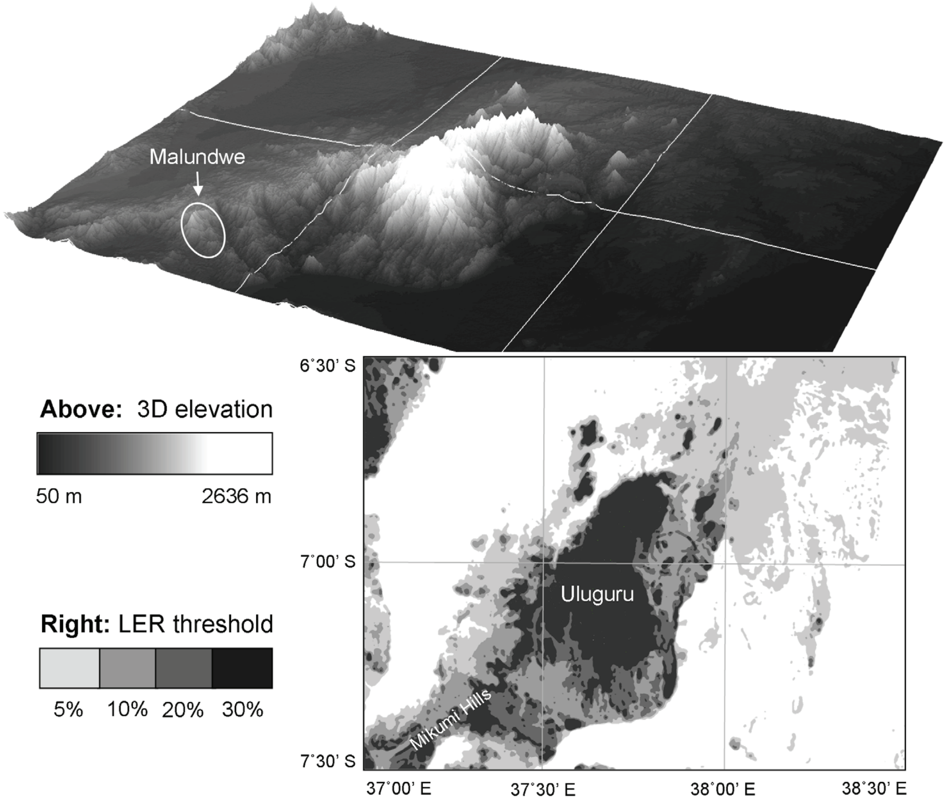


Figure 2.2. Sensitivity of the regional mountain typology to local elevation range (LER, radius = 500 m). Upper pane: elevation in the Uluguru bloc (×5 vertical exaggeration). Lower pane: variations in extent resulting from different LER thresholds (% of radius).

Table 2.1. Mountain area and forest sites included as a result of different calibrations of the local elevation range parameter (LER radius/threshold, mountain classes 1-2). Analyses were conducted within 56 half-degree grid squares (Fig. 2.1a) using 90 m elevation data and slope thresholds of 5° (class 2) and 10° (class 3); otherwise mountain classification follows UNEP-WCMC, but with no lower limit in class 1.

	Mountain area (10^3 km^2)	Forest † (% included)		Plant endemism records (% included)	
		Present-day	Preclearance	Strict-endemics	Near-endemics
Threshold = 0.05 × LER radius					
LER_{250m}	86	99.8	99.7	99.4	96.3
LER_{500m}	98	99.9	99.8	99.4	96.3
LER_{1km}	95	99.7	99.8	99.5	97.4
LER_{2km}	88	99.8	99.8	99.5	97.4
Threshold = 0.1 × LER radius					
LER_{250m}	55	98.7	98.3	99.1	95.4
LER_{500m}	62	99.2	99.1	99.1	95.4
LER_{1km}	62	98.7	99.1	99.1	95.2
LER_{2km}	61	97.8	98.8	99.3	95.4
Threshold = 0.2 × LER radius					
LER_{250m}	39	91.2	91.6	98.7	92.1
***LER_{500m}**	**43**	**93.9**	**94.4**	**98.6**	**93.2**
LER_{1km}	41	92.4	93.4	98.6	91.9
LER_{2km}	37	90.2	91.2	93.8	85.3
Threshold = 0.3 × LER radius					
LER_{250m}	30	83.9	81.7	98.4	91.7
LER_{500m}	33	87.3	86.8	98.4	91.2
LER_{1km}	31	86.0	84.5	98.1	86.6
LER_{2km}	28	82.7	79.7	92.9	82.5
EAM boundaries	48	99.0	99.3	99.1	95.2
[Including plateaus]	*[52]*	*[99.0]*	*[99.3]*	*[99.7]*	*[96.7]*

* Mountain typology from which EAM boundaries were derived (Appendix 2B)
† Preclearance forest follows Hall *et al.* (2009); Tanzanian blocs only

Boundary placement

Polygons corresponding to the south-east slopes of Mt. Kilimanjaro were spatially distinct and so straightforward to remove (Fig. 2.3). In the south, the EAMs are distinguished from adjacent highlands by climatic influence and vegetation type rather than landform. Consequently, the Udzungwa bloc had to be divided manually from moorlands south-east of the Makambako Gap: we followed the Mpanga River upstream from the Kilombero Valley towards the southern perimeter of Mufindi Scarp Forest Reserve, and then traced this west to dissect the polygon fully.

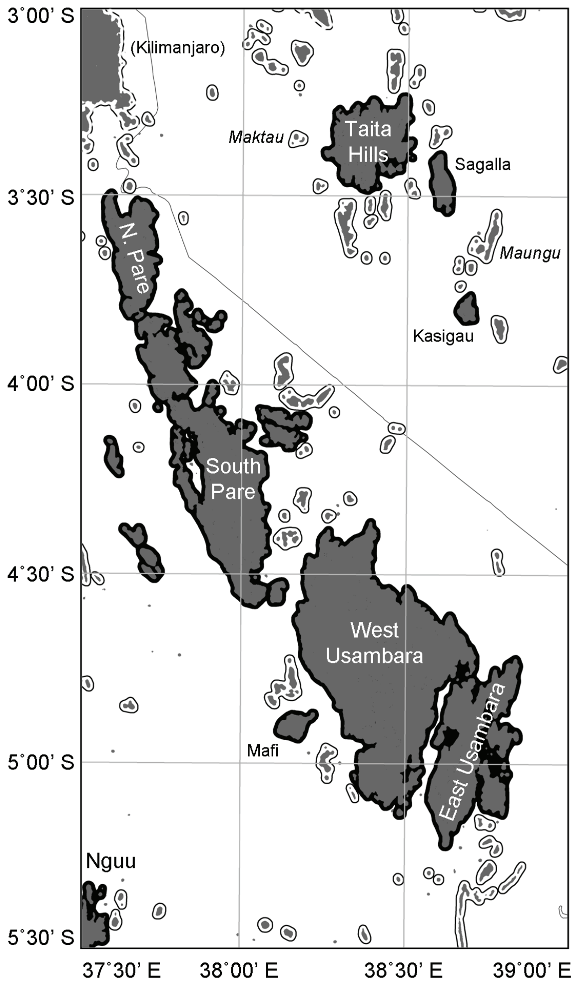


Figure 2.3. Boundary placement in the northern blocs. Features identified as ‘mountainous’ by the chosen typology were aggregated within simplified boundaries and distinguished as mountains or hills/escarpments depending on their prominence relative to adjacent elevations. Italicised hill names indicate inclusion on the basis of plant endemism.

Within the EAM chain, the majority of range boundaries were clearly distinguished (Fig. 2.4). Exceptions occurred for those blocs separated by steep-sided river valleys (Ukaguru *vs.* Rubeho *vs.* Udzungwa), in which cases the divisions were imposed *post-hoc* by overlaying the respective river paths (Mkondoa and Great Ruaha Rivers). The most problematic bloc to delineate from surrounding terrain was Malundwe, a 1259 m peak within the Mikumi Hills (Fig. 2.2). The explicit delineation of this mountain and its small area of forest (< 3 km^2) was important for consistency with previous studies, but parameter combinations strict enough to isolate it resulted in the oversimplification or omission of marked topography elsewhere (e.g., LER$_{2km}$ with 30% threshold). Since the Mikumi Hills in their entirety are rarely considered part of the EAM chain, we distinguished the small Malundwe peak and its forest patch using a 900 m contour, buffered and simplified as for the other blocs (Fig. 2.4). An alternative delineation that retains the Mikumi relief is provided (Appendix 2A).

The proposed boundaries enclose a total area of *c.* 48,000 km^2 (52,000 km^2 including plateaus). Elevations range from 121 m to 2636 m above mean sea level, with both extremes corresponding to the highly prominent Uluguru range (Table 2.2). The 600 m cut-off for relative relief proved appropriate, with only five ‘hilly’ features warranting inclusion on the basis of forest cover and/or plant endemism: two in Matundu (south-east Udzungwa), one west of the Nguu Range, and two in the Taita Hills. The latter support locally endemic plant species: *Encephalartos kisambo* on the Maungu Hills and *Monadenium guentheri* on Maktau Hill, east and west of the main Taita bloc, respectively (Fig. 2.3). Handeni Hill, 50 km equidistant from West Usambara and Nguu, might have been included on similar grounds, but these forests are more usually associated with coastal vegetation.

The boundaries capture over 99% of all present-day and preclearance forest (Table 2.2), the lowest parts of Matundu being the only exception (beyond the south-central perimeter of Kilombero Nature Reserve). This area could be included *post-hoc* as suggested in the supplementary material (Appendix 2A). Non-woody species collected from mid-elevation plateaus in south-west Udzungwa (e.g., *Oldenlandia oxycoccoides*) might be considered strict-endemics by elevational criteria, but are not included by either the global or regional mountain typologies (slopes shallower than 2º and 5º, 1 km and 90 m resolutions respectively). At both scales, steep plateau margins are considered mountainous; thus, depending on the application, Udzungwa plateaus could be included *post-hoc* as suggested in Fig. 2.4. Alternatively, dry western margins could be removed, restricting Udzungwa to just those forested slopes benefiting from high orographic rainfall.

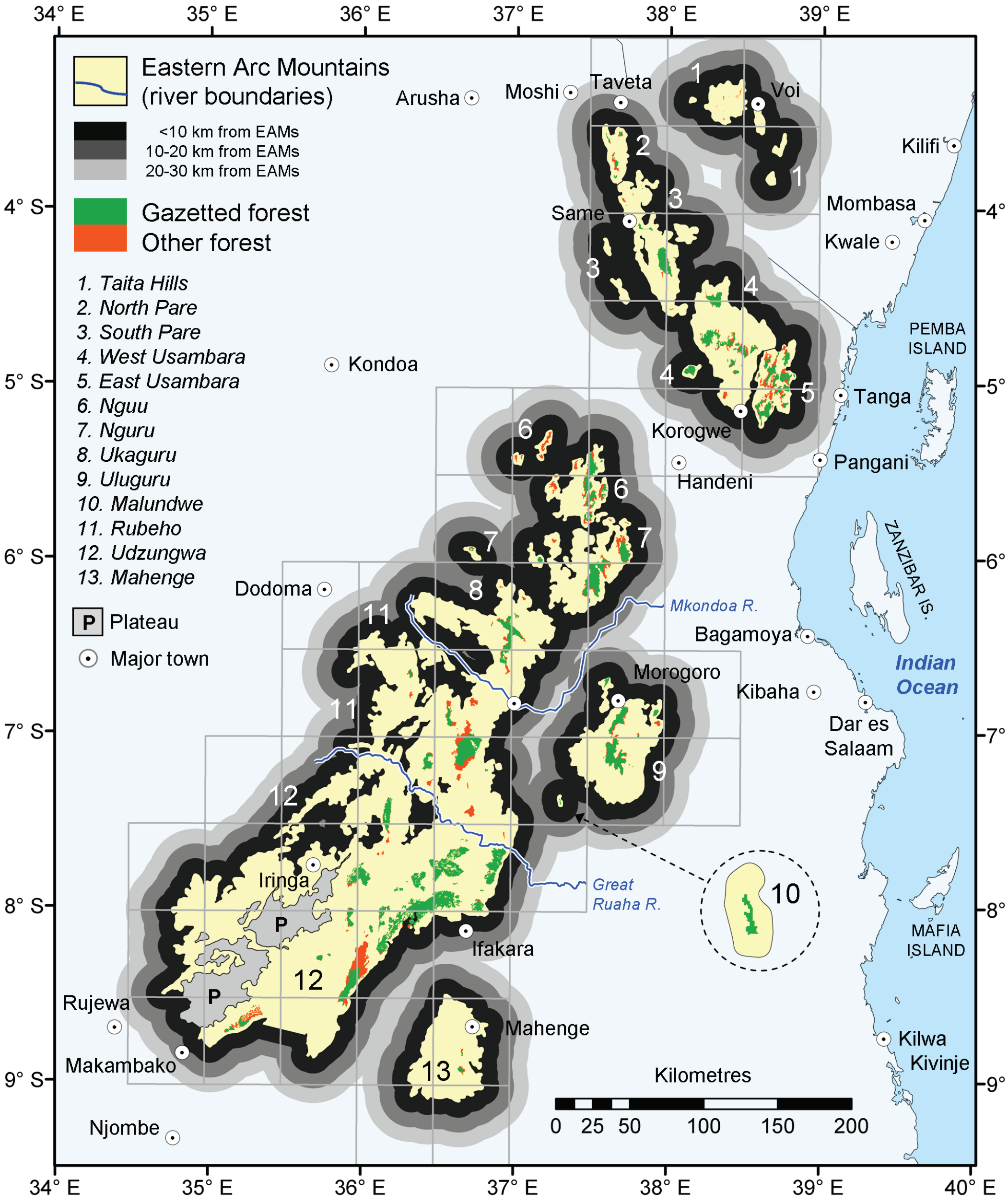


Figure 2.4. Ecoregion boundary, overlaid with forest distributions and protection status. Mid-elevation plateaus in Udzungwa are an option for inclusion (grasslands/heathlands but no natural forest). The majority of 'near-endemic' plant species have been collected within 10 or 20 km of the ecoregion boundary. Density of rural persons is highest within 30 km. Projection is Africa Albers.

Of the 53 near-endemic plant species, at least 30 have been recorded exclusively within the topographical limits defined here (up to 40 if we allow for spatial error in collection localities). Unless additional specimen data suggest otherwise, these species could be revised to strictly endemic. Most other near-endemics have been collected within 10 km of the EAM boundary, and all within 20 km. Magombera forest, for example, occupies a lowland position 6 km east of the Udzungwa bloc. Vegetation is predominantly of the kind common in coastal forests, but the presence of some characteristic EAM flora (e.g., *Dialium holtzii* and *Isoberlinia scheffleri*) and fauna (e.g., Udzungwa Red Colobus monkey *Procolobus gordonorum*) suggests past connectivity to higher elevation forests to the west. Such species persist in lowland areas like Magombera due to their proximity to the coast rather than altitude or slope *per se*; instead of a qualitative definition of near-endemism, the vast majority could be effectively defined as occurring only within the EAM perimeter and up to 10 km or 20 km beyond.

Protected areas

Less than 10% of the EAMs remain forested, compared with 37% preclearance cover. Notwithstanding spatial errors and misclassifications in the protected area and land cover data, we estimated that 75% of present-day forest lies within the protected area network (Table 2.2 and Fig. 2.4). A further 50 km^2 corresponds to forest reserves proposed but not yet gazetted, mainly in South Pare, East Usambara and Udzungwa. Other areas are traditionally managed or in private estates, and so are not represented in the database of protected areas used here (UNEP-WCMC, 2009). In Kenya, forests are gazetted only within Kasigau Forest Reserve. Forests in the Tanzanian blocs are more widely protected by the state, with forest reserves containing 1909 km^2 and nature reserves 785 km^2 (Nilo and Amani in East Usambara, Kilombero in Udzungwa, Uluguru). A further 586 km^2 of EAM forest lies within three national parks, namely Udzungwa Mountains (north-east Udzungwa), Mikumi (all of Malundwe) and Mkomazi (north-east South Pare).

Forest estimates may contain some woodland, up to an eighth of the total, accounting for the lack of protection in some blocs (e.g., North Pare). The forest area within reserves may also be overestimated due to fragmentation resulting from mixed forest cultivation plots and long-lived fire-maintained grasslands (Finch and Marchant, 2011). Conversely, forest regrowth in depopulated areas outside reserves is likely to be underrepresented.

Table 2.2. Summary of the EAMs ecoregion, detailed by mountain bloc.

Mountain bloc	Total area (km^2)	Base height (m)	Summit height (m)	Mountain forest (km^2) †			Mountain population ††		
				Preclearance†	Present-day	% Gazetted	Total (000's)	Mean pp.km^{-2}	Median pp.km^{-2}
Taita Hills	941.4	509	2198	–	10.0	30	60	64	25
North Pare	510.3	697	2099	323.0	40.7	52	69	136	35
South Pare	2327.5	459	2454	1088.7	129.9	89	131	56	11
West Usambara	2945.2	290	2294	2362.2	328.4	80	555	188	38
East Usambara	1145.0	123	1501	807.0	384.3	63	96	84	22
Nguu	1562.9	676	1998	667.9	326.7	56	39	25	7
Nguru	2564.9	351	2382	919.9	357.1	76	95	37	10
Ukaguru	3242.6	415	2259	1075.6	191.0	79	142	44	20
Uluguru	3057.3	121	2636	1627.8	308.6	84	219	72	25
Malundwe	32.8	476	1259	24.1	2.3	100	0	0	0
Rubeho	7984.4	272	2345	2647.8	520.9	57	167	21	10
Udzungwa	19,375.3	249	2556	5790.7	1726.3	82	572	30	10
Mahenge	2606.4	320	1501	557.1	20.2	56	52	20	9
All EAMs	48,296.1	121	2636	17,891.6	4346.3	75	2197	45	12
[Including plateaus]	*[51,628.2]*	*[121]*	*[2636]*	*[17,891.6]*	*[4346.3]*	*[75]*	*[2353]*	*[46]*	*[13]*

† Preclearance forest follows Hall *et al*. (2009). Percent gazetted is according to UNEP-WCMC (2009)

†† Human populations were based on LandScan (2006) estimates, which we corrected according to the protected area data (no people live in National Parks or Game Reserves) and ward-level household surveys from the 2002 Tanzanian census (NBS 2002). See Appendix 2C for details.

Human populations

At the time of the 2002 Tanzania census, we estimate that the EAMs had a population of *c.* 2.2 million people (2.35 million people if including the Udzungwa plateaus) and that a further 5.4 million people lived within 30 km (Table 2.2). Mean and median population densities are around 15% and 250% above the national averages for Tanzania respectively (Fig. 2.5). Median density – more representative of rural communities than the mean – decreases with increasing distance from the bloc perimeters, up to a distance of 30 km (Figs. 2.4 and 2.5). Major towns situated within the boundaries include Same (South Pare), Korogwe (West Usambara), Kilosa (Ukaguru), Mpwapwa (Rubeho) and Iringa (Udzungwa), with Morogoro and Ifakara just outside the Uluguru and Udzungwa blocs, respectively. According to these estimates, population densities are highest in West Usambara and North Pare, followed by East Usambara and Uluguru (Table 2.2).

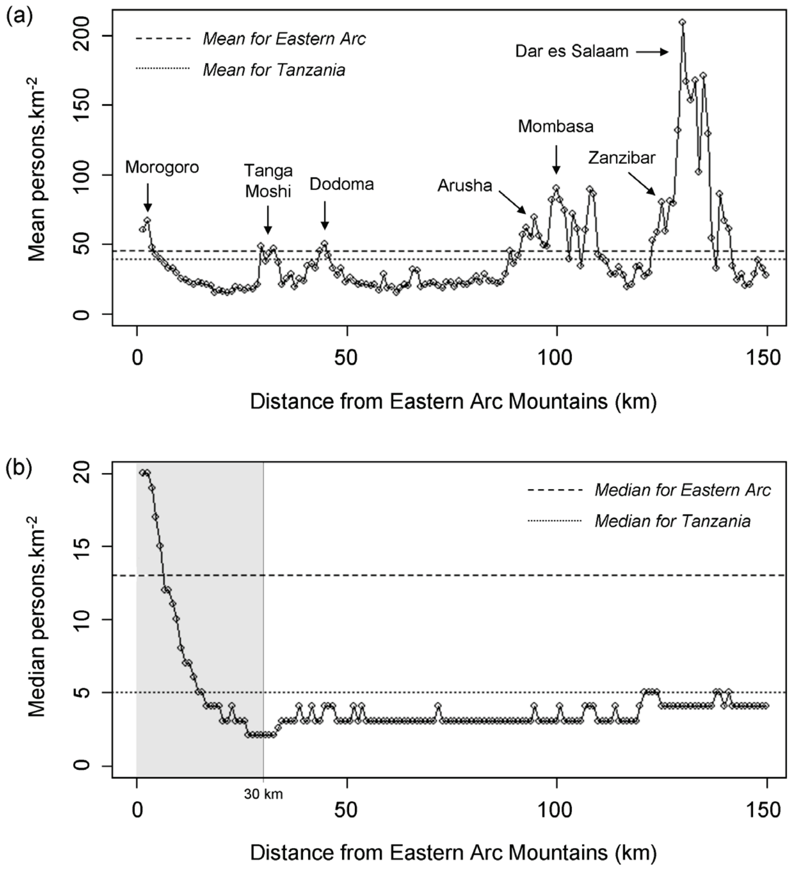


Figure 2.5. Human population density *vs*. distance to the EAMs (includes Udzungwa plateaus). (a) Peaks in mean density correspond to towns and cities with populations exceeding 100,000. (b) Median density better portrays the distribution of rural persons in relation to the mountain resource (0-30 km).

Discussion

Ecoregions are useful tools for making explicit the strong biogeographical affinities that exist within many regions of the world. Where consistently defined, they provide spatial platforms for monitoring, managing and forecasting the health of constituent ecosystems, as well as the people reliant upon them for natural resources. Communities living in and around tropical mountains benefit from lower ambient temperatures, access to forest products, clean potable water, improved agricultural potential and fewer vectors for disease (e.g., mosquitoes). Advantages such as these are evident from the decrease in rural population density with increasing distance from the EAM boundary, a trend also affected by restrictions on human settlement in adjacent game reserves (Appendix 2C). Large-scale studies suggest a general pattern of human pressures in biologically important regions (Cincotta *et al.*, 2000; Balmford *et al.*, 2001; but see Joppa *et al.*, 2009), driven by real or perceived benefits or by extraction frontiers, such as mining or logging, which open up previously remote environments for human settlement (Joppa *et al.*, 2010; Scholte and de Groot, 2010).

In common with much of tropical Africa, the population of Tanzania has increased dramatically over the last half century, from ten million people in 1960 to 42 million people in 2008 (http://data.worldbank.org/). Kenya has experienced a similar population boom, from eight million to 39 million people over the same period. Population growth is exerting increasing pressure on water supplies, energy production and land for agriculture. Demand for timber, poles and charcoal is also increasing as stocks in more accessible Miombo woodlands and coastal forests diminish under pressure from urban centres such as Dar es Salaam (Ahrends *et al.*, 2010). Increased resource demand and intensified land use, exacerbated by climate change, have direct implications for forest health and local livelihoods, as well as more diffuse impacts such as biodiversity loss and the release of sequestered carbon into the atmosphere (Kohler *et al.*, 2010).

Global typologies derived from digital elevation data provide a welcome platform for large-scale studies of mountain environments and the people they support (Huddleston *et al.*, 2003). At more local scales, regional biogeography, naming conventions and micro-relief gain importance, and so targeted case studies are required to obtain relevant delineations of region extent. Here, we provide a consistent topographical foundation for delimiting the EAMs, one of the world's most important ecoregions for conservation (Olson and

Dinerstein, 2002). The approach we present is generic and could be readily calibrated for application to other mountain regions.

Mountain ecosystems, like all biological assemblages, are moving targets: the products of ancient evolutionary processes, recent climatic conditions and, in the case of the EAMs, on-going disturbance by humans, fire and large herbivores such as elephants. Ecoregion extent inferred directly from climate, vegetation and/or land use therefore requires frequent revision in response to new data or changes in those variables. Moreover, if considered independently of surrounding habitats and without historical perspective, contemporary snapshots of forest mosaics and local climates have restricted potential to aid understanding of the system as a whole (Fjeldså and Lovett, 1997). In cases where mountainous relief is central to the historical affinities under consideration, boundaries defined by topographical means are in our view preferable. Here, they provide a geographically coherent framework for monitoring, which is likely to incorporate potentially important but as yet undocumented sites for conservation.

When the EAMs were first proposed as phytogeographically distinct in the 1980s, the biological importance of many of the blocs was unconfirmed. Forests in Rubeho, for instance, having only recently received funding for botanical surveys, have long been under-valued in terms of conservation priority (Doggart *et al.*, 2006). Similarly, recent focus in Nguru has revealed a number of species new to science, particularly amongst herpetofauna (Menegon *et al.*, 2008). Nguu remains largely unsurveyed, but recent field visits and bioclimatic modelling suggest it too could be species rich (Chapter 4). The majority of forest in eastern parts of Nguru and Nguu are within reserves, but their western outliers remain ungazetted, leaving them open to degradation.

The size of the ecoregion is here defined to be 48,000–52,000 km^2, depending on the inclusion of mid-elevation plateaus in Udzungwa. This is higher than a previous estimate of 37,000 km^2 (Tanzania only) published by the Forestry and Beekeeping Division (Mbilinyi *et al.*, 2006), which imposes a 500 m lower limit, omits the western margins of Udzungwa and includes fewer outlying peaks. In Kenya, we identify the main Taita bloc plus Mt. Kasigau and the Sagalla Hills to the south and east, as well as two lower elevation hills (Maungu and Maktau) known to support locally endemic plant species. The explicit inclusion of these outliers is especially important given the plight of other Taita forests, now restricted to a few remnant patches (Pellikka *et al.*, 2009), threatening the persistence of many rare species (Rogo and Oguge, 2000). Although farthest from the main bloc, Kasigau forests are relatively undisturbed and so may provide refugia for important flora and fauna (e.g. Taita White-eye *Zosterops silvanus*; Mulwa *et al.*, 2007). Elsewhere, outlying features were identified by the regional mountain typology but omitted from the current definition on account of low relative elevation and lack of data on the presence/absence of characteristic EAM vegetation. Because new data may yet justify their inclusion, we provide the spatial extents of all identified features in the supplementary material (Appendix 2A).

Although these boundaries were not placed directly according to forest distributions, we did use vegetation data to indicate appropriate terrain parameters. Our estimation of present-day forest cover at over 4300 km^2 is more than some previous studies, but less than the 5700 km^2 of natural forest according to Newmark (2002). Other sources put the figure closer to 3500 km^2 (Mbilinyi *et al.*, 2006; Burgess *et al.*, 2007 and references therein), but are similar to Newmark's estimate if woodlands are included. The land cover data used here have been iteratively improved during a series of workshops in Tanzania through the Valuing the Arc Programme (http://valuingthearc.org/), and were further corrected for the current application by reference to forest change estimates, government reports and our own field notes. Unless ground-truthed, all such estimates are subject to uncertainty, especially as regards the distinction between closed woodland and deciduous/degraded forest, which can be easily confused in remotely sensed images. Levels of disturbance and fragmentation in the forest interior are also difficult to assess remotely.

To reduce our reliance on these data, which at best provide a present-day snapshot of forest distribution, we further considered extrapolations of historical forest cover (Hall *et al.*, 2009). Such estimates are not directly applicable as ecoregion boundaries as sediment cores are sparsely distributed and an evidence-based assessment of historical extent is only starting to become possible (Mumbi *et al.*, 2008; Finch *et al.*, 2009; Finch and Marchant, 2011), but they do provide a broad indication of preclearance cover. The fact that our boundaries fully enclose these kinds of estimates suggests that they are well placed to withstand future changes in distribution, at least over time scales relevant for management. If the boundaries were to require adjustment in the future, this could be addressed multilaterally according to the framework presented here.

Genetic evidence from a range of focal taxa shows that EAM vegetation has persisted over many millions of years. Phylogenetic analysis of the tree *Macaranga capensis* indicates long-term separation of populations on different mountain blocs, but also past connectivity (A. S. Jump, personal communication 2010); i.e., species presently restricted to montane areas may have once encroached on tropical lowlands. Strong affinities have also been observed between the flora and fauna of the EAMs and those of west and central Africa, suggesting remnants of a pan-African forest belt (Couvreur *et al.*, 2008). More locally, the Lake Nyasa Highlands, Mt. Kilimanjaro and coastal forests all contain plant taxa otherwise restricted to the EAMs, as do the Shimba Hills in south-east Kenya, again indicating past connections or possibly long-distance dispersal.

Because of these historical overlaps in composition, the dichotomy of EAM and coastal vegetation can sometimes be contentious (Burgess and Clarke, 2000). An academic reason for maintaining some form of distinction between mountain and coastal vegetation in East Africa is that since 2004 they have been classified as belonging to different biodiversity hotspots (Eastern Afromontane *vs*. Coastal Forests of East Africa; Mittermeier *et al.*, 2004). In addition, the Tanzanian mountains contain central government reserves administered by the catchment forest office, whereas the districts administer coastal forest reserves. From a human perspective, mountain people face different challenges to those living in towns or rural lowlands (Kreutzmann, 2001); thus ensuring their social welfare and right to the traditional use of natural resources, whilst also fulfilling conservation objectives, demands a specific focus on the mountain region.

The 500 m threshold for plant endemism, although a pragmatic response to the need for spatially consistent mountain limits, does not account for geographical differences in baseline elevation, nor is it consistent with other taxonomic groups (*cf*. Poynton *et al.*, 2007). Using the altitudinal range of forest within a given mountain bloc is an appealing alternative, but one that is particularly sensitive to recent patterns of deforestation. The spatial limits presented here might be an appropriate basis for a new set of endemism criteria. Our results suggest that well over half of the plant species here tentatively considered ‘near-endemic’ would be classified as strictly endemic according to the derived ecoregion boundary, and that all could be consistently defined according to an inclusion zone of up to 20 km. Moreover, in a preliminary retabulation of plant endemism, which compares the recorded elevations of specimens with the altitudinal limits of the corresponding mountain blocs (as detailed in Table 2), nearly all ‘near-endemics’ can be considered strictly endemic. We encourage similar tests for other taxonomic groups (spatial data available online; see Appendix 2A).

In providing a spatially explicit definition for this area of endemism, we hope to prompt research and conservation in lesser-studied parts of the EAMs, which could be biologically or politically undervalued due to spatial bias in the data (Chapter 4; also Ahrends *et al*., 2011 in Appendix I). The boundaries are also relevant for reforestation strategies, particularly ahead of the REDD pilot in Tanzania.

Acknowledgements

Our thanks to providers of data to the Missouri Botanical Garden's TROPICOS database, local experts for advice on land cover classification, and for the comments of two anonymous reviewers. This work was funded by the Marie-Curie programme of the European 6th Framework (MEXT-CT-2004-517098), and by grants from the Academy of Finland, Leverhulme Trust, Packard Foundation and Royal Society.

Author contributions

The study was conceived by P.J.P., who conducted all analyses and prepared the manuscript. R.E.G. advised on endemism status. A.R.M. and P.K.E.P. provided estimates of forest cover in the Udzungwa Mountains and Taita Hills, respectively. N.D.B. and R.D.S. advised on ecosystem service provision, conservation priorities and land classification. J.C.L. advised on historical depictions of the EAMs and on policy implications. The work was supervised by R.M. and C.J.M.

Appendix 2A. Description of spatial data for boundary placement

The following files are available online at http://www.journals.cambridge.org/enc2011004. The projection is Africa Albers equal-area conic (WGS 1984 spheroid, central meridian 25E, standard parallels 20N-23S). Boundaries were derived using elevation data from the CGIAR-SRTM DEM (Jarvis *et al.*, 2008). For general use, we recommend version [9] because it includes all highland habitats as well as the lowest parts of Matundu forest in Udzungwa. Two alternative coordinate systems are provided: UTM zone 37S and geographic WGS 1984, available at http://www.journals.cambridge.org/enc2011002 and http://www.journals.cambridge.org/enc2011003, respectively.

	File name (dbf, prj, shp, shx)	Description
[1]	EasternArc_halfdegSquares	56 half-degree grid squares that intersect with the EAM region. Defines the spatial limits of data in [2-9]
[2]	RegionalMountainTypology	Mountain classification as defined in Appendix 2B
[3]	MountainClusters_prominence	Mountainous features from [2] aggregated within simplified bounds. Prominence is summit height minus lowest boundary elevation. Small, isolated fragments not included (< 1 km^2 and > 1 km from nearest neighbour)
[4]	EasternArc_byPart	EAM boundary definition, according to topographic prominence (≥ 600 m) and phytogeographical divisions described by Lovett (1990)
[5]	EasternArc_byBloc	As in [4], dissolved by mountain bloc
[6]	EasternArc_byBloc_incMikumi	As in [5], retaining Mikumi relief that surrounds Malundwe Mountain
[7]	EasternArc_byBloc_incPlateaus	As in [5], with the addition of mid-elevation plateaus in Udzungwa
[8]	EasternArc_inclusionZones	As in [7], including 10, 20 and 30 km buffers around the perimeter. Some lowland remnants of characteristic EAM forest persist within 10 km (e.g., Magombera); all 'near-endemic' plant taxa (as defined in the main article text) have been collected within 20 km; the density of rural persons is highest within 30 km
[9]	EasternArc_byBloc _incPlateaus&LowMatundu	As in [7], with the addition of low-lying forest in Matundu. This version was created for the Valuing the Arc Programme (http://www.valuingthearc.org/) to investigate ecosystem service provision. It is also used by the KITE project (http://www.york.ac.uk/res/kite/) to assess biodiversity and land cover change, and by Missouri Botanical Garden to investigate the spatial limits of plant endemism.

Appendix 2B. Mountain classes in the regional typology

Mountain classes defined in the regional mountain typology (90 m raster resolution). LER is local elevational range, defined across a circle of 500 m radius. Areas include south-west slopes of Mt. Kilimanjaro and northern parts of the Lake Nyasa Highlands (see file [2]).

Mountain Class	Elevational zone	Slope and LER constraints	Area in half-degree grid cells (km^2)
1	0–1000 m	$LER_{500\ m} \geq 100$ m	15,890
2	1000–1500 m	Slope > 10° or $LER_{500\ m} \geq 100$ m	13,570
3	1500–2500 m	Slope > 5°	11,579
4	2500–3500 m	No constraint	96
5	3500–4500 m	No constraint	11
6	≥ 4500 m	No constraint	0
7	[Isolated inner basins and plateaus < 25 km^2 in area]		1913

Appendix 2C. Derivation of the population surface

The distribution of persons at 1 km resolution was based on a modelled population surface, provided by LandScan (LandScan, 2006; 30 arc-sec resolution). These estimates were rescaled on a ward-by-ward basis, such that ward totals matched empirical data from the 2002 Tanzania census (NBS 2002). An additional correction was made, such that no person should be permanently resident in a national park or game reserve (World Database on Protected Areas; UNEP-WCMC, 2009). It was not appropriate to use previous LandScan releases, closer to the census year, because of errors and shortfalls in those datasets (see http://www.ornl.gov/sci/landscan/). Census data were not available for Kenya, and so population estimates for the Taita Hills were mapped directly from LandScan.

ArcMap procedure for deriving the high-resolution population surface in Tanzania

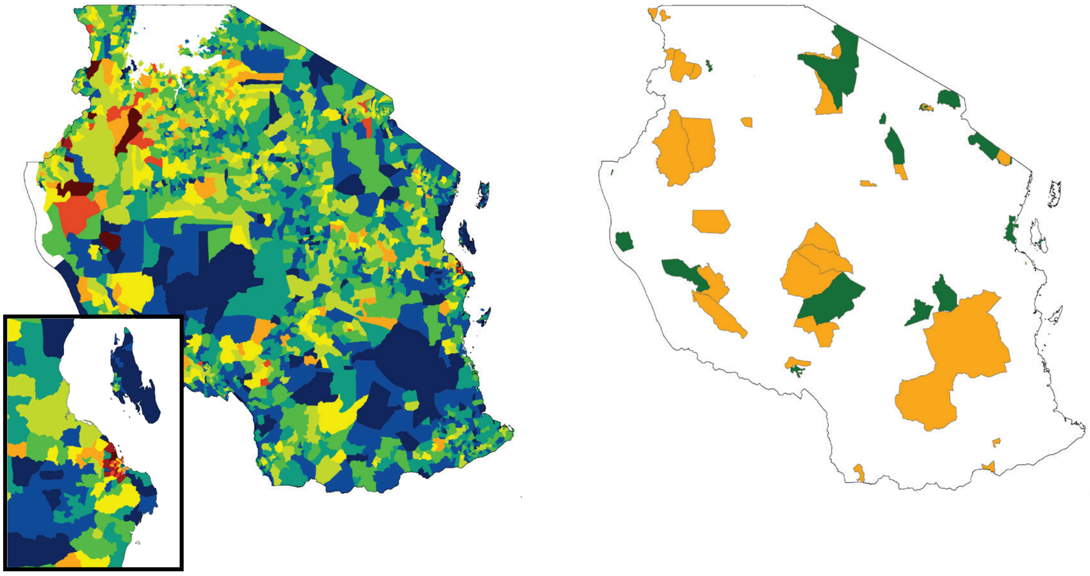
1. Using the total population field in the Tanzania census (2002) shapefile, grid the wards polygon data at 30 arc-sec resolution (WGS1984 coordinate system; Feature to Raster Tool). Let this raster be called ‘wardPop02’.

2. Extract national parks and game reserves from the WDPA (2009) shapefile. Grid these polygons at 30 arc-sec resolution and reclassify NoData values such that national parks and game reserves have zeroes and all other cells have ones (Reclassify Tool). Multiply this grid and the LandScan (2006) grid to remove persons from these protected areas (Single Map Algebra Tool). Let the resultant raster be called ‘LS06_WDPA’.

3. Using the census ward number as the zone marker, calculate the total ‘LS06_WDPA’ population in each census ward (Zonal Statistics Tool). Depending on computational power, it may be necessary to perform this operation separately for each region of Tanzania (merge results using Mosaic to Raster Tool). Let the resultant raster be called ‘LS06_WDPA_tot’.

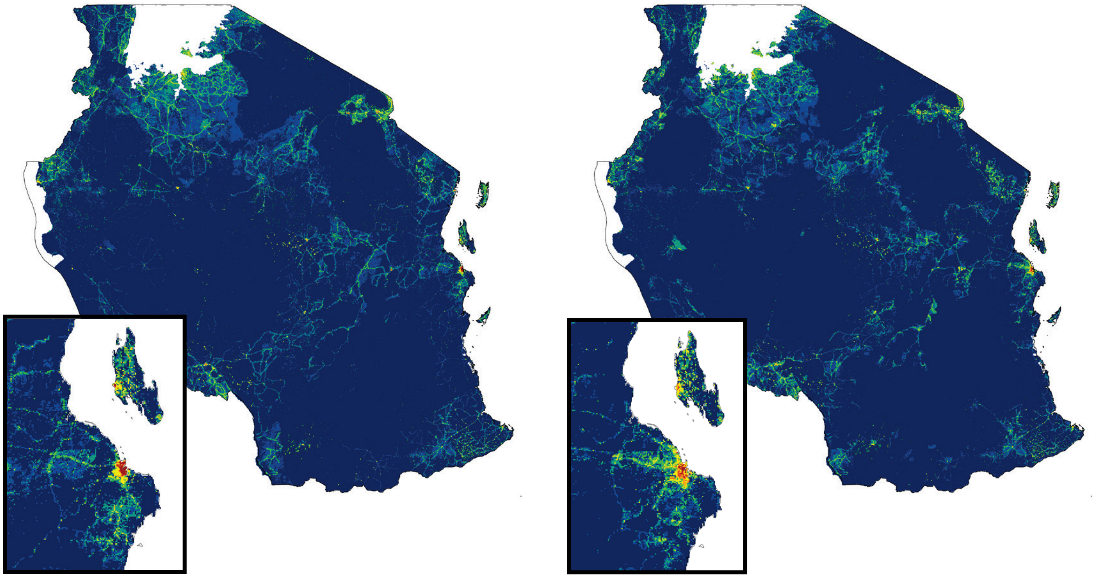
4. Using the Single Map Algebra Tool, divide ‘LS06_WDPA’ by ‘LS06_WDPA_tot’ (allocate zero values where the quotient is undefined) and multiply by ‘wardPop02’. The result is a 30 arc-sec population surface where ward totals match the 2002 census counts, and where no people are resident in game reserves or national parks. Let this grid be called ‘popgrid02_dd’.

5. Population density in the form persons.km^{-2} (rather than persons per 30 arc-sec) can be obtained as follows. First, calculate the areas of all 30 arc-sec grid squares in Tanzania. This can be achieved by generating a 30 arc-sec raster in which each cell has a unique value, converting this to polygons, adding an ‘area_km2’ field to the attribute table (calculate geometry using Africa Albers Equal Area Conic coordinate system), and then converting this shapefile back to raster format (snap to ‘popgrid02_dd’). Second, divide ‘popgrid02_dd’ by the area grid. Finally, project the resultant raster from the geographic coordinate system (WGS1984) to UTM zone 37S or Africa Albers, using nearest neighbour assignment.



(a) Ward populations, 2002 Tanzania census

(b) Protected areas where no people live



(c) LandScan 2006 (pp.km^{-2})

(d) High-resolution estimates for 2002 (pp.km^{-2})

(a) Ward populations according to the Tanzania census in 2002. (b) Game reserves (orange) and national parks (green) from the World Database on Protected Areas. (c) Population distribution according to LandScan 2006. (d) As in [c] but with cell values adjusted to match the ward census counts and protected area data. Respectively, cool colours (blue-green) and warm colours (yellow-red) correspond to areas of low and high population. Insets show zoomed perspectives of Dar es Salaam and Zanzibar Island.

References

Ahrends, A., Burgess, N. D., Milledge, S. A. H., Bulling, M. T., Fisher, B., Smart, J. C. R., Clarke, G. P., Mhoro, B. E. and Lewis, S. L. (2010) Predictable waves of sequential forest degradation and biodiversity loss spreading from an African city. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 14556-14561.

Ahrends, A., Burgess, N. D., Gereau, R. E., Marchant, R., Bulling, M. T., Lovett, J. C., Platts, P. J., Kindemba, V. W., Owen, N., Fanning, E. and Rahbek, C. (2011) Funding begets biodiversity. *Diversity and Distributions*, **17**, 191-200.

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

Blyth, S., Groombridge, B., Lysenko, I., Miles, L. and Newton, A. (2002) *Mountain Watch*. UNEP World Conservation Monitoring Centre, Cambridge.

Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Rylands, A. B., Konstant, W. R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G. and Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909-923.

Browne, T., Fox, R. and Funnell, D. (2004) The "invisible" mountains - Using GIS to examine the extent of mountain terrain in south Africa. *Mountain Research and Development*, **24**, 28-34.

Burgess, N., Mwakalila, S., Madoffe, S., Ricketts, T., Olwero, N., Swetnam, R., Mbilinyi, B., Marchant, R., Mtalo, F., White, S., Munishi, P., Marshall, A., Malimbwi, R., Jambiya, G., Fisher, B., Kajembe, G., Morse-Jones, S., Kulindwa, K., Green, J. and Balmford, A. (2009) Valuing the Arc - A Programme to Map and Value Ecosystem Services in Tanzania. *Mountain Research Initiative Newsletter*, **18**, 18-21.

Burgess, N. D., Bahane, B., Clairs, T., Danielsen, F., Dalsgaard, S., Funder, M., Hagelberg, N., Harrison, P., Haule, C., Kabalimu, K., Kilahama, F., Kilawe, E., Lewis, S. L., Lovett, J. C., Lyatuu, G., Marshall, A. R., Meshack, C., Miles, L., Milledge, S. A. H., Munishi, P. K. T., Nashanda, E., Shirima, D., Swetnam, R. D., Willcock, S., Williams, A. and Zahabu, E. (2010) Getting ready for REDD+ in Tanzania: a case study of progress and challenges. *Oryx*, **44**, 339-351.

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. and Stuart, S. N. (2007) The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, **134**, 209-231.

Burgess, N. D. and Clarke, G. P. (2000) *Coastal Forests of Eastern Africa*, IUCN, Gland, Switzerland and Cambridge, UK.

Burgess, N. D., Hales, J. D., Ricketts, T. H. and Dinerstein, E. (2006) Factoring species, non-species values and threats into biodiversity prioritisation across the ecoregions of Africa and its islands. *Biological Conservation*, **127**, 383-401.

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

Clark, B. J. F. and Pellikka, P. K. E. (2009) Landscape analysis using multiscale segmentation and object-orientated classification. In *Recent Advances in Remote Sensing and Geoinformation Processing for Land Degradation Assessment* (Eds, Röder, A. and Hill, J.) Taylor & Francis Group 2009, pp. 323-342.

Conservation-International (2008) Forest cover and change in Eastern Arc Mountain and Coastal Forests, 1990-2000 (http://www.cepf.net/where_we_work/regions/africa/eastern_arc_coastal_forests/).

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

Doggart, N., Perkin, A., Kiure, J., Fjeldså, J., Poynton, J. and 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.

Elenga, H., Peyron, O., Bonnefille, R., Jolly, D., Cheddadi, R., Guiot, J., Andrieu, V., Bottema, S., Buchet, G., de Beaulieu, J. L., Hamilton, A. C., Maley, J., Marchant, R., Perez-Obiol, R., Reille, M., Riollet, G., Scott, L., Straka, H., Taylor, D., Van Campo, E., Vincens, A., Laarif, F. and Jonson, H. (2000) Pollen-based biome reconstruction for southern Europe and Africa 18,000 yr BP. *Journal of Biogeography*, **27**, 621-634.

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

Finch, J. M. and Marchant, R. (2011) A palaeoecological investigation into the role of fire and human activity in the development of montane grasslands in East Africa. *Vegetation History and Archaeobotany*, **20**, 109-124.

Fjeldså, J. and Lovett, J. C. (1997) Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation*, **6**, 325-346.

Gerrard, A. J. (1990) *Mountain Environments: An Examination of the Physical Geography of Mountains*, Belhaven Press, London.

Hall, J., Burgess, N. D., Lovett, J., Mbilinyi, B. and Gereau, R. E. (2009) Conservation implications of deforestation across an elevational gradient in the Eastern Arc Mountains, Tanzania. *Biological Conservation*, **142**, 2510-2521.

Huddleston, B., Ataman, E., de Salvo, P., Zanetti, M., Bloise, M., Bel, J., Franceschini, G. and Fè d'Ostiani, L. (2003) *Towards a GIS-based Analysis of Mountain Environments and Populations*, Environment and Natural Resources Working Paper No. 10, Food and Agriculture Organisation of the United Nations, Rome, Italy.

Jarvis, A., Reuter, H. I., Nelson, A. and Guevara, E. (2008) Hole-filled seamless SRTM data. Version 4. International Centre for Tropical Agriculture (CIAT). http://srtm.csi.cgiar.org/.

Joppa, L. N., Loarie, S. R. and Nelson, A. (2010) Measuring population growth around tropical protected areas: current issues and solutions. *Tropical Conservation Science*, **3**, 117-121.

Joppa, L. N., Loarie, S. R. and Pimm, S. L. (2009) On population growth near protected areas. *Plos One*, **4**, e4279.

Kapos, V., Rhind, J., Edwards, M., Price, M. F. and Ravilious, C. (2000) Developing a map of the world's mountain forests. In *Forests in Sustainable Mountain Development: a State of Knowledge Report for 2000* (Eds, Price, M. F. and Butt, N.) CAB International, Wallingford, UK, pp 4-9.

Kohler, T., Giger, M., Hurni, H., Ott, C., Wiesmann, U., von Dach, S. W. and Maselli, D. (2010) Mountains and Climate Change: A Global Concern. *Mountain Research and Development*, **30**, 53-55.

Kreutzmann, H. (2001) Development indicators for mountain regions. *Mountain Research and Development*, **21**, 132-139.

LandScan (2006) High Resolution global Population Data Set copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy.

Larigauderie, A. and Mooney, H. A. (2010) The Intergovernmental science-policy Platform on Biodiversity and Ecosystem Services: moving a step closer to an IPCC-like mechanism for biodiversity. *Current Opinion in Environmental Sustainability*, **2**, 9-14.

Lovett, J. C. (1985) Moist forests of eastern Tanzania. *Swara*, **8**, 8-9.

Lovett, J. C. (1990) Classification and status of the moist forests of Tanzania. *Notifications from the Institute for General Botany Hamburg*, **23a**, 287–300.

Lovett, J. C. (1992) *Classification and Affinities of the Eastern Arc Moist Forests of Tanzania*, PhD Thesis, University of Wales, Bangor, UK.

Lovett, J. C., Hansen, J. R. and Hørlyck, V. (2000) Comparison with Eastern Arc forests. In *Coastal Forests of Eastern Africa* (Eds, Burgess, N. D. and Clarke, G. P.) IUCN, Gland, Switzerland and Cambridge, UK.

Lovett, J. C. and Pócs, T. (1993) *Assessment of the condition of the Catchment Forest Reserves, a botanical appraisal*. Catchment Forest Report (http://www.easternarc.or.tz/).

Marshall, A. R., Jorgensbye, H. I. O., Rovero, F., Platts, P. J., White, P. C. L. and Lovett, J. C. (2010) The species area relationship and confounding variables in a threatened monkey community. *American Journal of Primatology*, **72**, 325-336.

Mbilinyi, B. P., Malimbwi, R. E., Shemwetta, D. T. K., Songorwa, Zahabu, E., Katani, J. Z. and Kashaigili, J. (2006) *Forest Area Baseline for the Eastern Arc Mountains*, Conservation and Management of the Eastern Arc Mountain Forests, Forestry and Beekeeping Division, Dar es Salaam, The United Republic of Tanzania.

McMahon, G., Wiken, E. B. and Gauthier, D. A. (2004) Toward a scientifically rigorous basis for developing mapped ecological regions. *Environmental Management*, **34**, S111-S124.

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

Messerli, B. and Ives, J. D. E. (1997) *Mountains of the World: A Global Priority*, Parthenon, New York.

Meybeck, M., Green, P. and Vorosmarty, C. (2001) A new typology for mountains and other relief classes: An application to global continental water resources and population distribution. *Mountain Research and Development*, **21**, 34-45.

Mittermeier, R. A., Robles Gil., P., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., Lamoreux, J. and da Fonseca, G. A. B. (2004) *Hotspots Revisited*, CEMEX, Mexico.

MNRT (1997) *National Reconnaissance Level Land Use and Natural Resources Mapping Project*, Ministry of Natural Resources and Tourism, The United Republic of Tanzania.

Mulwa, R. K., Bennun, L. A., Ogol, C. and Lens, L. (2007) Population status and distribution of Taita White-eye Zosterops silvanus in the fragmented forests of Taita Hills and Mount Kasigau, Kenya. *Bird Conservation International*, **17**, 141-150.

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

NBS (2002) *Tanzania Population and Housing Census 2002*, National Bureau of Statistics, Dar es Salaam, The United Republic of Tanzania.

Newmark, W. D. (2002) *Conserving Biodiversity in East African Forests: a Study of the Eastern Arc Mountains*, Springer, Berlin, Germany.

Nordregio (2004) *Mountain Areas in Europe: Analysis of Mountain Areas in EU Member States, Acceding and Other European Countries*, Nordic Centre for Spatial Development, Stockholm, Sweden.

OED (1989) *Oxford English Dictionary* 2nd Ed, Oxford University Press, Oxford, UK.

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

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P. and Kassem, K. R. (2001) Terrestrial ecoregions of the worlds: A new map of life on Earth. *Bioscience*, **51**, 933-938.

Pellikka, P. K. E., Lötjönen, M., Sijander, M. and Lens, L. (2009) Airborne remote sensing of spatiotemporal change (1955-2004) in indigenous and exotic forest cover in the Taita Hills, Kenya. *International Journal of Applied Earth Observation and Geoinformation*, **11**, 221-232.

Poynton, J. C., Loader, S. P., Sherratt, E. and Clarke, B. T. (2007) Amphibian diversity in east african biodiversity hotspots: Altitudinal and latitudinal patterns. *Biodiversity and Conservation*, **16**, 1103-1118.

Pócs, T. (1976) Vegetation mapping in the Uluguru mountains (Tanzania, East Africa). *Boissiera*, **24**, 477-498.

Rogo, L. and Oguge, N. (2000) The Taita Hills forest remnants: A disappearing world heritage. *Ambio*, **29**, 522-523.

Schlüter, T. (1997) Geology of East Africa. In *Contributions to the Regional Geography of the Earth* (translated from German), Borntraeger, Berlin, Germany.

Scholte, P. and de Groot, W. T. (2010) From Debate to Insight: Three Models of Immigration to Protected Areas. *Conservation Biology*, **24**, 630-632.

Stanley, W. T. and Olson, L. E. (2005) Phylogeny, phylogeography, and geographic variation of Sylvisorex howelli (Soricidae), an endemic shrew of the Eastern Arc Mountains, Tanzania. *Journal of Zoology*, **266**, 341-354.

Thompson, W. F. (1964) How and why to distinguish between mountains and hills. *Professional Geographer*, **16**, 6-8.

UNEP-WCMC (2009) *The World Database on Protected Areas*, UNEP-WCMC, Cambridge, UK (http://www/wdpa.org/).

Vincens, A., Garcin, Y. and Buchet, G. (2007) Influence of rainfall seasonality on African lowland vegetation during the Late Quaternary: pollen evidence from Lake Masoko, Tanzania. *Journal of Biogeography*, **34**, 1274-1288.

Chapter 3 – Predicting Tree Distributions

Published in *Ecological Modelling* (Oct. 2008), Vol. 218, Issues 1-2, Pages 121-134

Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty

Philip J. Platts [a,]*, Colin J. McClean [b], Jon C. Lovett [b], Rob Marchant [a]

[a] *The York Institute for Tropical Ecosystem Dynamics (KITE), Environment Department, University of York, Heslington, York YO10 5DD, UK.* [b] *Centre for Ecology Law and Policy, Environment Department, University of York, Heslington, York YO10 5DD, UK.*

* *Corresponding author. Email: philip.platts@gmail.com*

Abstract

The Eastern Arc Mountains (EAMs) of Tanzania and Kenya support some of the most ancient tropical forest on Earth. The forests are a global priority for biodiversity conservation and provide vital resources to the Tanzanian population. Here, we make a first attempt to predict the spatial distribution of 40 EAM tree taxa (38 species), using generalised additive models, plot data and environmental predictor maps at 1 km resolution. The results of three modelling experiments are presented, investigating predictions obtained by (1) two different procedures for the stepwise selection of predictors, (2) down-weighting absence data, and (3) incorporating autocovariate terms to describe fine-scale spatial aggregation. In response to recent concerns regarding the extrapolation of model predictions beyond the restricted environmental range of training data, we also demonstrate a novel graphical tool for quantifying envelope uncertainty in restricted range niche-based models (envelope uncertainty maps). We find that even for taxa with very few documented occurrences useful estimates of distribution can be achieved. Initiating selection with a null model is found to be useful for explanatory purposes, whilst beginning with a full predictor set can over-fit the data. We show that a simple multimodel average of these two best-model predictions yields a superior compromise between generality and precision. Down-weighting absences shifts the balance of errors in favour of higher sensitivity, reducing the number of serious mistakes (i.e., falsely predicted absences); however, response functions are more complex, exacerbating uncertainty in larger models. Spatial autocovariates help describe fine-scale patterns of occurrence and significantly improve explained deviance, though if important environmental constraints are omitted then model stability and explanatory power can be compromised. We conclude that the best modelling practice is contingent both on the intentions of the analyst (explanation or prediction) and on the quality of distribution data. Generalised additive models have potential to provide valuable information for conservation in the EAMs, but methods must be carefully considered, particularly if occurrence data are scarce.

Keywords: Eastern Arc Mountains; tropical trees; generalised additive models; stepwise selection; model averaging; prevalence; spatial autocorrelation; extrapolation uncertainty.

Introduction

Research into the habitat requirements of species plays a fundamental role in planning for their future conservation, particularly if external pressures such as disturbance and climatic change threaten their persistence. Vegetation surveys provide point data for many taxa, but invariably survey sites are too sparse or spatially biased for species distributions to be estimated directly (Küper *et al.*, 2006). One solution is to model the likelihood of occurrence as a function of the local environment, using the available distribution data and environmental variables as predictors of habitat suitability. Species distribution models have been used previously for biodiversity analysis (Austin, 1999; Ferrier *et al.*, 2002b), improved sampling of rare and endangered species (Engler *et al.*, 2004; Guisan *et al.*, 2006), determination of reserve boundaries (Ferrier *et al.*, 2002a; Araújo *et al.*, 2004), historical reconstruction (Richards *et al.*, 2007) and assessment of climate change impacts (Thomas *et al.*, 2004; McClean *et al.*, 2005). All of these applications could prove extremely useful for the Eastern Arc Mountains of Tanzania and Kenya (EAMs; Lovett, 1985), one of the most important regions for conservation in the world (Olson and Dinerstein, 1998; Stattersfield *et al.*, 1998; Myers *et al.*, 2000), yet to our knowledge no regional-scale predictive model for tree distributions in this area has been published.

The EAMs are a particularly challenging environment to model, characterised by steep climatic gradients that must be portrayed at a high spatial resolution if the environmental tolerances of taxa are to be properly described. The study presented here uses generalised additive models (GAMs; Hastie and Tibshirani, 1990) to parameterise the responses of 38 large tree species (40 taxa, including subspecies and varieties) to a number of climatic and topographic gradients. GAMs are a semi-parametric class of regression model, chosen because of their ability to describe highly non-linear responses (Yee and Mitchell, 1991; Austin, 2007). The aim is to assess the potential of this data-driven tool for assisting research and conservation in the EAMs – the application of GAMs to small environmental datasets is increasingly common, but often due consideration is not given to pitfalls such as over-fitting.

As is common for studies of this nature, the distribution data available to us are not well suited to high-resolution raster-based regression analysis. Impediments to model performance may include mislocated or misidentified samples, low sample size and prevalence, and a biased or restricted distribution of occurrence data. In order to obtain robust estimates of species distributions, and for the benefit of other studies faced with similar challenges, we compare baseline model predictions with those that incorporate down-weighted absences (Maggini *et al.*, 2006) and spatial autocovariates (Augustin *et al.*, 1996). Given that predictions can be highly sensitive to the predictor sets used for modelling (e.g., Dormann *et al.*, 2008), we also calibrate and compare three different methods for model selection: two best-model stepwise procedures and one multimodel.

Model selection

The goal of selection is to construct as parsimonious a predictor set as possible whilst retaining sufficient information to predict the given distribution. A widely used procedure is to select predictors in a stepwise manner, beginning with either a null model (forward selection) or a full model (backward selection) and adding or removing predictors according to their impact on a global measure of model performance (Eberhardt, 2003). Marginal statistics can be biased by the inevitable collinearity amongst environmental predictors (Cohen *et al.*, 2003; Graham, 2003), and so the use of null hypothesis tests during selection is best avoided. Issues of multiple testing (Pearce and Ferrier, 2000a; Whittingham *et al.*, 2006) and arbitrary levels of statistical significance (Mickey and Greenland, 1989; Rushton *et al.*, 2004) further enforce this standpoint. Multimodel inference has been proposed as an alternative to best-model stepwise procedures. Anderson *et al.* (2000) for instance describe an approach called information-theoretic (IT), in which a number of good models are identified from an *a priori* set of hypotheses (predictor sets) and then compared using Akaike Information Criterion (AIC; Akaike, 1973), or combined in a model-average using Akaike weights. Although not strictly adhering to the IT philosophy of multimodel inference, many studies now adopt the use of AIC in stepwise procedures.

Data bias

With absences often far outweighing presences, particularly for rare and less well-known species, low sample prevalence is a common problem that can lead to misleading evaluations (Manel *et al.*, 2001; Engler *et al.*, 2004; McPherson *et al.*, 2004). A standardised prevalence can be achieved by applying weights to the absence data prior to parameterisation, as demonstrated by Maggini *et al.* (2006) in their modelling of Switzerland's forest communities. The technique was shown to perform well, improving both the accuracy and stability of predictions. Maggini *et al.* found that the application of weights increased the overall probabilities of occurrence, and also report that the balance of model fit may have been altered. It is the latter in which we see potential for improving our predictions: absence 'observations' are inherently unreliable (Anderson, 2003), and since misclassifications distort the modelled relationship between species and environment it follows that a strategic reduction in the dependence of models on absence data could be beneficial. Simulations based on use-availability data (resource selection function modelling; Johnson *et al.*, 2006) suggest that logistic regression is relatively robust to contamination rates of below 20% – a level that could well be exceeded in our data.

Another source of error is the tendency for nearby locations to be alike in terms of the communities they support, a trend known as spatial autocorrelation (SAC). If a regression model cannot explain fully the observed spatial clustering then its residuals exhibit spatial structure, violating the assumption that they should be independent and identically distributed. There are two reasons why this kind of error is common in niche models. First, predictors rarely contain sufficient information to describe fully the observed species aggregation (Guisan and Thuiller, 2005); missing pieces of the puzzle include dispersal patterns, competition/mutualism and disturbance. Second, ecologists are inclined to visit sites in more accessible locations and areas of particular interest, yielding a spatially clustered sampling distribution that may not be representative of SAC in the candidate predictors. Over recent years, the number of ecological studies to address SAC in models has increased, with a majority reporting significant improvements in model fit (Dormann, 2007b). Augustin *et al.* (1996) modelled deer populations using autologistic regression, a form of auto-model (Besag, 1974) that has since been applied to a variety of species distribution model (Miller *et al.*, 2007). In previous application to GAMs, this method for describing localised spatial dependence has performed well (e.g., Segurado and Araujo, 2004); however recent studies warn that autologistic (autocovariate) models may underestimate the environmental controls on a species distribution (Dormann, 2007a; Dormann *et al.*, 2007).

Envelope uncertainty

The breadth of niche-space spanned by the distribution data may not be sufficient to fully represent the study region, particularly if projecting models under climate change scenarios. This is a common problem in the estimation of species distributions, though there are few tools available for estimating the associated uncertainty in predictions (Pearson and Dawson, 2003; Thuiller *et al.*, 2004; Pearson *et al.*, 2006; Dormann, 2007c). For GAMs specifically, model uncertainty arises because response shapes are constructed using non-parametric smoothers – each smoother focuses on a specific portion of the data, and so the modelled response does not naturally extend past the limits of the training data. In essence, the problem is the same for all predictive models: that attempts to predict species occurrence beyond the documented niche-breadth are subject to high uncertainty, particularly if more than one environmental factor is under-represented (Thuiller *et al.*, 2004). At a time when extrapolations into unknown climate-space are increasingly in demand, the development of methodologies to address this issue has been identified as a priority for research (Araujo and Guisan, 2006). The solution we present is to accompany model predictions with envelope uncertainty maps (EUMs), which allow the analyst to identify geographical locations where the profile of environmental conditions at sample sites results in high model uncertainty.

Methods

Study region

The EAMs are part of the Eastern Afromontane Biodiversity Hotspot (Mittermeier *et al.*, 2004) and are defined as those ancient crystalline mountains under the direct climatic influence of the Indian Ocean (Lovett, 1990). Beginning in the Taita Hills of southern Kenya, they extend down through eastern Tanzania to the Udzungwa Mountains in the south (Fig. 3.1). The mountains are a chain of 13 disjoint blocs, isolated from the surrounding lowlands since the Miocene about 30 million years ago (Schlüter, 1997). Today they support 3300-5700 km^2 of moist tropical forest, though it has been estimated that this may be less than 30% of the original forest cover (Burgess *et al.*, 2007b). Much of the remaining area is protected by forest and nature reserves, national parks and community-based management, many covering critical water catchments; the EAMs are a source of drinking water and hydroelectric power for over half of Tanzania's urban population. The archipelago-like distribution of mountain blocs promotes significantly higher range-size rarity than is found in some other high biodiversity tropical ecosystems (Taplin and Lovett, 2003; Burgess *et al.*, 2007a), rendering EAM flora particularly sensitive to further fragmentation. Species richness scores are high and the concentrations of endemism are exceptional (Burgess *et al.*, 2007b), though many hundreds of endemic plants and animals are threatened by extinction. Around 500 vascular plant species are putatively endemic, including over 80 tree species (Lovett *et al.*, 2006). Subject to significant anthropogenic pressure and harbouring such high biodiversity per unit area, the EAMs have been identified as one of Earth's most fragile biodiversity hotspots (Brooks *et al.*, 2002).

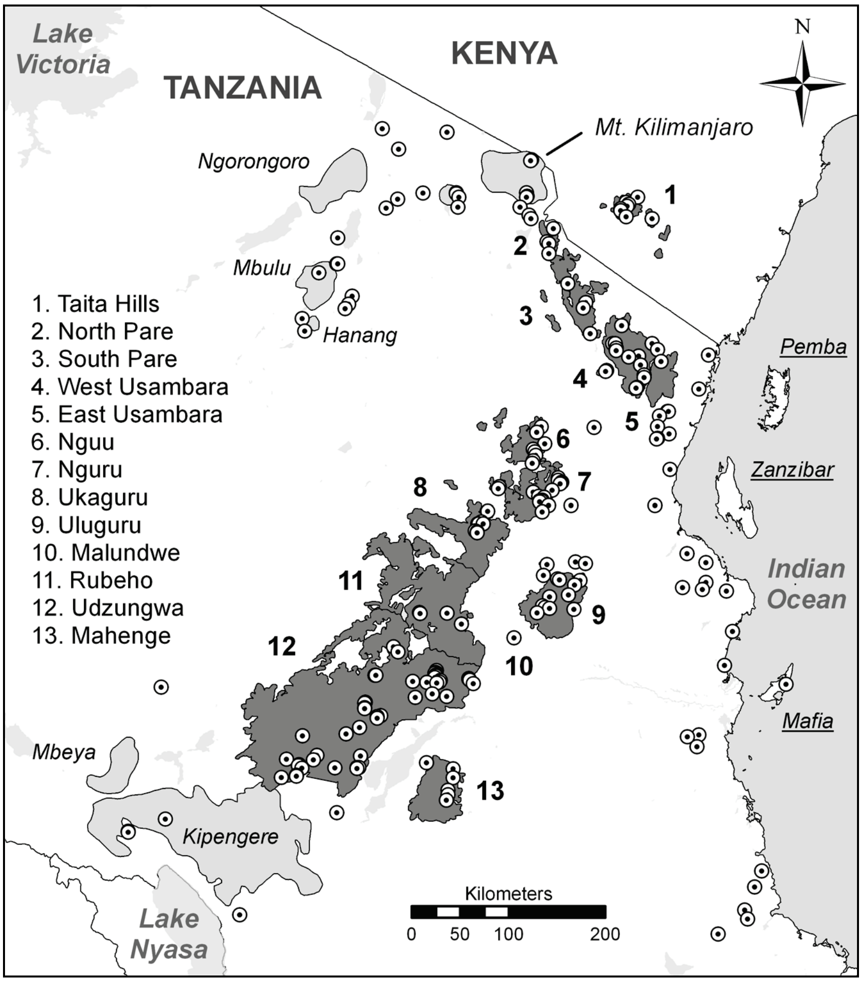


Figure 3.1. Map showing the 13 crystalline blocs that comprise the Eastern Arc Mountain chain. Encircled dots locate the 201 modelling points. Note the clustered distribution of samples – a classic problem in species distribution modelling. Region for model extrapolation was the full map extent: 32.5°E–40.5°E, 1.5°S–10.5°S.

Tree data

The tree database collates observations from 363 variable area plots visited between 1979 and 1994. Since some of our target species' ranges extend beyond the EAMs (e.g., Hemp, 2006), we included plots from other forested mountains such as Mt. Kilimanjaro, and also from the nearby coastal forests (Coastal Forests of Eastern Africa Biodiversity Hotspot; Mittermeier *et al.*, 2004). All plots share a common sampling method, whereby a focal point is chosen and the nearest 20 trees measuring at least 20 cm diameter at breast height are recorded. Lists of trees occurring outside the plots were also included in the database where available. Using field notes of location and altitude, we identified 201 distinct 30 arc-sec (920 m) grid cells across which the samples were collected (Fig. 3.1). The choice of scale is an important consideration for modelling; in the EAMs climatic conditions vary rapidly over short spatial scales, and so we used the finest cell size allowed by the field data. A coarser grid would aggregate more sites, reducing the impact of SAC and mislocation errors in the data, but critically for our study area might omit important changes in habitat across the altitudinal gradient. A species was recorded absent from a grid cell if there was no record of presence in either the plot data or the tree lists. These absences should be considered 'pseudo-absences' because the lists are not exhaustive and the 20-tree method is unlikely to capture all species present at a given site. A full list of the tree species modeled, including their sample prevalence, is presented in Appendix 3A. For further details of the field data we refer the reader to Lovett (1998).

Environmental predictor variables

Climatic and topographic predictor maps were used to estimate the environmental conditions at each site and to extrapolate predictions to the wider Eastern Arc region. Climate surfaces were obtained from the Centre for Resource and Environmental Studies, Australian National University (http://cres.anu.edu.au/). The grids are based on climate station data collected between 1920 and 1980, and provide estimates of mean monthly rainfall and mean daily temperature extremes at a spatial resolution of three arc minutes (5.5 km). To achieve consistency with the 30 arc-sec resolution of the tree data, we interpolated the surfaces using a distance-weighted average of the 16 nearest neighbours. Derived predictors were then calculated to better represent the climatic gradients directly affecting species distributions (Table 3.1). Absolute minimum temperature is as described by Prentice *et al.* (1992), potential evapotranspiration follows the Thornthwaite (1948) method, and annual moisture index is the ratio of mean annual precipitation to potential evapotranspiration. Our temperature-days variable is derived similarly to the growing degree-days measure commonly applied in temperate zones. Its inclusion here provides species-specific information on climatic suitability across all 12 months of the year. Since the phenologies of modelled taxa are not known, we bounded suitable conditions for growth using the presence records: for each species *i*, the upper bound was the maximum value of *tmeanw* across all sites where species *i* occurs; the lower bound was the minimum *tmeanc* (refer to Table 3.1 for abbreviations).

Table 3.1. Summary of environmental predictor variables. Climatic range is high because of Mt. Kilimanjaro (5895 m AMSL).

Predictor	Description	Units	Mean		Range	
			Sites	All cells	Sites	All cells
gradient*	Angle from horizontal	°	7.242	1.533	28.88	64.58
trasp*	Wetness/radiation index	–	0.6720	0.5366	1.000	1.000
trange*	Annual temperature range	°C	15.85	16.07	8.672	13.10
pptdry*	Precipitation driest month	mm	11.95	4.949	54.11	94.80
pptann^1	Mean annual precipitation	cm	107.4	87.14	121.1	194.1
pptwet^1	Precipitation wettest month	mm	229.1	186.4	389.3	437.0
tmean^2	Mean temperature	°C	21.20	22.71	14.01	36.44
tmeanw^2	Mean temp. warmest month	°C	23.23	24.69	14.03	36.45
tmeanc^2	Mean temp. coldest month	°C	18.44	20.23	14.75	36.23
tmaxw^2	Max. temp. warmest month	°C	28.83	30.46	13.58	38.50
tminc^2	Min. temp. coldest month	°C	13.15	14.52	17.56	33.75
tabsmin^2	Absolute minimum temp.	°C	4.486	7.228	22.45	51.94
tdays$^{(2)}$	Temperature-days	°-days	–	–	–	–
pet^2	Potential evapotranspiration	cm/year	105.4	115.1	99.77	223.9
ami^2	Annual moisture index	cm/year	1.093	0.7918	2.041	4.997

* = independent [abs(*r*) < 0.7]; 1 = first correlation group; 2 = second correlation group; () = species-specific

Topographic data were from the Shuttle Radar Topography Mission, U.S. Geological Survey (http://srtm.usgs.gov/). Two predictors were derived from the partial derivatives of elevation (Horn, 1981): gradient of the slope and a transformation of aspect (Table 3.1). The digital elevation model was supplied at a resolution of three arc-sec (92 m); derived predictors were rescaled to 30 arc-sec for compatibility with the tree data. In order to overcome the problems associated with using a circular predictor variable (i.e.,, 0° ≡ 360° on a compass) we used a cosine transformation of aspect to obtain a symmetric wetness/radiation index (Roberts and

Cooper, 1989). Plots of aspect against rainfall showed that on average slopes facing east-south-east receive the most rainfall during the dry season, when moisture carried by the trade winds is most critical, and so these slopes were allocated the highest wetness indices, and west-north-west facing slopes the lowest.

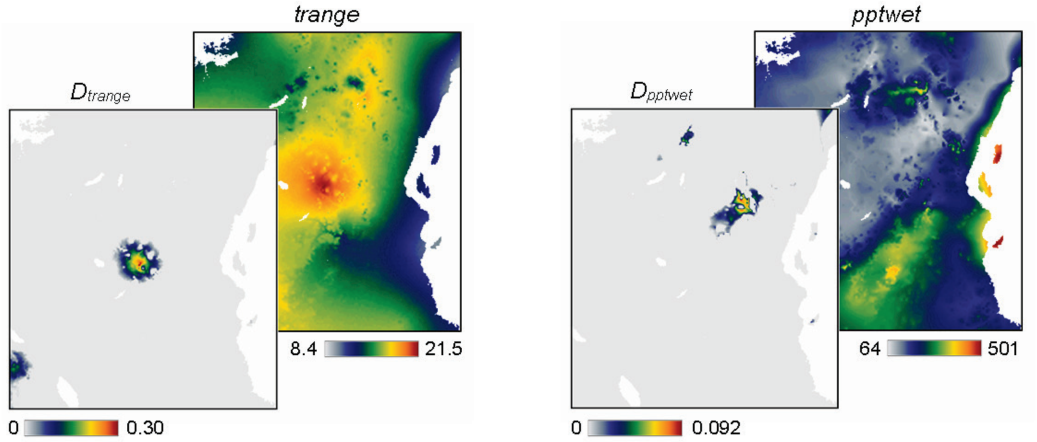


Figure 3.2. Environmental coverage of predictor variables. Background: annual temperature range (left) and precipitation during the wettest month (right). Foreground: proportional distance (D_X) of these predictors from the calibration envelope. Distance maps can be combined in a contribution-weighted average to yield envelope uncertainty maps (EUMs).

In order to evaluate how well the 201 modelling points captured the environmental range of our study region, the proportional 'distance' of each grid cell from the calibration envelope was calculated with respect to each predictor (Fig. 3.2). Envelope uncertainty maps (EUMs) estimate the associated model uncertainty using an average of these distance maps, weighted according to the relative contributions of predictors in a model. Cell *i* is given by

$$\mathrm{EUM}_i = \frac{\sum_{X\in S} C_X D_{X_i}}{\sum_{X\in S} C_X} ,$$

where S is the predictor set, C_X is the contribution of predictor X, and D_{Xi} is the proportional distance of X_i from the calibration envelope:

$$D_{X_i} = \frac{\max\left\{\hat{X}_{\min} - X_i, X_i - \hat{X}_{\max}, 0\right\}}{\hat{X}_{\max} - \hat{X}_{\min}},$$

where $\hat{X}$ denotes the calibration subset. In this paper we define predictor contributions to be the percentage drop in explained deviance when predictor X is removed from the final model. As a rule of thumb, Dormann (2007c) recommends that one should not extrapolate further than 1/10th of the parameter range (i.e., D_X should not exceed 0.1). Particular caution is therefore recommended for grid cells where the EUM > 0.1, since this indicates that at least one predictor has been extrapolated beyond the 1/10th level.

Statistics for calibration and evaluation

Model performance was assessed using the proportion of explained deviance (D^2), area under the receiver-operating characteristic curve (AUC; Green and Swets, 1974) and an associated measure of generalisation error (GE; see below). Predictions of occurrence were on a continuous scale, from zero to one. For direct comparison with the tree data, these were dichotomised by maximising the sum of sensitivity (proportion of presences correctly predicted) and specificity (proportion of absences correctly predicted), a method shown to perform well by Liu *et al*. (2005) in their comparative study. The AUC is a threshold-independent measure, incorporating both type I (false positive) and type II (false negative) error rates, and is largely unaffected by sample prevalence (McPherson *et al.*, 2004). In a recent critique of the AUC (Lobo *et al.*, 2008), the lack of spatial information and validity of symmetric error-weights are questioned – two weaknesses that could be mitigated by the use of EUMs and absence-weights, respectively. For testing the significance of differences between models we used the Wilcoxon rank sum (Mann–Whitney) test, a statistic closely related to the AUC (Pearce and Ferrier, 2000b).

Since occurrence data were too scarce to partition into independent sets for training and testing, we used cross-validation (CV; Stone, 1974) to assess generality. First the data were partitioned into ten disjoint subsets of roughly equal size. The model was then fitted to nine of them and assessed using the withheld fraction as pseudo-independent test data – this step was repeated ten times, each time omitting a different fraction of data. The entire procedure was repeated 20 times and results were averaged to give the final cross-validation index (Kohavi, 1995). To ensure that the subsets of data used to train and test models reflected the true sample prevalence, partitions were stratified such that prevalence was approximately equal between folds (Parker *et al.*, 2007; see also Appendix 3B).

For a particular model, the severity of generalisation error can be gauged by comparison of the cross-validated and resubstituted AUC (subscripted ‘CV’ and ‘RS’), where resubstitution refers to the reuse of training data for testing. For a standardised comparison across all models, we used the following measure.

$$\mathrm{GE} = \frac{\mathrm{AUC_{RS}} - \mathrm{AUC_{CV}}}{\mathrm{AUC_{RS}} - 0.5}\ ,\quad \mathrm{AUC_{CV}} < \mathrm{AUC_{RS}}\ (> 0.5)$$

That is, the proportion of above chance AUC that is lost under cross-validation. Lower values are best: GE ≈ 0 indicates a very stable model, provided that extrapolation sites are within the environmental range of the training data; GE ≈ 1 warns that discriminatory ability at unvisited sites could be no better than that of a null model; GE > 1 only when $\mathrm{AUC_{CV}} < 0.5$ (worse than chance).

Statistical calculations were performed in R 2.3.1 (R-Development-Core-Team, 2009) using functions from the 'gam' and 'ROC' packages, together with custom R code. For spatial analyses we used GRASS GIS 6.0 (GRASS-Development-Team, 2009). The manipulation of map layers and calibration of models were automated using shell scripts and executed in Windows XP *via* a Linux emulation layer (http://cygwin.com/).

Modelling experiments

The 201 sites were located on predictor maps and the corresponding cell values were extracted for model calibration. GAMs were then fitted to the data using a logit link and binomial error term (Yee and Mitchell, 1991). Given that response shapes can vary greatly in natural systems, both between species (Pearce and Ferrier, 2000a) and with respect to different environmental gradients (Austin, 2002), we determined the effective number of parameters for smoothers (degrees of freedom, df) separately for each species-predictor pair, such that df in [1, 4] at intervals of 0.25. Where df = 1 we fitted parametric curves in order to reduce the uncertainty of extrapolating smooth functions (Hastie and Tibshirani, 1990); for higher degrees of freedom the smooth terms were retained. A schematic summary of the modelling procedure is shown in Fig. 3.3; details of the experiments are as follows.

Selecting predictors

All predictor pairs were tested for collinearity using Pear- son's correlation coefficient (*r*). If two predictors were highly collinear [abs(r) > 0.7] then the one that yielded the highest univariate $\mathrm{AUC_{CV}}$ was entered for selection. The motivation for this step was to allow predictors conveying subtly different information (e.g., *tminc* and *tabsmin*) to be available for all species, without excessive overlap in the data – highly correlated data are not parsimonious and may bias selection (Cohen *et al.*, 2003; Graham, 2003). Other studies have used factor analysis to similar effect, reducing the full predictor set to a smaller number of uncorrelated factors (e.g., Bakkenes *et al.*, 2002). We experimented with a range of thresholds before deciding on the appropriate level [abs(r)] at which predictors should be separated. Fixing the threshold at 0.7 was found to create three distinct subsets, such that predictors were either uncorrelated with all others or belonged to one of two mutually exclusive correlation groups (Table 3.1). To dampen sensitivity to weaker correlations, and those too non-linear to be detected by the Pearson coefficient, we cross-validated stepwise procedures (Hastie *et al.*, 2001; Maggini *et al.*, 2006) and avoided hypothesis tests in favour of global measures of model performance (Anderson *et al.*, 2000).

Two stepwise selection procedures were employed to further promote parsimony amongst solutions. The first, forward-backward selection (denoted 'FB'), began with an empty predictor set, sequentially added/removed variables according to the resultant change in AIC, and was complete when AIC ceased to improve. After each selection step the generality of predictions was assessed, and the final model was that which achieved the highest AUC_{CV}. The formula for AIC consists of two terms: the first evaluates model fit using a log-likelihood function; the second is a penalty term proportional to the number of predictors in the model. Its purpose here was to identify a set of candidate models from which the most robust could be selected by cross-validation.

The second method was backward-forward selection (denoted 'BF'). This time the procedure began with a full model and variables were removed/added according to BIC (Bayesian Information Criterion; Akaike, 1978). Again, the final model was determined by AUC_{CV}. BIC was preferred here because it penalises large models more heavily than AIC, encouraging the removal of noise variables and the selection of more parsimonious solutions. A simple multimodel solution (denoted 'MM') was achieved by weighting the two best-model predictions according to their respective above chance AUC_{CV} values (AUC_{CV} − 0.5), and taking the average.

Weighting absences

In baseline models presence records and absence records were treated with equal confidence, assuming no contamination of one class by the other. With weighted models we attempted a more realistic portrayal of the data by placing greater emphasis on observed presences (P)

than on absences inferred from plot data (*A*). This was achieved by the weighting absence data by a factor of *P*/*A* (< 1 for all species), forcing a standardised prevalence of 0.5. The intention was to tilt the balance of errors away from false negatives and toward false positives (McPherson *et al.*, 2004). Such a shift is desirable because a presence observation necessitates suitable conditions for growth, whilst an absence record could be a consequence of the restricted sampling regime, or of ecological factors beyond the scope of the model (Anderson, 2003).

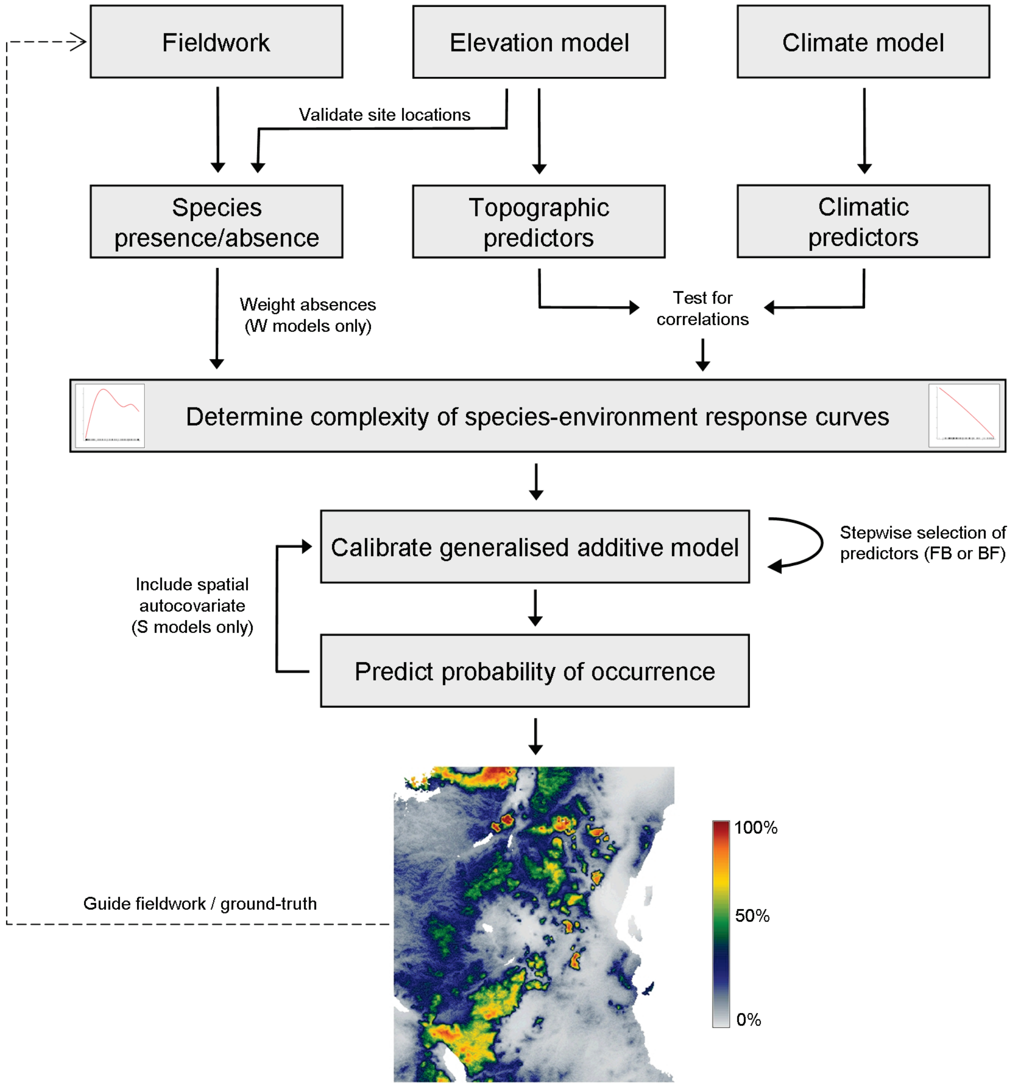


Figure 3.3. Schematic of the modelling procedure.

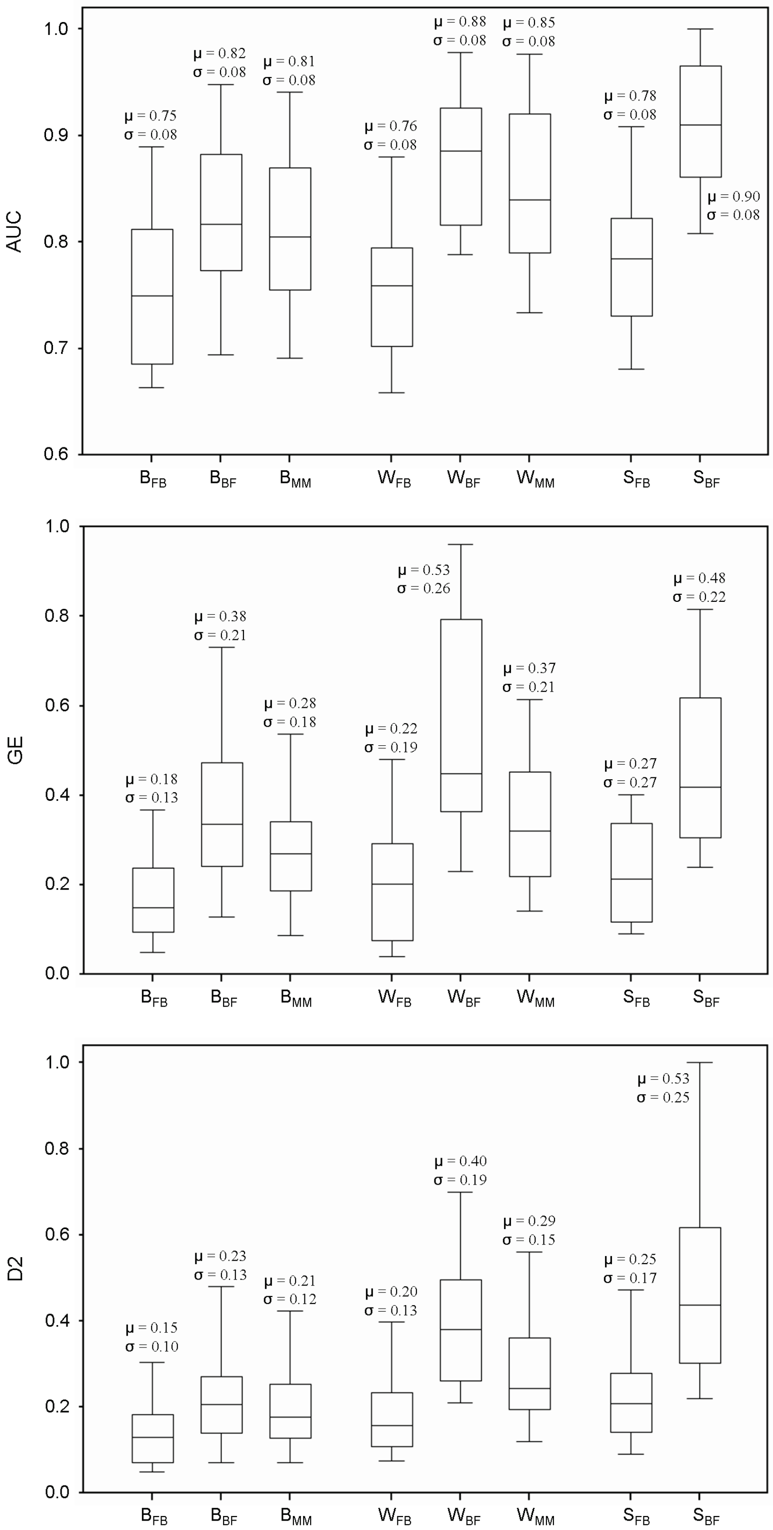


Figure 3.4. Boxplot comparisons of model performance in the different experiments. Box whiskers extend up to 150% of the interquartile range of each box. Upper: area under the receiver-operating characteristic curve (AUC). Middle: generalisation error (GE). Lower: proportion of deviance explained (D^2). Model type: B, baseline; W, weighted; S, spatial. Selection: FB, forward-backward; BF, backward-forward; MM, multimodel.

Spatial autocovariates

Autocovariate terms were used to describe fine-scale spatial clustering in species distributions. The first step was to obtain preliminary estimates of the distributions, for which we used weighted model predictions. Autocovariate terms were then derived such that each grid cell (i) was a distance-weighted average across a set of neighbours (k_i):

$$A_i = \frac{\sum_{j \in k_i} p_j d_{ij}^{-1}}{\sum_{j \in k_i} d_{ij}^{-1}}$$

where p_j is the probability of occurrence in neighbouring cell j, and d_{ij} is the Euclidean distance between i and j (Augustin *et al.*, 1996). Four autocovariates were calculated for each model, with neighbourhoods represented by squares of side 3, 5, 7 and 9 cells (2.8, 4.6, 6.4 and 8.3 km, respectively). The autocovariate included in the final model was that which led to the greatest improvement in explained deviance. Larger neighbourhoods were not included because seeds are typically heavy, limiting wind dispersal. Birds and mammals may carry fruits further, but successful establishment would be fragmented by the rapidly changing landscape and restricted environmental ranges of taxa. Spatial models were not constructed for multimodels because there was no formula to which to append the autocovariate.

Results

Baseline models (B)

Our interpretation of Swets (1988) analysis of the AUC measure is that for ecological studies a value in the range [0.7, 0.9) indicates a reasonable or good model and a value in the range [0.9, 1.0] indicates an excellent model, although any model with AUC > 0.5 should provide some discriminatory power. Following this classification for each of the 40 taxa, 27 forward-backward (FB) models and 36 backward-forward (BF) models produced reasonable, good or excellent predictions. For two species, *Syzygium cordatum* and *Tabernaemontana pachysiphon*, the FB solution was a null model (no predictors were added to the formula). For the same two species, BF returned non-trivial but highly unstable solutions (GE = 0.82 and 0.68, respectively). In general the FB method selected more parsimonious models with fewer predictors and better generalisation error. Conversely, BF models tended to be larger and better equipped to capture the observed deviance, though performance suffered under cross-validation (Table 3.2; Fig. 3.4). These two selection methods agreed for just five species. Baseline multimodels achieved a useful compromise, returning higher precision than FB models but with lower generalisation error than BF models. Moreover, for nearly half the taxa the multimodel produced higher AUC_{CV} than either of the best-model predictions.

Weighted models (W)

Applying weights to the absence data significantly altered the distribution of smoother complexity ($p < 0.01$): for baseline models the distribution was positively skewed, with simpler curves constructed for most predictors; for weighted models, the distribution was shifted in favour of more complex response shapes. This altered the predictor sets chosen by selection, and ultimately resulted in different spatial predictions (Table 3.2; Figs. 4 and 5). For FB selection, predictor sets chosen during weighted and baseline model calibration differed for 27 of the 40 taxa, with seven null models; for BF selection they differed for 26. Inspection of response curves showed that the change in smoother complexity had increased uncertainty, especially near the limits of the training data (Fig. 3.5).

Under FB selection the impact on the AUC was not significant, though other statistics revealed important differences: the mean proportion of errors that were false negatives decreased by 23% compared with baseline models (increased sensitivity), and the mean value of D^2 was significantly higher. Under BF selection, weighted models tended to fit the training data very well – all but one species (*S. cordatum*) achieved reasonable to excellent AUC and the mean value of D^2 was particularly high (Fig. 3.4). As for FB models the error distribution was also much improved, with a 46% reduction in the proportion of errors that were false negatives. BF models remained prone to over-fitting though, a problem that appears to have been exacerbated by the weights. Prediction error was again dampened by model averaging, with the weighted multimodel returning the highest mean AUC under cross-validation.

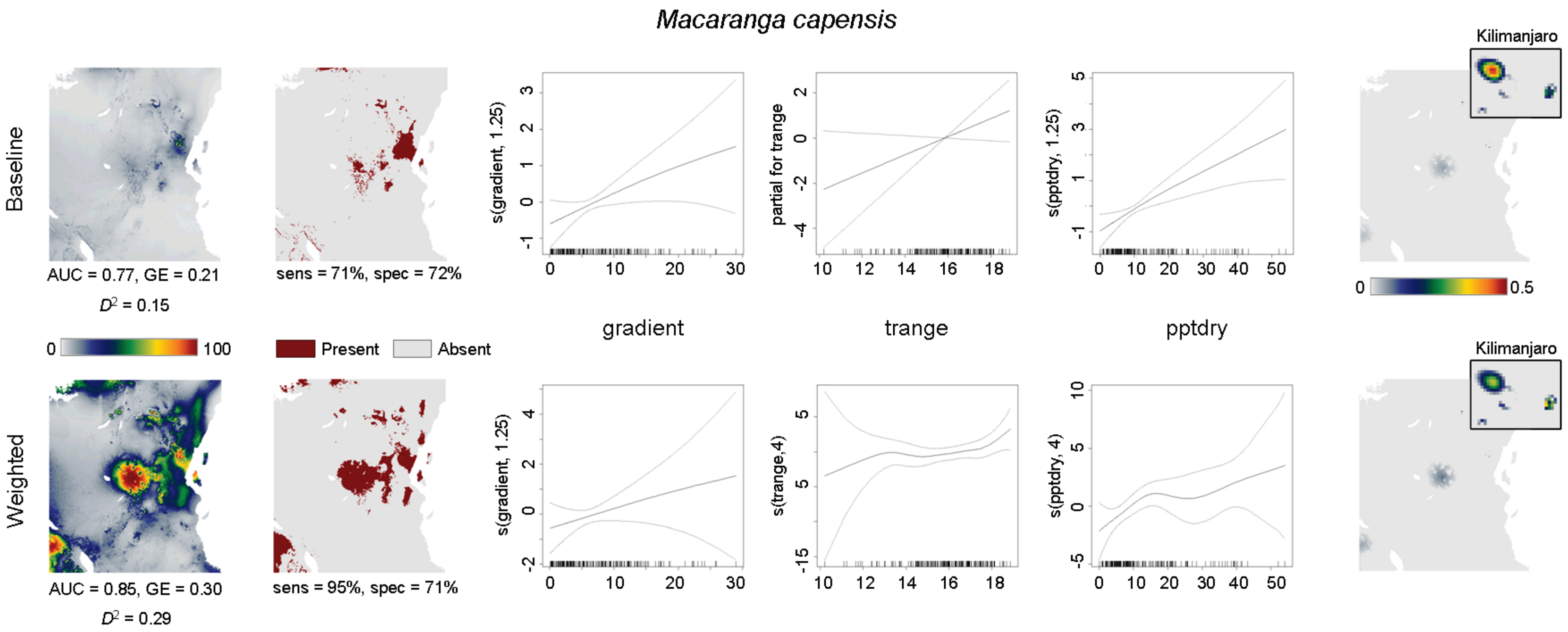


Figure 3.5. Baseline and weighted model predictions for *Macaranga capensis* (forward-backward selection). From left to right: probability of occurrence (%), presence-absence (maximising sum of sensitivity and specificity), response curves for selected predictors (including standard errors) and envelope uncertainty.

Table 3.2. Significance of differences between modelling experiments. With respect to the area under the receiver-operating characteristic curve (AUC), generalisation error (GE) and proportion of explained deviance (D^2). Model type: B, baseline; W, weighted; S, spatial. Selection method: FB, forward-backward; BF, backward-forward; MM, multimodel.

	AUC			GE			D^2		
Baseline selection									
	FB	BF	MM	FB	BF	MM	FB	BF	MM
FB	–	***	***	–	***	***	–	***	***
BF	–	–	ns	–	–	**	–	–	ns
Forward-backward models									
	B	W	S	B	W	S	B	W	S
B	–	ns	ns	–	ns	**	–	*	***
W	–	–	ns	–	–	ns	–	–	ns
Backward-forward models									
	B	W	S	B	W	S	B	W	S
B	–	***	***	–	***	**	–	***	***
W	–	–	ns	–	–	ns	–	–	*
Multimodels									
	B	W		B	W		B	W	
B	–	*		–	**		–	***	

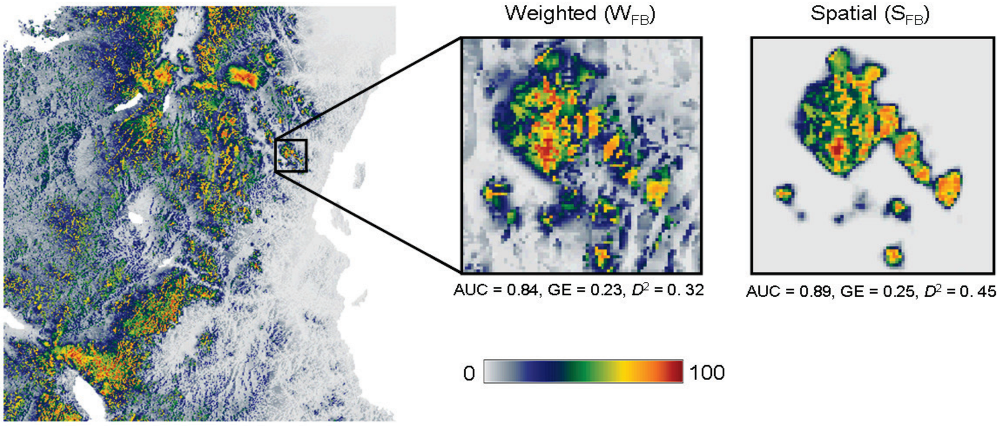
***, $p \leq 0.01$; **, $p \leq 0.05$; *, $p \leq 0.1$; ns, not significant (Wilcoxon rank sum tests, two-sided)

Spatial models (S)

The inclusion of a spatial autocovariate increased the pro- portion of explained deviance in all cases. Spatial models were significantly better at correctly predicting presences and absences (Table 3.3), and for BF selection they were also more stable (Fig. 3.4). Model size was typically larger in BF models and so climatic and topographic constraints were better represented alongside the autocovariate: the mean collective contribution of environmental predictors was 11% in FB models and 24% in BF models; the mean contribution of the autocovariate was 20% and 21%, respectively.

The neighbourhood size chosen for the autocovariate varied between species and no particular scale was superior overall (Appendix 3C). Visual inspection of the predicted distributions showed the environmental preferences of taxa to be more clearly delineated in spatial models: fine- scale patchiness in weighted model predictions had been smoothed, revealing what appear to be more realistic spatial patterns (e.g., Fig. 3.6a). There were some cases however where the incorporation of an autocovariate led to over-fitting. The spatial model for *Syzygium micklethwaitii* yielded a prediction with perfect discriminatory ability but high generalisation error (Fig. 3.6b). In this example the non-spatial model is more useful for inference since it retains a realistic gradient of suitability.

(a) *Neoboutonia macrocalyx*



(b) *Syzygium micklethwaitii*

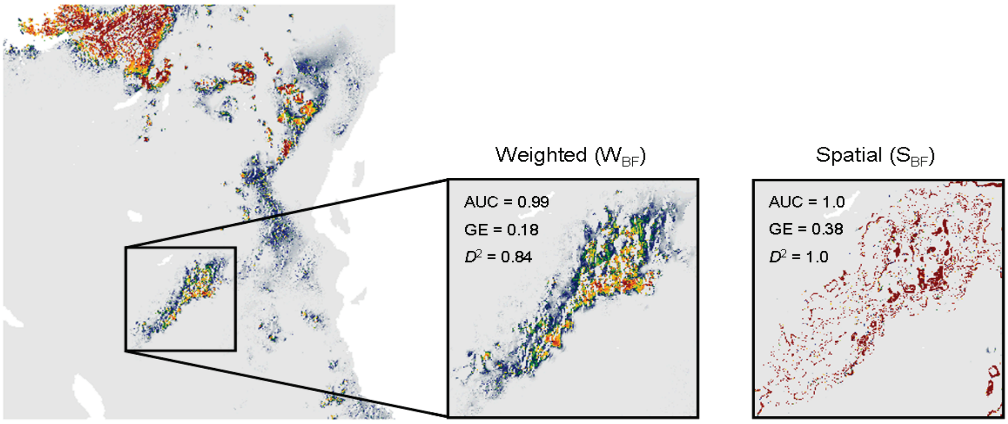


Figure 3.6. Weighted and spatial predictions for (a) *Neoboutonia macrocalyx*, focussing on the Usambara Mountains, and (b) *Syzygium micklethwaitii*, focussing on the Udzungwa Mountains. Scale bar shows probability of occurrence (%).

Table 3.3. Performance of predictors in non-spatial (baseline and weighted) models. Percentage selection rate in group, mean contribution to model, mean contribution of other covariates when selected. Predictor contributions are defined as the percentage drop in explained deviance when removed from the final model, standard errors in parentheses.

Predictor	Selection rate	Contribution (s.e.)	Covariate contribution
Independent			
gradient	–	19.61 (1.44)	18.34
trasp	–	18.75 (1.46)	18.50
trange	–	19.66 (1.47)	18.54
pptdry	–	19.85 (1.45)	19.58
mean	–	19.47 (1.46)	18.74
Group 1			
pptann	49.41	22.30 (2.52)	19.09
pptwet	50.59	18.05 (2.10)	18.14
mean	–	20.18 (2.31)	18.62
Group 2			
tmean	6.35	33.83 (6.12)	27.59
tmeanw	5.56	25.31 (4.88)	21.21
tmeanc	3.17	14.60 (5.16)	28.57
tmaxw	7.14	23.68 (5.28)	20.00
tminc	15.87	18.28 (2.00)	24.10
tabsmin	8.73	18.53 (3.10)	26.19
tdays	7.94	36.09 (6.01)	26.32
pet	26.98	19.50 (2.39)	26.56
ami	18.25	16.15 (2.39)	22.12
mean	–	22.89 (4.15)	24.74

Total times selected: Independent, 343; Group 1, 85; Group 2, 126

Predictors and envelope uncertainty

Both topographic variables were independent of correlation groups, as were *trange* and *pptdry*. These four predictors were the most frequently selected for inclusion in the final model, and each contributed similarly to model performance. The most popular predictor overall was *pptdry*. In the first correlation group, *pptann* and *pptwet* were chosen a similar number of times, though *pptann* was marginally better at explaining deviance. In the second correlation group *tmeanc* contributed the least to D^2 and was the least frequently selected, often dropped in favour of similar measures such as *tminc* and *tabsmin*. Predictors that describe moisture availability, such as *pet* and *ami*, were often included but their mean contributions were below average. The most successful descriptor of deviance was the temperature-days variable, which when removed from the final models resulted in a mean drop in D^2 of 36% (Table 3.3).

The environmental range of survey sites was generally good, with the exception that climatic predictors lacked coverage near the summit of Mt. Kilimanjaro (Table 3.1; Fig. 3.5). The only notable shortfalls were for *trange* and *pptwet*: both were within the calibration envelope for most of the study region, but *trange* was up to 30% beyond the envelope near the Maasai Steppe, and *pptwet* was up to 9.2% beyond the envelope for a small area south of the Pare Mountains (Fig. 3.2).

Discussion

The potential of GAMs to estimate the distribution of EAM trees shows promise, with a number of models achieving a high level of predictive success. However, it is clear from our experiments that the distributions predicted are highly sensitive to the modelling method employed. Selection procedures frequently disagreed, produced different spatial predictions, and yet often returned similar validation scores. These findings illustrate the importance of understanding the biases imposed by the selection procedure in use, and of not relying solely on validation scores as evidence of good model performance – consideration should also be given to the chosen predictor set and spatial patterns predicted. Whittingham *et al*. (2006) advise against the use of stepwise procedures, arguing that there is rarely a true ‘best model’ for selection to identify and that different predictor sets are likely to explain the response equally well. This conjuncture is supported by our experiments, though we suggest that studies with access to more extensive distribution data are likely to find greater agreement between selection methods.

Whilst forward-backward models often lacked precision, particularly if neither topographic predictor was selected, they invariably produced stable predictions using minimal predictor sets, and are therefore likely to be more useful than backward-forward methods for inferring causal relation- ships. Backward-forward selection described the data well but retained too many predictors to avoid over-fitting. Given the disagreement between selection procedures, there is a good argument for favouring expert opinion over computer selection (but see:

Pearce *et al.*, 2001; Seoane *et al.*, 2005). Regardless, automated procedures remain necessary when deriving models for a large number of taxa, particularly if their ecologies are not well known. The multimodel, averaging forward-backward and backward-forward predictions according to their relative cross-validated performance, identified a superior trade-off between generality and precision that in many cases outperformed both conventional selection procedures (higher AUC_{CV}). The weakness of this method is the need to compute two sets of predictions, increasing computation time. However, if models are lacking, either in fit or stability, we think it prudent to investigate other selection options as a matter of course, in which case the computation of a model average would be trivial. Other kinds of multimodel have also produced favourable results (Anderson *et al.*, 2000; Johnson and Omland, 2004; Hartley *et al.*, 2006; Dormann *et al.*, 2008), and appear to be generally superior to best-model approaches for predictive purposes.

Both the performance and reliability of models were correlated with sample prevalence, such that low prevalence led to more discriminative but less stable models. This may reflect the fact that restricted range tree species in the EAMs typically cover a narrow altitudinal range (Lovett, 1996; Lovett *et al.*, 2001), making their climatic preferences easier for models to capture but rendering them highly sensitive to errors in the distribution data. The relationship was found to be stronger for weighted models than for baseline models due to the higher levels over-fitting. In contrast to our results, Maggini *et al.* (2006) found that weighting absences improved model performance without impairing stability. A probable reason for the difference in our results is the comparatively high instability of our baseline models: the more robust baseline predictions tended to remain stable in the weighted experiment, whereas those with high GE suffered from further over-fitting. We did observe a useful shift in the error distribution, such that weighted model predictions were much less likely to contain false negative errors. During recent field expeditions to Nguu and North Pare (Fig. 3.1), we found that the higher sensitivity of weighted models gave a better indication of the actual forest distributions, especially for spatial models. Further ground-truthing of this result is encouraged.

In the spatial experiment we aimed to improve weighted model predictions by including an autocovariate to account for fine-scale spatial clustering. In agreement with previous applications of this technique (e.g., Augustin *et al.*, 1996; Segurado and Araújo, 2004), spatial models fitted the training data more accurately than non-spatial models and were superior for describing fine-scale patterns in distribution. Where model size was large (five or six predictors) spatial models also had lower generalisation error. For smaller model sizes though the contribution of autocovariates in explaining deviance was around twice that of environmental predictors, which may be a cause for concern given recent suggestions that autocovariates can lead to biased predictions (Dormann, 2007a; Dormann *et al.*, 2007). One should certainly be sceptical of extrapolations into different points in time (e.g., historical reconstruction or climate change studies) since spatial dependencies could well be different (Guisan and Thuiller, 2005). The degree to which predictions of this kind can be truly representative of the actual distributions will always be uncertain, because we cannot be sure to what extent a species realised ecological niche is restricted by its environmental tolerance and to what extent by competition/mutualism with other species (Pulliam, 2000). Community interactions are expected to play an important role in such an ancient ecosystem, though the relevant spatial scales are not well understood. In New Zealand's old-growth forests, attempts have been made to model competition between tree species using logistic regression: Leathwick and Austin (2001) found that including the presence/absence of one species as a covariate alongside climatic constraints could improve the predicted distribution of another. This study was based on community compositional data for just two competing species; here we are concerned with a web of interactions involving tens, possibly hundreds of taxa, presenting a far greater challenge for modelling.

The optimum neighbourhood size for a particular tree often differed according to the selection procedure used to obtain the initial prediction, and so did not provide much insight into the processes underpinning SAC. This might simply reflect the high variability in predictor sets chosen by selection and the associated omission of different autocorrelates (Lichstein *et al.*, 2002). It may also be the case that no single range-specific autocovariate could carry sufficient information to identify the true scales at which aggregation occurs (van Teeffelen and Ovaskainen, 2007). Dispersal is one factor known to drive spatial patterns, but this mechanism is under-researched in the EAMs and few empirical data are currently available for parameterising/validating models. We also draw attention to the fact that clustering was assumed by autocovariates to be roughly isotropic, i.e.,, apparent in equal measure in all directions (neighbourhoods were approximated by squares). However, spatial patterns may actually be elongated in some areas as a result of elevational migration. It was further assumed that SAC operates similarly in different regions, which is unlikely to be the case given the high topographic heterogeneity of the study area. The possibility that regression parameters could reflect local rather than global trends has been investigated by Fotheringham *et al.* (2002) and appears to work well in some settings (e.g., Brunsdon *et al.*, 2007), though we have reservations as to the suitability of geographically weighted regression for our dataset (*cf.* Austin, 2007).

Dichotomising probabilities of occurrence using the sensitivity-specificity threshold, we compared the different areas of occupancy forecast by models (Appendix 3D). On average, the number of cells predicted present was similar across the baseline, weighted and spatial experiments, despite often large differences in occurrence probabilities (e.g., Fig. 3.5). In the selection experiment, forward-backward models were the least well constrained by environmental variables, resulting in the greatest areas of predicted occupancy. Interestingly, the number of cells predicted present by the multimodels was similar to that forecast by backward-forward models. Multimodels contained the most information and also gave the highest AUC under cross-validation, and so we are inclined to trust the magnitude of backward-forward area pre- dictions more than those yielded by the forward-backward method.

The frequency with which moisture related variables were selected by models is not surprising given that the EAMs are under the direct climatic influence of the Indian Ocean (Lovett, 1990; Marchant *et al.*, 2007). The gradient of the slope, in addition to indicating moisture availability *via* its relationship with run-off, also helps to distinguish montane habitats from the surrounding lowlands. The wetness/radiation index performed well, reflecting the importance of moisture carried by trade winds during the drier months. Response curves constructed for climatic predictors were not calibrated with respect to the environmental extremes found near the summit of Mt. Kilimanjaro, and so predictions of occurrence in these grid cells are subject to high uncertainty. The sample sites were otherwise found to cover a wide breadth of environmental conditions and were generally representative of the Eastern Arc region. Where extrapolations beyond the envelope occurred, the worst affected models were those that relied heavily on the predictor contributions of annual temperature range and rainfall during the wettest month. The shortfall in these predictors had the greatest impact on forward-backward models, where model size was smaller. For *Macaranga capensis* we found that the highest occurrence probabilities were obtained by extrapolating beyond the range of the training data. The weighted model in particular predicts that this pioneer tree, usually associated with submontane and riverine forests (Lovett *et al.*, 2006), should also be suited to the Maasai Steppe, a lowland savannah habitat. The EUM confirms that the grid cells with the highest envelope uncertainty correspond precisely with the region deemed most favourable by the model. Here inference can only be made after inspecting response curves beyond the limits of the training data. Because EUMs pinpoint the locations where a model may be weakest, we suggest they might also be useful in targeting field sampling in a way that most improves data quality.

Conclusions

GAMs can provide useful information for conservation in the EAMs, even when the frequency of documented occurrence is low. Indeed it was the modelling method employed and the quality not quantity of distribution data that mattered most. However, there were a number of instances where over-fitting seriously compromised the generality of predictions, and we recommend that the application of GAMs to small datasets be approached with care. If over-fitting cannot be avoided, then the parametric terms of generalised linear models should be considered in preference to data-driven smoothers. With respect to the different methodologies investigated, our main observations are as follows.

1. Forward-backward selection is less discriminative than backward-forward selection, but is the more useful of the two for explanatory purposes. Backward-forward selection retains more ecologically relevant detail but can suffer from high prediction error. Multimodels provide a useful compromise, and are arguably the best choice for predictive purposes.
2. Models calibrated with weighted absence data are superior in terms of overall accuracy and have better sensitivity, though they can be especially vulnerable to over-fitting if the distribution data are not well described by environmental predictors.
3. Including a spatial autocovariate improves model fit and better represents spatial clustering in predictions; the stability of models may however suffer if environmental constraints are inadequately represented.
4. Envelope uncertainty maps display important information that should be taken into account when drawing inference from predictions, especially if a model is to be extrapolated into novel parameter space.

This work involved the parameterisation of environmental response functions for 40 taxa of large tree, targeted for modelling because of historical patterns occurrence (Mumbi *et al.*, 2008), endemism (Lovett *et al.*, 2006) and conservation interest (http://iucnredlist.org/). It is hoped that further analysis of response shapes will add to our understanding of their habitat preferences, and specifically the degree to which environmental controls restrict their distributions.

Acknowledgements

This research was funded by the Marie-Curie Excellence programme of the European 6th Framework under contract MEXT-CT-2004-517098 to R. Marchant. We are grateful to A. Ahrends, M. Bulling, J. Finch and A. Jump for commenting on previous versions of this manuscript, and for the helpful feedback of two anonymous reviewers.

Author contributions

The study was conceived by P.J.P., who programmed and implemented the modelling experiments, analysed the results and prepared the manuscript. Plot data was provided by J.C.L. The work was supervised by R.M. and C.J.M.

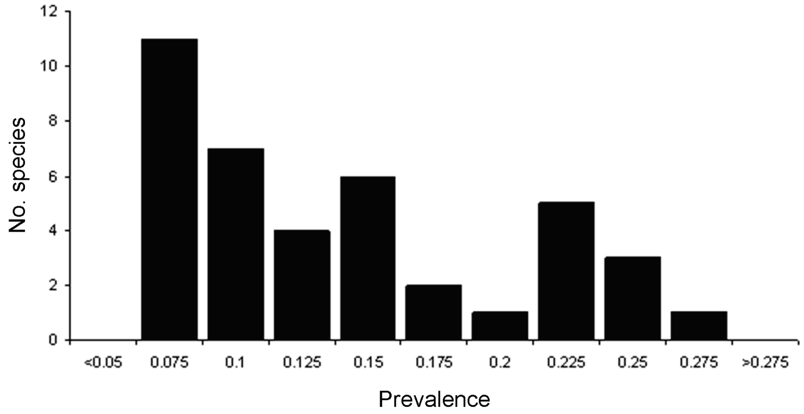
Appendix 3A. Tree species modelled

Species code	Family	Species/taxon name	Presences (prevalence)
sp1	Anacardiaceae	*Sorindeia madagascariensis* Thouars ex DC.	41 (0.20)
sp2	Annonaceae	*Lettowianthus stellatus* Diels	11 (0.05)
sp3	Apocynaceae	*Tabernaemontana pachysiphon* Stapf	21 (0.10)
sp4	Aquifoliaceae	*Ilex mitis* (L.) Radlk.	18 (0.09)
sp5	Araliaceae	*Polyscias stuhlmannii* Harms	15 (0.07)
sp6	Araliaceae	*Polyscias fulva* (Hiern) Harms	41 (0.20)
sp7	Celastraceae	*Maytenus acuminata* (L.f.) Loes.	29 (0.14)
sp8	Chrysobalanaceae	*Parinari excelsa* Sabine	41 (0.20)
sp9	Combretaceae	*Terminalia sambesiaca* Engl. & Diels	15 (0.07)
sp10	Euphorbiaceae	*Croton macrostachyus* Hochst. ex Delile	17 (0.08)
sp11	Euphorbiaceae	*Drypetes gerrardii* Hutch.	28 (0.14)
sp12	Euphorbiaceae	*Drypetes natalensis* (Harv.) Hutch.	11 (0.05)
sp13	Euphorbiaceae	*Drypetes usambarica* (Pax) Hutch.	16 (0.08)
sp14	Euphorbiaceae	*Macaranga capensis* (Baill.) Benth. ex Sim var. *capensis*	21 (0.10)
sp15	Euphorbiaceae	*Macaranga capensis* (Baill.) Benth. ex Sim var. *kilimandscharica* (Pax) Friis & M. G. Gilbert	41 (0.20)
sp16	Euphorbiaceae	*Neoboutonia macrocalyx* Pax	17 (0.08)
sp17	Lauraceae	*Ocotea usambarensis* Engl.	41 (0.20)
sp18	Leguminosae (Fabaceae): Mimosoideae	*Newtonia buchananii* (Baker f.) G. C. C. Gilbert & Boutique	48 (0.24)
sp19	Loganiaceae	*Anthocleista grandiflora* Gilg	21 (0.10)
sp20	Loganiaceae	*Nuxia congesta* R. Br. ex Fresen.	40 (0.20)
sp21	Meliaceae	*Trichilia emetica* Vahl.	10 (0.05)
sp22	Monimiaceae	*Xymalos monospora* (Harv.) Warb.	46 (0.23)
sp23	Myricaceae	*Morella salicifolia* (Hochst. ex A. Rich.) Verdc. & Polhill	28 (0.14)
sp24	Myrsinaceae	*Myrsine melanophloeos* (L.) R.Br.	51 (0.25)
sp25	Myrtaceae	*Syzygium guineense* (Willd.) DC. subsp. *afromontanum* F. White	47 (0.23)
sp26	Myrtaceae	*Syzygium micklethwaitii* Verdc.	14 (0.07)
sp27	Myrtaceae	*Syzygium cordatum* Hochst.	13 (0.06)
sp28	Oleaceae	*Olea capensis* (L.) subsp. *macrocarpa* (C. H. Wright) I. Verd.	29 (0.14)
sp29	Oleaceae	*Olea europea* (L.) subsp. *cuspidata* (Wall. ex G. Don) Cif.	12 (0.06)
sp30	Palmae (Arecaceae)	*Phoenix reclinata* Jacq.	27 (0.13)
sp31	Podocarpaceae	*Afrocarpus falcatus* (Thunb.) C. N. Page	12 (0.06)
sp32	Podocarpaceae	*Podocarpus latifolius* (Thunb.) R.Br. ex Mirb.	35 (0.17)
sp33	Rosaceae	*Hagenia abyssinica* J.F. Gmel.	18 (0.09)
sp34	Rosaceae	*Prunus africana* (Hook.f.) Kalkman	29 (0.14)
sp35	Sapindaceae	*Dodonea viscosa* Jacq.	14 (0.07)
sp36	Sapotaceae	*Chrysophyllum gorungosanum* Engl.	33 (0.16)
sp37	Sterculiaceae	*Dombeya torrida* (J.F. Gmel.) Bamps	18 (0.09)
sp38	Sterculiaceae	*Leptonychia usambarensis* K. Schum.	19 (0.09)
sp39	Ulmaceae	*Celtis africana* Burm.f.	11 (0.05)
sp40	Ulmaceae	*Trema orientalis* (L.) Blume	21 (0.10)

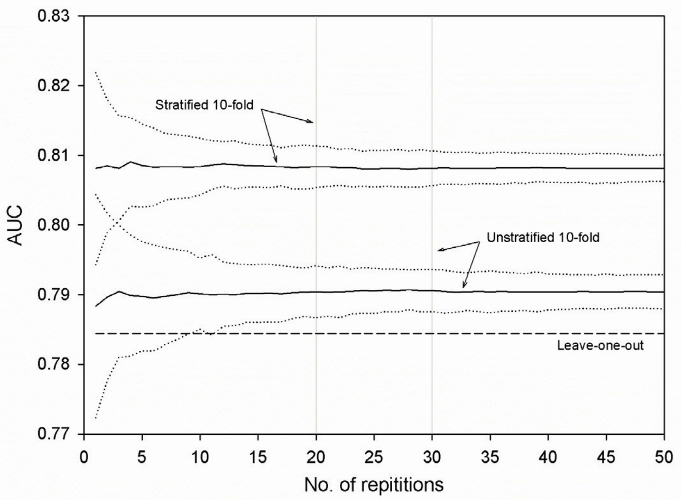
Nomenclature follows Lovett *et al*. (2006).

Appendix 3B. Stratified cross-validation

Given the low number of presences for many taxa and the use of cross-validation during model calibration, it was important to ensure that the subsets of data used to train and test models reflected the true sample prevalence. Stratifying cross-validation such that prevalence was consistent between partitions reduced both variance and pessimistic bias in the estimation of the AUC. The ten-fold method was preferred to leave-one-out cross-validation (LOOCV) because it provides a more robust measure of model stability, particularly when repeated multiple times. Our tests with LOOCV showed that for baseline models it was also marginally more pessimistic than ten-fold, despite of the greater proportion of data available for training (Appendix 3D). For weighted models the values of AUC_{CV} and AUC_{LOOCV} were closer, suggesting that down-weighting absences might reduce stratification bias.



Above. Histogram of prevalence across the 201 sample locations. The lowest prevalence permitted by our modelling procedure was 0.05 (10 presences), because stratified ten-fold cross-validation requires that each fold contain at least one presence record.



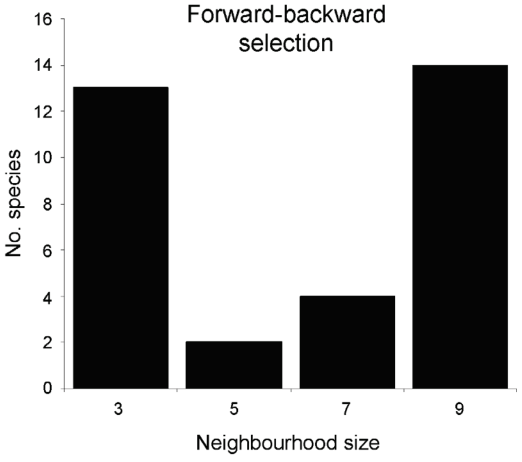
Left. Comparison of cross-validation methods, using *Drypetes natalensis* as an example. How many repetitions are required for consistent results? Averaged over 100 runs for stratified and unstratified cross-validation, results show mean AUC values ± ten standard errors. Dashed line shows the result from leave-one-out cross-validation (*N*-fold). Vertical lines indicate the number of repetitions required for s.e. < 0.0003.

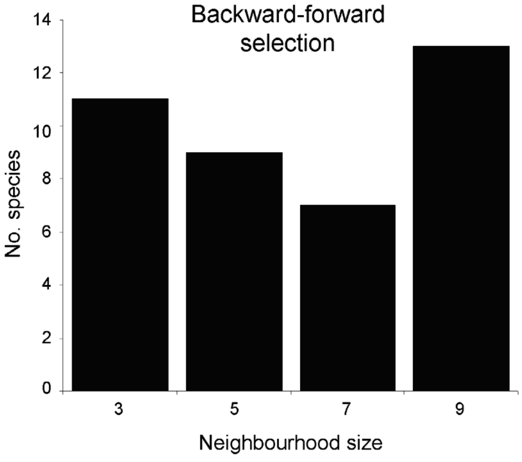
Appendix 3C. Neighbourhood size for spatial autocovariates

Neighbourhood sizes that minimised explained deviance in the spatial experiment (width of neighbourhood square, number of cells). Refer to Appendix 3A for species names.

Species code	Neighbourhood size		Species code	Neighbourhood size	
	Forward-backward	Backward-forward		Forward-backward	Backward-forward
sp1	5	9	sp21	9	9
sp2	3	3	sp22	9	9
sp3	5	9	sp23	9	3
sp4	–	5	sp24	3	3
sp5	9	7	sp25	3	5
sp6	3	5	sp26	–	3
sp7	7	9	sp27	9	3
sp8	9	9	sp28	7	5
sp9	3	9	sp29	–	3
sp10	9	5	sp30	3	9
sp11	9	5	sp31	3	7
sp12	–	7	sp32	3	3
sp13	3	3	sp33	–	5
sp14	3	5	sp34	7	9
sp15	9	9	sp35	9	7
sp16	9	9	sp36	–	7
sp17	7	3	sp37	3	7
sp18	9	5	sp38	–	9
sp19	9	7	sp39	3	3
sp20	3	3	sp40	9	9

Forward-backward: mean = 6.16, median = 7
Backward-forward: mean = 6.10, median = 6





Appendix 3D. Species-specific results

Results from modelling experiments, detailed by species (see Appendix 3 for species names). From left to right: area under the receiver-operating characteristic curve (AUC_{RS}, resubstituted training data; AUC_{CV}, stratified 10-fold cross-validation; AUC_{LOOCV}, leave-one-out cross-validation); generalisation error (GE); explained deviance (D^2); presence/absence cut-point; sensitivity and specificity at this cut-point; number of grid cells predicted suitable for the species.

Baseline models: forward-backward selection

Species code	AUC_{RS}	AUC_{CV}	AUC_{LOOCV}	GE	D^2	Cut-point	Sensitivity	Specificity	Suitable cells
sp1	0.74	0.67	0.66	0.29	0.12	0.11	0.81	0.60	21394
sp2	0.72	0.65	0.63	0.30	0.07	0.05	0.82	0.61	421148
sp3	0.67	0.63	0.61	0.24	0.06	0.16	0.79	0.51	223438
sp4	0.68	0.61	0.58	0.40	0.06	0.09	0.71	0.64	654270
sp5	0.69	0.67	0.64	0.10	0.07	0.08	0.71	0.59	681076
sp6	0.67	0.64	0.61	0.20	0.06	0.08	0.83	0.51	250372
sp7	0.80	0.75	0.75	0.17	0.18	0.13	0.86	0.61	86333
sp8	0.84	0.81	0.78	0.09	0.20	0.09	0.91	0.81	170296
sp9	0.71	0.58	0.55	0.62	0.11	0.10	0.44	0.89	5435
sp10	0.92	0.86	0.86	0.15	0.43	0.14	0.89	0.92	293153
sp11	0.68	0.65	0.62	0.16	0.05	0.11	0.56	0.73	52242
sp12	0.81	0.70	0.68	0.36	0.17	0.10	0.79	0.70	62036
sp13	0.88	0.83	0.82	0.15	0.25	0.05	0.91	0.80	502100
sp14	0.77	0.71	0.70	0.21	0.15	0.10	0.71	0.72	34261
sp15	0.75	0.74	0.73	0.03	0.14	0.10	0.98	0.39	568572
sp16	0.66	0.65	0.63	0.08	0.07	0.10	0.93	0.37	565894
sp17	0.76	0.71	0.70	0.16	0.13	0.10	0.96	0.47	336861
sp18	0.69	0.68	0.67	0.05	0.09	0.22	0.84	0.50	425298
sp19	0.83	0.80	0.79	0.10	0.21	0.07	0.94	0.67	132737
sp20	0.75	0.70	0.69	0.21	0.14	0.22	0.85	0.61	32667
sp21	0.78	0.74	0.74	0.12	0.17	0.24	0.80	0.71	171969
sp22	0.82	0.79	0.79	0.09	0.25	0.16	0.88	0.63	94454
sp23	0.76	0.72	0.72	0.14	0.16	0.25	0.55	0.90	205934
sp24	0.89	0.88	0.88	0.03	0.30	0.15	0.83	0.91	295372
sp25	0.71	0.67	0.66	0.20	0.11	0.28	0.54	0.86	10426
sp26	0.71	0.61	0.58	0.46	0.07	0.15	0.78	0.61	439504
sp27	0.67	0.63	0.59	0.24	0.05	0.05	0.83	0.47	345107
sp28	0.78	0.71	0.70	0.25	0.17	0.20	0.74	0.67	157091
sp29	0.91	0.86	0.86	0.13	0.43	0.22	0.80	0.92	27642
sp30	0.78	0.73	0.72	0.17	0.13	0.30	0.56	0.89	217405
sp31	0.82	0.78	0.77	0.13	0.23	0.11	0.90	0.61	448853
sp32	0.75	0.74	0.73	0.03	0.13	0.18	0.83	0.60	550821
sp33	–	–	–	–	–	–	–	–	–
sp34	0.65	0.64	0.63	0.08	0.05	0.20	0.77	0.48	342629
sp35	0.89	0.84	0.84	0.12	0.33	0.09	0.86	0.82	24942
sp36	–	–	–	–	–	–	–	–	–
sp37	0.74	0.72	0.70	0.11	0.11	0.05	1.00	0.45	680240
sp38	0.60	0.57	0.52	0.32	0.02	0.09	0.81	0.36	273575
sp39	0.65	0.64	0.64	0.05	0.07	0.10	0.40	0.93	107119
sp40	0.72	0.70	0.69	0.10	0.11	0.17	0.89	0.47	355242
mean	0.75	0.71	0.70	0.18	0.15	0.14	0.79	0.66	270208
median	0.75	0.71	0.69	0.15	0.13	0.11	0.82	0.62	236905

Baseline models: backward-forward selection

Species code	AUC_{RS}	AUC_{CV}	AUC_{LOOCV}	GE	D^2	Cut-point	Sensitivity	Specificity	Suitable cells
sp1	0.84	0.71	0.70	0.39	0.22	0.08	0.90	0.63	43052
sp2	0.72	0.65	0.63	0.30	0.07	0.05	0.82	0.61	421148
sp3	0.69	0.63	0.61	0.32	0.07	0.18	0.73	0.61	178928
sp4	0.80	0.59	0.58	0.69	0.15	0.11	0.71	0.78	403058
sp5	0.69	0.67	0.64	0.13	0.07	0.08	0.71	0.59	681076
sp6	0.78	0.58	0.53	0.74	0.14	0.10	0.89	0.62	365584
sp7	0.84	0.73	0.72	0.34	0.23	0.15	0.82	0.72	63670
sp8	0.84	0.81	0.78	0.09	0.20	0.09	0.91	0.81	170296
sp9	0.81	0.53	0.51	0.90	0.22	0.06	0.88	0.59	145052
sp10	0.95	0.85	0.85	0.23	0.51	0.17	0.89	0.94	288678
sp11	0.92	0.79	0.78	0.31	0.37	0.12	0.94	0.80	51715
sp12	0.88	0.69	0.68	0.51	0.27	0.10	0.95	0.72	80009
sp13	0.88	0.82	0.81	0.16	0.26	0.05	0.91	0.77	472874
sp14	0.88	0.74	0.72	0.38	0.30	0.14	0.81	0.84	47054
sp15	0.78	0.69	0.68	0.33	0.19	0.17	0.88	0.57	274456
sp16	0.76	0.62	0.61	0.53	0.13	0.18	0.66	0.76	46086
sp17	0.82	0.71	0.70	0.34	0.20	0.17	0.79	0.73	104347
sp18	0.78	0.69	0.68	0.32	0.20	0.28	0.78	0.64	143382
sp19	0.90	0.79	0.79	0.26	0.31	0.11	0.94	0.79	56596
sp20	0.80	0.69	0.69	0.37	0.19	0.21	0.85	0.65	70541
sp21	0.82	0.73	0.72	0.28	0.23	0.24	0.83	0.72	94813
sp22	0.89	0.79	0.79	0.25	0.36	0.18	0.93	0.76	85960
sp23	0.82	0.73	0.73	0.28	0.24	0.31	0.55	0.94	138977
sp24	0.95	0.83	0.83	0.26	0.49	0.22	0.92	0.96	265119
sp25	0.78	0.66	0.66	0.42	0.18	0.24	0.66	0.79	25125
sp26	0.83	0.66	0.66	0.51	0.21	0.13	0.85	0.68	99850
sp27	0.95	0.75	0.74	0.44	0.43	0.16	0.92	0.91	15136
sp28	0.84	0.69	0.69	0.44	0.24	0.14	0.94	0.60	223287
sp29	0.97	0.86	0.87	0.24	0.59	0.10	1.00	0.87	33842
sp30	0.82	0.70	0.69	0.36	0.20	0.23	0.85	0.68	251412
sp31	0.85	0.77	0.77	0.23	0.27	0.11	0.93	0.63	448802
sp32	0.81	0.74	0.73	0.23	0.21	0.14	0.93	0.59	301476
sp33	0.76	0.55	0.52	0.82	0.13	0.06	0.85	0.59	200723
sp34	0.65	0.64	0.63	0.09	0.05	0.20	0.77	0.48	342629
sp35	0.97	0.90	0.90	0.14	0.55	0.12	0.93	0.90	53470
sp36	0.73	0.57	0.56	0.68	0.13	0.11	0.71	0.68	248870
sp37	0.81	0.66	0.65	0.48	0.18	0.06	0.93	0.58	369912
sp38	0.77	0.55	0.53	0.83	0.13	0.12	0.86	0.73	186184
sp39	0.65	0.65	0.64	0.04	0.07	0.10	0.40	0.93	107119
sp40	0.76	0.67	0.67	0.34	0.15	0.29	0.65	0.77	94322
mean	0.82	0.70	0.69	0.38	0.23	0.15	0.83	0.72	192366
median	0.82	0.69	0.69	0.34	0.21	0.14	0.86	0.72	144217

Baseline multimodels

Species code	AUC_{RS}	AUC_{CV}	AUC_{LOOCV}	GE	D^2	Cut-point	Sensitivity	Specificity	Suitable cells
sp1	0.81	0.70	0.69	0.36	0.19	0.12	0.76	0.73	17777
sp2	0.72	0.65	0.63	0.31	0.07	0.05	0.82	0.61	421148
sp3	0.68	0.63	0.61	0.28	0.07	0.15	0.82	0.49	258707
sp4	0.76	0.61	0.57	0.60	0.11	0.09	0.82	0.66	629437
sp5	0.69	0.67	0.64	0.09	0.07	0.08	0.71	0.59	681076
sp6	0.74	0.61	0.58	0.53	0.09	0.08	0.94	0.56	353643
sp7	0.83	0.74	0.73	0.27	0.21	0.14	0.86	0.68	70896
sp8	0.84	0.81	0.78	0.09	0.20	0.09	0.91	0.81	170296
sp9	0.76	0.55	0.52	0.83	0.15	0.07	0.63	0.74	83035
sp10	0.94	0.86	0.86	0.19	0.48	0.15	0.89	0.93	303473
sp11	0.92	0.79	0.78	0.31	0.30	0.12	0.89	0.84	40495
sp12	0.87	0.70	0.70	0.45	0.23	0.12	0.89	0.78	53927
sp13	0.88	0.83	0.81	0.12	0.26	0.05	0.91	0.78	490520
sp14	0.85	0.74	0.72	0.32	0.25	0.15	0.71	0.83	18404
sp15	0.77	0.72	0.71	0.19	0.16	0.17	0.85	0.56	294318
sp16	0.73	0.64	0.62	0.37	0.11	0.16	0.79	0.60	102968
sp17	0.80	0.72	0.70	0.27	0.18	0.11	0.96	0.56	230123
sp18	0.76	0.69	0.68	0.26	0.16	0.24	0.88	0.55	272413
sp19	0.88	0.81	0.79	0.19	0.27	0.09	0.94	0.75	76974
sp20	0.80	0.71	0.71	0.30	0.17	0.28	0.75	0.76	18647
sp21	0.81	0.74	0.73	0.20	0.21	0.25	0.83	0.75	118974
sp22	0.88	0.80	0.80	0.20	0.32	0.20	0.88	0.75	65364
sp23	0.80	0.73	0.72	0.23	0.22	0.25	0.59	0.89	176967
sp24	0.91	0.88	0.88	0.07	0.42	0.14	0.92	0.92	332387
sp25	0.76	0.67	0.67	0.34	0.15	0.27	0.56	0.82	11078
sp26	0.81	0.67	0.66	0.45	0.17	0.13	0.85	0.66	161996
sp27	0.94	0.79	0.77	0.34	0.36	0.14	0.92	0.94	13563
sp28	0.82	0.71	0.71	0.35	0.21	0.19	0.80	0.69	154268
sp29	0.97	0.87	0.88	0.20	0.54	0.18	0.93	0.90	27309
sp30	0.82	0.72	0.72	0.31	0.18	0.28	0.73	0.83	184573
sp31	0.84	0.78	0.78	0.17	0.25	0.11	0.93	0.60	450491
sp32	0.79	0.75	0.74	0.14	0.18	0.22	0.73	0.73	280462
sp33	–	–	–	–	–	–	–	–	–
sp34	0.65	0.64	0.63	0.08	0.05	0.20	0.77	0.48	342629
sp35	0.95	0.88	0.88	0.16	0.48	0.11	0.86	0.89	41431
sp36	–	–	–	–	–	–	–	–	–
sp37	0.78	0.70	0.69	0.30	0.15	0.04	1.00	0.43	680447
sp38	0.72	0.56	0.52	0.74	0.07	0.10	0.86	0.54	228553
sp39	0.65	0.64	0.64	0.05	0.07	0.10	0.40	0.93	107119
sp40	0.75	0.70	0.69	0.21	0.13	0.18	0.89	0.53	308557
mean	0.81	0.72	0.71	0.28	0.21	0.15	0.82	0.71	217749
median	0.81	0.71	0.71	0.27	0.18	0.14	0.86	0.73	173632

Weighted models: forward-backward selection

Species code	AUC_{RS}	AUC_{CV}	AUC_{LOOCV}	GE	D^2	Cut-point	Sensitivity	Specificity	Suitable cells
sp1	0.74	0.65	0.63	0.37	0.14	0.51	0.81	0.56	45546
sp2	0.65	0.61	0.61	0.26	0.08	0.49	0.82	0.52	751681
sp3	0.64	0.61	0.60	0.22	0.05	0.43	0.88	0.37	457182
sp4	–	–	–	–	–	–	–	–	–
sp5	0.70	0.68	0.68	0.10	0.12	0.53	0.79	0.56	694878
sp6	0.77	0.67	0.66	0.39	0.20	0.52	0.89	0.58	387197
sp7	0.77	0.73	0.73	0.14	0.15	0.45	0.82	0.66	83594
sp8	0.81	0.79	0.79	0.05	0.25	0.44	0.91	0.69	351505
sp9	0.71	0.59	0.58	0.54	0.12	0.56	0.44	0.89	5440
sp10	0.90	0.85	0.85	0.13	0.44	0.49	0.89	0.79	407486
sp11	0.69	0.66	0.66	0.15	0.07	0.34	1.00	0.32	372303
sp12	–	–	–	–	–	–	–	–	–
sp13	0.94	0.83	0.82	0.27	0.58	0.51	1.00	0.83	229045
sp14	0.85	0.74	0.75	0.30	0.29	0.46	0.95	0.71	90049
sp15	0.75	0.74	0.74	0.03	0.16	0.37	0.88	0.51	338540
sp16	0.73	0.66	0.66	0.28	0.12	0.35	0.93	0.42	277180
sp17	0.82	0.73	0.73	0.29	0.27	0.41	0.93	0.61	194677
sp18	0.69	0.68	0.68	0.07	0.10	0.44	0.84	0.50	461690
sp19	0.84	0.77	0.77	0.23	0.32	0.51	0.82	0.74	97072
sp20	0.77	0.70	0.70	0.25	0.17	0.47	0.85	0.63	32773
sp21	0.77	0.75	0.75	0.08	0.22	0.47	0.88	0.58	215701
sp22	0.76	0.73	0.72	0.12	0.20	0.51	0.61	0.77	91447
sp23	0.72	0.71	0.71	0.06	0.11	0.48	0.72	0.67	147475
sp24	0.94	0.81	0.81	0.30	0.59	0.48	1.00	0.83	434070
sp25	0.69	0.65	0.64	0.22	0.09	0.51	0.59	0.78	17084
sp26	–	–	–	–	–	–	–	–	–
sp27	0.73	0.58	0.57	0.65	0.16	0.54	0.75	0.63	308626
sp28	0.77	0.71	0.71	0.20	0.20	0.43	0.89	0.55	257171
sp29	–	–	–	–	–	–	–	–	–
sp30	0.78	0.74	0.73	0.15	0.16	0.56	0.71	0.73	324872
sp31	0.77	0.76	0.75	0.04	0.15	0.34	0.93	0.51	388590
sp32	0.75	0.74	0.74	0.03	0.14	0.45	0.85	0.59	557024
sp33	–	–	–	–	–	–	–	–	–
sp34	0.65	0.64	0.64	0.08	0.06	0.45	0.77	0.50	329889
sp35	0.78	0.76	0.76	0.07	0.20	0.63	0.71	0.78	57192
sp36	–	–	–	–	–	–	–	–	–
sp37	0.84	0.74	0.74	0.29	0.33	0.59	1.00	0.66	378559
sp38	–	–	–	–	–	–	–	–	–
sp39	0.67	0.52	0.52	0.91	0.10	0.61	0.40	0.93	113265
sp40	0.71	0.70	0.70	0.04	0.11	0.47	0.80	0.57	316507
mean	0.76	0.70	0.70	0.22	0.20	0.48	0.82	0.64	279252
median	0.76	0.71	0.71	0.20	0.16	0.48	0.85	0.63	308626

Weighted models: backward-forward selection

Species code	AUC_{RS}	AUC_{CV}	AUC_{LOOCV}	GE	D^2	Cut-point	Sensitivity	Specificity	Suitable cells
sp1	0.88	0.70	0.69	0.47	0.37	0.39	1.00	0.68	73568
sp2	0.96	0.65	0.64	0.67	0.60	0.58	1.00	0.88	246162
sp3	0.76	0.55	0.55	0.82	0.17	0.47	0.79	0.63	90106
sp4	0.86	0.51	0.50	0.97	0.32	0.54	0.88	0.79	303555
sp5	0.90	0.52	0.51	0.95	0.39	0.46	0.93	0.75	282147
sp6	0.90	0.58	0.58	0.81	0.39	0.54	0.94	0.78	382911
sp7	0.86	0.72	0.71	0.40	0.33	0.47	0.86	0.74	66309
sp8	0.91	0.56	0.54	0.86	0.50	0.64	0.91	0.87	104118
sp9	0.90	0.54	0.52	0.91	0.42	0.50	1.00	0.78	110231
sp10	0.97	0.79	0.79	0.37	0.64	0.63	0.94	0.93	282084
sp11	0.94	0.77	0.78	0.39	0.52	0.56	0.94	0.87	29023
sp12	0.90	0.70	0.69	0.50	0.43	0.45	1.00	0.72	67646
sp13	0.97	0.61	0.59	0.76	0.68	0.63	1.00	0.91	166910
sp14	0.92	0.78	0.79	0.33	0.48	0.63	0.90	0.87	54375
sp15	0.82	0.70	0.70	0.38	0.26	0.52	0.78	0.71	136965
sp16	0.82	0.63	0.63	0.58	0.26	0.40	0.97	0.54	124153
sp17	0.85	0.72	0.72	0.38	0.36	0.43	1.00	0.66	116149
sp18	0.82	0.70	0.71	0.36	0.26	0.62	0.67	0.83	64097
sp19	0.93	0.74	0.74	0.44	0.56	0.57	1.00	0.85	39553
sp20	0.82	0.70	0.70	0.38	0.23	0.48	0.85	0.71	61985
sp21	0.81	0.74	0.74	0.23	0.27	0.46	0.93	0.64	181472
sp22	0.89	0.78	0.79	0.28	0.40	0.53	0.83	0.84	61078
sp23	0.86	0.70	0.70	0.45	0.32	0.44	0.90	0.70	401593
sp24	0.98	0.77	0.76	0.45	0.79	0.60	1.00	0.94	354581
sp25	0.81	0.66	0.65	0.49	0.23	0.36	0.95	0.51	121408
sp26	0.89	0.65	0.66	0.61	0.41	0.46	1.00	0.75	125538
sp27	0.98	0.69	0.68	0.60	0.70	0.59	1.00	0.92	18822
sp28	0.85	0.70	0.71	0.43	0.32	0.44	0.94	0.68	165059
sp29	0.99	0.87	0.88	0.25	0.81	0.72	1.00	0.97	33751
sp30	0.83	0.70	0.71	0.39	0.26	0.55	0.78	0.78	166891
sp31	0.87	0.75	0.75	0.32	0.36	0.39	1.00	0.64	519086
sp32	0.80	0.73	0.74	0.23	0.23	0.38	0.95	0.57	303805
sp33	0.92	0.50	0.49	0.99	0.48	0.55	1.00	0.83	169248
sp34	0.65	0.64	0.64	0.10	0.06	0.46	0.77	0.50	329887
sp35	0.99	0.90	0.90	0.18	0.84	0.77	1.00	0.97	23081
sp36	0.84	0.55	0.54	0.85	0.27	0.58	0.71	0.82	195811
sp37	0.92	0.64	0.63	0.67	0.49	0.58	1.00	0.81	169939
sp38	0.79	0.50	0.49	0.99	0.21	0.42	1.00	0.55	180849
sp39	0.93	0.52	0.48	0.96	0.49	0.55	1.00	0.84	164666
sp40	0.72	0.70	0.70	0.10	0.12	0.40	0.89	0.45	380214
mean	0.88	0.67	0.67	0.53	0.40	0.52	0.93	0.76	171721
median	0.89	0.70	0.70	0.45	0.38	0.53	0.95	0.78	150816

Weighted multimodels

Species code	AUC_{RS}	AUC_{CV}	AUC_{LOOCV}	GE	D^2	Cut-point	Sensitivity	Specificity	Suitable cells
sp1	0.87	0.70	0.70	0.46	0.28	0.37	1.00	0.62	46384
sp2	0.95	0.68	0.66	0.61	0.40	0.64	0.91	0.93	206072
sp3	0.71	0.59	0.59	0.57	0.09	0.52	0.73	0.62	185036
sp4	–	–	–	–	–	–	–	–	–
sp5	0.75	0.67	0.67	0.32	0.14	0.54	0.79	0.65	647126
sp6	0.86	0.64	0.63	0.62	0.24	0.59	0.78	0.83	302319
sp7	0.83	0.73	0.73	0.31	0.24	0.48	0.82	0.74	61834
sp8	0.84	0.78	0.78	0.19	0.27	0.47	0.91	0.75	292247
sp9	0.87	0.55	0.53	0.87	0.24	0.44	0.94	0.76	145776
sp10	0.95	0.85	0.85	0.22	0.52	0.59	0.89	0.91	337941
sp11	0.93	0.78	0.78	0.36	0.38	0.56	0.89	0.87	26316
sp12	–	–	–	–	–	–	–	–	–
sp13	0.96	0.80	0.80	0.34	0.57	0.55	1.00	0.87	193690
sp14	0.90	0.78	0.79	0.30	0.38	0.54	0.90	0.84	56897
sp15	0.79	0.73	0.73	0.22	0.20	0.40	0.88	0.58	292191
sp16	0.79	0.66	0.66	0.45	0.18	0.47	0.90	0.58	131372
sp17	0.85	0.73	0.73	0.34	0.27	0.41	0.96	0.65	143825
sp18	0.79	0.71	0.71	0.28	0.19	0.48	0.88	0.59	247869
sp19	0.80	0.68	0.68	0.41	0.30	0.47	0.82	0.67	81663
sp20	0.81	0.71	0.71	0.31	0.20	0.47	0.88	0.65	38198
sp21	0.80	0.75	0.75	0.17	0.22	0.52	0.83	0.70	142424
sp22	0.88	0.78	0.78	0.27	0.33	0.46	0.88	0.73	94834
sp23	0.82	0.72	0.72	0.32	0.23	0.36	0.97	0.56	473737
sp24	0.97	0.81	0.80	0.35	0.66	0.56	1.00	0.91	377097
sp25	0.79	0.66	0.66	0.43	0.18	0.41	0.83	0.60	47630
sp26	–	–	–	–	–	–	–	–	–
sp27	0.98	0.72	0.72	0.53	0.55	0.56	1.00	0.91	17233
sp28	0.83	0.72	0.72	0.33	0.24	0.45	0.91	0.66	180995
sp29	–	–	–	–	–	–	–	–	–
sp30	0.83	0.73	0.73	0.32	0.20	0.54	0.83	0.74	211851
sp31	0.84	0.77	0.77	0.20	0.25	0.37	1.00	0.60	496825
sp32	0.78	0.75	0.75	0.13	0.19	0.51	0.76	0.73	271233
sp33	–	–	–	–	–	–	–	–	–
sp34	0.65	0.64	0.64	0.07	0.06	0.46	0.74	0.51	315043
sp35	0.99	0.88	0.89	0.22	0.64	0.59	1.00	0.97	26794
sp36	–	–	–	–	–	–	–	–	–
sp37	0.91	0.69	0.68	0.53	0.34	0.60	1.00	0.81	204731
sp38	–	–	–	–	–	–	–	–	–
sp39	0.92	0.48	0.45	1.06	0.32	0.48	1.00	0.81	202659
sp40	0.71	0.70	0.70	0.08	0.11	0.45	0.85	0.51	315015
mean	0.85	0.71	0.71	0.37	0.29	0.49	0.89	0.72	206511
median	0.84	0.72	0.72	0.32	0.24	0.48	0.89	0.73	193690

Spatial models: forward-backward selection

Species code	AUC_{RS}	AUC_{CV}	AUC_{LOOCV}	GE	D^2	Cut-point	Sensitivity	Specificity	Suitable cells
sp1	0.73	0.64	0.63	0.37	0.15	0.45	0.90	0.51	47155
sp2	0.66	0.60	0.59	0.35	0.08	0.44	0.91	0.44	799290
sp3	0.68	0.61	0.61	0.38	0.09	0.53	0.61	0.65	267164
sp4	–	–	–	–	–	–	–	–	–
sp5	0.71	0.63	0.62	0.41	0.14	0.45	0.86	0.51	704839
sp6	0.76	0.68	0.68	0.32	0.20	0.51	0.89	0.57	402397
sp7	0.80	0.74	0.73	0.21	0.24	0.57	0.79	0.75	64565
sp8	0.82	0.78	0.79	0.12	0.26	0.39	0.91	0.66	361760
sp9	0.81	0.69	0.69	0.39	0.23	0.46	0.81	0.71	602218
sp10	0.92	0.88	0.88	0.09	0.49	0.54	0.89	0.85	382879
sp11	0.68	0.63	0.62	0.30	0.11	0.41	0.94	0.45	198565
sp12	–	–	–	–	–	–	–	–	–
sp13	0.95	0.86	0.87	0.21	0.70	0.70	1.00	0.91	117676
sp14	0.89	0.81	0.82	0.19	0.41	0.65	0.86	0.86	55147
sp15	0.77	0.75	0.75	0.07	0.18	0.66	0.59	0.81	75668
sp16	0.80	0.72	0.72	0.27	0.24	0.56	0.76	0.73	183307
sp17	0.85	0.72	0.73	0.36	0.32	0.53	0.89	0.69	140616
sp18	0.74	0.71	0.71	0.11	0.15	0.41	0.90	0.47	528320
sp19	0.89	0.80	0.80	0.25	0.45	0.42	1.00	0.66	136213
sp20	0.80	0.70	0.70	0.32	0.22	0.49	0.83	0.64	27964
sp21	0.78	0.76	0.75	0.10	0.22	0.54	0.78	0.70	148039
sp22	0.83	0.79	0.79	0.12	0.29	0.55	0.68	0.83	317731
sp23	0.73	0.70	0.70	0.13	0.12	0.46	0.72	0.64	171792
sp24	0.98	0.87	0.85	0.24	0.82	0.81	1.00	0.95	164066
sp25	0.70	0.64	0.64	0.31	0.10	0.51	0.63	0.75	14221
sp26	–	–	–	–	–	–	–	–	–
sp27	0.75	0.64	0.63	0.45	0.19	0.51	0.83	0.60	444505
sp28	0.81	0.72	0.71	0.31	0.27	0.52	0.86	0.65	204149
sp29	–	–	–	–	–	–	–	–	–
sp30	0.79	0.74	0.74	0.18	0.22	0.56	0.68	0.78	292962
sp31	0.77	0.74	0.74	0.09	0.18	0.42	0.83	0.60	260912
sp32	0.75	0.74	0.73	0.05	0.15	0.43	0.83	0.61	543616
sp33	–	–	–	–	–	–	–	–	–
sp34	0.65	0.63	0.63	0.16	0.07	0.48	0.79	0.49	320161
sp35	0.80	0.76	0.76	0.13	0.21	0.42	0.86	0.63	212122
sp36	–	–	–	–	–	–	–	–	–
sp37	0.85	0.81	0.81	0.12	0.39	0.53	0.93	0.73	361628
sp38	–	–	–	–	–	–	–	–	–
sp39	0.74	0.35	0.31	1.65	0.16	0.41	0.90	0.54	420617
sp40	0.71	0.69	0.68	0.11	0.12	0.39	0.89	0.48	481060
mean	0.78	0.71	0.71	0.27	0.25	0.51	0.83	0.66	286464
median	0.78	0.72	0.72	0.21	0.21	0.51	0.86	0.65	260912

Spatial models: backward-forward selection

Species code	AUC$_{RS}$	AUC$_{CV}$	AUC$_{LOOCV}$	GE	D^2	Cut-point	Sensitivity	Specificity	Suitable cells
sp1	0.94	0.82	0.81	0.28	0.62	0.55	1.00	0.82	38252
sp2	1.00	0.60	0.62	0.80	1.00	0.01	1.00	1.00	91235
sp3	0.77	0.59	0.59	0.66	0.20	0.41	0.91	0.57	172857
sp4	0.89	0.54	0.54	0.90	0.39	0.56	0.88	0.79	414473
sp5	0.91	0.53	0.52	0.93	0.44	0.50	0.93	0.82	240624
sp6	0.94	0.71	0.72	0.52	0.57	0.56	0.94	0.84	346245
sp7	0.88	0.75	0.75	0.35	0.38	0.39	0.93	0.72	71253
sp8	0.97	0.64	0.63	0.71	0.76	0.51	1.00	0.96	55450
sp9	0.93	0.58	0.57	0.82	0.51	0.53	1.00	0.80	128832
sp10	0.98	0.85	0.85	0.27	0.76	0.48	1.00	0.91	295523
sp11	0.95	0.79	0.79	0.36	0.61	0.57	0.94	0.91	26828
sp12	0.99	0.71	0.71	0.56	0.78	0.66	1.00	0.96	32757
sp13	1.00	0.64	0.64	0.72	1.00	0.01	1.00	1.00	125498
sp14	0.94	0.81	0.81	0.28	0.59	0.38	1.00	0.81	134333
sp15	0.84	0.74	0.74	0.30	0.30	0.52	0.83	0.75	141835
sp16	0.90	0.74	0.74	0.41	0.43	0.48	0.97	0.74	124958
sp17	0.86	0.74	0.75	0.33	0.38	0.43	0.96	0.70	100539
sp18	0.89	0.72	0.73	0.43	0.40	0.52	0.84	0.79	155040
sp19	0.92	0.71	0.72	0.50	0.59	0.45	1.00	0.84	48672
sp20	0.83	0.72	0.72	0.33	0.26	0.41	0.92	0.64	89729
sp21	0.83	0.76	0.76	0.21	0.30	0.44	0.93	0.63	181975
sp22	0.90	0.79	0.80	0.27	0.43	0.53	0.88	0.83	65657
sp23	0.90	0.71	0.71	0.46	0.43	0.52	0.86	0.78	376909
sp24	1.00	0.73	0.73	0.55	1.00	0.01	1.00	1.00	161628
sp25	0.85	0.67	0.67	0.51	0.29	0.40	0.93	0.66	116575
sp26	0.91	0.70	0.70	0.51	0.48	0.50	0.96	0.80	111526
sp27	0.99	0.73	0.74	0.52	0.82	0.77	1.00	0.96	12280
sp28	0.91	0.76	0.76	0.36	0.48	0.49	0.97	0.72	173739
sp29	1.00	0.83	0.85	0.33	0.98	0.76	1.00	1.00	31599
sp30	0.86	0.76	0.77	0.28	0.34	0.49	0.85	0.77	161457
sp31	0.87	0.74	0.74	0.35	0.40	0.46	0.97	0.69	456731
sp32	0.81	0.73	0.74	0.24	0.24	0.35	0.95	0.56	273061
sp33	0.95	0.60	0.58	0.79	0.62	0.69	1.00	0.91	177411
sp34	0.65	0.63	0.62	0.18	0.07	0.49	0.77	0.49	329138
sp35	1.00	0.81	0.81	0.38	1.00	0.01	1.00	1.00	95358
sp36	0.91	0.70	0.69	0.53	0.47	0.41	1.00	0.76	212241
sp37	0.93	0.75	0.75	0.40	0.59	0.57	1.00	0.85	132383
sp38	0.81	0.52	0.52	0.95	0.27	0.50	0.90	0.64	111054
sp39	0.98	0.68	0.66	0.64	0.76	0.66	1.00	0.94	100337
sp40	0.72	0.69	0.69	0.14	0.12	0.38	0.96	0.45	347581
mean	0.90	0.71	0.71	0.48	0.53	0.46	0.95	0.80	161589
median	0.91	0.72	0.72	0.42	0.47	0.49	0.96	0.80	130608

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium of Information Theory* (Ed, Csaki, F. and Petrov, B. N.) Akademiai Kiado, Budapest, Hungary.

Akaike, H. (1978) A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, **30**, 9-14.

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

Anderson, R. P. (2003) Real vs. artefactual absences in species distributions: tests for Oryzomys albigularis (Rodentia : Muridae) in Venezuela. *Journal of Biogeography*, **30**, 591-605.

Araujo, M. B. and Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688.

Araújo, M. B., Cabeza, M., Thuiller, W., Hannah, L. and Williams, P. H. (2004) Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, **10**, 1618-1626.

Augustin, N. H., Mugglestone, M. A. and Buckland, S. T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339-347.

Austin, M. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1-19.

Austin, M. P. (1999) The potential contribution of vegetation ecology to biodiversity research. *Ecography*, **22**, 465-484.

Austin, M. P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.

Bakkenes, M., Alkemade, J. R. M., Ihle, F., Leemans, R. and Latour, J. B. (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, **8**, 390-407.

Besag, J. (1974) Spatial interaction and statistical-analysis of lattice systems. *Journal of the Royal Statistical Society Series B-Methodological*, **36**, 192-236.

Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Rylands, A. B., Konstant, W. R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G. and Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909-923.

Brunsdon, C., Fotheringham, S. and Charlton, M. (2007) Geographically weighted discriminant analysis. *Geographical Analysis*, **39**, 376-396.

Burgess, N. D., Balmford, A., Cordeiro, N. J., Fjeldså, J., Kuper, W., Rahbek, C., Sanderson, E. W., Scharlemann, J. P. W., Sommer, J. H. and Williams, P. H. (2007a) Correlations among species distributions, human density and human infrastructure across the high biodiversity tropical mountains of Africa. *Biological Conservation*, **134**, 164-177.

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. and Stuart, S. N. (2007b) The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, **134**, 209-231.

Cohen, J., Cohen, P., West, S. G. and Aiken, L. S. (2003) *Applied Multiple Regression/Correlation Analysis For The Behavioural Sciences*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA.

Dormann, C. F. (2007a) Assessing the validity of autologistic regression. *Ecological Modelling*, **207**, 234-242.

Dormann, C. F. (2007b) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129-138.

Dormann, C. F. (2007c) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, **8**, 387-397.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kuhn, I., Ohlemuller, R., Peres-Neto, P. R., Reineking, B., Schroder, B., Schurr, F. M. and Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609-628.

Dormann, C. F., Schweiger, O., Arens, P., Augenstein, I., Aviron, S., Bailey, D., Baudry, J., Billeter, R., Bugter, R., Bukacek, R., Burel, F., Cerny, M., De Cock, R., De Blust, G., DeFilippi, R., Diekotter, T., Dirksen, J., Durka, W., Edwards, P. J., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., Le Coeur, D., Liira, J., Maelfait, J. P., Opdam, P., Roubalova, M., Schermann-Legionnet, A., Schermann, N., Schmidt, T., Smulders, M. J. M., Speelmans, M., Simova, P., Verboom, J., van Wingerden, W. and Zobel, M. (2008) Prediction uncertainty of environmental change effects on temperate European biodiversity. *Ecology Letters*, **11**, 235-244.

Eberhardt, L. L. (2003) What should we do about hypothesis testing? *Journal of Wildlife Management*, **67**, 241-247.

Engler, R., Guisan, A. and Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263-274.

Ferrier, S., Drielsma, M., Manion, G. and Watson, G. (2002a) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, **11**, 2309-2338.

Ferrier, S., Watson, G., Pearce, J. and Drielsma, M. (2002b) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation*, **11**, 2275-2307.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002) *Geographical Weighted Regression: The Analysis of Spatially Relationships*, Wiley, Chichester, UK.

Graham, M. H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology*, **84**, 2809-2815.

GRASS-Development-Team (2009) Geographic Resources Analysis Support System (GRASS), GNU General Public License (http://grass.osgeo.org/).

Green, D. M. and Swets, J. A. (1974) *Signal Detection Theory and Psychophysics*, Krieger, New York, USA.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A. and Zimmermann, N. E. (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**, 501-511.

Guisan, A. and Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Hartley, S., Harris, R. and Lester, P. J. (2006) Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters*, **9**, 1068-1079.

Hastie, T. J., Friedman, J. H. and Tibshirani, R. J. (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York, USA.

Hastie, T. J. and Tibshirani, R. J. (1990) *Generalized Additive Models*, Chapman and Hall, London, UK.

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

Horn, B. K. P. (1981) Hill shading and reflectance map. *Proceedings of the IEEE*, **69**, 14-47.

Johnson, C. J., Nielsen, S. E., Merrill, E. H., McDonald, T. L. and Boyce, M. S. (2006) Resource selection functions based on use-availability data: Theoretical motivation and evaluation methods. *Journal of Wildlife Management*, **70**, 347-357.

Johnson, J. B. and Omland, K. S. (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution*, **19**, 101-108.

Kohavi, R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*.

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

Leathwick, J. R. and Austin, M. P. (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, **82**, 2560-2573.

Lichstein, J. W., Simons, T. R., Shriner, S. A. and Franzreb, K. E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, **72**, 445-463.

Liu, C. R., Berry, P. M., Dawson, T. P. and Pearson, R. G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385-393.

Lobo, J. M., Jimenez-Valverde, A. and Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145-151.

Lovett, J. C. (1985) Moist forsts of eastern Tanzania. *Swara*, **8**, 8-9.

Lovett, J. C. (1990) Classification and status of the moist forests of Tanzania. *Notifications from the Institute for General Botany Hamburg*, **23a**, 287–300.

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. (1998) Importance of the Eastern Arc Mountains for vascular plants. *Journal of East African Natural History*, **87**, 37-58.

Lovett, J. C., Clarke, G. P., Moore, R. and Morrey, G. H. (2001) Elevational distribution of restricted range forest tree taxa in eastern Tanzania. *Biodiversity and Conservation*, **10**, 541-550.

Lovett, J. C., Ruffo, C. K., Gereau, R. E. and Taplin, J. R. D. (2006) *Field guide to the moist forest trees of Tanzania*, The Society for Environmental Exploration, UK and the University of Dar es Salaam, The United Republic of Tanzania.

Maggini, R., Lehmann, A., Zimmermann, N. E. and Guisan, A. (2006) Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography*, **33**, 1729-1749.

Manel, S., Williams, H. C. and Ormerod, S. J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-931.

Marchant, R., Mumbi, C., Behera, S. and Yamagata, T. (2007) The Indian Ocean dipole - the unsung driver of climatic variability in East Africa. *African Journal of Ecology*, **45**, 4-16.

McClean, C. J., Lovett, J. C., Kuper, W., Hannah, L., Sommer, J. H., Barthlott, W., Termansen, M., Smith, G. E., Tokamine, S. and Taplin, J. R. D. (2005) African plant diversity and climate change. *Annals of the Missouri Botanical Garden*, **92**, 139-152.

McPherson, J. M., Jetz, W. and Rogers, D. J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811-823.

Mickey, R. M. and Greenland, S. (1989) The impact of confounder selection criteria on effect estimation. *American Journal of Epidemiology*, **129**, 125-137.

Miller, J., Franklin, J. and Aspinall, R. (2007) Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, **202**, 225-242.

Mittermeier, R. A., Robles Gil., P., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., Lamoreux, J. and da Fonseca, G. A. B. (2004) *Hotspots Revisited*, CEMEX, Mexico.

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

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

Olson, D. M. and Dinerstein, E. (1998) The global 200: A representation approach to conserving the Earth's most biologically valuable ecoregions. *Conservation Biology*, **12**, 502-515.

Parker, B. J., Guenter, S. and Bedo, J. (2007) Stratification bias in low signal microarray studies. *BMC Bioinformatics*, **8**, 16.

Pearce, J. and Ferrier, S. (2000a) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127-147.

Pearce, J. and Ferrier, S. (2000b) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225-245.

Pearce, J. L., Cherry, K., Drielsma, M., Ferrier, S. and Whish, G. (2001) Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. *Journal of Applied Ecology*, **38**, 412-424.

Pearson, R. G. and Dawson, T. P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361-371.

Pearson, R. G., Thuiller, W., Araujo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P. and Lees, D. C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704-1711.

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A. and Solomon, A. M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117-134.

Pulliam, H. R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349-361.

R-Development-Core-Team (2009) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org/).

Richards, C. L., Carstens, B. C. and Knowles, L. L. (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, **34**, 1833-1845.

Roberts, D. W. and Cooper, S. V. (1989) Concepts and techniques of vegetation mapping. In *Land Classifications Based on Vegetation: Applications for Resource Management* (Eds, Ferguson, D., Morgan, P. and Johnson, F. D.) USDA For. Serv. Gen. Tech. Rep. INT-257, Ogden, Utah, USA, pp. 90-96.

Rushton, S. P., Ormerod, S. J. and Kerby, G. (2004) New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**, 193-200.

Schlüter, T. (1997) Geology of East Africa. In *Contributions to the Regional Geography of the Earth* (translated from German), Borntraeger, Berlin, Germany.

Segurado, P. and Araujo, M. B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555-1568.

Segurado, P. and Araújo, M. B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555-1568.

Seoane, J., Bustamante, J. and Diaz-Delgado, R. (2005) Effect of expert opinion on the predictive ability of environmental models of bird distribution. *Conservation Biology*, **19**, 512-522.

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

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **36**, 111-147.

Swets, J. A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293.

Taplin, J. R. D. and 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.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L. and Williams, S. E. (2004) Extinction risk from climate change. *Nature*, **427**, 145-148.

Thornthwaite, C. W. (1948) An Approach toward a Rational Classification of Climate. *Geographical Review*, **38**, 55-94.

Thuiller, W., Brotons, L., Araújo, M. B. and Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165-172.

van Teeffelen, A. J. A. and Ovaskainen, O. (2007) Can the cause of aggregation be inferred from species distributions? *Oikos*, **116**, 4-16.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B. and Freckleton, R. P. (2006) Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, **75**, 1182-1189.

Yee, T. W. and Mitchell, N. D. (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science*, **2**, 587-602.

Chapter 4 – Distribution Models and Conservation Priority

Published in *Diversity and Distributions* (Jul. 2010), Vol. 16, Issue 4, Pages 628-642

Can distribution models help refine inventory-based estimates of conservation priority? A case study in the Eastern Arc forests of Tanzania and Kenya

Philip J. Platts [a,*], Antje Ahrends [b], Roy E. Gereau [c], Colin J. McClean [a], Jon C. Lovett [d], Andrew R. Marshall [a,e], Petri K. E. Pellikka [f], Mark Mulligan [g], Eibleis Fanning [h], Rob Marchant [a]

[a] *The York Institute for Tropical Ecosystem Dynamics (KITE), Environment Department, University of York, Heslington, York YO10 5DD, UK.* [b] *Royal Botanic Garden Edinburgh, 20A Inverleith Row, EH3 5LR, UK.* [c] *Missouri Botanical Garden, P.O. Box 299, Saint Louis, MO 63166-0299, USA.* [d] *Centre for Clean Technology and Environmental Policy, University of Twente/CSTM, P.O. Box 217, 7500 AE Enschede, The Netherlands.* [e] *Flamingo Land Ltd, Kirby Misperton, Malton, North Yorkshire YO17 6UX, UK.* [f] *Department of Geography, University of Helsinki, P.O. Box 64, 00014 Helsinki, Finland.* [g] *Department of Geography, King's College London, Strand, London, WC2R 2LS, UK.* [h] *Society for Environmental Exploration (Frontier), 50-52 Rivington Street, London EC2A 3QP, UK.*

** Corresponding author. Email: philip.platts@gmail.com*

Abstract

Data shortages mean that conservation priorities can be highly sensitive to historical patterns of exploration. Here, we investigate the potential of regionally focussed species distribution models to elucidate fine-scale patterns of richness, rarity and endemism in the Eastern Arc Mountains (Tanzania and Kenya). Generalised additive models and land cover data are used to estimate the distributions of 452 forest plant taxa (trees, lianas, shrubs and herbs). Presence records from a newly compiled database are regressed against environmental variables in a stepwise multimodel. Estimates of occurrence in forest patches are collated across target groups and analysed alongside inventory-based estimates of conservation priority. We find that predicted richness is higher than observed richness, with the biggest disparities in regions that have had the least research. North Pare and Nguu in particular are predicted to be more important than the inventory data suggest. Environmental conditions in parts of Nguru could support as many range-restricted and endemic taxa as Uluguru, although realised niches are subject to unknown colonisation histories. Concentrations of rare plants are especially high in the Usambaras, a pattern mediated in models by moisture indices, whilst overall richness is better explained by temperature gradients. Tree data dominate the botanical inventory; we find that priorities based on other growth forms might favour the mountains in a different order. We conclude that distribution models can provide conservation planning with high-resolution estimates of richness in well-researched areas, and predictive estimates of conservation importance elsewhere. Spatial and taxonomic biases in the data are essential considerations, as is the spatial scale used for models. We caution that predictive estimates are most uncertain for the species of highest conservation concern, and advocate using models and targeted field assessments iteratively to refine our understanding of which areas should be prioritised for conservation.

Keywords: biodiversity; conservation planning; endemism; rare species; sampling bias; spatial prediction.

Introduction

Limited resources for conservation dictate identification of priority regions to achieve effective conservation action (Margules and Pressey, 2000; Myers *et al.*, 2000; Eken *et al.*, 2004; Wilson *et al.*, 2006). A major constraint, particularly at the site scale, is the scarcity of fine-scale data on the distribution of biodiversity (da Fonseca *et al.*, 2000; Küper *et al.*, 2006). Given the urgency of conservation action and the fact that much-needed biodiversity inventories are costly and underfunded (Lawton *et al.*, 1998), the application of distribution models to species occurrence data could provide a practical way forward.

Conservation action is most often driven by decisions at the site scale (Mace *et al.*, 2000; Ferrier, 2002). Such prioritisations can be highly sensitive to the inventory data available at the time, resulting in bias towards sites with a good history of biological exploration (Reddy and Davalos, 2003). Early explorations in the Eastern Arc Mountains (hereafter, EAMs) focused almost exclusively on the Uluguru and Usambara ranges (1880–1980). Over the last 30 years, funding has continued to be spread unevenly, favouring some mountain blocs such as the Usambaras and Udzungwas, whilst others such as North Pare and Nguu remain under-surveyed (Ahrends *et al*., 2011 in Appendix I). Recent investment in the Nguru and Rubeho Mountains has resulted in the discovery of new species, altering conservation priorities still further (Doggart *et al.*, 2006; Menegon *et al.*, 2008). Spatially referenced inventory data for regions such as the EAMs have become increasingly accessible in recent years (e.g., http://www.tropicos.org); however, for use in a modelling framework, it is necessary to consider the historical, artifactual and biological processes that underlie them (Graham *et al.*, 2004). For instance, inventory data are often biased not only in geographical space but also towards particular taxonomic groups – in the case of vascular plants, trees tend to be the dominant growth form recorded. Since plant diversity is sometimes employed as an indicator of overall biodiversity value (Bladt *et al.*, 2008; Larsen *et al.*, 2009), it is important to consider whether models predict similar patterns for the different growth forms within this group.

Historical habitat and climate configurations are also important for understanding species distributions, especially for endemic taxa (Jetz *et al.*, 2004; Possingham and Wilson, 2005; Graham *et al.*, 2006). Climatic conditions in the EAMs are thought to have been relatively stable, their proximity to the Indian Ocean providing a buffer against global trends in climate (Lovett, 1990; Marchant *et al.*, 2007). Similar ecoclimatic stability is evident in other regions where highland habitats abut warm tropical oceans, such as the Atlantic rainforests in South America and the Queensland rainforests in Australia (Lovett *et al.*, 2005) and has been suggested as a key driver of endemism in biodiversity hotspots (Fjeldså *et al.*, 1997). Historical and evolutionary processes are particularly pertinent in the EAMs, which are geologically much older than adjacent mountains (Griffiths, 1993; Schlüter, 1997). Recently, however, they have suffered significant deforestation, reducing forest cover by around 70% (Burgess *et al.*, 2007; Hall *et al.*, 2009).

The aims of this article are to investigate the extent to which modelled richness is affected by historical and taxonomic bias in inventory data and to highlight the potential conservation importance of under-researched areas. Present-day climatic conditions, topography and soil parameters are combined with remotely sensed land cover data to estimate the spatial distributions of 452 plant taxa (species, subspecies, varieties), including 71 that are endemic to the EAMs and/or threatened with extinction. Our discussion of results explores the potential of distribution models to help refine conservation priorities in a region where confounding factors are typical of those found in many biodiversity hotspots.

Methods

Study region

The EAMs are part of the Eastern Afromontane Biodiversity Hotspot (Mittermeier *et al.*, 2004), extending from the Taita Hills in south-east Kenya to the Makambako Gap in south-central Tanzania (Fig. 4.1 and Table 4.1). Around 500 vascular plant species are putatively endemic, of which over 80 are trees. Endemism amongst birds is also high (ICBP, 1992; Stattersfield *et al.*, 1998) and a number of mammals and amphibians are endemic or near-endemic (Burgess *et al.*, 2007; Poynton *et al.*, 2007). Preservation of this region is a priority for biodiversity conservation (Olson and Dinerstein, 1998; Brooks *et al.*, 2002) and crucial to Tanzania's population, for whom the forests provide ecosystem services such as water, electricity, building materials, medicine and revenue from tourism (Mwakalila *et al.*, 2009).

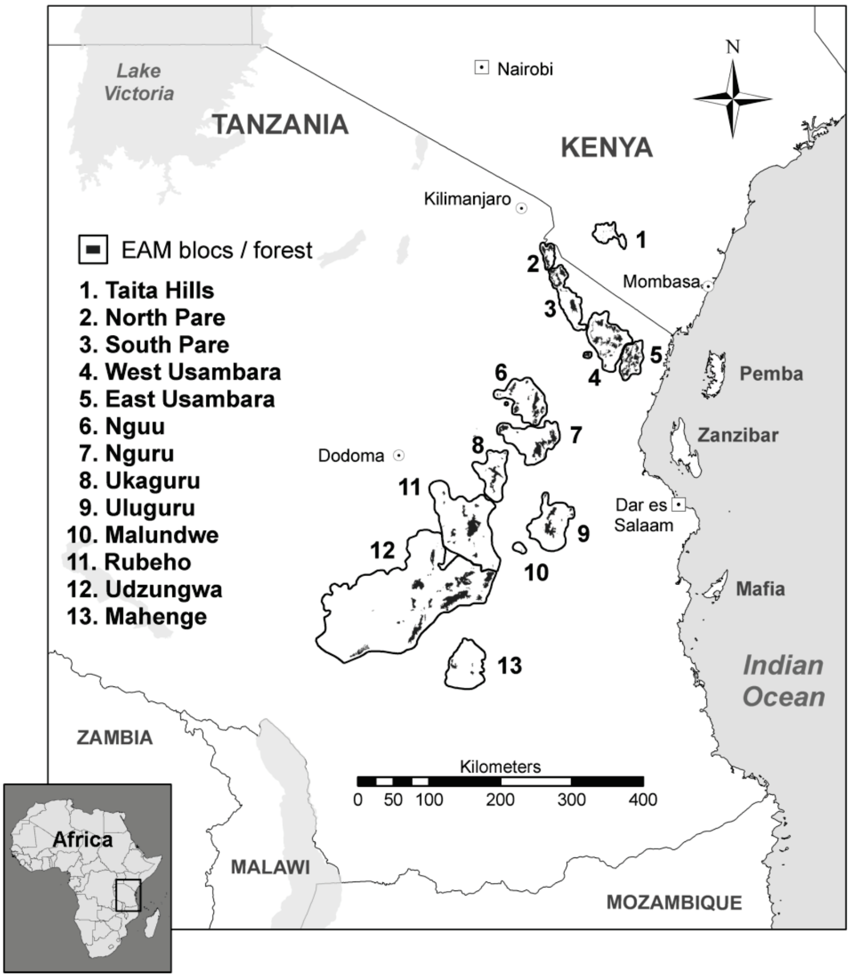


Figure 4.1. Map of the 13 mountain blocs that comprise the Eastern Arc chain, including forest cover at 1-ha resolution (see also Table 4.1). Projection (except inset) is UTM zone 37 south. Note that forest estimates and boundary placement pre-date Chapter 2, although divisions between blocs are consistent – the delineation pictured is a topologically simpler version of Fig. 2.4.

Plant inventory data

The plant database (*c*. 70,000 records) combines our own field data with two large datasets contributed by the Missouri Botanical Garden (http://www.tropicos.org) and Frontier-Tanzania (http://www.frontier.ac.uk). Botanical identifications were verified by herbaria (Royal Botanic Gardens, Kew, Missouri Botanical Garden, and the University of Dar es Salaam); nomenclature was standardised by reference to the African Flowering Plants Database (AFPD, 2009). Threatened and potentially threatened taxa were identified according to an ongoing assessment of the conservation status of the combined EAM and Coastal Forest flora (Gereau *et al.*, 2010). Endemism in the context of this article refers to taxa that have been found only in the EAMs at and above 500 m elevation. We modelled all taxa with records of occurrence in ten or more distinct 1 km or 2 km grid squares, favouring the higher resolution where specimen locality data allowed (Appendix 4A). The modelling subset targets 452 taxa in 90 plant families: 304 trees, 12 lianas, 62 shrubs and 74 herbs. Of these, 319 were modelled at 1 km resolution and 133 at 2 km resolution; 68 are threatened, and 25 are endemic.

Environmental data

Point patterns observed for our target taxa were regressed against twelve predictor variables, each representing an aspect of the environment thought to directly affect plant distributions in the EAMs (Tables 1 and 2). For temperature, we used interpolated climate surfaces based on records from the period 1950-2000 (Hijmans *et al.*, 2005). These data provide monthly temperature means and extremes at a spatial resolution of 1 km, from which we derived the annual mean and range, potential evapotranspiration (Thornthwaite, 1948) and an associated moisture index (annual rainfall / potential evapotranspiration). Rainfall grids were based on analysis of data from the Tropical Rainfall Measuring Mission (TRMM 2B31 combined PR, TMI profile): first, mean monthly 1 km gridded atmosphere rainfall was calculated from observations spanning the period 1997-2006 (Mulligan, 2006a); surface-received orographic rainfall was then modelled using wind velocity, slope, aspect and topographic exposure (Mulligan and Burke, 2005). Maximum water deficit represents the length and severity of the dry season and was calculated as the highest cumulative deficit in mean monthly rainfall, where a deficit is less than 100 mm month^{-1}. Estimates of cloud frequency were based on a 1 km climatology derived from the MODIS MOD35 Cloud Mask Product (Mulligan, 2006b).

Beside climate, we also considered topographic and edaphic factors. From a high-resolution (three arc-sec) digital elevation model (CIAT-CSI SRTM; Jarvis *et al.*, 2008), we derived gradient of the slope and two cosine transformations of slope aspect, the latter being oriented such that slopes facing towards prevailing winds (dry season, south-easterly; wet season, northerly) were allocated the highest values, and opposing slopes the lowest. Soil parameters were obtained from the Soil and Terrain Digital Database (SOTER) and include soil reaction (pH), cation exchange capacity and available water capacity (Batje, 2004).

Table 4.1. Forest area, including number of patches (> 1 km apart) and spatial variations in altitude, temperature and rainfall (mean values in parentheses). Estimates of forest cover in Tanzania are based on those of MNRT (1997), updated using expert knowledge and imagery from 2000 onwards by the Remote Sensing and GIS Laboratory, Sokoine University of Agriculture. Forests in Kenya were identified from SPOT multi-spectral satellite images (Clark and Pellikka, 2009). These data pre-date Chapter 2, in which estimated forest area is revised to 4346 km^2.

Mountain bloc (north to south)	Forest (km^2)	No. patches	Altitude (m)	Mean annual temperature (°C)	Mean annual rainfall (mm/year)
Taita Hills	7.0	14	1102 – 2208* (1585)	16 – 22 (19)	253 – 1208 (630)
North Pare	147.0	2	755 – 2099 (1274)	16 – 24 (20)	158 – 1677 (770)
South Pare	331.0	6	541 – 2454 (1384)	13 – 24 (20)	359 – 2947 (1100)
West Usambara	528.8	14	408 – 2294 (1365)	13 – 25 (18)	393 – 3126 (1005)
East Usambara	391.4	5	124 – 1484 (628)	17 – 26 (22)	529 – 2788 (1176)
Nguu	416.8	13	709 – 1998 (1232)	16 – 23 (20)	333 – 3543 (1243)
Nguru	471.8	7	350 – 2382 (1243)	14 – 26 (20)	222 – 3814 (1706)
Ukaguru	197.3	6	885 – 2259 (1693)	15 – 23 (18)	634 – 2352 (1537)
Uluguru	308.5	9	255 – 2636 (1691)	12 – 27 (18)	579 – 2352 (1482)
Malundwe	2.3	1	793 – 1259 (1054)	20 – 23 (21)	978 – 1469 (1132)
Rubeho	530.7	16	565 – 2334 (1700)	15 – 25 (18)	281 – 1415 (822)
Udzungwa	1673.2	32	278 – 2555 (1390)	13 – 26 (20)	388 – 2470 (1346)
Mahenge	70.5	3	347 – 1478 (749)	18 – 26 (23)	1100 – 3238 (1813)
All EAMs	5076.4	130	124 – 2636 (1352)	12 – 27 (20)	158 – 3814 (1257)

* Pellikka *et al.* (2009)

Table 4.2. Environmental predictor variables used for modelling plant distributions. Correlation matrix shows Pearson coefficients (1 km resolution, bloc extent plus 25 km buffer to include all data points); bold values indicate highly correlated variables that were separated prior to model selection. Spearman *rho* correlations were similar, as were those calculated at 2 km resolution. Far right columns summarise the contribution of predictors in explanatory models (forward-backward selection): times chosen and median decrease in explained deviance with predictor removed (↓D^2).

	Environmental predictor	1	2	3	4	5	6	7	8	9	10	11		All taxa		Endemics	
														Chosen	↓D^2	Chosen	↓D^2
1	Mean annual temperature												[	149	0.13	15	0.16
2	Potential evapotranspiration	**0.96**												112	0.16	12	0.17
3	Annual temperature range	-0.42	-0.52											197	0.12	18	0.16
4	Annual moisture index	-0.43	-0.36	-0.19										141	0.13	19	0.16
5	Maximum water deficit	-0.09	-0.18	0.43	-0.58									121	0.10	13	0.04
6	Cloud frequency	0.39	0.44	-0.61	0.31	-0.56								130	0.11	16	0.15
7	Soil: pH	-0.07	-0.07	0.27	-0.11	0.20	-0.18							64	0.16	9	0.19
8	Soil: cation exchange capacity	0.09	0.10	0.18	-0.19	0.15	-0.02	0.49						118	0.11	12	0.15
9	Soil: available water capacity	-0.01	0.00	0.03	0.00	-0.04	-0.02	0.19	0.45					106	0.10	15	0.13
10	Slope: angle from horizontal	-0.32	-0.33	0.00	0.39	-0.23	0.06	0.03	-0.13	0.06				107	0.09	15	0.06
11	Slope: orientation, northness	-0.02	-0.01	-0.03	-0.04	0.07	-0.13	0.02	0.04	0.05	-0.05		[	73	0.09	9	0.08
12	Slope: orientation, south-eastness	0.10	0.11	-0.05	0.04	-0.09	0.22	0.01	0.03	-0.04	0.02	**-0.72**		50	0.10	11	0.11

Model calibration

Spatial data were projected to UTM zone 37 south and resampled to 1 km or 2 km, depending on the taxon. Observed distributions were related to environmental predictors using generalised additive models (GAMs), calibrated using logit link functions and binomial error terms and allowing between one and four degrees of freedom for smoothers (Yee and Mitchell, 1991). For statistical calculations and the manipulation of map layers, we used R 2.10.0 (R-Development-Core-Team, 2009) and GRASS GIS 6.3.0 (GRASS-Development-Team, 2009).

Background data

As is often the case when working with plot and herbarium data, ground-truthed absences were not available. Instead, we generated pseudo-absence (background) data to constrain the models. Because presence localities were spatially biased, it was appropriate to impose similar bias on the background data (Phillips *et al.*, 2009). In a previous application of this approach, we targeted pseudo-absences for EAM tree species towards locations known to have been surveyed using similar methods (Chapter 3). Here, we extend this methodology to consider separately the four different growth forms of plants – tree data are more plentiful than herb data, for example, not because tree species are necessarily more abundant but because vegetation plot assessments (*c*. 70% of our data) often target plants of a minimum size (e.g., ≥ 10 cm diameter at breast height *c*. 1.3 m). Thus, background data were placed only in locations where a matching growth form of plant has been sampled in the past (excluding presence sites for that taxon), using a ratio of five absence points for every presence point (Appendix 4B).

Predictor selection

Two pairs of predictors were strongly collinear: mean annual temperature *vs*. potential evapotranspiration, and aspect north *vs*. aspect south-east (Table 4.2). These were reduced prior to modelling by constructing additive models separately for each taxon-predictor pair and retaining whichever yielded the strongest prediction. Minimal predictor sets were then identified using forward–backward selection, beginning with a null model and adding or removing terms iteratively according to Akaike Information Criterion. Next, we sought alternative solutions using backward–forward selection, beginning with a full model and removing or adding terms according to Bayesian Information Criterion. In each case, the most powerful predictive model was selected by cross-validating the area under the receiver-operating characteristic curve (AUC) – a threshold independent measure that incorporates both type I and type II error rates (Green and Swets, 1974). We used a five-fold cross-validation procedure (80:20 training:testing split) stratified with respect to prevalence and averaged over ten independent runs (Parker *et al.*, 2007). These 'best-model' solutions were combined in performance-weighted averages to give multimodel estimates of occurrence.

Spatial autocorrelation

A common problem with using regression techniques in ecology is that environmental variables are rarely sufficient to explain fully spatial dependence in species data (Dormann *et al.*, 2007; Miller *et al.*, 2007). Consequently, model residuals exhibit spatial structure, violating the statistical assumption that they are independent and identically distributed. Spatial autocorrelation in model predictions was parameterised by appending autocovariate terms to the GAM formulae (Augustin *et al.*, 1996):

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 \operatorname{cov}_1 + \ldots + \beta_n \operatorname{cov}_n + \beta_{n+1} auto\operatorname{cov}_i$$

where p_i is the probability of occurrence in focal cell i, and $auto\mathrm{cov}_i$ is a distance-weighted average of occurrence probabilities in surrounding grid cells (neighbourhood size = 10 km). There is a risk, however, that autocovariate models may underestimate environmental controls on species distributions, resulting in less stable predictions (Dormann, 2007a; Chapter 3). Autocovariate terms were therefore retained if and only if they improved predictive performance on unseen data (five-fold AUC).

Testing and validation

In addition to the measures of model performance employed during calibration, final model predictions were further validated using a fully independent test set. These presence data were omitted from calibration because of low or uncertain spatial accuracy but remained useful for gauging the sensitivity of predictions, and in particular the ability of models to predict occurrence in novel mountain blocs, i.e., those within a plant's documented range but that were not represented in the presence data for that taxon. Test data accurate to *c.* 2 km were available for 286 taxa (1956 records); data with lower spatial accuracy were available for 341 (1578) and were assumed accurate only at the mountain bloc resolution.

The extent to which sampling distributions captured the range of environmental conditions in EAM forests was investigated using envelope uncertainty maps – spatial representations of where and to what extent particular models were extrapolated beyond the niche-breadth of the training data (Appendix 4C; see also Chapter 3).

Richness estimates

Plant richness was calculated by summing maps of estimated presence-absence over all taxa in a target group (e.g., trees or endemics). Distribution models predicted occurrence on a continuous scale, from 0 to 1; these predictions were dichotomised using taxon-specific occurrence thresholds, chosen by maximising the sum of sensitivity and specificity (Appendix 4B).

Because of uncertain colonisation histories, we produced three versions of each richness map. First, model predictions were extrapolated to all forested grid squares, regardless of location. Richness maps derived from these estimates are tentative predictions, because they assume no historical barriers to dispersal. Second, models were extrapolated only to mountain blocs within a plant's documented range. Derived richness is less speculative but biased by the level of research. Third, we map the disparity between predicted and confirmed richness, giving an indication of which areas should be prioritised for future exploration.

Results

Model performance

According to validation statistics, models performed well and were rarely forced to extrapolate far beyond the niche-breadth used for calibration (Table 4.3 and Appendix 4C). The balance of errors favoured correctly predicted presences (higher sensitivity), which is preferable because presence locations have been ground-truthed whereas background data are likely to contain genuine misclassifications. Even so, fully independent tests revealed that models for endemic taxa often failed to predict known occurrences accurately (median error = 4.24 km), especially in blocs beyond the spatial range of training data (Table 4.3). The sensitivity of novel-bloc predictions was also comparatively low for threatened taxa.

When training data were reused for testing, models calibrated at 2 km resolution outperformed those calibrated at 1 km resolution, but for unseen data 1 km models were significantly better (five-fold AUC, $P < 0.001$; Wilcoxon rank sum). The pattern was similar across growth forms, but only significant for trees. Tree models were particularly stable, retaining significantly more of the AUC under cross-validation than models for lianas, shrubs or herbs (Appendix 4C).

Table 4.3. Summary of model performance: explained deviance (D^2), area under the receiver-operator characteristic curve (AUC) including a five-fold cross-validation, and the proportion of presences (sensitivity) and pseudo-absences (specificity) classified correctly. Figures shown are median values because of negative skew. Using high-resolution independent test data (*c.* 2 km accuracy), we present the median distance to the nearest predicted occurrence (km). Using all available test data (bloc-level accuracy), we assess the ability of models to predict occurrence successfully in novel mountain blocs (those with no presence points in the training data): mean sensitivity ± one standard deviation (medians = 1). See also Appendix 4C.

							Independent test data	
	N	D^2	AUC*	5-fold AUC	Sens.	Spec.	Distance to occurrence	Novel-bloc sensitivity
Trees	302	0.66	0.95	0.87	0.94	0.87	1.00	0.91 ± 0.25
Lianas	12	0.60	0.94	0.82	0.90	0.91	0.00	1.00 ± 0.00
Shrubs	62	0.67	0.95	0.83	0.94	0.92	0.00	0.92 ± 0.25
Herbs	74	0.62	0.93	0.79	0.93	0.89	0.00	0.94 ± 0.20
Endemic	25	0.73	0.98	0.89	1.00	0.89	4.24	0.87 ± 0.26
Threatened	68	0.71	0.96	0.88	0.99	0.91	1.00	0.85 ± 0.30
All species	452	0.65	0.95	0.86	0.94	0.89	1.00	0.92 ± 0.24

* AUC: 0.5-0.7, better than chance; 0.7-0.9, good performance; 0.9-1.0, excellent performance (Swets, 1988)

The two alternative stepwise models frequently returned different solutions (21% agreement), but predictive performance was similar. On average, forward–backward models were smaller than backward–forward models (mean number of predictors = 3 and 4, respectively) and so were preferred for inferring causal relationships (Table 4.2). Temperature variables were the most often selected, reflecting the importance of altitude in determining species distributions in mountainous regions. Predictors of moisture availability, including cloud frequency, were also important, as were slope orientation and cation exchange capacity. The least selected predictor was soil acidity, although it contributed highly when included (Table 4.2). Response shapes for soil variables were not always sensible, indicating that they captured broad geographical patterns rather than functional relationships (see also Appendix 4E).

Spatial autocovariates were retained in 30% of cases, more often in larger (backward–forward selection) and more stable (1 km) models. The median increase in explained deviance was only 6%, so environmental constraints were well represented alongside spatial terms.

Sampling bias

Bias in exploration history was quantified by survey intensity, which we calculated at bloc level using all available data. The East Usambaras and Udzungwas are by far the best-researched blocs, each with 20,000-30,000 data points. There is a steep drop to the Ulugurus and West Usambaras (6000-8000), followed by Nguru and Rubeho (3000-4000), South Pare, Mahenge then Ukaguru (1000-3000). The Taita Hills, North Pare, Nguu and Malundwe have fewer than 500 records amongst them. Tree species dominate the data, accounting for over 80% of specimens in most blocs (60% in Taita and South Pare); the remainder is mainly shrub and herb records, with lianas accounting for less than 5%.

The relationship between the numbers of modelled taxa observed in each mountain bloc and the number predicted to have potential niche-space was highly significant (Fig. 4.2a), reflecting both genuine biogeographical patterns and spatial bias in exploration history. Survey intensity explained 89% of the deviance in observed plant richness (log-linear relationship). The fit was lower for predicted richness (66%) with a shallower gradient, but still highly significant; Malundwe Mountain was an outlier with models predicting fewer taxa than expected (Cook's distance = 1.2).

For species of conservation concern, the fit was stronger for predicted richness than for observed richness, and the gradient of the slope remained comparatively steep (Fig. 4.2c). This may be a consequence of non-climatic factors such as isolation: survey intensity and environmental correlates predict similar richness in Rubeho and South Pare, yet observed richness is very different. Combined with lower performance in independent tests (Table 4.3), we find endemic and narrow-ranged taxa to be particularly sensitive to sampling bias.

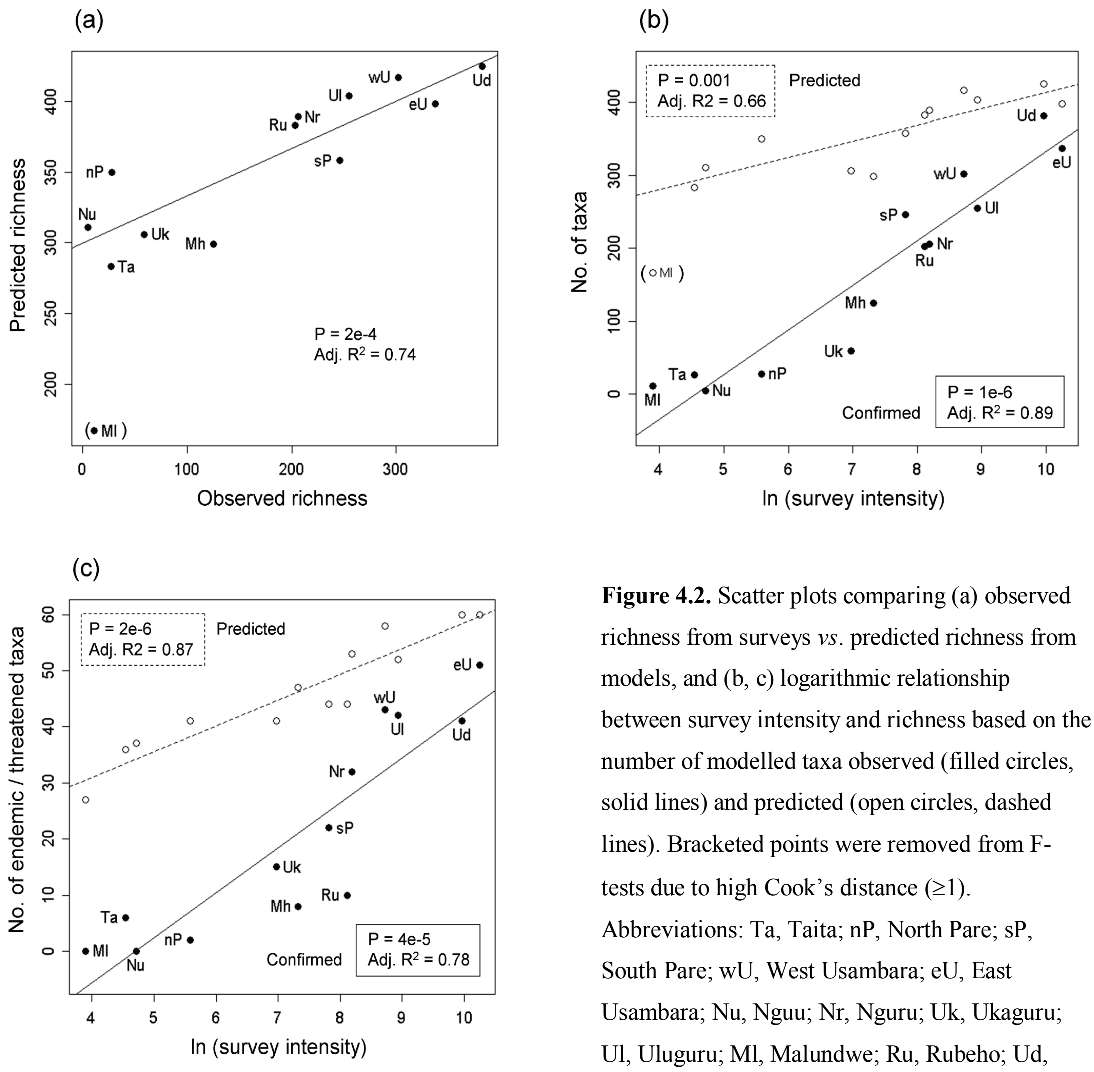


Figure 4.2. Scatter plots comparing (a) observed richness from surveys *vs*. predicted richness from models, and (b, c) logarithmic relationship between survey intensity and richness based on the number of modelled taxa observed (filled circles, solid lines) and predicted (open circles, dashed lines). Bracketed points were removed from F-tests due to high Cook's distance (≥1). Abbreviations: Ta, Taita; nP, North Pare; sP, South Pare; wU, West Usambara; eU, East Usambara; Nu, Nguu; Nr, Nguru; Uk, Ukaguru; Ul, Uluguru; Ml, Malundwe; Ru, Rubeho; Ud, Udzungwa; Mh, Mahenge.

Richness

Confirmed at bloc level

Extrapolating predictions within but not between mountain blocs, Fig. 4.3a shows a clear bias towards better-studied regions, especially the East Usambaras and Udzungwas. Localised richness was also high in parts of South Pare, Uluguru and Rubeho. Average richness across grid cells in West Usambara was comparatively low given that it ranked second at the bloc resolution (modelling subset, Table 4.4). Fig. 4.4 shows that many taxa in this bloc were not predicted to be widespread in larger forests, suggesting high species turnover. The same may be true of Nguru, which is also ranked higher than the 1 km map suggests (*cf.* Fig. 4.3a and Table 4.4). In South Pare, richness was concentrated mainly in Chome Forest Reserve, reflecting a bias in collection localities.

Endemic and threatened taxa were most prevalent across grid cells in the Uluguru and Usambara Mountains (Fig. 4.3b), with the South Pares and parts of Udzungwa also important. Compared with overall richness, relative concentrations were higher in Nguru and Ukaguru, and lower in Rubeho and Udzungwa, although the bloc total for Udzungwa was still high (ranked fourth in Table 4.4). In Table 4.5, we provide details of 18 taxa that are both endemic to the EAMs and threatened with extinction, including area-based recommendations for the IUCN Red List.

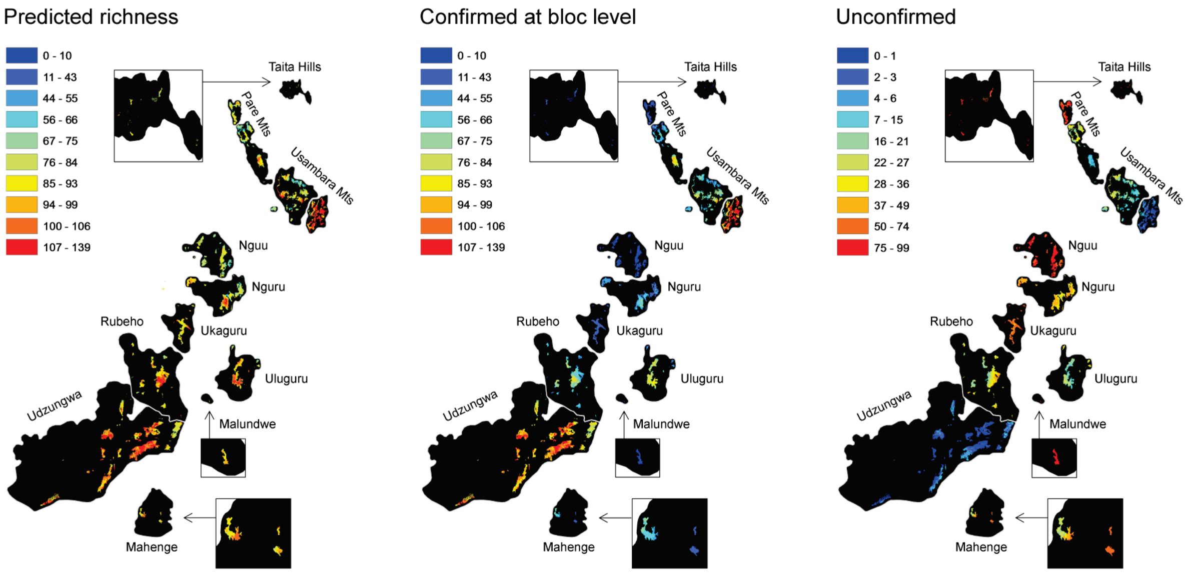
Predictive estimates

Predicted richness was greater than observed richness in all cases, with the size of the disparity showing a negative correlation with survey intensity (Fig. 4.2b-c). Unconfirmed but potentially suitable habitat was therefore most common in Taita, North Pare, Nguu and Malundwe. Environmental conditions in Nguru, Ukaguru, Rubeho and Mahenge suggest these areas could also support higher concentrations of species than currently documented (Figs 4.3 and 4.4).

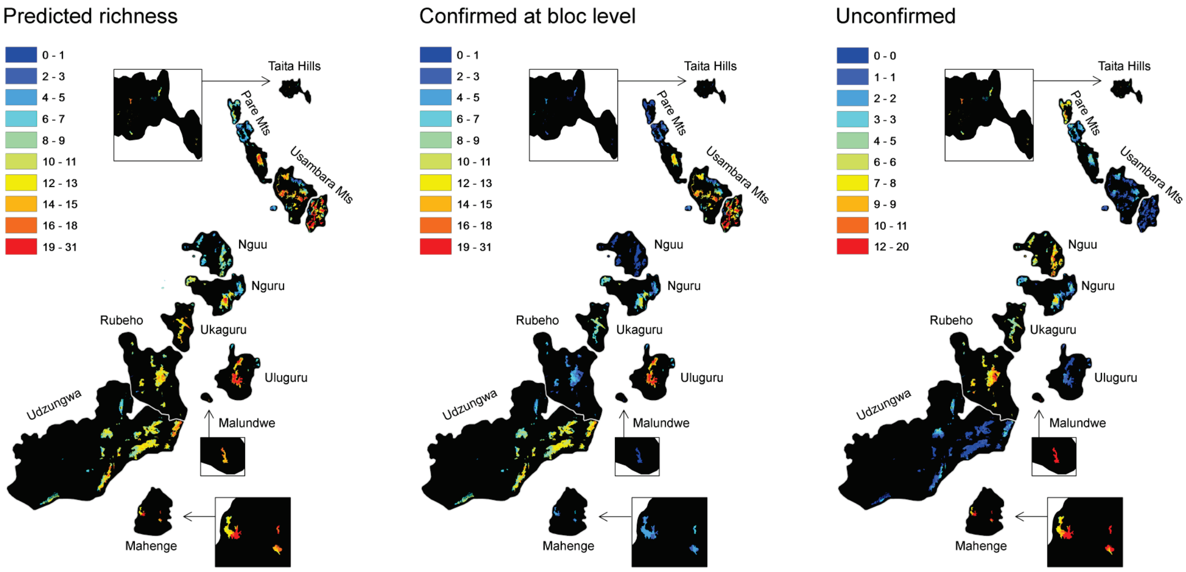
Compared with observed richness at the bloc level, predicted richness ranked the Nguru and Rubeho Mountains above South Pare (Table 4.4). Also, North Pare and Nguu were ranked above Ukaguru and Mahenge despite sampling bias in favour of the latter. Predictive rankings for endemic and threatened taxa followed a similar pattern, except that the Ulugurus were ranked slightly lower, and the importance of Mahenge is predicted to be higher than inventory data suggest (Table 4.4 and Appendix 4D).

Growth form

Fig. 4.5 shows patterns of richness to be similar across growth forms, with the notable exception that tree richness is highest in the two most researched mountain blocs (East Usambara and Udzungwa), whereas lianas, shrubs and herbs have equally high (confirmed) or higher (predicted) richness in other areas, particularly the West Usambara and Rubeho Mountains.



(a) All taxa



(b) Endemic and / or threatened taxa

Figure 4.3. Spatial estimates of plant richness calculated across (a) all taxa and (b) taxa of conservation concern. Scale bars show the number of taxa predicted to have potential niche-space in 1 km grid squares. In the left panel, modelled distributions are extrapolated to all forest patches with suitable environmental conditions. In the centre panel, predictions are restricted to just those mountain blocs where the respective taxa have been confirmed present. The right panel shows predictions of occurrence in unconfirmed blocs (left panel minus centre panel) – we suggest this map can be helpful in selecting future sites for exploration.

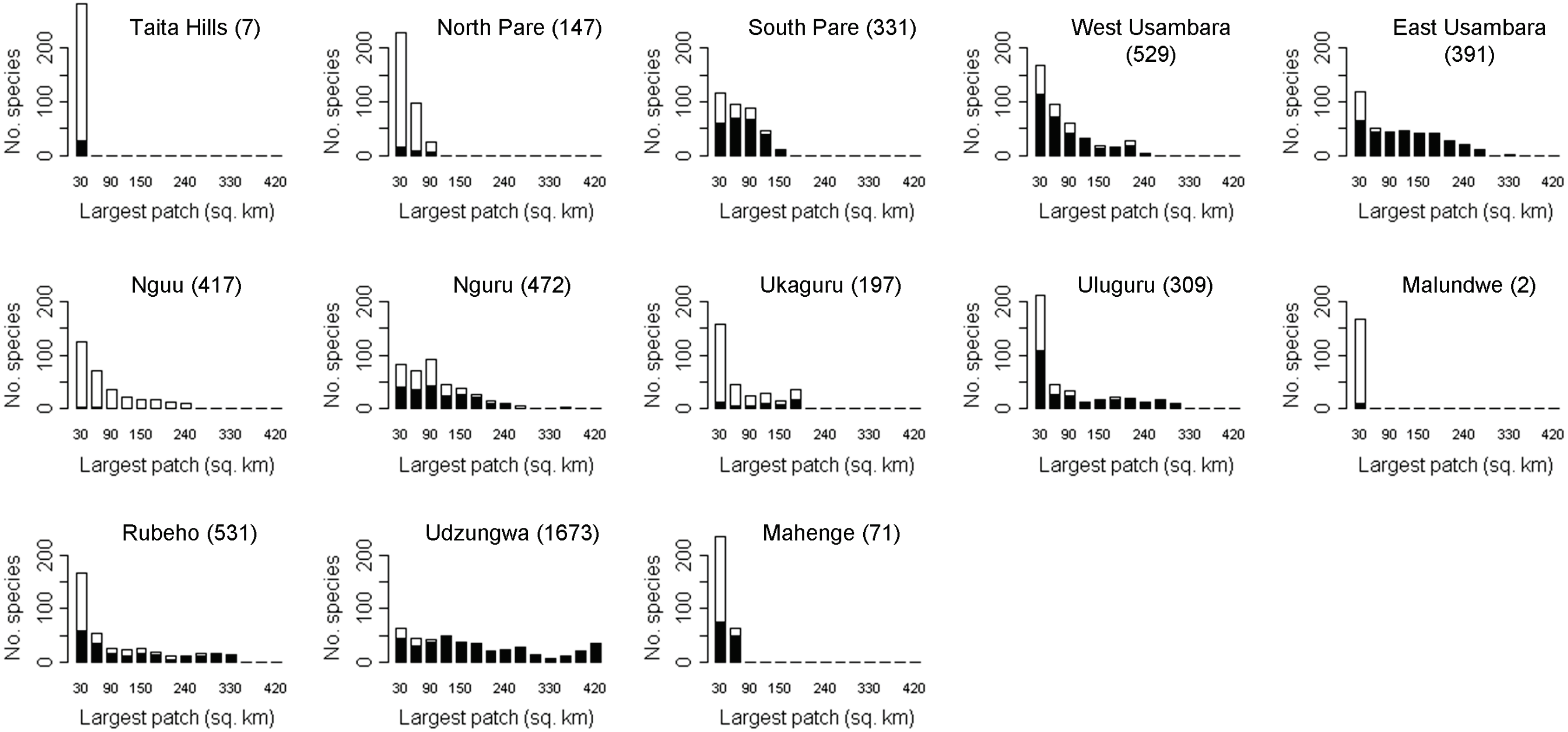


Figure 4.4. Histograms showing patterns of within-bloc patch occupancy. Horizontal axes represent the largest contiguous area of forest providing environmentally suitable conditions for a particular taxon. Vertical axes show the number of taxa in each 30 km^2 patch size interval. Filled bars relate to confirmed occurrence at the bloc level; open bars relate to predictions of occurrence in novel mountain blocs. Total forest area in each bloc shown in parentheses (km^2). Patterns for endemic/threatened taxa are presented in Appendix 4D.

Table 4.4. Conservation priorities, based on the number of plant taxa confirmed or predicted in each mountain bloc. These rankings are not corrected for forest area, and therefore favour larger mountain blocs such as Udzungwa. See Figs 4.3 and 4.4 for fine-scale richness estimates.

	All taxa						Endemic and/or threatened taxa					
Rank	Full inventory		Modelled taxa (confirmed)		Modelled taxa (predicted)		Full inventory		Modelled taxa (confirmed)		Modelled taxa (predicted)	
1	Udzungwa	2546	Udzungwa	382	Udzungwa	425	Udzungwa	319	E. Usambara	51	E. Usambara	60
2	E. Usambara	1108	W. Usambara	337	W. Usambara	417	Uluguru	233	W. Usambara	43	Udzungwa	60
3	South Pare	894	E. Usambara	302	Uluguru	404	E. Usambara	187	Uluguru	42	W. Usambara	58
4	Uluguru	835	Uluguru	255	E. Usambara	398	W. Usambara	162	Udzungwa	41	Nguru	53
5	W. Usambara	713	South Pare	246	Nguru	389	Nguru	159	Nguru	32	Uluguru	52
6	Rubeho	665	Nguru	206	Rubeho	383	South Pare	75	South Pare	22	Mahenge	47
7	Nguru	658	Rubeho	203	South Pare	358	Rubeho	58	Ukaguru	15	South Pare	44
8	Mahenge	583	Mahenge	125	North Pare	350	Mahenge	58	Rubeho	10	Rubeho	44
9	North Pare	108	Ukaguru	59	Nguu	311	Ukaguru	39	Mahenge	8	North Pare	41
10	Ukaguru	103	North Pare	28	Ukaguru	306	Taita Hills	23	Taita Hills	6	Ukaguru	41
11	Taita Hills	57	Taita Hills	27	Mahenge	299	North Pare	13	North Pare	2	Nguu	37
12	Malundwe	31	Malundwe	11	Taita Hills	283	Nguu	3	Nguu	0	Taita Hills	36
13	Nguu	29	Nguu	5	Malundwe	167	Malundwe	1	Malundwe	0	Malundwe	27

Table 4.5. Model estimates of the habitat available for 18 plant taxa endemic to the Eastern Arc Mountains, presented in descending order of rarity. Also shown are the current IUCN Red List designations (no designation for eight taxa; IUCN, 2009), the proposed Red List status of each taxon in an ongoing assessment of plant conservation in East Africa (Gereau *et al.*, 2010), and recommendations based solely on environmentally suitable habitat in mountain blocs where taxa are known to occur: critically endangered (CR), area of occupancy < 10 km^2; endangered (EN), area of occupancy < 500 km^2; vulnerable (VU), area of occupancy < 2000 km^2 (see also Hall *et al.*, 2009). Note that areas of occupancy are not the only consideration in determining the level of threat: *Englerodendron usambarense* has a very narrow range but is proposed as Not Threatened because it is well protected within Amani Nature Reserve. For full details of Red List categories and criteria, visit http://www.iucnredlist.org/.

Endemic species or infra-specific taxon	Growth form	Mountain bloc coverage	Suitable habitat (km^2)	Current IUCN listing (version 2.3 or 3.1)	Proposed threat status	Area-based Recommendation
Cynometra longipedicellata	tree	eU	132	VU [B1+2b], ver. 2.3	CR	EN
Englerodendron usambarense	tree	eU	156	VU [B1+2c], ver 2.3	NT	EN
Mammea usambarensis	tree	sP, wU	157	VU [B1+2b], ver. 2.3	VU	EN
Allophylus melliodorus	tree	wU, eU, Nr	214		PT	EN
Eugenia toxanatolica	tree	sP, wU, Mh	233		PT	EN
Cola usambarensis	tree	eU	243		PT	EN
Mussaenda microdonta subsp. *microdonta*	tree	wU, Nr, Ul	295	VU [B1+2b], ver. 2.3	PT	EN
Memecylon cogniauxii	shrub	sP, wU, eU, Nr, Ul	302		PT	EN
Casearia engleri	tree	sP, wU	328	VU [B1+2b], ver. 2.3	VU	EN
Syzygium micklethwaitii	tree	sP, wU, Nr, Uk, Ul	468		PT	EN
Coffea mongensis	tree	wU, eU, Nr, Ud	535	VU [B1+2b], ver. 2.3	LC	VU
Impatiens palliderosea	herb	Uk, Ul, Ru	543		VU	VU
Craterispermum longipedunculatum	tree	Ul, Ud	712	VU [B1+2b], ver. 2.3	PT	VU
Lasianthus pedunculatus	tree	Nr, Uk, Ul, Ud	867	VU [B1+2b], ver. 2.3	PT	VU
Zenkerella capparidacea	tree	wU, eU, Nr, Ul	872		VU	VU
Polyscias stuhlmannii	tree	sP, wU, Uk, Ul	933	EN B2ab(iii), ver. 3.1	EN	VU
Dicranolepis usambarica	tree	Ta, sP, wU, eU, Nr, Ul, Ud	996		PT	VU
Allanblackia ulugurensis	tree	Nr, Uk, Ul, Ud	1133	VU [B1+2c], ver. 2.3	VU	VU

CR, critically endangered; EN, endangered; VU, vulnerable; NT, near threatened; LC, least concern; PT, potentially threatened but not yet evaluated; B1, restricted extent of occurrence; 2b, area of occupancy continuing to decline; 2c, extent and/or quality of habitat declining; Ta, Taita; nP, North Pare; sP, South Pare; wU, West Usambara; eU, East Usambara; Nu, Nguu; Nr, Nguru; Uk, Ukaguru; Ul, Uluguru; Ml, Malundwe; Ru, Rubeho; Ud, Udzungwa; Mh, Mahenge

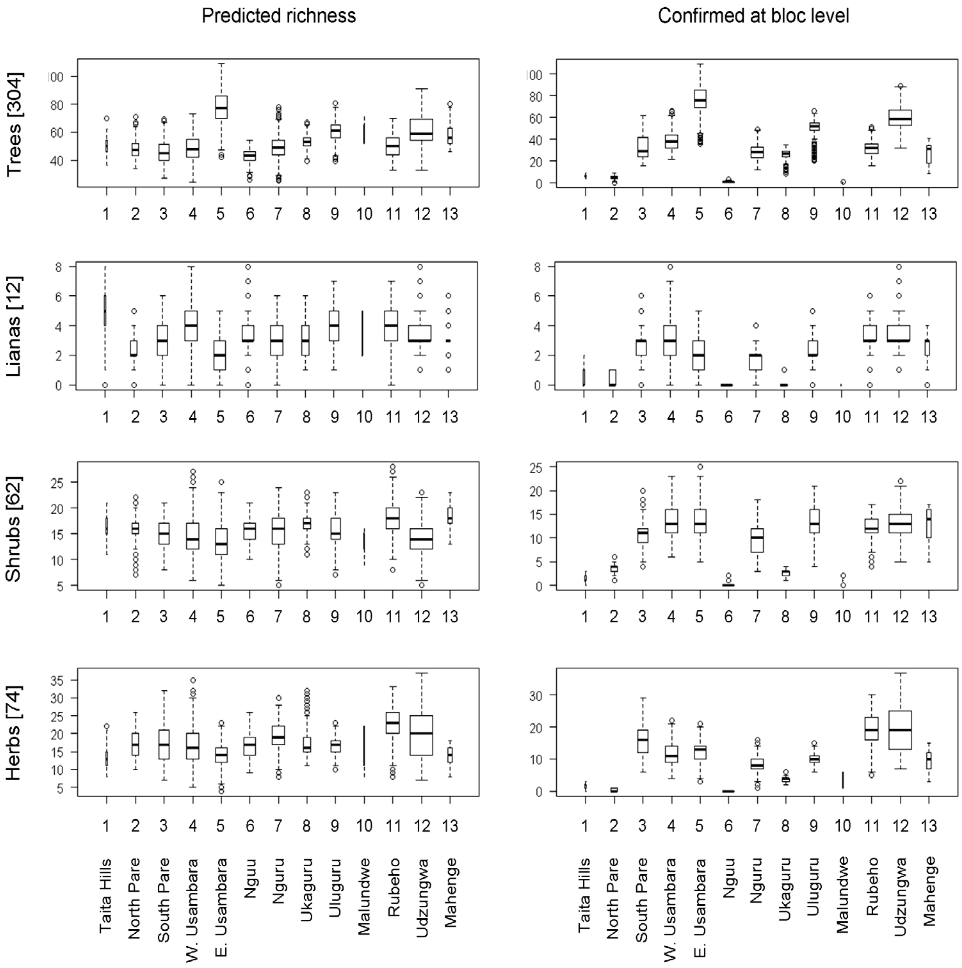


Figure 4.5. Box plots detailing how plant richness varies according to growth form. In the left panel, modelled distributions are extrapolated to all forest patches with suitable environmental conditions. In the right panel, predictions are restricted to just those mountain blocs where the respective taxa have been confirmed present. Box widths are proportional to the area of forest remaining in each mountain bloc.

Discussion

The prioritisation of areas for conservation within the EAMs has tended to change with the availability of new field data. First, the Usambaras and Ulugurus were ranked most important; subsequently, the importance of Udzungwa was recognised, followed by Nguru and now Rubeho (CEPF, 2003; Doggart *et al.*, 2006). This reshuffling of conservation priorities is symptomatic of a paucity of survey data common to many high biodiversity regions and highlights the need for strategically targeted field sampling. Distribution models are an appealing tool for obtaining high-resolution estimates of richness in well-researched areas, and tentative estimates of conservation importance elsewhere. Alongside other considerations such as threats to habitat, richness in other taxonomic groups and ecosystem value (e.g., carbon stocks, hydrology, natural resources, ecotourism; Naidoo *et al.*, 2008), they could form part of a more consistent approach to conservation priority setting and strategic planning of surveys.

In many cases, the data available for modelling are biased both in geographical space and towards particular groups of organisms. Here, tree data were the most abundant and tree models the most stable. Our results suggest that if the bias were towards lianas, shrubs or herbs, instead of trees, then we might favour the mountain blocs in a slightly different order. Faced with insufficient data, conservation planners must determine the degree to which different taxonomic groups and growth forms can serve as surrogates for each other in the prioritisation of areas for conservation (Burgess *et al.*, 2006). We find that even within the group of vascular plants, it is preferable to consider all growth forms in the analysis of conservation priority. Low levels of congruence have also been reported for vertebrates (Grenyer *et al.*, 2006) and when comparing patterns of endemism across a range of taxonomic groups (Kremen *et al.*, 2008).

Because of broad-scale geographical bias in the occurrence data, coupled with uncertain colonisation histories, we have been careful to distinguish between those mountain blocs where a taxon is known to occur and those where it is to-date undocumented. When dispersal limitations are not considered, models predict that richness could be more evenly distributed across the mountains than is currently documented (Fig. 4.3). In the 2003 Ecosystem Profile of the EAMs and Coastal Forests (CEPF, 2003), the Usambaras, Ulugurus and Udzungwas were identified as being the most species-rich blocs. Predictive estimates largely confirm this ranking, whilst indicating that the importance of Nguru and Rubeho may still be underestimated, particularly for rare species (see also Doggart *et al.*, 2006). Lesser-researched blocs, especially North Pare and Nguu, could also be important, following higher rankings despite low survey intensity. Predictions such as these could be verified and subsequently refined by ongoing and targeted field assessments (Guisan *et al.*, 2006).

Using independent test data, we found that models were generally quite successful at predicting occurrence in novel mountain blocs. These validations were, however, limited to bloc-level sensitivity, so the extent of over-prediction remains uncertain. Models for threatened and endemic taxa were most likely to under-predict when extrapolated into novel blocs, indicating gaps in the documented environmental niche. This could be a problem for wider-ranging taxa too, for it is difficult to know whether or not the complete range of conditions under which a taxon exists has been sampled. Further, we suspect that in some cases the soil predictors, which vary broadly by mountain bloc, simply identified spatial biases in the sampling distribution, rather than truly casual factors. Given the sensitivity of predictions to survey intensity and the fact that realised distributions of EAM endemics are highly dependent on past connectivity, we caution that it is for the taxa of highest conservation concern that predictive estimates are most uncertain.

Restricting analyses to confirmed blocs only, we find that environmental conditions across most forests in Udzungwa have potential to support large numbers of plant taxa; concentrations of rare and endemic taxa, meanwhile, are predicted to be lower than in the Usambaras and Ulugurus – possibly a real pattern given the close proximity of non-EAM habitats. Mahenge is predicted to be suitable for many of the rare plants modelled here, but occurrence is unconfirmed in most cases. The Usambaras and Ulugurus are better known centres of endemism (Iversen, 1991; Temu and Andrew, 2008), promoted by geographical isolation and exposure to rain bearing ocean winds. High levels of endemism have also been recorded in the Taita Hills (Beentje, 1988; Beentje, 1994); however, this bloc is not well represented in our database, leading models to under-estimate its importance. Forests in Taita are of particular conservation concern, having been reduced to just a few remnant patches (Rogo and Oguge, 2000; Pellikka *et al.*, 2009).

Human activity has resulted in widespread fragmentation and degradation of many tropical forests, yet modelled estimates of diversity often do not consider the minimum forest area required for species persistence, nor the vulnerability of small fragments to degradation. Here, we map forest cover using remotely sensed land cover data. Whilst these estimates are not without error, they can at least be indicative of potential threats. We show that many taxa, especially those predicted to occur in blocs beyond their documented range, have suitable conditions only in relatively small forest patches (Fig. 4.4). Species across many taxonomic groups are less likely to persist in smaller and more isolated habitats, even if environmental conditions are favourable (MacArthur and Wilson, 1967; Lomolino, 2000; also Marshall *et al.*, 2010 in Appendix I). Around one fifth of the forests we identify from the land cover map are both smaller than 1 km^2 and more than 1 km from another patch. Much of this fragmentation is relatively recent, so in many cases the extinction debt has yet to be realised. In less isolated fragments, long-term persistence might be possible *via* seed recruitment from neighbouring populations (Lehouck *et al.*, 2009) – it is therefore imperative to conserve forests of all sizes to maintain connectivity (Fjeldså and Lovett, 1997a). Although not considered here, there is scope to address such patch dynamics post hoc by linking predicted distributions with spatially explicit population models (Keith *et al.*, 2008).

Exacerbated by forest loss, the extinction risk for narrow-range endemics is considerable (Brooks *et al.*, 2002). The tree *Platypterocarpus tanganyikensis* Dunkley & Brenan was once found in the West Usambara Mountains, but collections show no record of its presence since 1953, even before high levels of forest clearance in the 1960s. Distribution models for rare species require particular scrutiny, but as part of a wider assessment they can be useful for indicating the appropriate level of threat on the IUCN Red List (Table 4.5). One of the rarest endemics modelled here is the tree *Cynometra longipedicellata* Harms, known only from the East Usambaras. Models identify potential niche-space in Mahenge, but this species is more likely endemic to north-eastern Tanzania. We estimate the area of occupancy to be *c*. 132 km^2, probably less given competition for niche-space and other factors beyond the scope of our models (Pulliam, 2000). Based on the tree's observed altitudinal range, Hall *et al*. (2009) estimate that *C. longipedicellata* may have only 70 km^2 of habitat remaining, a decrease of over 70% since 1955. This species is currently assessed as Vulnerable (IUCN, 2009); we recommend elevating the threat status to Endangered, EN B1ab(iii) + B2ab(iii) (extent of occurrence < 5000 km^2, area of occupancy < 500 km^2, extent and/or quality of habitat declining) or Critically Endangered, CR B1ab(iii) (extent of occurrence < 100 km^2).

Patterns of endemism are often complex (Jetz *et al.*, 2004). Our perceptions of these patterns and our ability to identify causal factors are likely to be influenced by the spatial resolution used for modelling (Whittaker *et al.*, 2001; Rahbek, 2005). We find that higher resolution models are more stable, presumably because micro-climatic conditions are better represented. High levels of endemism in the EAMs have been attributed to historical isolation coupled with long-term climatic stability, with persistent orographic rainfall and mist having minimised climatically linked extinctions (Fjeldså *et al.*, 1997; Fjeldså and Lovett, 1997b). Recent pollen analyses confirm that whilst there were shifts in abundance, few if any plant taxa were lost during the last glacial maximum (Mumbi *et al.*, 2008; Finch *et al.*, 2009). Analysis of model predictions also suggests that moisture is a key driver for concentrations of endemism, with the annual moisture index explaining 31% of deviance across forested grid squares (Appendix 4E). Similarly, other studies have found contemporary rainfall to be a good predictor of endemism in the EAMs (Fjeldså and Lovett, 1997b) and of range-size rarity in West Africa (Holmgren and Poorter, 2007). Cloud cover explains little of the spatial variation in endemism but was an important predictor for some of the rarest plants (e.g., *C. longipedicellata*). The correlation between cloud frequency and overall richness was higher (13% explained deviance), with frequencies over 50% promoting climatic suitability for the most taxa (Appendix 4E). Annual temperature range was the best climatic predictor of modelled richness (24%), with lower seasonality correlating with higher diversity. Given the importance of the moisture index, these results suggest that measures of seasonal constancy in the water balance might be worth including in future studies.

Conclusions

The application of distribution models to plant inventory data can provide useful indications of which areas may be important for biodiversity conservation, and offers a means to estimate the niche-space available for species of conservation concern. Whilst models are highly sensitive to spatial bias in the inventory data, especially for rare species, we suggest that predictive definitions of conservation priority could be systemically improved by targeting field sampling towards locations with large discrepancies between observed and predicted diversity. As improvements in data quality cease to increase model stability, the limits of environmental controls on species' distributions will become clearer, providing a baseline by which to quantify the roles of historical and non-climatic factors in shaping contemporary patterns of biodiversity. Our results indicate that it is necessary to consider all growth forms of plants in the prioritisation of sites for conservation, and so we draw attention to the sometimes-excessive dominance of tree species in botanical inventories.

Acknowledgements

Many thanks to all who contributed to the species database, including field researchers at Frontier-Tanzania, data providers of the Missouri Botanical Garden's TROPICOS database and John Hall (University of Wales, UK). Thanks also to Ruth Swetnam and Jon Green (University of Cambridge, UK), Neil Burgess and Jon Fjeldså (University of Copenhagen, Denmark), Boniface Mbilinyi (Sokoine University of Agriculture, Tanzania) and Barnaby Clark (University of Helsinki, Finland) for their work on land cover classifications. Logistical support and permissions for fieldwork in Tanzania were provided by the Tanzanian Commission for Science and Technology, the Forestry and Beekeeping Division of the Ministry of Natural Resources and Tourism, the University of Dar es Salaam and WWF Tanzania. This research was funded by the Marie-Curie programme of the European

6th Framework under contract MEXT-CT-2004–517098 to R. Marchant, and contributes to the Global Land Project (http://www.globallandproject.org/) and to the Valuing the Arc Programme (http://www.valuingthearc.org/). The Leverhulme Trust funds the latter with additional support from the Royal Society. Marshall was funded by Natural Environment Research Council grant NER/S/A/2002/11177 and by the National Geographic Society and Margot Marsh Biodiversity Foundation. Funding from the Academy of Finland for the Taita Project (http://www.helsinki.fi/science/taita/) is also gratefully acknowledged. Finally, we thank Janet Franklin (Associate Editor), Alistair Jump (University of Stirling, UK), Daniel Kissling (Aarhus University, Denmark) and two anonymous reviewers, whose comments greatly improved this manuscript.

Author contributions

The study was conceived by P.J.P., who programmed and implemented the modelling experiments, analysed the results and prepared the manuscript. R.E.G. provided details of endemism and proposed Red List designations. A.A., E.B., R.E.G., J.C.L, A.R.M. and P.J.P. contributed to the botanical database, which was complied and cleaned by A.A., R.E.G. and P.J.P. Estimates of forest cover in the Taita Hills were provided by P.K.E.P. Rainfall and cloud data were provided by M.M. The study was supervised by R.M. and C.J.M.

Appendix 4A. Details of plant data

Species location data were based on a large dataset totalling *c*. 70 000 records, 30% of which were from the Missouri Botanical Garden's TROPICOS database and 70% from vegetation plot assessments (Frontier-Tanzania, A.A., A.R.M., J.C.L. and P.J.P). Occurrence data were collated and modelled at species level, except when only one infra-specific taxon of a species is known to occur in the EAMs, in which cases the subspecies or variety was modelled.

A project sponsored by the Critical Ecosystem Partnership Fund has recently undertaken an updated assessment of the conservation status of the combined EAM and Coastal Forest flora (Gereau *et al.*, 2010). Pending publication on the IUCN Red List, we did not distinguish between threat categories, but simply identified as "Threatened" the modelled taxa that either have a proposed assessment in one of the globally threatened categories (Vulnerable, Endangered or Critically Endangered) or are considered as potentially threatened and remain to be evaluated. For purposes of endemism used a uniform lower altitudinal limit of 500 m. This procedure, although somewhat over-simplified given complexities in the altitudinal limits of coastal vegetation, was the most pragmatic given the data available (but see Chapters 2, 5 and 6, which post-date this article). Of the 452 taxa modelled, 68 are proposed as threatened and 25 are endemic to the EAMs.

For model calibration purposes, we reviewed the locality information of all specimen records, assigning each to one of four spatial categories according to our confidence in the coordinates provided: 150 m or higher (42%), 1 km (21%), 2 km (30%) or lower (7%). Taxa with records of occurrence in ten or more distinct 1 km grid squares were modelled at 1 km resolution, using all available 150 m and 1 km records. The remaining taxa were modelled at 2 km, using all available 150 m, 1 km and 2 km records, provided that these localities spanned ten or more 2 km grid squares. Records not trusted to within 2 km were omitted from model calibration, but were retained as independent test data. In some cases, there was scope to calibrate models at the very highest resolution (i.e., records available in ten or more 150 m grid squares), potentially giving a superior representation of microclimate; this however was beyond the spatial precision of the climate and soils data. Moreover, specimens were often clustered within the same 1 km grid square, so running models at such a fine-scale would have exacerbated fine-scale spatial dependence in the training data.

Appendix 4B. Occurrence thresholds and sensitivity to prevalence

Using a test set of 16 taxa (four of each growth form) we investigated the sensitivity of models to prevalence and to the chosen method for selecting occurrence thresholds (see table below; chosen method in bold font). We first tried an intermediate prevalence of 0.5, allocating absences at a ratio of 1:1 against presences. This approach resulted in spatial predictions that were poorly constrained and that varied considerably between runs. For our data, a presence-absence ratio of 1:5 was more appropriate. Lower prevalence (< 0.2) led to similar spatial patterns but slightly lower validation scores. Previous studies confirm that a prevalence in the range 0.2-0.8 minimises bias in validation metrics (Manel *et al.*, 2001; McPherson *et al.*, 2004) and allows optimal occurrence thresholds to be more easily identified (Liu *et al.*, 2005). In our study, a prevalence below 0.2 also hindered comparison across growth forms, because for lianas the required number of absences sometimes exceeded the number of target sites available.

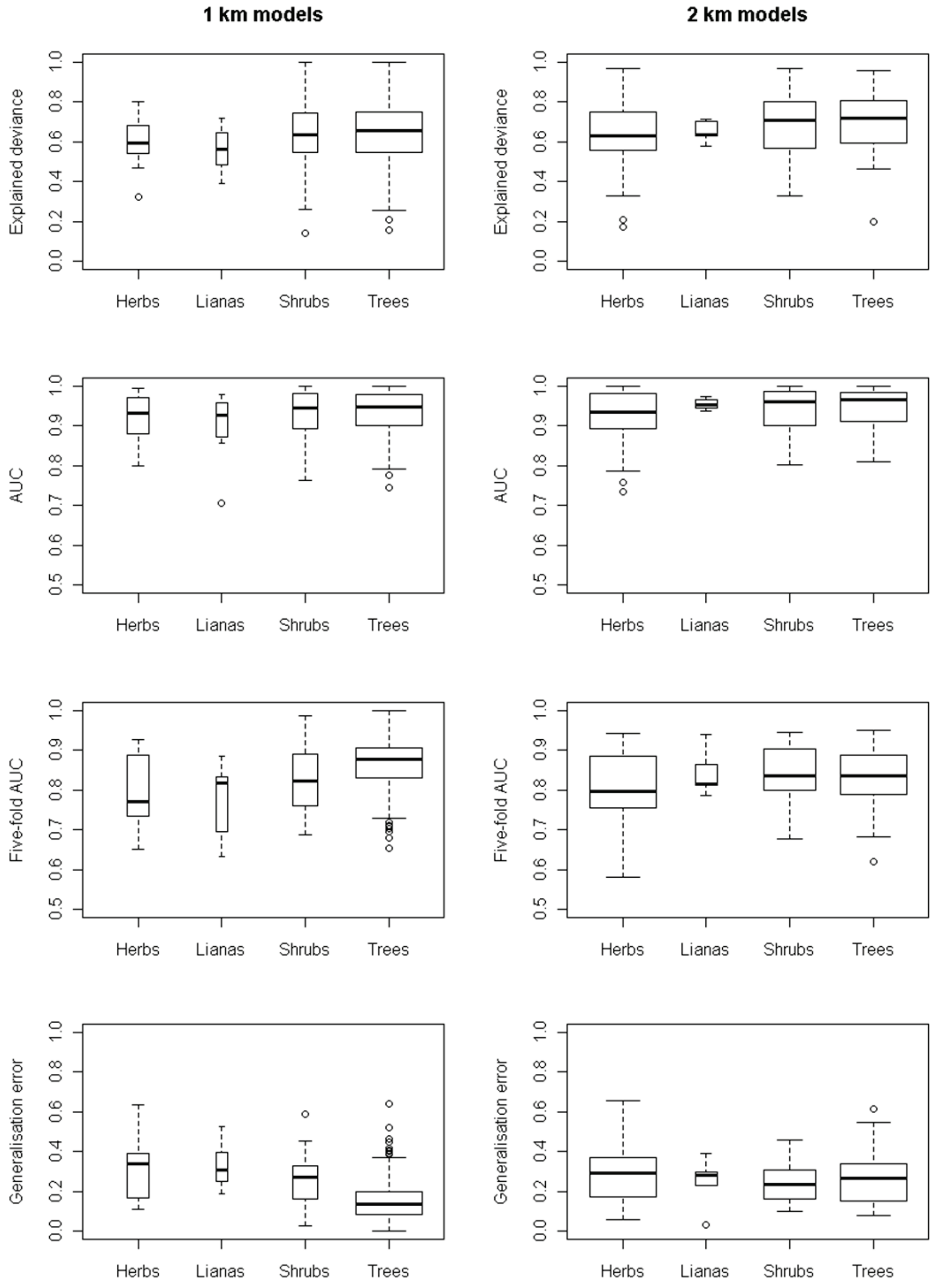
Once calibrated at the chosen prevalence, models predicted occurrence on a continuous scale, from zero to one. Maps of estimated presence-absence were obtained by imposing taxon-specific occurrence thresholds, chosen by maximising the sum of sensitivity and specificity (Cantor *et al.*, 1999). This approach was shown to perform well in a comparative study by Liu *et al.* (2005), who recommend it alongside two other techniques: the prevalence approach (threshold = model prevalence) and the sensitivity-specificity equality approach. All three methods produced similar results, but that maximising the sum of sensitivity and specificity yielded the most constrained predictions with minimal type II error.

	Threshold	Sensitivity	Specificity
Presences 1 : 5 Absences			
Prevalence of training data	0.21 (0.01)	0.93 (0.02)	0.88 (0.02)
Sensitivity-specificity sum maximisation	**0.37 (0.05)**	**0.94 (0.02)**	**0.93 (0.01)**
Sensitivity-specificity equality	0.35 (0.04)	0.91 (0.02)	0.90 (0.02)
Presences 1 : 10 Absences			
Prevalence of training data	0.14 (0.02)	0.92 (0.02)	0.85 (0.02)
Sensitivity-specificity sum maximisation	0.27 (0.05)	0.93 (0.02)	0.90 (0.02)
Sensitivity-specificity equality	0.25 (0.04)	0.88 (0.02)	0.89 (0.02)
Presences : All target sites			
Prevalence of training data	0.09 (0.02)	0.92 (0.02)	0.84 (0.02)
Sensitivity-specificity sum maximisation	0.17 (0.04)	0.92 (0.02)	0.90 (0.02)
Sensitivity-specificity equality	0.14 (0.03)	0.89 (0.02)	0.89 (0.01)

Appendix 4C. Analysis of model performance

Box-plot comparisons of resolution and growth form

Box widths are proportional to the number of taxa. From the top: proportion of deviance explained, area under the receiver-operator characteristic curve (AUC) including a five-fold cross-validation, and generalisation error (GE). The latter is defined as the proportion of above-chance AUC lost under cross-validation; GE ≈ 0 indicates a very stable model, whilst GE ≈ 1 warns that discriminatory ability at unvisited sites may be no better than chance.



Significance of differences between models

Model resolution

Wilcoxon rank sum tests (one-sided), comparing the performance of models calibrated at 1 km resolution (254 trees, 7 lianas, 33 shrubs, 25 herbs) with those calibrated at 2 km resolution (50 trees, 5 lianas, 29 shrubs, 49 herbs).

	Explained deviance		AUC		5-fold AUC		Generalisation error	
	1 km > 2 km	2 km > 1 km	1 km > 2 km	2 km > 1 km	1 km > 2 km	2 km > 1 km	1 km > 2 km	2 km > 1 km
Trees	ns	*	ns	*	***	ns	ns	***
Lianas	ns	ns	ns	ns	ns	ns	ns	ns
Shrubs	ns	ns	ns	ns	ns	ns	ns	ns
Herbs	ns	ns	ns	ns	ns	ns	ns	ns
All taxa	ns	*	ns	*	***	ns	ns	***

***, $p \leq 0.001$ (extremely significant); **, $p \leq 0.01$ (highly significant); *, $p \leq 0.05$ (significant); ns, not sig.

Growth form

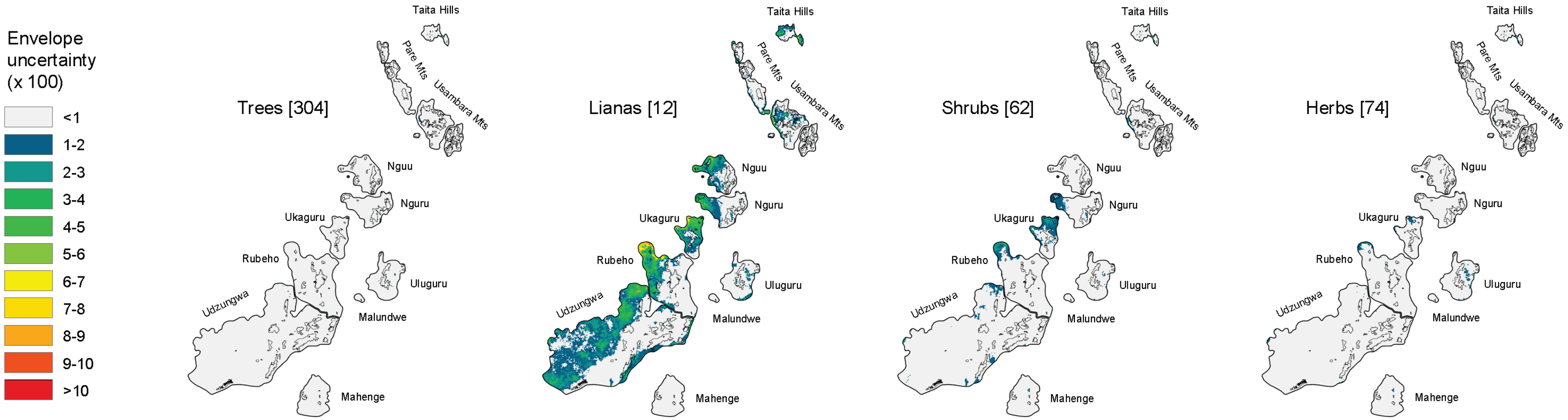
Wilcoxon rank sum tests (one-sided: rows > columns), comparing the performance of models calibrated for the different growth forms of plants (304 trees, 12 lianas, 62 shrubs, 74 herbs).

	Explained deviance				AUC				5-fold AUC				Generalisation error			
	Trees	Lianas	Shrubs	Herbs	Trees	Lianas	Shrubs	Herbs	Trees	Lianas	Shrubs	Herbs	Trees	Lianas	Shrubs	Herbs
Trees	-	ns	ns	ns	-	ns	ns	ns	-	ns	ns	***	-	ns	ns	ns
Lianas	ns	-	ns	ns	ns	-	ns	ns	ns	-	ns	ns	***	-	ns	ns
Shrubs	ns	ns	-	ns	ns	ns	-	ns	ns	ns	-	*	***	ns	-	ns
Herbs	ns	ns	ns	-	ns	ns	ns	-	ns	ns	ns	-	***	ns	*	-

***, $p \leq 0.001$ (extremely significant); **, $p \leq 0.01$ (highly significant); *, $p \leq 0.05$ (significant); ns, not significant

Envelope uncertainty maps (EUMs)

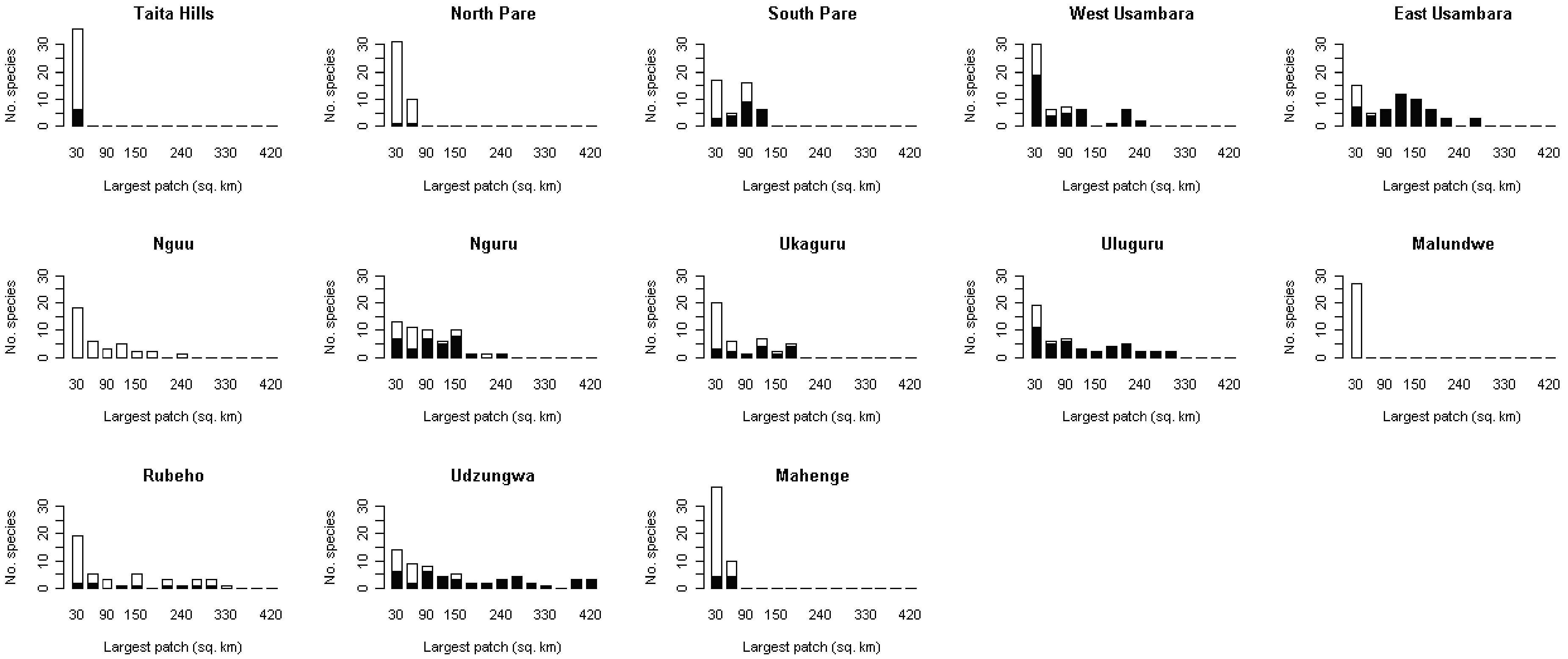
The proportional ‘distance’ of each grid cell from the calibration envelope was mapped with respect to each environmental predictor. Prediction uncertainty resulting from extrapolation to novel parameter space was estimated using an average of these maps, weighted according to the relative contributions of predictors in models (drop in explained deviance with predictor removed). The EUMs below show mean values for different growth forms of plant (number of taxa in parentheses). Dormann (2007b) recommends that one should not extrapolate further than 1/10th of the parameter range – caution is therefore recommended where the EUM > 0.1 (Chapter 3).



Models rarely extrapolated far beyond the niche-breadth used for calibration. Environmental coverage of tree and herb data was particularly good. Coverage for shrubs was slightly less comprehensive with respect to western Nguru and northern Ukaguru, but models were not seriously affected (EUM < 0.1). Uncertainty was highest for lianas, generally increasing with distance from the coast. With the exception of western Nguru near Talagwe Forest Reserve, the areas of highest uncertainty did not coincide with present-day forest cover, and so our results were unaffected.

Appendix 4D. Patch occupancy for endemic/threatened taxa

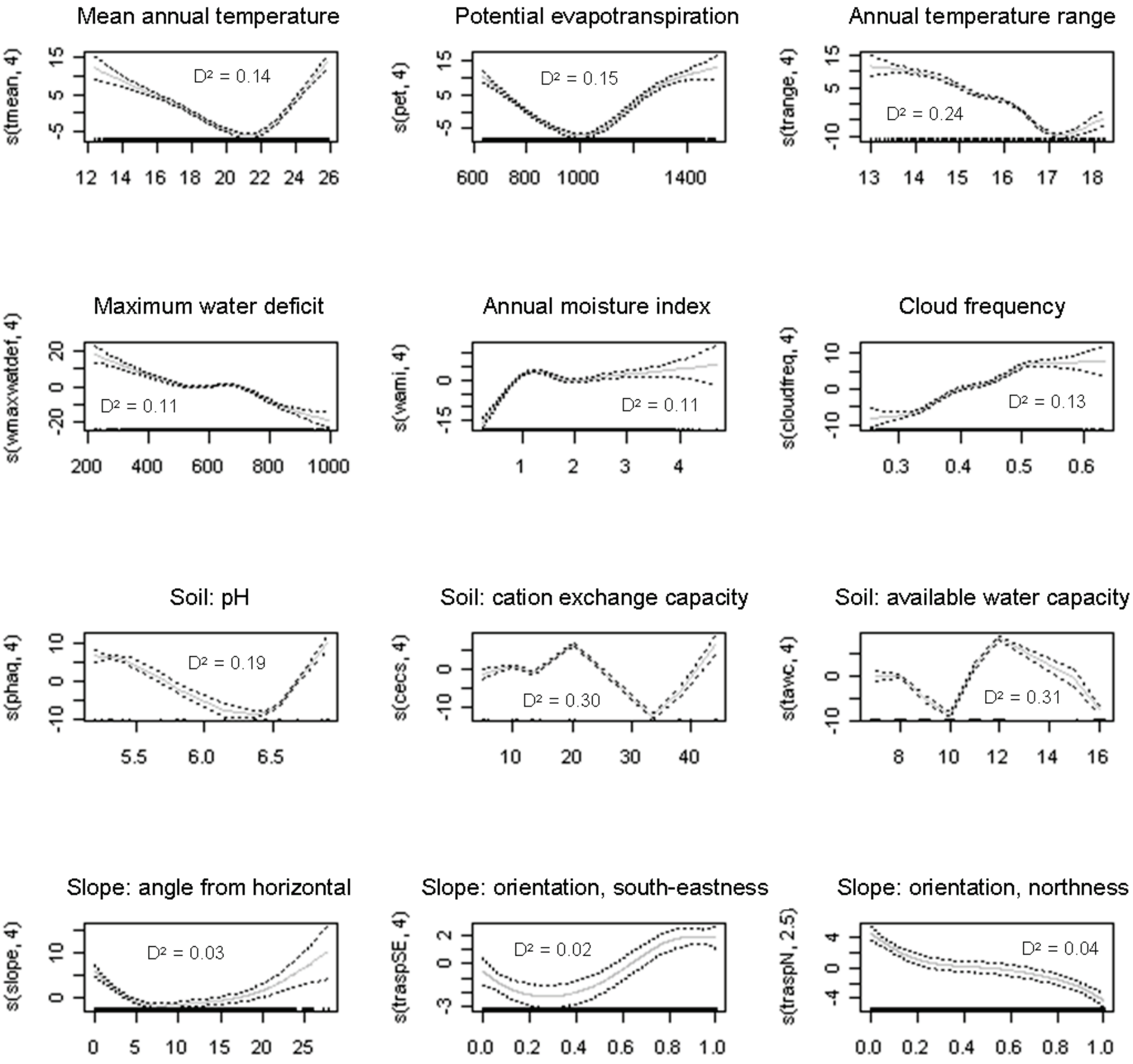
Histograms showing patterns of within-bloc patch occupancy for taxa that are endemic and/or threatened. Horizontal axes represent the largest contiguous area of forest providing environmentally suitable conditions for a particular species. Vertical axes show the number of species in each 30 km^2 patch size interval. Filled bars relate to confirmed occurrence at the bloc level; open bars relate to predictions of occurrence in novel mountain blocs.



Appendix 4E. Correlates of richness and endemism

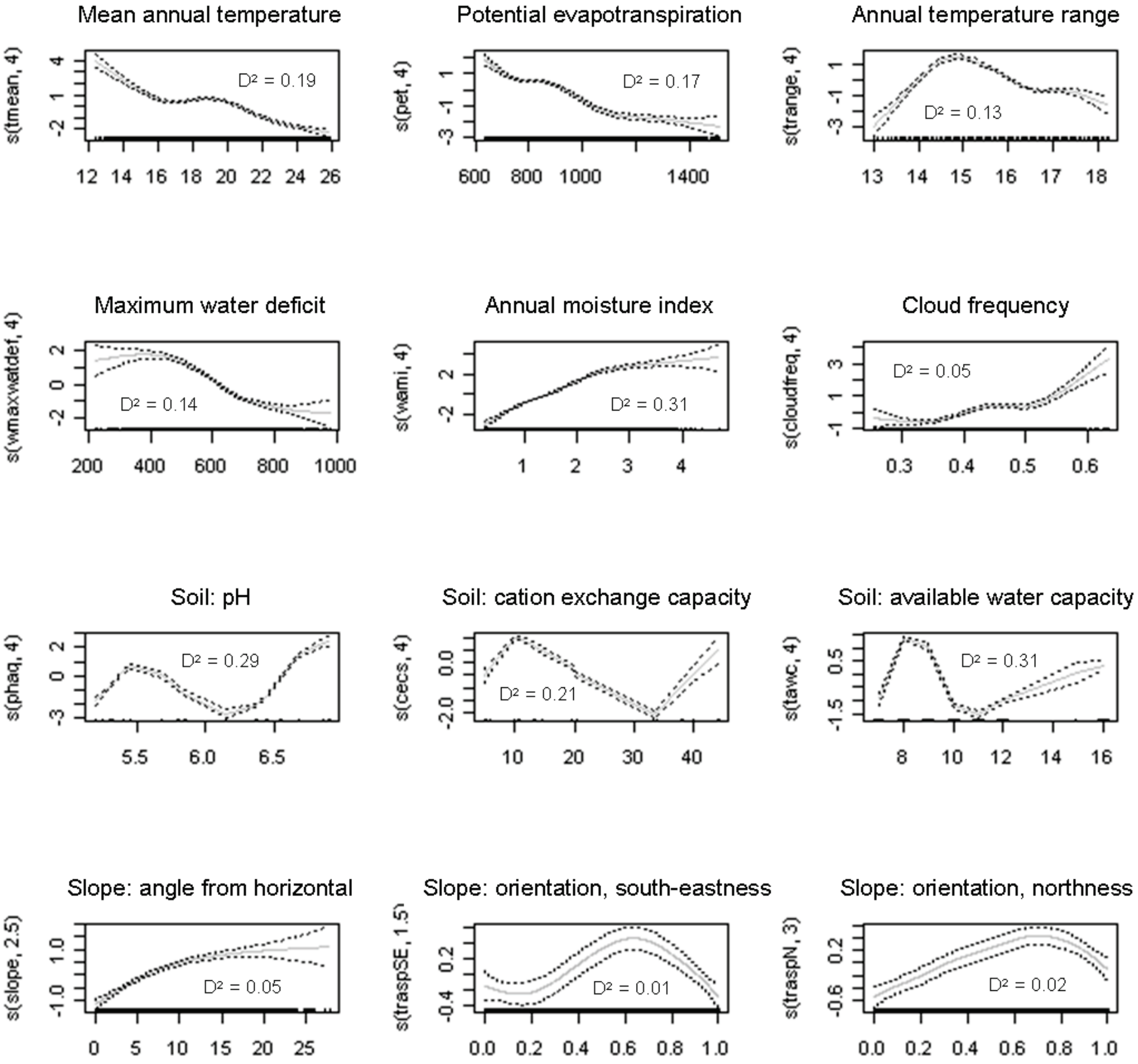
Richness

Response of plant richness (across forested grid cells) to the environmental variables used in modelling, including the proportion of deviance explained in an additive model (D^2). Temperature range is probably the strongest functional predictor. Soil variables appear to be important, but irrational response shapes suggest that these are not casual factors. Conversely, responses to slope and aspect appear sensible (overall richness higher on south and south-easterly slopes > 15˚; endemism higher on north-easterly slopes > 0˚) but explain little of the deviance in modelled richness.



Endemism

Response of endemic plant richness (across forested grid cells) to the environmental variables used in modelling, including the proportion of deviance explained in an additive model (D^2). Annual moisture index is the strongest predictor. As above, response shapes for soil predictors suggest spurious relationships.



References

AFPD (2009) *African Flowering Plants Database (version 3.2)*, Conservatoire et Jardin botaniques de la Villle de Genève and South African National Biodiversity Institute, Pretoria, South Africa (http://www.ville-ge.ch/musinfo/bd/cjb/africa/).

Ahrends, A., Burgess, N. D., Gereau, R. E., Marchant, R., Bulling, M. T., Lovett, J. C., Platts, P. J., Kindemba, V. W., Owen, N., Fanning, E. and Rahbek, C. (2011) Funding begets biodiversity. *Diversity and Distributions*, **17**, 191-200.

Augustin, N. H., Mugglestone, M. A. and Buckland, S. T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339-347.

Batje, N. H. (2004) *SOTER-based soil parameter estimates for Southern Africa*, Report 2004/04, ISRIC - World Soil Information, Wageningen, The Netherlands.

Beentje, H. J. (1988) An ecological and floristic study of the forests of the Taita hills, Kenya. *Utafiti - the Occational Papers of the National Museum of Kenya*, **1**, 23-66.

Beentje, H. J. (1994) *Kenya Trees, Shrubs and Lianas*, National Museums of Kenya, Nairobi, Kenya.

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

Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Rylands, A. B., Konstant, W. R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G. and Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909-923.

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. and Stuart, S. N. (2007) The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, **134**, 209-231.

Burgess, N. D., Hales, J. D., Ricketts, T. H. and Dinerstein, E. (2006) Factoring species, non-species values and threats into biodiversity prioritisation across the ecoregions of Africa and its islands. *Biological Conservation*, **127**, 383-401.

Cantor, S. B., Sun, C. C., Tortolero-Luna, G., Richards-Kortum, R. and Follen, M. (1999) A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *Journal of Clinical Epidemiology*, **52**, 885-892.

CEPF (2003) *Ecosystem Profile for the Eastern Arc Mountains & Coastal Forests of Tanzania and Kenya*, Final version 31 July 2003, updated March 2005, Critical Ecosystem Partnership Fund (http://www.cepf.net/).

Clark, B. J. F. and Pellikka, P. K. E. (2009) Landscape analysis using multiscale segmentation and object-orientated classification. In *Recent Advances in Remote Sensing and Geoinformation Processing for Land Degradation Assessment* (Eds, Röder, A. and Hill, J.) Taylor & Francis Group 2009, pp. 323-342.

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. and 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.

Doggart, N., Perkin, A., Kiure, J., Fjeldså, J., Poynton, J. and 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.

Dormann, C. F. (2007a) Assessing the validity of autologistic regression. *Ecological Modelling*, **207**, 234-242.

Dormann, C. F. (2007b) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, **8**, 387-397.

Dormann, C. F., McPherson, J. M., Araujo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kuhn, I., Ohlemuller, R., Peres-Neto, P. R., Reineking, B., Schroder, B., Schurr, F. M. and Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609-628.

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

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

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

Fjeldså, J., Ehrlich, D., Lambin, E. and 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. and Lovett, J. C. (1997a) Biodiversity and environmental stability. *Biodiversity and Conservation*, **6**, 315-323.

Fjeldså, J. and Lovett, J. C. (1997b) Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation*, **6**, 325-346.

Gereau, R. E., Taylor , C. M., Bodine, S. and Kindeketa, W. J. (2010) Plant Conservation Assessment in the Eastern Arc Mountains and Coastal Forests of Tanzania and Kenya (http://www.mobot.org/MOBOT/Research/tanzania/cepf.shtml/).

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

Graham, C. H., Moritz, C. and Williams, S. E. (2006) Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 632-636.

GRASS-Development-Team (2009) *Geographic Resources Analysis Support System (GRASS)*, GNU General Public License (http://grass.osgeo.org/).

Green, D. M. and Swets, J. A. (1974) *Signal Detection Theory and Psychophysics*, Krieger, New York.

Grenyer, R., Orme, C. D. L., Jackson, S. F., Thomas, G. H., Davies, R. G., Davies, T. J., Jones, K. E., Olson, V. A., Ridgely, R. S., Rasmussen, P. C., Ding, T. S., Bennett, P. M., Blackburn, T. M., Gaston, K. J., Gittleman, J. L. and Owens, I. P. F. (2006) Global distribution and conservation of rare and threatened vertebrates. *Nature*, **444**, 93-96.

Griffiths, C. J. (1993) The geological evolution of East Africa. In *Biogeography and ecology of the rain forests of Eastern Africa* (Eds, Lovett, J. C. and Wasser, S.) Cambridge University Press, Cambridge, UK, pp. 9-21.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A. and Zimmermann, N. E. (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**, 501-511.

Hall, J., Burgess, N. D., Lovett, J., Mbilinyi, B. and Gereau, R. E. (2009) Conservation implications of deforestation across an elevational gradient in the Eastern Arc Mountains, Tanzania. *Biological Conservation*, **142**, 2510-2521.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Holmgren, M. and Poorter, L. (2007) Does a ruderal strategy dominate the endemic flora of the West African forests? *Journal of Biogeography*, **34**, 1100-1111.

ICBP (1992) *Putting Biodiversity on the Map: Priority Areas for Global Conservation*, International Council for Bird Preservation, Cambridge.

IUCN (2009) IUCN Red List of Threatened Species. Version 2009.2. (http://www/iucnredlist.org/). Downloaded on 20 January 2010.

Iversen, S. T. (1991) The Usambara Mountains, NE Tanzania: Phytogeography of the vascular plant flora. *Acta Universitatis Upsaliensis Symbolae Botanicae Upsalienses*, **29**, 1-234.

Jarvis, A., Reuter, H. I., Nelson, A. and Guevara, E. (2008) Hole-filled seamless SRTM data (version 4), International Centre for Tropical Agriculture, CIAT (http://srtm.csi.cgiar.org/).

Jetz, W., Rahbek, C. and 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.

Keith, D. A., Akcakaya, H. R., Thuiller, W., Midgley, G. F., Pearson, R. G., Phillips, S. J., Regan, H. M., Araujo, M. B. and Rebelo, T. G. (2008) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, **4**, 560-563.

Kremen, C., Cameron, A., Moilanen, A., Phillips, S. J., Thomas, C. D., Beentje, H., Dransfield, J., Fisher, B. L., Glaw, F., Good, T. C., Harper, G. J., Hijmans, R. J., Lees, D. C., Louis, E., Nussbaum, R. A., Raxworthy, C. J., Razafimpahanana, A., Schatz, G. E., Vences, M., Vieites, D. R., Wright, P. C. and Zjhra, M. L. (2008) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, **320**, 222-226.

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

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

Lawton, J. H., Bignell, D. E., Bolton, B., Bloemers, G. F., Eggleton, P., Hammond, P. M., Hodda, M., Holt, R. D., Larsen, T. B., Mawdsley, N. A., Stork, N. E., Srivastava, D. S. and Watt, A. D. (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, **391**, 72-76.

Lehouck, V., Spanhove, T., Gonsamo, A., Cordeiro, N. and Lens, L. (2009) Spatial and temporal effects on recruitment of an Afromontane forest tree in a threatened fragmented ecosystem. *Biological Conservation*, **142**, 518-528.

Liu, C. R., Berry, P. M., Dawson, T. P. and Pearson, R. G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385-393.

Lomolino, M. V. (2000) Ecology's most general, yet protean pattern: the species-area relationship. *Journal of Biogeography*, **27**, 17-26.

Lovett, J. C. (1990) Classification and status of the moist forests of Tanzania. *Notifications from the Institute for General Botany Hamburg*, **23a**, 287–300.

Lovett, J. C., Marchant, R., Taplin, J. and Küper, W. (2005) The oldest rainforests in Africa: stability or resilience for survival and diversity? In *Phylogeny and Conservation* (Eds, Purvis, A., Gittleman, J. L. and Brooks, T. M.) Cambridge University Press, Cambridge, UK.

MacArthur, R. H. and Wilson, E. O. (1967) *The Theory of Island Biogeography*, Princeton University Press, Princeton, USA.

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. and van Jaarsveld, A. S. (2000) It's time to work together and stop duplicating conservation efforts. *Nature*, **405**, 393-393.

Manel, S., Williams, H. C. and Ormerod, S. J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-931.

Marchant, R., Mumbi, C., Behera, S. and Yamagata, T. (2007) The Indian Ocean dipole - the unsung driver of climatic variability in East Africa. *African Journal of Ecology*, **45**, 4-16.

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

Marshall, A. R., Jorgensbye, H. I. O., Rovero, F., Platts, P. J., White, P. C. L. and Lovett, J. C. (2010) The species-area relationship and confounding variables in a threatened monkey community. *American Journal of Primatology*, **72**, 325-336.

McPherson, J. M., Jetz, W. and Rogers, D. J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811-823.

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

Miller, J., Franklin, J. and Aspinall, R. (2007) Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, **202**, 225-242.

Mittermeier, R. A., Robles Gil., P., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., Lamoreux, J. and da Fonseca, G. A. B. (2004) *Hotspots Revisited*, CEMEX, Mexico.

MNRT (1997) National Reconnaissance Level Land Use and Natural Resources Mapping Project. Ministry of Natural Resources and Tourism, The United Republic of Tanzania.

Mulligan, M. (2006a) Global gridded 1 km TRMM rainfall climatology and derivatives. Version 1 (http://www.ambiotek/1kmrainfall/).

Mulligan, M. (2006b) MODIS MOD35 pan-tropical cloud climatology. Version 1 (http://www.ambiotek.com/clouds/).

Mulligan, M. and Burke, S. M. (2005) Global cloud forests and environmental change in a hydrological

context. Final report to DfID of project ZF0216 (http://www.ambiotek.com/cloudforests/).

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

Mwakalila, S., Burgess, N. D., Ricketts, T., Olwero, N., Swetnam, R., Mbilinyi, B., Marchant, R., Mtalo, F., White, S., Munishi, P., Marshall, A., Malimbwi, R., Smith, C., Jambiya, G., Madoffe, S., Fisher, B., Kajembe, G., Morse-Jones, S., Kulindwa, K., Green, J. and Balmford, A. (2009) Linking Science with Stakeholders to sustain Natural Capital. *The Arc Journal*, 22-27.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. and 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. and 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.

Olson, D. M. and Dinerstein, E. (1998) The global 200: A representation approach to conserving the Earth's most biologically valuable ecoregions. *Conservation Biology*, **12**, 502-515.

Parker, B. J., Guenter, S. and Bedo, J. (2007) Stratification bias in low signal microarray studies. *BMC Bioinformatics*, **8**, 16.

Pellikka, P. K. E., Lötjönen, M., Sijander, M. and Lens, L. (2009) Airborne remote sensing of spatiotemporal change (1955-2004) in indigenous and exotic forest cover in the Taita Hills, Kenya. *International Journal of Applied Earth Observation and Geoinformation*, **11**, 221-232.

Phillips, S. J., Dudik, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. and Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181-197.

Possingham, H. P. and Wilson, K. A. (2005) Biodiversity - Turning up the heat on hotspots. *Nature*, **436**, 919-920.

Poynton, J. C., Loader, S. P., Sherratt, E. and Clarke, B. T. (2007) Amphibian diversity in east african biodiversity hotspots: Altitudinal and latitudinal patterns. *Biodiversity and Conservation*, **16**, 1103-1118.

Pulliam, H. R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349-361.

R-Development-Core-Team (2009) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org/).

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

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

Rogo, L. and Oguge, N. (2000) The Taita Hills forest remnants: A disappearing world heritage. *Ambio*, **29**, 522-523.

Schlüter, T. (1997) Geology of East Africa. In *Contributions to the regional geography of the Earth* (translated from German), Borntraeger, Berlin, Germany.

Stattersfield, A. J., Crosby, M. J., Long, A. J. and Wege., D. C. (1998) *Endemic bird areas of the world: priorities for biodiversity conservation*, BirdLife International, Cambridge, UK.

Swets, J. A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293.

Temu, R. P. C. and Andrew, S. M. (2008) Endemism of plants in the Uluguru Mountains, Morogoro, Tanzania. *Forest Ecology and Management*, **255**, 2858-2869.

Thornthwaite, C. W. (1948) An Approach toward a Rational Classification of Climate. *Geographical Review*, **38**, 55-94.

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

Wilson, K. A., McBride, M. F., Bode, M. and Possingham, H. P. (2006) Prioritizing global conservation efforts. *Nature*, **440**, 337-340.

Yee, T. W. and Mitchell, N. D. (1991) Generalized additive models in plant ecology. Journal of Vegetation *Science*, **2**, 587-602.

Chapter 5 – Climate Change

Published in *Ecography* (in press), doi: 10.1111/j.1600-0587.2012.07805.x

Spatial heterogeneity of climate change in an Afromontane centre of endemism

Philip J. Platts [a,b,*], Roy. E. Gereau [c], Neil D. Burgess [b,d,e], Rob Marchant [a]

[a] *The York Institute for Tropical Ecosystem Dynamics (KITE), Environment Department, University of York, Heslington, York YO10 5DD, UK.* [b] *Conservation Science Group, Zoology Department, University of Cambridge, Cambridge CB2 3EJ, UK.* [c] *Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA.* [d] *Centre for Macroecology, Evolution and Climate, Department of Biology, Copenhagen University, Universitetsparken 15, DK-2100, Copenhagen, Denmark.* [e] *WWF-US Conservation Science Program, 1259 24th Street NW, Washington DC, USA*

** Corresponding author. Email: philip.platts@gmail.com*

Abstract

Broad-scale assessments of how climate change might impact mountain ecosystems, especially in areas of high biodiversity and endemism, are compromised by the lack of localised climate feedback in global circulation models. Here, we use regionally downscaled climate models to highlight how spatial variation in forecast change could impact rare plant distributions differentially across the Eastern Arc Mountains of Tanzania and Kenya, part of the Eastern Afromontane Biodiversity Hotspot. Concordant with the theory that climatic stability facilitates the accumulation of rare species, we find significant positive correlations between endemic plant richness and future climatic persistence within the dispersal-limiting sky islands of this mountain archipelago. Further, we explore the hypothesis that mountain plants will move upslope in response to climate change and find that, conversely, some species are predicted to tend downslope, despite warmer annual conditions, driven by changes in seasonality and water availability. Importantly, two thirds of the modelled plant species are predicted to respond in different directions in different parts of their ranges, exemplifying the potential for individualistic responses of species and disjunct populations to environmental change, and the need for regional focus in climate change impact assessment. Conservation planners, and more broadly those charged with developing climate adaption policy, are advised to take caution in inferring local patterns of change from zoomed perspectives of broad-scale models. Moreover, a preoccupation with mean annual temperature as the principal driver of ecosystem change is misguided and could compromise efforts to make conservation plans resilient to future climate change. Faced with spatially complex and inherently uncertain future conditions, sensible priorities are to restore forest connectivity and to underpin adaption strategies with knowledge of how ecosystems and people have adapted to previous episodes of rapid change.

Keywords: altitudinal migration; climatic stability; conservation; downscaling; Eastern Arc Mountains; global warming; plants; regional climate models; refugia, resilience, richness gradients

Introduction

Global and continental-scale modelling assessments warn of climate-induced range contraction and potential extinction for many species by the end of the 21st century (Thomas *et al.*, 2004; McClean *et al.*, 2005; Pereira *et al.*, 2010). With respect to globally important but locally complex mountain systems, these studies have limited potential to guide ground-level adaption. The datasets employed are too coarse-grained to detect accurately changes in niche-space or potential refugia for all but the most generalist and wide-ranging of species, risking a bias in climate-sensitive priorities toward the least vulnerable components of biodiversity. Spatially downscaled, high-resolution climate projections are therefore in great demand for conservation planning, especially in tropical Africa where species data and observational time-series are scarce (Wilby *et al.*, 2004).

Addressing this need, a recently popularised technique is change-factor downscaling, whereby anomalies forecast by global circulation models (GCMs) are added to higher-resolution contemporary baselines (Ramirez and Jarvis, 2008; Tabor and Williams, 2010; Kou *et al.*, 2011). The resulting surfaces convey locally important detail such as orographic rainfall, affording investigations of ecological impacts at scales more relevant to management (e.g., 1 km). A serious shortcoming, however, is that spatial heterogeneity in change is crudely approximated by interpolation of GCM squares (typically of side 200-600 km). Governments, conservation organisations and other bodies concerned with climate change adaption are therefore provided with seemingly high-resolution scenarios of change, upon which policy is developed, but that omit key regional dynamics that are fundamental to the conservation of species, habitats and associated natural resources. Incorporation of more localised climate feedback is especially critical in highly heterogeneous landscapes and areas with strong land-ocean climate gradients, both of which are exemplified in East Africa.

An increasingly accessible means to minimise this spatial uncertainty is through the use of dynamically downscaled forecasts, obtained from regional climate models (RCMs). RCMs employ thermodynamic processes, interactions and feedbacks similar to those of large-scale coupled atmosphere-ocean models, but operate on horizontal resolutions of tens rather than hundreds of kilometres. Although not currently available for all parts of the world, the derivation of change factors from regional models is likely to become increasingly feasible within the timeline of the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC-AR5, scheduled for 2014). In this paper, we combine dynamic and change-factor downscaling to investigate patterns of change forecast for the Eastern Arc Mountains (EAMs) of Tanzania and Kenya (Fig. 5.1a), part of the Eastern Afromontane Biodiversity Hotspot (Mittermeier *et al.*, 2004). Of the 26,848 vascular plant species indigenous to tropical Africa, a remarkable 3834 (14%) have been documented in the EAMs (R. E. Gereau, unpublished data) – habitats representing just 0.24% of the land area (Chapter 2; Gereau *et al.*, in prep.). The assessment of climate change impacts on these species is, per unit area of investment, one of the most important and cost effective foci for biodiversity conservation and adaption planning globally (Brooks *et al.*, 2002), but has to date been hindered by a lack of data and analytical methods at spatial scales relevant for conservation decision making.

The EAMs encapsulate many characteristics that confound site-scale inference from broad-scale assessments: they are topographically complex, fragmented across a series of ‘sky islands’ and historically dependent upon maritime climatic influence for species persistence (Chapter 2; Lovett, 1990; Marchant *et al.*, 2007). Using an RCM already available for the region, we identify sites where baseline climates might be lost and where novel conditions could emerge. Implications for conservation are discussed in relation to plant species strictly endemic to the mountain ecoregion, revealing how spatially diverse shifts in seasonality and rainfall have potential to drive ecosystem change in directions unforeseen by broader-scale models. With this new level of spatial detail, it is possible for the first time to question whether future change may be indicative of past ecosystem stability in this globally important centre of endemism.

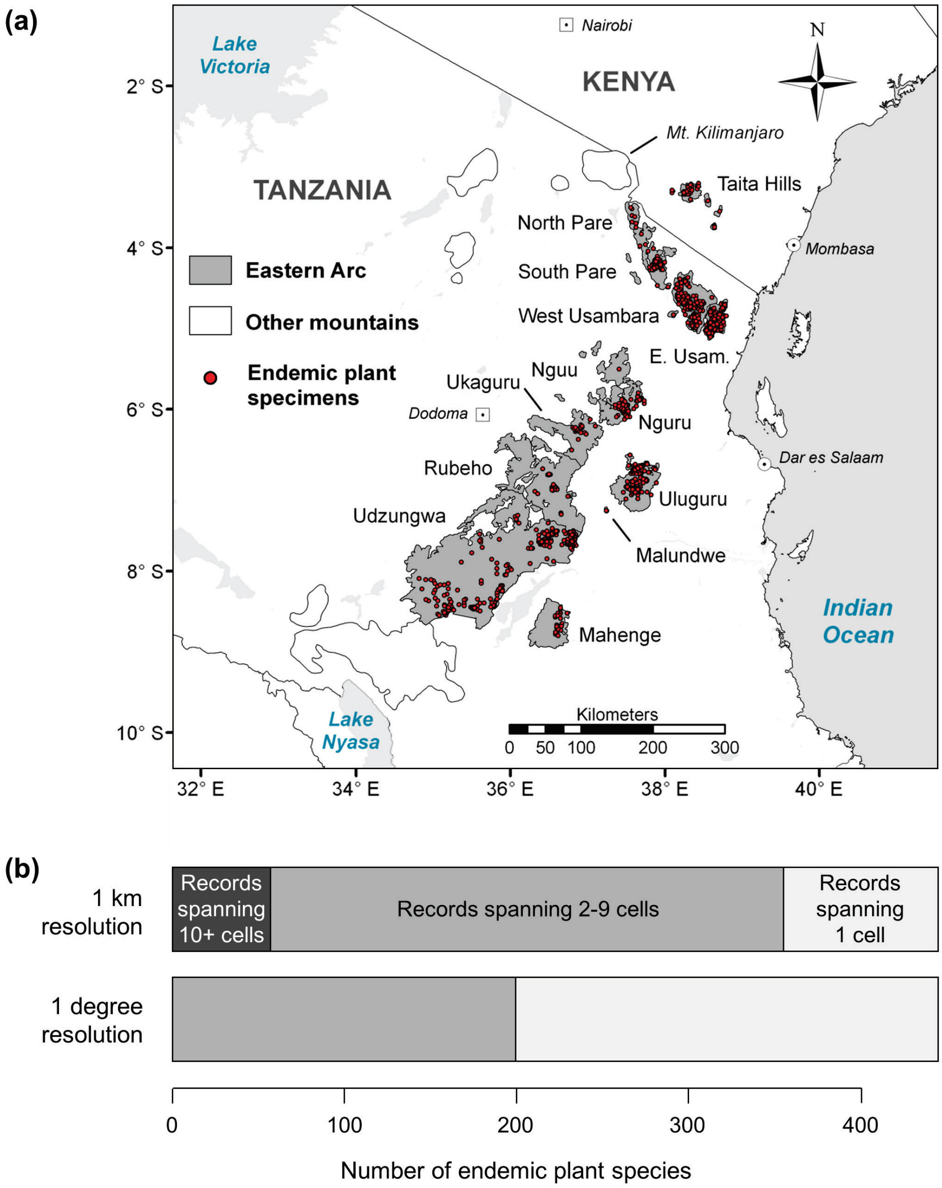


Figure 5.1. (a) The Eastern Arc Mountains, spanning 5.2 million ha (10% forest) and 13 distinct blocs. Volcanoes to the north and highlands to the south are floristically distinct and contain fewer species of restricted distribution, attributed to geological age and climatic influence (Lovett, 1990). Red dots locate the origins of endemic plant specimen data used in this study. (b) Bar plot summarises sample prevalence at resolutions suitable for species distribution modelling at local (1 km) and continental / global (1° ~111 km) scales. Simple envelope models are possible for 354 and 198 species at the higher and lower resolutions, respectively. For more sophisticated kinds of distribution model (e.g., statistical regression), records of occurrence in at least ten cells are required (Stockwell and Peterson, 2002). At 1 km resolution, 56 species satisfy this condition; no species has sufficient coverage at 1° resolution.

Methods

Selection of study region

The EAMs (Fig. 5.1a) harbour remnants of a pan-African forest belt, which prior to the uplift of the Central Tanganyika Plateau (25 MY BP) periodically extended to the Guineo-Congolian forests in west and central Africa (Couvreur *et al.*, 2008). Even older possible links to Madagascar and southeast Asia indicate speciation events dating back over 100 MY to the breakup of Gondwana (Dinesen *et al.*, 1994). Species persistence is attributed to long-term moisture stability under the influence of the Indian Ocean, coupled with potential for populations to 'cling on' in topoclimatic refugia or else track transient conditions up or down slope (Fjeldså *et al.*, 1997; Lovett *et al.*, 2005). In the language of Williams *et al.* (2007), familiar conditions may have been slow to 'disappear' despite more pronounced climate change at broader scales, giving species longer to adapt via selection or dispersal. The pollen record, although incomplete, appears to support relatively low climate velocities in the EAMs during the Late Quaternary period (Mumbi *et al.*, 2008; Finch *et al.*, 2009). Similarly, analysis of macroclimatic shifts since the last glacial maximum suggests low climate velocities in centres of endemism more generally (Sandel *et al.*, 2011). These deeper time perspectives link to key questions for biogeography and conservation planning, such as which sites might remain buffered under the present warming trend, and whether linking past and future change through theory could help planners to mitigate biodiversity loss.

Plant data

Point distributions for plant species currently understood to be endemic to the EAMs were downloaded from TROPICOS (http://www.tropicos.org/). The raw data consisted of 3526 herbarium specimens spanning 471 endemic species in 80 vascular plant families, of which 262 species (56%) are endemic to a single mountain bloc. Due to narrow ranges and threats to habitat, the majority of these species are being assessed for possible inclusion on the IUCN Red List of Threatened Species (Gereau *et al.*, 2010; IUCN, 2011). Botanical identifications were verified at the Royal Botanic Gardens Kew (UK), the Missouri Botanical Garden (USA) and the University of Dar es Salaam (Tanzania). Discrepancies between recorded (specimen labels) and remotely-sensed (SRTM version 4; Jarvis *et al.*, 2008) elevations were investigated alongside field notes, ensuring that each specimen was reliably georeferenced to the correct 100 m altitudinal band. Unambiguous errors were corrected (e.g., confusion between UTM zones or between eastings and northings);

unresolved records were discarded. This resulted in a dataset spanning 444 species, of which 354 had records of occurrence in at least two distinct 1 km grid squares and 56 in at least ten distinct grid squares (Fig. 5.1b).

Climate data

Regional climate forecasts were obtained from the REMO RCM (Potsdam Institute for Climatic Impacts Research), which operates at a horizontal resolution of 55 km (0.5°). Boundary conditions were from the ECHAM5 GCM (210 km horizontal resolution, 1.9°), which shows good agreement with observed 20th century climate in East Africa (Doherty *et al.*, 2010). As for other GCMs, ECHAM5 predicts hotter and wetter conditions in East Africa (17% increase in both annual mean temperature and annual rainfall by 2090). Two IPCC-AR4 scenarios were available for REMO. The first, scenario A1B, describes a world in which economic growth and globalisation prevail, and where energy sources are balanced across fossil-intensive and no-fossil technologies. Global population peaks mid-century and declines thereafter. The second scenario, B1, assumes similar population and convergence among regions as A1B, but with more rapid improvements in public services and economic structures. The emphasis is on clean and resource-efficient technologies, leading to a reduced warming trend (IPCC, 2007). Given recent patterns in global energy consumption and sluggish rates of decarbonisation, we note that both these storylines may now be considered optimistic (Peters *et al.*, 2012).

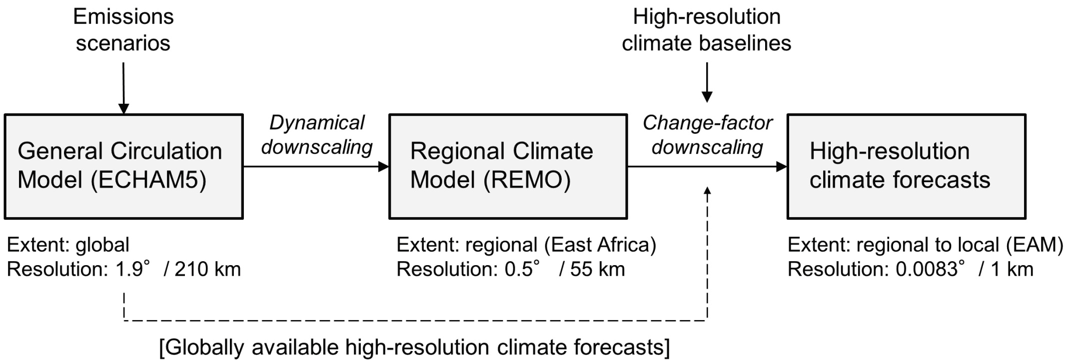


Figure 5.2. Two-step climate downscaling procedure exploits mechanistic models to their spatial limits, beyond which fine-scale variation is inferred from contemporary estimates. Step 1: parameterise a regional climate model using boundary conditions from a subset of GCM squares. Step 2: calculate 21st century anomalies, interpolate and add to higher-resolution baseline grids.

Regional anomalies for the years 2055 (mean of 2046-2065) and 2090 (2081-2100) were calculated relative to present-day conditions before being spatially interpolated to 1 km using regularised splines with tension (Mitasova and Mitas, 1993). These smoothed surfaces were then added to high-resolution baseline grids (Fig. 5.2). For temperature baselines, we used 1 km monthly interpolations of records from the period 1950-2000 (mean 1975; WorldClim, Hijmans *et al.*, 2005). Rainfall baselines were from the Tropical Rainfall Measuring Mission (1997-2006, mean 2001), post-processed to obtain 1 km monthly grids of surface-received orographic rain (Mulligan, 2006). Due to the sparse distribution of rain gauges in the EAMs, these satellite data provide more realistic rainfall patterns than interpolated climatologies such as WorldClim. Mist- and fog-affected zones could not be identified directly, but might to some extent reflect patterns in the rainfall grids, which include corrections for wind velocity, slope aspect and topographic exposure (Mulligan, 2006).

From these monthly grids, we derived five climatic gradients known to correlate well with plant distributions in the study region (Chapters 3 and 4; Greve *et al.*, 2011): mean annual temperature and temperature seasonality (annual range); mean annual rainfall and dry season water stress; and a moisture index, which provides a parsimonious measure of annual conditions (ratio of annual rainfall to potential evapotranspiration, according to Thornthwaite, 1948). Water stress is defined as the cumulative deficit in mean monthly rainfall throughout the longest dry season, where a deficit is < 10 mm.month^{-1}.

Climates lost and gained

Sites were identified where future conditions might extend beyond climate-space accessible to species in the present (climates ‘gained’), as well as the reciprocal: areas where current climates may no longer be accessible in the future (climates ‘lost’). The search for climate analogues extended to all grid cells within, but not beyond, each of the 13 mountain blocs that comprise the EAM chain: lowland habitats and large rivers are significant barriers to dispersal. Results were mapped according to the proportional ‘distance’ of climates gained, or lost, beyond present or future extremes, respectively (Chapter 3; Williams *et al.*, 2007). These values were summed over four of the five summary variables described above (moisture index was omitted to provide an equal balance between temperature and rainfall, and between annual and seasonal conditions):

$$Gain_{i,b} = \sum_{X \in S} \frac{\max\left\{ F_{X,b,i} - \max(P_{X,b}), \min(P_{X,b}) - F_{X,b,i}, 0 \right\}}{\max(P_{X,b}) - \min(P_{X,b})},$$

$$Loss_{i,b} = \sum_{X \in S} \frac{\max\left\{ P_{X,b,i} - \max(F_{X,b}), \min(F_{X,b}) - P_{X,b,i}, 0 \right\}}{\max(F_{X,b}) - \min(F_{X,b})},$$

where S is the set of four climate variables, $P_{X,b,i}$ and $F_{X,b,i}$ are present and future values of variable X in cell i of mountain bloc b, and $P_{X,b}$ and $F_{X,b}$ are the sets of present and future values of X across all cells in the same bloc.

Overlaying the plant data with maps of climate loss revealed how many species will be forced to adapt their climatic niche during the 21st century, or else persist in refugia at the sub-1 km scale, regardless of forest connectivity or dispersal capacity within their current mountain bloc distributions. Under the ecoclimatic stability hypothesis, areas with historically stable climates are thought to contain higher levels of endemism (Fjeldså *et al.*, 1997; Lovett *et al.*, 2005). By extension, one might expect climates lost in the 21st century to exhibit some negative correlation with patterns of endemic plant richness, a possibility we investigated by plotting the former against the latter at known endemism sites.

Endemism* vs. *altitude

Univariate response

The popular application of temperature lapse-rate models in preference to spatially explicit, multivariate climate projections, particularly in mountain regions (e.g., Colwell *et al.*, 2008; Kreyling *et al.*, 2010; La Sorte and Jetz, 2010), carries an implicit assumption that seasonality and rainfall are less important than annual mean temperatures in limiting species' altitudinal (latitudinal) distributions, or at least that changes in these variables will act upon distributions in a similar way. Accordingly, we tested the null hypothesis that peak plant endemism would shift upslope in response to forecast changes in seasonality and rainfall, as well as to warming along the altitudinal gradient. For each of the 354 species with two or more spatially distinct populations recorded (on a 1 km grid; Fig. 5.1b), we inferred continuous climatic suitability, subject to forest and bloc distributions, within the range of conditions at collection sites. Postulating that each climate variable, in turn, is solely important for determining species' range limits, these one-dimensional envelopes were forced under climate scenarios and then summed over species to reveal net changes in the altitudinal distribution of endemism.

Multivariate response

For 56 species with ten or more spatially distinct occurrence records (Fig. 5.1b), we used generalised additive modelling to investigate the relative importance of temperature, seasonality and moisture in determining range limits. For each species, performance-weighted multi-models were obtained using stepwise selection (Chapter 3). Highly correlated variables (annual rainfall *vs*. moisture index, Pearson's $r = 0.92$) were separated prior to selection based on univariate model performance. Where possible, statistical normality was improved using power transforms (Appendix 5A). Each predictor was allowed between one and four effective degrees of freedom, optimised according to a cross-validation of the sum of squared residuals (Yee and Mitchell, 1991). Linear fits were preferred where smooth terms did not improve predictive performance under cross-validation. Background data were distributed within the same mountain blocs as the presence data (prevalence = 0.2), and specifically at locations where other endemic plant species have been recorded. Thus, absences exhibited similar spatial, environmental and taxonomic bias as presences (Phillips *et al.*, 2009; Ahrends *et al.*, 2011 in Appendix I). Each regression model was iterated ten times using different realisations of background data. The final model for a given grid square was the median prediction across these ten runs. Post-model analysis was restricted to those species for which robust predictive fits were achieved (five-fold cross-validation of the area under the ROC curve, $AUC_{CV} \geq 0.7$). We recorded which climate variables were most often retained during selection and the extent to which species-specific contractions/expansions and altitudinal shifts varied between populations on different mountain blocs.

Results

Spatial variation in change

Annual mean temperature was found to increase similarly across the study region (Appendix 5B), reaching a median 21.3º C (scenario B1) or 22.5º C (A1B) by 2090 (interquartile range [IQR] in percentage change from baseline conditions: B1 [12, 15] %; A1B [18, 22] %). Forecasts for other important determinants of species distributions were not so readily anticipated (Fig. 5.3 and Appendix 5B). For example, in West Usambara forests, temperatures become less seasonal (B1 [-4, -3] %; A1B [-5, -4] %) and annual rainfall is predicted to increase (B1 [9, 16] %; A1B [14, 33] %), resulting in a stable moisture index and little change in dry season water stress, despite the rising temperatures. In Uluguru forests, seasonality in temperature and rainfall remain constant, whilst increased potential for evapotranspiration tips the moisture index toward drier conditions by the end of the century. In Udzungwa forests, farther south and more distant from the coast, temperature seasonality is predicted to increase (IQR 2090: B1 [3, 4] %; A1B [8, 10] %) and annual rainfall to decrease (IQR 2090: B1 [-16, 0] %; A1B [-23, 2] %), resulting in a lower moisture index and an increasingly harsh dry season.

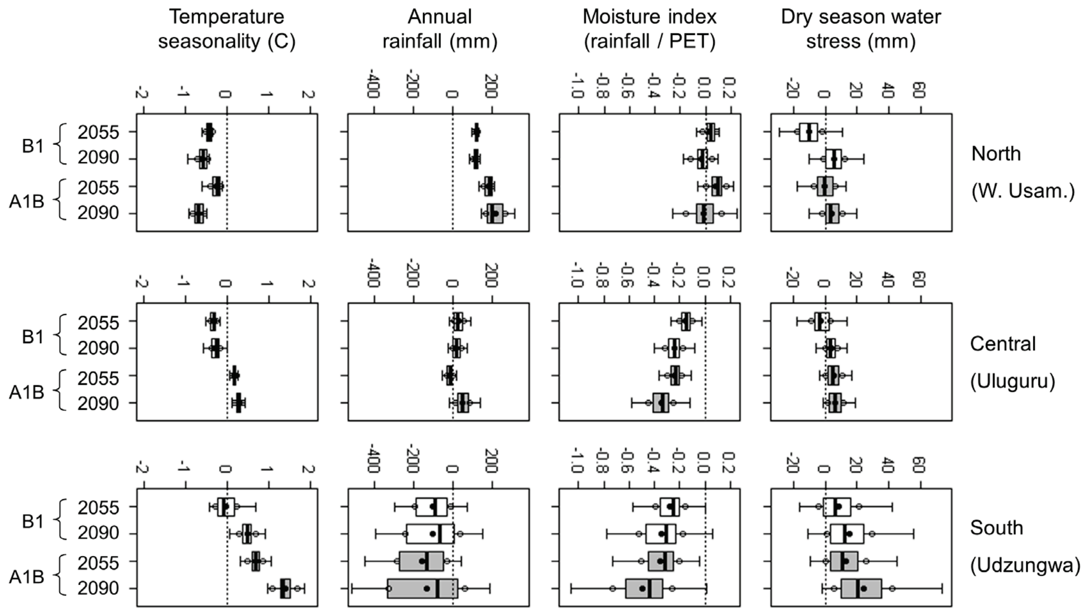


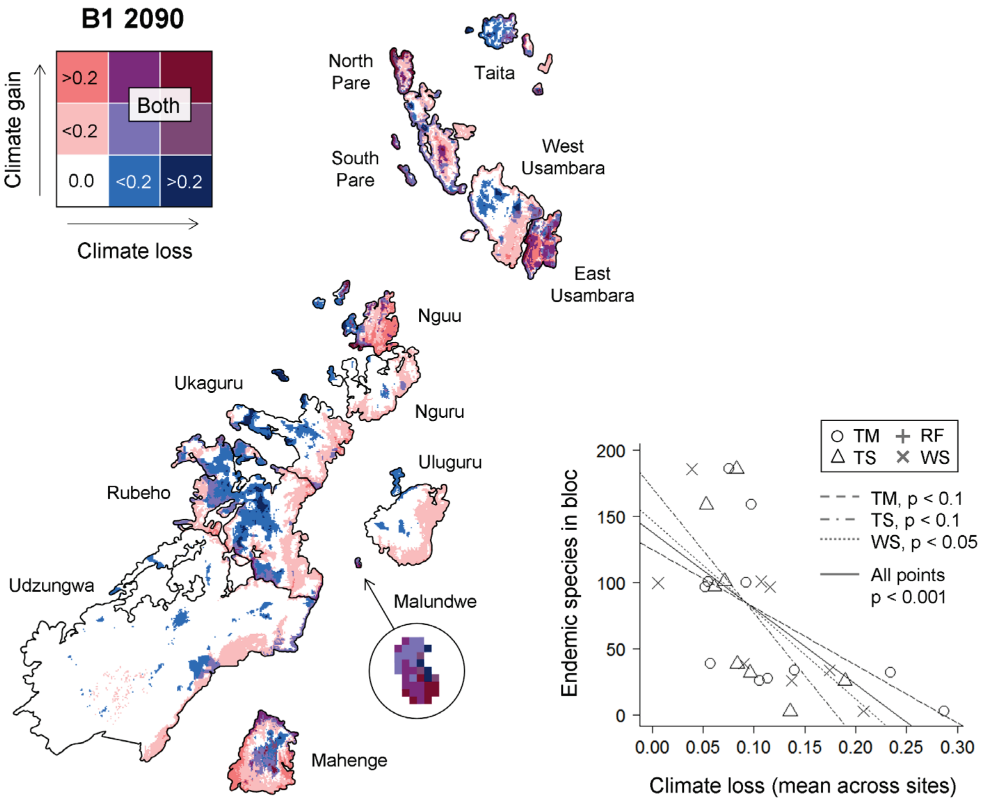
Figure 5.3. Spatial variation in climate anomalies at sub-regional scales (future minus present, IPCC-AR4 emissions scenarios B1 [white] and A1B [grey]), focussing on forests in three of the best studied and most biologically important mountain blocs: West Usambara, Uluguru and Udzungwa. Dotted lines correspond to baseline conditions. Whiskers extend up to 1.5 times the interquartile range. Filled and open circles show means ± standard deviations, respectively. Forecasts for mean annual temperature are similar across all mountain blocs (Appendix 5B), whereas measures of seasonality and moisture vary considerably from north to south and with distance from the coast.

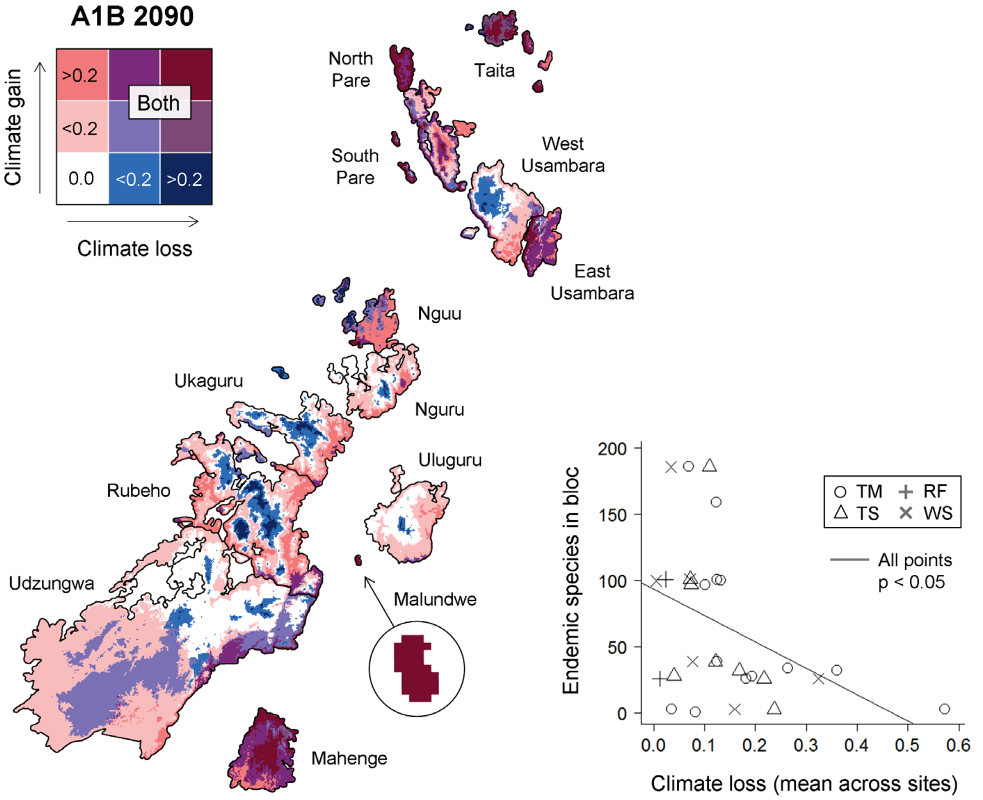
Climates lost and gained

Under both emissions scenarios, cool summit conditions are lost to rising temperatures (Fig. 5.4). Contemporary conditions on the highest peaks of West Usambara, Uluguru and, to a lesser extent, South Pare and Nguru, have no climate analogue anywhere in the EAM chain by 2090. Some such sites are grassland, lacking forest cover and associated levels of endemism due to factors such as frost, fire and lower incidence of mist (Finch and Marchant 2011). In some cases, higher temperatures could potentially lead to more forest cover due to reduced chance of frost, whilst other areas could maintain or expand grassland habitat through increased fire frequency.

The coincidence of climates lost and gained (purple shades in Fig. 5.4) indicates that not only will a species *in situ* be subject to conditions currently found nowhere else within the bloc (climate gain), but that it may also have no opportunity to colonise sites with familiar climate analogues by way of within-bloc dispersal (climate loss). Such conditions are predictably extreme in Malundwe (an isolated peak with just 2 km^2 of forest) due to its small extent and present lack of climatic heterogeneity. However, especially under A1B, such a forecast is also apparent for Taita, North and South Pare, East Usambara, Nguu, Udzungwa and Mahenge. For the first five, higher annual temperatures coincide with reduced seasonality and, in the case of East Usambara, increased dry season water stress on lower slopes (despite higher annual rainfall; Appendix 5B). Increased annual rainfall (loss of dry annual conditions) contributes to purple shading in the Pares and Nguu. In Mahenge and on the high plateaus of Udzungwa, higher temperature seasonality is gained, coupled with the loss of cooler mean annual conditions, whilst high temperature seasonality is lost in Matundu forest (lower slopes of east Udzungwa).

Figure 5.4. Climatic conditions lost and gained in the EAMs under scenarios B1 and A1B by the year 2090. Search for climate analogues is restricted to individual mountain blocs, reflecting the range-restrictions of many endemic plant species (median recorded range = 1 bloc). Climate loss is scaled by the dissimilarity of present-day conditions compared with future climatic extremes within the same mountain bloc, with respect to: mean annual temperature (TM), temperature seasonality (TS), mean annual rainfall (RF) and dry season water stress (WS). Climate gain is the proportional dissimilarity of future conditions beyond present-day bloc extremes. Scatter plots compare no-zero loss (mean across known endemism sites) with endemic plant richness. Regression lines and corresponding p-values (F-tests) show significant relationships. Results for 2055 are presented in Appendix 5C.


B1 2090
Climate gain
>0.2
<0.2
0.0
Both
<0.2
>0.2
Climate loss
North Pare
Taita
South Pare
West Usambara
East Usambara
Nguu
Ukaguru
Nguru
Uluguru
Rubeho
Udzungwa
Malundwe
Mahenge
Endemic species in bloc
0
50
100
150
200
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Climate loss (mean across sites)
TM
RF
TS
WS
TM, p < 0.1
TS, p < 0.1
WS, p < 0.05
All points
p < 0.001


A1B 2090
Climate gain
>0.2
<0.2
0.0
Both
<0.2
>0.2
Climate loss
North Pare
Taita
South Pare
West Usambara
East Usambara
Nguu
Ukaguru
Nguru
Uluguru
Rubeho
Udzungwa
Malundwe
Mahenge
Endemic species in bloc
0
50
100
150
200
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Climate loss (mean across sites)
TM
RF
TS
WS
All points
p < 0.05

Of the 719 distinct 1 km sites where endemic plant species have been sampled, 22% (B1) or 17% (A1B) are forecast to experience climate loss by 2055, and 29% (B1) or 51% (A1B) by 2090. A climate lost means that an endemic species must migrate to an entirely different mountain bloc in order to avoid extinction, assuming it cannot adapt *in situ*. Concentrating on locations where loss is non-zero, we found that endemism sites in mountain blocs with greater endemic richness (> 50 species) are generally less susceptible to extreme change (Fig. 5.4). The most significant relationship is for climates lost by 2090 under the B1 scenario ($p < 0.001$, $R^2 = 0.38$; F-test on 25 df), which is arguably more representative of past change than A1B, but also evident for the latter ($p < 0.05$, $R^2 = 0.17$; F-test on 27 df). The observed relationship, of lower endemic richness at sites where scope for within-bloc persistence via dispersal is least likely, is consistent with the ecoclimatic stability hypothesis, especially with respect to dry season water stress (B1 scenario; $p < 0.05$, $R^2 = 0.57$; F-test on 6 df). Relationships between climate loss and endemic richness were also significant for the year 2055 under both emissions scenarios. Climate gain, less directly associated with species persistence (suitable conditions may be accessible elsewhere in the bloc) although still indicative of extreme change, was significant only for A1B 2090 (Appendix 5C).

Endemism* vs. *altitude

Univariate response

Documented plant endemism exhibits an altitudinal peak between 1000 and 1700 m above mean sea level. Constructing one-dimensional climate envelope maps for 354 species and forcing these under scenarios of change shows peak endemism shifting upslope in response to mean annual warming and, to a lesser extent, the annual moisture index (Fig. 5.5). The latter is perhaps a conservative estimate of the potential for downslope movement, given that Thornthwaite PET is insensitive to forecast increases in humidity. A similar analysis focussed on seasonal measures illustrates the opposite trend, with species limited in their distributions by temperature seasonality and/or dry season rainfall expected, on average, to migrate downslope in order to maintain their climatic niches (Fig. 5.5). With annual and seasonal forces acting in opposite directions at the extremes of the altitudinal gradient, it seems that mid-altitude forests will maintain (and possibly further develop) high concentrations of endemic species into the future, although floristic composition at these sites will undoubtedly be different and subject to uncertain outcomes from novel biotic interactions.

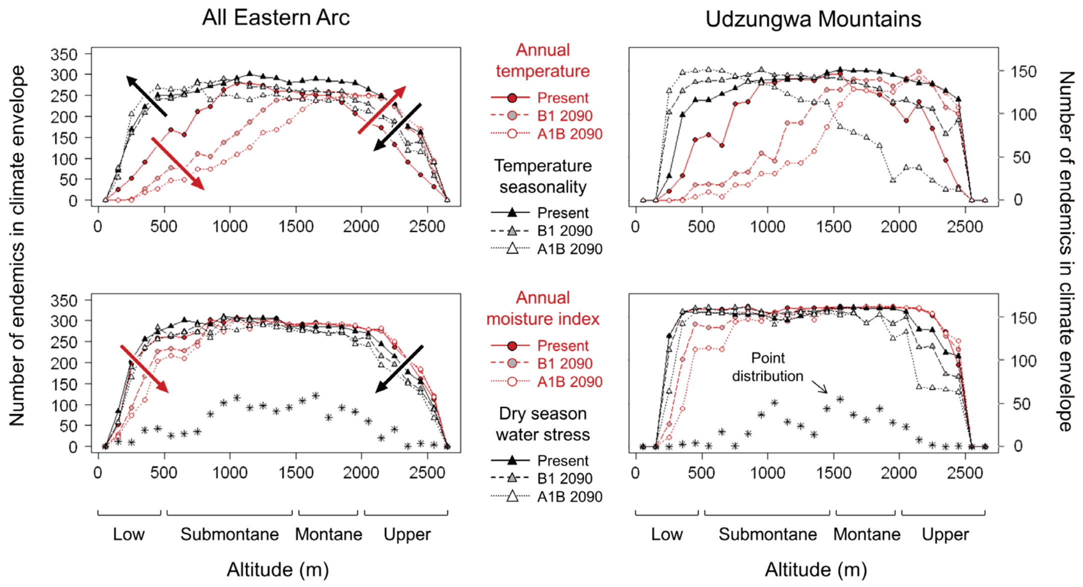


Figure 5.5. Altitudinal distribution of endemism, now and in the future, according to one-dimensional climate envelopes for 354 plant species. Arrows emphasise directions of change. If mean annual conditions (red) are most important for limiting species' distributions, then peak endemism is expected to shift upslope in response to warming. Conversely, if seasonality (black) is more important, then peak endemism could shift downslope. Patterns are stronger for temperature gradients (upper) than for moisture gradients (lower), the latter being more variable within a given altitudinal band. Patterns are broadly consistent across mountain blocs, but especially clear in Udzungwa forests (right panel) where forecast change is most pronounced.

Multivariate response

Of the 56 endemic plant species eligible for multivariate regression, 34 were modelled with sufficient accuracy for further analysis (median AUC = 0.84, range [0.76, 0.96]; median AUC_{CV} = 0.75, range [0.70, 0.86]; median D^2 = 0.45, range [0.19, 0.66]). On average, each climate variable was selected a similar number of times and contributed similarly to model performance, although there was considerable variation among species (Appendix 5D). Temperature seasonality was selected marginally more often (55% of model runs, *cf.* 53% annual temperature, 52% water stress, 49% rainfall or moisture index), and explained more deviance when selected (mean over species = 15%, *cf.* 13% annual temperature, 12% water stress, 11% rainfall or moisture index). Up to nine of the 34 species, depending on the emissions scenario, were forecast to undergo a net increase in climatic suitably across occupied mountain blocs during the 21st century. Around twice as many were forecast with increasing suitability in one part of their range, but decreasing suitability in another, especially under scenario B1 by 2055 (Table 5.1 and Appendix 5D).

Table 5.1. Directions of change in climatic suitability and altitudinal optima (100 m bands) summarised for 34 multivariate species models. Numbers show how many endemic plant species are predicted to experience a net increase or decrease in suitability/altitude across all EAM blocs, and how many respond in opposite directions on different blocs within their range. Species-specific results are presented in Appendix 5D.

		Climatic suitability			Optimum altitudinal band			
		▲	▼	▲▼	▲	◄►	▼	▲▼
B1	2055	9	25	20	12	16	6	23
	2090	6	28	11	11	15	8	23
A1B	2055	6	28	13	10	15	9	21
	2090	6	28	11	14	10	10	22

▲ net increase; ▼ net decrease; ◄► no net change, ▲▼ varies by mountain bloc

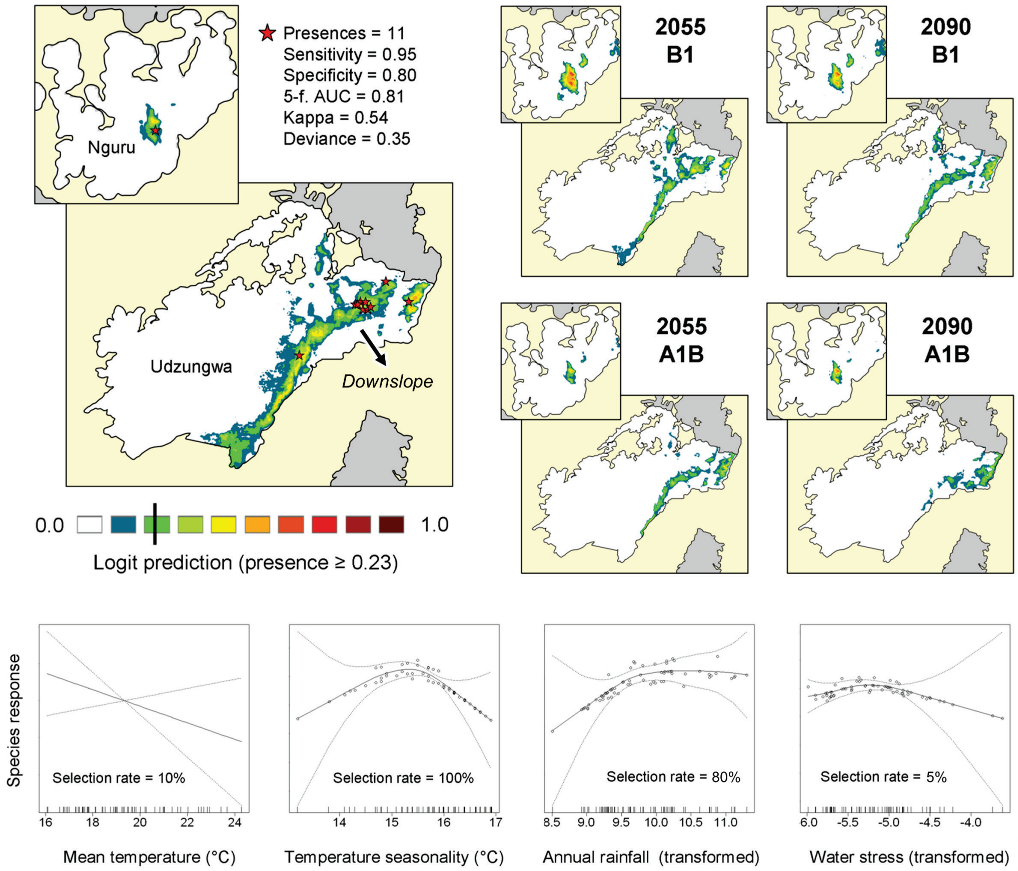
As suggested by univariate analysis (Fig. 5.5), multivariate species models did not predict ubiquitous upslope shifts in species' altitudinal optima. Of the 34 species, 10-14 were projected to undergo upslope shifts, 6-10 downslope, and 10-16 no net change. In order to maintain their current climatic niches, around two thirds of the species are required to migrate upslope in some parts of their range, but downslope in others (Table 5.1 and Appendix 5D). A spatially explicit illustration of how directions of change could vary both within and between species is provided in Fig. 5.6, which maps the contrasting responses of *Tricalysia aciculiflora* and *Danais xanthorrhoea* (both Rubiaceae) in Udzungwa, whilst both species maintain relatively stable ranges in Nguru.

Discussion

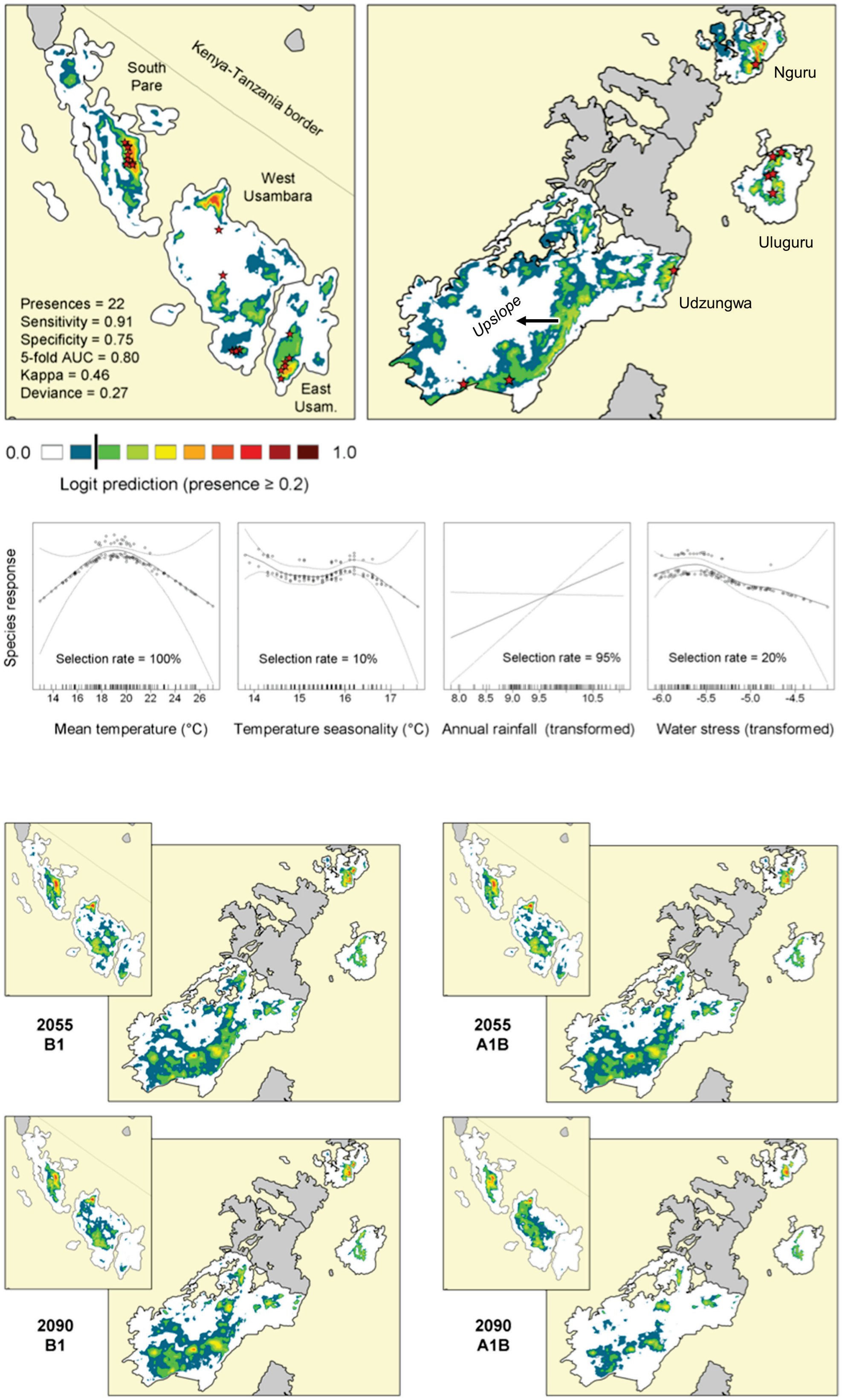
Progress in bridging the gap between coarse-resolution global datasets and the high-resolution, regionally focussed information necessary for the development of practical climate change adaption strategies has, to date, exhibited bias toward lowland regions and temperate biomes (Wilby *et al.*, 2004). The highly biodiverse and habitat-heterogeneous tropical zone could experience climate change impacts sooner than temperate regions (Beaumont *et al.*, 2011), but many countries, particularly in Africa, lack the species information, climate data and monitoring systems necessary to guide environmental policy (IPCC, 2007). Since much of the rapidly growing African population depends directly on natural resources for fuel, building materials, medicine and food, such a data gap and lack of fundamental understanding on the possible impacts of climate change on the natural world is a major concern. The local, national and global goals of poverty reduction, sustainable development, biodiversity conservation and climate change mitigation are highly interwoven (Rapport *et al.*, 1998), and predicting outcomes for any and all is strongly hampered by a scarcity of meaningful data at the relevant spatial scales, especially in mountainous regions.

Figure 5.6. Present day distribution of *Tricalysia aciculiflora* (a) is best explained by temperature seasonality and annual rainfall. Climatic suitability is stable (A1B) or increasing (B1) in Nguru. In Udzungwa, increased seasonality and less rainfall may push the species downslope toward the north-east of the bloc. Conversely, *Danais xanthorrhoea* (b) is predicted to have a similar Udzungwa range to *T. aciculiflora* at present but climate suitability tends in the opposite direction over time (upslope to the south-west). Suitability for *D. xanthorrhoea* is stable in Nguru, expands upslope in South Pare, West Usambara and Uluguru, but contracts in East Usambara.

(a) *Tricalysia aciculiflora* Robbr.



(b) *Danais xanthorrhoea* (K. Schem.) Bremek.



GCM forecasts are the backbone of climate change assessment, but at their native or interpolated resolutions they provide limited insight into the spatial heterogeneity of change important for site-scale conservation. As RCM predictions become more widely available through initiatives such as CORDEX (Coordinated Regional Climate Downscaling Experiment; http://wcrp.ipsl.jussieu.fr/), more studies can take account of the unique regional forcings that underpin species persistence and ecosystem change in their study areas. In this paper, we have shown that climate change impacts on endemic mountain plants could vary considerably from site to site and from species to species. Further, by looking at variation in endemic species richness between the dispersal-limiting island systems of a mountain archipelago, it has been possible to correlate an uncertain future with hypotheses of past ecosystem stability. Our findings demonstrate that, despite the plethora of global climate impact assessments in the scientific literature, conservation planners working on the ground would be ill advised to base adaption strategies on zoomed perspectives of these broad-scale models. Overlooking fine-scale variation, regional climate dynamics or other factors relevant to historical stability, especially in the highly biodiverse tropical mountains of Africa, could result in misguided policy and wasted conservation/adaption investment.

In the downscaling methodology presented here, we obtained high-resolution forecasts from a RCM by applying change-factors to contemporary climate grids. In regions with sufficient observational time-series data, an alternative to change-factor downscaling would be to correlate large-scale atmospheric variables with local climatic conditions (statistical downscaling). However, the statistical downscaling of climatic change in mountain regions is a complex and data-hungry procedure (Dobrowski *et al.*, 2009) and few, if any, such studies exist for the tropics (Wilby *et al.*, 2004). Change-factor analysis applied to regionally focussed climate models is therefore a more realistic goal for the near term. At local scales, it is important to note that both procedures are insensitive to changes in land-surface feedback. Anecdotal evidence from residents and long-term researchers in the EAMs tells of the cloud line shifting upslope during the 20th century, as forest at lower elevations has been cleared (see also Fairman *et al.*, 2011). In East Usambara, predicted here to experience higher annual temperatures but lower temperature seasonality, and increased annual rainfall but harsher dry seasons, there are reports of reduced regeneration of some tree species, attributed to a lower incidence of mist in recent decades (Hamilton and Bensted-Smith 1989). In anticipation of how socio-economic forces might play out across landscapes, there is scope to embed detailed scenarios of land use change within sub-national climate downscaling projects. There exists a wealth of local knowledge on vegetation-climate feedbacks in mountain regions, which should not be overlooked in such a process. In particular, greater understanding is needed on how micro-climatic variation responds to broader-scale change, especially with regard to fog, mist and conditions beneath the forest canopy.

Some previous studies have minimised downscaling complications by reducing the forcing mechanism in species' distributions to a monotonic relationship between altitude and temperature, which predicts upward range-shifts, mountaintop extinction and lowland biotic attrition in accord with broader-scale warming trends (Colwell *et al.*, 2008). We caution, however, that findings based solely on annual temperature are confounded by changes in seasonality and rainfall (Knapp *et al.*, 2008; McCain and Colwell, 2011), shown here to be equally important for explaining observed patterns of occurrence, if not spatial and temporal variation in the lapse rates themselves (Peyron *et al.*, 2000). We find that two thirds of multivariate species models predict populations of endemic flora to decrease in elevation in at least one mountain bloc, and that any population whose range is predominantly governed by water stress or low temperature seasonality might be expected to tend downslope rather than up.

Tropical species, subject to lesser seasonal overlap in thermal regimes between low and high altitudes, are thought to be especially sensitive to temperature change (Janzen, 1967; see also Ghalambor *et al.*, 2006; Cadena *et al.*, 2012), but empirical studies do not necessarily support upward shifts at the magnitude forecast by warming models (Forero-Medina *et al.*, 2011; Scherrer and Koerner, 2011). Whilst there exist numerous examples of upslope migration in response to climate change in the 20th century (e.g., Lenoir *et al.*, 2008; Jump *et al.*, 2012), it would be premature to suppose this pattern will hold true for all species in all regions (Lenoir *et al.*, 2010). Alongside upward shifts for 118 plant species in Europe, Lenoir *et al.* (2008) also report downward shifts for 53 species, suggesting that climatic factors other than mean annual temperature may be mediating their distributions. Similarly in California, coastal flora have shifted their altitudinal optima downhill, despite warming along the gradient, due to changes in the water balance (Crimmins *et al.*, 2011) – phenomena that cannot be explained by adiabatic lapse rates alone. Most fundamentally, when anticipating climate change impacts on species, it is preferable that predictor variables have direct, or at least physiologically relevant, influence over observed patterns of occurrence: relations to proximal measures such as altitude are mediated by factors beyond the model's scope, which themselves could be subject to change (Austin, 2007).

At continental scales, low seasonality correlates with higher species richness and rarity (Jetz *et al.*, 2004). Here we find this variable to be an important predictor of intra-specific distribution at the landscape scale. Simple climate envelope models suggest that the downslope influence of changes in seasonality, combined with the upslope forces of warming, could maintain, if not accentuate, a mid-altitudinal peak in endemic species richness. This pattern has parallels with the mid-domain effect, whereby hard boundaries at opposite ends of the altitudinal gradient are said to promote greater species overlap at intermediate elevations (McCain and Colwell, 2011). The past contraction of species range limits toward mid-elevations, leading to novel vegetation assemblages composed of both high and low altitude taxa, has been recorded in the pollen record, for example on Mt. Kenya following the last glacial maximum (Rucina *et al.*, 2009). Further, the predicted accumulation of plant species at mid-altitudes coincides with a peak in the richness of ancient bird species (Fjeldsa *et al.*, 2010). Based on these findings, we suggest that submontane refugia, and the migration pathways to and from them, be given high priority in conservation plans.

For clarity in assessing climate change impacts on rare mountain flora, we have focussed on species strictly endemic to the EAMs proper (Chapter 2; Gereau *et al.*, in prep.). The possibility of currently low-elevation flora migrating upslope in response to warming, or otherwise, was not considered explicitly. In today's human-dominated landscape, the lower forest edge is defined both by climatic factors and by reserve boundaries and challenging terrain, which limit agricultural encroachment and production. Opportunities for gradual migration from lowland to mountain habitats, or *vice versa*, are therefore limited mainly to species capable of long distance dispersal. Since we restricted our analysis to accurately georeferenced, and therefore mostly recent, collections, some putative mountain endemics may in fact be persisting at the upper altitudinal limits of previously more extensive distributions, curtailed at lower elevations by land use change (especially forest loss). Warming might conceivably benefit such species by restoring their thermal optima, a possibility missed by climatic regression against post-clearance species distributions.

Macroclimatic shifts during the Late Quaternary suggest that extreme episodes of past climate change impacted species in some regions less severely than in others (Sandel *et al.*, 2011), promoting accumulation of now relictual taxa in centres of endemism such as the EAMs. Concordant with hypotheses of long-term ecoclimatic stability (Fjeldså *et al.*, 1997; Lovett *et al.*, 2005), we find that endemic richness in the EAMs is significantly higher where species have the potential to maintain familiar conditions throughout the 21st century by within-bloc dispersal, suggesting that spatial heterogeneity in future climatic change could be similar to that of the past. The current rate of climate change, although exacerbated by anthropogenic effects, is not without precedent (McInerney and Wing, 2011). For example, the Younger Dryas (*c.* 11,000 yr BP) was characterised by a very rapid warming episode, with temperatures in some regions increasing by as much as 7 °C in just 50 years (Dansgaard *et al.*, 1989). However, on a Quaternary time-scale there has never been a change in climate from a warm to warmer state concomitant with high concentrations of atmospheric CO_2, such as that which is currently being experienced. Above all, the ability of narrow-ranged mountain plants to respond to future change will be very different to that of the past: the processes by which organisms were previously able to adapt, adjust and migrate have been massively curtailed in a landscape dominated by people and agricultural/pastoral land (Fjeldså and Lovett, 1997).

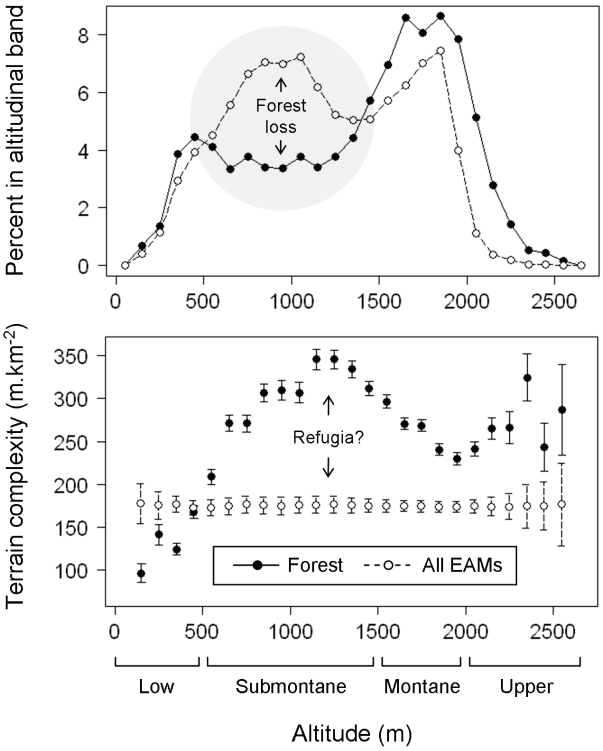


Figure 5.7. Altitudinal distribution (100 m bands) of forest compared with all mountainous land in the EAMs (upper panel). Shaded portion corresponds to extensive forest loss in the submontane zone during the 20th century, impeding the ability of species to track climatic change along the altitudinal gradient. Lower panel shows higher than average terrain complexity (altitudinal range within 1 km squares) in the forests that remain, suggesting potential for topoclimatic refugia. Error bars plot 10 standard errors around the means.

Migration speeds of long-lived tropical trees, which define the forest canopy and mediate fog-capture and micro-climatic conditions in the understory, are relatively slow. Even for species with reproductive cycles in accord with future climate velocities, anthropogenic fragmentation of the forest mosaic has compromised migration pathways and agents for dispersal (e.g., birds; Cordeiro and Howe, 2003). In the EAMs, there has been a peak in deforestation at low- to mid-altitudes (Hall *et al.*, 2009), presenting a serious obstacle to the patterns of within-bloc movement considered here. In favour of *in situ* persistence, the forests that remain occupy significantly more complex terrain (altitudinal range within 1 km squares) than the background mountain area ($p < 0.001$, Wilcoxon rank sum; Fig. 5.7), suggesting the potential for small populations to be maintained in topoclimatic refugia, at least in the short term (Randin *et al.*, 2009; Austin and Van Niel, 2011). The upper montane zone, characterised by complex micro-habitat mosaics, has more often survived agricultural encroachment, again providing narrow-ranged species with the possibly of localised refuge from thermal shifts (Scherrer and Koerner, 2011).

RCM ensemble forecasts, anticipated to become more widely available in time for IPCC-AR5 (which will convey a revised set of emissions scenarios; Rogelj *et al.*, 2012), should help to quantify uncertainty in the REMO/ECHAM5 forcings presented here. In the meanwhile, sensible priorities for management are to protect all remaining forest, regardless of patch size, and to seek to establish migration corridors between isolated fragments, especially in the submontane zone. Where available, information at the genetic level could help to ensure that natural variation is maintained, facilitating species' adaption to uncertain future conditions (Kahindo *et al.*, 2007). At the heart of any successful conservation strategy will be the need to find ways to manage sustainably, and where possible restore, forest ecosystems, whilst addressing the resource needs of increasingly numerous human populations. Solutions to the huge challenges of climate change, biodiversity loss, population growth and pressure on ecosystem services will require further targeted research on both the spatial and temporal character of environmental change, and how this impacts ecosystems and associated livelihoods. It is not sufficient to apply models that are too coarse, or to apply scientific understanding that is not rooted in the ecosystem under investigation.

Acknowledgements

We thank the many providers of plant data to the Missouri Botanical Garden's TROPICOS database, Jon Newell for heroic amounts of specimen data entry, and Stefan Hagemann and Daniela Jacob (Max Planck Institute for Meteorology, Hamburg) for providing REMO climate forecasts for East Africa. Thanks also to Colin McClean for useful discussion and to two anonymous reviewers for their insightful critiques. This work was funded by the Marie-

Curie programme of the European 6th Framework (MEXT-CT-2004-517098), with additional support from the Leverhulme Trust (http://www.valuingthearc.org/), the Ministry for Foreign Affairs of Finland (http://www.chiesa.icipe.org/) and the British Institute in East Africa (http://www.biea.ac.uk/).

Author contributions

The study was conceived by P.J.P. and R.M. P.J.P conducted all analyses and prepared the manuscript. R.E.G. and P.J.P. cleaned the plant data. R.M. advised on historical perspectives and N.B. on management implications.

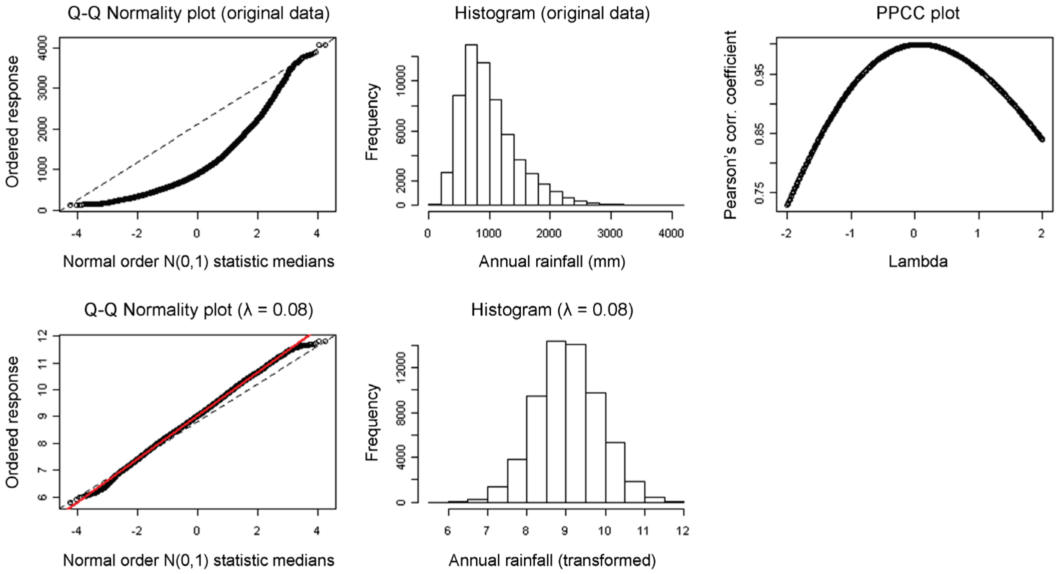
Appendix 5A. Improving normality for statistical regression

For use in multivariate regression (generalised additive models), we attempted to improve the normality of climatic predictors using square root, logarithmic (bases 2, 10 and *e*), inverse and power transformations. For mean annual temperature and temperature seasonality, no improvement was gained using these methods, but departures from normality were not considered serious (Pearson's correlation coefficient $r > 0.98$, ordered response *vs*. normal order statistic medians). For moisture predictors, Box-Cox power transforms were applied. Power transformations work by compressing the right side of the distribution more than the left. For negatively skewed data, we reflected the distribution, adding a constant such that all values were greater than zero, applied the Box-Cox transform [$T(X) = (X^{\lambda} - 1) / \lambda$] and then reflected once more to restore the original order of the variable. The chosen λ was that which optimised linear dependence between the (transformed) ordered response and normal order statistic medians (probability plot correlation coefficient; PPCC plot).

Annual rainfall

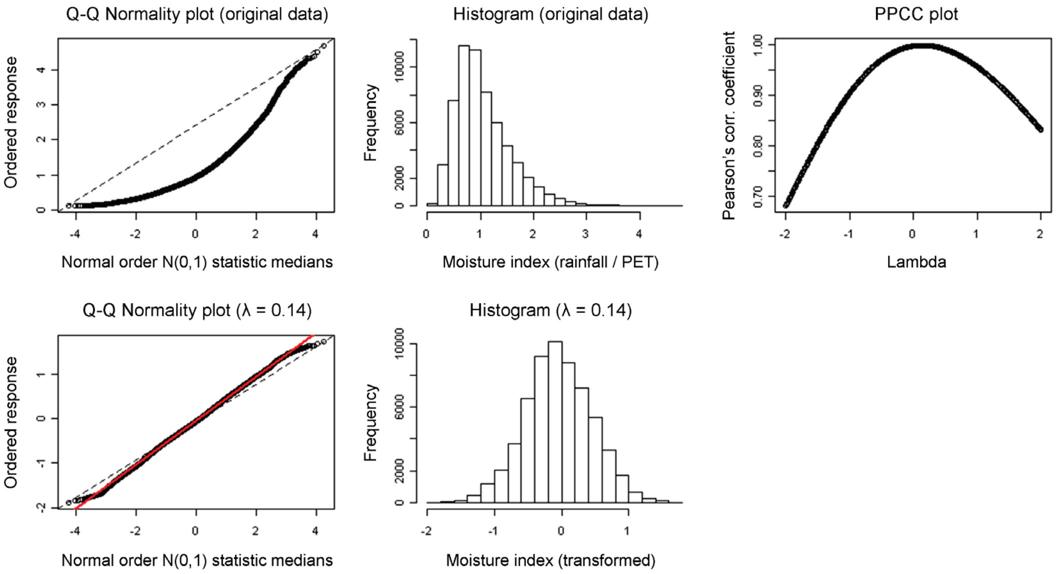
Negative skew; Box-Cox, $\lambda = 0.08$

Pearson's correlation with normal order statistic medians, $r = 0.9997$



Annual moisture index

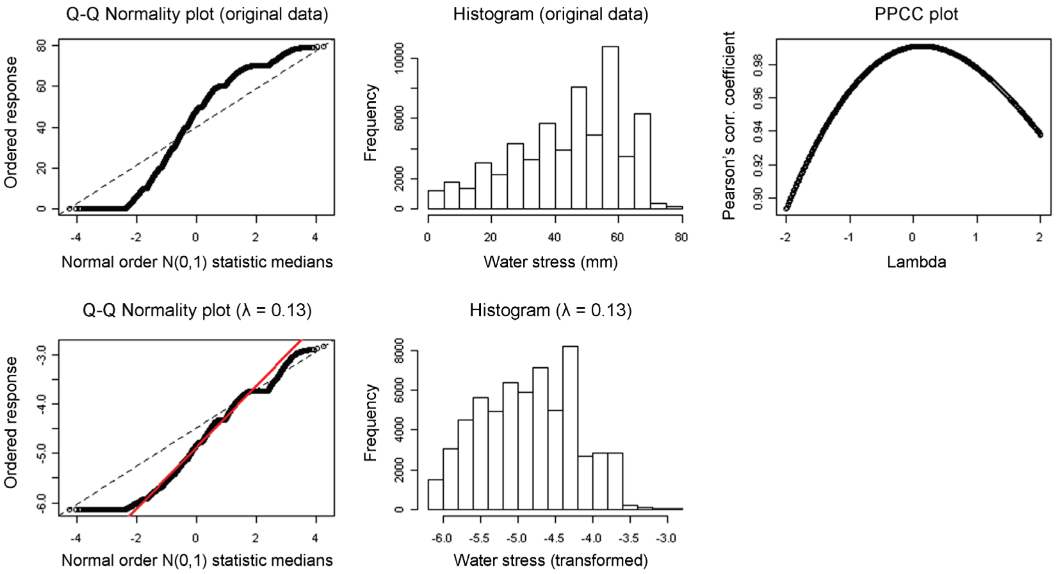
Box-Cox, λ = 0.14. Pearson's correlation with normal order statistic medians, *r* = 0. 9997



Dry season water stress

Positive skew (data reflected, constant = 91). Box-Cox, λ = 0.13

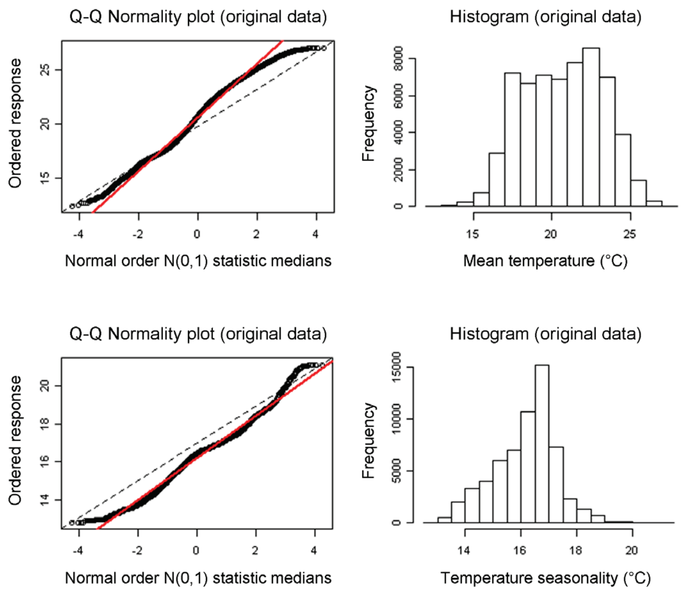
Pearson's correlation with normal order statistic medians, *r* = 0.9913



Mean annual temperature and temperature seasonality

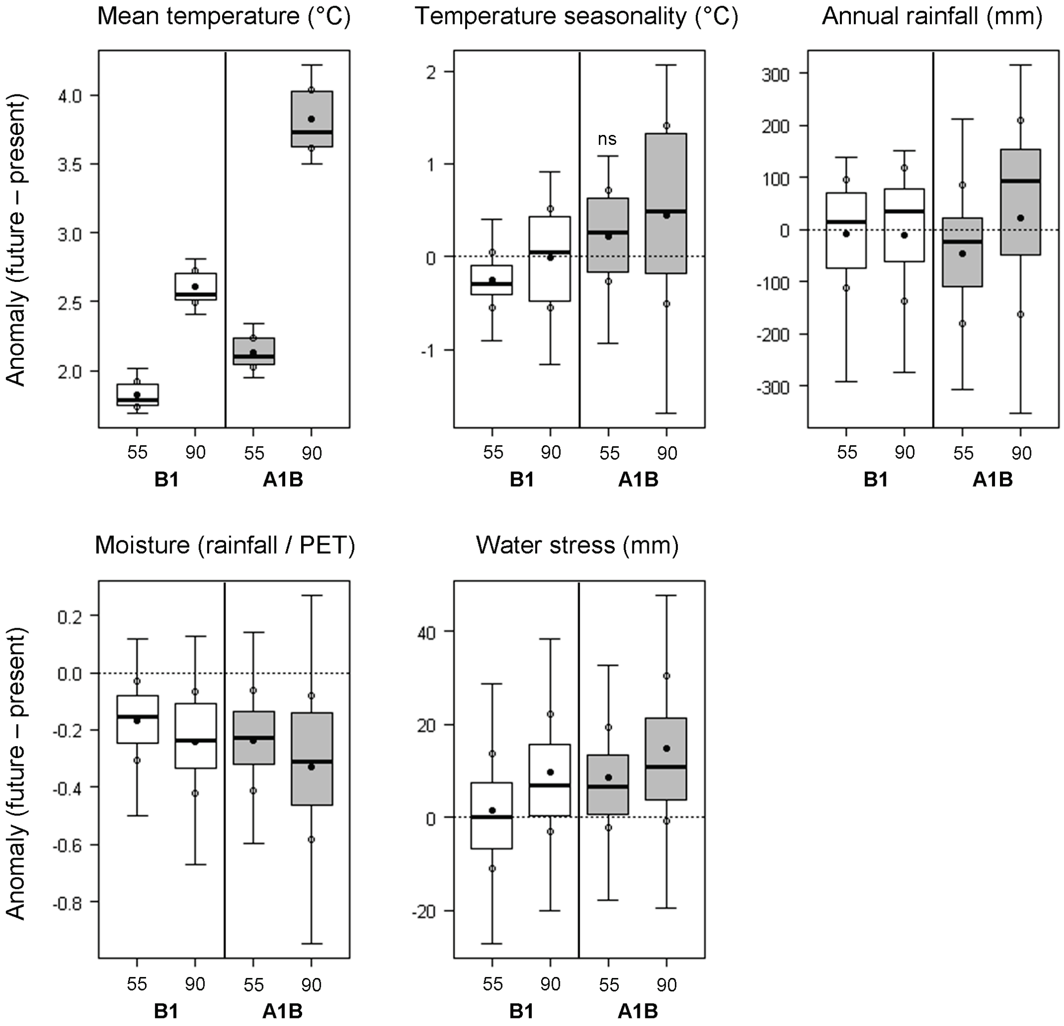
No transformation improved normality

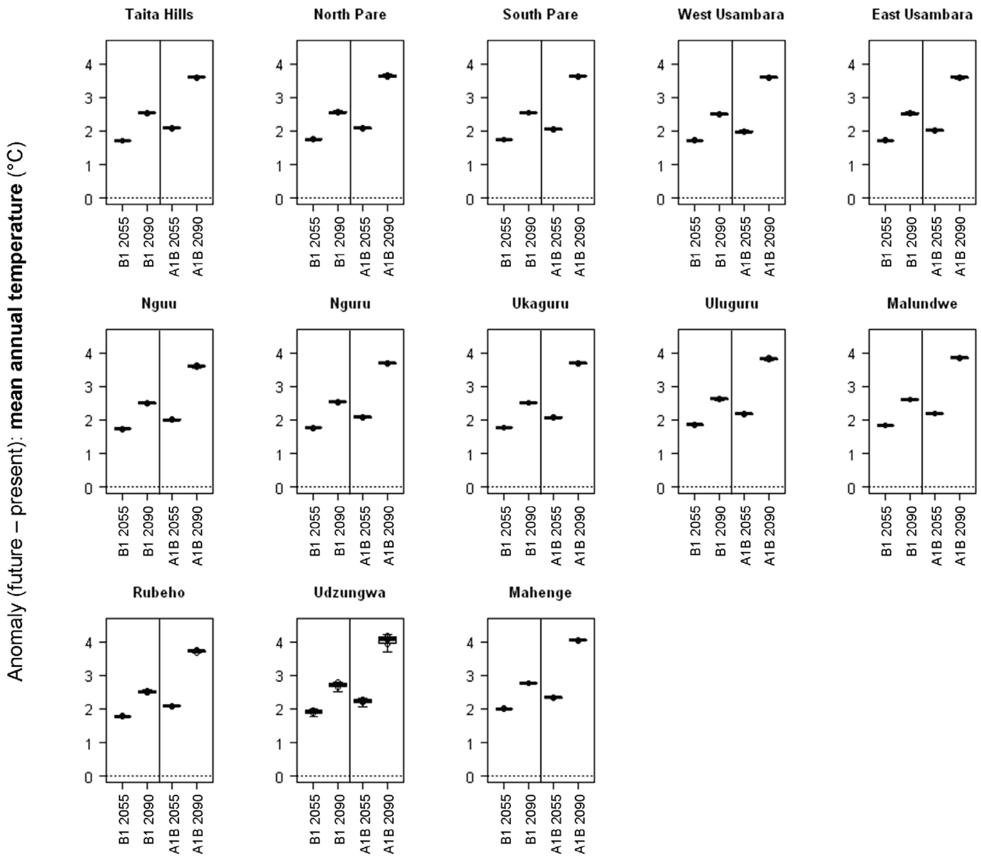
Pearson’s correlation with normal order statistic medians, *r* = 0. 9915 (annual mean) and *r* = 0. 9893 (annual range)

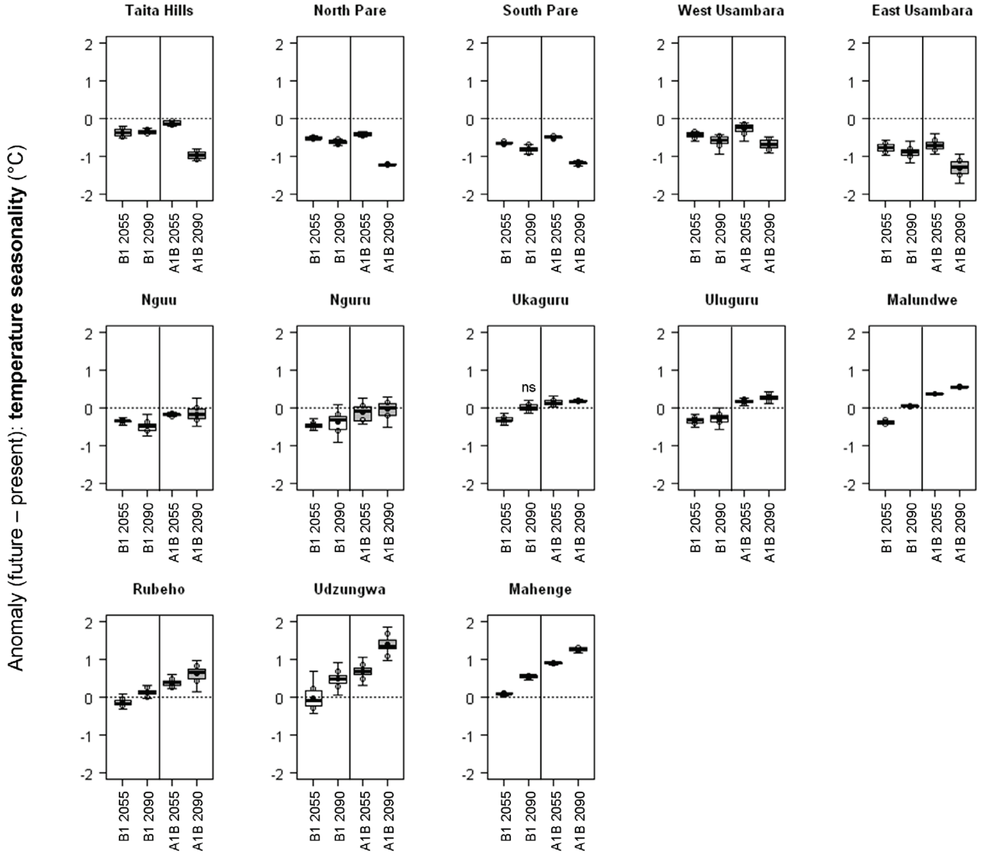


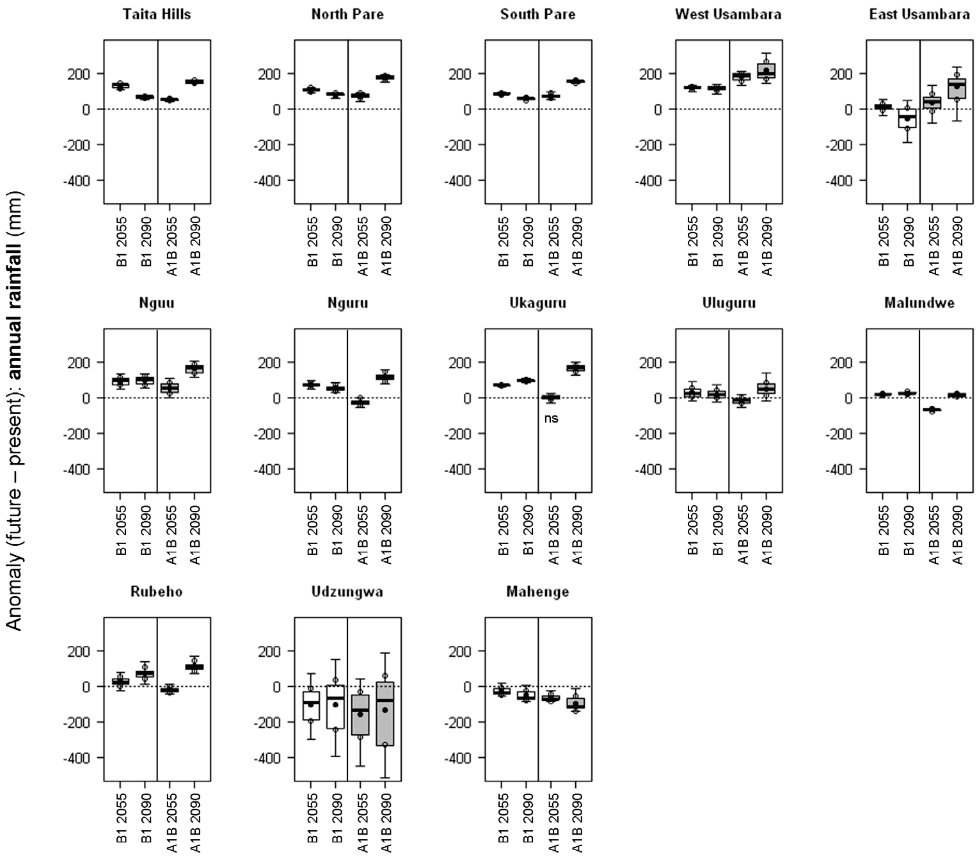
Appendix 5B. 21st century climate anomalies

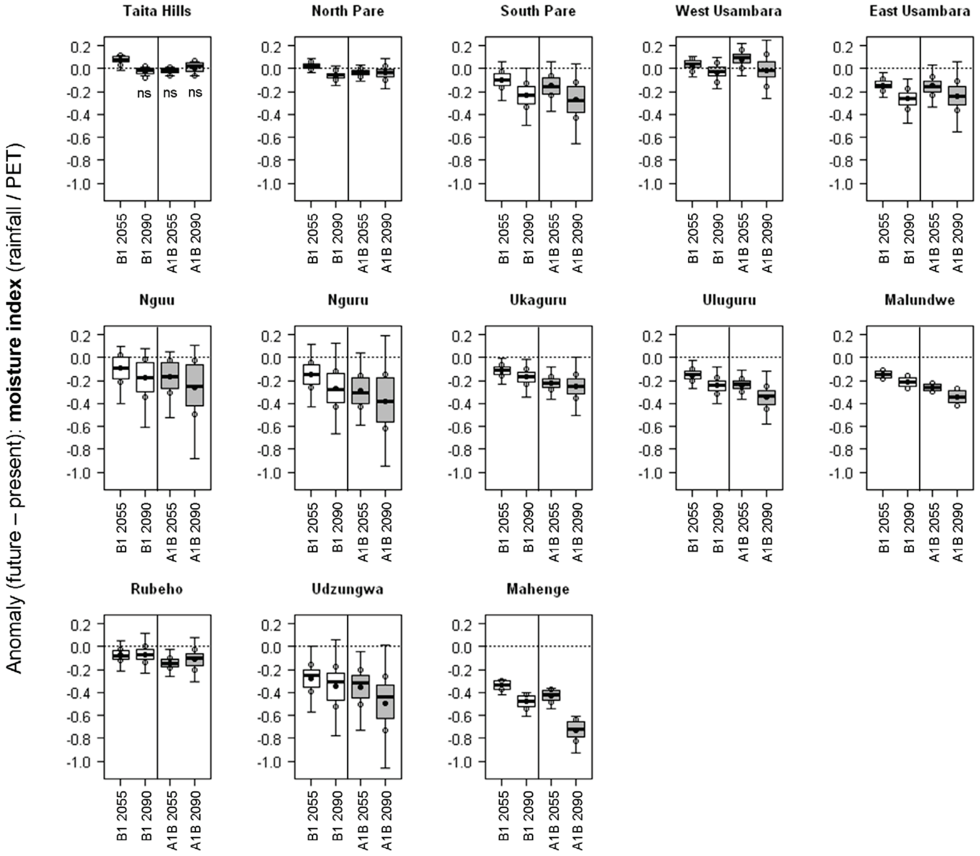
Future climate anomalies summarised across all EAM forests, and individually by mountain bloc. Forecasts are according to IPCC-AR4 emissions scenarios B1 (white) and A1B (grey), downscaled from a regional climate model (55 km, 0.5°, ECHAM5 boundary data) using WorldClim (temperature) and TRMM (rainfall) surfaces as present-day baselines (1 km, 0.00833°). Whiskers extend up to 150% of the interquartile range from each box. Filled and open circles show means ± standard deviations, respectively. All climates are significantly different to the present day ($p < 0.05$; Wilcoxon Signed-Rank), unless denoted "ns" in the plots (e.g., temperature seasonality, B1 2090).

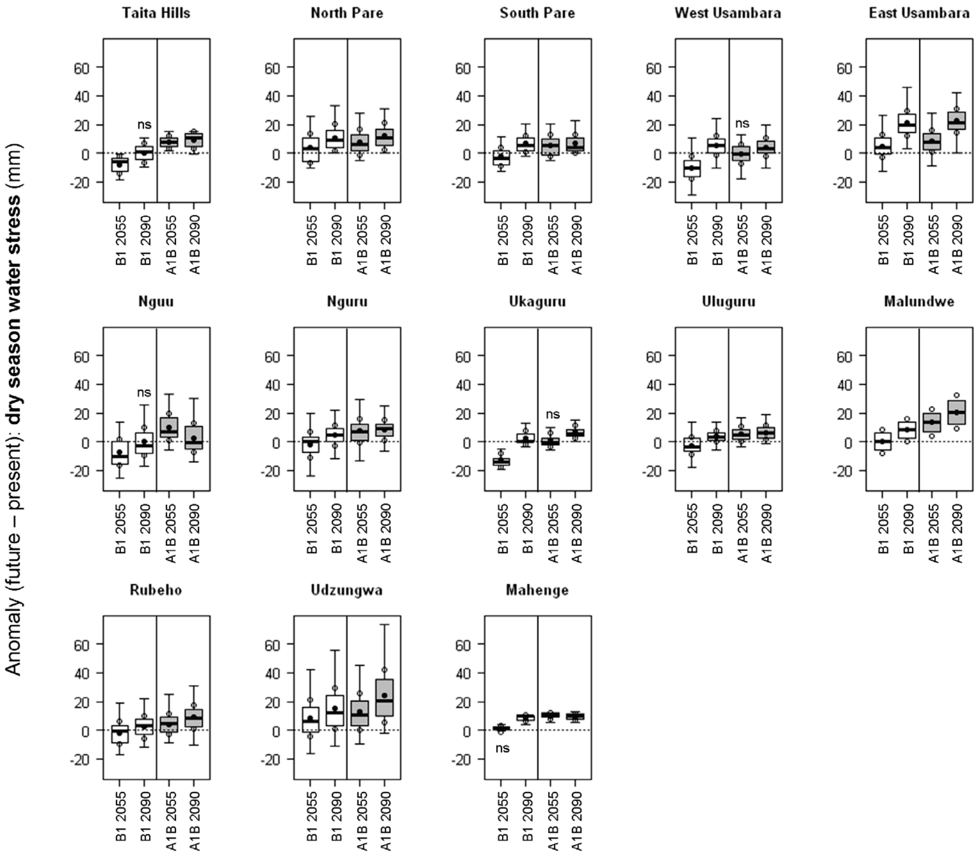



Anomaly (future – present): **mean annual temperature** (°C)
Taita Hills
North Pare
South Pare
West Usambara
East Usambara
Nguu
Nguru
Ukaguru
Uluguru
Malundwe
Rubeho
Udzungwa
Mahenge
0
1
2
3
4
B1 2055
B1 2090
A1B 2055
A1B 2090


Anomaly (future – present): **temperature seasonality** (°C)
Taita Hills
North Pare
South Pare
West Usambara
East Usambara
Nguu
Nguru
Ukaguru
ns
Uluguru
Malundwe
Rubeho
Udzungwa
Mahenge
-2
-1
0
1
2
B1 2055
B1 2090
A1B 2055
A1B 2090

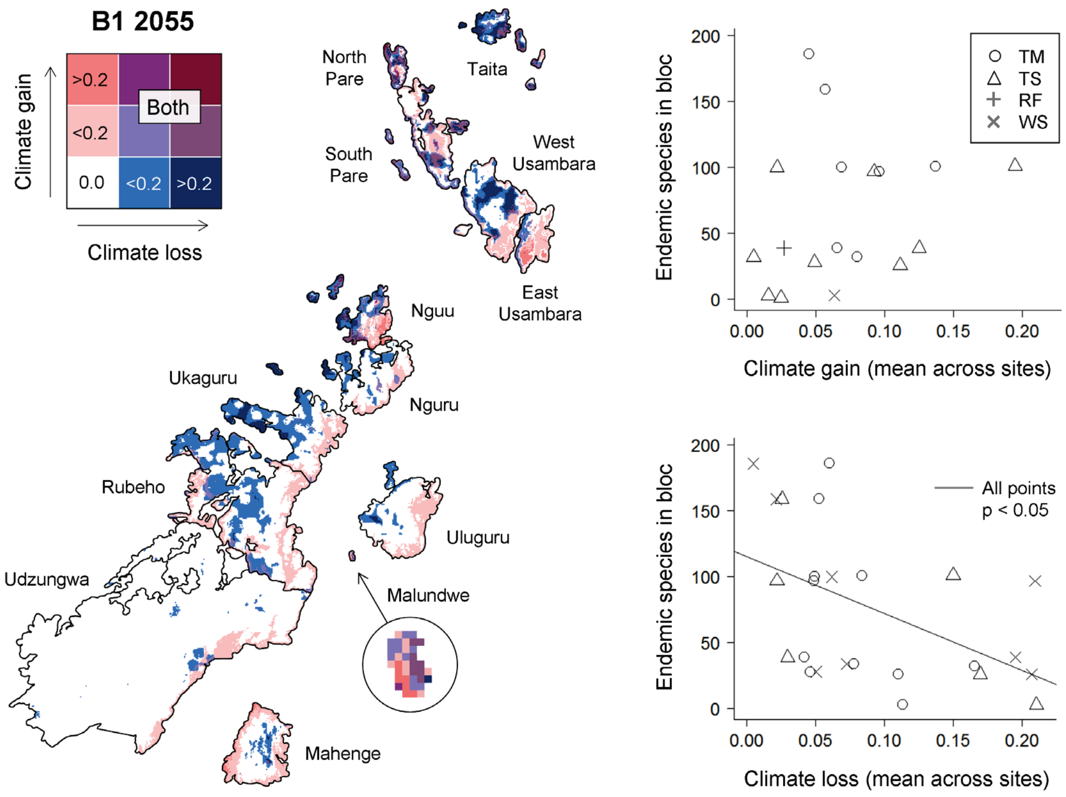

Anomaly (future – present): annual rainfall (mm)
Taita Hills
North Pare
South Pare
West Usambara
East Usambara
Nguu
Nguru
Ukaguru
Uluguru
Malundwe
Rubeho
Udzungwa
Mahenge
200
0
-200
-400
B1 2055
B1 2090
A1B 2055
A1B 2090
ns

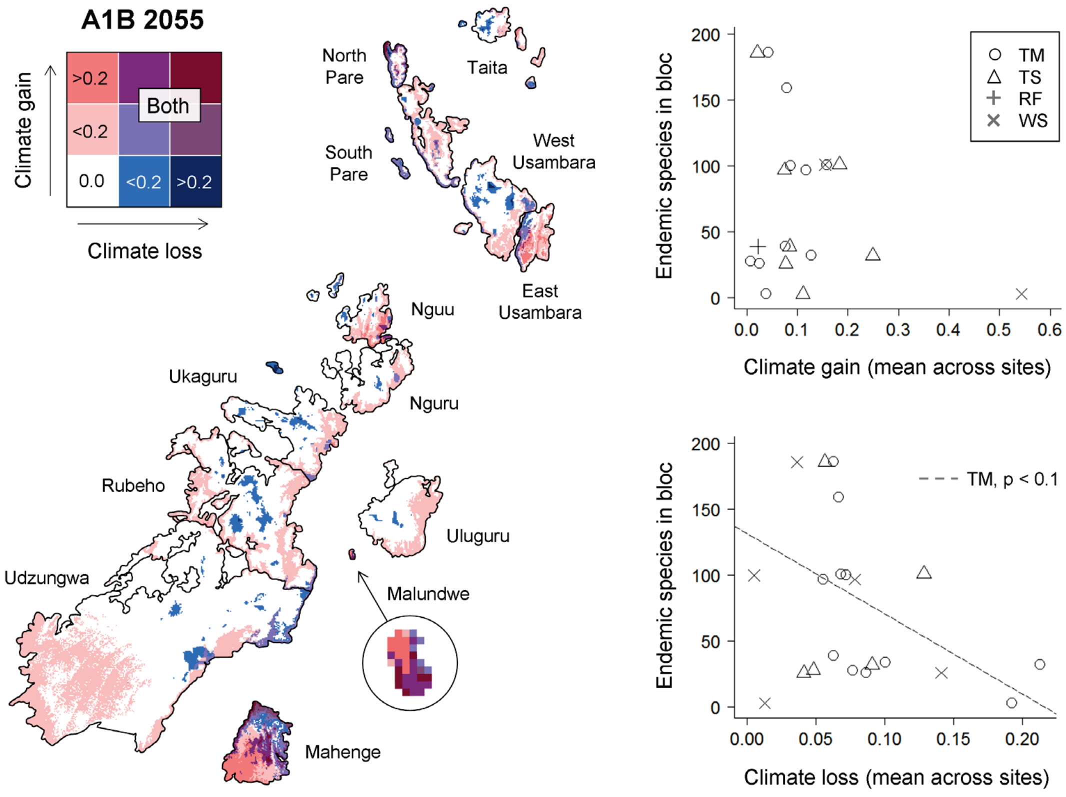

Anomaly (future – present): moisture index (rainfall / PET)
Taita Hills
North Pare
South Pare
West Usambara
East Usambara
Nguu
Nguru
Ukaguru
Uluguru
Malundwe
Rubeho
Udzungwa
Mahenge
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
B1 2055
B1 2090
A1B 2055
A1B 2090
ns

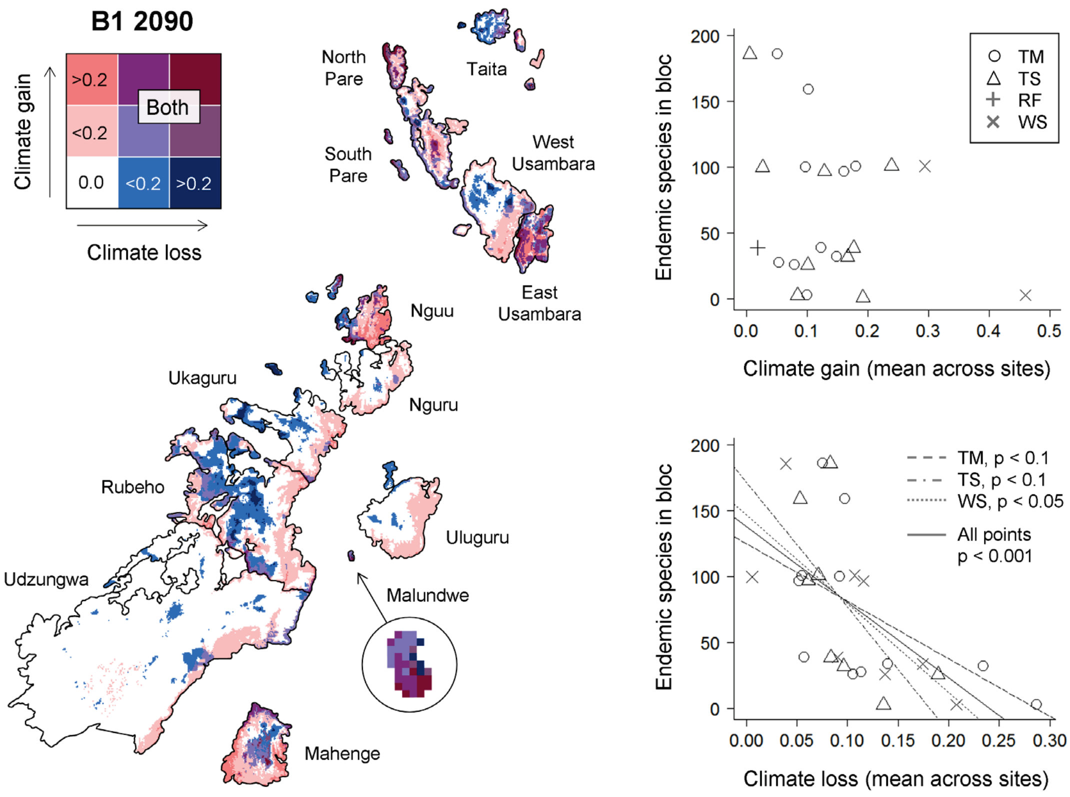

Anomaly (future – present): dry season water stress (mm)
Taita Hills
North Pare
South Pare
West Usambara
East Usambara
Nguu
Nguru
Ukaguru
Uluguru
Malundwe
Rubeho
Udzungwa
Mahenge
60
40
20
0
-20
ns
B1 2055
B1 2090
A1B 2055
A1B 2090

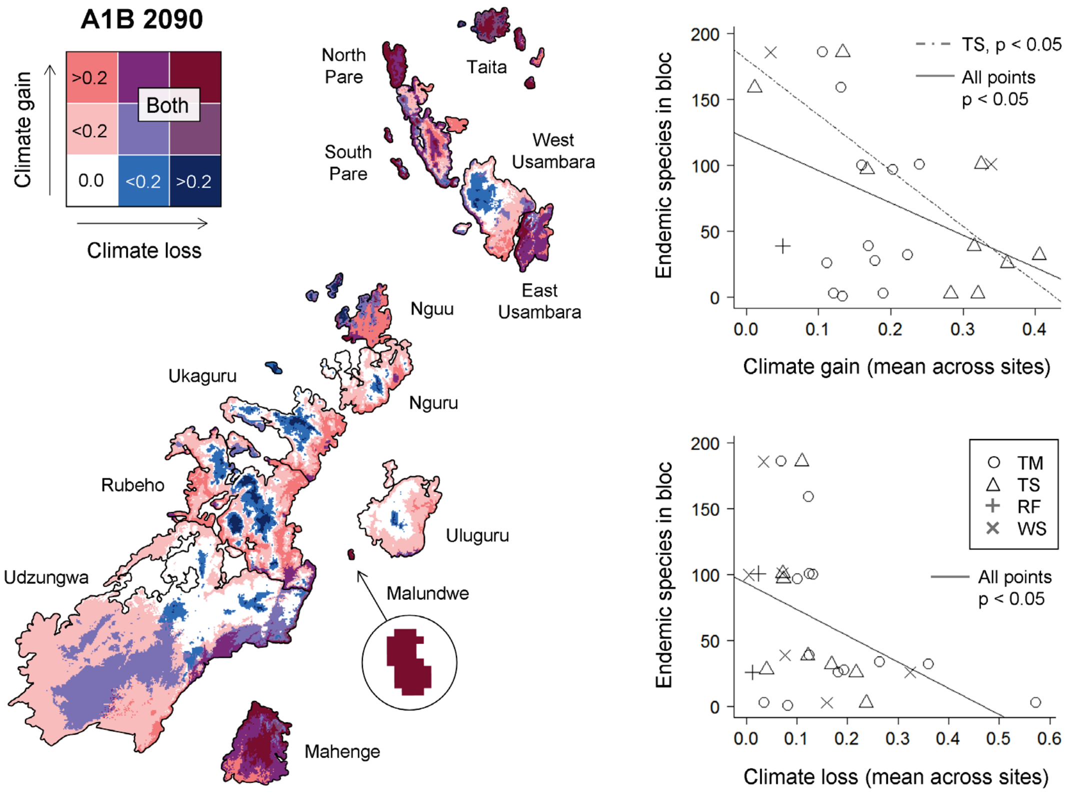
Appendix 5C. Climates lost and gained

Climatic conditions lost and gained in the EAMs, relative to present-day conditions, by the years 2055 and 2090. Due to the dispersal limitations of narrow-ranged mountain plants, the search for climate analogues is restricted to individual mountain blocs. Climate lost is the proportional distance of present-day conditions beyond future extremes with respect to, and summed over: mean annual temperature (TM), temperature seasonality (TS), mean annual rainfall (RF) and dry season water stress (WS). Climate gain sums the proportional distance of *future* conditions beyond *present-day* extremes. Scatter plots compare no-zero loss and gain (mean across known endemism sites) with endemic plant richness. Regression lines and corresponding p-values (F-tests) show significant negative relationships between the degree of climate loss and endemic richness for both 2055 and 2090 and under both emissions scenarios. Climate gain, less directly associated with species persistence (because familiar conditions may be accessible in neighbouring cells), is significant only for A1B 2090.


B1 2055
Climate gain
>0.2
<0.2
0.0
Both
<0.2
>0.2
Climate loss
North Pare
Taita
South Pare
West Usambara
East Usambara
Nguu
Ukaguru
Nguru
Rubeho
Uluguru
Udzungwa
Malundwe
Mahenge
Endemic species in bloc
200
150
100
50
0
0.00
0.05
0.10
0.15
0.20
Climate gain (mean across sites)
TM
TS
RF
WS
Climate loss (mean across sites)
All points
p < 0.05


A1B 2055
Climate gain
>0.2
<0.2
0.0
Both
<0.2
>0.2
Climate loss
North Pare
Taita
South Pare
West Usambara
East Usambara
Nguu
Ukaguru
Nguru
Rubeho
Uluguru
Udzungwa
Malundwe
Mahenge
Endemic species in bloc
200
150
100
50
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Climate gain (mean across sites)
TM
TS
RF
WS
0.00
0.05
0.10
0.15
0.20
Climate loss (mean across sites)
TM, p < 0.1


B1 2090
Climate gain
>0.2
<0.2
0.0
Both
<0.2
>0.2
Climate loss
North Pare
Taita
South Pare
West Usambara
East Usambara
Nguu
Nguru
Ukaguru
Rubeho
Uluguru
Udzungwa
Malundwe
Mahenge
Endemic species in bloc
200
150
100
50
0
0.0
0.1
0.2
0.3
0.4
0.5
Climate gain (mean across sites)
TM
TS
RF
WS
200
150
100
50
0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Climate loss (mean across sites)
TM, p < 0.1
TS, p < 0.1
WS, p < 0.05
All points
p < 0.001


A1B 2090
Climate gain
>0.2
<0.2
0.0
Both
<0.2
>0.2
Climate loss
North Pare
Taita
South Pare
West Usambara
East Usambara
Nguu
Nguru
Ukaguru
Rubeho
Uluguru
Udzungwa
Malundwe
Mahenge
Endemic species in bloc
200
150
100
50
0
0.0
0.1
0.2
0.3
0.4
Climate gain (mean across sites)
TS, p < 0.05
All points
p < 0.05
200
150
100
50
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Climate loss (mean across sites)
TM
TS
RF
WS
All points
p < 0.05

Appendix 5D. Multivariate species models

Of the 56 endemic plant species eligible for multivariate regression (generalised additive modelling with ten or more presence records), 34 produced models with an average cross-validated (five-fold) AUC of at least 0.7. This level of model accuracy, balancing type I and type II error rates on unseen data, was considered sufficiently robust to provide a good indication of the relative importance of climatic gradients in explaining observed distributions. It is worth noting that background (pseudo-absence) data outnumbered presence data (5:1) and were biased towards sites close to known occurrences, providing particularly constrained models. That is, commission error may be lower than the AUC suggests. For these 34 species, we summarise the relative importance of model predictors and directions of change, with respect to range size and altitude, under future climate scenarios.

Relative important of predictor variables

Selection rate was calculated over 20 model runs: forward-backward and backward-forward stepwise procedures for ten independent realisations of background data. Contribution is the average drop in explained deviance when a variable was removed from the final model.

		Selection rate (%)				Contribution (%)			
Species	Growth form	TM	TS	RF	WS	TM	TS	RF	WS
Allanblackia ulugurensis	Tree	70	5	90	100	5	5	10	12
Allophylus melliodorus	Tree	100	5	0	0	22	15	-	-
Alsodeiopsis schumannii	Tree	85	70	10	100	5	5	0	7
Casearia engleri	Tree	100	65	70	30	18	14	17	3
Chassalia albiflora	Shrub	5	90	25	100	2	10	3	8
Chassalia zimmermannii	Shrub	45	95	15	80	11	17	4	16
Cola stelechantha	Tree	25	90	80	10	4	13	15	17
Craterispermum longipedunculatum	Tree	90	25	20	75	17	11	2	10
Danais xanthorrhoea	Herb	100	10	95	20	16	9	6	9
Diplazium pseudoporrectum	Herb	90	5	45	100	14	8	11	32
Dissotis polyantha	Shrub	40	80	10	100	3	8	1	19
Eugenia toxanatolica	Tree	55	15	100	10	9	6	10	0
Gravesia pulchra	Shrub	45	30	60	85	10	11	7	6
Isolona linearis	Tree	20	100	10	90	8	26	2	11
Lasianthus cereiflorus	Tree	90	60	75	100	23	12	14	32
Lasianthus glomeruliflorus	Shrub	25	15	90	85	13	5	18	6
Lasianthus macrocalyx	Shrub	100	0	0	0	23	-	-	-
Lasianthus pedunculatus	Tree	80	45	40	75	9	9	5	5
Lobelia longisepala	Herb	30	100	85	10	9	14	9	14
Memecylon cogniauxii	Shrub	50	25	45	95	3	5	4	9
Pauridiantha coalescens	Shrub	25	95	75	25	9	20	12	1
Pavetta amaniensis	Shrub	40	65	70	40	12	8	7	9
Pavetta holstii	Shrub	20	85	30	5	9	18	4	14
Plectranthus triangularis	Herb	20	95	95	55	1	9	8	6
Polysphaeria macrantha	Tree	35	0	95	15	9	-	34	15
Psychotria brucei	Tree	100	50	65	15	17	4	8	8
Psychotria megalopus	Tree	95	45	85	90	11	3	7	6
Psychotria pandurata	Herb	45	25	20	95	9	8	4	12
Psychotria porphyroclada	Shrub	60	100	0	10	11	21	-	14
Psychotria triclada	Shrub	10	100	5	5	5	20	5	2
Sorindeia calantha	Tree	60	15	5	100	9	5	8	9
Tarenna roseicosta	Shrub	5	100	5	10	6	17	0	0
Tricalysia aciculiflora	Shrub	10	100	80	5	3	24	8	9
Vangueria fuscosetulosa	Shrub	35	80	80	20	22	37	15	5
All 34 species	Mean	53	55	49	52	13	15	11	12
	s.d.	± 32	± 37	± 36	± 40	± 8	± 9	± 10	± 9

TM, mean annual temperature; TS, temperature seasonality
RF, annual rainfall (moisture index [RF/PET] was preferred in 46% of model runs)
WS, dry season water stress

Direction of change in climatic suitability

Increase (↑) or decrease (↓) in modelled climatic suitability across mountain blocs within each species' range. Emission scenarios B1 and A1B. Pairs of arrows correspond to the years 2055 (left) and 2090 (right).

	Taita Hills	North Pare	South Pare	West Usambara	East Usambara	Nguu	Nguru	Ukaguru	Uluguru	Malundwe	Rubeho	Udzungwa	Mahenge	All EAM
Allanblackia ulugurensis														
B1	-	-	-	-	-	-	↑↓	-	↓↓	-	-	↓↓	-	↓↓
A1B	-	-	-	-	-	-	↓↓	-	↓↓	-	-	↓↓	-	↓↓
Allophylus melliodorus														
B1	-	-	-	↑↑	↓↓	-	↓↓	-	-	-	-	-	-	↓↓
A1B	-	-	-	↑↑	↓↓	-	↓↓	-	-	-	-	-	-	↓↓
Alsodeiopsis schumannii														
B1	-	-	-	↑↓	↓↓	-	↑↓	-	↓↓	-	-	↓↓	-	↓↓
A1B	-	-	-	↑↓	↓↓	-	↓↓	-	↓↓	-	-	↓↓	-	↓↓
Casearia engleri														
B1	-	-	↓↓	↓↓	-	-	-	-	-	-	-	-	-	↓↓
A1B	-	-	↓↓	↓↓	-	-	-	-	-	-	-	-	-	↓↓
Chassalia albiflora														
B1	-	-	↑↓	↑↓	↑↑	-	↑↑	-	-	-	-	-	-	↑↑
A1B	-	-	↑↑	↑↑	↑↑	-	↓↓	-	-	-	-	-	-	↑↑
Chassalia zimmermannii														
B1	-	-	-	↑↓	↓↓	-	-	-	↑↓	-	-	-	-	↑↓
A1B	-	-	-	↑↓	↓↓	-	-	-	↓↓	-	-	-	-	↓↓
Cola stelechantha														
B1	-	-	-	-	-	-	↑↑	-	-	-	-	↓↓	-	↓↓
A1B	-	-	-	-	-	-	↑↑	-	-	-	-	↓↓	-	↓↓
Craterispermum longipedunculatum														
B1	-	-	-	-	-	-	↓↓	-	↓↓	-	-	↓↓	↓↓	↓↓
A1B	-	-	-	-	-	-	↓↓	-	↓↓	-	-	↓↓	↓↓	↓↓
Danais xanthorrhoea														
B1	-	-	↓↓	↑↑	↓↓	-	↓↓	-	↓↓	-	-	↓↓	-	↓↓
A1B	-	-	↓↓	↑↑	↓↓	-	↓↓	-	↓↓	-	-	↓↓	-	↓↓
Diplazium pseudoporrectum														
B1	-	-	-	-	-	-	↓↓	↑↓	↓↓	-	-	↓↓	-	↓↓
A1B	-	-	-	-	-	-	↓↓	↓↓	↓↓	-	-	↓↓	-	↓↓
Dissotis polyantha														
B1	-	-	-	-	-	-	↓↓	↓↑	↓↑	-	↑↑	↓↓	-	↓↓
A1B	-	-	-	-	-	-	↑↓	↑↓	↓↓	-	↓↓	↓↓	-	↓↓
Eugenia toxanatolica														
B1	-	-	↑↓	↑↓	-	-	-	-	-	-	↓↓	↓↓	↓↓	↓↓
A1B	-	-	↓↓	↑↑	-	-	-	-	-	-	↓↓	↓↓	↓↓	↓↓

[climatic suitability continued]	Taita Hills	North Pare	South Pare	West Usambara	East Usambara	Nguu	Nguru	Ukaguru	Uluguru	Malundwe	Rubeho	Udzungwa	Mahenge	All EAM
Gravesia pulchra														
B1	-	-	-	-	-	-	↑↓	-	↓↓	-	↓↓	↓↓	↓↓	↓↓
A1B	-	-	-	-	-	-	↓↓	-	↓↓	-	↓↓	↓↓	↓↓	↓↓
Isolona linearis														
B1	-	-	-	↑↓	-	-	-	-	-	-	-	↓↓	-	↓↓
A1B	-	-	-	↓↓	-	-	-	-	-	-	-	↓↓	-	↓↓
Lasianthus cereiflorus														
B1	-	-	-	-	-	-	↓↓	-	↓↓	-	-	-	-	↓↓
A1B	-	-	-	-	-	-	↓↓	-	↓↓	-	-	-	-	↓↓
Lasianthus glomeruliflorus														
B1	-	-	-	-	-	-	↑↓	-	↓↓	-	-	-	-	↓↓
A1B	-	-	-	-	-	-	↓↓	-	↓↓	-	-	-	-	↓↓
Lasianthus macrocalyx														
B1	-	-	-	-	-	-	-	-	↓↓	-	-	-	-	↓↓
A1B	-	-	-	-	-	-	-	-	↓↓	-	-	-	-	↓↓
Lasianthus pedunculatus														
B1	-	-	-	-	-	-	↓↓	↓↓	↓↓	-	↓↓	↓↓	-	↓↓
A1B	-	-	-	-	-	-	↓↓	↓↓	↓↓	-	↓↓	↓↓	-	↓↓
Lobelia longisepala														
B1	-	-	-	-	↑↑	-	↑↑	-	↑↑	-	-	↑↓	↓↓	↑↓
A1B	-	-	-	-	↑↑	-	↓↓	-	↓↓	-	-	↓↓	↓↓	↓↓
Memecylon cogniauxii														
B1	-	-	↑↓	↑↓	↓↓	-	↑↓	-	↑↓	-	-	↓↓	-	↓↓
A1B	-	-	↓↓	↓↓	↓↓	-	↓↓	-	↓↓	-	-	↓↓	-	↓↓
Pauridiantha coalescens														
B1	-	-	-	-	-	-	-	-	-	-	-	↓↓	-	↓↓
A1B	-	-	-	-	-	-	-	-	-	-	-	↓↓	-	↓↓
Pavetta amaniensis														
B1	-	-	-	↑↑	↑↑	-	-	-	-	-	-	-	-	↑↑
A1B	-	-	-	↑↑	↑↑	-	-	-	-	-	-	-	-	↑↑
Pavetta holstii														
B1	-	-	↓↓	↑↑	↑↑	-	↓↓	-	↑↑	-	-	-	-	↑↑
A1B	-	-	↓↓	↑↑	↑↑	-	↓↓	-	↓↓	-	-	-	-	↑↑
Plectranthus triangularis														
B1	↑↑	-	↑↑	↑↑	-	-	↑↑	-	↑↑	-	-	-	-	↑↑
A1B	↑↑	-	↑↑	↑↑	-	-	↑↑	-	↓↓	-	-	-	-	↑↑
Polysphaeria macrantha														
B1	-	-	↑↑	-	↑↓	-	-	-	-	-	-	↓↓	-	↓↓
A1B	-	-	↑↑	-	↑↑	-	-	-	-	-	-	↓↓	-	↓↓
Psychotria brucei														
B1	-	-	↓↓	-	-	-	-	-	↓↓	-	-	↓↓	-	↓↓
A1B	-	-	↓↓	-	-	-	-	-	↓↓	-	-	↓↓	-	↓↓
Psychotria megalopus														
B1	-	-	-	-	-	-	-	-	↓↓	-	-	↓↓	↓↓	↓↓
A1B	-	-	-	-	-	-	-	-	↓↓	-	-	↓↓	↓↓	↓↓

[climatic suitability continued]	Taita Hills	North Pare	South Pare	West Usambara	East Usambara	Nguu	Nguru	Ukaguru	Uluguru	Malundwe	Rubeho	Udzungwa	Mahenge	All EAM
Psychotria pandurata														
B1	-	-	-	-	↓↓	-	-	-	↑↓	-	-	↓↓	-	↓↓
A1B	-	-	-	-	↓↓	-	-	-	↓↓	-	-	↓↓	-	↓↓
Psychotria porphyroclada														
B1	-	-	-	↑↑	-	-	-	-	-	-	-	-	-	↑↑
A1B	-	-	-	↑↑	-	-	-	-	-	-	-	-	-	↑↑
Psychotria triclada														
B1	-	-	-	↑↑	↑↑	-	↑↑	-	-	-	-	-	-	↑↑
A1B	-	-	-	↑↑	↑↓	-	↑↑	-	-	-	-	-	-	↑↑
Sorindeia calantha														
B1	↑↑	-	↑↓	-	-	-	↑↑	-	-	-	-	↓↓	-	↑↓
A1B	↓↓	-	↓↓	-	-	-	↓↓	-	-	-	-	↓↓	-	↓↓
Tarenna roseicosta														
B1	-	-	-	-	-	-	-	-	-	-	↑↑	↓↓	-	↓↓
A1B	-	-	-	-	-	-	-	-	-	-	↓↓	↓↓	-	↓↓
Tricalysia aciculiflora														
B1	-	-	-	-	-	-	↑↑	-	-	-	-	↓↓	-	↓↓
A1B	-	-	-	-	-	-	↑↑	-	-	-	-	↓↓	-	↓↓
Vangueria fuscosetulosa														
B1	-	-	-	-	-	-	-	-	-	-	-	↓↓	-	↓↓
A1B	-	-	-	-	-	-	-	-	-	-	-	↓↓	-	↓↓

Direction of change in altitudinal optima

Increase (↑), decrease (↓), or stasis (o) in the altitudinal optima of species' ranges under future compared with present-day climatic conditions, calculated by summing climatic suitability within 100 m altitudinal bands. Emission scenarios B1 and A1B. Pairs of arrows correspond to the years 2055 (left) and 2090 (right).

	Taita Hills	North Pare	South Pare	West Usambara	East Usambara	Nguu	Nguru	Ukaguru	Uluguru	Malundwe	Rubeho	Udzungwa	Mahenge	All EAM
Allanblackia ulugurensis														
B1	-	-	-	-	-	-	oo	-	oo	-	-	↑↑	-	↓↓
A1B	-	-	-	-	-	-	oo	-	oo	-	-	↑↑	-	↓↓
Allophylus melliodorus														
B1	-	-	-	↑↑	↑↑	-	o↑	-	-	-	-	-	-	↑↑
A1B	-	-	-	↑↑	↑↑	-	↑↑	-	-	-	-	-	-	↑↑
Alsodeiopsis schumannii														
B1	-	-	-	↑↑	↓↓	-	↑↑	-	oo	-	-	↑↑	-	↑↑
A1B	-	-	-	↑↑	↓↓	-	↑↑	-	oo	-	-	↑↑	-	↑↑
Casearia engleri														
B1	-	-	oo	oo	-	-	-	-	-	-	-	-	-	oo
A1B	-	-	oo	oo	-	-	-	-	-	-	-	-	-	o↑
Chassalia albiflora														
B1	-	-	↓↓	↑o	o↓	-	o↑	-	-	-	-	-	-	oo
A1B	-	-	↓↓	↑↑	o↓	-	↑↑	-	-	-	-	-	-	o↓
Chassalia zimmermannii														
B1	-	-	-	oo	oo	-	-	-	oo	-	-	-	-	oo
A1B	-	-	-	oo	o↑	-	-	-	oo	-	-	-	-	oo
Cola stelechantha														
B1	-	-	-	-	-	-	oo	-	-	-	-	o↓	-	↓↓
A1B	-	-	-	-	-	-	oo	-	-	-	-	↓↓	-	↓↓
Craterispermum longipedunculatum														
B1	-	-	-	-	-	-	o↑	-	↑↑	-	-	↑↑	oo	↑↑
A1B	-	-	-	-	-	-	o↑	-	↑↑	-	-	↑↑	oo	↑↑
Danais xanthorrhoea														
B1	-	-	↑↑	o↑	o↑	-	↑↑	-	↑↑	-	-	↑↑	-	↑↑
A1B	-	-	↑↑	o↑	o↑	-	↑↑	-	↑↑	-	-	↑↑	-	↑↑
Diplazium pseudoporrectum														
B1	-	-	-	-	-	-	↑↑	↑↓	↑↑	-	-	↑↑	-	↑↑
A1B	-	-	-	-	-	-	↑↑	↓↓	↑↑	-	-	↑↑	-	↑↑
Dissotis polyantha														
B1	-	-	-	-	-	-	↓↓	↓↓	oo	-	↑o	↓o	-	↓o
A1B	-	-	-	-	-	-	oo	↓o	oo	-	oo	o↓	-	↓↓
Eugenia toxanatolica														
B1	-	-	↓↓	oo	-	-	-	-	-	-	oo	oo	↑↑	oo
A1B	-	-	↓↓	oo	-	-	-	-	-	-	oo	oo	↑↑	oo

[altitudinal optima continued]	Taita Hills	North Pare	South Pare	West Usambara	East Usambara	Nguu	Nguru	Ukaguru	Uluguru	Malundwe	Rubeho	Udzungwa	Mahenge	All EAM
Gravesia pulchra														
B1	-	-	-	-	-	-	↓↓	-	oo	-	↑↑	o↑	o↓	↑↑
A1B	-	-	-	-	-	-	↓↓	-	oo	-	oo	↑↑	o↓	↑↑
Isolona linearis														
B1	-	-	-	oo	-	-	-	-	-	-	-	o↓	-	o↓
A1B	-	-	-	oo	-	-	-	-	-	-	-	↓↓	-	↓↓
Lasianthus cereiflorus														
B1	-	-	-	-	-	-	↑↑	-	↑↑	-	-	-	-	↑↑
A1B	-	-	-	-	-	-	↑↑	-	↑↑	-	-	-	-	↑↑
Lasianthus glomeruliflorus														
B1	-	-	-	-	-	-	↑↑	-	oo	-	-	-	-	oo
A1B	-	-	-	-	-	-	↑↑	-	o↑	-	-	-	-	o↑
Lasianthus macrocalyx														
B1	-	-	-	-	-	-	-	-	↑↑	-	-	-	-	↑↑
A1B	-	-	-	-	-	-	-	-	↑↑	-	-	-	-	↑↑
Lasianthus pedunculatus														
B1	-	-	-	-	-	-	oo	oo	↑↑	-	oo	oo	-	oo
A1B	-	-	-	-	-	-	oo	oo	o↑	-	oo	oo	-	oo
Lobelia longisepala														
B1	-	-	-	-	↓↓	-	oo	-	oo	-	-	oo	oo	oo
A1B	-	-	-	-	↓↓	-	oo	-	oo	-	-	oo	oo	oo
Memecylon cogniauxii														
B1	-	-	↓↓	↑↑	↓↓	-	↓↓	-	oo	-	-	↓↑	-	↑o
A1B	-	-	o↓	o↑	↓↓	-	↓↓	-	oo	-	-	↑↑	-	oo
Pauridiantha coalescens														
B1	-	-	-	-	-	-	-	-	-	-	-	oo	-	oo
A1B	-	-	-	-	-	-	-	-	-	-	-	↓↓	-	↓↓
Pavetta amaniensis														
B1	-	-	-	↑↑	↓↓	-	-	-	-	-	-	-	-	oo
A1B	-	-	-	↑↑	↓↓	-	-	-	-	-	-	-	-	oo
Pavetta holstii														
B1	-	-	oo	↑↑	oo	-	oo	-	oo	-	-	-	-	o↓
A1B	-	-	o↓	o↑	o↓	-	o↓	-	oo	-	-	-	-	oo
Plectranthus triangularis														
B1	oo	-	↓↓	oo	-	-	oo	-	oo	-	-	-	-	oo
A1B	oo	-	↓↓	oo	-	-	oo	-	oo	-	-	-	-	o↑
Polysphaeria macrantha														
B1	-	-	oo	-	oo	-	-	-	-	-	-	oo	-	oo
A1B	-	-	oo	-	oo	-	-	-	-	-	-	oo	-	oo
Psychotria brucei														
B1	-	-	oo	-	-	-	-	-	↑↑	-	-	oo	-	oo
A1B	-	-	oo	-	-	-	-	-	↑↑	-	-	oo	-	oo
Psychotria megalopus														
B1	-	-	-	-	-	-	-	-	↑↑	-	-	↑↑	oo	↑↑
A1B	-	-	-	-	-	-	-	-	↑↑	-	-	↑↑	oo	↑↑

[altitudinal optima continued]	Taita Hills	North Pare	South Pare	West Usambara	East Usambara	Nguu	Nguru	Ukaguru	Uluguru	Malundwe	Rubeho	Udzungwa	Mahenge	All EAM
Psychotria pandurata														
B1	-	-	-	-	oo	-	-	-	oo	-	-	↑↑	-	oo
A1B	-	-	-	-	oo	-	-	-	oo	-	-	↑↑	-	oo
Psychotria porphyroclada														
B1	-	-	-	↓↓	-	-	-	-	-	-	-	-	-	↓↓
A1B	-	-	-	↓↓	-	-	-	-	-	-	-	-	-	↓↓
Psychotria triclada														
B1	-	-	-	↑↑	oo	-	oo	-	-	-	-	-	-	↑↑
A1B	-	-	-	o↑	o↑	-	o↑	-	-	-	-	-	-	o↑
Sorindeia calantha														
B1	↓↓	-	oo	-	-	-	↑↓	-	-	-	-	↑↑	-	o↑
A1B	oo	-	oo	-	-	-	↓↓	-	-	-	-	↑↑	-	↑↑
Tarenna roseicosta														
B1	-	-	-	-	-	-	-	-	-	-	↑o	↓↓	-	↓↓
A1B	-	-	-	-	-	-	-	-	-	-	oo	↓↓	-	↓↓
Tricalysia aciculiflora														
B1	-	-	-	-	-	-	↓↓	-	-	-	-	↑↓	-	↑↓
A1B	-	-	-	-	-	-	↓↓	-	-	-	-	↓↓	-	↓↓
Vangueria fuscosetulosa														
B1	-	-	-	-	-	-	-	-	-	-	-	↓↓	-	↓↓
A1B	-	-	-	-	-	-	-	-	-	-	-	↓↓	-	↓↓

References

Ahrends, A., Rahbek, C., Bulling, M. T., Burgess, N. D., Platts, P. J., Lovett, J. C., Kindemba, V. W., Owen, N., Sallu, A. N., Marshall, A. R., Mhoro, B. E., Fanning, E. and Marchant, R. (2011) Conservation and the botanist effect. *Biological Conservation*, **144**, 131-140.

Austin, M. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1-19.

Austin, M. P. and Van Niel, K. P. (2011) Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, **38**, 1-8.

Beaumont, L. J., Pitman, A., Perkins, S., Zimmermann, N. E., Yoccoz, N. G. and Thuiller, W. (2011) Impacts of climate change on the world's most exceptional ecoregions. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 2306-2311.

Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Rylands, A. B., Konstant, W. R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G. and Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909-923.

Cadena, C. D., Kozak, K. H., Gomez, J. P., Parra, J. L., McCain, C. M., Bowie, R. C. K., Carnaval, A. C., Moritz, C., Rahbek, C., Roberts, T. E., Sanders, N. J., Schneider, C. J., VanDerWal, J., Zamudio, K. R. and Graham, C. H. (2012) Latitude, elevational climatic zonation and speciation in New World vertebrates. *Proceedings of the Royal Society B-Biological Sciences*, **279**, 194-201.

Colwell, R. K., Brehm, G., Cardelus, C. L., Gilman, A. C. and Longino, J. T. (2008) Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science*, **322**, 258-261.

Cordeiro, N. J. and Howe, H. F. (2003) Forest fragmentation severs mutualism between seed dispersers and an endemic African tree. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 14052-14056.

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

Crimmins, S. M., Dobrowski, S. Z., Greenberg, J. A., Abatzoglou, J. T. and Mynsberge, A. R. (2011) Changes in Climatic Water Balance Drive Downhill Shifts in Plant Species' Optimum Elevations. *Science*, **331**, 324-327.

Dansgaard, W., White, J. W. C. and Johnsen, S. J. (1989) The abrupt termination of the Younger Dryas climate event. *Nature*, **339**, 532-534.

Dinesen, L., Lehmberg, T., Svendsen, J. O., Hansen, L. A. and Fjeldsa, J. (1994) A new genus and species of perdicine bird (*Phasianidae*, *Perdicini*) from Tanzania - a relict form with Indio-Malayan affinities. *Ibis*, **136**, 3-11.

Dobrowski, S. Z., Abatzoglou, J. T., Greenberg, J. A. and Schladow, S. G. (2009) How much influence does landscape-scale physiography have on air temperature in a mountain environment? *Agricultural and Forest Meteorology*, **149**, 1751-1758.

Doherty, R. M., Sitch, S., Smith, B., Lewis, S. L. and Thornton, P. K. (2010) Implications of future climate and atmospheric CO2 content for regional biogeochemistry, biogeography and ecosystem services across East Africa. *Global Change Biology*, **16**, 617-640.

Fairman, J. G., Nair, U. S., Christopher, S. A. and Molg, T. (2011) Land use change impacts on regional climate over Kilimanjaro. *Journal of Geophysical Research-Atmospheres*, **116**, 24.

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

Finch, J. and Marchant, R. (2011) A palaeoecological investigation into the role of fire and human activity in the development of montane grasslands in East Africa. *Vegetation History and Archaeobotany*, **20**, 109-124.

Fjeldså, J., Ehrlich, D., Lambin, E. and 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. and Lovett, J. C. (1997) Biodiversity and environmental stability. *Biodiversity and Conservation*, **6**, 315-323.

Fjeldsa, J., Kiure, J., Doggart, N., Hansen, L. A. and Perkin, A. (2010) Distribution of highland forest birds across a potential dispersal barrier in the Eastern Arc Mountains of Tanzania. *Steenstrupia*, **32**, 1-43.

Forero-Medina, G., Terborgh, J., Socolar, S. J. and Pimm, S. L. (2011) Elevational ranges of birds on a tropical montane gradient lag behind warming temperatures. *PloS one*, **6**, e28535.

Gereau, R. E., Hall, J., Hemp, A. and Platts, P. J. (in prep.) Distribution and endemism of vascular plants in the Eastern Arc Mountains.

Gereau, R. E., Taylor , C. M., Bodine, S. and Kindeketa, W. J. (2010) Plant Conservation Assessment in the Eastern Arc Mountains and Coastal Forests of Tanzania and Kenya (http://www.mobot.org/MOBOT/Research/tanzania/cepf.shtml/).

Ghalambor, C. K., Huey, R. B., Martin, P. R., Tewksbury, J. J. and Wang, G. (2006) Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integrative and Comparative Biology*, **46**, 5-17.

Greve, M., Lykke, A. M., Blach-Overgaard, A. and Svenning, J.-C. (2011) Environmental and anthropogenic determinants of vegetation distribution across Africa. *Global Ecology and Biogeography*, **20**, 661-674.

Hall, J., Burgess, N. D., Lovett, J., Mbilinyi, B. and Gereau, R. E. (2009) Conservation implications of deforestation across an elevational gradient in the Eastern Arc Mountains, Tanzania. *Biological Conservation*, **142**, 2510-2521.

Hamilton, A. C. and Bensted-Smith, R. (1989) *Forest Conservation in the East Usambara Mountains, Tanzania*, International Union for the Conservation of Nature, Gland and Cambridge, UK.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

IPCC (2007) *Climate change 2007: Synthesis report*, Intergovernmental Panel on Climate Change.

IUCN (2011) IUCN Red List of Threatened Species. Version 2011.2 (http://www.iucnredlist.org/).

Janzen, D. H. (1967) Why mountain passes are higher in the tropics. *The American Naturalist*, **101**, 233-249.

Jarvis, A., Reuter, H. I., Nelson, A. and Guevara, E. (2008) Hole-filled seamless SRTM data. Version 4. International Centre for Tropical Agriculture, CIAT (http://srtm.csi.cgiar.org/).

Jetz, W., Rahbek, C. and 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.

Jump, A. S., Huang, T.-J. and Chou, C.-H. (2012) Rapid altitudinal migration of mountain plants in Taiwan and its implications for high altitude biodiversity. *Ecography*, **35**, 204-210.

Kahindo, C., Bowie, R. C. K. and Bates, J. M. (2007) The relevance of data on genetic diversity for the conservation of Afro-montane regions. *Biological Conservation*, **134**, 262-270.

Knapp, A. K., Beier, C., Briske, D. D., Classen, A. T., Luo, Y., Reichstein, M., Smith, M. D., Smith, S. D., Bell, J. E., Fay, P. A., Heisler, J. L., Leavitt, S. W., Sherry, R., Smith, B. and Weng, E. (2008) Consequences of More Extreme Precipitation Regimes for Terrestrial Ecosystems. *Bioscience*, **58**, 811-821.

Kou, X., Li, Q. and Liu, S. (2011) High-resolution bioclimatic dataset derived from future climate projections for plant species distribution modeling. *Ecological Informatics*, **6**, 196-204.

Kreyling, J., Wana, D. and Beierkuhnlein, C. (2010) Potential consequences of climate warming for tropical plant species in high mountains of southern Ethiopia. *Diversity and Distributions*, **16**, 593-605.

La Sorte, F. A. and Jetz, W. (2010) Projected range contractions of montane biodiversity under global warming. *Proceedings of the Royal Society B-Biological Sciences*, **277**, 3401-3410.

Lenoir, J., Gegout, J. C., Marquet, P. A., de Ruffray, P. and Brisse, H. (2008) A significant upward shift in plant species optimum elevation during the 20th century. *Science*, **320**, 1768-1771.

Lenoir, J., Gegout, J. C., Guisan, A., Vittoz, P., Wohlgemuth, T., Zimmermann, N. E., Dullinger, S., Pauli, H., Willner, W. and Svenning, J. C. (2010) Going against the flow: potential mechanisms for unexpected downslope range shifts in a warming climate. *Ecography*, **33**, 295-303.

Lovett, J. C. (1990) Classification and status of the moist forests of Tanzania. *Notifications from the Institute for General Botany Hamburg*, **23a**, 287–300.

Lovett, J. C., Marchant, R., Taplin, J. and Küper, W. (2005) The oldest rainforests in Africa: stability or resilience for survival and diversity? In *Phylogeny and Conservation* (Eds, Purvis, A., Gittleman, J. L. and Brooks, T. M.) Cambridge University Press, Cambridge, UK.

Marchant, R., Mumbi, C., Behera, S. and Yamagata, T. (2007) The Indian Ocean dipole - the unsung driver of climatic variability in East Africa. *African Journal of Ecology*, **45**, 4-16.

McCain, C. M. and Colwell, R. K. (2011) Assessing the threat to montane biodiversity from discordant shifts in temperature and precipitation in a changing climate. *Ecology Letters*, **14**, 1236-1245.

McClean, C. J., Lovett, J. C., Kuper, W., Hannah, L., Sommer, J. H., Barthlott, W., Termansen, M., Smith, G. E., Tokamine, S. and Taplin, J. R. D. (2005) African plant diversity and climate change. *Annals of the Missouri Botanical Garden*, **92**, 139-152.

McInerney, F. A. and Wing, S. L. (2011) The Paleocene-Eocene Thermal Maximum: A Perturbation of Carbon Cycle, Climate, and Biosphere with Implications for the Future. *Annual Review of Earth and Planetary Sciences*, **39**, 489-516.

Mitasova, H. and Mitas, L. (1993) Interpolation by regularized spline with tension. 1. Theory and implementation. *Mathematical Geology*, **25**, 641-655.

Mittermeier, R. A., Robles Gil., P., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C. G., Lamoreux, J. and da Fonseca, G. A. B. (2004) *Hotspots Revisited*, CEMEX, Mexico.

Moritz, C., Patton, J. L., Conroy, C. J., Parra, J. L., White, G. C. and Beissinger, S. R. (2008) Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science*, **322**, 261-264.

Mulligan, M. (2006) Global gridded 1km TRMM rainfall climatology and derivatives. Version 1 (http://www.ambiotek.com/1kmrainfall).

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

Pereira, H. M., Leadley, P. W., Proenca, V., Alkemade, R., Scharlemann, J. P. W., Fernandez-Manjarres, J. F., Araujo, M. B., Balvanera, P., Biggs, R., Cheung, W. W. L., Chini, L., Cooper, H. D., Gilman, E. L., Guenette, S., Hurtt, G. C., Huntington, H. P., Mace, G. M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R. J., Sumaila, U. R. and Walpole, M. (2010) Scenarios for Global Biodiversity in the 21st Century. *Science*, **330**, 1496-1501.

Peters, G. P., Marland, G., Le Quere, C., Boden, T., Canadell, J. G. and Raupach, M. R. (2012) CORRESPONDENCE: Rapid growth in CO2 emissions after the 2008-2009 global financial crisis. *Nature Climate Change*, **2**, 2-4.

Peyron, O., Jolly, D., Bonnefille, R., Vincens, A. and Guiot, J. (2000) Climate of east Africa 6000 C-14 yr BP as inferred from pollen data. *Quaternary Research*, **54**, 90-101.

Phillips, S. J., Dudik, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. and Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181-197.

Ramirez, J. and Jarvis, A. (2008) High Resolution Statistically Downscaled Future Climate Surfaces. International Centre for Tropical Agriculture, CIAT (http://www.ccafs-climate.org/).

Randin, C. F., Engler, R., Normand, S., Zappa, M., Zimmermann, N. E., Pearman, P. B., Vittoz, P., Thuiller, W. and Guisan, A. (2009) Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, **15**, 1557-1569.

Rapport, D. J., Costanza, R. and McMichael, A. J. (1998) Assessing ecosystem health. *Trends in Ecology & Evolution*, **13**, 397-402.

Rucina, S. M., Muiruri, V. M., Kinyanjui, R. N., McGuiness, K. and Marchant, R. (2009) Late Quaternary vegetation and fire dynamics on Mount Kenya. *Palaeogeography Palaeoclimatology Palaeoecology*, **283**, 1-14.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J. and Svenning, J. C. (2011) The Influence of Late Quaternary Climate-Change Velocity on Species Endemism. *Science*, **334**, 660-664.

Scherrer, D. and Koerner, C. (2011) Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography*, **38**, 406-416.

Stockwell, D. R. B. and Peterson, A. T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1-13.

Tabor, K. and Williams, J. W. (2010) Globally downscaled climate projections for assessing the conservation impacts of climate change. *Ecological Applications*, **20**, 554-565.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L. and Williams, S. E. (2004) Extinction risk from climate change. *Nature*, **427**, 145-148.

Thornthwaite, C. W. (1948) An Approach toward a Rational Classification of Climate. *Geographical Review*, **38**, 55-94.

Wilby, R., Charles, S., Zorita, E., Timbal, B., Whetton, P. and Mearns, L. (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods. Intergovernmental Panel on Climate Change (http://www.ipcc-data.org/guidelines/).

Williams, J. W., Jackson, S. T. and Kutzbacht, J. E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5738-5742.

Yee, T. W. and Mitchell, N. D. (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science*, **2**, 587-602.

Chapter 6 – Summary Discussion

Part of this chapter (endemism criteria) in preparation for: Gereau, R. E., Hall, J., Hemp, A., Platts, P. J. Distribution and endemism of vascular plants in the Eastern Arc Mountains. For submission to *East African Journal of Natural History*

This thesis explored how species and climate modelling can contribute to our understanding of phytogeographical relationships in the Eastern Arc Mountain (EAMs), providing direction for conservation and botanical survey. Chapter 2 used terrain analysis and vegetation data to provide a coherent delineation for the ecoregion; Chapter 3 investigated some emerging methodologies for species distribution modelling (SDM); and Chapters 4 and 5 explored spatial patterns of plant diversity and endemism under present and future climatic conditions. The following discussion syntheses findings in the context of ongoing research and conservation in the EAMs, and suggests some priorities for fieldwork and management.

Mountain limits

Surprisingly, a universally accepted method for deciding where a mountain begins, and whether a ‘mountain’ should be better considered a ‘hill’, does not exist (OED, 1989; Gerrard, 1990). Whilst collating spatial data and statistics for the EAMs, I frequently encountered the problem of inconsistently defined mountain extent: the study region, whilst defined qualitatively by geology and climate (Lovett, 1990), had yet to be precisely delimited for spatial analyses. In Chapter 2, mountain delineation considered not just elevation, as in a variety of previous definitions, but also steepness of slope and terrain roughness – key criteria that distinguish low mountains near a coastline from similarly elevated plateaus farther inland (Gerrard, 1990). The chosen parameterisation maximised inclusion of preclearance and present-day forest extent, and known distributions of putative endemic plant species within a minimal mountain area. The terrain units deemed optimal are similar to those used to classify landform in the European soil database (SOTER; Dobos *et al.*, 2007), suggesting potential for a global standard in high-resolution mountain delineation.

Endemism criteria

With EAM extent well defined in a GIS, ecological and anthropogenic metrics can be monitored with consistency through time and between organisations. Chapter 2 suggested that plant endemism could be more effectively defined by the ecoregion boundary than by a fixed 500 m altitudinal cut-off, as was the previous convention. Since the publication of these findings in *Environmental Conservation* (Platts *et al.*, 2011), work has continued in collaboration with Roy Gereau (Missouri Botanical Garden), Andreas Hemp (University of Bayreuth) and Jaclyn Hall (University of Stanford), and will culminate in 2012/13 with a definitive plant checklist of distribution and endemism (Gereau *et al.*, in prep.). A total of

3845 indigenous vascular plant species (4684 plant taxa) are known from the EAMs. In a major advance for plant taxonomy in the region, each of these species, subspecies and varieties has now been classified, based on the ecoregion boundary, into one of seven phytogeographical categories: (1) strictly endemic to the EAMs; or occurring only in the EAMs *and* (2) Coastal Forests of Eastern Africa, (3) Neogene Volcanoes of northern Tanzania, (4) Lake Nyasa mountain system of Tanzania, (5) two or more of the above areas, (6) elsewhere in continental Africa, or (7) outside continental Africa.

	EAM endemics		Single-bloc endemics	
	All taxa	Spp. only	All taxa	Spp. only
Taita Hills	35	26	9	7
North Pare	12	6	1	1
South Pare	50	40	5	3
West Usambara	142	105	38	26
East Usambara	121	104	41	36
Nguu	1	1	0	0
Nguru	133	112	29	22
Ukaguru	33	27	5	4
Uluguru	224	187	95	80
Malundwe	3	3	0	0
Rubeho	34	29	0	0
Udzungwa	215	188	80	70
Mahenge	44	40	12	12
All EAMs	570	471	315	261

Table 6.1. Endemism of vascular plants in the EAMs, according to the ecoregion boundary defined in Chapter 2. Distributions are as known to science in March 2012 (Gereau *et al.*, in prep.). Taxonomy was standardised to the African Flowering Plants Database (AFPD, 2010), with updates according to taxonomic revisions and monographs.

In an early analysis of the world's hotspots of biodiversity, Myers (1990) estimated that the EAMs contain 1600 plant species, of which 535 endemic, equating to under- and over-estimates of 58% and 13%, respectively. Although the latter was quite accurate given the data available at the time, the inferred level of endemism was more than 2.5 times higher than calculated here (33% *cf.* 12.2%). At the time of writing, many publications and websites still report 800 plant species (40%) endemic to the EAMs, citing Lovett and Wasser (1993), alongside a wide range of estimates for forest and mountain area. It is intended that the ecoregion definition presented in Chapter 2, combined with rigorous taxonomy and widespread collaboration, will ease such confusion by providing a consistent spatial platform upon which the state of knowledge can be continually updated.

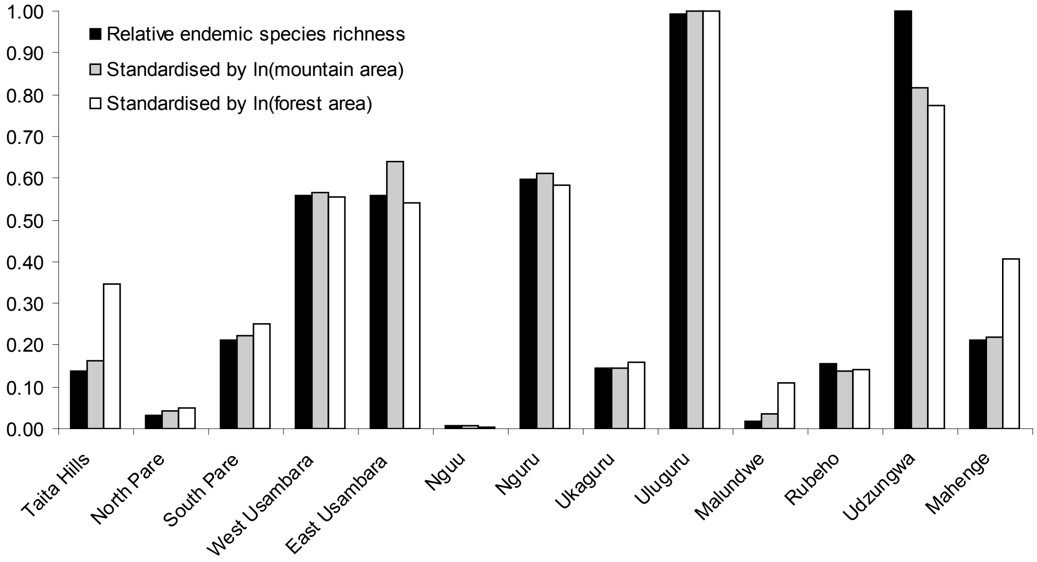


Figure 6.1. Number of endemic species in each EAM bloc, compared with numbers standardised by land or forest area, according to the ecoregion boundary defined in Chapter 2. Bar heights are relative to the maximum value across blocs.

Human populations

One quarter of the global population lives in close proximity to mountain environments (Meybeck *et al.*, 2001; Huddleston *et al.*, 2003). Countless more benefit environmentally, economically or culturally from the myriad services they provide. As decision-makers look to incorporate the economic value of well-functioning ecosystems into mainstream policy (Fisher *et al.*, 2008; Naidoo *et al.*, 2008), of particular interest is whether biological importance coincides spatially with more tangible ecosystem services such as forest products, revenue from nature-based tourism, water provision and carbon storage (Anderson *et al.*, 2009). Studies aiming to assess such a link are however hampered by the fact that administrative boundaries are rarely well-placed for collating and analysing socio-economic data within topographically distinct areas (Kreutzmann, 2001; Browne *et al.*, 2004).

Chapter 2 provided the first well-defined estimates of human population within each topographically distinct EAM bloc. Population density was shown to be around 2.5 times higher than the national average for rural Tanzania, peaking near the base of the mountains. This demonstrates the pull of resource-rich, biologically important areas for human settlement, and a potential conflict between conservation objectives and the resource needs of growing human populations – an issue that extends across the African continent (Balmford *et al.*, 2001). Population densities are highest in West Usambara and North Pare, followed by East Usambara and Uluguru (Table 6.2). Three of these blocs are also amongst the most important in terms of species richness and endemism, both for plants (Table 6.1, Figures 6.1 and Chapter 4) and for other groups of organism (Burgess *et al.*, 2007b).

Table 6.2. Predictions for the number of people living in each EAM bloc by the year 2025 (Tanzanian only). Baseline populations are as described in Chapter 2, but updated with LandScan (2008). Future scenarios impose ward-level growth (according to 1988-2002 census data), scaled by national trends (2% or 3%) and adjusted for scenarios of protected area governance (Appendix II).

	Total persons (thousands)			Mean density (pp.km^{-1})			Median density (pp.km^{-1})		
	2002	2025[a]	2025[b]	2002	2025[a]	2025[b]	2002	2025[a]	2025[b]
North Pare	67	64	69	132	127	136	34	28	41
South Pare	140	170	186	60	73	80	13	14	18
West Usambara	533	647	712	181	219	241	38	34	48
East Usambara	96	105	114	84	92	100	19	15	25
Nguu	40	57	71	25	36	45	8	10	14
Nguru	92	142	167	36	55	65	10	14	18
Ukaguru	154	212	246	48	65	76	22	30	37
Uluguru	247	336	428	81	110	140	27	30	35
Malundwe	0	0	0	0	0	0	0	0	0
Rubeho	181	255	301	23	32	38	11	15	17
Udzungwa	714	969	1141	31	42	50	12	14	17
Mahenge	58	74	83	22	28	32	9	11	13
Total	2322	3032	3517	46	60	69	13	16	20
All Tanzania	34,470	53,611	67,061	38	59	74	6	8	10

[a] Moderate growth scenario: national rate = 2 %.y^{-1}, no people in any centrally governed protected area

[b] Rapid growth scenario: national rate = 3 %.y^{-1}, growth continues outside game reserves and national parks

The ecoregion boundary has been adopted as the basis for all spatial analyses by the Valuing the Arc Programme (VTA; http://www.valuingthearc.org/), a five-year collaboration between institutions and stakeholders in Tanzania, the UK and USA, which aims to map and value ecosystem services, biodiversity and conservation cost across the EAMs, both now and under scenarios of future population growth and land use change. Floristic diversity for the VTA biodiversity module are based on the modelled (bloc-restricted) distributions for endemic/threatened taxa derived in Chapter 4, supplemented with simpler multidimensional niche envelopes for species that have too few data for multivariate regression (Chapter 5). This thesis also provides VTA (and WWF-US) with high resolution, spatially explicit forecasts for future population (Table 6.2 and Appendix II). Time series data for the national (Tanzanian) growth rate were used to construct two scenarios: moderate growth (rates begin to stabilise by 2025), and rapid growth (exponential rates continue). Spatial variation in change was extrapolated from the estimates in Chapter 2 (updated with the latest LandScan release), based on anomalies in ward-level census counts between 1988-2002 (NBS, 2002).

Methods for species distribution modelling

Fine-scale data on the distribution of biodiversity are increasingly costly to collect (Lawton *et al.*, 1998) and even in well-surveyed regions provide conservation planners with point-based estimates, rather than spatially complete information. SDM is an appealing tool for conservation planning, with potential to elucidate fine-scale patterns of biodiversity and to forecast response to environmental change. Model predictions, however, can vary depending on the methods employed (Chapter 3). Here, I used generalised additive models (GAMs) to describe the relationship between species presence and environmental gradients. As described in the General Introduction, GAMs can parameterise a limitless variety of response shapes, as determined by the data. With this freedom comes a risk of over-fitting, and therefore particular sensitivity to the methods for model calibration. Chapter 3 explored the sensitivity of GAM predictions to sample and variable selection, data weights and spatial autocorrelation. Latter chapters on conservation priority and climate change implemented models with these findings in mind, as summarised below.

Sample selection

A common problem in SDM is that the available distribution data were not collected with high-resolution, logistic regression in mind (Graham *et al.*, 2004). Key obstacles include a rarely satisfied assumption that observations be independent and identically distributed (Graham, 2003; Miller *et al.*, 2007), and the need for absence data to constrain predictions. Absence data cannot be 'collected' in the literal sense. Instead, pseudo-absences are inferred from plot data, targeted or exhaustive sampling, or computer-generated. Comparative analyses find that predictive models constructed using some kind of absence data, even if distributed randomly, are generally superior to presence-only methods, which tend to over-predict areas of occupancy (Meynard and Quinn, 2007; Chefaoui and Lobo, 2008).

Chapter 3 used data from variable area plot assessments to relate known distributions of 40 large tree taxa to climatic and topographic conditions at sample sites. In order to constrain a logistic response, absences were generated at sites that had been surveyed by the same field botanists (Jon Lovett and Jon Hall) but where the target species had not been recorded. Chapter 4, which modelled 452 taxa using all available herbarium and plot data (Ahrends, 2010), placed absences at sites where a corresponding growth form of plant had been collected, so as to account for sampling bias towards certain structural groups (usually trees). In Chapter 5, absences were targeted toward known endemism sites, because of the skill and confidence required to correctly identify such species and the spatial bias this may impose on known distributions (Ahrends *et al.*, 2011b in Appendix I). In order to exclude sites likely to be dispersal limited, rather than environmentally unsuitable *per se*, and to negate bloc-level bias in the sampling distribution, Chapters 4 and 5 sampled absence data only from mountain blocs where the target species is known to occur.

Engler *et al.* (2004) have suggested a two-step procedure, whereby absence data for logistic regression are placed in areas deemed unsuitable by a prior presence-only model (e.g., ecological-niche factor analysis; Hirzel *et al.*, 2002). Model accuracy in this case may however be artificially inflated because absences are necessarily further from the observed niche (Chefaoui and Lobo, 2008; Wisz and Guisan, 2009), without accounting for spatial bias in the sampling distribution (Phillips *et al.*, 2009). Somewhat counter-intuitively, absence data biased toward the same geographical (and therefore environmental) domain as the presence data – as exemplified throughout this thesis – have been shown to yield the most robust models (Zaniewski *et al.*, 2002; Chefaoui and Lobo, 2008; Phillips *et al.*, 2009). Such data are more likely to identify unsuitable sites in accessible locations than they are the reciprocal – i.e., suitable sites that are dispersal limited (Soberon and Nakamura, 2009). Also, spatial structure in presence data for a particular taxonomic group or sampling design is reflected in the absence distribution, increasing confidence that predicted occurrence is not an artifact of non-independence of field observations (e.g., Chapman, 2010).

Data weights

Sample sizes for most plants in the EAMs are small, especially amongst rare species of high conservation concern – a ubiquitous problem in biodiversity hotspots – and so choosing absence data to outnumber presence data is often necessary to provide adequate representations of environmental variation across the region for extrapolation (Chapter 4). Because low sample prevalence can bias validation scores (McPherson *et al.*, 2004), Chapter 3 explored the possibility of down-weighting absence data to impose a standardised sample prevalence of 0.5. Weighting the data was found to shift the error distribution in favour of higher sensitivity (reduced type II error), which, given the inherent uncertainty of pseudo-absence data compared with verified presence data, might be a desirable bias to impose (Anderson, 2003). This method had been tested previously in regions where species distribution data are more complete (e.g., Europe; Maggini *et al.*, 2006), but critical applications to smaller datasets, more typical of tropical forest species, were lacking.

Chapter 3 found that down-weighting absence data increased the degrees of freedom used by GAMs (i.e., response shapes were more complex). If non-weighted models were robust under cross-validation, then weighted models produced favourable results. If unweighted models were unstable, however, then forcing a standardised prevalence exacerbated prediction uncertainty. Before applying this method to derive spatial estimates of conservation priority in Chapter 4, a preliminary analysis was conducted to assess the relative merits of data weights under various ratios of presence/absence. A model prevalence of 0.2 (five absence points for every presence), implemented without weights, was found to be the most robust approach, representing a sufficiently wide range of environmental conditions for extrapolation, without negative impact on validation scores or model stability.

Variable selection

A popular method for selecting model predictors from an *a priori* set is to employ a stepwise procedure, whereby candidate variables are removed or added according to some metric of model fit or predictive power. Such a procedure begins with either a full model (all variables) or a null model (no variables), and culminates in a ‘best’ (most parsimonious) solution for describing the response. Although rarely discussed in the literature, Chapter 3 found that the direction of stepwise reduction/addition can have a marked effect on which model is considered optimal, and thus on the inference one draws from the patterns predicted. In cases where strong species-environment relationships were difficult to detect, initiating selection with a full predictor set typically resulted in larger, less stable, final models than if selection began with a null set. Chapter 4 found greater agreement between stepwise procedures at higher spatial resolutions (1 km *vs*. 2 km), presumably because environmental gradients relevant for plant distributions were more effectively captured. Since different solutions can often explain the same species distribution equally well, averaging over more than one model is a sensible approach to spatial prediction. The method suggested in Chapter 3, and implemented in Chapters 4 and 5 (see also Marshall *et al.*, 2012 in Appendix I), is to obtain two or more 'best-model' solutions *via* bidirectional selection, beginning with first a null and then a full model, in each case testing all possible additions and removals at each selection step. These solutions are then weighted in a multimodel according to their relative cross-validated performance.

Alongside environmental predictors, Chapter 3 tested the inclusion of spatial autocovariates to account for fine-scale spatial aggregation in species distributions. The analysis showed that autocovariates had potential to significantly increase explained deviance, but that if environmental constraints were underrepresented then model stability and explanatory power could be compromised (see also Dormann, 2007). In Chapter 4, autocovariates were retained only if they improved model performance on unseen data – true in around one third of cases. In Chapter 5, autocovariates were not included, because climate change might alter the observed aggregation of individuals, for example due to changes in a competitor's range, or other unseen factors blindly captured by the autocovariate under contemporary conditions. Topographic variables were excluded from climate change models for similar reasons, these being proximate factors with no direct relation to the species response.

Environmental relationships

Five climatic variables were found to be of particular importance for predicting plant distributions in the EAMs: mean annual temperature; annual range in temperature; annual rainfall; an associated moisture index (rainfall / potential evapotranspiration); and cumulative water deficit, which measures the length and severity of the dry season. Other physiologically relevant variables, such as minimum and maximum temperature (affecting respiration and photosynthesis), are highly correlated with the above set and so do not add substantially to predictive power (Chapter 3). Where available, estimates of fog/mist should be considered alongside temperature and rainfall. Data from the MODIS MOD35 Cloud Mask Product (Mulligan, 2006) were found to be good predictors of some species distributions in Chapter 4. Fog estimates are currently too crude to be useful in the EAMs. For example, FIESTA (Mulligan and Burke, 2005) imposes an arbitrary lower altitudinal bound, such that no fog at all is predicted for low coastal ranges such as East Usambara. The inclusion of topographic variables such as slope and aspect may help to represent microclimatic variation and fog capture, albeit indirectly. As demonstrated throughout this thesis, it is crucial to include measures of moisture and seasonality alongside mean annual temperature (Stephenson, 1990; Crimmins *et al.*, 2011; Greve *et al.*, 2011), particularly when extrapolating models under scenarios of future change (Chapter 5).

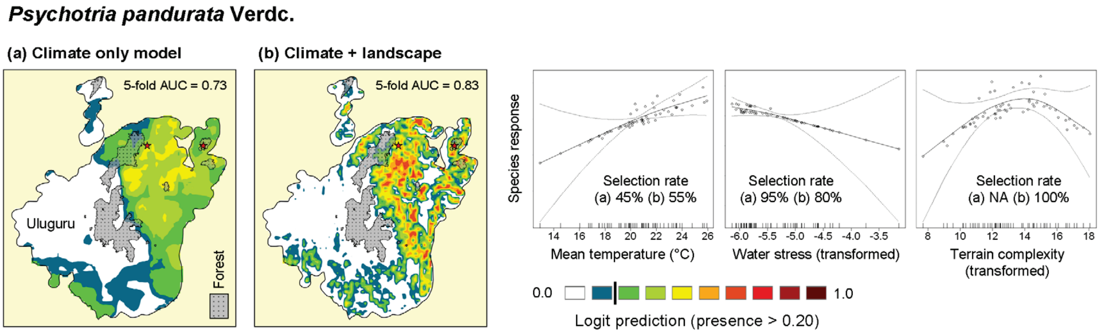


Figure 6.2. Modelled distribution of *Psychotria pandurata* in Uluguru, using (a) climate variables only and (b) climate plus terrain complexity (defined as the altitudinal range within 1 km squares).

Beside climate, geological and edaphic factors are important for determining the suitability of a site for a particular plant species (Quesada *et al.*, 2009; Ashcroft *et al.*, 2011). The EAMs, by definition, share a common Precambrian crystalline substrate, including some metamorphosed limestone at the base of Uluguru. Species composition is markedly different to forests on volcanic soils (Lovett, 1990). Spatial variability in soil properties could nonetheless be high across the EAMs (e.g., water capacity and pH), and so the application of broad-scale estimates may be problematic. Until such time as soil properties can be accurately derived from extensive local survey data, it is preferable that species distributions be modelled using topoclimatic factors. For example, Fig. 6.2 shows how the inclusion of terrain complexity improves the model for *Psychotria pandurata* (Rubiaceae), an endemic herb found on damp ground in dense forest. The best climatic predictors are mean temperature, which helps to explain its altitudinal range (300-850 m), and dry season water stress, which correctly identifies its preference for moist windward flanks (now largely deforested in Uluguru; Fig. 6.2). Terrain complexity adds fine-scale detail to the prediction, improving performance under cross-validation, most likely due to its proximal relation with soil moisture at sub 1-km scales.

In Chapter 4, models for tree species performed significantly better under cross-validation than did models for lianas, shrubs or herbs. Key information missing from models for understory species such as *P. pandurata*, both in the EAMs and elsewhere, is how climatic conditions vary beneath the canopy. In order to address this data gap, the KITE research group (York) have begun installing temperature data loggers and taking infrared thermometer readings at a number of forest and non-forest sites across the EAMs. Also, automated weather stations, which log 35 meteorological measurements every ten seconds, have been installed along transects in the Taita Hills, on Mt. Kilimanjaro and in the Ethiopian Highlands, as part of the CHIESA project (Climate Change Impacts on Ecosystem Services and Food Security in Eastern Africa; http://chiesa.icipe.org/). These kinds of data will help to ground-truth the remotely sensed and interpolated climatologies used here for species distribution modelling. Long-term monitoring, or using space as a surrogate for time, should help to better predict how future changes in climate and adjacent land use could affect species beneath the canopy.

Combining modelled distributions across multiple species, high richness was found to correlate most strongly with low temperature range (Chapter 4), echoing trends previously documented at larger spatial scales (e.g., Jetz *et al.*, 2004). Concentrations of rare and endemic plants were better explained by the moisture index, a result consistent with the theory of long-term constancy in the water balance, mediated by trade winds from the Indian Ocean (Marchant *et al.*, 2007), being an important factor underlying the region's high levels of endemism (Fjeldså and Lovett, 1997). Model extrapolations in Chapter 4, especially between mountain blocs, were found to be more uncertain for rare taxa (even when correcting for sample size) – possibly a consequence of lower dispersal capacities, unknown colonisation histories and the importance of past climates in shaping contemporary patterns of occurrence (Jansson, 2003; Jetz *et al.*, 2004; Graham *et al.*, 2006; Dobrowski *et al.*, 2011). Intriguingly, projected spatial variation in future climate suggests that the absence of extreme change, particularly with respect to seasonality and moisture, coupled with the ability of species to track transient conditions within the limits of a given mountain bloc, may be indicative of rare species persistence in the past (Chapter 5).

Socioeconomic relationships

Chapter 2 identified a strong relationship between the density of people (on a 1 km grid) and proximity to the EAMs. In order to quantify the impact of these populations on forest health and biodiversity, it is appropriate to consider a derived measure of population 'pressure'. This is because disturbance and degradation, through demand for natural resources, is mediated by pressure from both local settlements (usually situated outside protected areas) and remote demand centres, such as towns and cities. Assuming that the pressure on a forest patch *i* increases linearly with the number of persons (*p*) in a remote location *j*, and that the relative influence (*w*) of the remote population decreases with distance (*d*), then:

$$pressure_i = \sum_{j=1}^{N} p_j \cdot w_{ij}, \qquad w_{ij} = \exp\left(-\left(d_{ij}/\sigma\right)^2\right)$$

where *N* is the number of locations across which pressure accumulates and σ changes the shape of the half-normal distance decay. A sigma value of 25 (pressure halves over a distance of 20 km) is the most significant predictor in a model that explains over 40% of variation in the percentage of trees cut along transects (fully independent test data; Ahrends and Platts, unpublished). The pressure grids are also significant predictors of carbon stock (negative correlation, indicative of degradation; Willcock *et al.*, in prep.), as well as of expenditure on protected area management and estimated spend required to meet conservation objectives (positive correlations; Green *et al.*, 2012).

Plotting population pressure against modelled distributions of endemic/threatened plant taxa, it is apparent that, within the EAMs, pressures are often greater in sites with more taxa of conservation concern (Fig. 6.3). Similar correlations have been observed at larger spatial scales, comparing the EAMs, Albertine Rift and Cameroon-Nigeria Mountains with the rest of sub-Saharan Africa (Burgess *et al.*, 2007a). Within the EAM chain, there are likely to be a number of factors at play. First, both the plant data and their modelled distributions could be biased by sampling effort toward more accessible locations (Chapter 4), which means closer to villages, roads and towns, and thus higher population pressure. Second, people tend to settle in areas with more reliable rainfall, both seasonally and annually – conditions also hypothesised to be important for species persistence and endemism (Chapter 5). Both these factors are influenced by proximity to the Indian Ocean: coastal areas are both densely populated (historical connections to the Indian subcontinent *via* Dar es Salaam and Zanzibar) and also the source of consistent orographic rainfall and mist.

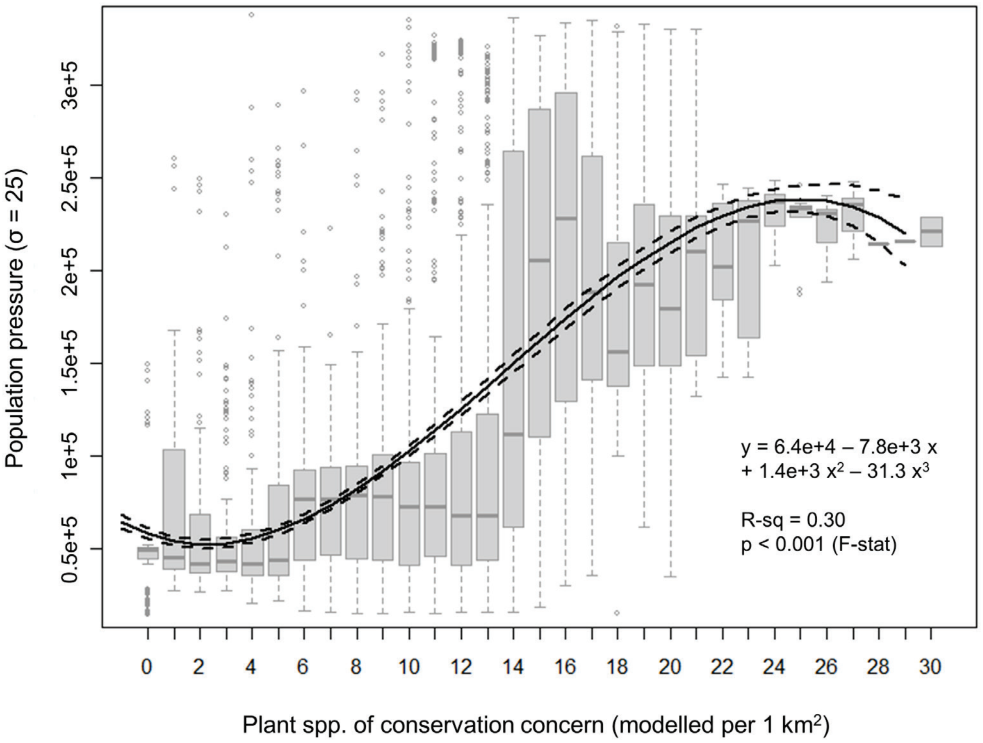


Figure 6.3. Population pressure *vs*. number of endemic/threatened plant taxa with environmentally suitable conditions in EAM forests (bloc-restricted distributions, Chapter 4). Population pressure was accrued across the Tanzanian watershed of the EAMs, with the weights given to remote populations (relative to the local population) halving over a distance of 20 km. Dashed lines show 95% bootstrap confidence intervals.

Together, these patterns imply that the costs of protected area management in the EAMs happen to be highest in the most biologically important areas, and that strict forest protection based only on global conservation objectives could impact negatively on the livelihoods of many local people (Balmford and Whitten, 2003), resulting in either (1) poverty, or (2) illegal resource use and/or displacement of forest degradation and biodiversity loss to unprotected sites (leakage). Protected areas therefore provide an important framework for species conservation and should be invested in, but not without communication with local communities, potentially involving them in management, and finding ways to address their continuing resource requirements (Burgess *et al.*, 2007a; Blomley *et al.*, 2008).

Directions for fieldwork

Conservation priorities are sensitive to perceived biological importance, which in turn is sensitive to under-collection and spatial bias in the inventory data (Chapter 4). Moreover, it may be the case that priorities are self-perpetuating, with sites overlooked during early explorations being less attractive for fieldwork than those already known to be species-rich (Ahrends *et al.*, 2011a in Appendix I). Whilst I believe that, to some extent, climatic and geographic factors have influenced species rarity and sampling effort alike (acting through distance from coastal climates/conurbations), it is sensible to keep an open mind regarding the relative biological importance of under-researched mountain blocs, as well as lesser-known (and under-funded) forest reserves in general, as exemplified by recent fieldwork in Rubeho (Doggart *et al.*, 2006). Even in Uluguru, new plant species are still being discovered (e.g., Haston *et al.*, 2009). At the global scale, analysis of taxonomic effort over time indicates that biodiversity hotspots do indeed house the most undiscovered plant species, and that Tanzania in particular has much to gain from further exploration (Joppa *et al.*, 2011).

In the best-surveyed mountain blocs (East and West Usambara, Uluguru and Udzungwa), the species distributions modelled in Chapter 4 provide useful information for conservation planning. A sensible way forward in other areas is to use SDM and fieldwork iteratively, both to improve knowledge of species' ranges and to maximise new species discoveries. In practice, this could work in one of two ways. Either fieldwork would be strategically targeted toward sites predicted to be viable for a particular species of conservation concern (e.g., Menon *et al.*, 2010), or toward sites where many more species are predicted than are currently recorded. If the primary purpose of fieldwork is something other than biodiversity assessment (e.g., measurement of carbon), then one could still obtain predicted distributions for the region's rare species prior to fieldwork. In both cases, ground-truthed presence *and* absence data should be recorded, as well as climate and soils data whenever feasible, especially at sites identified as novel in these regards by envelope uncertainty maps (Chapters 3 and 4). This new information would improve model accuracy, so that refined predictions could be made, ready for further validation on subsequent field trips. As the accuracy of correlative environmental SDM ceases to improve, one obtains an increasingly coherent basis upon which to investigate other factors, such as patch-dynamics and dispersal limitation, biotic interactions and non-equilibrium states, and potentially incorporate them in a nested modelling framework (Zurell, 2011; Godsoe and Harmon, 2012). In an application of the SDM and climate downscaling techniques developed here (Chapters 3-5), the KITE research group is currently embarking on a program of iterative improvement in models for the genus *Acacia* (Fabaceae) across East Africa. Early signs are that northern and eastern Kenya remain underrepresented in existing field guides – predictions to be investigated on the ground in the summer of 2012 (Marshall *et al.*, 2012 in Appendix I).

Anomalies between observed and predicted richness in the EAMs suggest that the importance of Nguru and Rubeho remains underestimated, despite much investment in botanical inventories since 2005. There is also potential for much higher biological diversity in Nguu than apparent from inventory data (Chapter 4), a consequence of being almost entirely overlooked by previous vegetation assessments. Fieldwork undertaken as part of this PhD confirmed that Nguu forests are in generally good health, and contain at least one endemic plant species (*Anisotes tangensis*) as well as many rare herpetofauna (Michele Menegon, pers. comm.). Nguu is subject to the lowest population pressure of all the EAM blocs, being sparsely populated (Table 6.2) and relatively distant from major roads and urban centres. Hydrological runoff is high (Table 6.3 and Appendix III), providing an important ecosystem service to more densely populated areas downstream (mountains drain to the Wami River). Indeed, many of the forest reserves in eastern Tanzania are gazetted for catchment protection, although they are often poorly funded and have few staff. Since population pressure is low in Nguu, especially on western peaks (currently unprotected, Chapter 2), the necessary spend to fulfil conservation objectives is likely to be lower than elsewhere in the EAMs (Green *et al.*, 2012).

Further to considering spatial patterns in overall plant diversity, this thesis has highlighted the tendency for vegetation plot assessments to favour larger growth forms of plant. Woody species are by far the most often recorded, with trees accounting for over 80% of data in most mountain blocs. Importantly, models ranked the mountains slightly differently in terms of species richness and rarity, depending on the growth form considered (Chapter 4). Given that over 80% of endemic plant species, as here defined, are *not* trees, this is an important consideration for conservation planning (Gentry, 1992). Sampling protocols that target smaller stems (< 10 cm dbh), at least in subplots, are therefore encouraged to rebalance the biodiversity data. Gentry (1982) suggests recording all stems ≥ 2.5 cm dbh (in ten randomly placed 20×50 m plots) in order to gain a fair representation of diversity patterns. Recent simulation modelling has confirmed the efficacy of this method for tropical forests, but finds that a simple hump-shaped peak in species richness is rarely captured if only stems ≥ 10 cm dbh are sampled, as is more often the convention (Ahrends *et al.*, in prep.).

Plants in context

As one of the better-studied taxonomic groups, vascular plants are sometimes employed as indicators of overall biodiversity. However, since distributional patterns vary even within this group (Chapter 4), it is prudent to consider all available biodiversity data in management plans (Gentry, 1992; Bladt *et al.*, 2008; Larsen *et al.*, 2009). Incorporation of non-species factors, such as threats to habitat, resource provision and forecasts of future environmental change are also relevant (Burgess *et al.*, 2006). To this end, Table 6.3 places the plant priorities derived in Chapter 4 in the context of a more holistic management perspective. The geographical units for this analysis reflect the existing protected area network of forest reserves, nature reserves and national parks, which combined cover around 75% of the EAM forest area (Chapter 2). I further distinguish between State and District governance in Tanzania, information not yet fully covered in the WDPA (UNEP-WCMC, 2010), but which has been collated, mapped and submitted for inclusion in the next release of the dataset.

Modelled distributions of rare, threatened and (near-) endemic animal species were provided by Green (2012). These include 41 species of mammal (Global Mammal Assessment plus African Elephant Database), 76 species of bird (extents of occurrence from Birdlife International, 2008; areas of occupancy by expert opinion), 57 species of amphibian (Global Amphibian Assessment) and 14 species of chameleon (extents of occurrence from Spawls *et al.*, 2004; areas of occupancy by expert opinion). When comparing conservation priorities across taxonomic groups, it is important to note differences in modelling approach. The plant models used in Table 6.3 required at least ten spatially distinct presence points for calibration, and thus favour blocs where many vegetation assessments have been conducted (notably East Usambara, *cf.* Table 6.3 and Figure 6.1). By contrast, estimates for animal distribution required only one presence point: the method was to refine extents of occurrence (quarter degree grid square) by expert opinion, rather than statistical regression against environmental factors (Rondinini *et al.*, 2006). Accordingly, priorities for plant and animal conservation (standardising by reserve/park area) are highest in the East and West Usambaras, respectively, with reserves in Uluguru and Udzungwa also ranked in the top ten. Three protected areas are in the top ten for both groups: Nilo Nature Reserve and Baga 1 (Mzinga) Forest Reserve (both East Usambara), and Uluguru Nature Reserve. Baga 1 is also in the top ten for population pressure, indicating high conservation cost in the face of agricultural encroachment and demand for natural resources such as firewood and poles. Similarly, Kitara Ridge, Ndelemai and Mkussu Forest Reserves (West Usambara) are all in the top ten for both animals and population pressure.

Table 6.3. Ranked priorities for protected area management in the EAMs, standardised by area, with respect to biological importance, service provision and the degree of threat. Bold names identify reserves/parks that appear in the top ten for two or more metrics (from a total of 115 reserves/parks, each ≥ 5 km^2). FR = Forest Reserve.

	Biological importance[1]		Service provision[2]		Degree of threat[3]	
Rank	Plants (N=71)	Animals (N=204)	Carbon storage	Hydrological runoff	Population pressure	Climate change
1	Magoroto (eU) State FR	**Baga 1 / Mzinga** (wU) State FR	Ihanga (Ud) District FR	**Mselezi** (Mh) State FR	Mindu (Ul) State FR	Udzungwa Scarp (Ud) State FR
2	Amani (eU) Nature Reserve	**Kitara Ridge** (wU) District FR	Iwande (Ud) State FR	Pumula (Nu) State FR	**Ndelemai** (wU) State FR	Lulanda (Ud) State FR
3	Kwamgumi (eU) State FR	**Ndelemai** (wU) State FR	Nyanganje (Ud) State FR	Muhulu (Mh) State FR	**Mkussu** (wU) State FR	Mufindi Tea Estate (Ud) State FR
4	**Nilo** (eU) Nature Reserve	**New Dabaga-Ulong.** (Ud) State FR	Mvuha (Ul) State FR	Nguru South (Nr) St ate FR	Balangai West (wU) State FR	Mufindi Scarp (Ud) State FR
5	Mtai (eU) State FR	**Mkussu** (wU) State FR	Ruvu (Ul) State FR	Mahenge Scarp (Mh) State FR	**Kitara Ridge** (wU) District FR	Mninga (Ud) State FR
6	**Baga 1 / Mzinga** (wU) State FR	**Nilo** (wU) Nature Reserve	**Mselezi** (Mh) State FR	Mbegere (Nu) State FR	**Baga 1 / Mzinga** (wU) State FR	Kigogo (Ud) State FR
7	Segoma (eU) State FR	Kisima Gonja (wU) State FR	Udzungwa Mts. (Ud) National Park	Nderema (Nu) State FR	Mpalalu (wU) District FR	Gulosilo (Ud) State FR
8	Kambai (eU) State FR	Idewa (Ud) District FR	Mkungwe (Ul) State FR	Mwalugulu (Mh) State FR	Muhezangulu (wU) State FR	**New Dabaga-Ulong.** (Ud) State FR
9	**Uluguru** (Ul) Nature Reserve	**Uluguru** (Ul) Nature Reserve	Kilombero (Ud) Nature Reserve	Mkuli (Nu) State FR	Mtumbi (wU) District FR	Irundu (Ud) State FR
10	Mkumbi (wU) State FR	Balangai West (wU) State FR	Pala Mountain (Ru) State FR	Chome (sP) State FR (proposed NR)	Shume Magamba (wU) State FR	Kiverenge (nP) State FR

[1] Modelled distributions of 71 endemic and/or threatened plant taxa (bloc-restricted predictions, Chapter 4) and 204 rare animal species (provided by Green, 2012)
[2] Carbon estimates according to Willcock *et al.* (in prep.). Water is annual hydrological runoff, described in detail in Appendix III
[3] Population pressure as described in the text (σ=25). Climate change is the relative anomaly by 2090 (IPCC-AR4 scenario A1B), summed over four climatic gradients (Chapter 5)

Carbon data are based on regression of field measurements of above ground live carbon against environmental and socioeconomic predictors, and ratios thereof for the remaining IPCC carbon pools (Willcock *et al.*, in prep.). Carbon storage, per unit area, is highest for forest reserves in Udzungwa, which, according to regional climate forecasts (Chapter 5), are the most likely to experience rapid climate change during the 21st century. These are amongst the larger tracts of forest remaining in the EAMs, and so niche persistence *via* dispersal may be more viable than elsewhere. Carbon stocks are similarly high in Uluguru, Rubeho and Mahenge. The United Nations REDD programme (Reduced Emissions from Deforestation and Forest Degradation) aims to mitigate global warming by offering incentives for the widespread conservation of forest carbon. As a co-benefit, REDD has potential to greatly reduce extinction rates for species the tropics (Venter *et al.*, 2009; Harvey *et al.*, 2010; Gardner *et al.*, 2012), perhaps by as much as 46-82% over a five year period, depending on the level of investment (Busch *et al.*, 2011). The costs of conservation, however, vary greatly from place to place (Balmford *et al.*, 2003; Naidoo and Ricketts, 2006) and in some regions exceed likely revenues from the carbon market (Fisher *et al.*, 2011). Building the case for forest conservation on the basis of biological importance remains a priority.

Mselezi Forest Reserve (Mahenge) is ranked sixth for carbon, and top for hydrological runoff (see Appendix III for details of the water model). Although not the highest ranking reserve in terms of runoff per unit area, Uluguru Nature Reserve is very important for water provision because of the large number of downstream beneficiaries in Dar es Salaam (*via* the Ruvu River). A nearby reserve, Mindu, is ranked top for population pressure, situated on an outlying peak adjacent to Morogoro town. Appendix III explores a hypothetical scenario of uncontrolled agricultural expansion in Uluguru, and finds that further forest clearance could lead to increased river flow during the wet season, with resultant risks such as flooding and loss of hydropower, whilst potentially reducing flow in the dry season because of less water discharging slowly as baseflow – a pattern consistent with observed trends in the second half of the 20th century (Lopa *et al.*, 2012). Alongside biological importance, forest products and payments for carbon storage, the value of mountain forests as the 'water towers' of Africa remains a powerful political incentive for investing in their conservation, especially given current rates of population growth and the uncertainty of climate change (Messerli *et al.*, 2004; Chapter 5).

References

AFPD (2010) *African Flowering Plants Database (version 3.3.2)*, Conservatoire et Jardin botaniques de la Villle de Genève and South African National Biodiversity Institute, Pretoria, South Africa (http://www.ville-ge.ch/musinfo/bd/cjb/africa/)

Ahrends, A. (2010) *Do we know enough? The impacts of data biases, insufficient sampling and degradation on biodiversity estimates in Tanzanian forests – implications for conservation planning*. Ph.D. Thesis. University of York, UK.

Ahrends, A., Bulling, M. T., Platts, P. J., Marchant, R., Burgess, N. D., Lovett, J. C. and Rahbek, C. (in prep.) Sampling protocols for assessing plant species richness in tropical forests.

Ahrends, A., Burgess, N. D., Gereau, R. E., Marchant, R., Bulling, M. T., Lovett, J. C., Platts, P. J., Kindemba, V. W., Owen, N., Fanning, E. and Rahbek, C. (2011a) Funding begets biodiversity. *Diversity and Distributions*, **17**, 191-200.

Ahrends, A., Rahbek, C., Bulling, M. T., Burgess, N. D., Platts, P. J., Lovett, J. C., Kindemba, V. W., Owen, N., Sallu, A. N., Marshall, A. R., Mhoro, B. E., Fanning, E. and Marchant, R. (2011b) Conservation and the botanist effect. *Biological Conservation*, **144**, 131-140.

Anderson, B. J., Armsworth, P. R., Eigenbrod, F., Thomas, C. D., Gillings, S., Heinemeyer, A., Roy, D. B. and Gaston, K. J. (2009) Spatial covariance between biodiversity and other ecosystem service priorities. *Journal of Applied Ecology*, **46**, 888-896.

Anderson, R. P. (2003) Real vs. artefactual absences in species distributions: tests for Oryzomys albigularis (Rodentia : Muridae) in Venezuela. *Journal of Biogeography*, **30**, 591-605.

Ashcroft, M. B., French, K. O. and Chisholm, L. A. (2011) An evaluation of environmental factors affecting species distributions. *Ecological Modelling*, **222**, 524-531.

Balmford, A. and Whitten, T. (2003) Who should pay for tropical conservation, and how could the costs be met? *Oryx*, **37**, 238-250.

Balmford, A., Gaston, K. J., Blyth, S., James, A. and 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. and Rahbek, C. (2001) Conservation conflicts across Africa. *Science*, **291**, 2616-2619.

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

Blomley, T., Pfliegner, K., Isango, J., Zahabu, E., Ahrends, A. and 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.

Browne, T., Fox, R. and Funnell, D. (2004) The "invisible" mountains - Using GIS to examine the extent of mountain terrain in south Africa. *Mountain Research and Development*, **24**, 28-34.

Burgess, N. D., Balmford, A., Cordeiro, N. J., Fjeldså, J., Kuper, W., Rahbek, C., Sanderson, E. W., Scharlemann, J. P. W., Sommer, J. H. and Williams, P. H. (2007a) Correlations among species distributions, human density and human infrastructure across the high biodiversity tropical mountains of Africa. *Biological Conservation*, **134**, 164-177.

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. and Stuart, S. N. (2007b) The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, **134**, 209-231.

Burgess, N. D., Hales, J. D., Ricketts, T. H. and Dinerstein, E. (2006) Factoring species, non-species values and threats into biodiversity prioritisation across the ecoregions of Africa and its islands. *Biological Conservation*, **127**, 383-401.

Busch, J., Godoy, F., Turner, W. R. and Harvey, C. A. (2011) Biodiversity co-benefits of reducing emissions from deforestation under alternative reference levels and levels of finance. *Conservation Letters*, **4**, 101-115.

Chapman, D. S. (2010) Weak climatic associations among British plant distributions. *Global Ecology and Biogeography*, **19**, 831-841.

Chefaoui, R. M. and Lobo, J. M. (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210**, 478-486.

Crimmins, S. M., Dobrowski, S. Z., Greenberg, J. A., Abatzoglou, J. T. and Mynsberge, A. R. (2011) Changes in Climatic Water Balance Drive Downhill Shifts in Plant Species' Optimum Elevations. *Science*, **331**, 324-327.

Dobos, E., Daroussin, J. and Montanarella, L. (2007) A Quantitative Procedure for Building Physiographic Units for the European SOTER Database. In *Digital Terrain Modelling. Development and Applications in a Policy Support Environment* (Eds, Peckham, R. J. and Jordan, G.) Springer, Berlin, Germany.

Dobrowski, S. Z., Thorne, J. H., Greenberg, J. A., Safford, H. D., Mynsberge, A. R., Crimmins, S. M. and Swanson, A. K. (2011) Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. *Ecological Monographs*, **81**, 241-257.

Doggart, N., Perkin, A., Kiure, J., Fjeldså, J., Poynton, J. and 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.

Dormann, C. F. (2007) Assessing the validity of autologistic regression. *Ecological Modelling*, **207**, 234-242.

Engler, R., Guisan, A. and Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263-274.

Fisher, B., Edwards, D. P., Giam, X. and Wilcove, D. S. (2011) The high costs of conserving Southeast Asia's lowland rainforests. *Frontiers in Ecology and the Environment*, **9**, 329-334.

Fisher, B., Turner, K., Zylstra, M., Brouwer, R., de Groot, R., Farber, S., Ferraro, P., Green, R., Hadley, D., Harlow, J., Jefferiss, P., Kirkby, C., Morling, P., Mowatt, S., Naidoo, R., Paavola, J., Strassburg, B., Yu, D. and Balmford, A. (2008) Ecosystem services and economic theory: integration for policy-relevant research. *Ecological Applications*, **18**, 2050-2067.

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

Gardner, T. A., Burgess, N. D., Aguilar-Amuchastegui, N., Barlow, J., Berenguer, E., Clements, T., Danielsen, F., Ferreira, J., Foden, W., V., K., Khan, S. M., Leesm, A. C., Parry, L., Roman-Cuesta, R. M., Schmitt, C. B., Strange, N., Theilade, I. and Vieira, I. C. G. (2012) A framework for integrating biodiversity concerns into national REDD+ programmes. *Biological Conservation*, in press.

Gentry, A. H. (1982) Patterns of neotropical plant-species diversity. *Evolutionary Biology*, **15**, 1-85.

Gentry, A. H. (1992) Tropical forest biodiversity - distributional patterns and their conservation significance. *Oikos*, **63**, 19-28.

Gereau, R. E., Hall, J., Hemp, A. and Platts, P. J. (in prep.) Distribution and endemism of vascular plants in the Eastern Arc Mountains.

Gerrard, A. J. (1990) *Mountain Environments: an examination of the Physical Geography of Mountains*, Belhaven Press, London, UK.

Godsoe, W. and Harmon, L. J. (2012) How do species interactions affect species distribution models? *Ecography*, in press.

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

Graham, C. H., Moritz, C. and Williams, S. E. (2006) Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 632-636.

Graham, M. H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology*, **84**, 2809-2815.

Green, J. M. H. (2012) *Incorporating Costs and Processes into Systematic Conservation Planning in a Biodiversity Hotspot*, PhD Thesis, University of Cambridge, UK.

Green, J. M. H., Burgess, N. D., Green, R. E., Madoffe, S. S., Munishi, P. K. T., Turner, R. K. and Balmford, A. (2012) Estimating management costs of protected areas: a novel approach from the Eastern Arc Mountains, Tanzania. *Biological Conservation*, in press.

Greve, M., Lykke, A. M., Blach-Overgaard, A. and Svenning, J.-C. (2011) Environmental and anthropogenic determinants of vegetation distribution across Africa. *Global Ecology and Biogeography*, **20**, 661-674.

Harvey, C. A., Dickson, B. and Kormos, C. (2010) Opportunities for achieving biodiversity conservation through REDD. *Conservation Letters*, **3**, 53-61.

Haston, E. M., Mejissa, J. and Watkins, C. (2009) Two new species of *Saintpaulia* from the Uluguru Mountains, Tanzania. *Curtis's Botanical Magazine*, **26**, 270-272.

Hirzel, A. H., Hausser, J., Chessel, D. and Perrin, N. (2002) Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027-2036.

Huddleston, B., Ataman, E., de Salvo, P., Zanetti, M., Bloise, M., Bel, J., Franceschini, G. and Fè d'Ostiani, L. (2003) *Towards a GIS-based Analysis of Mountain Environments and Populations*, Environment and Natural Resources Working Paper No. 10, Food and Agriculture Organisation of the United Nations, Rome, Italy.

Jansson, R. (2003) Global patterns in endemism explained by past climatic change. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 583-590.

Jetz, W., Rahbek, C. and 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.

Joppa, L. N., Roberts, D. L., Myers, N. and Pimm, S. L. (2011) Biodiversity hotspots house most undiscovered plant species. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 13171-13176.

Kreutzmann, H. (2001) Development indicators for mountain regions. *Mountain Research and Development*, **21**, 132-139.

LandScan (2008) High Resolution global Population Data Set copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy.

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

Lawton, J. H., Bignell, D. E., Bolton, B., Bloemers, G. F., Eggleton, P., Hammond, P. M., Hodda, M., Holt, R. D., Larsen, T. B., Mawdsley, N. A., Stork, N. E., Srivastava, D. S. and Watt, A. D. (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, **391**, 72-76.

Lopa, D., Mwanyoka, I., Jambiya, G., Massoud, T., Harrison, P., Ellis-Jones, M., Blomley, T., Leimona, B., van Noordwijk, M. and Burgess, N. D. (2012) Towards operational payments for water ecosystem services in Tanzania: a case study from the Uluguru Mountains. *Oryx*, **46**, 34-44.

Lovett, J. C. (1990) Classification and status of the moist forests of Tanzania. *Notifications from the Institute for General Botany Hamburg*, **23a**, 287–300.

Lovett, J. C. and Wasser, S. K. (1993) *Biogeography and Ecology of the Rainforests of Eastern Africa*, Cambridge University Press, Cambridge, UK.

Maggini, R., Lehmann, A., Zimmermann, N. E. and Guisan, A. (2006) Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography*, **33**, 1729-1749.

Marchant, R., Mumbi, C., Behera, S. and Yamagata, T. (2007) The Indian Ocean dipole - the unsung driver of climatic variability in East Africa. *African Journal of Ecology*, **45**, 4-16.

Marshall, A. R., Platts, P. J., Gereau, R. E., Kindeketa, W., Kang'ethe, S. and Marchant, R. (2012) The genus *Acacia* (Fabaceae) in East Africa: distribution, diversity and the protected area network. *Plant Ecology and Evolution*, in press.

McPherson, J. M., Jetz, W. and Rogers, D. J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811-823.

Menon, S., Choudhury, B. I., Khan, M. L. and Peterson, A. T. (2010) Ecological niche modeling and local knowledge predict new populations of Gymnocladus assamicus a critically endangered tree species. *Endangered Species Research*, **11**, 175-181.

Messerli, B., Viviroli, D. and Weingartner, R. (2004) Mountains of the world: Vulnerable Water Towers for the 21st century. *Ambio*, **13**, 29-34.

Meybeck, M., Green, P. and Vorosmarty, C. (2001) A new typology for mountains and other relief classes: An application to global continental water resources and population distribution. *Mountain Research and Development*, **21**, 34-45.

Meynard, C. N. and Quinn, J. F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**, 1455-1469.

Miller, J., Franklin, J. and Aspinall, R. (2007) Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, **202**, 225-242.

Mulligan, M. (2006) MODIS MOD35 pan-tropical cloud climatology. Version 1 (http://www.ambiotek.com/clouds).

Mulligan, M. and Burke, S. M. (2005) FIESTA: Fog Interception for the Enhancement of Streamflow in Tropical Areas (http://www.ambiotek.com/fiesta/)

Myers, N. (1990) The biodiversity challenge: expanded hot-spots analysis. *Environmentalist*, **10**, 243-256.

Naidoo, R., Balmford, A., Costanza, R., Fisher, B., Green, R. E., Lehner, B., Malcolm, T. R. and 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.

Naidoo, R. and Ricketts, T. H. (2006) Mapping the economic costs and benefits of conservation. *Plos Biology*, **4**, 2153-2164.

NBS (2002) *Tanzania Population and Housing Census 2002*. National Bureau of Statistics, Dar es Salaam, The United Republic of Tanzania.

OED (1989) *Oxford English Dictionary* 2nd Ed. Oxford University Press, Oxford, UK.

Phillips, S. J., Dudik, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. and Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181-197.

Platts, P. J., Burgess, N. D., Gereau, R. E., Lovett, J. C., Marshall, A. R., McClean, C. J., Pellikka, P. K. E., Swetnam, R. D. and Marchant, R. (2011) Delimiting tropical mountain ecoregions for conservation. *Environmental Conservation*, **38**, 312-324.

Quesada, C. A., Lloyd, J., Schwarz, M., Baker, T. R., Phillips, O. L., Patiño, S., Czimczik, C., Hodnett, M. G., Herrera, R., Arneth, A., Lloyd, G., Malhi, Y., Dezzeo, N., Luizão, F. J., Santos, A. J. B., Schmerler, J., Arroyo, L., Silveira, M., Priante Filho, N., Jimenez, E. M., Paiva, R., Vieira, I., Neill, D. A., Silva, N., Peñuela, M. C., Monteagudo, A., Vásquez, R., Prieto, A., Rudas, A., Almeida, S., Higuchi, N., Lezama, A. T., López-González, G., Peacock, J., Fyllas, N. M., Alvarez Dávila, E., Erwin, T., di Fiore, A., Chao, K. J., Honorio, E., Killeen, T., Peña Cruz, A., Pitman, N., Núñez Vargas, P., Salomão, R., Terborgh, J. and Ramírez, H. (2009) Regional and large-scale patterns in Amazon forest structure and function are mediated by variations in soil physical and chemical properties. *Biogeosciences Discussions*, **6**, 3993–4057.

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

Soberon, J. and Nakamura, M. (2009) Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19644-19650.

Spawls, S., Howell, K., Drewes, R. and Ashe, J. (2004) *A Field Guide to the Reptiles of East Africa*, Christopher Helm Publishers Ltd, London, UK.

Stephenson, N. L. (1990) Climatic control of vegetation distribution - the role of the water-balance. *American Naturalist*, **135**, 649-670.

UNEP-WCMC (2010) *The World Database on Protected Areas*, UNEP-WCMC, Cambridge, UK (http://www.wdpa.org/).

Venter, O., Meijaard, E., Possingham, H., Dennis, R., Sheil, D., Wich, S., Hovani, L. and Wilson, K. (2009) Carbon payments as a safeguard for threatened tropical mammals. *Conservation Letters*, **2**, 123-129.

Willcock, S., Phillips, O. L., Platts, P. J., Burgess, N. D., Balmford, A., Lovett, J. C., Ahrends, A., Bayliss, J., Doggart, N., Doody, K., Fanning, E., Green, J., Hall, J., Howell, K. L., Marchant, R., Marshall, A. R., Mbilinyi, B., Munishi, P. K. T., Owen, N., Pfeifer, M., Shirima, D., Swetnam, R. D., Topp-Jorgensen, E. J. and Lewis, S. L. (in prep.) The drivers of carbon storage and sequestration within a tropical biodiversity hotspot.

Wisz, M. S. and Guisan, A. (2009) Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecology*, **9**, 8.

Zaniewski, A. E., Lehmann, A. and Overton, J. M. C. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**, 261-280.

Zurell, D. (2011) *Integrating dynamic and statistical modelling approaches in order to improve predictions for scenarios of environmental change*, PhD Thesis, University of Potsdam, Germany.

Thesis Appendix

Appendix I. Co-authored paper abstracts

Journal articles, further to those presented for examination in the main thesis chapters, to which I contributed during my PhD studies. Each builds upon or has relevance to the PhD research. Abstracts are presented in reverse chronological order. Full text versions are available from the respective journal websites.

Research article – The genus* Acacia *(Fabaceae) in East Africa: distribution, diversity and the protected area network

Plant Ecology and Evolution (in press)

Andrew R. Marshall[†], *Philip J. Platts*[†], Roy E. Gereau, William Kindeketa, Simon Kang'ethe, Rob Marchant. [†]Equal contribution

Background and aims. Plants are often overlooked in conservation planning, yet they are the foundation of all terrestrial ecosystems. The East Africa region is used to investigate the effectiveness of protected areas for conserving plants. With a wide range of ecosystems and 771 protected areas covering nearly one quarter of the land area, East Africa is an ideal location to assess the effectiveness of protected areas through distribution modelling of the genus *Acacia*.

Methods. Herbarium specimen data (2,047 records) were collated from East Africa for 65 taxa (species, subspecies, varieties) from the genus *Acacia*. Generalised Additive Models were used to determine climatic drivers, and thence to extrapolate climatic suitability across the region. For two *Acacia* taxa, we investigated the potential for climate-induced range-shifts using a downscaled regional climate model under two IPCC scenarios.

Key results. Over half of *Acacia* diversity hotspots had <10% coverage by protected areas. Furthermore, the protected area network covered less of the predicted ranges of the *Acacia* taxa and contained fewer taxa per unit area than would be expected under randomised placement. Areas with suitable climate for high-elevation, moisture-dependent taxa such as *Acacia abyssinica* subsp. *calophylla* are predicted to contract their potential range by up to 80% toward mountain peaks, where protected areas are dominated by low-level protection forest reserves. Conversely the area of suitable environment for a xerophytic low-elevation species (*A. turnbulliana*) is predicted to increase by up to 77%.

Conclusions. East Africa's national parks may not be preserving an important component of ecosystem diversity; a situation exacerbated by climate change. Even within the genus *Acacia*, different plant species are predicted to respond differently to climate change. Priority areas for research and conservation are identified based on overlap between predicted high *Acacia* diversity and gaps in the collection record, with northern and eastern Kenya highlighted as particularly important. High elevation protected areas are also predicted to become increasingly important as climatic refugia in a warmer future.

Research article – Measuring and modelling above-ground carbon and tree allometry along a tropical elevation gradient

Biological Conservation (in press)
doi: 10.1016/j.biocon.2012.03.017

Andrew R. Marshall, Simon Willcock, *Philip J. Platts*, Jon C. Lovett, Andrew Balmford, Neil D. Burgess, Julia Latham, Pantaleo K. Munishi, Salter, R., Deo D. Shirima, Simon L. Lewis

Abstract. Emerging international policy aimed at reducing carbon emissions from deforestation and forest degradation (REDD+) in developing countries, has resulted in numerous studies on above-ground live carbon (AGC) in tropical forests. However, few studies have addressed the relative importance of disturbance, topography, climate, soil and methods for stem measurement, on the estimation of AGC. We established 18 one ha plots containing 7202 stems, stratified along forested elevation gradients in Tanzania. We recorded a broad set of physical, climatic and edaphic predictors of AGC and tree stature. AGC estimates using dbh, height and wood density, gave a mean value of 174.6t ha^{-1}, compared with 229.6t ha^{-1} when height was excluded. Regression models revealed that stems were tallest per unit diameter (dbh) at mid-elevation (1000–1250m), on south-facing slopes, and without past logging. High AGC was strongly associated with shallow slopes, followed by intermediate elevation, elephant absence, low potential evapotranspiration and low soil pH. Structural variables explained more AGC variation than environmental variables (93% *vs*. 70%), operating through stem size, stature and number, rather than wood density per species. Large stems (≥ 70 cm dbh; 4.6% of stems) contained 52% of AGC in all plots, declining to 36% in lowland plots. We discuss the cost:benefit of different measurements and recommend a tiered approach to AGC monitoring, depending on available resources. AGC assessments in African forests could exclude small stems, but should aim to record disturbance, topography and species. Stem height is vital for AGC estimation and valuation; when excluding height our 55t ha^{-1} over-estimation of AGC would have over-valued the carbon resource by 24% (mean US$3300 ha^{-1}).

Research article – Protected areas: mixed success in conserving East Africa's evergreen forests

PloS ONE (June 2012)
Volume 7, Issue 6, Article e39337, doi: 10.1371/journal.pone.0039337

Marion Pfeifer, Neil D. Burgess, Ruth D. Swetnam, *Philip J. Platts*, Simon Willcock, Rob Marchant

Abstract. In East Africa, human population growth and demands for natural resources cause forest loss contributing to increased carbon emissions and reduced biodiversity. Protected Areas (PAs) are intended to conserve habitats and species. Variability in PA effectiveness and 'leakage' (here defined as displacement of deforestation) may lead to different trends in forest loss within, and adjacent to, existing PAs. Here, we quantify spatial variation in trends of evergreen forest coverage in East Africa between 2001 and 2009, and test for correlations with forest accessibility and environmental drivers. We investigate PA effectiveness at local, landscape and national scales, comparing rates of deforestation within park boundaries with those detected in park buffer zones and in unprotected land more generally. Background forest loss (BFL) was estimated at -9.3% (17,167 km^2), but varied between countries (range: -0.9% to -85.7%; note: no BFL in South Sudan). We document high variability in PA effectiveness within and between PA categories. The most successful PAs were National Parks, although only 26 out of 48 parks increased or maintained their forest area (i.e. Effective Parks). Forest Reserves (Ineffective Parks, i.e. parks that lose forest from within boundaries: 204 out of 337), Nature Reserves (six out of 12) and Game Parks (24 out of 26) were more likely to lose forest cover. Forest loss in buffer zones around PAs exceeded background forest loss, in some areas indicating leakage driven by Effective National Parks. Human pressure, forest accessibility, protection status, distance to fires and long-term annual rainfall were highly significant drivers of forest loss in East Africa. Some of these factors can be addressed by adjusting park management. However, addressing close links between livelihoods, natural capital and poverty remains a fundamental challenge in East Africa's forest conservation efforts.

Research article – Funding begets biodiversity

Diversity and Distributions (March 2011)
Volume 17, Issue 2, Pages 191-200, doi: 10.1111/j.1472-4642.2010.00737.x

Antje Ahrends, Neil D. Burgess, Roy E. Gereau, Rob Marchant, Mark T. Bulling, Jon C. Lovett, *Philip J. Platts*, Victoria Wilkins Kindemba, Nisha Owen, Eibleis Fanning, Carsten Rahbek

Aim. Effective conservation of biodiversity relies on an unbiased knowledge of its distribution. Conservation priority assessments are typically based on the 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. Here, we test whether there is circularity between funding and perceived biodiversity, which may reinforce the conservation status of areas already perceived to be important while other areas with less initial funding may remain overlooked.

Location. Eastern Arc Mountains, Tanzania.

Methods. We analysed time series data (1980-2007) of funding (n = 134 projects) and plant species records (n = 75,631) from a newly compiled database. Perceived plant diversity, over three decades, is regressed against funding and environmental factors, and variances decomposed in partial regressions. Cross-correlations are used to assess whether perceived biodiversity drives funding or vice versa.

Results. Funding explained 65% of variation in perceived biodiversity patterns – six times more variation than accounted for by 34 candidate environmental factors. Cross-correlation analysis showed that funding is likely to be driving conservation priorities and not vice versa. It was also apparent that investment itself might trigger further investments as a result of reduced start-up costs for new projects in areas where infrastructure already exists. It is therefore difficult to establish whether funding, perceived biodiversity, or both drive further funding. However, in all cases, the results suggest that regional assessments of biodiversity conservation importance may be biased by investment. Funding effects might also confound studies on mechanisms of species richness patterns.

Main conclusions. Continued biodiversity loss commands urgent conservation action even if our knowledge of its whereabouts is incomplete; however, by concentrating inventory funds in areas already perceived as important in terms of biodiversity and/or where start-up costs are lower, we risk losing other areas of underestimated or unknown value.

Research article – Conservation and the botanist effect

Biological Conservation (January 2011)
Volume 144, Issue 1, Pages 131-140, doi: 10.1016/j.biocon.2010.08.008

Antje Ahrends, Carsten Rahbek, Mark T. Bulling, Neil D. Burgess, *Philip J. Platts*, Jon C. Lovett, Victoria Wilkins Kindemba, Nisha Owen, Albert Ntemi Sallu, Andrew R. Marshall, Boniface E. Mhoro, Eibleis Fanning, Rob Marchant

Abstract. Over the last few decades, resources for descriptive taxonomy and biodiversity inventories have substantially declined, and they are also globally unequally distributed. This could result in an overall decline in the quality of biodiversity data as well as geographic biases, reducing the utility and reliability of inventories. We tested this hypothesis with tropical tree records (n = 24,024) collected from the Eastern Arc Mountains, Tanzania, between 1980 and 2007 by 13 botanists, whose collections represent 80% of the total plant records for this region. Our results show that botanists with practical training in tropical plant identification record both more species and more species of conservation concern (20 more species, two more endemic and one more threatened species per 250 specimens) than untrained botanists. Training and the number of person-days in the field explained 96% of the variation in the numbers of species found, and training was the most important predictor for explaining recorded numbers of threatened and endemic species. Data quality was related to available facilities, with good herbarium access significantly reducing the proportions of misidentifications and misspellings. Our analysis suggests that it may be necessary to account for recorder training when comparing diversity across sites, particularly when assessing 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 and in the provisioning of good facilities would substantially increase recording efficiency and data reliability, thereby improving conservation planning and implementation on the ground.

Discussion piece – Palaeoenvironmental perspectives for sustainable development in East Africa

Climate of the Past Discussion (May 2010)
Volume 6, Issue 3, Pages 963-1007, doi: 10.5194/cpd-6-963-2010

Rob Marchant, Jemma Finch, Rahab N. Kinyanjui, Veronica M. Muiruri, Cassian T. Mumbi, *Philip J. Platts*, Stephen M. Rucina

Abstract. East African ecosystems are shaped by long-term interaction with changing climate, human population, fire and wildlife. There remains today a strong connection between people and ecosystems, a relationship that is being strained by the rapidly developing and growing East African population, and their associated resource needs. Predicted climatic and atmospheric change will further impact on ecosystems culminating in a host of challenges for their management and sustainable development, further compounded by a backdrop of political, land tenure and economic constraints. Given the many direct and indirect benefits that ecosystems provide to surrounding human populations, understanding how they have changed over time and space deserves a special place on the ecosystem management agenda. Such a perspective can be derived from palaeoecology, particularly where there are high resolution data through time and across space. The East African palaeoecological archive is reviewed, in particular to assess how it can meet this need. Although there remain crucial gaps, the number of palaeoecological archives from East Africa growing rapidly, some employing new and novel techniques to trace past ecosystem response to climate change. When compared to the archaeological record it is possible to disentangle human from climate change impacts, and how the former interacts with major environmental changes such as increased use of fire, changing herbivore densities and increased atmospheric CO_2 concentration. With this multi-dimensional perspective of environmental change impacts it is imperative that our understanding of past human-ecosystem interactions are considered to impart effective long term management strategies; such an approach will enhance possibilities for a sustainable future for East African ecosystems and maximise the livelihoods of the populations that rely on them.

Research article – The species-area relationship and confounding variables in a threatened monkey community

American Journal of Primatology (April 2010)
Volume 72, Issue 4, Pages 325-336, doi: 10.1002/ajp.20787

Andrew R. Marshall, Helle I. O. Jorgensbye, Francesco Rovero, *Philip J. Platts*, Piran C. L. White, Jon C. Lovett

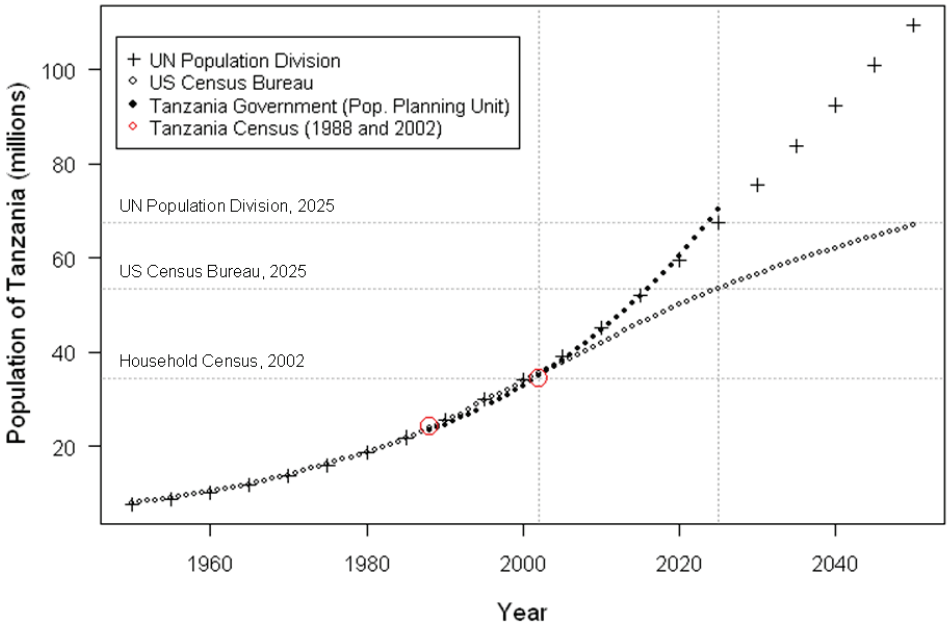
Abstract. This study investigates the species-area relationship (SAR) for forest monkeys in a biodiversity hotspot. The Udzungwa Mountains of Tanzania are well-suited to investigate the SAR, with seven monkey species in a range of fragment sizes (0.06-526 km^2). We test the relationship between species richness and forest fragment size, relative to human and environmental factors. We distinguish resident and transitory species because the latter have an "effective patch size" beyond the area of forest. Forest area was the strongest (log-linear) predictor of species richness. However, forest area, elevation range and annual moisture index were inter-correlated. Previous knowledge of the relationship between elevation and tree communities suggests that the SAR is largely a result of habitat heterogeneity. Isolation by farmland (matrix habitat) also had a significant negative effect on species richness, probably exacerbated by hunting in small forests. The effect of area and isolation was less for transitory species. The human influence on species' presence/absence was negatively related to the extent of occurrence. Weaker relationships with temperature and precipitation suggest underlying climatic influences, and give some support for the influence of productivity. A reduced area relationship for smaller forests suggests that fragment sizes below 12-40 km^2 may not be reliable for determining SAR in forest monkeys. Further practical implications are for management to encourage connectivity, and for future SAR research to consider residency, matrix classification and moisture besides precipitation.

Appendix II. Population projections for Tanzania

Appendix 2C describes a method for refining LandScan population estimates for consistency with Tanzanian census counts (NBS, 2002) and strictly enforced protected area governance (no people in games reserves or national parks). Here, I further develop spatially explicit projections for the Tanzanian population for the year 2025, based on spatial variation in subnational trends and two scenarios of national growth.

National growth

The Government of Tanzania predicts a fixed annual growth rate of 3.1 %.y^{-1}, leading to a Tanzanian population of 70 million by 2025 (http://www.tanzania.go.tz/populationf.html/). Other sources have explicitly modelled changes in the national growth rate over time, based upon assumptions of fertility, mortality and international migration. The Population Division of the United Nations (Department of Economic and Social Affairs) has projected a Tanzanian population of 67 million by 2025 (http://www.un.org/esa/population/unpop.htm), which is equivalent to fixed annual growth of *c*. 3%.y^{-1}. The United States Census Bureau projects a 2025 population of 53 million (http://www.census.gov/ipc/www/idb/), equivalent to fixed annual growth of *c*. 2%.y^{-1}.

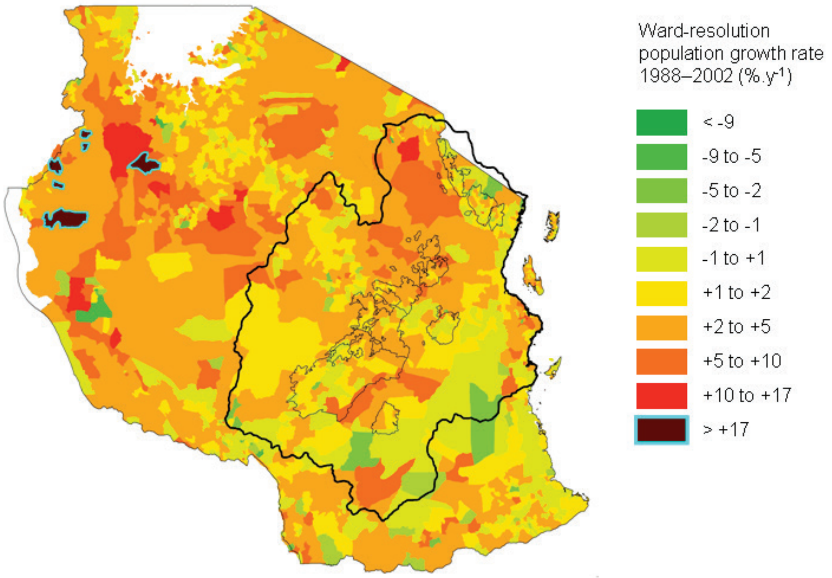


Spatial variation in ward rates

At the national scale, scenarios of moderate and rapid growth were based on the projections of the US Census Bureau and the United Nations Population Division, respectively. At sub-national scales, population trends were derived from anomalies in ward-level census data, comparing counts compiled in 1988 with those compiled in 2002. In 1988, the population of Tanzania was 24.1 million, increasing to 34.4 million by 2002. The national growth rate, *R*, over this 14 year period is given by:

$$R = 100 * \left(\left(\frac{P_{2002}}{P_{1988}} \right)^{1/14} - 1 \right) = 2.5782\ \%.y^{-1}$$

Similarly, one can derive ward-specific growth rates for this period, as mapped below. Six outliers were identified (highlighted in cyan), situated to rural parts of western Tanzania. In these wards, unusually high growth rates of up to 31%.y^{-1} were inferred from differences in census counts. Possible explanations include under- and/or over-enumeration during the census years, or genuine migrations into previously unpopulated areas, possibly a consequence of social conflict in Rwanda and the Democratic Republic of the Congo. The extrapolation of such high compound growth rates into the future would be unrealistic. Instead, these areas were allotted the mean rates from adjacent wards.



Population growth rates in Tanzania, 1988-2002. Outliers highlighted in cyan were allotted the mean rates from adjacent wards. Also shown are the EAMs (thin line) and their watershed basin (bold line).

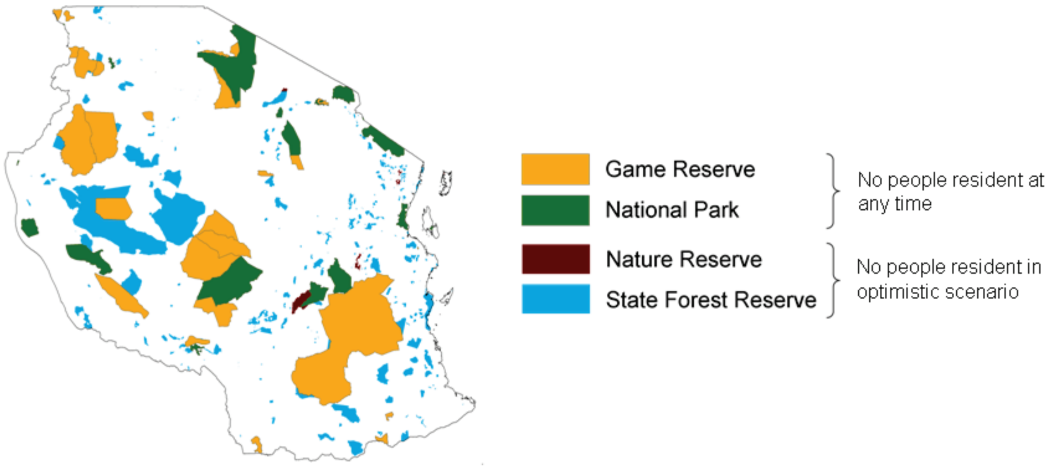
Between 1988 and 2002, populations increased in most wards (2520 of 2806), whilst decreasing (273) or remaining static (13) in the others. For scenarios of 2025 population, the relative differences between ward-level growth rates were maintained. The magnitude of change, however, was adjusted by a common factor, *a*, such that the national rates of 2% or 3% be satisfied:

$$P_{2002,TZ}\left(1+R\right)^{23} = \sum_{i=1}^{2806}\left(P_{2002,i}\left(1+ar_i\right)^{23}\right)$$

where $P_{2002,\ TZ}$ is the population of Tanzania in 2002, $P_{2002,\ i}$ is the population of ward *i* in 2002, and r_i is the annual growth rate of ward *i* between 1988 and 2002 (with outliers corrected). In the moderate growth scenario, ward rates were scaled such that R = 2%.y^{-1} (a = 0.59). For rapid growth, R = 3%.y^{-1} (a = 0.83). High-resolution (1 km) population grids for the year 2025 were then obtained by substituting, in the procedure described in Appendix 2C, baseline (2002) ward populations for projected (2025) ward populations. Thus, growth varied *between* wards according observed trends, but the distribution of persons *within* wards remained static through time (but see below).

Stronger governance

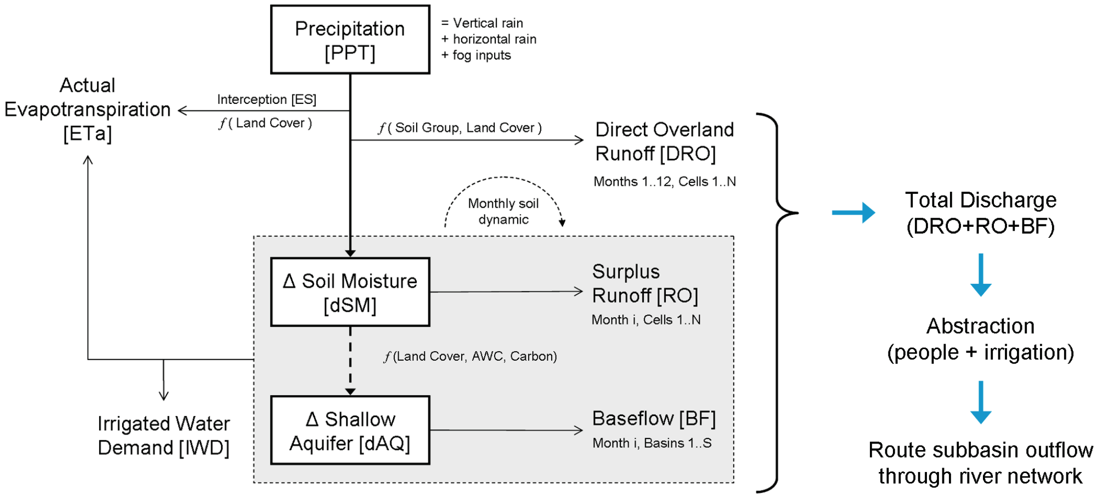
Under the moderate growth scenario, forest governance was assumed to be stronger than at present. Accordingly, the small numbers of people identified as currently residing in nature reserves and central government forest reserves were displaced to other locations in the corresponding wards, prior to the implementation of growth algorithms.



Appendix III. Hydrological modelling

Model structure

Table 6.3 in the Summary Discussion ranks protected areas in the EAMs according to the mean annual hydrological runoff generated across 1 km grid cells within those sites. These calculations were based on a simple water balance model, driven by TRMM rainfall data (Mulligan, 2006). The model was modified from a framework described by Alemaw and Chaoka (2003). To better understand how forest clearance in the mountains might impact seasonal flow in lowland areas, I extended the model to incorporate shallow aquifer and baseflow components.



Schematic of the hydrological model used in Chapter 6

The model was implemented across the Tanzanian watershed of the EAMs, using a monthly time-step at 1 km spatial resolution. Ground water was pooled in large subbasins (median 1660 km^2, r.watershed tool in GRASS GIS) and discharged to the stream network more gradually than overland flow. Monthly rainfall grids were as described in Chapters 4-5 (TRMM 1997-2006). Land cover data were based on a survey by Hunting Technical Services (1997), corrected and updated by local experts as part of the VTA Programme. Quarterly leaf area indices (LAI) for each land cover type were derived from MODIS 8-day composites, and adjusted upward by a factor of 1.5 to better match ground-truthed estimates in the EAMs (Pfeifer *et al.*, 2012). Potential evapotranspiration (PET) was derived from

WorldClim (1950-2000) using the Hargreaves method. These data were combined to estimate maximum transpiration rates: PET–ES, or (PET–ES) / 3 if LAI < 3, where ES is the evaporation of intercepted water (Neitsch *et al.*, 2005).

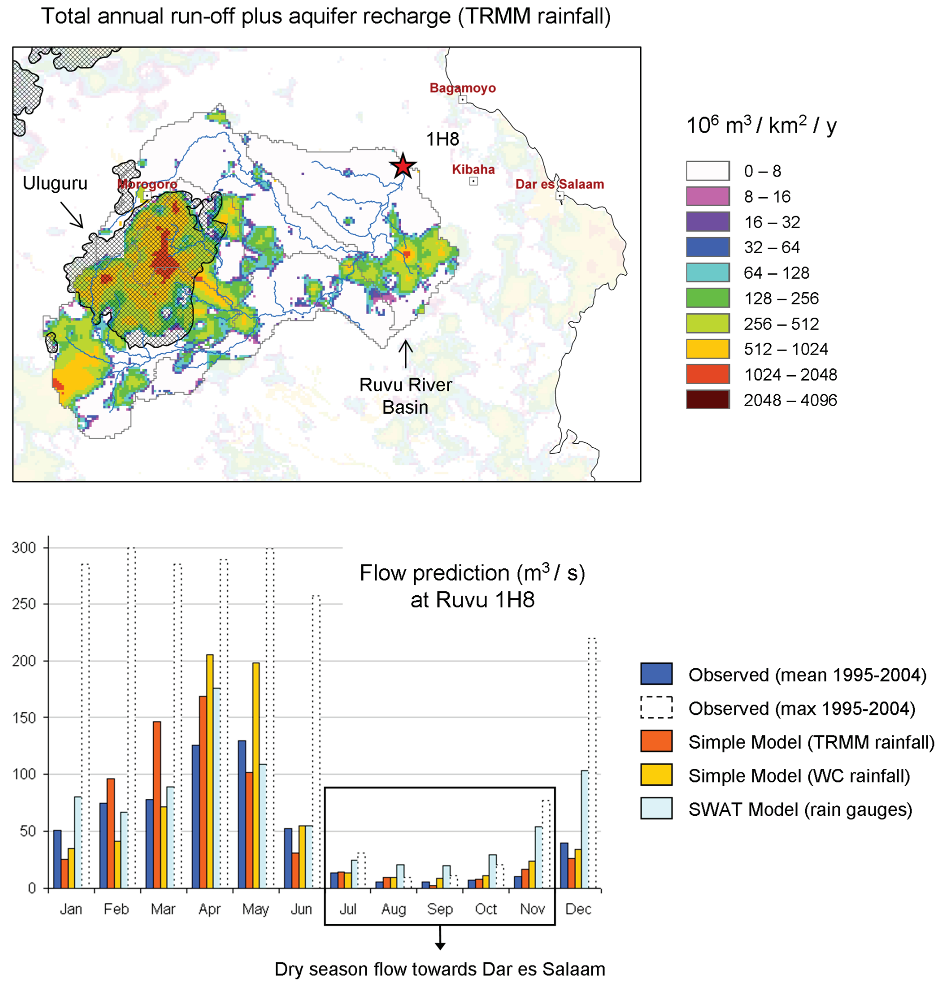
The amount of water available for transpiration in a given month was determined in the first instance by the SCS curve number method, which estimates direct overland runoff (leaves subbasin within one month) as a function of soil type and land cover (SCS, 1972). Soil type was according to SOTER, as used in Chapter 4. If the remainder was insufficient to meet transpiration requirements then plants absorbed moisture from the soil, down to wilting point. Conversely, if transpiration requirements were fulfilled, then any remaining water was allowed to infiltrate the soil column. When soils reached field capacity, surplus water either (1) discharged through the upper layers of the soil column, reaching the stream network within two months, or (2) infiltrated the soil, forcing an equivalent amount to seep through the bottom of the soil column, recharging the shallow aquifer. The ratio of surplus runoff to aquifer recharge was a function of soil carbon (Willcock *et al.*, in review): as organic matter increases, surplus water is held in a cell for longer, and so a higher proportion can seep through the soil. Discharge from the shallow aquifer to the main river channel (baseflow) occurred over a period of several months, according to an exponential decay.

Abstraction

Runoff generated at the grid cell resolution (direct overland runoff plus surplus runoff) was combined with baseflow at the outlet of each subbasin, and subsequently routed through the stream network toward the coast. On route, flow was adjusted to meet the demands of local populations (washing, drinking, cooking, etc.). Per capita daily demand was estimated at 19 litres in rural wards, 60 litres in urban wards, and 34 litres in mixed wards. Applying these volumes to the population grid (Chapter 2), I obtained estimates of total water demand per unit area. Irrigation demand was calculated according to the difference between maximum and actual transpiration at the grid cell resolution. Rice, sugar and vegetables were irrigated as required using water available at the subbasin resolution. Other crops, such as maize and bean, were assumed to be entirely rain-fed.

Calibration

Although a simplistic model compared with more intensively process-based hydrological assessment tools, such as SWAT (Neitsch *et al.*, 2005), the above procedure allowed rapid assessment of hydrological runoff and river flow across the EAMs using readily available data. The model was calibrated in the Ruvu River basin (Uluguru), favouring accuracy in dry season flow at the expense of some over-prediction during the wet season (but within the observed range). Annual flow at gauge 1H8 (mapped below) was captured effectively at 1.7-1.9 × 10^9 $m^3.y^{-1}$ (mean observed, 1995-2004: 1.6 × 10^9 $m^3.y^{-1}$). Modelled seasonal flow was also similar to gauge data: PBIAS = 1.2%, Pearson's $r = 0.87$, Nash-Sutcliffe = 0.62, RSR = 0.5 (Moriasi *et al.*, 2007). Independent testing in the Sigi Basin (East Usambara), which is much smaller and subject to different land use and climate (biannual rains), found that the same model parameters again captured annual flows effectively (1.6-1.7 × 10^8 $m^3.y^{-1}$), but that seasonal flow was more uncertain. For comparison, the figure below includes preliminary model predictions from SWAT (White and Ashagre, unpublished), as well as estimates derived using WorldClim (instead of TRMM) rainfall.

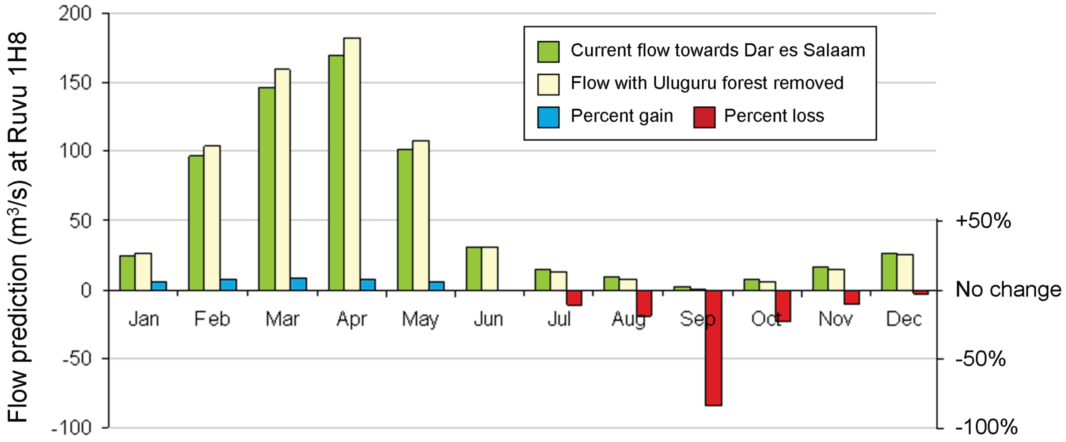


Upper: annual runoff plus aquifer recharge at 1 km resolution, used to rank protected areas in Table 6.3

Lower: modelled *vs.* observed river flow at the outlet of the Ruvu basin, emphasising low flows in the dry season

Scenario of forest loss

The figure below summarises the modelled response of river flow at 1H8 to a scenario of uncontrolled agricultural expansion in Uluguru. This involved switching 'forest' land cover types (lowland through to upper montane) to 'cultivation', with associated changes in LAI and soil carbon. In practice, this scenario is unrealistic because Uluguru Nature Reserve has recently been established and governance is strong. In the absence of conservation, however, uncontrolled encroachment may well have occurred – forests outside the nature reserve boundary (previously forest reserves) have already been cleared.



Forests ecosystems have high transpiration rates compared with cropland. Consequently, forest conversion was predicted to result in a 5% increase in annual flow at 1H8. However, the majority of this additional water is discharged as direct overland runoff, or as surplus runoff through the upper layers of soil, and so reaches the main channel within 1-2 months (i.e., during the wet season). Potential consequences are increased flood risk and reduced hydropower generation (both through increased sedimentation and because turbines are stopped during periods of high flow). Conversely, during the dry season when water shortages are most acute, the model predicts a slight reduction in flow, because less water infiltrates to recharge the aquifer during the wet season, reducing baseflow throughout the year. Whilst the methods employed here are simplistic, they do indicate an important regulatory role of mountain forests in terms of water provision downstream. Further, the patterns predicted here corroborate a widely held belief in Tanzania that deforestation in Uluguru, combined with increased abstraction by agriculture and people, has contributed to reduced dry season flow into Dar es Salaam, Tanzania's largest city (Lopa *et al.*, 2012).

References

Alemaw, B. F. and Chaoka, T. R. (2003) A continental scale water balance model: a GIS-approach for Southern Africa. *Physics and Chemistry of the Earth*, **28**, 957-966.

Hunting-Technical-Services (1997) *National Reconnaissance Level Land Use and Natural Resources Mapping Project*, Final Report to the Ministry of Natural Resources and Tourism, The United Republic of Tanzania.

Lopa, D., Mwanyoka, I., Jambiya, G., Massoud, T., Harrison, P., Ellis-Jones, M., Blomley, T., Leimona, B., van Noordwijk, M. and Burgess, N. D. (2012) Towards operational payments for water ecosystem services in Tanzania: a case study from the Uluguru Mountains. *Oryx*, **46**, 34-44.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. and Veith, T. L. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the Asabe*, **50**, 885-900.

Mulligan, M. (2006) Global gridded 1km TRMM rainfall climatology and derivatives. Version 1 (http://www.ambiotek.com/1kmrainfall).

NBS (2002) Tanzania Population and Housing Census 2002. National Bureau of Statistics, Dar es Salaam, The United Republic of Tanzania.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Srinivasan, R. and Williams, J. R. (2005) *Soil and Water Assessment Tool Theoretical Documentation, version 2005*, Grassland, Soil and Water Research Laboratory, Agricultural Research Service, Texas, USA, pp. 131.

Pfeifer, M., Gonsamo, A., Disney, M., Pellikka, P. and Marchant, R. (2012) Leaf area index for biomes of the Eastern Arc Mountains: Landsat and SPOT observations along precipitation and altitude gradients. *Remote Sensing of Environment*, **118**, 103-115.

SCS (1972) Section 4: Hydrology. In *National Engineering Handbook*, USDA Soil Conservation Service, Washington DC, USA.

Willcock, S., Phillips, O. L., Platts, P. J., Balmford, A., Burgess, N. D., Lovett, J. C., Ahrends, A., Bayliss, J., Doggart, N., Doody, K., Fanning, E., Green, J., Hall, J., Howell, K., Marchant, R., Marshall, A. R., Mbilinyi, B., Munishi, P., Owen, N., Swetnam, R. D., Topp-Jorgensen, E. and Lewis, S. L. (in review) Towards regional, error-bounded landscape carbon storage estimates for data deficient areas of the world. *PLoS ONE*.

Abbreviations

BP	Before Present
°C	Degrees Celsius
DD / °	Decimal Degrees
dbh	Diameter (of a tree) at Breast Height
DEM	Digital Elevation Model
df	Degrees of Freedom (in a model)
EAM	Eastern Arc Mountain chain
eU	East Usambara Mountains (in the Eastern Arc)
FAO	Food and Agriculture Organisation (of the United Nations)
GAM	Generalised Additive Model
GCM	General Circulation Model
GIS	Geographical Information System
GLM	Generalised Linear Model
ha	Hectares
IPCC	Intergovernmental Panel on Climate Change
IUCN	International Union for Conservation of Nature
KITE	York Institute for Tropical Ecosystem Dynamics
km	Kilometres
m	Metres
mm	Millimetres
Mh	Mahenge Mountains (in the Eastern Arc)
Ml	Malundwe Mountain (in the Eastern Arc)
MY	Million Years
NFTP	Non-Timber Forest Product
nP	North Pare Mountains (in the Eastern Arc)
Nr	Nguru Mountains (in the Eastern Arc)
Nu	Nguu Mountains (in the Eastern Arc)
RCM	Regional Climate Model
REDD	Reduced Emissions from Deforestation and Forest Degradation
Ru	Rubeho Mountains (in the Eastern Arc)
SAC	Spatial Autocorrelation
SDM	Species Distribution Model
sP	South Pare Mountains (in the Eastern Arc)
Ta	Taita Hills (in the Eastern Arc)
Ud	Udzungwa Mountains (in the Eastern Arc)
UDSM	University of Dar es Salaam
Uk	Ukaguru Mountains (in the Eastern Arc)
Ul	Uluguru Mountains (in the Eastern Arc)
UNEP	United Nations Environment Programme
UTM	Universal Transverse Mercator (coordinate system)
VTA	Valuing the Arc Programme
WCMC	World Conservation Monitoring Centre
WDPA	World Database on Protected Areas
wU	West Usambara Mountains (in the Eastern Arc)
WWF	World Wildlife Fund
y	Years