This is a repository copy of *Can distribution models help refine inventory-based estimates of conservation priority?:A case study in the Eastern Arc forests of Tanzania and Kenya*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/75267/

Version: Submitted Version

Article:

Platts, Philip J. orcid.org/0000-0002-0153-0121, Ahrends, Antje, Gereau, Roy E. et al. (7 more authors) (2010) Can distribution models help refine inventory-based estimates of conservation priority?:A case study in the Eastern Arc forests of Tanzania and Kenya. Diversity and Distributions. pp. 628-642. ISSN: 1472-4642

https://doi.org/10.1111/j.1472-4642.2010.00668.x

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

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).

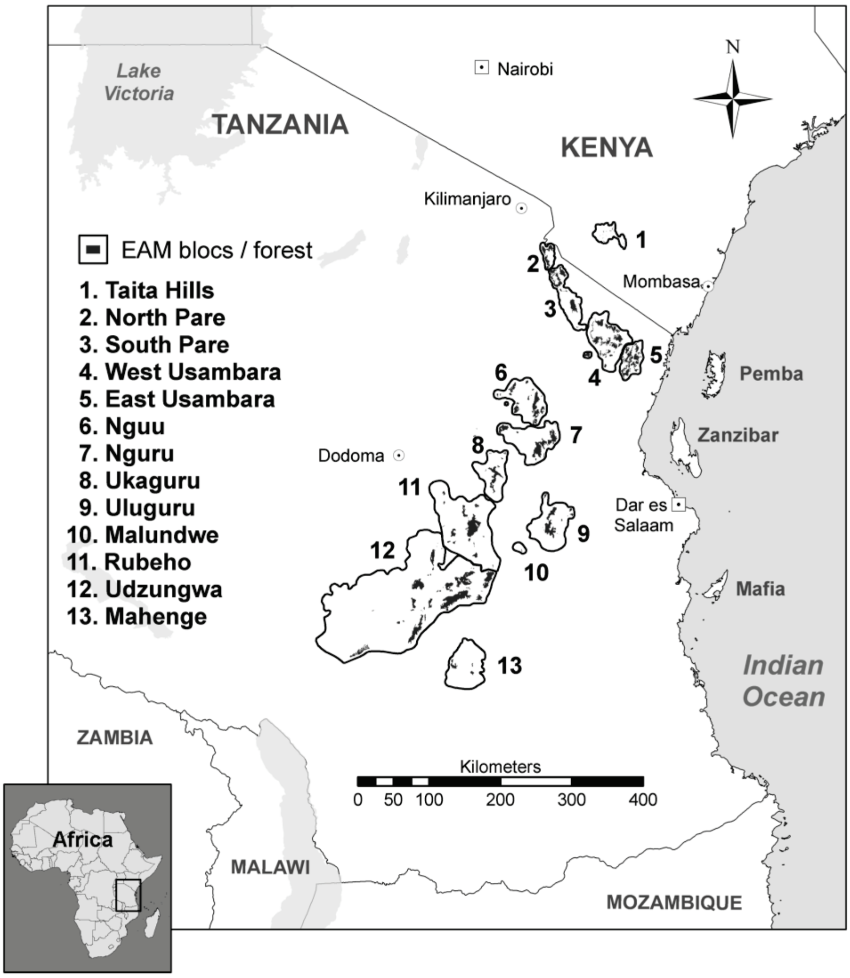


Figure 4.1. Map of the 13 mountain blocs that comprise the Eastern Arc chain, including forest cover at 1-ha resolution (see also Table 4.1). Projection (except inset) is UTM zone 37 south. Note that forest estimates and boundary placement pre-date Chapter 2, although divisions between blocs are consistent – 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 $auto\text{cov}_i$ 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.

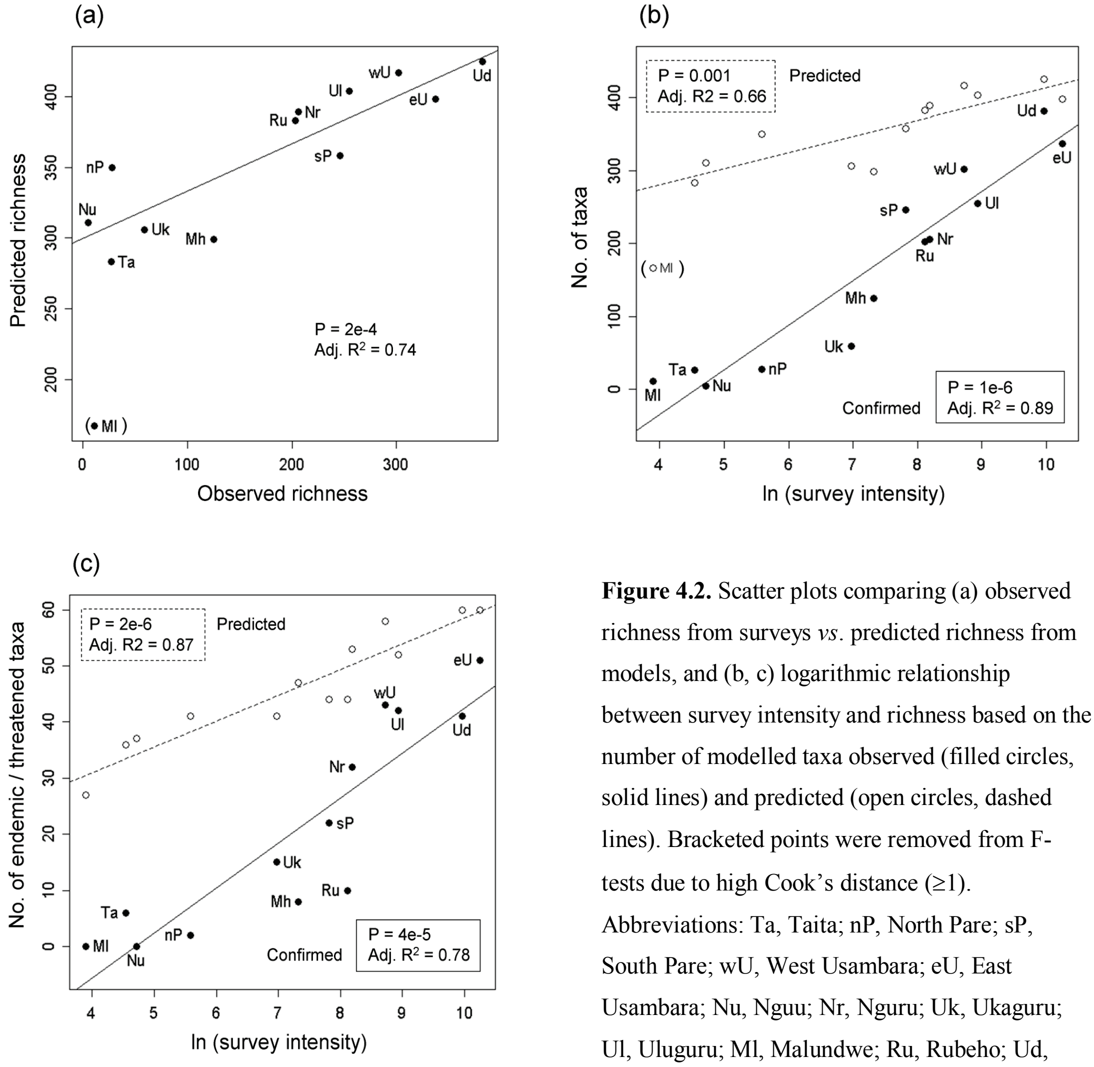


Figure 4.2. Scatter plots comparing (a) observed richness from surveys *vs*. predicted richness from models, and (b, c) logarithmic relationship between survey intensity and richness based on the number of modelled taxa observed (filled circles, solid lines) and predicted (open circles, dashed lines). 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

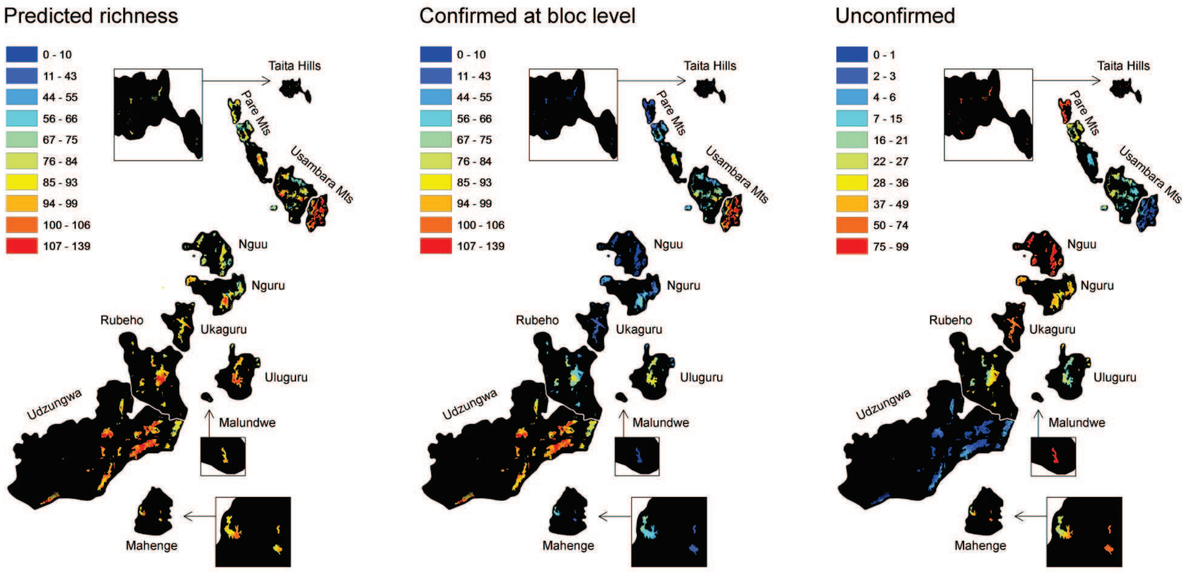
Predicted richness was greater than observed richness in all cases, with the size of the disparity showing a negative correlation with survey intensity (Fig. 4.2b-c). Unconfirmed but potentially suitable habitat was therefore most common in Taita, North Pare, Nguu and Malundwe. Environmental conditions 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

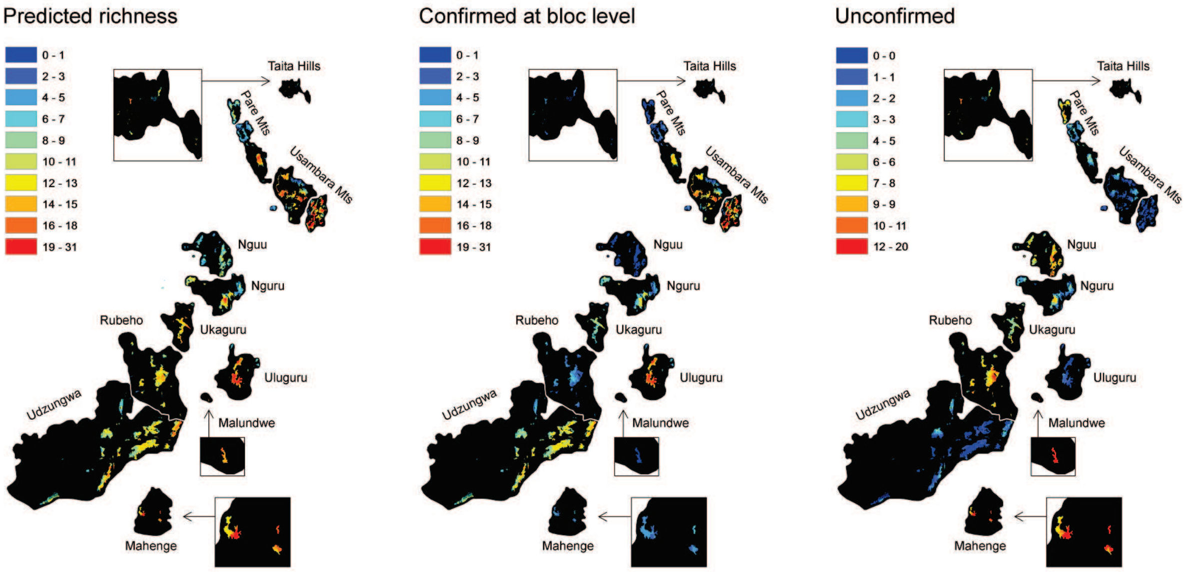


Figure 4.3. Spatial estimates of plant richness calculated across (a) all taxa and (b) taxa of conservation concern. Scale bars show the number of taxa predicted to have potential niche-space in 1 km grid squares. In the left panel, modelled distributions are extrapolated to all forest patches with suitable 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.

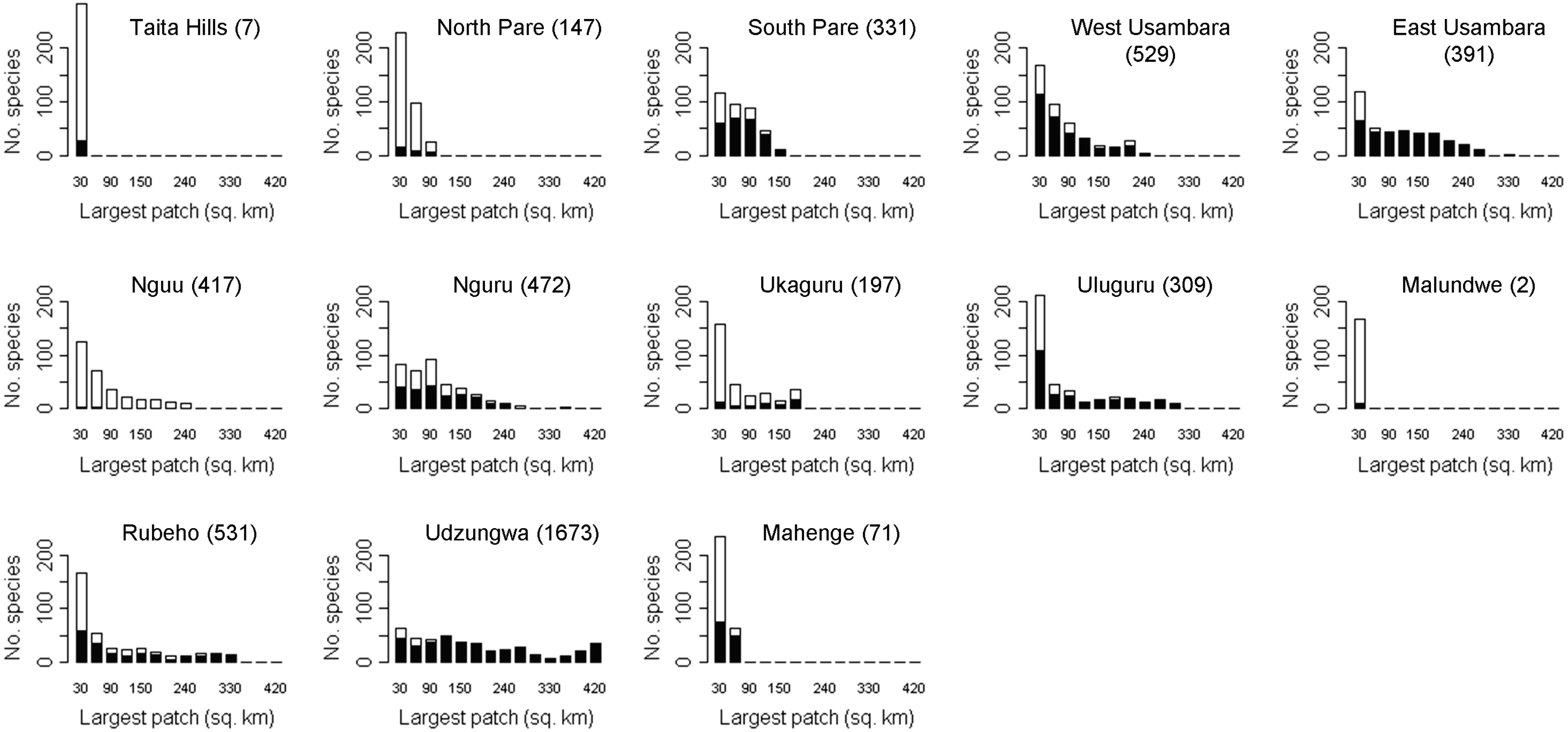


Figure 4.4. Histograms showing patterns of within-bloc patch occupancy. Horizontal axes represent the largest contiguous area of forest providing environmentally suitable conditions for a particular taxon. Vertical axes show the number of taxa in each 30 km^2 patch size interval. Filled bars relate to 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

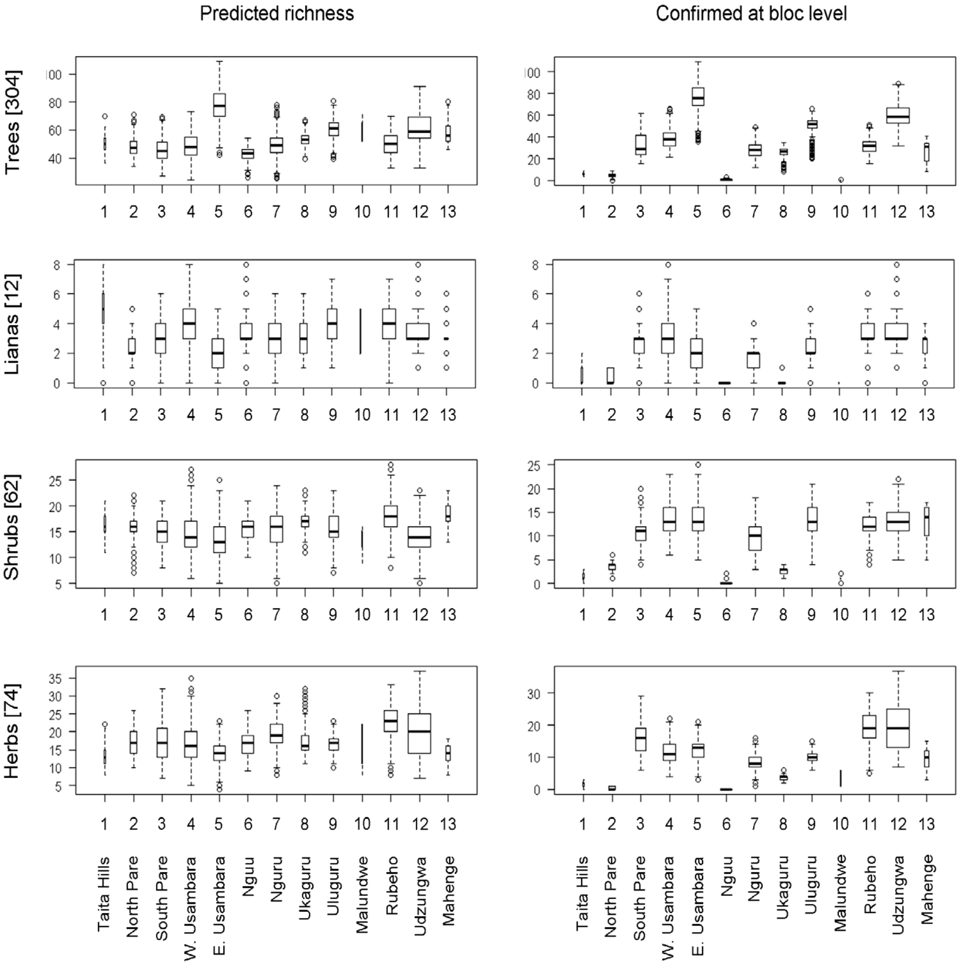


Figure 4.5. Box plots detailing how plant richness varies according to growth form. In the left panel, modelled distributions are extrapolated to all forest patches with suitable environmental conditions. In the right panel, predictions are restricted to just those mountain blocs where the respective 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.

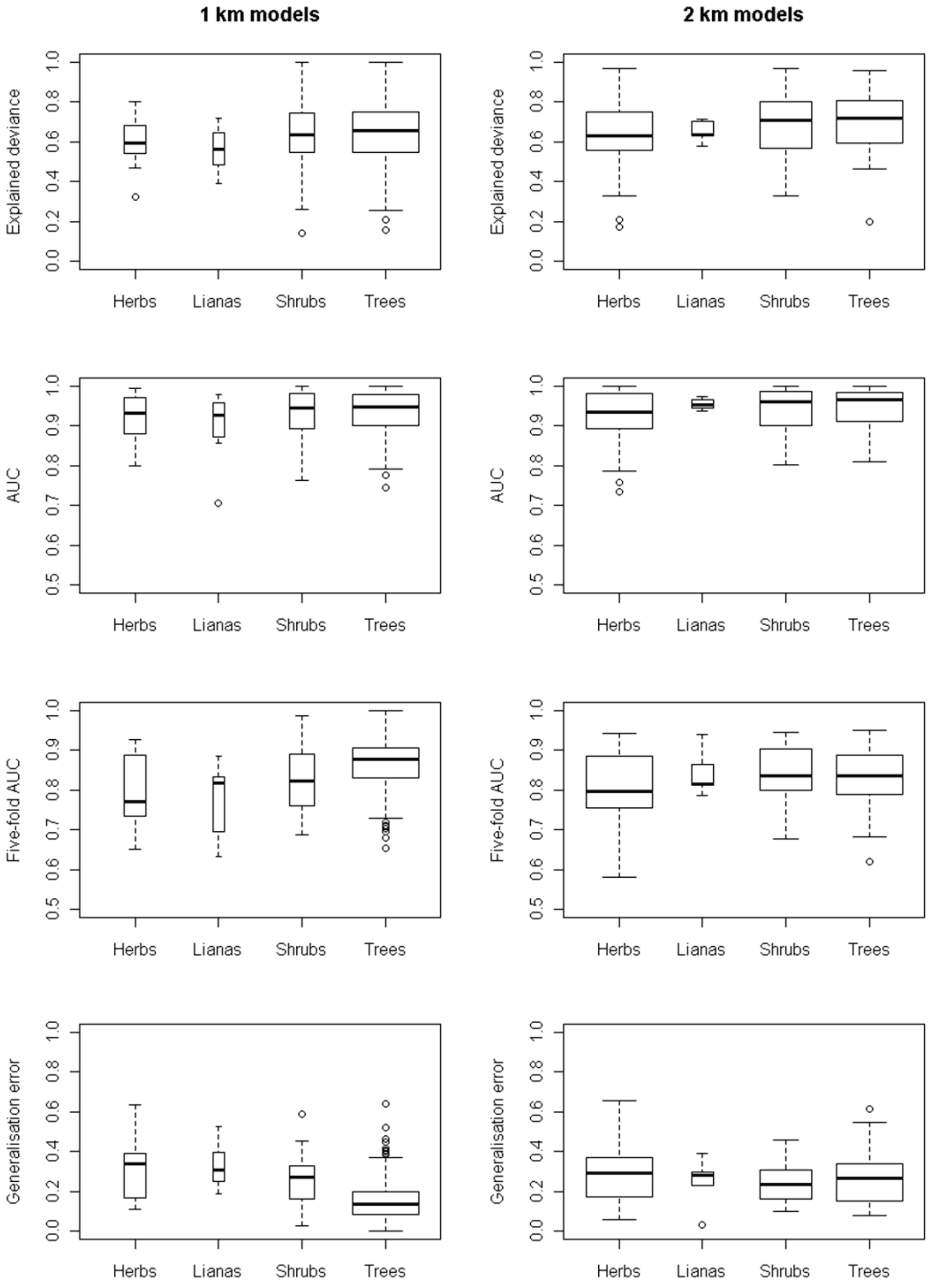
Once calibrated at the chosen prevalence, models predicted occurrence on a continuous scale, from zero to one. Maps of estimated presence-absence were obtained by imposing taxon-specific occurrence thresholds, chosen by maximising the sum of sensitivity and specificity (Cantor *et al.*, 1999). This approach 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 ≤ 0.001 (extremely significant); **, p ≤ 0.01 (highly significant); *, p ≤ 0.05 (significant); ns, not sig.

Growth form

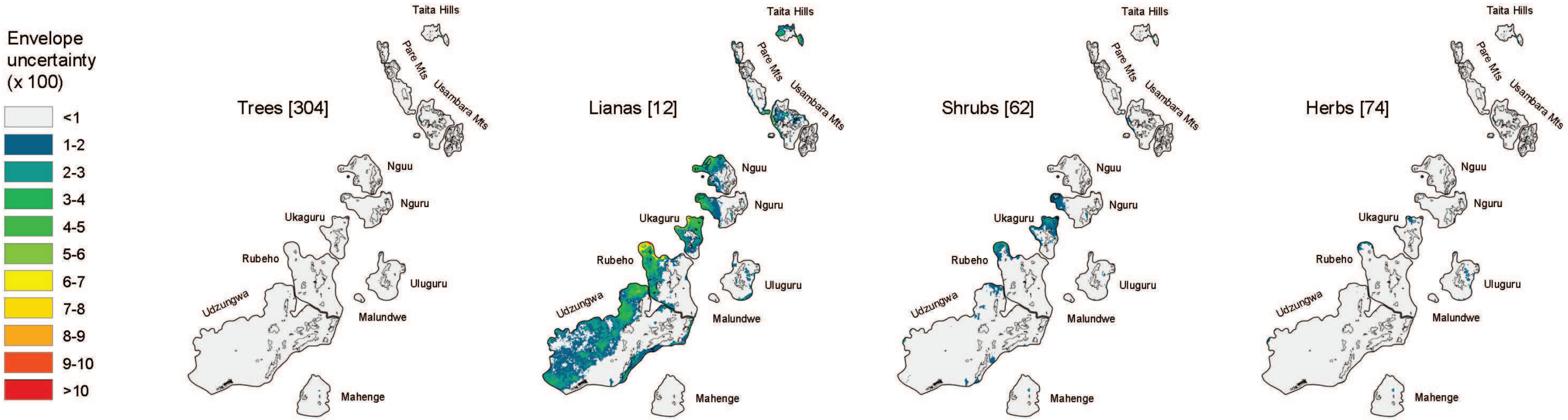
Wilcoxon rank sum tests (one-sided: rows > columns), comparing the performance of models calibrated for the different growth forms of plants (304 trees, 12 lianas, 62 shrubs, 74 herbs).

	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 ≤ 0.001 (extremely significant); **, p ≤ 0.01 (highly significant); *, p ≤ 0.05 (significant); ns, not significant

Envelope uncertainty maps (EUMs)

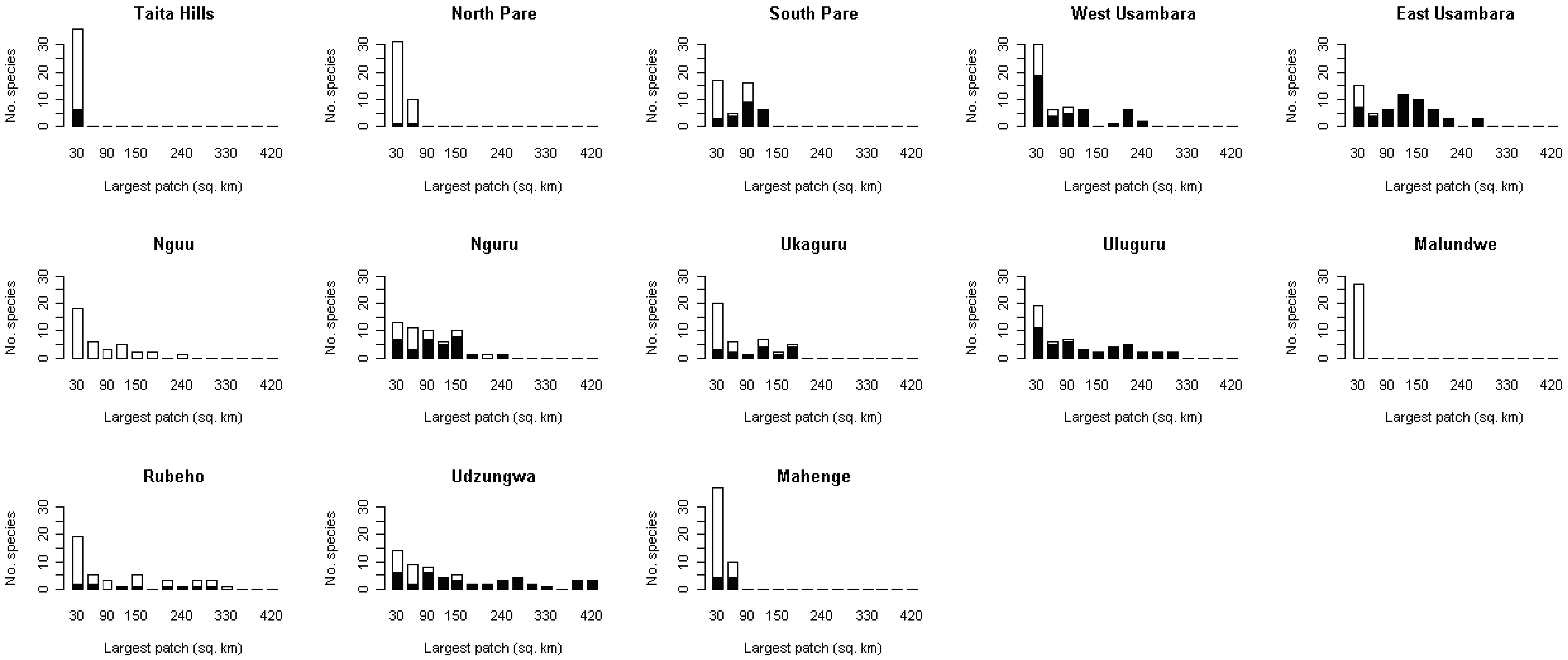
The proportional 'distance' of each grid cell from the calibration envelope was mapped with respect to each environmental predictor. Prediction uncertainty resulting from extrapolation to novel parameter space was estimated using an average of these maps, weighted according to the relative contributions 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

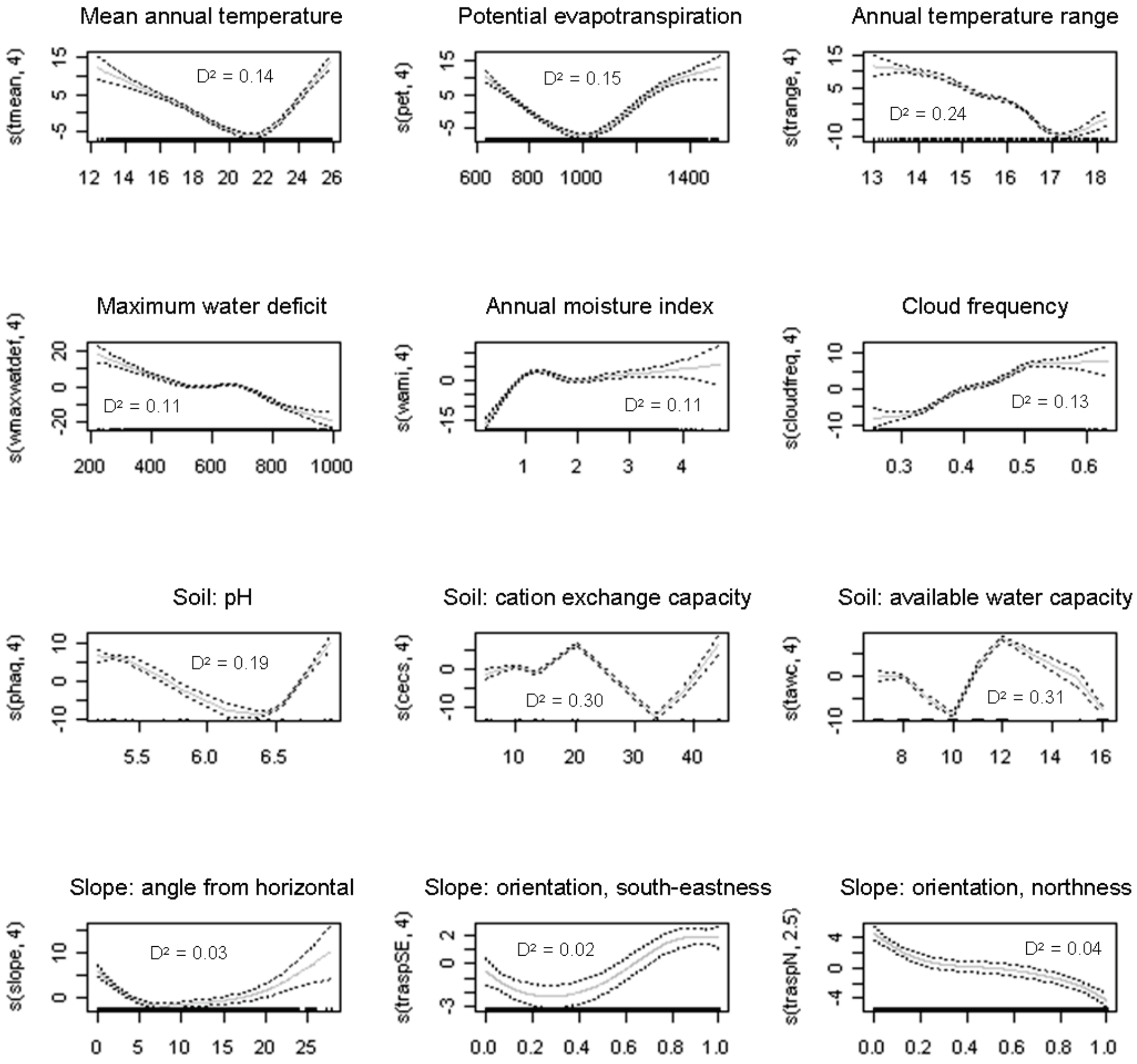
Histograms showing patterns of within-bloc patch occupancy for taxa that are endemic and/or threatened. Horizontal axes represent the largest contiguous area of forest providing environmentally suitable conditions for a particular species. Vertical axes show the number of species in each 30 km^2 patch 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

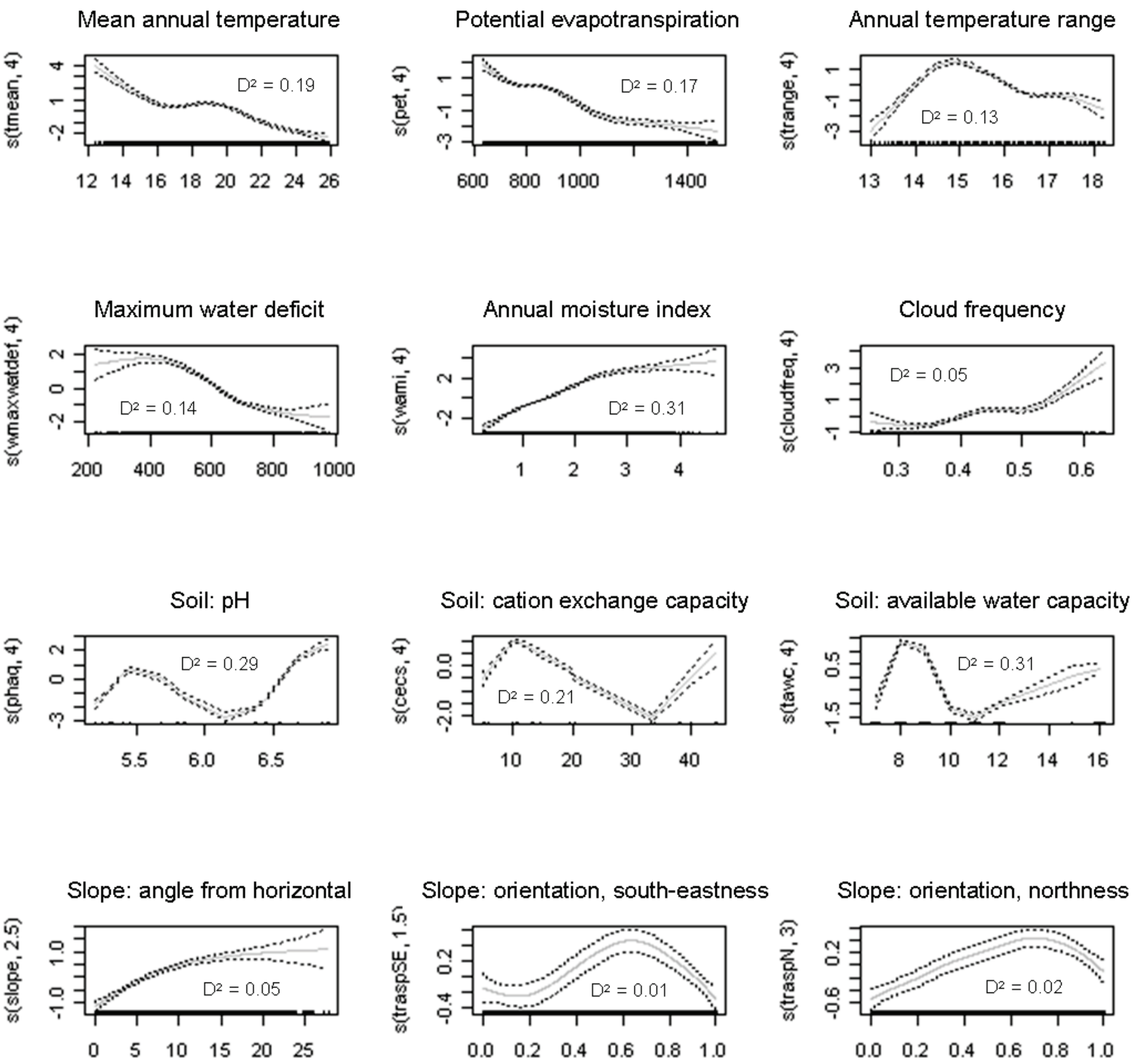
Richness

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.