Vegetation data analysis
Mselezi Forest Reserve

REPORT SUBMITTED TO THE FRONTIER TANZANIA BREAM PROJECT

Antje Ahrends, Alistair Jump, Jon C. Lovett and Rob Marchant

York Institute for Tropical Ecosystem Dynamics (KITE)
aa528@york.ac.uk; rm524@york.ac.uk

Environment Department
University of York
Heslington
York, YO10 5DD, UK

Tel: +44-(0)1904-404372
Fax: +44-(0)1904-432998
http://www.york.ac.uk/res/kite

TABLE OF CONTENTS

1 Summary

One of the most striking characteristics of the vegetation of the Eastern Arc Mountains in Tanzania is the immense variation in both species composition and vegetation structure that to date has eluded a systematic classification. Furthermore, the species' ecologies, their dispersal abilities and the evolutionary history of the area are poorly known. Attempts to tackle these questions - fundamental to developing sound conservation and management plans - are further constrained by large parts of the Eastern Arc Mountains being under-researched. This report has been produced for the Frontier Tanzania Biodiversity Research and Awareness in the Lesser-Known Eastern Arc Mountains project (BREAM), which will significantly contribute to filling gaps in the Eastern Arc Mountain biodiversity research through targeted field assessments. It classifies vegetation associations, the vegetation-shaping environmental gradients, regeneration properties and broad-scale floristic affinities of Mselezi Forest Reserve, Mahenge Mountains (Fig. 1).

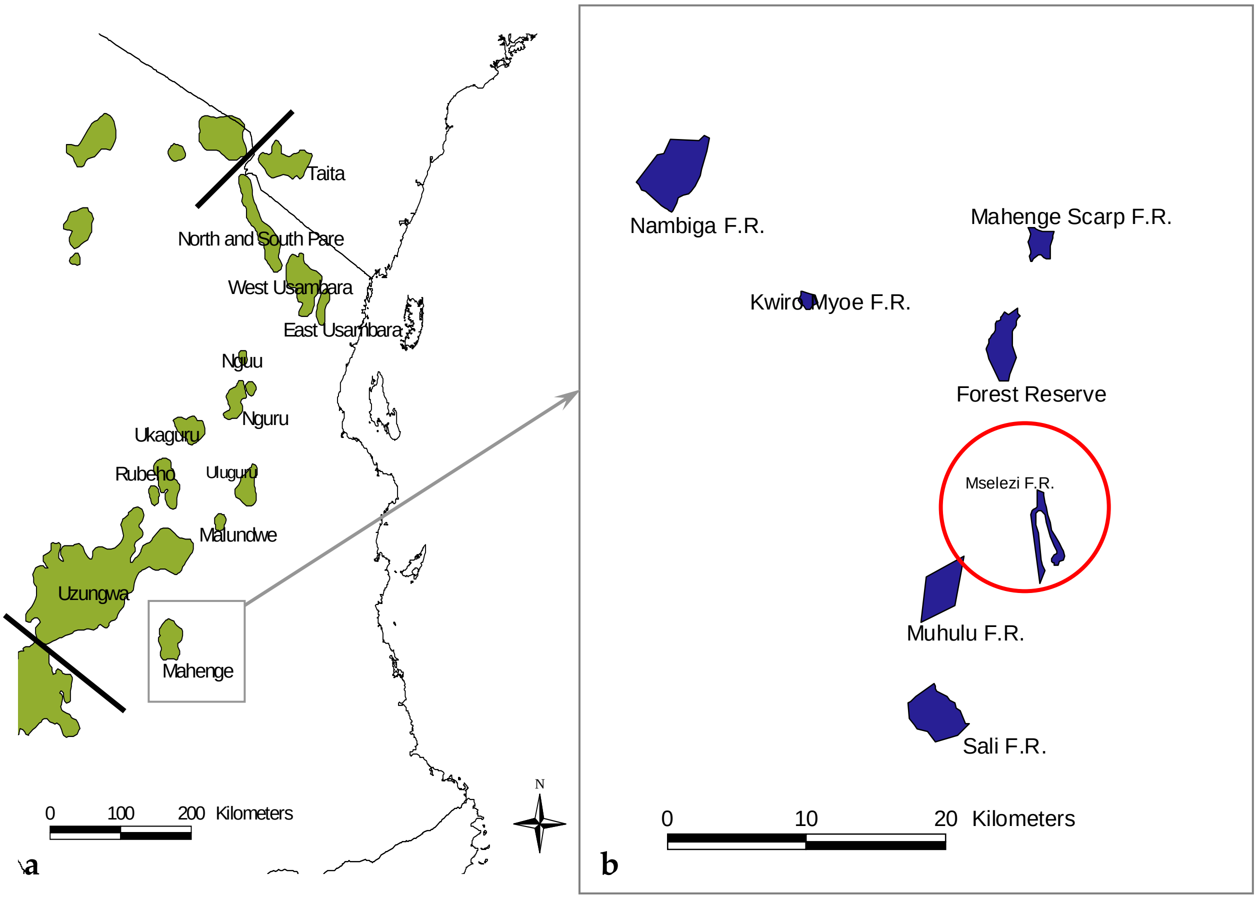

Fig. 1: Eastern Arc Mountains (a) and the Mahenge Mountains region (b)

A **vegetation classification** was attempted using a variety of different techniques including both conventional methods and experimental variations, encompassing (i) the establishment of phyto-sociological groups using TWINSPAN, (ii) a variation of the Braun-Blanquet approach, (iii) the establishment of phyto-ecological groups using species indicator values, and (iv) clustering of structural groups. All these methods proved of limited use to produce a convincing classification of the vegetation. Although species group based analyses (i and ii) coherently identified the same associations, and these results also were acceptably robust to data modifications for different sensitivity analyses, the structural groups only showed a limited overlap with the floristic groups. The assignment of strategy groups (e.g. forest specialist, forest generalist, pioneer) and further indicator properties (e.g. fire tolerant, moisture indicator, disturbance resistant, fire resistant) to the individual species and the analysis of the regeneration properties provided a possible cause for this divergence. Thus, conventional methods typically focusing on a single or a few vegetation properties failed to produce a coherent and widely applicable vegetation classification; the fuller picture only emerging when different approaches were

combined. The necessity for combining different approaches and maximising information that the classifications are based upon was further emphasised by the fact that each of the single approaches suffered substantial shortfalls, in particular with respect to the representativity of the data. The number of species that overlapped between the different plots was so low that any of the apparent groups had to be interpreted with great care. In addition, the high species richness, or rather the lack of clear dominance patterns, eluded a traditional classification using character and differential species. The feasibility of establishing phyto-ecological groups using species indicator values was limited by the majority of the species having poorly known ecologies. While analyses based on the vegetation structure alone tend to be relatively robust in areas where the species dynamics are poorly known, in this particular context the analysis was complicated by extensive human disturbance severely altering the vegetation structure - resulting in great variation. Such a situation means that a larger sampling size is necessary to clearly distinguish between the potential natural vegetation structure and the influence of human disturbance. As the conventional vegetation classification methods thus proved unsuitable due to the high variability of the vegetation and the various shortfalls in data availability, a more appropriate approach combining information on species composition, species' ecological requirements, vegetation structure and site characteristics is proposed. This classification initially is very coarse, but could be subject to further review and refinement once more species identifications (and more data in general through further fieldwork that Frontier Tanzania is conducting) have become available.

A noticeable characteristic of Mselezi plots apparent from the vegetation grouping exercise was the great variability, in both species composition and vegetation structure, preventing a coherent classification of the vegetation. This great variability presumably partly rooted in human disturbance as a vegetation shaping gradient to the natural variety of environmental site conditions. The established groups included (i) dry open canopy woodland on *dry rocky underground in an edaphically restrained climax state* characterised by the presence of *Bombax rhodognaphalon*, *Sterculia appendiculata* and *Drypetes reticulata*, (ii) *dense closed-canopy woodland on East facing (moisture receiving) steep and rocky slopes* with sub-canopy species *Sorindeia madagascariensis* and *Trilepisium madagascariense* underneath a canopy largely dominated by *Ficus sur* - portraying signals of a former disturbance (strongly developed shrub layer, canopy and regeneration dominated by non-forest dependent species, and the prominent presence of species thought to be comparatively disturbance resistant (*Trema orientalis* and *Hoslundia opposite*)) at sites that had been affected by fire, mining and timber cutting, (iii) *grassland savannah with sparse tree recruitment*, (iv) *open lowland woodland in a transitional stage*, and (v) heavily *anthropogencially altered vegetation*, with a number of sites showing signs of recovery in a very initial *pioneering stage* while other sites were still *under cultivation*. These groups were subject to autocorrelations, and their wider applicability is therefore doubtful.

Potential **environmental correlates** (altitude, topography, slope, aspect and water association in addition to past and present disturbance) were explored using different ordination, regression or general linear modelling techniques as appropriate. The analysis showed that land use/disturbance was the dominant factor shaping the vegetation. Due to both present and recent disturbances the vegetation had been significantly altered, entailing divergences between species based and structure based vegetation classifications. Further important vegetation shaping gradients appeared to be the presence of rocky outcrops, the slope inclination and the position along the slope. Vegetation exhibited a more forest-type character the more rocky outcrops present, the steeper the slope and the higher the sample unit position on the slope. However, these might be premature correlations: plots situated at higher altitudes and steeper slopes where rocky outcrops are present might simply be less attractive for human activities and therefore support a more original forest. Plots on East facing slopes generally supported more moisture dependent vegetation communities, apparently benefiting from the moisture laden South-Eastern trade winds that prevail during the dry season.

There was ample **regeneration** of pioneers in plots formerly cleared for agriculture. Plots that had been disturbed a longer time ago were characterised by tree recruitment, however, there was no indication of forest dependent species regeneration at formerly disturbed forests. Such a result might (i) be an artefact of insufficient data, (ii) indicate that the sample unit never supported such species, and/or (iii) that secondary forest re-growth is a long and potentially limited process.

Despite the apparent scarcity of primary forest and the fact that only two threatened and endemic species (*Lettowianthus stellatus* and *Bombax rhodognaphalon*) had been identified so far, Mselezi appeared to be of much scientific interest as the area might constitute a **transitional stage between the lowland coastal forests and the Eastern Arc forests**, indicating that there is an ecological continuum between the two areas. Such an interesting ecological concept has profound implications for understanding and managing the Eastern Arc Mountains ecosystem, and warrants further investigation. The close floristic affinity to the lowland coastal forests also raises an interesting issue apparent in most delineated boundaries, be they geographic, geologic or floristic; the static delineation of the Eastern Arc Mountains might be unable to account for various borderline cases. The Eastern Arc Mountain ecosystem definition should potentially be reviewed, in particular in the face of environmental change and the increased interest in vegetation migration theories and a development of a more dynamic protected area plan.

Further scientific interest could be attached to Mselezi due to the fact that the vegetation has apparently been altered by **disturbance**, in addition to which the forest has a highly fragmented structure. Longer term research could establish whether forest regeneration is at all possible given the various site specific constraints. The supposition that the Mselezi vegetation has been altered by disturbance was further supported by floristic affinity being greatest to lowland forests that similarly as Mselezi had been affected by agricultural encroachment and fire (Namakutwa and Ruvu South forests). Establishing the detailed impact of disturbance from this particular study was a difficult task as so few plots were available. The available data suggested that agricultural encroachment was the most dominant form of disturbance. The very nature of this form of disturbance being total clearance associated with the introduction of an edge effect and increased risk of fire, this constitutes a major threat.

In general, the analysis proved once again how (i) poorly Eastern Arc vegetation-environment relations are understood to date, (ii) how little information there is available on the species-specific ecological tolerances, (iii) how varied the vegetation is across the different Eastern Arc Mountain blocks and (iv) how many more vegetation samples will be needed to establish a robust and refined vegetation classification system that is applicable throughout the Eastern Arc Mountains. Furthermore, the analysis was complicated by the low number of workable plots and the as yet low taxonomic resolution. Nonetheless, the intensive fieldwork that Frontier Tanzania had undertaken - in particular with regard to the meticulous recording of site, vegetation structure and regeneration variables – allowed for a good preliminary analysis, and will certainly pay dividends as more plot data become available. The report concludes with recommendations on the sampling design and protocol that are aimed to add further value to Frontier Tanzania field data collections and our understanding of the Eastern Arc Mountain ecosystem.

2 Research questions

1. **Establishment of vegetation communities:**
    a. Explore options for the **most feasible and suitable vegetation classification(s)** for the data provided, using both conventional methods and experimental variations of conventional methods where applications of these fail, e.g. due to the poor taxonomic resolution of the data, the lack of information on indicator properties of species, the lack of a readily developed vegetation classification system for the broader region, the small plot size, and the high species richness potentially requiring much larger sampling sizes and eluding a traditional classification using character/dominant species.
    b. Where feasible, assign strategy types (e.g. forest specialist, forest generalist, pioneer) and further indicator properties (e.g. fire tolerant, moisture indicator, etc.) to the individual species. Subsequently, identify character (e.g. dominant canopy) and differential species (e.g. indicator species) to further **refine the characteristics of the established vegetation communities**.
    c. Identify the **limitations** of the vegetation data with a view to generate recommendations for future vegetation assessments.

2. **Identification of the vegetation shaping environmental gradients:**
    a. Explore potential environmental correlates (altitude, topography, slope, aspect and water association) using indirect **gradient analysis** with different ordination, regression or general linear modelling techniques as appropriate.
    b. Identify the **impact of disturbance** and fragmentation using residual analysis.
    c. Identify the **limitations** of the plot data with a view to generate recommendations for future plot assessments.

3. **Analysis of the properties of regeneration and associated factors:**
    a. Analyse **regeneration properties** with regard to regeneration stability, presence of potentially invasive species, endemic species and other target taxa.
    b. Explore the potential **environmental and vegetation structural correlates** (e.g. vegetation structure, disturbance, fragmentation, etc.) using Indirect gradient analysis with different ordination, regression or general linear modelling techniques as appropriate.
    c. Identify the **limitations** of the regeneration and regeneration plot data with a view to generate recommendations for future plot assessments.

4. **Further questions of conservation and scientific interest:**
    a. Analyse the factors associated with high **species richness** (e.g. topographic heterogeneity, habitat heterogeneity caused by disturbance, etc.)
    b. Quantify the share of species of conservation concern (i.e. endemic and threatened species), and their properties (e.g. disturbance resistant/resilient or disturbance sensitive, etc.) and assign **relative conservation priority** to the investigated forests (comparative to other Eastern Arc Mountain forest blocks).
    c. Summarise the apparent **impact of different forms of disturbance**.
    d. Derive **overall management implications**.

5. **Further questions of technical interest:**
    a. Summarise **recommendations** for future vegetation assessments.
    b. Analyse the **sensitivity** of study outputs to taxonomic resolution and quality of species id.

3 Methods

3.1 Data preparation

Prior to analysis the data underwent minor modifications. It was necessary to ensure that the plots have unique identifications[1], that diameter at breast height (dbh) measurements are entered in separate cells. The taxonomy was updated to match that of Lovett at al. (*in press*). The data was captured using pivot tables, and the following standard ecological parameters and variables were established[2]:

With respect to potentially **explanatory parameters**:

- **Environmental correlates:** altitude, slope inclination, position on the slope, aspect
- **Disturbance measures:** presence/absence of signs of past use, presence/absence of present use, weighted combined use measurement, number of cut poles, number of cut timbers, number of overall cuts, overall number of disturbances, presence/absence of particular types of disturbances, presence/absence of roads or tracks
- **Other features of interest:** water association, presence/absence of rocky outcrops

With respect to the **dependent vegetation variables**:

- **Vegetation structure related variables:** standing density of trees >10 cm dbh, standing density of trees >20 cm dbh, average dbh of trees >10 cm dbh, average dbh of trees >20 cm dbh, basal area, canopy cover, shrub cover, ground cover, canopy height, number of dead trees
- **Species composition related variables:** species abundance, species dominance, species combined abundance/dominance, species indicator value, species disturbance resistance, species fire resistance

With respect to **dependent regeneration variables:**

- **Regeneration structure related variables:** cover with herbaceous vegetation, cover with litter, cover with stones, extent of bare ground, degree of grass dominance
- **Regeneration species composition related variables:** regeneration stability, species indicator value, proportionate share of pioneering species

With respect to **dependent species richness and diversity variables:**

- **Dominance pattern related variables:** values for Berger-Parker index; values for Simpson's index
- **Diversity related variables:** values for Shannon index, values for Evenness index, total number of species

3.2 General statistical analytical procedures

All statistical analyses have been undertaken using the software packages Microsoft Excel, SPSS 11.5, R-2.3.1, and PC ORD 4.0. Significance was established at $P \leq 0.05$ (significant) and $P \leq 0.01$ (very significant) if not stated otherwise. The treatment of continuous data followed in all cases a standard procedure: For all variables **descriptive statistics** (mean, median, mode, range, standard deviation, variance, skewness, kurtosis and boxplots) were produced to screen for skewness, outliers, extreme values, and possible data entry errors. Subsequently, the variables were tested for normal distribution with a one-sample Kolmogorov-Smirnov test[3]. If a variable was found not to be normally distributed, an attempt was made to normalise it using logarithmic (skewed data), square root (Poisson distributions) and arcsine square root (percentages and proportions) data transformations (Zar, 1999). Where data could not be normalised,

[1] VP1 was termed Ms0101, VP2 renamed to Ms0102 and so forth. Ms stands for Mselezi, the following two digits for the number of the work unit and the last two digits for the number of vegetation plot within that work unit.

[2] How exactly those variables were captured and numerically coded is detailed in the relevant method sections and in Appendix A.

[3] Based on most extreme absolute differences, this assesses the significance of the departure of the given distribution from a theoretical distribution (here set to normal).

the variables were only subjected to non-parametric tests. Where appropriate, data was analysed using parametric or non-parametric **bivariate or multivariate techniques** (e.g. Dytham, 1999).

3.3 Exploration of different approaches to a vegetation classification

In order to identify the most feasible and suitable approach to a classification of different vegetation types for Mselezi forest, the following range of conventional methods and variations of these have been tested:

3.3.1 Two-Way Species Indicator Analysis (TWINSPAN)

Numeric classification using species abundance values alone

TWINSPAN (Hill, 1979; Gauch and Whittaker, 1981) is a numeric vegetation classification procedure that uses a divisive clustering algorithm to simultaneously classify both species and sample groups (plots). TWINSPAN is based on dichotomously dividing an ordination space produced with reciprocal averaging (RA)[4] using the following procedure: The RA ordination identifies the first axis[5], which is divided near its centre. Species preference scores are computed based on their frequency of occurrence on either side of the divided axis. The species are subsequently divided in 'positive preferentials', 'negative preferentials' or 'non-preferentials' with scores assigned. A subsequent refined ordination is based on these species scores excluding the non-preferentials. Borderline cases (plots or species that do not consistently appear on one side of the axis) are grouped based on a third ordination, where a simple discriminant function is used to reproduce the previously established dichotomy based on the most preferential species alone (indicator species)[6]. These three ordination steps are repeated at subsequent levels until the size of the established groups reaches the requested level. The result of TWINSPAN is a two-way dichotomy where both plots and species are consecutively divided into pairs with eigenvalues given for each division axis. Twinspan performs poorly when there is more than one gradient shaping the vegetation. It further fails to produce a convincing classification when the data is characterised by an underlying continuum.

For the Mselezi vegetation analysis the minimum group size for division was set to five. The analysis was repeated with numerous different pseudo-species cut-off levels[7] in order to test the robustness of the results. Further sensitivity analyses were conducted by introducing data modifications as they might occur as more species identifications become available. As such, randomly selected individuals[8] that had only been identified to the genus level were assigned dummy species names, e.g. *Grewia* spec. to *Grewia* spec. 1 and *Grewia* spec. 2. The analysis was based (i) on a dataset containing only those species that had been fully identified, (ii) on a dataset containing all entries were species had either been identified to the species or the genus level, and (iii) on several datasets where dummy species names have been assigned to randomly selected not fully identified individuals.

---

[4] The procedure of RA (Hill, 1973) is one of the most basic ordination techniques, and underlies many of early developed programmes such as TWINSPAN. RA results in an ordination space in which distances between sample points are proportional to their chi-squared distance values (McCune and Mefford, 1999), and thus constitutes a basic eigenvalue analysis technique. It has been described in much detail by Jongman (1995).

[5] The first axis can be understood as a vector in multidimentional space that captures as much of the variation in the data as possible. Whereas a linear model specifically tests for the influence of particular parameters, an ordination might identify a first axis that is uncorrelated with any of the parameters. As such, failures to assess an important parameter can be detected.

[6] TWINSPAPN owes its name to this particular procedure. Hill (1979) acknowledged that TWINSPAN should rather be referred to as 'Dichotomised Ordination Analysis' as the indicator ordination is an appendage and not the real base of the method.

[7] TWINSPAN being created to handle categorical data (species and plots), it cannot account for quantitative species abundance values. This limitation is overcome by the introduction of 'pseudo-species' that in fact represent abundance values for a particular species. For instance, when the pseudo-species cut-off levels are set to one and five, a species x that occurs once in a plot would be called species x 1, and the same species x occurring five or more times in another plot would be called species x 2. Typically, five or six pseudo-species cut-off levels are defined, and these are adapted to the overall species frequency (e.g. Jongman, 1995).

[8] Using a random number draw function in MatLab 6.0.

3.3.2 Phyto-sociological classification (Braun-Blanquet approach)

Subjective classification using combined species dominance-abundance values

The Braun-Blanquet (or Zürich-Montpellier) approach (Braun-Blanquet, 1964; Dierschke, 1994) was developed in continental Europe and constitutes one of the most widespread approaches to classifying vegetation. Contrary to the Russian or Finnish school of vegetation classification where vegetation associations are based on dominant species (e.g. Frey and Loesch, 2004), the Braun-Blanquet approach mainly builds on differential species (often with a narrow ecological amplitude) with character species (often dominant) being assigned subsequently. The classification is strictly hierarchical, and any newly identified association has to be embedded in the existing phyto-sociological system with a strict taxonomy. Associations are assigned names, which typically are based on one character and one differential species (e.g. *Luzula-Fagetum* for a forest characterised by the dominance of *Fagus sylvatica* where the presence of the differential species *Luzula sylvatica* indicates moist and acid soil conditions).

The Braun-Blanquet approach has a strict sampling protocol where species are recorded in separate layers and assigned a combined dominance-abundance value, with 'dominance' being a percentage cover value for each single species. As this had not been established in the field, the dominance component was based on basal area values instead, and a suitable scale developed (Tab. 1).

The species were then entered with their dominance-abundance values into a table; a frequency score was calculated for each species based on the number of occurrences in different plots. Subsequently, all those species occurring with a high frequency across the plots were sorted out, widespread taxa being less suitable as differential species. Furthermore, all individuals that had not yet been identified to the species level were excluded. This was followed by a manual sorting of the table until distinctive 'blocks' of species were identified that occurred across the same subsets of plots and that had low or no representation in any of the other plots (differential species). The possibility of establishing association character species was reviewed. It was not attempted to place the identified associations within a broader system, as - given the overwhelming diversity of species and potential associations in East Africa - attempts to classify the local vegetation with the Braun-Blanquet approach have been few (e.g. Clarke and Robertson, 2000; Hemp, 2005; Hemp, 2006; Schmidt, 1991), and the immense variety has hitherto eluded the development of a hierarchical classification scheme.

Combined dominance-abundance value	Species abundance	Species basal area (in cm)
1	1	<200
2	2-5	<1000
or	1	200 - <1000
3	6-20	<5000
or	1-5	1000 - <5000
4	>20	<10000
or	any	5000 - <10000
5	any	>10000

Tab. 1: Modification of the original Braun-Blanquet scale for the establishment of dominance-abundance values that has been developed for the particular context of this analysis.

3.3.3 Phyto-ecological classification (vegetation form approach)

Subjective classification using combined species-dominance abundance values and species indicator functions

The vegetation form approach roots in the traditional German forest site reconnaissance ('forstliche Standorterkundung') and was further developed by Kopp and Schwanecke (2003) and Succow and Joosten (2001). In central and East Europe this approach is often regarded as the closest rival to the Braun-Blanquet approach (e.g. Frey and Loesch, 2004). Whilst the data assessment in the field and the table preparation using combined dominance-abundance values and frequency scores are exactly the same as in the Braun-Blanquet approach, these two methods differ in that the vegetation form approach emphasises species with a narrow ecological amplitude (thus, a strong indicator value), and that the classification system is more flexible and not hierarchical.

Specifically, ecological vegetation associations are identified by assigning indicator values to individual species[9] followed by manual table sorting, where as in the Braun-Blanquet approach blocks of species are sought that occur across the same subset of plots, however, with the added condition that those species with overlapping indicator values appear within the same groups. As a first step, all available descriptions of the ecological requirements for each of the recorded species were collated (Beentje, 1994; Clarke, 1995a; Lovett et al., *in press*)[10] to decide which indicator categories could feasibly be established. Given that for the majority of the species occurring in East Africa the ecological amplitudes are unknown or established at a very coarse level, only two species indicator properties could be identified: the altitudinal amplitude[11] and the degree of forest dependence[12]. Further species properties that however could only be established nominally were disturbance resistance, fire resistance and moisture indicator. As for some of the species the above authors gave conflicting information, only those species where there appeared to be a consent were assigned into the respective categories. The collated information on the species ecological requirements and the assigned indicator values are presented in Appendix B. The category values were subsequently used for an attempt to identify phyto-ecological groups.

3.3.4 Classification based on the vegetation structure

Numeric classification using vegetation structural variables alone

With respect to the establishment of vegetation structural groups it was considered problematic that there were a total of ten vegetation structure variables, most of which were redundant. Sensitivity analyses showed that clusters were erratic and reacted highly sensitive to the inclusion or exclusion of single variables. When all variables were analysed simultaneously, those vegetation structural aspects that were represented by several collinear variables were overly emphasised.

Therefore, it was necessary to focus on a manageable subset of uncorrelated variables, and a factor analysis with principal component extraction (PCA) was undertaken to replace the ten original data files with their main axes (principal components). The first three axes (eigenvalues of λ = 7.82, λ = 1.89 and λ = 0.92 respectively) accounted for a cumulative 89% of the total variation in the data. A varimax rotated component matrix showed that axis 1 mostly represented average dbh, maximum dbh, density, basal area, tree density, dbh trees and minimum dbh, axis 2 was mostly correlated with canopy cover and canopy height and axis 3 with shrub cover and ground cover.

The vegetation structure cluster analysis was subsequently based on the standardised z-scores of these three components. Both Euclidean and Relativised Euclidean distances[13] were used as distance measures, and clusters were established using single-linkage, average linkage and complete linkage methods.

3.3.5 Ordinations in 'species space' and in 'vegetation structure space' and identification of interrelations

Simultaneous numeric analysis of species and vegetation structure variables

In order to deepen the understanding of how the vegetation structure and species composition are interrelated in this particular case and how both components mutually shape the vegetation, ordinations

---

[9] In continental Europe these are commonly based on the indicator values established by Ellenberg (1992), that assign species preference scores across a total of nine categories such as light, soil pH, moisture and plant available nitrogen and phosphate.

[10] Turrill and Milne-Redhead et al. (1952-) was not at the authors' disposal.

[11] As 'exclusive lowland species', 'exclusive (sub/upper) montane species', and 'species occurring in both lowland and montane areas'

[12] The species were categorised as 'non-forest species', 'forest non-dependent species', and 'forest dependent species', and the categories were defined as in Doggart et al. (2004).

[13] Euclidean distance is conceptually straightforward (the square root of the sum of the squared distances between all data points). While this distance measure in community ecology is generally not suitable, it was fine in this case as the vegetation structure principal components were standardised and normally distributed. However, Euclidean distance tends to overly emphasise outliers. Relativised Euclidean distance (or Chord distance) is similar conceptually, except that the data are normalised such that all of the data points fall on the surface of a unit quarter hypersphere (McCune and Mefford, 1999). Whereas the standardization was not necessary in this particular case, it was nevertheless of interest that Relativised Euclidean distance assigns less weight to quantitative aspects and therewith outliers. The use of distance measures ensured that a more comprehensive picture of the vegetation structural properties was produced.

both in 'species space' and in 'vegetation structure space' were carried out. In addition, the ordination in species space was overlaid with the vegetation structure variables and *vice versa* (i) to review the possibility of establishing a combined vegetation classification using both species and structural components, and (ii) to gain a quick overview of whether species and structural variables showed coherent or divergent signals.

The ordination in species space was carried out using Detrended Correspondence Analysis (DCA) (Hill and Gauch, 1980)[14] without down-weighting rare species. The ordination performance was judged with an after-the-fact coefficient of determination between Relative Euclidean distance in the unreduced species space and Euclidean distance in the ordination space (McCune and Mefford, 1999)[15]. Ordination in vegetation structure space was established using a PCA[16] that computed a cross-products matrix using Pearson's correlation coefficients and broken-stick-eigenvalues were used to decide how many components were worth interpreting (components >1). The correlation between the main axes and the overlaid parameters was established with simple parametric (Pearson's) and non-parametric (Kendall) correlation coefficients.

3.4 Identification of environmental gradients shaping vegetation

Environmental gradients shaping the vegetation were initially explored by linear or non-linear, simple and multiple regression analyses between all individual vegetation variables and all explanatory parameters (as detailed in 2.1)[17]. Subsequently, the vegetation variables were simultaneously analysed using ordinations, and overlaid with explanatory parameters that had been identified as the most important ones in the preceding linear models. The technical procedures applied were the same as detailed 2.3.5.

3.5 Analysis of regeneration properties

Regeneration was analysed with respect to the existence of invasive species and endemic species, and with respect to the regeneration 'stability': whether canopy and sub-canopy and potentially shade-loving species were regenerating or whether the regeneration was dominated by pioneering species. Following this, regeneration variables were categorised and numerically coded as detailed in 2.1 and Appendix A. The analysis of regeneration associated factors was carried out using linear model and ordination

---

[14] DCA is an eigenanalysis ordination technique geared towards ecological datasets. In its core it is based on RA ordination. Contrary to its parent technique, a DCA suppresses the 'arch effect' (the fact that in an RA the second axis almost inevitably is a second order polynomial of the first axis) by diving the first axis into segments, and then setting the average score on the second axis within each segment to zero. Furthermore, after each iteration the axes are rescaled, therewith correcting for the RA tendency to compress the axis ends relative to the middle (McCune and Mefford, 1999).

[15] Due to the processes of rescaling and detrending, the correspondence between the eigenvalue and the structure along that axis is destroyed, and the eigenvalue itself cannot be interpreted as proportion of variance explained.

[16] In its essence, a PCA (Goodall, 1954) is a multi-variable extension of a multiple regression analysis, and the eigenvector is identified by applying straight-line regression and calibration iteratively. The main difference to a DCA is that whereas a DCA accounts for the fact that species or vegetation communities generally exhibit a bell-shaped and not a linear relationship to a given gradient, a PCA assumes a linear monotoic model (e.g. Jongman, 1995; McCune and Mefford, 1999).

[17] The assumption of homocesdascity was tested with a Levene test, and the assumption of normality of the error term was examined in histograms with overlying normal curve. Whether a relation was best expressed linear or non-linear was established using scatter plots and a curve fit tool in CurveExpert 1.3. Multiple regression analyses were checked for multicollinearity, and if the variation inflation reached ≥15, the model was built stepwise with entry and removal probabilities set at $P = 0.05$[17]. The significance of both single and multiple linear regression models was assessed with *F*-statistics (ANOVA), and their performance evaluated based on the coefficient of determination (R^2). When the dependent variable was binary (presence/absence values), and a logistic regression model was appropriate, significance was tested using Wald-statistics and the –2 log likelihood, and different specifications of the model were compared with the Nagelkerke R^2 (Whitehead).

techniques as in 2.4. In addition, potential interrelations between the (mature) vegetation structure and the regeneration were investigated.

3.6 Analysis of further questions of conservation and scientific interest

3.6.1 Analysis of factors associated with high species richness and diversity

Once the remaining species identifications have become available, species richness and diversity will be computed using the Berger-Parker, Simpson's, Shannon and Evenness indices (Magurran, 1988). As required, significant difference between two sampling units in Shannon diversity will be tested with a Hutcheson *t* test (Zar, 1999). The analysis of associated factors with high species richness and diversity will be tested using the same technical procedures as in 2.4.

3.6.2 Broad-scale vegetation comparison

Beta diversity between the Mselezi vegetation and other forests both in the Eastern Arc Mountains and in the East African coastal forests was established as the total number of shared species, and corrected for the overall total of species using the Jaccard's Coefficient and the Sørensen Index (or Coefficient of Community)[18]. While it would have been desirable to correct for the overall collection effort within the different areas, this was a complicated task due to the fact that not all collectors used fixed size plots. Collection intensity could thus only be established in broad rank categories.

3.6.3 Assessment of the apparent impact of different forms of disturbances

Attempts were made to assess the importance of different disturbance types using factor and linear analysis. However, the number of available sample units was too low for meaningful analysis but useful for hypothesis generation.

[18] The reason for using both indices is that the Jaccard Coefficient ($SJ = c/(a+b+c)$) is slightly more sensitive to species richness – a potentially undesirable property that is less prominent in the Sørensen Index ($CC = 2C/(A+B)$).

4 Results

4.1 Exploration of the most feasible and suitable vegetation classification(s):

Although the taxonomic resolution of the data provided so far was low, with the majority of the assessed individuals only being identified to the genus level, there appeared to be relatively stable species grouping. These showed acceptably low variation in various sensitivity analyses, and were robust across the different methodologies applied: Species groups established using abundance values alone resembled those established using combined abundance-dominance values. Furthermore, within confines of poorly known ecological requirements of the respective species, the established groups reflected the species' ecologies. However, there was little overlap between the established species groups and groups established based on the vegetation structure. The following paragraphs illustrate the findings in further detail:

4.1.1 Two-Way Species Indicator Analysis (TWINSPAN)

Numeric classification using species abundance values alone

All analyses carried out in TWINSPAN on the original dataset and modifications of this (Methods) produced similar grouping outputs. It was apparent that a small group of four species (*Bombax rhodognaphalon, Drypetes reticulata, Sterculia appendiculata* and *Turraea mombassana*), recorded in plots Ms0102 and Ms0103, were isolated from the rest of all species (Tab. 2). This division was already established after the tenth reciprocal averaging iteration, and the resulting first axis carried an usually high eigenvalue (λ = 1). A second division was established after the 106th iteration (λ = 0.86) that distinguished a group of species (*Annona senegalensis, Commiphora africana, Ficus sycomorus, Lonchocarpus bussei, Markhamia zanzibarica, Piliostigma thonningii, Pteleopsis myrtifolia, Syzygium guineense, Vitex doniana* and *Xerroderis stuhlmannii*) mainly restricted to plots Ms0101 and Ms0110. Further divisions were established at two successive levels, however, the additional axes were less meaningful with extremely low eigenvalues (λ = 0.1 and λ = 0.01 respectively). Thus, there were three relative distinctive species groups with *Cordia africana, Harrisonia abyssinica* and *Stereospermum kunthianum* being non-preferential taxa, overlapping in their occurrence between these groups.

4.1.2 Phyto-sociological classification (Braun-Blanquet approach)

Subjective classification using combined species dominance-abundance values

The establishment of Braun-Blanquet phyto-sociological groups resulted in a very similar classification (Tab. 2 and Tab. 3)[19]. Although some species were more abundant and/or dominant than others within the identified groups[20], the number of workable plots and identified taxa was too low to allow for the identification of character and differential species in traditional Braun-Blanquet fashion.

4.1.3 Phyto-ecological classification (vegetation form approach)

Subjective classification using combined species-dominance abundance values and species indicator functions

As detailed in the Methods section, only two indicator properties could be established: the altitudinal amplitude and the degree of forest dependence. All but one species were either exclusively 'lowland' or occurred in both lowland and montane areas (Appendix C), and they did not show any particular distributional patterns across the plots. While it fitted that the single montane species (*Zanthoxylum leprieurii*) was found at a relatively high elevation (730 m), it was not possible to classify the vegetation on the available species altitudinal amplitudes. The degree of forest dependence proved a slightly more useful indicator property, however, with most species being 'forest generalists' (Appendix B) widely scattered across all plots, a classification would have had to be based on less than a dozen

---

[19] A negligible difference being that two of the species that in the TWINSPAN divisive clustering appeared as least preferential had been assigned to the second and third group respectively (*Stereospermum kunthianum* and *Cordia africana*).

[20] For instance, *Bombax rhodognaphalon, Pancovia* spec. and *Ricinodendron* spec. in the first group; *Indigofera* spec., *Piliostigma thonningii, Pteleopsis myrtifolia and Vitex doniana* in the second group; and *Ficus sur, Harrisonia abyssinica, Lettowianthus stellatus and Trilepisium madagascariense,* in the third group.

species with a more narrow ecological amplitude. Whilst the available information on the species' ecological requirements did not allow for a classification, the species groups established with both the above approaches were largely coherent with the species ecological indicator values (Tab. 3). Both the first and the third group comprised more forest dependent species, contrasting with the second group that largely embraced species known to prefer open area conditions and a number of species that are thought to be fire-resistant (e.g. *Annona senegalensis* and *Commiphora africana*). The average ecological indicator value for the first group suggested that the corresponding sites might have been characterised by climax forest or dense woodland conditions. The third group was the most heterogeneous group, encompassing forest species (*Sorindeia madagascariensis* and *Trilepisium madagascariense*) with pioneer species in disturbed habitat (*Hoslundia opposita* and *Trema orientalis*). The vast majority of the species in the third group were of the forest dependent or forest generalist type, and a number of those species were moisture indicators. Thus, the first group will be referred to as *forest/woodland*, the second group as *grassland/woodland*, and the third group as *heterogeneous*.

	Ms 0101	Ms 0110	Ms 0108	Ms 0109	Ms 0111	Ms 0112	Ms 0102	Ms 0103				
Turmom	-	-	-	-	-	-	-	2	1			
Steapp	-	-	-	-	-	-	1	-	1			
Dryret	-	-	-	-	-	-	-	1	1			
Bomrho	-	-	-	-	-	-	2	2	1			
Xerstu	1	-	-	-	-	-	-	-	0	1	1	1
Vitdon	1	1	-	-	-	-	-	-	0	1	1	1
Syzgui	1	-	-	-	-	-	-	-	0	1	1	1
Ptemyr	2	1	-	-	-	-	-	-	0	1	1	1
Piltho	3	-	-	-	-	-	-	-	0	1	1	1
Marzan	-	1	-	-	-	-	-	-	0	1	1	1
Lonbus	1	-	-	-	-	-	-	-	0	1	1	1
Ficsyc	1	-	-	-	-	-	-	-	0	1	1	1
Comafr	2	-	-	-	-	-	-	-	0	1	1	1
Annsen	3	-	-	-	-	-	-	-	0	1	1	1
Stekun	2	-	1	-	-	-	-	-	0	1	1	0
Haraby	1	2	2	2	2	-	-	-	0	1	0	
Corafr	-	1	-	1	-	1	-	-	0	1	0	
Zanlep	-	-	-	1	-	-	-	-	0	0		
Turhol	-	-	-	2	-	-	-	-	0	0		
Trimad	-	-	1	-	-	3	-	-	0	0		
Trieme	-	-	-	-	-	1	-	-	0	0		
Treori	-	-	1	-	-	-	-	-	0	0		
Tabodo	-	-	2	2	-	-	-	-	0	0		
Sormad	-	-	-	-	-	3	-	-	0	0		
Perang	-	-	-	1	-	-	-	-	0	0		
Ludaby	-	-	1	-	-	-	-	-	0	0		
Letste	-	-	-	3	4	-	-	-	0	0		
Leppla	-	-	-	1	-	-	-	-	0	0		
Hosopp	-	-	1	-	-	-	-	-	0	0		
Ficsur	-	-	-	-	-	3	-	-	0	0		
Engmag	-	-	2	-	-	1	-	-	0	0		
Dalmel	-	-	-	1	-	-	-	-	0	0		
Chaari	-	-	-	-	1	-	-	-	0	0		
Brimic	-	-	-	1	-	2	-	-	0	0		
	0 0	0 0	0 1	0 1	0 1	0 1	1	1				

Tab. 2: Two-way ordered site by species matrix for the dataset containing the fully identified taxa only (TWINSPAN). A key for the species codes is given in Appendix B. (The code operates with the first three letters of the genus and the first three letter of the species name.) The patterns of zeros and ones in the last columns respectively columns indicate the established group dichotomy in a dendrogram style. The matrix displays the pseudo-species values (Methods). The colour coding represents the groups established using the plant sociological approach and shows how remarkably coherent the

	Ms 0102	**Ms 0103**	**Ms 0101**	**Ms 0110**	**Ms 0108**	**Ms 0109**	**Ms 0111**	**Ms 0112**	**frequency**	**indicator values**	**average indicator value**

Steapp	3								1	1	
Turmom		3							1	2	**1.50**
Bomrho	2	3							2	2	
Drvret		1							1	1	
Annsen			3						1	3	
Comafr			2						1	3	
Ficsvc			3						1	2	
Lonbus			2						1	3	
Piltho			3						1	3	
Xerstu			3						1	2	**2.50**
Stekun			3		2				2	2	
Svzgui			1						1		
Ptemvr			2	2					2	2	
Vitdon			2	3					2	3	
Marzan				2					1	2	
Ludabv					2				1		
Treori					2				1	2	
Engmag					3			1	2	2	
Hosopp					1				1	2	
Leppla						1			1	2	
Tabodo					3	2			2	2	
Dalmel						2			1	3	
Perang						3			1		
Turhol						3			1	1	**1.87**
Zanlep						2			1	2	
Letste						3	5		2	2	
Brimic						1		2	2	2	
Chaari							2		1	2	
Trimad					2			3	2	1	
Ficsur								5	1	2	
Sormad								3	1	1	
Trieme								2	1	2	
Non-differential and non-identified taxa											
Corafr				2		2		3	3	3	
Haraby			2	2	3	3	3		5		
Acaspe					2			2	2		
Albspe	3							1	2		
Amospe			1						1		
Celspe						2			1		
Colspe	1				2	2			3		
Comspe			2				3		2		
Dalspe						3			1		
Ficspe				1	1	1		4	4		
Grespe						2			1		
Indspe				5					1		
Letspe							2		1		
Lonspe				2					1		
Majspe					1	2			2		
Marspe	1								1		
Milspe				3					1		
Panspe	2	4		2		1	2		5		
Ricspe	4				1				2		
Trispe					2				1		
Xylspe								4	1		

Tab. 3: Manually sorted table in typical Braun-Blanquet fashion. The matrix displays the combined dominance-abundance values for the species. A key for the species codes is given in Appendix B.

4.1.4 Classification based on the vegetation structure

Numeric classification using vegetation structural variables alone

Cluster analyses based on the first three components of the vegetation structural variables showed that the vegetation plots forming the *forest/woodland group* also exhibited a relatively homogeneous

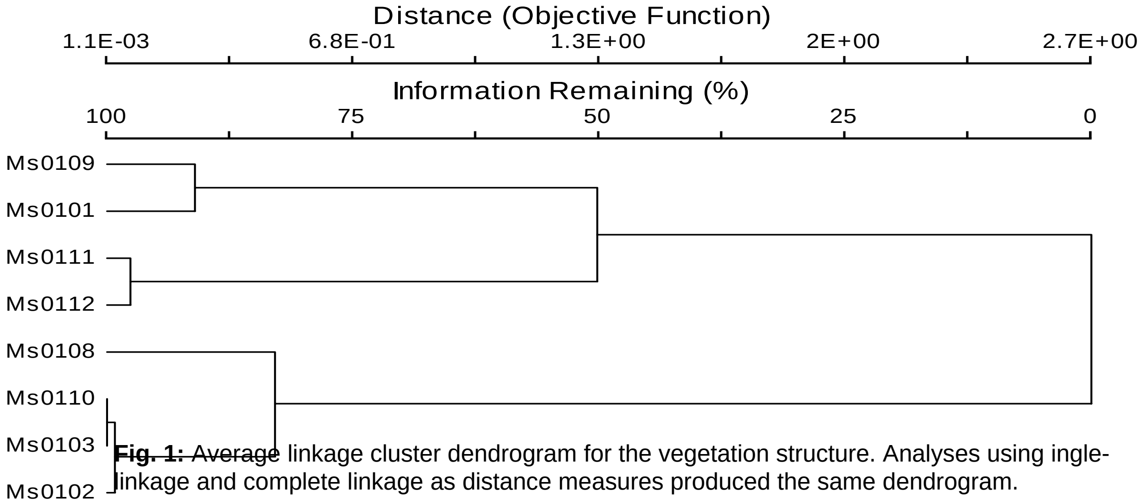

Fig. 1: Average linkage cluster dendrogram for the vegetation structure. Analyses using ingle-linkage and complete linkage as distance measures produced the same dendrogram.

vegetation structure (Fig. 2). Furthermore, those vegetation plots that had been grouped within the *heterogeneous group* also differed with respect to their vegetation physiognomy. The *grassland/woodland group,* that would be expected to exhibit a similarly open vegetation structure, did not cluster together.

Fig. 2: Vegetation structure average-linkage cluster. The same result was produced when using single-linage and complete-linkage methods.

The clusters were stable across the different distance measures.

Ms0102 and Ms0103 (*forest/woodland group*) were both characterised by a low density of trees (>10 cm dbh) and particularly big trees (>20 cm dbh). However, some of the trees present had a large dbh, the dbh varying substantially (Fig. 3). The canopy in this group was less than 10 m high and did not interlock; despite a presumably high percentage of light reaching the ground, both shrub and ground layer were not strongly developed. Further conspicuous features included the comparatively high number of dead trees. Ms0110 (*grassland/woodland group*) also clustered within the same group (with an information loss of less than 5%). However, Ms0110 exhibited a slightly larger number of trees with big trees of a slightly smaller dbh than those in Ms0102 and Ms0103 (with a narrow 95% confidence interval). The 'sum' being the same suggest that at this stage the analysis carries an artefact from merging the number of trees and their average dbh into a single principal component. Ms0110 was further distinguished by a higher canopy, medium canopy coverage, a low shrub density and high ground cover. At a later stage, and with an information loss of nearly 25%, the Ms0102-Ms0103-Ms0110 cluster was joined by Ms0108 (*heterogeneous group*). While the density of trees in Ms0108 was similar to that in Ms0110, the trees had a smaller average dbh, the canopy coverage was lower and the shrub layer much stronger developed.

Another cluster was that of the plots Ms0111 and Ms0112 (*heterogeneous group*): both sites exhibited a canopy height of 20 to 30 m and a medium to high canopy coverage. However, whilst Ms0112 was characterised by a large number of mostly small but a few very big trees, Ms0111 contained fewer trees with a slightly higher average dbh. Furthermore, the shrub cover in Ms0111 was considerably higher inhibiting the development of an extensive ground cover, whereas both shrub and ground cover reached medium values in Ms0112.

Ms0101 showed a pronounced physiognomic similarity to Ms0109: both plots were characterised by a low to medium canopy cover with a low coverage, a medium to high shrub and ground cover, a very small average dbh and little variation in this. The most conspicuous difference was that that Ms0109 had a much higher density of trees.

Vegetation plot number	Canopy cover (%)	Shrub layer (%)	Ground layer (%)	Canopy height (m)	average dbh (cm)	max dbh (cm)	tree density (>10 cm dbh)	tree density (>20 cm dbh)	basal area (cm)
Ms0101	<10	10-50	10-50	<10	24.51	47.4		27	14564.02
Ms0102	<10	<10	<10	<10	34.67	80.1	21	25	31674.11
Ms0103	<10	<10	<10	<10	29.79	69	12	15	14337.21
Ms0104	<10	<10	<10	<10	0	0	0	0	0
Ms0105	<10	>50	>50	<10	0	0	0	0	0
Ms0106	<10	>50	>50	<10	0	0	0	0	0
Ms0107	<10	10-50	<10	<10	0	0	0	0	0
Ms0108	10-50	10-50	10-50	10-20	17.96	42	34	29	9818.25
Ms0109	<10	>50	10-50	10-20	20.01	60.5	51	52	21133.03
Ms0110	10-50	<10	>50	20-30	29.17	62.2	33	30	26260.05
Ms0111	10-50	>50	<10	20-30	30.05	57.4	22	25	19803.65
Ms0112	>50	10-50	10-50	20-30	27.5	100	34	37	34778.73

Tab. 4: Vegetation structure details across sample units.

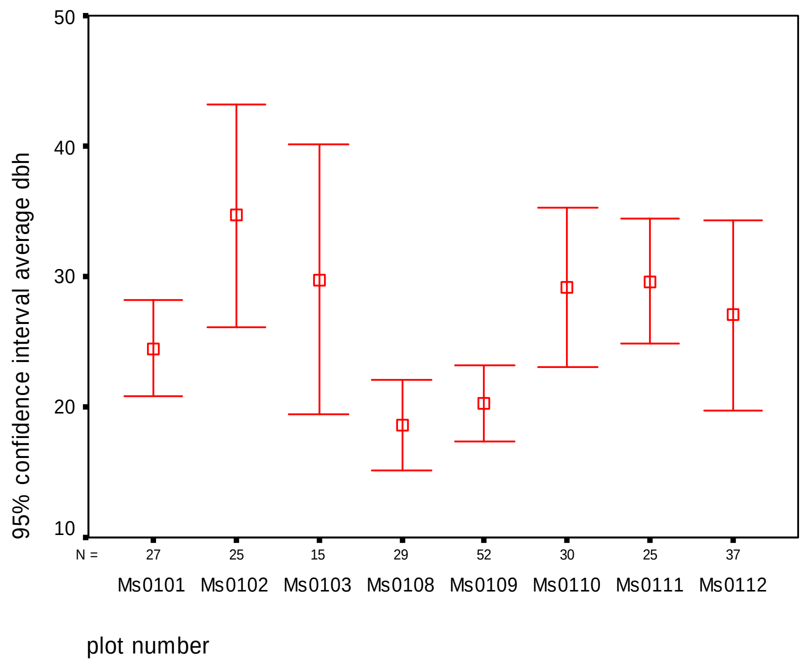

Fig. 3: 95% Confidence Intervals for average dbh across sample units.

4.1.5 Ordinations in species space and in vegetation structure space and identification interrelations

Simultaneous numeric analysis of species and vegetation structure variables

In order to gain a more detailed understanding of how the vegetation structure and the species composition are interrelated and mutually shape the vegetation, ordinations were carried out within species space and vegetation structure space and overlays performed.

Ordination in species space with vegetation structure overlay

With respect to ordinations in species space, DCAs were performed on the original dataset and on variations of this. As the case with TWINSPAN and Braun-Blanquet type analyses, the results were robust to these transformations of the original data. The three established species groups were clearly visible (Fig. 4), however, whilst the *forest/woodland group* and the *woodland/grassland group* were strongly separated from each other, there was a certain amount of overlap between the *heterogeneous group* and the *forest/woodland group*. Ms0109, Ms0111 and Ms0110 were situated relatively close in ordination space, indicating that only in the case of Ms0101 a separation from the heterogeneous group might be justifiable. The overlay with vegetation structural parameters confirmed that the *forest/woodland group* was distinguished by a low stem density and a high average dbh of trees. This characteristic strongly separated this group, particular the *grassland/woodland group*, that exhibited the opposite structure: a low stem density with a low average dbh. This latter group was mainly characterised by high ground cover (positively associated with the first axis, Tab. 5). The *heterogeneous group* comprised several vegetation types ranging from high stem density combined with low average dbh (Ms0108), medium stem density combined with low average dbh (Ms0111) to medium stem density combined with high average dbh (Ms0112). If these plots would form part of the same vegetation type, there are presumably strong environmental gradients present to account for these differences.

Axis	**1**			**2**		
	R	**R^2**	**tau**	**R**	**R^2**	**tau**
cancov	-0.381	0.146	-0.13	-0.402	0.162	-0.303
shrcov	0.084	0.007	0.206	-0.331	0.11	-0.206
grcov	**0.363**	**0.132**	**0.303**	-0.435	0.189	-0.477
caheight	-0.147	0.022	0.124	-0.295	0.087	-0.124
avdbh	-0.181	0.033	-0.143	**0.611**	**0.373**	**0.571**
treedens	0.056	0.003	0.036	**-0.565**	**0.319**	**-0.618**
basalarea	-0.492	0.242	-0.286	0.099	0.01	0
avdbhtr	**-0.561**	**0.314**	**-0.429**	0.488	0.238	0.286

Tab. 5: Correlation between the overlaid parameters (vegetation structure variables) and the species space ordination axes. Values are given for Pearson's correlation coefficient (R), coefficient of determination (R^2) and Kendall correlation coefficient (tau). The strongest positive and negative correlations are highlighted in bold and in grey and red colour respectively.

Key: cancov = canopy cover; shrcov = shrub cover; grcov = ground cover; caheight = canopy height; avdbh = average dbh of trees >10 cm dbh; treedens = density of trees >10 cm dbh; basalarea = basal area; avdbhtr = average dbh of trees >20 cm dbh

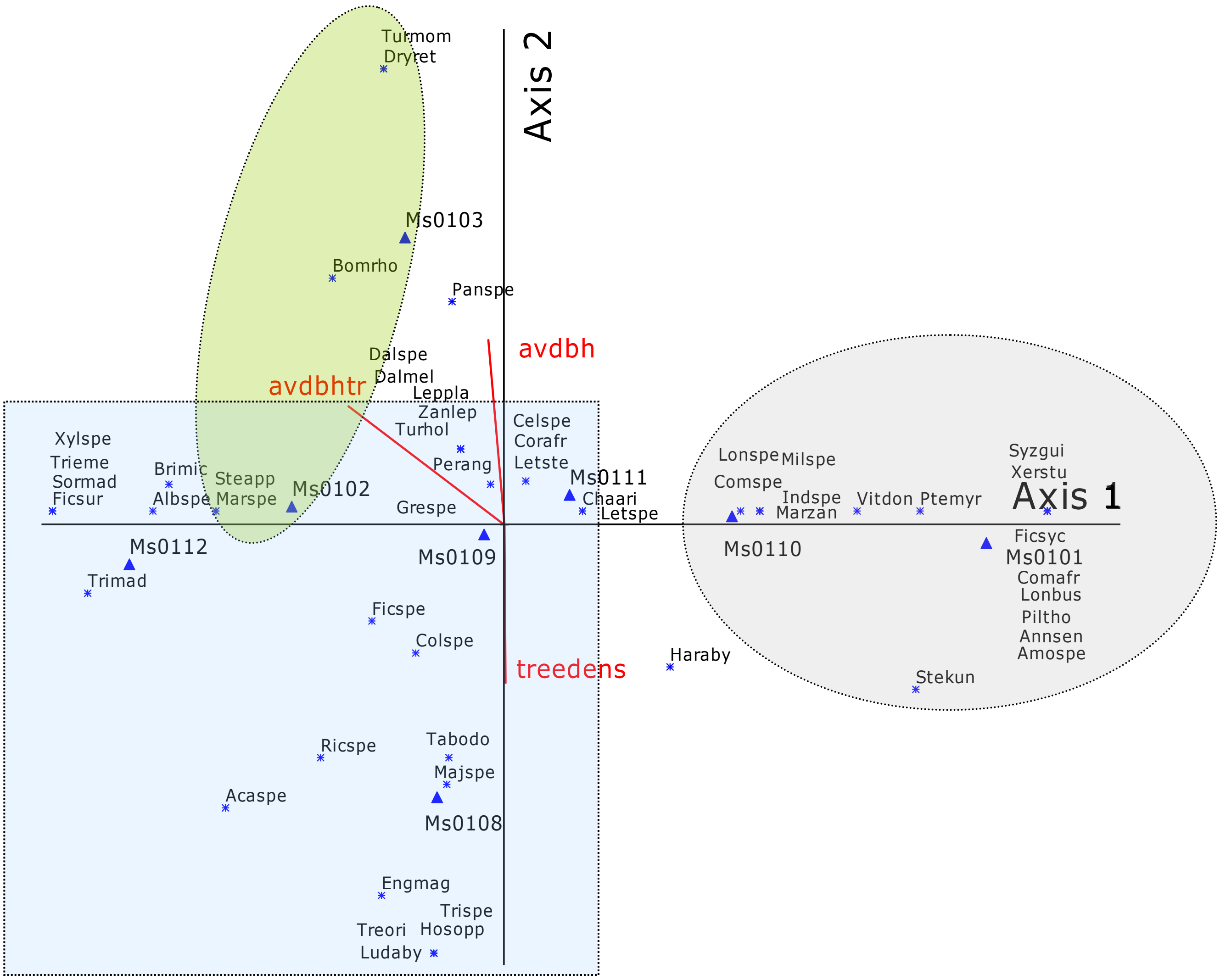

Fig. 4: Graphic ordination results of a DCA in species space. A key for the species codes is presented in Appendix B. Overlaid is a the are the vegetation structural variables. The R^2 cut-off level was set to 0.5 and the vector scaled to 100%.

Ordination in vegetation structure space with species overlay

PCA ordinations in vegetation structure space revealed three important axes with broken-stick eigenvalues of $\lambda = 2.72$, $\lambda = 1.72$ and $\lambda = 1.22$ respectively that combined explained approximately 87% of the total variation in the data. The first axis was mainly negatively associated with the average dbh, and positively associated with tree density and shrub cover (Tab. 6). axis 2 was negatively associated with canopy cover and canopy height and axis 3 with shrub cover and ground cover[21].

The ordination in vegetation structure space produced a different plot orientation than that in species space (Fig. 5). The only coherent grouping was that of Ms0102 and Ms0103 (*woodland/forest group*). Ms0101 and Ms0110 (*grassland/woodland group*) were separated along the second axis with Ms0101 being characterised by low (and Ms0110 by high) canopy cover and canopy height. A recurring phenomenon was that of the *heterogeneous group* being stretched out between the two main axes that represented average dbh and tree density along the first, and canopy cover and height along the

[21] These also are the three principal components that have been subjected to the vegetation structure cluster analysis (Methods).

second. With the exception of the *forest/woodland group* the vegetation structure based groups differed from the species based groups, implying that the identified associations were not robust. However, it was evident that there was an interrelation between vegetation structure and species composition: The proportionate share of open area species strongly increased as the average dbh decreased and the density of trees and the shrub cover increased (Fig. 5).

Axis:		**1**			**2**			**3**	
	R	**R^2**	**tau**	**R**	**R^2**	**tau**	**R**	**R^2**	**tau**
cancov	0.199	0.04	0.043	**-0.84**	**0.706**	**-0.65**	-0.211	0.045	-0.13
shrcov	0.739	0.545	0.701	0.054	0.003	-0.041	**-0.631**	**0.398**	**-0.454**
grcov	0.538	0.29	0.303	-0.443	0.196	-0.13	**0.677**	**0.458**	**0.477**
caheight	0.38	0.144	0.124	**-0.824**	**0.68**	**-0.701**	-0.266	0.071	-0.289
avdbh	**-0.891**	**0.794**	**-0.714**	-0.25	0.063	-0.143	-0.166	0.028	-0.143
treedens	0.834	0.696	0.618	-0.228	0.052	-0.255	0.237	0.056	0.036
basalarea	-0.359	0.129	-0.286	-0.755	0.57	-0.571	-0.001	0	-0.143
avdbhtr	**-0.858**	**0.737**	**-0.857**	-0.437	0.191	-0.286	0.118	0.014	0

Tab. 6: Correlation between the ordination variables (vegetation structure variables) and the ordination axes, indicating how the different variables are represented in the ordination space. For further explanations refer to Tab. 5.

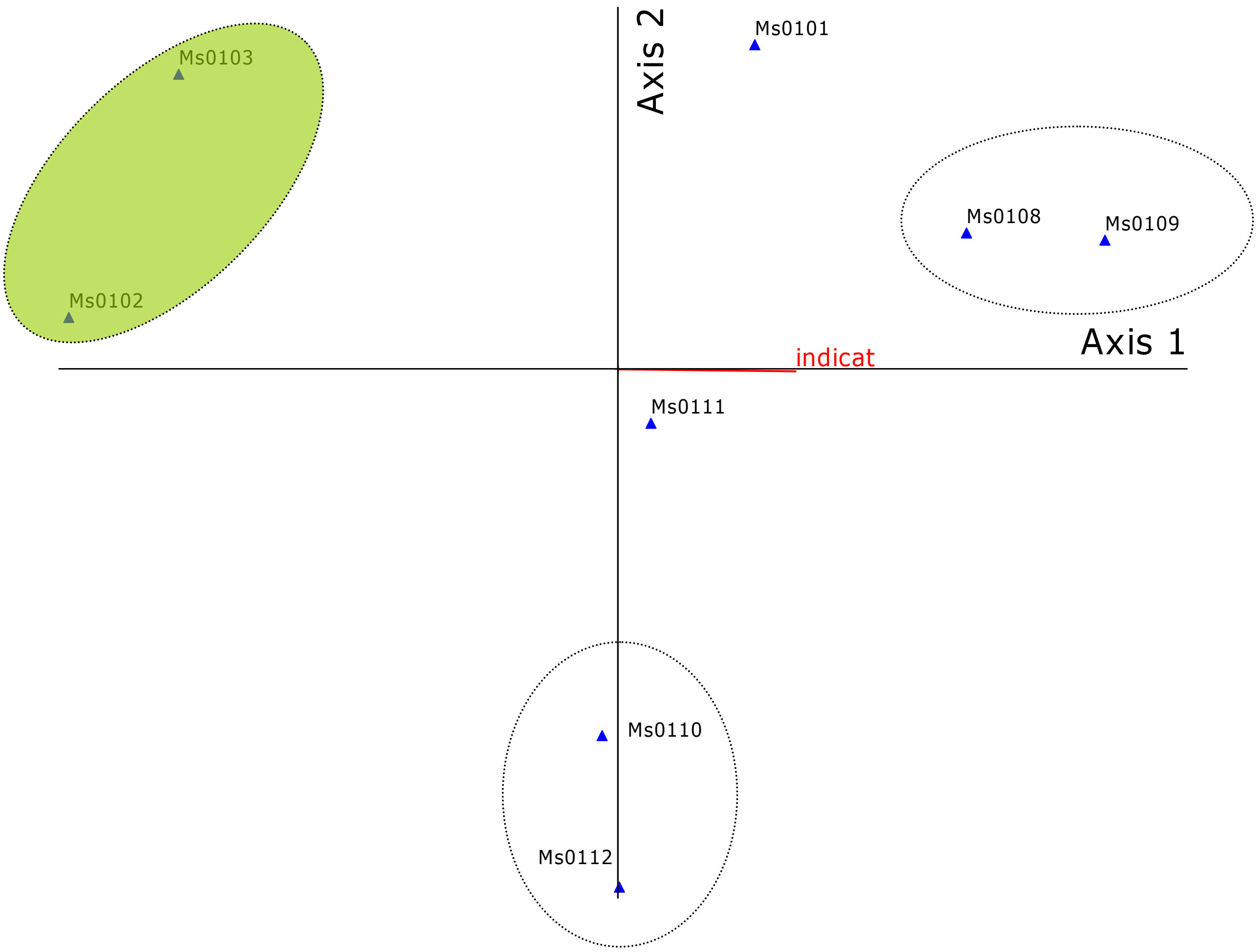

Fig. 5: Graphic ordination results of a PCA in vegetation structure space. Overlaid is a the species indicator value (Methods and Appendix B). It is evident that there is a strong positive correlation between the first axis and the species indicator value (R = 0.58; R2 = 0.34; tau = 0.47). The R^2 cut-off level was set to 0.5 and the vector scaled to 100%.

4.2 Identification of environmental gradients shaping vegetation

4.2.1 Environmental gradients shaping the species composition

A DCA ordination in species space with environmental parameter overlay revealed that the most important parameters were the presence of rocky outcrops and the slope inclination (Tab. 7 and Fig. 6). The most important parameters along the second ordination axis were aspect and a combined measure for past and present land use (Appendix A). In addition, the first axis was strongly positively correlated with the species ecological indicator value; the average share of open area species increased as the presence of rocky outcrops and the slope inclination decreased. Furthermore, both the first and the second axis were correlated with the presence of disturbance resistant species; the number of disturbance resistant species increasing as the presence of rocky outcrops and the slope inclination decreased and the land use measure increased.

Axis:	1			2		
	R	R^2	tau	R	R^2	tau
altitude	-0.374	0.14	-0.143	-0.245	0.06	0
slope	0.05	0.002	0.154	-0.231	0.054	-0.309
aspect	-0.394	0.155	-0.244	**-0.518**	**0.269**	**-0.342**
slopedes	**-0.518**	**0.268**	**-0.342**	-0.012	0	-0.049
position	-0.462	0.214	-0.303	-0.39	0.152	-0.303
rocky	**-0.54**	**0.291**	**-0.436**	0.057	0.003	0.218
use	-0.003	0	0.109	**-0.498**	**0.248**	**-0.109**
indicator	0.834	0.443	0.276	-0.200	0.040	-0.182
disturbance	0.661	0.437	0.265	-0.306	0.094	-0.718

Tab. 7 Correlation between the overlaid parameters and the ordination axes. For further explanations refer to Tab. 5 and Appendix A.

Key: altitude = altitude in m; slope = slope in degree; aspect = prevalence of eastern aspect yes/no; slopedes = slope inclination in four classes; position = position along the slope; rocky = presence/absence of rocky outcrops; use = combined variable for current and past land use

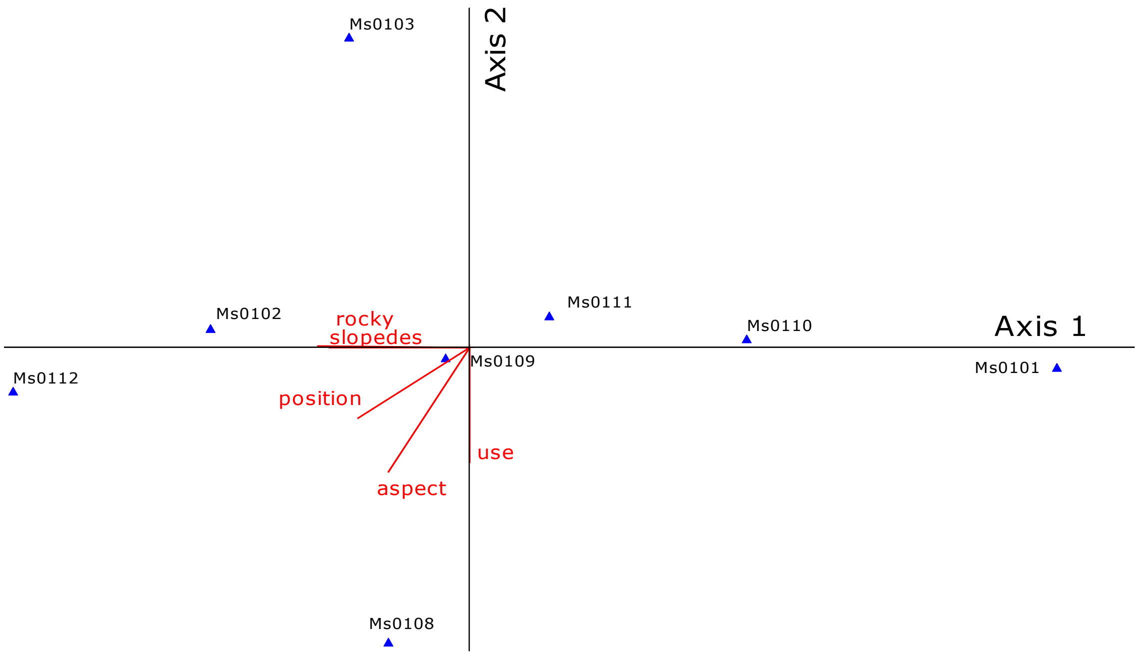

Fig. 6: Graphic ordination results of a DCA. A key to the coding of the overlaid parameters is given in Tab. 7. The R^2 cut-off level was set to 0.2 and vectors scaled to 100%.

4.2.2 Environmental gradients shaping the vegetation structure

The ordination in vegetation structure space (PCA) overlaying environmental parameters was carried out on the full dataset including those plots where, due to agricultural encroachment, no trees were present. Not surprisingly, the main gradient was the combined past and present land use measure (Tab. 8 and Fig. 7). A further apparently important gradient was the presence of rocky outcrops; the density of trees and their average dbh increasing with rocky outcrops present (Tab. 8 and Tab. 9). The same was true for sites situated at higher altitudes, sites with steeper slopes and sites at higher positions on the slope. The most important gradient along the second axis was aspect - sites facing east exhibited a higher canopy and denser canopy cover. The third axis was again associated with the presence of rocky outcrops, with the canopy cover increasing as more rocky outcrops were present (e.g. Ms0110). It was negatively correlated with slope - steeper slopes having a positive impact on the development of the shrub layer (e.g. Ms0109).

Axis:		**1**			**2**			**3**	
	R	**R^2**	**tau**	**R**	**R^2**	**tau**	**R**	**R^2**	**tau**
cancov	-0.687	0.473	-0.676	0.456	0.208	0.589	0.484	0.235	0.196
shrcov	-0.631	0.399	-0.544	0.301	0.09	0.262	**-0.6**	**0.36**	**-0.544**
grcov	-0.704	0.496	-0.641	0.272	0.074	0.103	0.073	0.005	-0.021
caheight	-0.802	0.643	-0.786	0.465	0.216	0.423	0.243	0.059	0.02
avdbh	-0.838	0.702	-0.467	**-0.488**	**0.238**	**-0.3**	0.116	0.013	0.1
density	**-0.919**	**0.844**	**-0.79**	0.033	0.001	0.084	-0.322	0.104	-0.118
basalarea	**-0.887**	**0.786**	**-0.8**	-0.307	0.094	-0.1	0.196	0.038	0.1
treesno	**-0.934**	**0.872**	**-0.824**	-0.041	0.002	0.05	-0.313	0.098	-0.05
avdbhtr	**-0.874**	**0.764**	**-0.533**	**-0.433**	**0.188**	**-0.233**	0.135	0.018	0.233

Tab. 8: Correlation between the ordination variables (vegetation structure variables) and the ordination axes, indicating how the different variables are represented in the ordination space. For further explanations refer to Tab. 5.

Axis:		**1**			**2**			**3**	
	R	**R^2**	**tau**	**R**	**R^2**	**tau**	**R**	**R^2**	**tau**
altitude	-0.433	0.188	-0.413	0.201	0.04	0.064	-0.27	0.073	-0.191
slope	-0.435	0.189	-0.364	0.239	0.057	0.331	**-0.341**	**0.117**	**-0.264**
aspect	-0.025	0.001	-0.065	0.539	0.291	0.676	0.181	0.033	-0.065
slopedes	-0.433	0.187	-0.302	0.142	0.02	0.302	-0.177	0.031	-0.181
position	-0.478	0.229	-0.307	0.041	0.002	0.271	-0.083	0.007	-0.163
rocky	**-0.651**	**0.423**	**-0.559**	-0.249	0.062	0.086	0.528	0.279	0.43
use	0.837	0.700	0.596	0.471	0.221	0.311	0.016	0	-0.078

Tab. 9: Correlation between the overlaid parameters and the ordination axes. For a key to the variable code refer to Tab. 7, for information on the variable calculation to Appendix A, and for further explanations to Tab. 5.

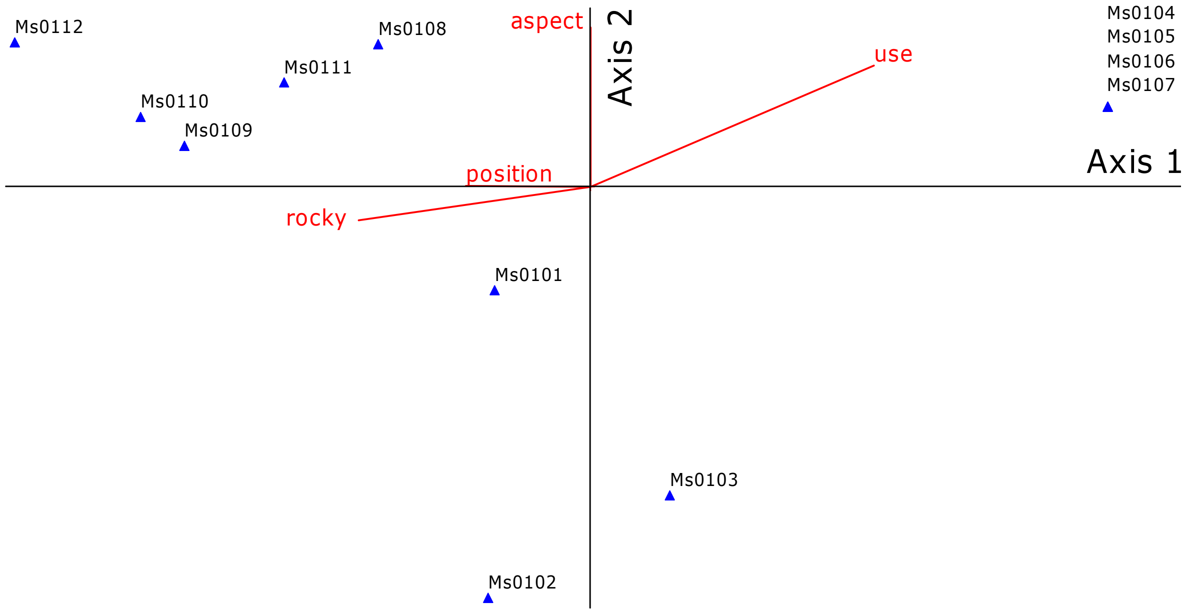

Fig. 7: Graphic ordination results of a PCA including all plots. A key to the coding of the overlaid parameters is given in Tab. 7. The R^2 cut-off level was set to 0.2 and vectors scaled to 100%.

4.3 Analysis of regeneration properties

4.3.1 Analysis of regeneration properties

The analysis of regeneration properties was constrained by the small regeneration plot size that might have captured little of the overall variation in species regeneration, and by a number of unidentified species. The overall number of recorded regenerating species was low with only 15 identified species and further seven unidentified genera; amongst these there were neither potentially invasive species nor endemic or threatened species. The available data suggested that regeneration was 'stable' in Ms0108 and Ms0110 with canopy species restocking as saplings[22], and presumably also stable in Ms0112 and Ms0111 with sub-canopy species being found amongst the samplings[23]. No regeneration was recorded for plots Ms0102 and Ms0103 where the rocky underground would hamper the establishment of seedlings or saplings. However, given the well developed tree layer, despite the unfavourable ground conditions, Ms0102 and Ms0103 might also represent vegetation in an edaphically restrained 'climax state'[24]. The divergence between canopy species and regenerating species In Ms0101 and Ms0109 suggested that the vegetation might be in a transitional state. Regenerating in Ms0105 and Ms0106 was species rich; in a typical post-canopy removal pioneering state. Only in Ms0112 there was substantial regeneration of shade loving species; one shade-loving species (*Mellera lobutlata*) was regenerating in Ms0109, and one (*Markhamia lutea*) in Ms0105, and virtually no shade-loving saplings were found in the remaining plots. Ms0105 and Ms0106 were characterised by ample regeneration of species that are known pioneers in disturbed habitats (*Deinbollia borbonica, Combretum pentagonum, Flueggea virosa, Lippia javanica* and *Triumfetta tomentosa*).

[22] *Englerophytum magalismontanum* in Ms0108 and *Indigofera* spec. in Ms0110

[23] *Sorindeia madagascariensis* and *Trilepisium madagascariense* in Ms0112 and Eng mag in Ms0111

[24] The term 'climax state' is used to describe a stage in which for the vegetation as a whole a directed succession (e.g. from woodland to forest) has been superseded by a more circular and often patchier dynamic. At this stage the vegetation constitutes the potential natural vegetation within the complexity of influences of environmental factors, unless its development is inhibited by human impacts. While this stage is more stable than a pioneering stage, vegetation dynamics naturally persist.

4.3.2 Analysis of factors associated with regeneration species composition properties

A species indicator value based PCA ordination (2.1) showed that there was only one main axis (broken-stick eigenvalue of λ = 1.83) accounting for a substantial overall variation in the data of 67%. This axis was mainly associated with the presence of pioneering species opposed to the degree of regeneration stability (Tab. 10 a). With respect to the vegetation structure, the first axis was positively correlated with the average dbh and negatively associated with the shrub and ground cover percentage (Tab. 10 b). Thus, the higher the average dbh the more stable the regeneration; and the higher the shrub and ground cover the higher the presence of pioneering species. Finally, with respect to potential environmental correlates there was a prominent relationship to land use; increasing land use resulting in a greater number of pioneers (Tab. 10 c and Fig. 8). In addition, the seemingly paradoxical relationship was evident: the steeper and higher the slope, especially when combined with the presence of rocky outcrops, the more stable the regeneration.

Axis:	1		
	R	R^2	tau
stability	0,937	0,878	0,625
indicator	-0,477	0,227	-0,756
pioneer	**-0,946**	**0,895**	**-0,644**

a

Axis:	1		
	R	R^2	tau
cancov	0,279	0,078	0,03
shrcov	**-0,569**	**0,323**	**-0,517**
grcov	**-0,736**	**0,541**	**-0,68**
caheight	0,31	0,096	-0,028
avdbh	0,897	0,805	0,542
treedens	0,654	0,428	0,167
basalarea	0,744	0,553	0,447
avdbhtr	0,927	0,859	0,589

b

Axis:	1		
	R	R^2	tau
altitude	0,573	0,329	0,442
slope	0,527	0,278	0,122
aspect	0,268	0,072	0,102
slopedes	0,661	0,437	0,493
position	0,782	0,611	0,449
rocky	0,607	0,369	0,446
use	**-0,875**	**0,734**	**-0,649**

c

Tab. 10 a: Correlation between the ordination variables and the ordination axes.

Key: stability = canopy or sub-canopy species regenerating yes/no; indicator = average species indicator value in regeneration; pioneer = number of pioneering species in the regeneration

Tab. 10 b: Correlation between vegetation structure variables and the ordination axes. For a key to the variable code refer to Tab. 3.4.

Tab. 10 c: Correlation between environmental correlates and a combined measure for past and present use and the regeneration ordination axes. For a key to the variable code refer to Tab. 7.

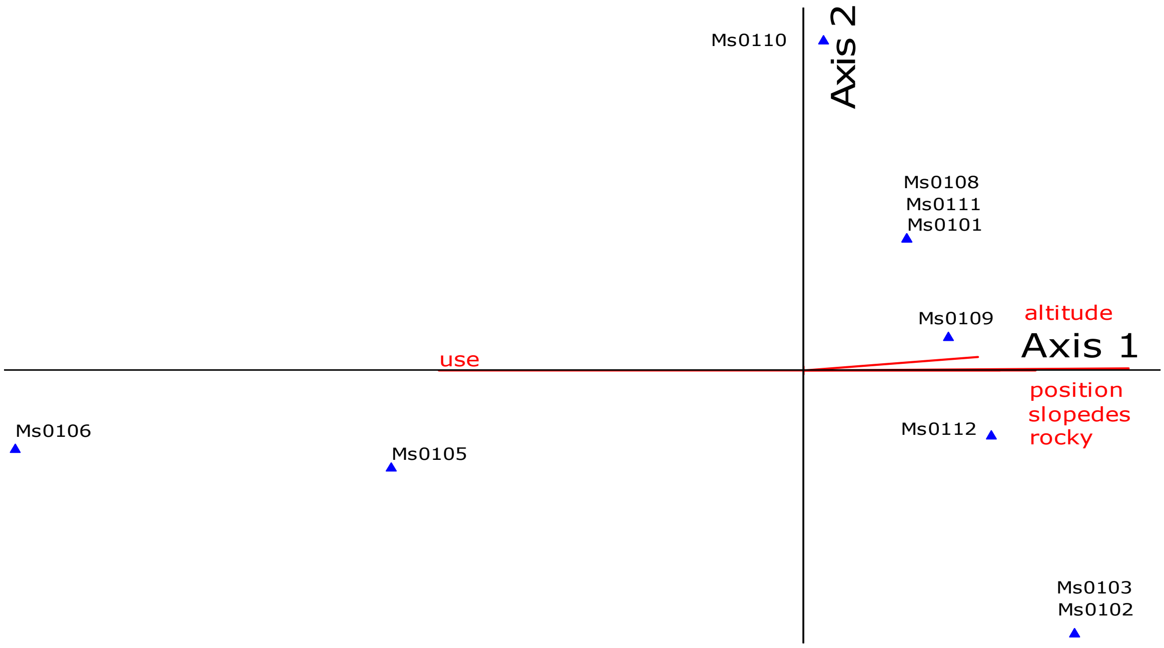

Fig. 8: Graphic ordination results of a PCA in regeneration species indicator values space. A key to the coding of the overlaid parameters is given in Tab. 7. The R^2 cut-off level was set to 0.3 and vectors scaled to 100%.

4.3.3 Analysis of factors associated with regeneration structure properties

A PCA ordination of the regeneration structure properties (2.1) produced two axes jointly accounting for 68% of the overall variation in the data (broken-stick eigenvalues of λ = 2.28 and λ = 1.28 respectively). The first axis mainly represented the percentage cover of litter and of bare soil (opposing each other), the second axis accounted for the herbaceous cover opposed to grass dominance (Tab. 11 a). The first relationship was expected (the more litter the less bare soil), although the second relationship was more interesting: the more herbaceous vegetation overall the less grass therein and *vice versa*. Both axes were correlated with the same vegetation structural properties: canopy height and canopy cover on the one hand, and shrub density on the other (Tab. 11 b); the higher the canopy and the greater the coverage, the more litter and the more herbaceous vegetation with low grass dominance. With regard to environmental correlates, the combined measure for past and present land use was again the most important predicting parameter (Tab. 11 c); the more intensive land use at a particular site, the less litter present and higher prevalence of grasses. This relationship was opposed by inclination and aspect of the slope: steep and East facing slopes being characterised by more litter) and the presence of rocky outcrops (rocky outcrops presence having a positive influence on the development of the herbaceous layer (Fig. 9).

a

Axis:	**1**			**2**		
	R	**R^2**	**tau**	**R**	**R^2**	**tau**
herbaceous	0,314	0,098	0,254	**-0,617**	**0,38**	**-0,56**
bare	**-0,83**	**0,689**	**-0,53**	-0,427	0,182	-0,354
litter	**0,929**	**0,863**	**0,493**	-0,29	0,084	-0,261
stones	0,573	0,328	0,435	0,153	0,024	0,145
grasses	0,024	0,001	0,122	**0,847**	**0,718**	**0,122**

b

Axis:	**1**			**2**		
	R	**R^2**	**tau**	**R**	**R^2**	**tau**
cancov	**0,648**	**0,419**	**0,452**	**-0,528**	**0,278**	**-0,452**
shrcov	**-0,206**	**0,042**	**-0,131**	**0,369**	**0,136**	**0,392**
grcov	-0,195	0,038	-0,196	0,033	0,001	-0,065
caheight	**0,683**	**0,467**	**0,523**	**-0,544**	**0,296**	**-0,327**
avdbh	0,524	0,275	0,366	-0,367	0,135	-0,141
treedens	0,451	0,203	0,057	-0,336	0,113	-0,171
basalarea	0,35	0,122	0,141	-0,524	0,275	-0,479
avdbhtr	0,53	0,28	0,197	-0,455	0,207	-0,423

c

Axis:	**1**			**2**		
	R	**R^2**	**tau**	**R**	**R^2**	**tau**
altitude	0,36	0,13	0,278	-0,29	0,084	0
slope	**0,609**	**0,37**	**0,354**	-0,058	0,003	0,059
aspect	**0,552**	**0,304**	**0,393**	-0,248	0,061	-0,157
slopedes	0,385	0,149	0,204	-0,425	0,181	-0,204
position	0,479	0,229	0,34	-0,369	0,136	-0,093
rocky	0,448	0,201	0,298	**-0,525**	**0,276**	**-0,447**
use	**-0,406**	**0,165**	**-0,204**	**0,346**	**0,120**	**0,272**

Tab. 11 a: Correlation between the ordination variables and the ordination axes.

Key: herbaceous = cover percentage of herbaceous vegetation; bare = cover percentage of bare soil; litter = cover percentage of litter; stones = cover percentage of stones; grasses = dominance percentage of grasses amongst the herbaceous vegetation

Tab. 11 b: Correlation between vegetation structure variables and the ordination axes. For a key to the variable code refer to Tab. 3.4.

Tab. 11 c: Correlation between environmental correlates and a combined measure for past and present use and the ordination axes. For a key to the variable code refer to Tab. 7.

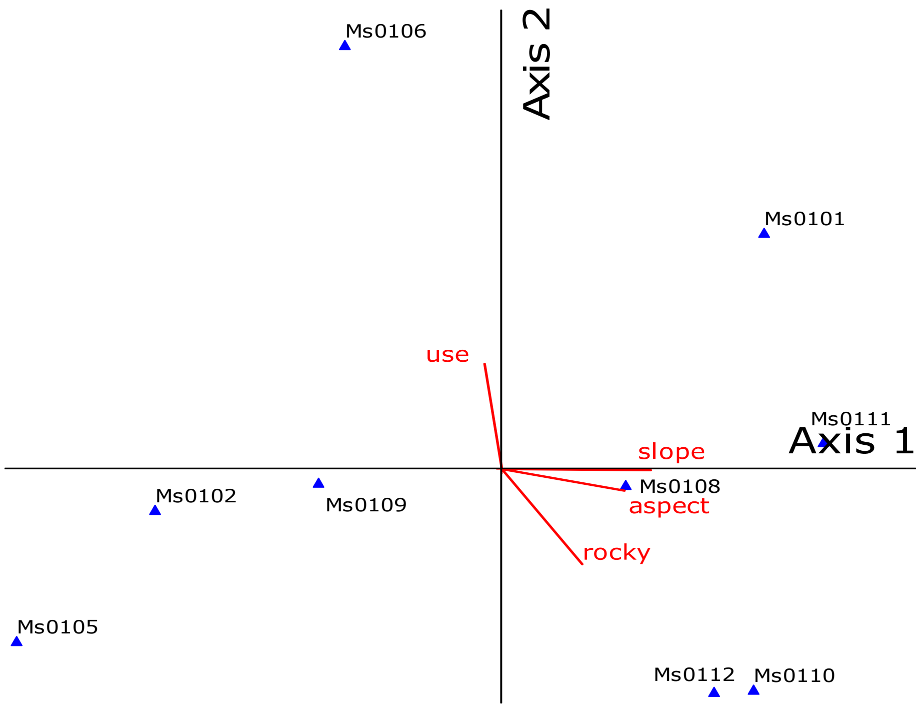

Fig. 9: Graphic ordination results of a PCA in regeneration structure space. A key to the coding of the overlaid parameters is given in Tab. 7. The R^2 cut-off level was set to 0.3 and vectors scaled to 100%.

4.4 Further questions of conservation and scientific interest:

4.4.1 Analysis of the factors associated with high species richness and diversity

With nearly half of the recorded individuals not yet being identified to the species level this analysis would not be very meaningful at this stage and will be postponed until the full species identifications are available.

4.4.2 Quantification of the share of species of conservation concern and their properties

As above, this analysis will be suspended untill full species identifications are available. Thus far, two species (*Lettowianthus stellatus* and *Bombax rhodognaphalon*) have been identified that are classified as vulnerable (IUCN, 2006; Oldfield et al., 1998) and also near endemic to the region (e.g. White, 1988; Clarke, 1995) with both species being generally found in lowland forest. *Bombax rhodognaphalon* can also occur in thicket and *Lettowianthus stellatus* is often also found in woodlands, thus both species are forest generalists and not known to be resistant or resilient to disturbances or fire.

4.4.3 Broad-scale vegetation comparison

Mselezi Forest is classified as one of the Eastern Arc Mountain forests (originally delineated based on climate and geology) (Lovett et al., 2000). In terms of its floristic composition it shows stronger affinity to the East African coastal forests (Appendix C). More than 50% of the species are predominantly lowland species, with another 30% being known to occur in both lowland and (sub) montane areas. Thus far only two species have been identified that are thought to be mainly restricted to mountainous areas (*Strombosia scheffleri* and *Zanthoxylum leprieurii*) in addition to *Markhamia lutea* that was found in the regeneration plots).

Both the Jaccard's Coefficient and Sørensen Index showed that the floristic affinity is greater between Mselezi and selected coastal forests[25] than to selected forest of the Eastern Arc region[26]. Beta diversity

[25] The following forests reserves were included in the comparison: Vikindu, Pugu, Pande, Ruvu South, Kisiju, Mchungu, Ngumburuni, Namakutwa and Kiwengoma

[26] Including East Usambara, West Usambara, Khihansi, Mufindi, Mazumbai, Nguru, Sanje, Chome, Taita, Uluguru North and South and Udzungwa

is apparently lowest between Mselezi and Namakutwa forests (SJ = 0,096 and CC = 0,212). Both share a great number of species that are predominantly forest generalist or woodland species such as *Dalbergia melanoxylon, Lettowianthus stellatus*, *Pteleopsis myrtifolia* and species that are additionally known to be fire/disturbance resistant, e.g. *Commiphora africana* and *Annona senegalensis*. Furthermore, there is a conspicuous floristic similarity between Mselezi and Ruvu South and Kiwengoma, which again is mainly due to common predominantly forest generalist and woodland species. Pugu forest shares the greatest number of species with Mselezi, these include forest dependent species to open area species. Despite the existence of a few shared species, the floristic affinity appaered very low to the Udzungwa and Taita forests in particular, and also to the Usambara and Uluguru areas. The collection intensity has been high in all these latter areas, and in some of those higher than in the coastal forests. When only those trees that reached a dbh greater than 20 cm were taken into account, the floristic similarities between Mselezi and the Udzunga, Taita, Usambara and Uluguru forests proved even lower.

With respect to the vegetation structure, the lack of closed canopy forest and the predominance of woodland and thicket type vegetation was conspicuous. However, the density of trees >20 cm dbh, the mean dbh and the basal areas were on average higher than in the majority of the coastal forests (compare e.g. Ahrends, 2006; Clarke, 1995b; Clarke and Dickinson, 1995; Clarke and Stubblefield, 1995; Lowe and Clarke, 2000).

4.4.4 Detailed summary of the apparent impact of different forms of disturbance.

A detailed break-down of the impact of different forms of disturbances proved infeasible as the sampling sizes of plots being affected by different disturbances was too low.

5 Discussion

5.1 Exploration of suitable approaches to a vegetation classification

5.1.1 Performance of conventional methods in general

Application of conventional methods, and variations of these, to classify the vegetation included (i) the establishment of phyto-sociological groups using TWINSPAN, (ii) a variation of the Braun-Blanquet approach, (iii) the establishment of phyto-ecological groups using species indicator values, and (iv) clustering of structural groups. These were of limited use to produce a convincing classification of the vegetation. Although species group based analyses (i and ii) coherently identified the same associations, and these results also were acceptably robust to data modifications for different sensitivity analyses, the structural groups only showed a limited overlap with the floristic groups. The assignment of strategy groups (e.g. forest specialist, forest generalist, pioneer) and further indicator properties (e.g. fire tolerant, moisture indicator, disturbance resistant, fire resistant) to the individual species and the analysis of the regeneration properties provided a possible cause for this divergence (see 4.3). Thus, conventional methods typically focusing on a single or a few vegetation properties failed to produce a coherent and widely applicable vegetation classification; the fuller picture only emerging when different approaches were combined.

5.1.2 Specific limitations to the performance of the conventional methods

The necessity for combining different approaches and maximising information that the classifications are based upon was further emphasised by the fact that each of the single approaches suffered substantial shortfalls, in particular with respect to the representativity of the data. The species group based analyses (i, ii, and iii) were impacted by poor taxonomic resolution and small number of plots: any vegetation classification based on only twelve (and eight workable) plots would be under-representative and arbitrary[27], and a classification would only be convincing if the established associations within the data were strong, and robust across different methodologies - this was not the case. The number of species that overlapped between the different plots was so low that any of the appearing groups had to be interpreted with great care. In addition, the high species richness, or rather the lack of clear dominance patterns, eluded a traditional classification using character and differential species. The feasibility of establishing phyto-ecological groups using species indicator values was limited by the majority of the species having poorly known ecologies. Typically, the species specific ecological amplitudes had been established on a very coarse level (e.g. *lowland to submontane*, or *submontane to upper montane*), if at all, and therefore only suitable for broad-scale comparisons. This rendered a phyto-ecological vegetation classification across a small gradient those established in the work unit relatively arbitrary. While analyses based on the vegetation structure alone tend to be relatively robust in areas where the species dynamics are poorly known, in this particular context the analysis was complicated by extensive human disturbance severely altering the vegetation structure - resulting is great variation. Such a situation means that a larger sampling size is necessary to clearly distinguish between the potential natural vegetation structure and the influence of human disturbance.[28]

[27] As a rule of thumb, in order to delineate a vegetation association this particular association should ideally be represented with at least five assessments. Thus, depending on how many vegetation associations appear to exist in the field, the number of plots should be fivefold or more, especially when the plots are not located subjectively but randomly or systematically.

[28] Of further technical interest was the fact that the analysis of vegetation structure types using different clustering techniques proved far less meaningful than those based on ordinations in 'vegetation structure' space. This was due to the fact that the very nature of cluster analysis is to merge information. Whilst the numerical information loss might have been low for instance when merging plots with a low average dbh and a high tree density on the one hand and with those that had a low tree density but a high average dbh, the actual information loss was very high. This limitation was further enhanced by priorly subjecting the data to a PCA in an attempt to remove redundant variables. The results of the cluster analysis were therefore interpreted with great care.

5.1.3 Suggestions for a classification scheme

With the conventional vegetation classification methods proving unsuitable due to the high variability of the vegetation and the various shortfalls in data availability, it seemed most appropriate to (i) opt for a coarser approach and (ii) combine information on species composition, species' ecological requirements, vegetation structure and site characteristics for the classification. This approach is typically featured in ecosystems where the vegetation is not known in enough detail, where high species richness combined with a low sample number elude a traditional classification focusing on species composition or vegetation structural properties alone. The inclusion of regeneration properties added further useful information and temporal depth to the interpretation. The scheme that was used here was the following:

1. Exploration of overlap and divergences between species composition groups and vegetation structure groups and preliminary delineation of those groups that appear consistent to attempt a broad classification identifying character species for the respective vegetation types.
2. Refinement and interpretation of the potential associations in the light of site conditions and species indicator functions.
3. Refinement and interpretation of the potential associations in the light of the regeneration properties.
4. Derivation of a name for the potential associations as exemplary:
   [most conspicuous vegetation structural characteristics, e.g. *open canopy*] + [characteristic and/or differential species e.g. *Bombax rhodognaphalon – Drypetes reticulata*] + [general vegetation type description according to Greenway (1973) e.g. *edaphic woodland*] + [description of prevalent and potentially vegetation shaping environmental site conditions e.g. *on dry rocky underground*] + [state as indicated by the regeneration properties e.g. *in an edaphically restrained climax state*]
   If the vegetation association is strongly dominated by certain processes, these might be placed right at the beginning, for example *disturbed, recovering,* or *fire induced*.

A relatively underdeveloped ecological classification concept, and potentially promising for the type of analysis Frontier Tanzania are undertaking, is that of Plant Functional Types (based on plant traits such as pollination mode, seed type, leaf shape, etc.). Within the framework of the KITE project, the potential of this classification will be explored in the context of the Eastern Arc flora and results be made available.

5.1.4 Establishment of the actual vegetation classification[29]

The different classification analyses coherently identified similarities in the vegetation between Ms0102 and Ms0103. The vegetation of these two plots was characterised by the presence of forest species (*Bombax rhodognaphalon, Sterculia appendiculata* and *Drypetes reticulata*), and the absence of species that are known to be resistant to disturbances or pioneers in disturbed habitat (thus resilient to disturbances), suggesting these plots represented a relatively undisturbed forest environment. Vegetation structure was characterised by a low density of trees with a comparatively high average dbh. Low values for canopy height, canopy, shrub and ground cover, the extremely low tree density (in particular in Ms0103), and the scare regeneration however implied environmental constraints. Indeed, a prominent site characteristic was the existence of rocky outcrops that might have inhibited tree establishment. In addition, both plots were situated on the middle part of a non-East-facing slope, thus in areas that might receive less moisture during the dry season when the moisture laden South-East trade winds constitute an added moisture source, and might be affected by a water run-off effect. The vegetation in Ms0102 and Ms0103 might consequently be described as *dry open canopy woodland on dry rocky underground in an edaphically restrained climax*[30] *state.*

[29] It should be noted that this classification initially is very coarse, but could be subject to further review and refinement once more species identifications (and more data in general through further fieldwork that Frontier Tanzania is conducting) have become available.

[30] This term is used to describe a dynamic stage in which for the vegetation as a whole a directed succession (e.g. from woodland to forest) has been superseded by a more circular and often patchier successional dynamic.

A further group consistently identified across the different classification exercises was that of Ms0111 and Ms0112. Both plots exhibited a relatively high canopy and a medium to high canopy coverage, and were situated on an East facing steep upper slope where rocky outcrops were present. Whilst Ms0102 was characterised by forest species and moisture indicators (*Sorindeia madagascariensis* and *Trilepisium madagascariense*) underneath a canopy largely dominated by the hydrophile *Ficus sur*, Ms0111 contained species such as *Lettowianthus stellatus* that do not strictly depend on a forest environment and are supported in a drier environment. Furthermore, the density of trees was lower in Ms0111 and the shrub layer stronger developed - regeneration seemed stable in both plots. Ms0111 and Ms0112 might represent *a dense closed-canopy woodland on East facing (moisture receiving) steep and rocky slopes*, with Ms0111 either portraying signals of a former disturbance[31] (strongly developed shrub layer, canopy and regeneration dominated by non-forest dependent species) or simply representing a slightly disadvantaged site being situated at an altitude >700 m on an upper very steep slope. A further plot that exhibits slightly similar conditions is Ms0108, it also shares a number so species with Ms0112 (some of which are moisture indicators), and resembles Ms0111 with respect to the vegetation structure (medium canopy cover, a well developed shrub layer and a low number of trees with a dbh greater than 20 cm). A further characteristic of Ms0108 is the prominent presence of species thought to be comparatively disturbance resistant (*Trema orientalis* and *Hoslundia opposita*) contrasting with the presence of forest species such as *Trilepisium madagascariense*. Being situated on an East facing steep upper slope, Ms0108 is characterised by very similar environmental conditions as Ms0111 and Ms0112. Added with the apparent overlap in terms of species composition to Ms0112 and the structural similarities to Ms0111, one might want to regard Ms0108 as part of the same vegetation type that however has been altered by disturbance; indeed, Ms0108 having been affected by fire, mining and timber cutting.

Ms0101 and Ms0109 are structurally somewhat similar (both plots exhibiting a low number of trees with an average dbh bigger than 20 cm, a sparsely developed canopy cover and relatively well developed ground and shrub cover). However, whilst Ms0101 represents a *grassland savannah with sparse tree recruitment*, there are forest species regenerating in Ms0109. Ms0109 might thus be characterised as an *open lowland woodland in a transitional stage*. There are a great number of possibilities why this might be the case, e.g. the site might have been a former forest environment subjected to natural or anthropogenic disturbances in the past, or it might simply be situated close to forest vegetation with occasional random forest tree recruitment. Ms0101 might either represent an edaphic woodland/grassland or portray signs of a more recent disturbance. Whatever, Ms0101 and Ms0109 cannot be regarded as representing the same vegetation as the similarities are restricted to the vegetation structure. The site conditions differ substantially (while Ms0109 is situated on a steep slope at high altitude Ms0101 is situated on a gentle lower slope), species with overlapping distribution seem to be restricted to those that are widespread anyway (e.g. *Harrisonia abyssinica*), and the vegetation structure differs in that the density of small trees is much higher in Ms0109 than in Ms0101, and Ms0101 exhibits a well developed grass layer. Ms0110 also exhibits a vegetation seemingly isolated from that of the rest. This plot is floristically similar to Ms0101, but structurally divergent with a high canopy and medium canopy coverage, a high density of trees and comparatively great basal area. Structural characteristics form a slight contrast with the presence of mainly woodland species, possibly explained by Ms0110 being bisected by a stream, thus characterised by a slightly azonal vegetation.

Ms0104, Ms0105, Ms0106 and Ms0107 represent *anthropogenically altered vegetation*, with Ms0105 and Ms0106 showing signs of recovery in a very initial *pioneer stage*, and Ms0104 and Ms0107 still being sites *under cultivation*.

5.1.5 Notes on the wider applicability of the groups established

A noticeable characteristic of Mselezi plots apparent from the vegetation grouping exercise was the great variability, in both species composition and vegetation structure, preventing a coherent

[31] Ms0111 is a former mining site

classification of the vegetation. This great variability is presumably due to human disturbance. With both vegetation structure and species composition either being directly, or indirectly, altered at the majority of the sites, the investigation fell back on largely using the environmental site conditions as a base for the group establishment with the regeneration providing some additional information. The established groups were subject to autocorrelations (with Ms0102 and Ms0103 as well as Ms0108, Ms0111 and Ms0112 being situated close to each other), and their wider applicability is therefore doubtful. Furthermore, Mselezi seemingly presents a transitional vegetation type between the lowland coastal forest vegetation and the Eastern Arc vegetation. This particular combination might not be found elsewhere.

5.2 Identification of environmental gradients shaping vegetation

Potential environmental correlates (altitude, topography, slope, aspect and water association in addition to past and present disturbance) were explored using different ordination, regression or general linear modelling techniques as appropriate. While it had initially been intended to conduct the analysis with the natural environmental gradients as main factors and analyse the residuals in the light of potential additional influences of use/disturbance, the apparent strong influence of disturbance/use necessitated the opposite approach: Residuals from land use/disturbance as fixed factor were interpreted in the light of other environmental correlates. These had a comparatively minor explanatory power. Use/disturbance alone was able to explain up to 90% of the variation in the data when the anthropogenically altered plots Ms0104, Ms0105, Ms0106 and Ms0107 were included. Even with these being excluded, use/disturbance accounted for approximately 50% of the variation in the data both in the vegetation structure as well as in the vegetation composition. Other important gradients were the presence of rocky outcrops and the inclination of the slope; implying that both the vegetation structure and the vegetation composition approximated a forest type environment with an increasing presence of rocky outcrops and a steeper inclination of the slope. Further parameters that pointed in the same direction (less significant) were the altitude and the position on the slope, thus, the plots were characterised by a higher tree density, greater basal area, higher and denser canopy cover and the presence of forest species with increasing altitude and a higher position on the slope. It is very likely that the apparent correlations are deceptive in that the actual factor they represent is that of accessibility. Plots where rocky outcrops are present, and that are situated at higher altitudes and steeper slopes might simply be less attractive for human activities and therefore preserve a more original forest environment.

A further important gradient was aspect, often correlated with the second axis: East facing slopes appeared to carry more voluptuous vegetation than non-East facing slopes. While this might be down to East facing aspects receiving more moisture laden air from the South Eastern trade winds prevalent during the dry season, this might also be an artefact of the small number of workable plots in the analysis.

5.3 Analysis of regeneration properties

The analysis of regeneration stability showed that whilst some of the vegetation plots seemed to be in a relatively stable state (with canopy and sub-canopy species regenerating), other plots seemed to be in a transitional stage, possibly following disturbance. Although this interpretation of the very small regeneration plots remains somewhat speculative, the former assumption seems at least plausible as the plots thought to represent a transitional stage were also characterised by a comparatively well developed shrub layer, a high density of trees with a yet small average dbh, and the presence of pioneering species (e.g. Ms0101 and Ms0109).

A question that is of much interest is whether once disturbed plots can regenerate to a forest environment. While a valid answer would have to based on time-series data, it could however be noted that in none of the previously disturbed plots forest species were regenerating, and tree recruitment appeared to be difficult due to the very densely developed shrub layer. No endemic, nor invasive, species were recorded in the regeneration. This however might simply be down to the fact that so few species were recorded overall.

5.4 Further questions of conservation and scientific interest:

With only two threatened and near-endemic species Mselezi appears to be of comparatively little interest as far as species conservation is concerned, however, (i) not all species identifications are available yet, (ii) the threat classification for the hotspot's forest species classification for the region is in urgent need of review (Gereau and Luke, 2006) as, for instance, locally forest loss and disturbance have been much more sever than was assumed (Ahrends, 2006) and (iii) an assignment of conservation interest based on threatened or endemic species alone would provide a very restricted view. An indication of the high level of endemism in the Mahenge region is given by the presence of *Allanblackia stuhlmannii, Garcinia semseii, Pterocarpus mildbraedii* subsp. *usambarensis* and *Octoknema orientalis* (Ministry of Natural Resources and Tourism, 2005). The area is fairly under-researched and further floristic inventory are needed. The Mahenge Mountains are also of great scientific interest due to their pivotal position at the south-eastern range of the Eastern Arc Mountains,

Mselezi in particular appears to have a closer floristic affinity to the lowland coastal forests, however, given its geographic position has been classified as belonging to the Eastern Arc forests. This raises an interesting issue apparent in most delineated boundaries, be they geographic, geologic or floristic; the static delineation of the Eastern Arc Mountains might be unable to account for various borderline cases. The Eastern Arc Mountain ecosystem definition should potentially be reviewed, in particular in the face of environmental change and the increased interest in vegetation migration theories and a development of a more dynamic protected area plan. Further scientific interest could be attached to Mselezi due to the fact that the vegetation has apparently been altered by disturbance, in addition to which the forest has a highly fragmented structure. Longer term research could establish whether forest regeneration is at all possible given the various site specific constraints.
The supposition that the Mselezi vegetation has been altered by disturbances receives further support from the apparent floristic affinity to Namakutwa and Ruvu South forests, in particular with respect to common species that typically occur in secondary vegetation. Both Namakutwa and Ruvu South forests have been affected by disturbances in the past. The Namakutwa plateau, for instance, had been cleared for agriculture in the 1950s (Clarke and Dickinson, 1995).

Establishing the detailed impact of disturbance from this particular study is a difficult task as so few plots are available and the analysis is preliminary. It can only be noted that the available data suggested that agricultural encroachment was the most dominant form of disturbance (affected four plots out of twelve). The very nature of this form of disturbance being total clearance associated with the introduction of an edge effect and increased risk of fire, this constitutes a major threatening impact. The regeneration data showed that pioneering species amply regenerated in formerly cultivated areas, and that tree recruitment was possible in areas where disturbances such as mining had taken place in the past, however, there were no indications for the re-establishment of forest dependent species. This might (i) be down to the fact that insufficient data was available, or (ii) indicate that the sample unit never supported such species, and/or (iii) that secondary forest re-growth is either lengthy or not possible.

5.5 Technical suggestions

In general, the analysis benefited from the outstanding field effort that Frontier Tanzania had undertaken. Despite the low number of workable plots, the fact that those plots were relatively large and that extensive data was available - most notably on topographic position, present disturbances, vegetation structure and regeneration quality – meant that extensive analysis was possible. While in this particular analysis it was generally not possible to distinguish between 'signal and noise', the high field effort will certainly pay off as more plot data become available. It is hoped that the following advice on sampling design and protocol might help to add further value to Frontier Tanzania's field data collection:

5.5.1 Sampling intensity, plot size and plot location

- With a total area of 0.012 km^2 being covered by vegetation plots, the **sampling intensity** was *c.* 0.35%[32]. This moderate to high sampling intensity appears perfectly adequate.

- As a general guideline, an adequate **plot size** for a tropical forest is 1000 - 10000 m^2. (e.g. Dierschke, 1994). While the plot size conformed to general recommendations, it is difficult to establish whether or not the plot size was sufficient in this particular case. A common approach to this issue is that of the 'minimum areal': starting with a small size, the plot area is consecutively doubled and any new species recorded. According to a general rule of thumb, the plot size is adequate when an area doubling yields 5% or less new species (e.g. Frey and Loesch, 2004). In a tropical forest this might not be achievable with reasonable effort (compare e.g. Schmidt, 1991), although the strong gradients prevalent in montane areas entail that the extent of homogeneous habitat and vegetation is somewhat smaller. If Frontier Tanzania was in a position to supply detailed spatial data on the distribution of the individuals within the plots, the minimum areal method could be applied retroperspectively.

- The **number of** (workable) **vegetation plots** was too small to capture the great variation in both floristic composition and vegetation structure. Presumably, a minimum of 30 plots would be needed to establish statistically meaningful results. This is a more pressing issue than the plots size, which might have been adequate. An alternative might be a nested approach that combines the establishment of large plots at subjectively chosen sites (where apparently there are different vegetation types) with smaller plots at randomly chosen sites (to capture the variation).

- The **regeneration plot size** was too small. As the assessment of regeneration might be of considerable interest, in particular in disturbed forests, it seems advisable to opt for a regeneration plot size of at least 10x10 m.

- Although the system is appealing in its clarity and efficiency, the authors would recommend to alter the current **model of work units** for two reasons: (1) the gradients captured tend to be small, and (2) both the vegetation composition and structure are subject to spatial autocorrelations. If field conditions allow, it would be more beneficial to locate the plots within farther distance from each other and across the entire forest. This would ensure that a maximum of variation is captured, which is particularly important in yet poorly known areas. Alternatively, several work units could be established, and if the sampling effort needs to be reduced, then the plots on the ends of the 'diagonal' transects could be spared.

5.5.2 Vegetation assessment

- As the height of the trees had not been assessed, the **canopy forming species** had to be deduced from the basal area. This was often ambiguous. The interpretation of the field data could be added to either by directly recording the canopy forming species (in the respective layers) or the individual heights of trees (e.g. in classes) in the field.

- It might be worthwhile to note the **density of lianas** as this might constitute an additional indicator for disturbances.

5.5.3 Assessment of environmental correlates and disturbances variables

- It was not possible to draw any conclusions on the regeneration plot **soil assessments** as only the substrate and substrate colour had been recorded. A darker soil colour might for instance indicate more moisture or a higher proportion of humic-acids indicating inhibited decomposition. In an area where the vegetation-soil relationships are as little known as they are in the Eastern Arc, it would be extremely insightful if additional analyses could be made on pH,

[32] Assuming that Mselezi has an area of c. 3.5 km^2, which had been established using ArcView GIS 3.3 and the 'hotspot profile' GIS data assembled by WWF.

moisture content, plant available phosphor and nitrogen content, calcium, potassium and cation exchange capacity of the upper soil layer. With the exception of calcium which could be estimated using 10% hydrochloric acid in the field this would require soil samples being taken to a laboratory. Such facilities are available at the Institute of Resource Assessment, University of Dar es Salaam.

- The assessment of **disturbances** should ideally be an integral part of the vegetation assessment. It was useful that the disturbance transects were located in line with the vegetation plots, however, for the plots on the 'diagonal' transects there was no such data. The authors would recommend that within a sample unit all stumps be routinely assessed (with dbh measurements and cutting age category), and an identification of the stump species be attempted. This might lead to useful conclusions about the particular types of disturbance (compare Ahrends, 2006). In addition, all other signs of disturbance should be recorded as already featured in the disturbance transects.
- It might further be useful to record all paths, tracks and roads that are encountered in the field during the survey work. This could be a useful **fragmentation** measure. Additional fragmentation measures are the circumference of the forest relative to its area and a combination of the extent of remaining primary forest with the distance between the prevailing primary forest parts.
- Finally, the authors would recommend to also assess the plots **distance to the nearest road and nearest village** as this might allow further conclusions about the forest disturbance patterns.

References

AHRENDS, A. (2006). *Patterns of degradation in lowland coastal forests in Coast Region, Tanzania.* A thesis submitted in candidacy for the degree of MSc. University of Greifswald

MINISTRY OF NATURAL RESOURCES AND TOURISM. FOREST AND BEEKEEPING DIVISION. GOVERNMENT OF TANZANIA (2005). *Prepared by*: Burgess, N.D., Butynski, T.M., CordeiroN., Doggart N.H., Fjeldså, J., Howell, K.M., Kilahama, F.B., Loador, S.B., Lovett, J.C., Mbilinyi B., Menegon, M., Moyer, D.C., Nashanda, E., Perkin, A., Rovero, F., Stanley, W.T., Stuart, S.N.. *Biodiversity baseline for the Eastern Arc Mountains of Tanzania and Kenya.* Unpublished report. Dar es Salaam

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

BRAUN-BLANQUET, J. (1964). *Pflanzensoziologie.* Springer. Wien.

CLARKE, G.P. (1995a). *Checklist of the Vascular Plants from 13 Coastal Forests in Tanzania*. Frontier-Tanzania Technical Report No.15 (unpublished). The Society for Environmental Exploration and The University of Dar es Salaam. Dar es Salaam

CLARKE, G.P. (1995b). *Status Reports for 6 Coastal Forests in Lindi Region, Tanzania*. Frontier Tanzania Technical Report No. 18. The Society for Environmental Exploration and The University of Dar es Salaam. Report

CLARKE, G.P. (1998). A new regional centre of endemism in Africa. In: HUXLEY, C.R., LOCK, J.M. AND CUTLER, D.F. (eds.). *Chorology, tanxonomy and ecology of the floras of Africa and Madagascar*. Royal Botanic Gardens, Kew

CLARKE, G.P. AND DICKINSON, A. (1995). *Status Reports for 11 Coastal Forests in Coast Region, Tanzania*. Frontier Tanzania Technical Report No. 17. The Society for Environmental Exploration and The University of Dar es Salaam. Report

CLARKE, G.P. AND ROBERTSON, S.A. (2000). Vegetation communities. In: BURGESS, N.D. AND CLARKE, G.P. (eds.) (2000). *Coastal Forests of Eastern Africa*. IUCN Publication Services Unit. Cambridge

CLARKE, G.P. AND STUBBLEFIELD, L.K. (1995). *Status Reports for 7 Coastal Forests in Tanga Region, Tanzania*. Frontier Tanzania Technical Report No. 16. The Society for Environmental Exploration and The University of Dar es Salaam. Report

DYTHAM, C. (1999). *Choosing and Using Statistics – A Biologist's Guide*. Blackwell Science Ltd. Oxford

ELLENBERG, H. (1992). *Zeigerwerte von Pflanzen in Mitteleuropa / Indicator values of plants in Central Europe.* Second Edition. Goltze. Göttingen

FREY, W. AND LOESCH, R. (2004). *Lehrbuch der Geobotanik: Pflanze in Raum und Zeit*. 2nd Ed. Spektrum Akademischer Verlag. Heidelberg.

GEREAU, R. AND LUKE, Q.L. (2006). *List of Potentially Threatened Plants in the Eastern Arc Mountains and Coastal Forest Biodiversity Hotspot of Kenya and Tanzania.* Unpublished report

GOODALL, D. W. (1954). Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Australian Journal of Botany* **2**: 304-324.

GREENWAY, P.J. (1973). A classification of the vegetation of East Africa. Kirkia.

HEMP, ANDREAS (2005). Climate change-driven forest fires marginalize the impact of ice cap wasting on Kilimanjaro. *Global Change Biology* **11**: 1013

HEMP, ANDREAS (2006). The Banana Forests of Kilimanjaro: Biodiversity and Conservation of the Chagga Homegardens. *Biodiversity and Conservation* **15**: 1193-1217.

HILL, M. O. (1973). Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology* **61**: 237-249

HILL, M. O. (1979). DECORANA - a FORTRAN program for detrended correspondence analysis and reciprocal averaging. *In*: *Ecology and Systematics*. Cornell University. Ithaca

HILL, M. O. AND GAUCH, H.G. (1980). Detrended correspondence analysis: an improved ordination technique. *Vegetatio* **42**: 47-58

JONGMAN, R.H.G., TER BRAAK, C.J.F. AND VAN TONGEREN, O.F.R. (1995). *Data Analysis in Community and Landscape Ecology*. Cambridge University Press. Cambridge

KOPP, D. AND SCHWANECKE, W. (2003). *Standörtlich-naturräumliche Grundlagen ökologiegerechter Forstwirtschaft : Grundzüge von Verfahren und Ergebnissen der forstlichen Standortserkundung in den fünf ostdeutschen Bundesländern.* Kessel. Remagen

LOVETT, J.C., HANSEN, J.R. AND HØRLYCK, V. (2000). Comparison with Eastern Arc Forests. BURGESS, N.D. AND CLARKE, G.P. (eds.) (2000). *Coastal Forests of Eastern Africa*. IUCN Publication Services Unit. Cambridge

LOVETT, J.C., RUFFO, C.K., GEREAU, R.E. AND TAPLIN, J.R.D. (*in press*). F*ield guide to the moist forest trees of Tanzania*. Currently accessible at URL :
http://www.york.ac.uk/res/celp/webpages/projects/ecology/tree%20guide/introduction.htm

LOWE, A.J. AND CLARKE, G.P. (2000). Vegetation structure. In: BURGESS, N.D. AND CLARKE, G.P. (eds.) (2000). *Coastal Forests of Eastern Africa*. IUCN Publication Services Unit. Cambridge

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

MCCUNE, B. AND MEFFORD, M.J. (1999). PC-ORD. *Multivariate Analysis of Ecological Data, Version 4.* MjM Software Design. Oregon

OLDFIELD, S., LUSTY, C., MACKINVEN, A. (eds.) (1998). *The World List of Threatened Trees.* Compiled by the World Conservation Monitoring Centre. World Conservation Press. Cambridge

SCHMIDT, R. (1991). Ecology of a tropical lowland rain forest – Plant communities, soil characteristics and nutrients relations of the forests in the Shimba Hills National Reserve, Kenya. *Dissertationes Botanicae* **179**, J. Cramer in der Gebrüder Borntraeger Verlagsbuchhandlung. Berlin – Stuttgart

SUCCOW, M. AND JOOSTEN, H. (2001). *Landschaftsökologische Moorkunde : mit 136 Tabellen.* Second Edition. Schweizerbart. Stuttgart

TURRILL, W.B. AND MILNE-REDHEAD, E. ET AL. (eds.) (1952-). *The Flora of Tropical East Africa*. A.A. Balkema, Rotterdam

WHITE, F. (1983). *The Vegetation of Africa*. UNESCO, Paris

WHITEHEAD, J. *An introduction to logistic regression*. Department of Economics, Appalachian State University. http://personal.ecu.edu.whiteheadj/data/logit/

ZAR, J.H. (1999). *Biostatistical Analysis – Fourth Edition*. Prentice Hall International. New Jersey

Appendix A

Summary and numerical assessment of variables supplied to the analysis

ENVIRONMENTAL CORRELATES:

- **altitude:** as measured in the field and as different classes (50 m and 100 m intervals)
- **slope inclination:** as measured in the field and as classes (5 degree intervals)
- **position on the slope:** coded as 0 = valley floor ; 1 = gentle; 2 = moderate; 3 = steep
- **aspect:** coded as 1 = eastern or south-eastern aspect; 0 = rest

DISTURBANCE MEASURES:

- **presence/absence of signs of past use:** coded as 1 = presence; 0 = absence
- **presence/absence of present use:** coded as 1 = presence; 0 = absence
- **weighted combined use measurement:** coded as 2 = present use; 1 = equal past use; 0 = no use
- **number of cut poles:** measured as the number of cut poles in plot itself and two adjacent plots
- **number of cut timbers:** measured as the number of cut timbers in plot itself and two adjacent plots
- **number of overall cuts:** measured as the number of cuts in plot itself and two adjacent plots
- **overall number of disturbances:** measured as simple count of disturbance present
- **presence/absence of particular types of disturbances:** coded as 1 = presence; 0 = absence
- **presence/absence of roads or tracks:** coded as 1 = presence; 0 = absence

OTHER FEATURES OF INTEREST:

- **water association:** coded as 1 = presence; 0 = absence
- **presence/absence of rocky outcrops:** coded as 1 = presence; 0 = absence

VEGETATION STRUCTURE RELATED VARIABLES:

- **standing density of trees >10 cm dbh**
- **standing density of trees >20 cm dbh**
- **average dbh of trees >10 cm dbh**
- **average dbh of trees >20 cm dbh**
- **basal area**
- **canopy cover; shrub cover; ground cover:** coded as 1 = <10%; 2 = 10-50%; 3 = >50%
- **canopy height:** coded as 1 = <10 m;2 = 10-20 m; 3 = 20-30 m; 4 = >30 m
- **number of dead trees:** measured as the number of dead trees in plot itself and two adjacent plots

SPECIES COMPOSITION RELATED VARIABLES:

- **species abundance:** measured as simple frequency count and abundance categories (pseudo-species in TWINSPAN)
- **species dominance:** measured as species basal area
- **species combined abundance/dominance:** see Tab. 1
- **species indicator value:** coded as 1 = forest specialist; 2 = forest generalist; 3 = predominantly grassland/woodland species
- **species disturbance resistance:** measured as 0 = not known to be resistant; 1 = resistant

- **species fire resistance:** measured as 0 = not known to be resistant; 1 = resistant

REGENERATION STRUCTURE RELATED VARIABLES:

- **cover with herbaceous vegetation:** measured as percentage
- **cover with litter:** measured as percentage
- **cover with stones:** measured as percentage
- **extent of bare ground:** measured as percentage
- **degree of grass dominance:** measured as percentage of grass cover in herbaceous layer

REGENERATION SPECIES COMPOSITION RELATED VARIABLES:

- **regeneration stability:** coded as 0 = regeneration of non-canopy species; 1 = canopy or sub-canopy species regenerating
- **species indicator value:** *as above in species composition related variables*
- **proportionate share of pioneering species**: measured as the share of pioneering species on the total number of species in regeneration

SPECIES RICHNESS AND DIVERSITY VARIABLES:

- **values for Berger-Parker index**
- **values for Simpson's index**
- **values for Shannon index**
- **values for Evenness index**
- **total number of species**

Appendix B

Species and their ecological requirements and distributions

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res.[35]	moi. ind.[35]	reg.[36]
Acanthaceae	Mellera	lobulata	Mellob		Forest only.	n/a	n/a		1				1
Anacardiaceae	Sorindeia	madagascariensis	Sormad	Riverine, lowland, submontane and montane forest.	Forest only.	Riverine forest, groundwater forest, on the coast also in forest not close to water.	C, EA, N, LN. South Eastern tropical Africa. Mascarenes. Madagascar.		1			1	1
Annonaceae	Annona	senegalensis	Annsen	Dry lowland forest. Woodland. Grassland. A woodland species that also occurs in open areas in forest.	Forest, woodland, thicket, grassland.	Wooded or bushed grassland, (riverine) woodland, secondary (fire-induced) bushland, on the coast also in evergreen forest.	Widespread in Tropical Africa, also in Madagascar and Comoro Islands.		3	1	1		
Annonaceae	Lettowianthus		Letspe										
Annonaceae	Lettowianthus	stellatus	Letste	Dry lowland and lowland forest. Riverine forest and woodland.	Forest, riverine forest, woodland.	n/a	C, EA.	x	2				
Annonaceae	Xylopia		Xylspe										1

[33] For key to distribution codes see end of table
[34] According to Beentje (1994); Burgess and Clarke (2000); Clarke (1995a); Lovett et al. (*in press*)
[35] Indicator value, disturbance resistance, fire resistance and moisture indicator (coded as in Appendix A)
[36] 1 = present in regeneration

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res. [35]	moi. ind. [35]	reg.[36]
Apocynaceae	Tabernaemontana	odoratissima	Tabodo	Riverine and lowland forest.	n/a	n/a	EA (Ma, Udz). Uganda, Central Africa.		2				
Araceae	Amorphophallus		Amospe										1
Bignonaceae	Stereospermum	kunthianum	Stekun	n/a	Forest, forest edge, woodland.	Rocky bushland, wooded grassland, on the coast always in forest (margins) and secondary bush.			2	1			
Bignoniaceae	Markhamia		Marspe										
Bignoniaceae	Markhamia	lutea	Marlut	Submontane forest.	n/a	(Riverine) forest (remnants).	C, EA, LT, LV. Widespread Tropical Africa.		1				1
Bignoniaceae	Markhamia	zanzibarica	Marzan	Dry lowland forest. Riverine. Woodland. Grassland.	Forest, riverine forest, woodland, bush.	Dry forest and secondary bush.	C, LT. Southern and Eastern Africa.		2	1			
Bombaceae	Bombax	rhodognaphalon	Bomrho	Dry lowland and lowland forest. Riverine. Coastal thicket.	Forest, woodland, bush, thicket, grassland.	Evergreen forest (margins, remnants), coastal bushland.	C, EA, LN. Eastern and Southeastern Tropical Africa.	x	2				
Boraginaceae	Cordia	africana	Corafr	Riverine, groundwater, dry montane and secondary forest. Grassland.	n/a	Wooded grassland, forest, riverine.	C, EA, N, LN, LT.		3				
Burseraceae	Commiphora		Comspe										

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res.[35]	moi. ind.[35]	reg.[36]
Burseraceae	Commiphora	africana	Comafr	n/a	Not forest.	*Distinguishes several varieties with differing ecological requirement.s*			3				
Combretaceae	Combretum		Comspe										
Combretaceae	Combretum	pentagonum	Compen	n/a	Forest, forest edge, thicket.	Coastal evergreen bushland and forest.			2				1
Combretaceae	Pteleopsis	myrtifolia	Ptemyr	Riverine, dry lowland forest. Woodland. Thicket.	Forest, riverine forest woodland, bush, grassland.	n/a	C, EA. Eastern and Southern Africa.		2				
Euphorbiaceae	Acalypha		Acaspe										1
Euphorbiaceae	Antidesma	venosum	Antven	Riverine and dry lowland forest. Woodland. Grassland. Thicket	Forest, riverine forest, forest edge, woodland, bush.	Wooded grassland, secondary bushland at forest edge, riverine forest, moist forest.	C, EA, N, LN, LT, LV. Tropical and Southern Africa.		2	1			
Euphorbiaceae	Bridelia	cathactica	Brican	n/a	Riverine forest, woodland, bush, thicket.	Forest margins, (secondary) bushland, littoral thicket.			2	1			
Euphorbiaceae	Bridelia	micrantha	Brimic	Edges and pioneer of lowland, montane, dry montane and riverine forest. Woodland.	Forest, forest edge, bush, thicket, shrub.	Usually riverine forest and in forest margins, less often in bushed for wooded grassland.	C, EA, N, LN, LT, LV, Tropical and Southern Africa.		2	1			

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res.[35]	moi. ind.[35]	reg.[36]
Euphorbiaceae	Drypetes	reticulata	Dryret	Riverine, dry lowland forest. Thicket.	0-500 m, forest, riverine forest, thicket.	Evergreen or semideciduous forest or on exposed coral.	C, EA (Ul). Eastern and Southern Africa.		1				
Euphorbiaceae	Flueggea	virosa	Fluvir	n/a	Riverine forest, forest edge, bushland, thicket, wasteland.	Riparian, in rocky bushland/bushed grassland, in wooded grassland; less often (western and coastal areas) in forest margins; also on black cotton soil.			2	1			1
Euphorbiaceae	Margaritaria	discoidea	Mardis	Dry lowland, montane and dry montane forest.	Forest, forest edge, woodland.	Moist or dry forest (margins) or forest remnants.	C, EA, N, LN, LT, LV. Tropical and Southern Africa.		2				1
Euphorbiaceae	Ricinodendron		Ricspe										
Lamiaceae	Hoslundia	opposita	Hosopp	n/a	Forest, riverine forest, forest edge, bush, wasteland.	(Secondary) bushland, bushed or wooded grassland, forest margins and disturbed habitats, no in very dry localities.			2	1			
Fabaceae: Caesalpinioideae	Cassia		Casspe										

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res. [35]	moi. ind. [35]	reg.[36]
Fabaceae: Caesalpinioideae	Piliostigma	thonningii	Piltho	n/a	n/a	(Combretum) wooded grassland or scattered tree grassland, often common or dominant.			3				
Fabaceae: Mimosoideae	Albizia		Albspe										
Fabaceae: Papilionoideae	Dalbergia		Dalspe										
Fabaceae: Papilionoideae	Dalbergia	melanoxylon	Dalmel	n/a	n/a	Deciduous woodland or - bushland, wooded grassland, often in rocky sites or on black cotton soil.			3				
Fabaceae: Papilionoideae	Indigofera		Indspe										1
Fabaceae: Papilionoideae	Lonchocarpus		Lonspe										
Fabaceae: Papilionoideae	Lonchocarpus	bussei	Lonbus	n/a	n/a	Wooded (palm) grassland, woodland, wooded bushland, thicket on dunes.			3				
Fabaceae: Papilionoideae	Millettia		Milspe										
Fabaceae: Papilionoideae	Pericopsis	angolensis	Perang	n/a	n/a	n/a							
Fabaceae: Papilionoideae	Xeroderris	stuhlmannii	Xerstu	n/a	n/a	n/a			2				
Meliaceae	Trichilia	emetica	Trieme	Riverine, dry lowland forest. Woodland.	Forest, riverine forest, woodland.	Riverine, or in sites with high groundwater.	C, EA, N, LN, LT, LV. Widespread in Africa.		2			1	

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res.[35]	moi. ind.[35]	reg.[36]
Meliaceae	Turraea	holstii	Turhol	Lowland, submontane and montane forests.	n/a	Forest.	C, EA, N, LN. Eastern Africa, Arabian Peninsula.		1				
Meliaceae	Turraea	mombassana	Turmom	n/a	Forest edge, woodland.	Dry forest (margins), semi-evergreen bushland.			2				
Moraceae	Ficus		Ficspe										
Moraceae	Ficus	sur	Ficsur	Riverine, lowland, submontane and montane forest. Left in cleared areas.	Forest, forest edge, riverine forest, grassland.	Riverine forest and bush, groundwater forest, less often in forest away from water.	C, EA, N, LN, LT, LV. Tropical and Southern Africa. Yemen.		2			1	
Moraceae	Ficus	sycomorus	Ficsyc	Riverine, lowland forest edge. Woodland	Riverine forest, forest edge, woodland, shrub.	Riparian, or in places with high groundwater table, ? Also in forest or bushland.	C, EA, N, LN, LT, LV. Eastern and Southern Africa. Arabian Peninsula, Madagascar, Comoros.		2			1	
Moraceae	Trilepisium	madagascariense	Trimad	Riverine, groundwater, lowland and submontane forests.	Forest only.	Moist forest.	C, EA, N, LN, LV. Tropical and Southern Africa. Madagascar, Seychelles		1			1	1

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res.[35]	moi. ind.[35]	reg.[36]
Myrtaceae	Syzygium	guineense	Syzgui	four subspecies	Forest only.	Ssp. guineense: riverine or in wooded grassland (1-2550m); ssp. afromontanum: forest (1500-2550m).	EA, N, LN, LT. Eastern, Central, and Southern Tropical Africa.						1
Olacaceae	Strombosia	scheffleri	Strsch	Submontane and montane forest.		Moist forest, sometimes dominant.	EA, N, LN, LT, LV. Tropical Africa.		1				
Onagraceae	Ludwigia	abyssinica	Ludaby	n/a	n/a	n/a							
Rubiaceae	Leptactina	platyphylla	Leppla	Lowland and submontane forest. Woodland.	Forest, woodland, bush.	n/a	C, EA, LT, LV. Eastern, Central and Southern Tropical Africa.		2				
Rubiaceae	Pavetta	crebrifolia	Pavcre	n/a	Forest , bushland.	Forest, bushland or littoral thicket.			2				1
Rubiaceae	Tricalysia		Trispe										1
Rubiaceae	Vangueria	infausta	Vaninf	n/a	n/a	Riverine forest or -woodland, rocky bushland or thickets.			2				
Rutaceae	Zanthoxylum	leprieurii	Zanlep	Submontane and montane forest.	n/a	n/a	C, EA, LV. Tropical and Southern Africa.		2				
Sapindaceae	Deinbollia	borbonica	Deibor	n/a	Forest, riverine forest, woodland, bushland, thicket.	Forest, secondary bush, evergreen coastal thicket, riverine bush.			2	1			1
Sapindaceae	Majidea		Majspe										1

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res.[35]	moi. ind.[35]	reg.[36]
Sapindaceae	Pancovia		Panspe										1
Sapotaceae	Englerophytum	magalismontanum	Engmag	Dry lowland, dry montane and riverine forest. On termite mounds in woodland.	n/a	n/a	C, EA, LN, LT. Tropical and Southern Africa.		2				1
Simaroubacaeae	Harrisonia	abyssinica	Haraby	Dry lowland and riverine forest. Woodland. Thicket.	Forest, riverine forest, forest edge, bush, thicket, grassland.	Dry bushland, wooded grassland, or riverine; on the coast also in forest margins.	C, EA, N, LN, LT, LV. Tropical Africa.		2				1
Sterculiaceae	Cola		Colspe										
Sterculiaceae	Dombeya	shupangae	Domshu	n/a	n/a	n/a							1
Sterculiaceae	Sterculia	appendiculata	Steapp	Lowland and riverine forest. Woodland.	Forest only.	(Riverine) forest.	C. Eastern Tropical Africa.		1				
Tiliaceae	Grewia		Grespe										
Tiliaceae	Triumfetta	tomentosa	Tritom	n/a	n/a	(Moist) forest margins, riverine forest; often in secondary vegetation in forest areas.			2	1			1
Ulmaceae	Celtis		Celspe										
Ulmaceae	Chaetacme	aristata	Chaari	Lowland and submontane forest edges, riverine forest.	n/a	Riverine (in forest or bushland), also in evergreen forest (edges).	EA, N, LT, LV. Tropical and Southern Africa and Madagascar.		2				

Family	Genus	Species	Species code	Ecology according to Lovett et al. (*in press*)	Ecology according to Clarke (1995a)	Ecology according to Beentje (1994)	distribution according to Lovett et al. (*in press*)[33]	(near) endemic[34]	ind. val.[35]	dist. res.[35]	fire res.[35]	moi. ind.[35]	reg.[36]
Ulmaceae	Trema	orientalis	Treori	Pioneer in riverine, lowland, submontane and montane forest.	Forst only.	Forest margines, riverine (secondary) bushland, woodland, wooded grassland, a pioneer where forest has been disturbed.	C, EA, N, LN, LT, LV. Tropical and Southern Africa to Madagascar and Asia.		2	1			
Verbenaceae	Lippia	javanica	Lipjav	n/a	n/a	Locally abundant in secondary bushland or grassland; less often in wooded grassland.			3	1			1
Verbenaceae	Premna		Prespe										
Verbenaceae	Vitex	doniana	Vitdon	n/a	Forest, woodland, grassland.	Wooded grassland or forest edge.			3				

Key for 'distribution according to Lovett et al. (*in press*):
Coastal (C). Eastern Arc (EA). Northern (N). Lake Nyasa (LN). Lake Tanganyika (LT). Lake Victoria (LV). Mountains (north to south): Teita Hills (Te), Pare (P), Usambara (Us), East Usambara (EUs), West Usambara (Wus), Northern Nguru (NNg), Southern Nguru (SNg), Nguru (Ng), Uluguru (Ul), Malundwe (Mal), Udzungwa (Udz), Mahenge (Ma).

Appendix C

Floristic affinity to other forests

	EUS and WUS[37]	Khi and Muf[37]	Maz[37]	Ngu[37]	San[37]	Cho[38]	Tai[39]	Ulu[40]	Udz[39]	Ks[41]	Ki[41]	Mch[41]	Na[41]	Ng[41]	Pa[41]	Pu[41]	Ru[41]	Vi[41]	Genus	Species
	150	45	66	62	145	77	318	112	320	7	105	41	94	104	59	131	71	4		
								1			1	1	1			1			Annona	senegalensis
											1			1		1	1		Antidesma	venosum
	1			1	1						1		1	1	1	1	1		Bombax	rhodognaphalon
													1	1		1	1		Bridelia	cathactica
	1		1	1		1		1	1		1		1	1		1	1		Bridelia	micrantha
																			Chaetacme	aristata
										1	1		1		1		1		Commiphora	africana
	1																		Cordia	africana
													1	1	1	1	1		Dalbergia	melanoxylon
	1															1			Drypetes	reticulata
																			Englerophytum	magalismontanum
	1		1	1	1		1	1	1										Ficus	sur
																1			Ficus	sycomorus
																			Harrisonia	abyssinica
									1					1	1	1	1		Hoslundia	opposita
					1				1										Leptactina	platyphylla
					1				1		1		1						Lettowianthus	stellatus
																			Lonchocarpus	bussei
																			Ludwigia	abyssinica

[37] Data J.C. Lovett
[38] Data Sokoine University
[39] Data H. Beentje
[40] Data Frontier Tanzania
[41] Data A. Ahrends

	EUS and WUS[37]	Khi and Muf[37]	Maz[37]	Ngu[37]	San[37]	Cho[38]	Tai[39]	Ulu[40]	Udz[39]	Ks[41]	Ki[41]	Mch[41]	Na[41]	Ng[41]	Pa[41]	Pu[41]	Ru[41]	Vi[41]	Genus	Species
																1			Markhamia	zanzibarica
									1		1		1	1			1		Pericopsis	angolensis
																			Piliostigma	thonningii
					1						1		1	1	1		1		Pteleopsis	myrtifolia
	1	1	1		1	1	1		1		1	1	1	1	1	1	1		Sorindeia	madagascariensis
	1																		Sterculia	appendiculata
					1				1		1		1			1			Stereospermum	kunthianum
	1	1	1	1	1	1	1												Strombosia	scheffleri
	1		1	1	1	1		1								1			Syzygium	guineense
																			Tabernaemontana	odoratissima
	1					1	1									1			Trema	orientalis
						1					1	1	1		1				Trichilia	emetica
	1	1	1	1	1														Trilepisium	madagascariense
							1		1							1			Turraea	holstii
																			Turraea	mombassana
									1		1		1	1		1	1		Vangueria	infausta
								1											Vitex	doniana
											1		1	1					Xeroderris	stuhlmannii
																			Zanthoxylum	leprieurii
shared species	11	3	6	6	10	6	5	5	10	1	13	3	14	11	7	16	11	0		
SJ	0,06	0,03	0,05	0,06	0,05	0,05	0,01	0,03	0,03	0,02	0,08	0,04	0,1	0,07	0,07	0,09	0,09	0		
CC	0,12	0,07	0,12	0,12	0,11	0,1	0,03	0,07	0,06	0,04	0,18	0,08	0,21	0,15	0,14	0,19	0,2	0		

Colour key: blue: lowland species; grey: montane species; white: species with overlapping distribution

Abbreviation key: EUS and WUS = East Usambara and West Usambara; Khi and Muf = Khihansi and Mufindi; Maz = Mazumbai; Ngu = Nguru; San = Sanje; Cho = Chome; Tai = Taita; Ulu = Uluguru; Usa = Usambara, Udz = Udzunga; Ks = Kisiju; Ki = Kiwengoma; Mch = Mchungu; Na = Namakutwa; Ng = Ngumburuni; Pa = Pande; Pu = Pugu; Vi = Vikindu