



An interpolated biogeographical framework for tropical Africa using plant species distributions and the physical environment

Cicely A.M. Marshall¹  | Jan J. Wieringa²  | William D. Hawthorne³ 

¹Department of Plant Sciences, University of Cambridge, Cambridge, UK

²Naturalis Biodiversity Center, National Herbarium of the Netherlands, Leiden, The Netherlands

³Department of Plant Sciences, University of Oxford, Oxford, UK

Correspondence

Cicely A. M. Marshall, ¹Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, UK.
Email: cm997@cam.ac.uk

Funding information

Merton College, University of Oxford; James Martin 21st Century School, University of Oxford; Clarendon Fund, University of Oxford; Dutch Economic Structure Enhancement Fund; King's College Cambridge, University of Cambridge

Handling Editor: Ole Vetaas

Abstract

Aim: Existing phytogeographical frameworks for tropical Africa lack either spatial completeness, unit definitions smaller than the regional scale or a quantitative approach. We investigate whether physical environmental variables can be used to interpolate floristically defined vegetation units, presenting an interpolated, hierarchical, quantitative phytogeographical framework for tropical Africa, which is compared to previously defined regions.

Location: Tropical mainland Africa 24°N to 24°S.

Taxon: 31,046 vascular plant species and infraspecific taxa.

Methods: We calculate a betasim dissimilarity matrix from a comprehensive whole-flora database of plant species distributions. We investigate environmental correlates of floristic turnover with local non-metric multidimensional scaling. We derive a hierarchical biogeographical framework by clustering the dissimilarity matrix. The framework is modelled using a classification decision tree method and 12 physical environmental variables to interpolate and increase the resolution of the framework across the study region.

Results: Floristic turnover is related strongly to water availability and temperature, with smaller contributions from land cover, topographic ruggedness and lithology. Region can be predicted with 90% accuracy by the model. We define 19 regions and 99 districts. We find a novel arrangement of the arid regions. Regional subdivision within the savanna biome is supported with minor variation to borders. Within the forests of west and central Africa, our whole-flora gridded regionalization supports the divisions identified by a previous analysis of trees only.

Main conclusions: Physical environmental variables can be used to predict floristically defined vegetation units with very high accuracy, and the approach could be pursued for other incompletely sampled taxa and areas outside of tropical Africa. Geographical coherence is higher than in previous quantitative phytoregional definitions. For most tropical African vascular plant species, we provide predictions of which species will occur within each mapped district and region of tropical Africa. The framework should be useful for future studies in ecology, evolution and conservation.

KEYWORDS

biogeographical region, CART, cluster analysis, ecoregion, floristic district, phytogeography, plants, Random Forest, species distribution, tropical Africa, vegetation classification

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Biogeography* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

By defining areas of relatively homogeneous species composition, biogeographical frameworks provide spatial units of analysis that are useful in macroecological, evolutionary and systematic studies of the processes which shaped or maintain species distributions (Morrone, 2018). We can use theory generated by such studies to improve the biogeographical framework itself, predicting where species are likely to occur for areas which are incompletely collected, providing increasingly complete geographical and taxonomic coverage (Holt et al., 2013). By predicting which species occur in which place, biogeographical frameworks are useful tools in conservation and land management: the units are convenient for monitoring trends in species abundance, range or habitat condition (Dauby et al., 2017), can provide shortlists of native species for restoration (Lillesø et al., 2011), and facilitate geographical searching of online identification resources (Brunken et al., 2008).

Africa has both a rich history and contemporary research front in plant biogeography. The most influential framework for African vegetation is that of White (White, 1979, 1983, 1993). Building on the work of others (Aubreville, 1962; Lebrun, 1960, 1961; Monod, 1957), White used species distributions of a subset of the flora to define a chorological framework at regional scale. Where others had used thresholds of endemism at successively finer taxonomic ranks to distinguish choria at different ranks (Engler, 1879; Good, 1947; Takhtajan, 1986), endemism thresholds were important to White only for distinguishing between *types* of region, and did not preclude the recognition of species poor or endemic poor choria at the rank of region. White overlaid on his regional chorology, a detailed map of vegetation units (1983), synthesizing many previous local vegetation maps. These mapping units are labelled with respect to their structure as well as the regional chorological situation (e.g. Congolian swamp forest). WWF's African ecoregions were derived from White's physiognomic vegetation map, though the ecoregion borders were simplified and differ where they were redrawn to respect animal distributions (Olson et al., 2001). More recently, the local vegetation maps that White synthesized have been resurrected and digitized for east Africa (Lillesø et al., 2011).

White's regional framework was produced from limited species distribution datasets and without the aid of multivariate statistics, prompting six quantitative assessments of tropical Africa's phyto-geographical regions (Denys, 1980; Droissart et al., 2018; Fayolle et al., 2014, 2019; Linder et al., 2005, 2012). All have shown broad similarities to White's regions, but with discrepancies from each other and White. Most have analysed gridded distribution data of species and infraspecific taxa at one degree square resolution; Fayolle et al. analysed local checklists and plot data for trees (2014) or woody taxa (2019). For the gridded analyses, there has been a trend towards more taxonomically complete datasets: 494 taxa (Denys, 1980), 5,438 (Linder et al., 2005), 5,881 (Linder et al., 2012), 24,719 (Droissart et al., 2018); the present dataset includes 31,046 species and infraspecific taxa. Most have used ordination and cluster analysis to produce regionalizations, as in the present analysis, while

Droissart et al. used bipartite network analysis (Edler, Guedes, Zizka, Rosvall, & Antonelli, 2017). These quantitative floristic frameworks have left much of the area of tropical Africa unassigned to a spatial unit, due to the insufficiency of plant species distribution data from many parts of the continent, and because there is currently no comprehensive set of species distribution models for African plants. A lack of geographical completeness makes it difficult to use the quantitative phytoregionalizations as spatial frameworks in further analyses, as they apply only to the idiosyncratic portion of the continent that was used to derive them. For tropical Africa, we also lack quantitatively defined floristic units similar in size to ecoregions. Using lower cluster solutions to define districts or provinces below regional level isprecedented for African birds (De Klerk, Crowe, Fjeldså, & Burgess, 2002), and for plants a hierarchical system has been described as a useful and natural way to depict floristic relationships (McLaughlin, 1992).

Taxonomic information on biological composition must be derived from field survey, and is more limited in spatial and temporal resolution, and geographical coverage, than physical environmental variables like climate, which have been interpolated at high resolution across the globe (Kriticos et al., 2012). Land cover maps derived from remotely sensed data, like GlobCover 2009 (Arino et al., 2012), have provided 90m resolution continuous characterizations of vegetation physiognomy, in terms of canopy height, openness, and deciduousness, for the world. For Africa, there is also A New Map of Standardised Terrestrial Ecosystems (Sayre et al., 2013), presenting a predicted vegetation classification at 90 m resolution across the continent using a classification and regression tree (CART) method. The vegetation classification that was predicted was described as a draft. Training points were supplied by experts or derived from previous maps, and were reconciled into a hierarchical physiognomic framework following principles developed for the USA.

The prediction of vegetation physiognomy from physical environmental variables, especially climate, is a long established and current activity (Arino et al., 2012; Holdridge, 1947). Some African plant species distributions have been predicted with physical environmental variables via species distribution models (Deblauwe et al., 2016; McClean et al., 2005). Plant communities in Africa are typically ordinated and correlated with physical environmental variables, to reveal strong relationships with rainfall (Bongers, Poorter, & Hawthorne, 2004; Fayolle et al., 2014, 2019; Hall & Swaine, 1976), lithology (Fayolle et al., 2012), soil (Réjou-Méchain et al., 2008), or a combination of soil and rainfall (Swaine, 1996), temperature and altitude (Fayolle et al., 2014, 2019). Within climatic envelopes, fire, biotic interactions and feedback processes are important (Favier et al., 2012). The manner and extent to which floristically defined units can be predicted by physical environmental variables is of great interest (Tuomisto, Ruokolainen, & Yli-Halla, 2003). If the predictive relationship is strong, we could use the relationship to (a) 'complete' our maps, filling holes left by patchy species distribution data; (b) increase the resolution of the mapped units; (c) interpret the relative importance of particular physical environmental variables for determining floristic turnover and (d) make the relationship

between physiognomy, environment and chorology more explicit and objective.

The aim of this study is to present a spatially complete quantitative phytogeographical framework for tropical mainland continental Africa (24°N to 24°S) that can predict which plant species occur at regional to more local scale across the study area, useful for future studies in ecology, evolution and conservation, and to facilitate (on-line) identification tools. We define a hierarchical phytogeographical framework via well-established multivariate methods from a comprehensive whole-flora assemblage of available plant species distribution data. We investigate the strength of the predictive relationship between physical environmental variables and these floristically defined spatial units, seeking to use this relationship to render the framework geographically complete within its scope. We ask:

1. Can physical environmental variables be used to predict floristically defined vegetation units?
2. How do our phytogeographical regions compare with previously defined regions?

2 | MATERIALS AND METHODS

2.1 | Species distribution data

We analyse 533,383 records of 31,046 tropical African species and infraspecific taxa in 1,197 degree squares of tropical mainland Africa between 24°N and 24°S. To prepare the dataset, the TRAFRICA dataset (Marshall, Wieringa, & Hawthorne, 2016) was supplemented with the FLOTROP dataset (Taugourdeau et al., 2019) and the RAINBIO dataset (Dauby et al., 2016), and is the largest yet compiled for tropical Africa. Species names and synonymy follow the taxonomic backbone of the TRAFRICA database, derived initially from the tropical African section of the African Plants Database (Conservatoire et Jardin botaniques de la Ville de Genève and South African National Biodiversity Institute, Pretoria, 2016). All distribution record identifications were updated against this framework for analysis. Records were checked and cleaned for geographical errors using custom Microsoft Visual FoxPro routines. Each record's supplied coordinates, if any, were compared against supplied textual locality information, with contradictions resolved or the record excluded. Records with textual locality information and without coordinates were parsed to the bounding boxes of the stated locality. We checked all records from centres of administrative areas, all records for any locality with > 25 records and suspiciously round coordinates. The geographical resolution of the record was respected via the use of bounding boxes: unlikely coordinates were removed in favour of administrative polygon geolocation, and records which straddled one degree squares were dropped. Distribution data were summarized at the commonly used resolution of one degree square; the difference in area between the largest and smallest degree square is not large (8.8%). We excluded from analysis vague names, hybrids, cultivars and taxa we knew to be introduced or cultivated. Analyses

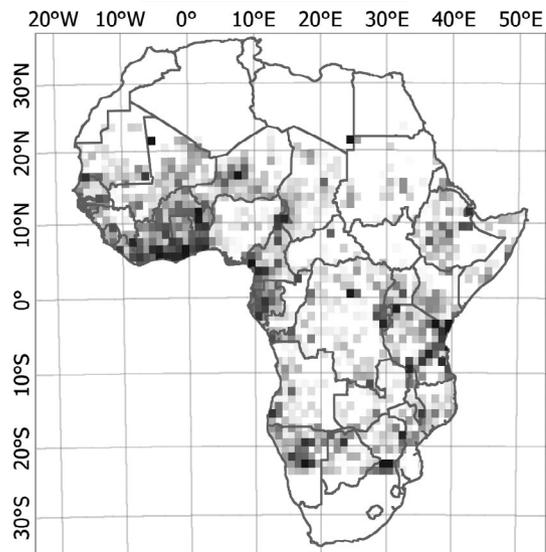


FIGURE 1 Geographical distribution of 533,383 records of 31,046 vascular plant taxa across 1,197 degree squares of tropical Africa included in the analysed dataset. Degree squares are coloured by sampling completeness from $\geq 6\%$ (lightest grey) to $\geq 100\%$ (black) in equal intervals of 1%. Sampling completeness is calculated by comparing the number of species recorded as present with richness estimates of Barthlott et al., 2005. Map projection Africa Albers Equal Area Conic

are conducted at the lowest named taxonomic rank, the most commonly used rank used in phytogeographical studies in tropical Africa (Denys, 1980; Fayolle et al., 2014, 2019; Linder et al., 2005, 2012). Only 4.5% of informative records are resolved to infraspecific rank. We excluded from analysis degree squares with < 15 taxon records, and $\leq 5\%$ of taxa sampled. These thresholds were set as low as possible to maximize floristic representation with respect to environmental gradients, without analysing noise, and are slightly more stringent than the 5 species/100 km² used by Kreft and Jetz (2010) and Linder et al. (2012). A description of data sources and cleaning is in Appendix S1. The geographical pattern of sampling is shown in Figure 1, and is included for each degree square in Appendix S2.

2.2 | Environmental data

Environmental data were summarized at one degree square and half degree square using QGIS 3.4.3. We summarized mean altitude from GMTED2010 at 30 arc second resolution (Danielson & Gesch, 2011). We derived topographic ruggedness from GMTED2010 using the GDAL Terrain Ruggedness Index tool via QGIS. Mean values for the climatic variables Bio1 to Bio35 at 30-min resolution for the years 1961–1990 were downloaded from the CliMond database (Kriticos et al., 2012). We used the surficial lithology classification of Sayre et al. (2013). We consulted the Harmonized World Soil Database, but could not use its classification as there were too many classes for the Random Forest algorithm to handle. We summarized the majority land cover class from GlobCover 2009 (Arino et al., 2012). We

estimated completeness of taxon sampling for each degree square by comparing the number of species recorded as present with richness estimates of Barthlott, Mutke, Rafiqpoor, Kier, & Kreft, 2005. Environmental data summarized at half degree square resolution can be found in Appendix S3.

2.3 | Gradients in species composition

Analysis was conducted in R 3.6.3 (R Core Team, 2020), scripts are found in Appendix S9. Dissimilarity in taxon composition between degree cells was measured using the Simpson index of beta diversity (β_{sim}) in the R package 'vegan' (Oksanen et al., 2019). β_{sim} down-weights dissimilarity between cells based on differences in species richness, making it appropriate for unevenly sampled datasets (Kreft & Jetz, 2010). The β_{sim} dissimilarity matrix was ordinated using locally constrained NMDS with seven axes specified (stress = 0.094), in *vegan*. Local NMDS is recommended as a robust technique for indirect gradient analysis (Minchin, 1987). The first three axis scores were visualized in RGB colour space using the R package 'plotrix' (Lemon, 2006). Environmental variables were fitted to the ordination space as fitted vectors using *vegan*'s *envfit* function.

2.4 | Deriving the biogeographical framework

We clustered the β_{sim} matrix using Ward's algorithm in the R package 'cluster' (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2019). Ward's algorithm gave the highest agglomerative coefficient with this dataset (0.95). We inspected 2–30 cluster solutions, every fifth cluster solution thereafter up to 95, and the 99th cluster solution. The highest node was always selected and no degree squares were reassigned. To help us choose the number of clusters to present, we calculated the decay in four different evaluation metrics for each of the clustering levels from 2 to 99, identifying the position of the knee using a geometric knee method: (a) dendrogram node height, where lower nodes are less informative than higher nodes; (b) ANOSIM R, a test statistic with the null that similarity between groups is greater than or equal to the similarity within the groups; (c) mean regional endemism, the average proportion of taxa restricted to each cluster and (d) total endemism, the percentage of all taxa that are restricted to any one cluster. The two endemism metrics were calculated from the tropical African dataset only, so regions particularly at the margins of the dataset will include some 'endemics' that are found elsewhere in the world.

2.5 | Modelling the framework

Random forest classification models were built using the environmental data as predictors, using the R package *randomForest* (Liaw & Wiener, 2002). We trained one model on the 19 regions to predict the regional framework. We subsequently trained 19 models to predict the distribution of the 99 districts within each of the 19 regions,

using the same selection of predictor variables. Continuous variables were rescaled between 0 and 1. Accuracy of the models was assessed using the out-of-bag error rate. Variable importance was assessed using the mean decrease in the Gini Index. Only half degree squares absent from the training dataset were classified using the model (half degree squares in the training dataset retained their classification). To reduce multicollinearity, overfitting and increase interpretability, indirect predictor variables were excluded and bioclim variables were restricted to the annual mean, minimum, and maximum values of temperature and precipitation, and annual mean value of the moisture index and radiation (Fig. S8.2 for correlations between variables). Including more bioclim variables decreased the OOB error rate by a negligible amount. Half degree squares were merged by region and district to produce two shapefiles (Appendix S4). The 19 regions were coloured and the 99 districts are cross-hatched using code written in Microsoft Visual Foxpro 9 for SVG output (Figure 5).

2.6 | Characterizing and comparing the framework

Characteristics of each region are summarized in Table 2 & Appendix S5, and for each district in Appendix S6. Districts were named after their largest population centre using Natural Earth's populated places layer. Floristic data were summarized by the number of taxa, number of endemic taxa, per cent endemism, per cent sampling completeness, number of indicator species and number of significant indicator species. Indicator species were identified in the R package 'indicspecies' (De Cáceres, Legendre, & Moretti, 2010). Species attributes, including occurrence by district and region, are in Appendix S7. Continuous environmental data used in the Random Forest model were summarized by their mean and standard deviation, minimum, median, maximum, interquartile range; lower and upper confidence intervals of the median are calculated using $\pm 1.58 \text{ IQR} / \sqrt{n}$. Categorical data were summarized by their majority class. We tested for significant differences in MAT and MAP among regions and districts with the multiple comparison Kruskal-Wallis test from package 'pgirmess' (Giraudoux, 2018). We used three of the evaluation metrics (mean endemism, total endemism, and ANOSIM R) to compare our framework floristically against the regional frameworks of White (1983) and Droissart et al. (2018). We compared the spatial congruence of our regions with the White and Droissart et al. regionalizations, and the spatial congruence of our districts with WWF's ecoregions, using the V-measure implemented by the R package 'sabre' (Nowosad & Stepinski, 2018).

3 | RESULTS

3.1 | Gradients in species composition

Floristic turnover at one degree square resolution is continuous, with no sharp disjunctions in ordination space (Figure 2). Floristic

turnover is well explained by both geographical and climatic variables while the estimated percentage of species sampled explains almost none of the one degree cells' position in ordination space ($r^2 = 0.027$) (Table 1). NMDS axis 1 represents the moisture gradient, and NMDS axis 2 the altitude/temperature gradient.

The soil moisture index (bio28) is the most strongly correlated variable with axis 1, and can explain a very large proportion of the one degree cells' positions in ordination space ($r^2 = 0.86$). The soil moisture index is a modelled environmental variable derived from precipitation, solar radiation (via pan evaporation) and water holding capacity of the soil; pan evaporation includes the effects of temperature, humidity, drought dispersion and wind (Hutchinson, Xu, Nix, & McMahon, 2009). Precipitation makes the most important contribution to the moisture index: the moisture index is very well correlated with mean annual precipitation (MAP) (Fig. S8.2). MAP only explains slightly less of the one degree cells' positions in ordination space than the soil moisture index ($r^2 = 0.83$), and is also very strongly parallel with axis 1.

Altitude is the most strongly parallel variable with axis 2, and explains a large amount of variation in ordination space ($r^2 = 0.69$). Altitude is well correlated with mean annual temperature (MAT) (Fig. S8.2), as is axis 2. Of the two variables, MAT explains more of the variation in ordination space than altitude ($r^2 = 0.76$). The zero point of axis 2 separates the continent rather precisely into the two parts which White (1983) recognized as 'high' and 'low' Africa: a separation of cooler, higher altitude southern and eastern Africa from the hotter, lower altitude northern, western and central regions; the 'line' is drawn from Angola in the southwest, around the Congo to western Ethiopia in the northeast (Figure 2b).

3.2 | Deriving the biogeographical framework

The continuous species turnover apparent in the ordination is also apparent in the cluster analysis. All four evaluation metrics have

TABLE 1 Squared correlation coefficient of the environmental vector fits for the 7 axis ordination of 31,046 vascular plant taxa across 1,197 degree squares of tropical Africa shown in Figure 2. All variables are significantly correlated with the ordination configuration ($p < .001$)

| Variable name | r^2 |
|-----------------------------------|-------|
| Latitude | 0.883 |
| Soil moisture index (Bio28) | 0.858 |
| Mean annual precipitation (Bio12) | 0.828 |
| Mean annual temperature (Bio01) | 0.763 |
| Longitude | 0.715 |
| Altitude | 0.691 |
| Solar radiation (Bio20) | 0.63 |
| Terrain Ruggedness Index | 0.442 |
| Lithology | 0.159 |
| % sampled | 0.027 |

smooth curves, and the knee detection tests suggest different optimum cluster numbers from each other: between 9 and 20 clusters (Fig. S7.1). The decision about how many clusters to recognize at the regional level is necessarily arbitrary within the range of break-points (Kreft & Jetz, 2010). No particular level is true or false, and so the results should be judged on the usefulness of the result (De Klerk et al., 2002). We chose to recognize the 19 cluster solution as regions. The 19th highest cluster is region 9 of Figure 3, a floristically distinct area which has previously been recognized at regional level (~ White's Karoo Namib). The 20th cluster is the division of the Sudanian region 2 into two latitudinal bands. The species distribution data through this area are relatively poor towards the east, weakening our confidence in the assignment of this cluster to regional level. We recognized the 99 cluster solution as districts: the choice of 99 is arbitrary, but is informed by the popularity of WWF's ecoregions (89 unique ecoregions in an equivalent area), and White's Vegetation Map's 100 cartographic units.

One degree squares in ordination space are coloured by the 19 clusters in Figure 3a. Hierarchical relationships between the 19 regions are visualized in Figure 3b and are mapped in Figure 3c. Three high level groups are apparent: (i) the rainforests of western and central Africa (1, 7, 11); (ii) low altitude northern tropical Africa (2–6) and (iii) high altitude southern and eastern Africa, (8–10, 12–19). High altitude southern and eastern Africa is the most floristically heterogeneous of the three primary groups, and can be divided as (iiia) Zambebian forests and woodlands (8, 10, 14, 13, 16, 17); (iiib) southern arid (9, 12) and (iiic) Horn of Africa (15, 18, 19).

3.3 | Modelling the framework

The regional Random Forest model has an out of bag error rate of 10.4%, that is, region could be accurately predicted for c. 90% of half degree square cells during cross validation. The error rates by region range from 2.94% (Region 12) to 25.4% (Region 3), and are higher for regions where sampling rates are lower (confusion matrix Table S8.2). Regional half degree square cell classifications are shown in Figure 4a. The relative importance of the model variables in the regional model is shown in Figure 4b. As in the ordination results, the most informative variables are those related to water availability (MAP, rad, soilMI and maxP), followed by temperature/altitude (maxT, minT, alt and MAT), with lower contributions from lithology, landcover and topographic ruggedness. The out of bag error rates for the district models range from 4.7% (Region 3) to 23.6% (Region 16). Floristic districts, derived from the 99-cluster solution, are mapped within the regional framework in Figure 5.

3.4 | Characterizing and comparing the framework

The regions are different from each other with respect to their total taxa (921 to 10,555), and mean endemism rate (1%–22%) (Table 2, Appendix S5). We have not adopted endemism thresholds to distinguish between

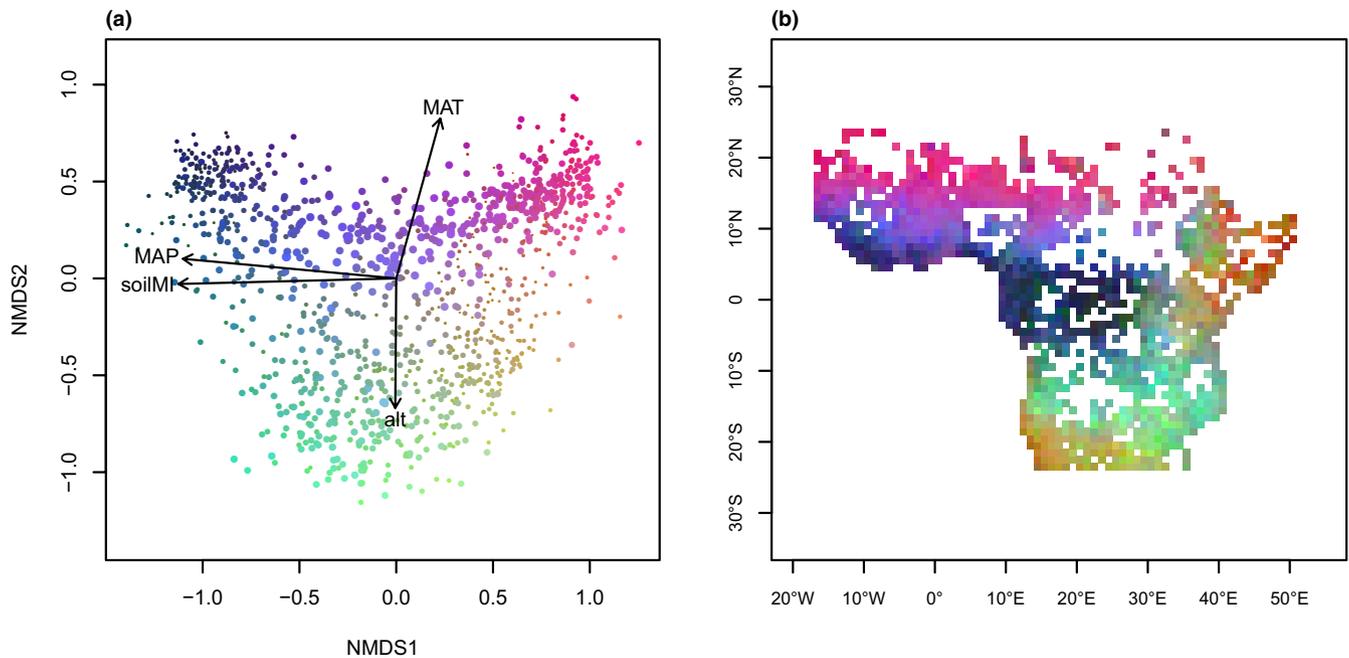


FIGURE 2 (a) NMDS ordination scores for 1,197 degree squares of tropical Africa. A betasim dissimilarity matrix constructed from 31,046 vascular plant records in each degree square was ordinated. Each degree square is represented in RGB, with red for axis 1, green for axis 2, and blue for axis 3 (third axis not shown). High axis 1 scores have maximum red values, low axis 2 scores have maximum green values and high axis 3 scores have maximum blue values. The third axis is also represented by size, with larger points having higher axis 3 scores. Environmental variables are fitted as vectors with length proportional to r^2 (Table 1). MAT = mean annual temperature (bio1), MAP = mean annual precipitation (bio12), soilMI = soil moisture index (bio28), alt = altitude. (b) Schematic representation of the location of cells, coloured by the same scheme

types of region. Such an approach is contingent upon a fair assessment of endemism rates across regions, and while the dataset is restricted to tropical Africa such assessments are imbalanced for internal versus external regions. The regions are distinct from each other climatically: a Kruskal–Wallis test of the difference in MAP across the 19 regions was significant overall (chi-square = 6,945, $df = 18$, $p < 0.005$). A multiple comparison Kruskal–Wallis test identified 13 statistically different groups of regions by median MAP (Figure S8.3). For example, MAP is not significantly different among the three rainforest regions. There are eight statistically different groups of regions by median MAT (Figure S8.4; chi-square = 5,179, $df = 18$, $p < .005$).

Our regions outperform White's regions by all evaluation metrics (mean endemism, total endemism, ANOSIM R; Table 3). Our regions outperform Droissart et al.'s bioregions by the ANOSIM R measure, that is, the ratio of similarity within groups to similarity between groups is higher in our regionalization. Mean regional and total endemism rates are higher in Droissart et al.'s bioregions because endemism thresholds were applied in that study, and because the number of degree squares we can use for the Droissart calculations is much less than the number of cells we can use to diagnose our own framework, as the Droissart framework covers a smaller area. Overall, our regions are slightly more congruent with Droissart et al.'s bioregions (V -measure = 0.67) than White's regions (V -measure = 0.63). We calculated taxon richness and endemism rates for White's tropical African Regions (Table S7.1). Our values are similar to previous reassessments (Droissart et al., 2018; Linder et al., 2005). White underestimated the number of 'chorological

transgressors'; his richness estimates for each chorion were too low, and his endemism rates too high.

The number of taxa per district varies between 100 and 8,579, and endemism rates between 0% and 25% (Appendix S6). The 99 districts are supported environmentally: there are 90 different median MAP values among the 99 districts (Figure S8.5, Kruskal–Wallis test chi-square = 7,401, $df = 98$, $p < 0.005$), and 74 different median MAT values among the 99 districts (Figure S8.6, Kruskal–Wallis test chi-square = 6,387, $df = 98$, $p < 0.005$). Floristically, they are at least moderately well supported: 32% of species are significant indicator species for one of the districts, compared with 56% of species which are significant indicators for one of the regions. Our districts are as homogeneous with respect to WWF's ecoregions as our regions are to White's regional framework (V -measure = 0.63, Figure 6c). At least in well-sampled areas, like our Guineo-Congolian (West) and Guineo-Congolian (West-Central), the districts would seem plausible based on field observation.

4 | DISCUSSION

4.1 | Can physical environmental variables be used to predict floristically defined vegetation units?

One of White's guiding principles in the assembly of his 1983 map was 'Vegetation, in the first instance, should be classified without

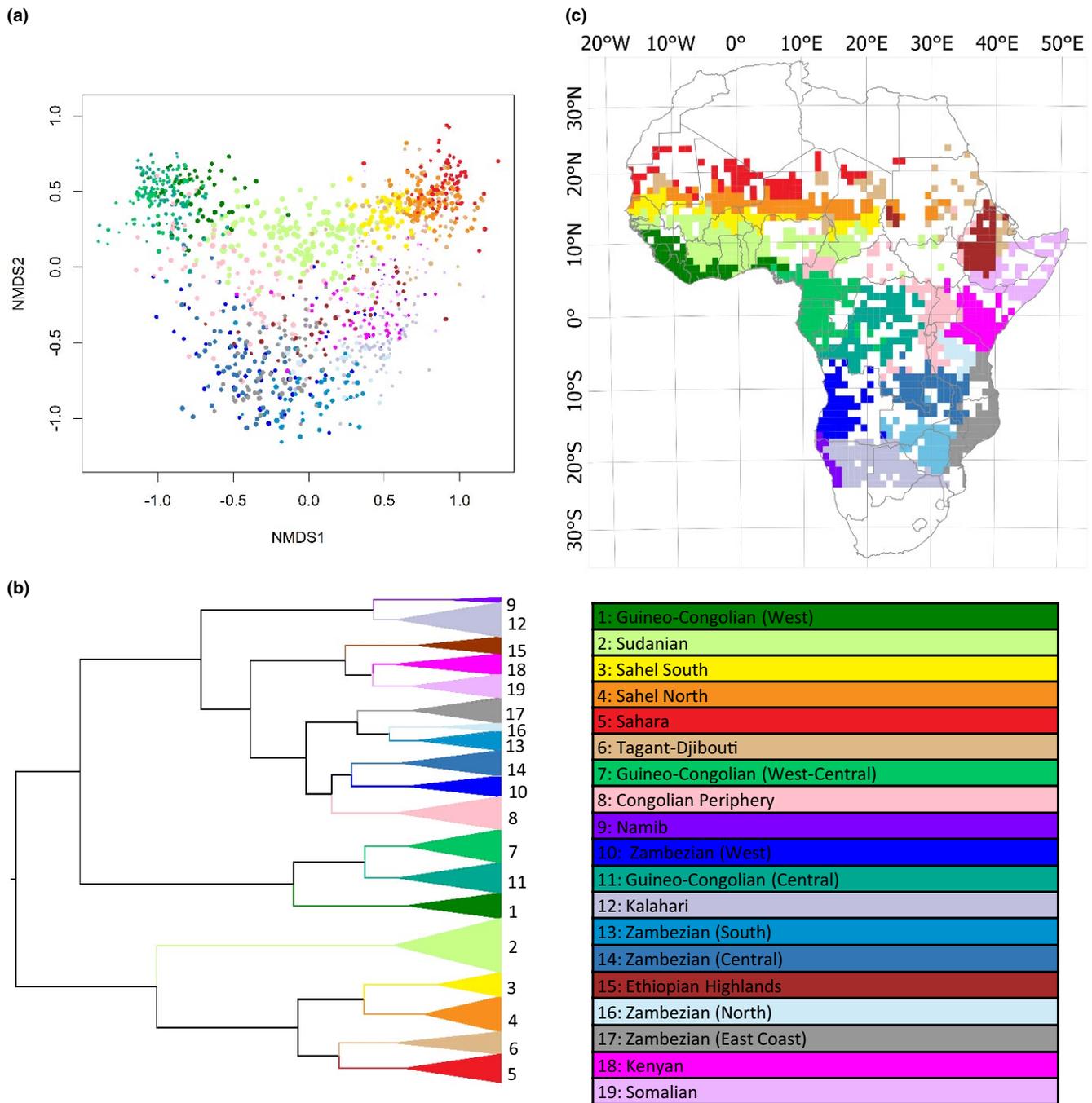


FIGURE 3 The 19 highest clusters given by Ward clustering of the betasim dissimilarity matrix of 31,046 tropical African vascular plant taxa. No cells have been reassigned and no clusters have been collapsed. In panel (a) the NMDS ordination as in Figure 2 is coloured by the clusters shown in panel b. (b) Hierarchical relationship between the 19 clusters (Table 2 for region attributes). (c) Geographical location of the 19 clusters. Grey polygon boundary lines delimit countries. Map projection Africa Albers Equal Area Conic

reference to the physical environment, including climate, or to animals. The extent to which environmental factors and the associated fauna can be used to diagnose vegetation types should be evaluated independently' (White, 1993). Our results suggest overwhelmingly that floristic turnover is well correlated with environmental variables at one degree square resolution across the range of tropical Africa's vegetation, and that environmental variables can be used to

predict floristically defined vegetation units empirically. At this scale and scope, we find that the principle variable to structure floristic turnover is water availability (explaining 86% of ordination position), followed by temperature, with smaller contributions from lithology, topographic ruggedness and land cover. Floristically defined regions can be predicted with 90% accuracy using a decision tree model and 12 readily available physical environmental predictors.

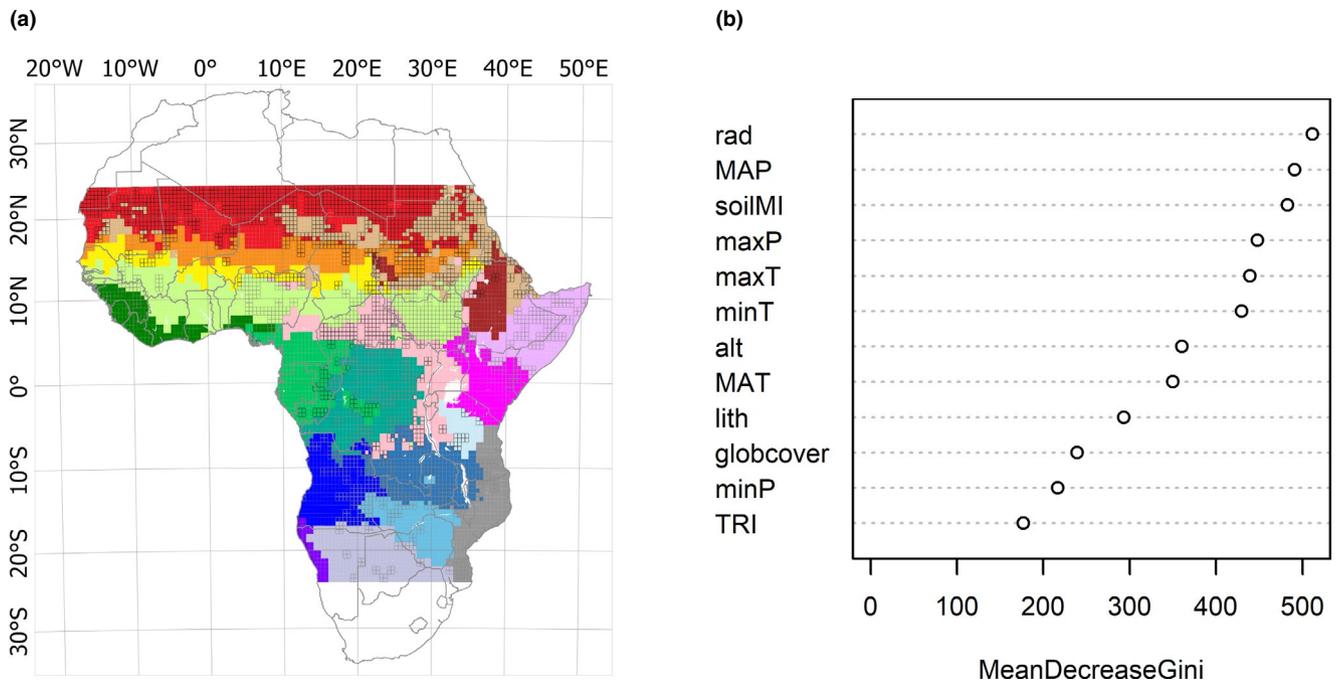


FIGURE 4 The 19 clusters derived from Ward clustering of the betasim dissimilarity matrix of 31,046 vascular plant taxa of tropical Africa (Figure 3) are modelled using Random Forest to yield the regional framework. The model has an out of bag error rate of 10.4%. (a) Cells with no border constitute the training data, predicted cell classifications with regional classification error rates of 10 > 25% are shown with a black border, predicted cell classifications with regional classification error rates of 3%–10% are shown with a grey border. Lakes are shown in white. (b) Variables included in the model are ranked by their importance. rad = radiation (bio20); MAP = mean annual precipitation (bio12); soilMI = annual mean moisture index (bio28); maxT = maximum temperature of the warmest month (bio5); maxP = precipitation of the wettest month (bio13); minT = minimum temperature of the coldest month (bio6); alt = mean altitude; MAT = mean annual temperature (bio1); lith = majority lithological category (Sayre et al., 2013); globcover = majority GlobCover category (Arino et al., 2012); minP = precipitation of the driest month (bio14); TRI = terrain ruggedness index. Map projection Africa Albers Equal Area Conic

Our model validation shows that cells whose region is known can be predicted with 90% accuracy. The ultimate test of the method can be conducted by collecting species occurrence data from currently unsampled parts of tropical Africa. As far as we are aware, the use of a decision tree (CART) method to interpolate a biotically derived biogeographical framework over a large spatial extent is novel, and the approach could be pursued for incompletely sampled taxa and areas outside of tropical Africa. The success of the CART approach is perhaps unsurprising, given the predictive relationship shown between climate and individual plant species distributions in tropical Africa (Blach-Overgaard, Balslev, Dransfield, Normand, & Svenning, 2015; Deblauwe et al., 2016; Maharjan et al., 2011; McClean et al., 2005), between climate and vegetation physiognomy (Arino et al., 2012), between climate and local-scale species assemblages across west African forest (Bongers et al., 2004), and between climate and local-scale species assemblages across savanna and forest biomes of tropical Africa (Fayolle et al., 2014, 2019). Our model is empirical (Guisan & Zimmermann, 2000), employed for the pragmatic purpose of rendering a biogeographical map spatially complete in a way that is plausible and useful. The model is not a formal test of association, and does not imply a causal or deterministic relationship between climate and floristic units, nor that the same relationships would be

found outside of tropical Africa, at different scale or scope in Africa, or under climate change scenarios.

By predicting floristically defined spatial units from environmental variables, our method offers a way to render, objectively and explicitly, the (imperfect) correspondence between chorology, vegetation physiognomy and physical environment. The modelled regions have been characterized by these physical variables. Ecosystems are defined as assemblages of biotic communities interacting with each other and their physical environment (Tansley, 1935). This definition draws together two principle approaches that have been used to divide the biosphere: the first using affinities and discontinuities in the distribution of taxa, the second using patterns of abiotic variation or direct measurement to classify the biosphere physiognomically (Mackey, Berry, & Brown, 2008). At finer spatial scales, it is rarely justified to separate the two approaches very distinctly, whereas in studies at larger spatial scales and scope the two disciplinary approaches have diverged. We have used a method popular in physiognomic spatial classification and applied it to a floristic biogeographical framework, producing spatial unit definitions that are closer to that of an ecosystem. Our framework has the advantage over physiognomic classifications (Arino et al., 2012; Sayre et al., 2013) that the units are diagnosed by, and fully characterized by, plant species distributions as currently represented by the plant biological record at

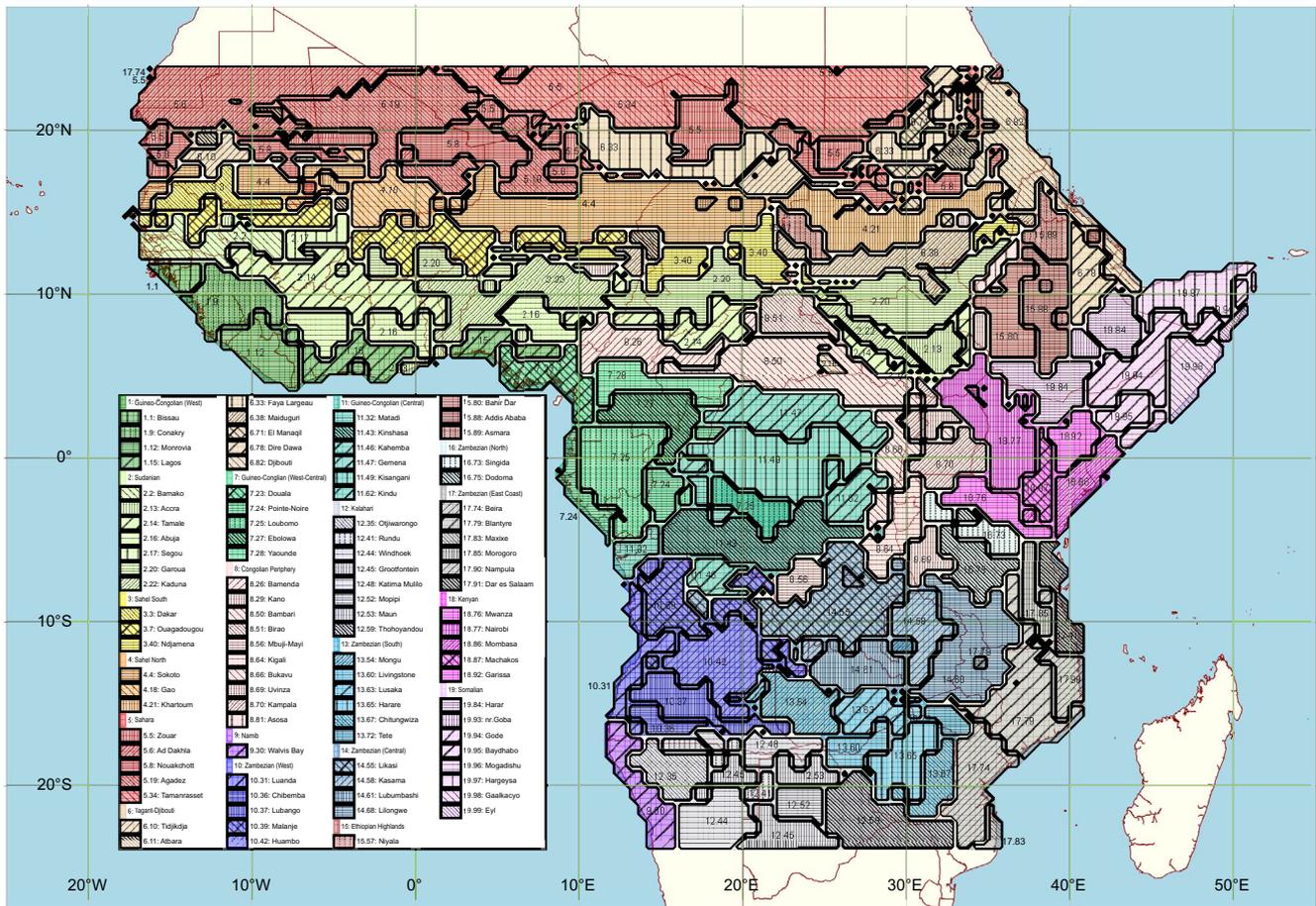


FIGURE 5 Final framework showing the 19 regions and 99 districts for tropical Africa, derived from cluster analysis of the betasim dissimilarity matrix calculated from 531,314 records of 31,046 vascular plant taxa across 1,197 degree squares and modelled with Random Forest. Regions are coloured and districts are crosshatched

one degree square resolution. It has the advantage over previous quantitative phytogeographical classifications (Droissart et al., 2018; Fayolle et al., 2014, 2019; Linder et al., 2005, 2012) that it is spatially complete across its scope. It has the advantage over qualitatively interpolated biogeographical frameworks (Olson et al., 2001; White, 1983) that our unit boundaries are drawn in a reproducible, objective and quantitative manner. Of course, the degree of advantage presented by our framework is dependent on the intended use.

4.2 | How do our phytogeographical regions compare with previously defined regions?

Each region is interpreted with respect to previous region definitions in Appendix S8. Overall, our regions are very geographically coherent before modelling (Figure 2, Figure 3): not a single cell was manually reassigned, and there are almost no cells which have a dubious cluster assignment. This is in contrast to previous quantitative gridded chorological studies, which showed more geographically fractured regions (Droissart et al., 2018; Linder et al., 2005; plants in Linder et al., 2012). More comprehensive species distribution data have helped to resolve the phytogeographical affinities across

the tropical continent. The bipartite method (Edler et al., 2017) employed in the Droissart et al. (2018) analysis produced rather small bioregions in some areas, which we suggest are not easily reconciled to the regional scale. The bipartite method recognized the montane flora of southern Africa distinctly from surrounding vegetation, and the approach might be well suited to the definition or validation of the district level presented here. Our gridded, whole-flora regionalisation for the forest biome is congruent with the tree-only plot data results of Fayolle et al. (2014), and divergent from White's (1983) division of Guineo-Congolia into Upper Guinea, Lower Guinea and Congo at the Dahomey Gap and Sangha River. This finding contradicts the conclusion of Droissart et al. (2018) that tree species cannot be used as a proxy of the whole flora for delimitating phytogeographical boundaries in tropical Africa, at least for the forests of west and central Africa. We recover no High/Low African disjunction in ordination space as in Fayolle et al. (2019) and our NMDS axis 1 represents the moisture gradient, but our NMDS ordination axis 2 very precisely divides the continent into High and Low Africa. Grid squares create assemblages of species which are found in distinct vegetation at finer scale (Kreft & Jetz, 2010), and the failure to resolve archipelago and mosaic regions with gridded data is a perennial issue (Linder et al., 2005). Even White (1993), with the option

| Name of region | No. taxa | ER (%) | MAP (mm/yr) | MAT (°C) | alt (m) |
|------------------------------------|----------|--------|---------------|--------------|------------------|
| 1: Guineo-Congolian (West) | 6,027 | 8 | 1776 ± 71.89 | 26.05 ± 0.12 | 177.49 ± 29.99 |
| 2: Sudanian | 5,893 | 4 | 1,005 ± 17.78 | 27.3 ± 0.06 | 377.16 ± 9.01 |
| 3: Sahel South | 2,107 | 1 | 532 ± 18.19 | 28.44 ± 0.1 | 297.3 ± 13.65 |
| 4: Sahel North | 1,475 | 1 | 218 ± 10.31 | 28.46 ± 0.11 | 344.61 ± 12.39 |
| 5: Sahara | 921 | 11 | 27 ± 2.46 | 26.28 ± 0.16 | 445.14 ± 13.72 |
| 6: Tagant-Djibouti | 2,930 | 4 | 92 ± 15.64 | 27.25 ± 0.12 | 430.65 ± 13.61 |
| 7: Guineo-Congolian (West-Central) | 9,194 | 22 | 1712 ± 22.6 | 24.72 ± 0.11 | 447.12 ± 21.64 |
| 8: Congolian Periphery | 9,873 | 6 | 1,366 ± 21.62 | 23.82 ± 0.15 | 815.3 ± 31.58 |
| 9: Namib | 1,495 | 5 | 102 ± 26.91 | 18.3 ± 0.58 | 570.57 ± 138.59 |
| 10: Zambezian (West) | 4,678 | 15 | 1,093 ± 42.49 | 21.24 ± 0.16 | 1,157.36 ± 32.92 |
| 11: Guineo-Congolian (Central) | 6,199 | 7 | 1661 ± 13.98 | 24.82 ± 0.07 | 484.44 ± 14.84 |
| 12: Kalahari | 5,111 | 17 | 416 ± 11.17 | 21.56 ± 0.13 | 1,095.73 ± 18.36 |
| 13: Zambezian (South) | 4,978 | 5 | 755.5 ± 17.38 | 21.84 ± 0.15 | 1,068.03 ± 30.1 |
| 14: Zambezian (Central) | 9,541 | 14 | 1,171 ± 18.3 | 21.62 ± 0.18 | 1,114.59 ± 28.62 |
| 15: Ethiopian Highlands | 4,472 | 9 | 1,149 ± 82.96 | 19.39 ± 0.5 | 1737.69 ± 83.78 |
| 16: Zambezian (North) | 4,267 | 2 | 787 ± 39.37 | 22 ± 0.24 | 1,206.42 ± 24.37 |
| 17: Zambezian (East Coast) | 8,587 | 9 | 1,112 ± 32.2 | 24.2 ± 0.14 | 285.45 ± 37.27 |
| 18: Kenyan | 10,555 | 12 | 716 ± 45.24 | 25.34 ± 0.51 | 669.6 ± 93.19 |
| 19: Somalian | 4,135 | 21 | 363 ± 24.62 | 26.37 ± 0.26 | 535.15 ± 52.57 |

Abbreviations: alt, altitude; ER, endemism rate; MAP, mean annual precipitation; MAT, mean annual temperature.

TABLE 3 Quantitative comparison between the tropical African regionalization derived from vascular plant taxa proposed here, White, 1983, and Droissart et al., 2018 frameworks. Rates have been calculated at the intraspecific level using the TRAFRICA dataset; rates published by the authors are given in brackets

| Framework | No. regions | Mean endemism | Total endemism rate | Total endemic taxa | Anosim R |
|------------------------|-----------------|-------------------|---------------------|-----------------------|----------|
| This publication | 19 | 9.10% | 34.3% | 10,635 | 0.79 |
| Droissart et al., 2018 | 15 ^a | 10.1% (13.5%) | 37.5% (32.1%) | 10,844 (8,144) | 0.69 |
| White, 1983 | 13 | 8.31% (c. 23%) | 33.0% | 10,238 (c. 15,808) | 0.60 |

^aCalculation from our dataset excludes São Tomé & Príncipe and transition regions

of drawing regions freehand, bemoaned the 'special problems' associated with mapping archipelago regions like the Afromontane representatively. Superimposing finer scale floristic units derived via quantitative analysis at variably finer resolutions where the data allow (Droissart et al., 2018; Fayolle et al., 2014, 2019) onto a

regional map as White did (1983), but allowing the regional identity of the smaller cells to diverge from that of their enclosing larger cell, would be a solution.

Of note is our reinterpretation of the arid flora affinities in tropical Africa: our results divide White's Somali-Masai into two

TABLE 2 Characterization of the 19 tropical African regions derived from vascular plant taxa. Continuous variables are summarized as medians with confidence intervals calculated by ± 1.58 IQR/sqrt(n)

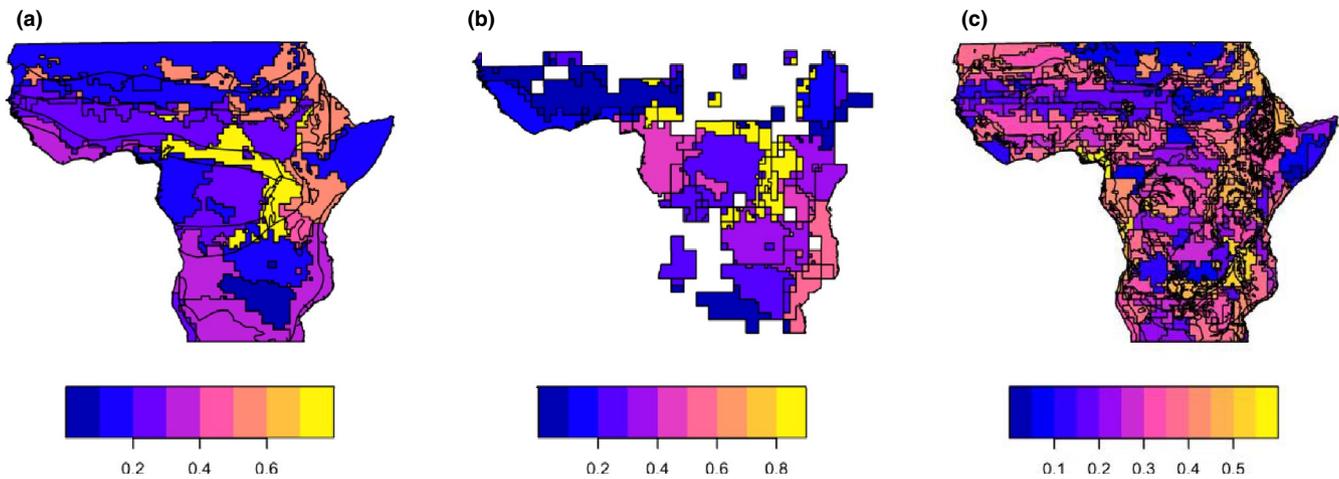


FIGURE 6 Spatial congruence of the tropical African vascular plant biogeographical framework derived here, with previous biogeographical frameworks for tropical Africa. Areas identified in the present analysis are coloured in proportion to the number of previously defined areas which they intersect or include. (a) Extent of inhomogeneity of the TRAFRICA 19 regions with respect to White's, 1983 regional classification, v -measure = 0.67; (b) Extent of inhomogeneity of TRAFRICA 19 regions with respect to Droissart et al.'s bioregions, v -measure = 0.63; (c) Extent of inhomogeneity of TRAFRICA 99 districts with respect to WWF Ecoregions, v -measure 0.63

regions for the first time (18: Kenyan, 19: Somalian), and separate the arid flora of Djibouti, north-eastern Ethiopia and Eritrea away from this region. Instead, the Djibouti flora clusters with cells under the same aridity regime right across to Mauritania (Tagant plateau) in the west of the continent. The new region is named 6: Tagant-Djibouti. The arid flora of eastern Namibia (9: Namib) is supported at regional level for the first time in a floristic quantitative regionalization of tropical African plants. The 'arid track' – shared species of plants (de Winter, 1971) and animals (Balinsky, 1962) across the geographically disjunct arid floras of the south west (Namibia) and north east (Djibouti and environs) – can be seen in ordination space (Figure 2), though overall their strongest affinity is to their geographically adjacent regions (Figure 3c). In the forest biome, we define three regions. Our division of White's Guineo-Congolian region is most congruent with Fayolle et al.'s subdivision of White's Guineo-Congolia (2014), with neither the Dahomey Gap nor the Sangha River interval recovered; floristic affinities instead follow climatic (physiognomic) patterns (before and after modelling). In the savanna biome north of the forest, the resolution of continuous floristic turnover along climatic gradients into useful regions is achieved (region 2: Sudanian, 3: Sahel South, 4: Sahel North). This is likely due to the inclusion of a large floristic dataset from northern tropical Africa (Taugourdeau et al., 2019). Our very literal definition of tropical Africa as the area between the tropics of Cancer and Capricorn has probably helped by defining both ends of the moisture gradient. We recover four main regions within White's Zambezi region. Droissart et al. (2018) found that the Zambezi region of White can be divided, something White also suggested. A fifth region mixes Zambezi vegetation with coastal forests and does not include the Kenyan coast (17: Zambezi East Coast). The signal from small-scale enclaves of distinctive vegetation has been lost in the matrix of the broader

grid cell (Hawthorne & Marshall, 2019). Our 'Afromontane' region (15: Ethiopian Highlands) is restricted to Ethiopia and the mountains of western Sudan. Smaller sized Afromontane sites, such as Erkowit in northeastern Sudan (Vetaas, 1993; Vetaas, Salih, & Jurasinski, 2012), and other outposts further south and north are not recovered within the region.

4.3 | Future directions

We limited the scope of our analysis to continental tropical Africa. The floras of northern Africa, tropical Africa, southern extra-tropical Africa and Madagascar are large and distinct from each other floristically and environmentally. The analysis of herbarium datasets south and north of 24° alongside tropical African datasets would be an important step. Our regions and districts are defined by gross floristic patterns. Considering how emphasis may be placed on patterns of endemism rather than gross floristic content, acknowledging that those patterns differ (De Klerk et al., 2002), would be an important development for biogeographically informed conservation, probably requiring separate map layers. Incorporating the evolutionary history of species into tropical African biogeographical frameworks is an exciting prospect (Daru et al., 2016; Slik et al., 2018), especially as vascular plant phylogenies improve (Janssens et al., 2020). Some areas of tropical Africa remain poorly represented by our dataset, for example, Sudan and South Sudan, Somalia and Nigeria, and our regionalization is inevitably more speculative where training data are sparser. Notwithstanding, the biogeographical framework presented here should help with the interpretation of tropical Africa's floristic patterns from a historical and evolutionary perspective in future studies, a current research interest of

ours (Dagallier et al., 2020). For almost all vascular plant taxa of tropical Africa, species occurrences and predicted occurrences within regions and districts are made available, and should be useful for monitoring trends in habitat condition, providing shortlists of native species for restoration, or filtering possible species during plant identification.

ACKNOWLEDGEMENTS

C.M. acknowledges support from King's College Cambridge, University of Cambridge; the Clarendon Fund, University of Oxford, and Merton College, University of Oxford; W.H. acknowledges support from the James Martin 21st Century School, University of Oxford. J.W. acknowledges support from the Dutch Economic Structure Enhancement Fund (FES) facilitating data capture of the Dutch herbaria. Cyrille Chatelain provided the African Plants Database Tropical Africa species list. Data capture, editing and broadcast of the African Plants Database is the product of a collaboration between the South African National Biodiversity Institute, the Conservatoire et Jardin botaniques de la Ville de Genève, Tela Botanica and the Missouri Botanical Garden. This paper has benefited from reviews by Thomas Couvreur, Mike Swaine, and others anonymously.

DATA AVAILABILITY STATEMENT

Dryad data deposit DOI: <https://doi.org/10.5061/dryad.rfj6q5786>. Sources of distribution data are summarized in Appendix S1. All results for each one degree square, including sampling levels, are made available in Appendix S2. All results for each half degree square, including sampling levels, are made available in Appendix S3. Shape files of the Regions and Districts are made available in Appendix S4. Summary data for regions, districts and species are made available in Appendices S5, S6 and S7. Code is available in S9.

ORCID

Cicely A.M. Marshall  <https://orcid.org/0000-0002-7397-6472>

Jan J. Wieringa  <https://orcid.org/0000-0003-0566-372X>

William D. Hawthorne  <https://orcid.org/0000-0002-6926-1722>

REFERENCES

- Arino, O., Ramos Perez, J. J., Kalogirou, V., Bontemps, S., Defourny, P., & Van Bogaert, E. (2012). Global Land Cover Map for 2009 (GlobCover 2009) © European Space Agency (ESA) & Université catholique de Louvain (UCL). PANGAEA. <https://doi.org/10.1594/PANGAEA.787668>
- Aubreville, A. (1962). Position chorologique du Gabon. *Flore du Gabon* 3 (pp. 3–11). ParisMuséum Hist. Nat.
- Balinsky, B. I. (1962). Patterns of animal distribution of the African continent. *Annals of the Cape Province Museum*, 2, 299–310.
- Barthlott, W., Mutke, J., Rafiqpoor, D. M., Kier, G., & Kreft, H. (2005). Global centers of vascular plant diversity. *Nova Acta Leopoldina NF*, 92(342), 61–83.
- Blach-Overgaard, A., Balslev, H., Dransfield, J., Normand, S., & Svenning, J. C. (2015). Global-change vulnerability of a key plant resource, the African palms. *Scientific Reports*, 5, 1–10. <https://doi.org/10.1038/srep12611>
- Bongers, F., Poorter, L., & Hawthorne, W. D. (2004). *The forests of Upper Guinea: Gradients in large species composition*. In Biodiversity of West African forests CABI Publishing, UK.
- Brunken, U., Schmidt, M., Dressler, S., Janssen, T., Thiombiano, A., & Zizka, G. (2008). West African plants - A Photo Guide. Forschungsinstitut Senckenberg, Frankfurt/Main, Germany. www.westafricanplants.senckenberg.de.
- Conservatoire et Jardin botaniques de la Ville de Genève and South African National Biodiversity Institute, Pretoria (2016). African Plants Database (version 3.4.0). URL: <http://africanplantdatabase.ch>
- Dagallier, L. P. M. J., Janssens, S. B., Dauby, G., Blach-Overgaard, A., Mackinder, B. A., Droissart, V., ... Couvreur, T. L. P. (2020). Cradles and museums of generic plant diversity across tropical Africa. *New Phytologist*, 225(5), 2196–2213. <https://doi.org/10.1111/nph.16293>
- Danielson, J. J., & Gesch, D. B. (2011). Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073. 2010, iv, 26.
- Daru, B. H., van der Bank, M., Maurin, O., Yessoufou, K., Schaefer, H., Slingsby, J. A., & Davies, T. J. (2016). A novel phylogenetic regionalization of phytogeographical zones of southern Africa reveals their hidden evolutionary affinities. *Journal of Biogeography*, 43(1), 155–166. <https://doi.org/10.1111/jbi.12619>
- Dauby, G., Stévant, T., Droissart, V., Cosiaux, A., Deblauwe, V., Simo-Droissart, M., ... Couvreur, T. L. P. (2017). ConR: An R package to assist large-scale multispecies preliminary conservation assessments using distribution data. *Ecology and Evolution*, 7(24), 11292–11303. <https://doi.org/10.1002/ece3.3704>
- De Cáceres, M., Legendre, P., & Moretti, M. (2010). Improving indicator species analysis by combining groups of sites. *Oikos*, 119(10), 1674–1684. <https://doi.org/10.1111/j.1600-0706.2010.18334.x>
- De Klerk, H. M., Crowe, T. M., Fjeldså, J., & Burgess, N. D. (2002). Biogeographical patterns of endemic terrestrial Afrotropical birds. *Diversity and Distributions*, 8(3), 147–162. <https://doi.org/10.1046/j.1472-4642.2002.00142.x>
- de Winter, B. (1971). Floristic relationships between the northern and southern arid areas in Africa. *Mitteilungen Botanische Staatssammler Mit Munchen*, 10, 424–437.
- Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J. C., ... Couvreur, T. L. P. (2016). Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography*, 25(4), 443–454. <https://doi.org/10.1111/geb.12426>
- Denys, E. (1980). A tentative phytogeographical division of tropical Africa based on a mathematical analysis of distribution maps. *Bulletin Du Jardin Botanique National De Belgique*, 50(3/4), 465. <https://doi.org/10.2307/3667842>
- Droissart, V., Dauby, G., Hardy, O. J., Deblauwe, V., Harris, D. J., Janssens, S., ... Couvreur, T. L. P. (2018). Beyond trees: Biogeographical regionalization of tropical Africa. *Journal of Biogeography*, 45(5), 1153–1167. <https://doi.org/10.1111/jbi.13190>
- Edler, D., Guedes, T., Zizka, A., Rosvall, M., & Antonelli, A. (2017). Infomap bioregions: Interactive mapping of biogeographical regions from species distributions. *Systematic Biology*, 66(2), 197–204. <https://doi.org/10.1093/sysbio/syw087>
- Engler, A. (1879). *Versuch einer Entwicklungsgeschichte der Pflanzenwelt*. LeipzigEngelman.
- Favier, C., Aleman, J., Bremond, L., Dubois, M. A., Freycon, V., & Yangakola, J. M. (2012). Abrupt shifts in African savanna tree cover along a climatic gradient. *Global Ecology and Biogeography*, 21(8), 787–797. <https://doi.org/10.1111/j.1466-8238.2011.00725.x>
- Fayolle, A., Engelbrecht, B., Freycon, V., Mortier, F., Swaine, M., Réjou-Méchain, M., ... Gourlet-Fleury, S. (2012). Geological substrates shape tree species and trait distributions in African moist



- forests. *PLoS One*, 7(8), 12–14. <https://doi.org/10.1371/journal.pone.0042381>
- Fayolle, A., Swaine, M. D., Aleman, J., Azihou, A. F., Bauman, D., te Beest, M., ... Woollen, E. (2019). A sharp floristic discontinuity revealed by the biogeographic regionalization of African savannas. *Journal of Biogeography*, 46(2), 454–465. <https://doi.org/10.1111/jbi.13475>
- Fayolle, A., Swaine, M. D., Bastin, J. F., Bourland, N., Comiskey, J. A., Dauby, G., ... Plumptre, A. J. (2014). Patterns of tree species composition across tropical African forests. *Journal of Biogeography*, 41(12), 2320–2331. <https://doi.org/10.1111/jbi.12382>
- Gilles, D., Zaiss, R., Blach-Overgaard, A., Catarino, L., Damen, T., Deblauwe, V., ... Couvreur, T. L. P. (2016). RAINBIO: A mega-database of tropical African vascular plants distributions. *PhytoKeys*, 74(1), 1–18. <https://doi.org/10.3897/phytokeys.74.9723>
- Giraudoux, P. (2018). *pgirmess: Spatial Analysis and Data Mining for Field Ecologists. R Package Version*, 1(6), 9. <https://cran.r-project.org/package=pgirmess>
- Good, R. (1947). *The Geography of the Flowering Plants*. Longmans. Green and Co., Ltd.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Hall, J. B., & Swaine, M. D. (1976). Classification and ecology of closed-canopy forest in Ghana. *The Journal of Ecology*, 64(3), 913. <https://doi.org/10.2307/2258816>
- Hawthorne, W. D., & Marshall, C. A. M. (2019). Rapid Botanic Survey, Bioquality and improving botanical inventory in the tropics by integrating across spatial scales. *Gardens' Bulletin Singapore*, 71(suppl.2), 315–333. [https://doi.org/10.26492/gbs71\(suppl.2\).2019-21](https://doi.org/10.26492/gbs71(suppl.2).2019-21)
- Holdridge, L. R. (1947). Determination of World Plant Formations from Simple Climatic Data Author (s): L. R. Holdridge. *Science*, 105(2727), 367–368.
- Holt, B. G., Lessard, J. P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., ... Rahbek, C. (2013). An update of Wallace's zoogeographic regions of the world. *Science*, 339(6115), 74–78. <https://doi.org/10.1126/science.1228282>
- Hutchinson, M., Xu, T., Nix, H., & McMahon, J. (2009). ANUCLIM 6.0 User's Guide.
- Janssens, S., Couvreur, T. L. P., Mertens, A., Dauby, G., Dagallier, L.-P., Abeele, S. V., Vandeloek, F., Mascarello, M., Beeckman, H., Sosef, M., Droissart, V., van der Bank, M., Maurin, O., Hawthorne, W., Marshall, C., Rejou-Mechain, M., Beina, D., Baya, F., Merckx, V., ... Hardy, O. (2020). A large-scale species level dated angiosperm phylogeny for evolutionary and ecological analyses. *Biodiversity Data Journal*, 8, e39677. <https://doi.org/10.1101/809921>
- Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, 37(11), 2029–2053. <https://doi.org/10.1111/j.1365-2699.2010.02375.x>
- Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J., & Scott, J. K. (2012). CliMond: Global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, 3(1), 53–64. <https://doi.org/10.1111/j.2041-210X.2011.00134.x>
- Lebrun, J. (1960). *Sur la richesse de la flore de divers territoires africains*. BruxellesAcademie Royale de Sciences d'Outre Mer.
- Lebrun, J. (1961). Les deux Flores d'Afrique tropicale. In Acad. R. de Belgique, Cl. des sciences Mémoires : Coll. in -8-o : sér. 2; 32,6. [Bruxelles] [Acad.].
- Lemon, J. (2006). Plotrix: A package in the red light district of R. *R-News*, 6(4), 8–12.
- Liaw, A., & Wiener, M. (2002). Classification and Regression by random-Forest. *R News*, 2(3), 18–22.
- Lillesø, J.-P.-B., van Breugel, P., Kindt, R., Bingham, M., Demissew, S., Dudley, C., ... Gaudal, L. (2011). Potential Natural Vegetation of Eastern Africa. Volume 1: The Atlas: Vol. Forest & L. www.sl.life.ku.dk.
- Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldså, J., & Rahbek, C. (2012). The partitioning of Africa: Statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography*, 39(7), 1189–1205. <https://doi.org/10.1111/j.1365-2699.2012.02728.x>
- Linder, H. P., Lovett, J., Mutke, J. M., Barthlott, W., Jürgens, N., Rebelo, T., ... Küper, W. (2005). A numerical re-evaluation of the sub-Saharan phytochoria of mainland Africa. *Biol. Skr*, 55, 229–252.
- Mackey, B. G., Berry, S. L., & Brown, T. (2008). Reconciling approaches to biogeographical regionalization: A systematic and generic framework examined with a case study of the Australian continent. *Journal of Biogeography*, 35(2), 213–229. <https://doi.org/10.1111/j.1365-2699.2007.01822.x>
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2019). *cluster: Cluster Analysis Basics and Extensions (R package version 2.0.8)*.
- Maharjan, S. K., Poorter, L., Holmgren, M., Bongers, F., Wieringa, J. J., & Hawthorne, W. D. (2011). Plant functional traits and the distribution of West African rain forest trees along the rainfall gradient. *Biotropica*, 43(5), 552–561. <https://doi.org/10.1111/j.1744-7429.2010.00747.x>
- Marshall, C. A. M., Wieringa, J. J. J., & Hawthorne, W. D. W. D. (2016). Bioquality hotspots in the tropical african flora. *Current Biology*, 26(23), 3214–3219. <https://doi.org/10.1016/j.cub.2016.09.045>
- McClean, C. J., Lovett, J. C., Termansen, M., Tokumine, S., Taplin, J. R. D., Küper, W., ... Smith, G. F. (2005). African plant diversity and climate change. *Annals of the Missouri Botanical Garden*, 92(2), 139–152. <https://doi.org/10.2307/3298511>
- McLaughlin, S. P. (1992). Are floristic areas hierarchically arranged? *Journal of Biogeography*, 19(1), 21. <https://doi.org/10.2307/2845617>
- Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Theory and models in vegetation science*, Vol. 69 (pp. 89–107). NetherlandsSpringer. https://doi.org/10.1007/978-94-009-4061-1_9
- Monod, T. (1957). Les grandes divisions chorologiques de l'Afrique : rapport prés. à la réunion de spécialistes sur la phytogéographie; <Yangambi, 29 juil.- 8 août 1956>. Londres.
- Morrone, J. J. (2018). The spectre of biogeographical regionalization. *Journal of Biogeography*, 45(2), 282–288. <https://doi.org/10.1111/jbi.13135>
- Nowosad, J., & Stepinski, T. F. (2018). Spatial association between regionalizations using the information-theoretical V-measure. *International Journal of Geographical Information Science*, 32(12), 2386–2401. <https://doi.org/10.1080/13658816.2018.1511794>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., ... Wagner, H. (2019). *vegan: Community Ecology Package (R package version 2.5-5)*. <https://cran.r-project.org/package=vegan>.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., ... Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, 51(11), 933. [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:teotwa\]2.0.co;2](https://doi.org/10.1641/0006-3568(2001)051[0933:teotwa]2.0.co;2)
- R Core Team, (2020). *R: A language and environment for statistical computing*. Vienna, AustriaR Foundation for Statistical Computing. <https://www.r-project.org/>.
- Réjou-Méchain, M., Pélissier, R., Gourlet-Fleury, S., Couteron, P., Nasi, R., & Thompson, J. D. (2008). Regional variation in tropical forest tree species composition in the Central African Republic: An assessment based on inventories by forest companies. *Journal of Tropical Ecology*, 24(6), 663–674. <https://doi.org/10.1017/S0266467408005506>
- Sayre, R., Comer, P., Hak, J., Josse, C., Bow, J., Warner, H., ... Waruingi, L. (2013). A new map of standardized terrestrial ecosystems of Africa. http://www.aag.org/cs/publications/special/map_african_ecosystems.



- Slik, J. W. F., Franklin, J., Arroyo-Rodríguez, V., Field, R., Aguilar, S., Aguirre, N., ... Zang, R. (2018). Phylogenetic classification of the world's tropical forests. *Proceedings of the National Academy of Sciences of the United States of America*, 115(8), 1837–1842. <https://doi.org/10.1073/pnas.1714977115>
- Swaine, M. D. (1996). Rainfall and soil fertility as factors limiting forest species distributions in Ghana. *The Journal of Ecology*, 84(3), 419. <https://doi.org/10.2307/2261203>
- Takhtajan, A. L., Crovello, T. J., & Cronquist, A. (1986). *Floristic regions of the world*. University of California Press.
- Tansley, A. (1935). The use and abuse of vegetational concepts and terms. *Ecology*, 16(3), 284–307. <https://doi.org/10.2307/1930070>
- Taugourdeau, S., Daget, P., Chatelain, C., Mathieu, D., Juanes, X., Huguenin, J., & Ickowicz, A. (2019). FLOTROP, a massive contribution to plant diversity data for open ecosystems in northern tropical Africa. *Scientific Data*, 6(1), 4–11. <https://doi.org/10.1038/s41597-019-0120-8>
- Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003). Dispersal, Environment, and Floristic Variation of Western Amazonian Forests. Published by : American Association for the Advancement of Science Stable, 299(5604), 241–244. URL: <http://www.jstor.org/stable/2995604>
- Vetaas, O. R. (1993). Spatial and Temporal Vegetation Changes Along a Moisture Gradient in Northeastern Sudan. Published by : The Association for Tropical Biology and Conservation Stable URL. <http://www.jstor.org/stable/2389180> Spatial and Temporal Vegetation Changes Along a M. *Biotropica*, 25(2), 164–175.
- Vetaas, O. R., Salih, E. A., & Jurasinski, G. (2012). Vegetation changes in the Red Sea Hills: From mist oasis to arid shrub. *Plant Ecology and Diversity*, 5(4), 527–539. <https://doi.org/10.1080/17550874.2012.749954>
- White, F. (1979). The Guineo-Congolian Region and Its Relationships to Other Phytochoria. *Bulletin Du Jardin Botanique National De Belgique*, 49(1/2), 11. <https://doi.org/10.2307/3667815>
- White, F. (1983). *The Vegetation of Africa, a descriptive memoir to accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa*. Scientific and Cultural Organization United Nations Educational.
- White, F. (1993). The AETFAT chorological classification of Africa: history, methods and applications. *Bulletin Du Jardin Botanique National De Belgique*, 62(1/4), 225. <https://doi.org/10.2307/3668279>

BIOSKETCH

Cicely Marshall is a botanist and ecologist interested in mapping and modelling floristic patterns in tropical Africa, particularly with a view to improving conservation outcomes; William Hawthorne is a tropical botanist and plant ecologist, concentrating mostly on the interface between the academic and the practical management of tropical biodiversity, particularly in forest and mostly in Africa and the Caribbean. Jan Wieringa is a systematist and curator of both plant and animal records, with a geographical focus on Europe and Africa. CM and WH conceived the ideas; WH, CM and JW collected and compiled the data; CM analysed the data; CM led the writing with assistance from WH and JW.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Marshall CAM, Wieringa JJ, Hawthorne WD. An interpolated biogeographical framework for tropical Africa using plant species distributions and the physical environment. *J Biogeogr.* 2020;00:1–14. <https://doi.org/10.1111/jbi.13976>