FRONTIERS OF BIOGEOGRAPHY
The scientific journal of
The International Biogeography Society

Frontiers of Biogeography 18, 2025,
e145498 | DOI 10.21425/fob.18.145498

RESEARCH ARTICLE

Satellite remote sensing can operationalise the IUCN Global Ecosystem Typology in the biome-diverse north-east of Brazil

Lucy H. Wells[1], Kyle G. Dexter[1,2,3], R. Toby Pennington[2,4], Ítalo Antônio Cotta Coutinho[5], Desiree Ramos[1,6], Oliver L. Phillips[7], Tim Baker[7], Casey M. Ryan[1]

1 *School of Geosciences, University of Edinburgh, Edinburgh, UK*
2 *Royal Botanic Garden Edinburgh, Edinburgh, UK*
3 *Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy*
4 *Department of Geography, University of Exeter, Exeter, UK*
5 *Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-graduação em Sistemática, Uso e Conservação da Biodiversidade, Fortaleza, Brazil*
6 *Department of Biodiversity, Bioscience Institute, São Paulo State University UNESP, Rio Claro, São Paulo, Brazil*
7 *School of Geography, University of Leeds, Leeds, UK*

Corresponding author: Lucy H. Wells (l.h.wells@sms.ed.ac.uk)

Editor Janet Franklin
Received 28 December 2024 ♦ **Accepted** 12 February 2025 ♦ **Published** 14 March 2025

Abstract

Accurate biome delineation is difficult where biomes occupy the same climatic space, as is the case for tropical dry forest and savanna. The resulting confusion limits our ability to understand and manage impacts of global change on these biomes. To address this, we developed an unsupervised, repeatable method to delineate biomes and their component functional ecosystems, based on landscape-level vegetation structure measured using remote sensing and an understanding of the ecology of the region. This approach contrasts with previous definitions, based on climate differences amongst savanna, dry forest and rain forest.

Using the heterogeneous north-east Brazil, where several biomes interdigitate, as a case study, a hierarchical functional ecosystem classification is proposed that aligns with both the IUCN Global Ecosystem Typology (GET) and previous work. Based on fuzzy clustering of remotely sensed vegetation attributes, seven groups were found, identified as rain forest, cerrado (savanna) and five caatinga vegetation groups. These groups broadly align with the literature, for example, sedimentary and arboreal caatinga. These groups align with three 'Ecosystem Functional Groups' (EFGs) described by the IUCN GET and, additionally, suggest there is a new, fourth EFG in the region: non-pyric shrublands. Random Forest models showed soil pH was the most important environmental variable distinguishing these vegetation groups.

These results suggest a remotely sensed structure-based approach is an effective method for operationalising the IUCN GET. North-East Brazil – where many EFGs are interdigitated – serves as a challenging case study and, therefore, we hope our approach will have generality for other regions globally.

Highlights

- There are seven vegetation groups in northeast Brazil, including savanna, rain forest and five types of caatinga.
- Most of these vegetation groups align with the IUCN Global Ecosystem Typology 2.0, but non-pyric shrubland (caatinga) vegetation may represent a new Ecosystem Functional Group.
- Soil pH is the strongest determinant of vegetation distribution in northeast Brazil.
- Remote sensing can provide objective, spatially explicit information on vegetation types in the region, largely consistent with previous vegetation classifications.
- Accurate biome mapping is vital for management, as biomes differ in ecosystem function and consequently require different management.

Keywords

Biome, Brazil, caatinga, IUCN, remote sensing, soil, vegetation structure

PENSOFT®

Introduction

Biomes are a key concept in ecology and biogeography and are generally now understood to be vegetation units that occupy a large geographical area across continents and have distinct ecosystem functioning (Moncrieff et al. 2016; Mucina 2019). Most biome definitions are part of hierarchical classification systems, grouping ecosystems together into broader categories that can be used at a global scale. Biome classifications, therefore, provide a framework for understanding similarities and differences between major ecosystems and additionally for monitoring change and managing these ecosystems. Biome classifications and maps are commonly applied in research on global conservation, land-use dynamics, fluxes of matter and energy and climate change (Mucina 2019). They are also a useful tool for communication outside academia, given the intuitive nature of many biome maps, which show vegetation formations, such as tropical rain forests and deserts, which are familiar to the general public.

Classically, biome definitions and classifications have often been climate-based, with an emphasis on precipitation and temperature to delineate global biomes (Schimper 1903; Whittaker 1970; Mucina 2019). However, this does not reflect the modern concept of biomes defined by distinct ecosystem functioning (Moncrieff et al. 2016; Mucina 2019). If biomes could be defined on climate alone, accurate mapping would be straightforward. Although macroclimate is important in structuring biomes, it fails as a single predictor in many cases, especially where multiple states are possible under the same climate conditions (e.g. rain forest – savanna transitions; Bond et al. (2005); Moncrieff et al. (2014)). The situation of multiple biomes in the same climate space is particularly the case in dry tropical regions (Holdridge 1967; Whittaker 1970; Bond 2005; Higgins et al. 2016; Mucina 2019; Ocón et al. 2021) such as the South American ‘dry diagonal’, which encompasses seasonally dry systems from north-eastern Brazil to Argentina, Paraguay and Bolivia. In this region, several studies have found climate to be a poor predictor of biome distribution (Murphy and Bowman 2012; Silva de Miranda et al. 2018; Castro Oliveira et al. 2019). Here, edaphic differences and disturbance regime (fire and herbivory) are important drivers as in other dry tropical regions (Dantas et al. 2016; Dexter et al. 2018; Silva de Miranda et al. 2018; Castro Oliveira et al. 2019). However, other drivers such as soil properties and herbivory either lack accurate global scale maps (e.g. SoilGrids global data are highly interpolated, Poggio et al. (2021)) or are intrinsically hard to quantify, as in the case of herbivory. This means fine scale biome maps are unavailable to users designing appropriate land management and conservation planning, whilst modelling the response of vegetation to climate and other global challenges is compromised.

The IUCN have recently designed a global, hierarchical classification system, the Global Ecosystem Typology 2.0 (GET), to provide a global foundation for ecosystem assessments, sustainable management and conservation (Keith et al. 2020b, 2022). The IUCN GET describes the first three tiers of the hierarchical system as: 1) ‘realms’ (terrestrial, freshwater, marine, subterranean and atmosphere); 2) ‘biomes’ and 3) ‘ecosystem functional groups’ (EFGs). Division of the IUCN biomes into EFGs is based on relevant ‘assembly drivers’ such as resource filters, disturbance regime filters and biotic interactions. These EFGs are equivalent to biomes in many other systems and general usage, for example, the ‘T1 Tropical-subtropical forests’ biome includes the EFGs ‘T1.1 Tropical subtropical lowland rainforests’ and ‘T1.2 Tropical subtropical dry forests and thickets’, which elsewhere have been distinguished as distinct biomes (e.g. by Dexter et al. (2018); Pennington et al. (2018)). EFGs are functionally (and, in practice, structurally) different groups determined by the expression of ecological drivers along temporally variable multidimensional gradients and can be geographically scattered in patches across continents and the world. As an example, within the GET ‘T1 Tropical-subtropical forests’ biome, the ‘T1.1 Tropical subtropical lowland rain forest’ and ‘T1.2 Tropical subtropical dry forests and thickets’, are separated by both contrasting vegetation structure and phenology and differences in drivers, such as water availability, linked to soil and substrate properties (Pennington et al. 2020; Keith et al. 2020a, 2022). EFGs are the hierarchical level that allows correspondence between local and global classification systems (UN 2021) and have recently been compared to Brazilian biome maps by the Brazilian Institute of Geography and Statistics (IBGE 2021).

The IUCN GET tries to address many of the critiques of biome classification noted above, presenting a hierarchical approach to global vegetation delimitation which emphasises the processes that shape ecosystem properties and the interactions between processes and vegetation form. It aims to meet six criteria: 1) incorporate ecosystem functions and ecological processes; 2) encapsulate characteristic biota of ecosystems; 3) conceptual consistency at the global scale; 4) scalability; 5) provide spatially explicit units; and 6) parsimony (Keith et al. 2020b; Keith et al. 2022). There has been some work towards the fifth criterion for the IUCN GET (providing spatially explicit units as a wall-to-wall map at the level of EFGs), for example, by providing indicative maps of the likely occurrence of each EFG (Keith et al. 2020b). However, further development in mapping of EFGs is required, at a suitable resolution for monitoring change and designing effective conservation measures (Nicholson et al. 2021; Keith et al. 2022). This has been done for some individual EFGs including tidal mud flats (Murray et al. 2019) and tropical cloud forests (Karger et al. 2021), but not over a contiguous area with several contrasting and interdigitating EFGs. This work aims to address this gap, by mapping multiple interdigitated EFGs in northeast Brazil.

Here, we identify vegetation structural groups using remotely-sensed metrics that describe ‘vegetation expression’. This is an unsupervised, bottom-up approach that does not make any *a priori* assumptions about what vegetation types are found in the region (see Higgins et al. (2016, 2023) for examples of similar work). This contrasts with supervised classification (“top down”) approaches whereby maps are made of vegetation types expected in the region (Olson et al. 2001; Dinerstein et al. 2017). Our approach is more objective as it is not based upon a preconceived idea of what the resulting vegetation groups should be and is repeatable, as the maps are not dependent upon the authors’ ideas of a particular vegetation type (Higgins et al. 2016; Conradi et al. 2020). The term ‘vegetation expression’ is used to describe the overall behaviour of the vegetation, as a combination of attributes representing vegetation structure, phenology and, as an indicator of fuel load, fire regime. We identify vegetation structural groups, which are subsequently aggregated into EFGs according to the IUCN GET. The approach is designed to fulfil the fifth criterion of the GET (Keith et al. 2022), providing spatially explicit maps, based upon an understanding of vegetation ecology, structure and function, using recent advances in remote sensing that have improved our ability to distinguish vegetation types in the dry tropics. Overall, this paper aims to address three questions:

- Which structural vegetation groups are present in the NE of Brazil and how do they differ in terms of previously hypothesised biome determinants, such as climate, geology, soil and human influence?
- How do vegetation groups that emerge from an unsupervised classification relate to previous descriptions of the vegetation in the region?
- How do the observed vegetation groups correspond to EFGs in the IUCN GET?

We focus on NE Brazil because of its well-known heterogeneity of vegetation types and biomes, which has made it a useful case study to test biome conceptualisations and mapping in previous work (Beuchle et al. 2015; Silva de Miranda et al. 2018; Moonlight et al. 2020; Cardoso et al. 2021). The region largely comprises the caatinga – a complex set of vegetation types variously – and controversially – ascribed to dry forest, scrub, savanna and the succulent biome (Ringelberg et al. 2020). In addition, the study region includes areas of tropical rain forest and pyric savanna (cerrado), which provide useful, less controversial, groups for comparison. The study region also includes highly human-disturbed areas, which are likely to have been altered in terms of vegetation structure, some of which is pre-Colombian and extends to more recent intensive cattle and soybean agriculture, especially in cerrado vegetation. It is estimated that up to 60% of the Caatinga Region has been converted into anthropogenic ecosystems (Silva et al. 2017, chapter 13). Small-scale shifting agriculture and free grazing by livestock, predominantly cattle and goats, are an important land use in the region, in addition to fuelwood harvesting, mining and charcoal production (Schulz et al. 2017). The generally small-scale land use has created a fuzzy mosaic of land cover across the region, which is difficult to study and map. Chronic anthropogenic disturbance is complex, mediated by water availability (Silva et al. 2017, chapter 13) and spatially uneven, although it is generally higher in the east of the region (Antongiovanni et al. 2020).

Methods

Description of the study region

The study was undertaken in NE Brazil, including the edges of the Amazon and Atlantic Rain Forest regions and parts of the cerrado savanna, but consisting primarily of the Caatinga Region. In the scientific literature, “caatinga” refers to both the region and various types of vegetation. In the Brazilian lexicon, the Atlantic Forest, Amazon Forest, Cerrado and Caatinga were previously referred to as Domains and are now referred to as Biomes (IBGE 1993; Souza Jr and Azevedo 2017), but here we use the term ‘region’ to not confuse with specific vegetation types, EFGs or biomes. Caatinga vegetation is highly variable, but is generally seasonal with physiognomies ranging from open shrublands to tall seasonally dry tropical forests with a closed canopy (Andrade-Lima 1981; Quieroz et al. 2017). A grassy understorey is absent and fire is rare. Transition zones around the core Caatinga are estimated to be 14% larger than the core itself (Cardoso et al. 2021).

NE Brazil has globally high levels of species endemism and floristic compositional heterogeneity (Santos et al. 2012; DRYFLOR 2016; Bueno et al. 2018). In addition, the region is likely to experience temperature increases and increased aridity with climate change (Pörtner et al. 2022). Interaction between human disturbance and climate change is likely to have a strong negative impact on woody plant diversity (Rito et al. 2017). Furthermore, NE Brazil is the world’s most populated semi-arid land, whilst also one of Brazil’s poorest regions (Melo 2017). Consequently, expected effects of climate change will impact many of the country’s most vulnerable citizens. Caatinga biodiversity has a long and complex relationship with humans and is often considered to be highly anthropogenically disturbed, predominantly as a result of small-scale agriculture, but increasingly through irrigated crop production (Silva et al. 2017, chapter 13).

Remote sensing of vegetation expression and clustering

The characteristics of tropical biomes in terms of attributes describing vegetation expression were identified via a literature review of all biomes potentially present within the region of interest. Remote sensing products were identified and obtained that could describe a subset of these attributes (Table 1).

Table 1. Sources of vegetation attribute data used in the clustering analysis and justification for their inclusion.

Vegetation Attribute	Data Source	Justification
Above ground woody biomass density (AGB, Mg/ha)	Globbiomass AGB data product for the year 2010 (+/- 1 year) (Santoro et al. 2018), with 150 m resolution.	Above ground woody biomass is highly correlated to variation in caatinga physiognomies (Castanho et al. 2020) and so was included as a measure of amount of woody vegetation.
AGB Heterogeneity	Coefficient of Variation (CV) of the Globbiomass AGB product (Santoro et al. 2018), measured using a 5 × 5 pixel moving window (using the function 'focal' from the package 'terra' (Hijmans 2024).	Captures spatial variation, to help distinguish areas which may have uniform biomass in comparison to those with high heterogeneity. This may capture distinct vegetation formations like cerrado savanna, which have high variability in tree cover.
Seasonality	Calculated from NDVI data from MODIS MOD13A1 V6.1 product, between 10/06/2000 and 10/06/2022 (Didan 2021). Raw resolution of the NDVI product is 500 m. Seasonality was calculated as ($NDVI_{90th}$ – $NDVI_{10th}$)/ $NDVI_{50th}$ using Google Earth Engine. $NDVI_{90th}$ $NDVI_{50th}$ and $NDVI_{10th}$ are the 90^{th}, 50^{th} (median) and 10^{th} percentile of all the images within the time period. Only good quality pixels were used for these calculations, as determined using the products' quality reliability flags (Didan 2021).	Deciduousness of caatinga is conspicuous, with leaf flushing at the start of the wet season, but there is phenological variation across caatinga vegetation. For example, caatinga on sedimentary soils is less strongly influenced by rainfall (Rocha et al. 2004; Queiroz et al. 2017). Therefore, including seasonality in the analysis may be able to distinguish different ecosystems, including different forms of caatinga.
Wet Season Leaf Area Index (LAI_{wet}). LAI represents half the total area of green elements of the canopy per unit horizontal ground area. This incorporates all the canopy layers, including the understorey.	Data from the Copernicus Global Land Service, over the period 2000–2020, at 1 km resolution. December to mid-June were considered to be the wet season. All images within these months were used to make a mean wet season LAI data layer.	Inclusion of LAI in different seasons aimed to capture both bulk difference in total leaf area and responses to seasonal conditions.
Dry Season Leaf Area Index (LAI_{dry})	As above, but for mid-June to November.	This layer aimed to represent the LAI in the driest part of the seasonal cycle of each pixel.
Fire Count (no. fires in 21 years)	Calculated from burned area data from the MODIS MCD64A1.061 product between 01/01/2001 and 11/01/2021, with the raw resolution of 500 m (Giglio et al. 2021). We used the burn date information to calculate fire frequency as the total number of days that the pixel had experienced fire during the time period, using Google Earth Engine.	A proxy for the presence of a grassy underlayer, which is otherwise difficult to determine directly using remote sensing. It is assumed areas with a grassy understorey represent a regime with fire disturbance. This also aims to reduce the reliance upon tree structure in defining dry tropical biomes, which should reduce misclassification of grassy biomes as forest (Parr et al. 2014).
Canopy Height (m)	Global canopy height product for 2019. The product was made by interpolating GEDI LiDAR estimates of canopy height (RH95) with 2019 Landsat analysis-ready time-series data. The data was accessed via Google Earth Engine (Potapov et al. 2021). Raw resolution of the data was 27 m.	Torello-Raventos et al. (2013) found canopy height was important for differentiating between dry vegetation types.

Data cleaning

All data were resampled to 1 km resolution using bilinear interpolation, calculated using the 'resample' function in the R package 'raster' (Hijmans 2022). Resampling allowed analysis to be of fine enough resolution to identify transitions between vegetation types, whilst allowing reasonably fast computation. The burn count data was highly right-skewed and so was transformed by taking the natural logarithm (after adding one) to increase the normality of the data.

Exclusion of non-natural vegetation is important to prevent conflation of human-dominated areas with naturally different structural vegetation groups. We thus removed areas with non-natural land covers using the MapBiomas land cover product for 2015 (along with water covered areas; MapBiomas 2022). However, there remains potential for areas with some human modification to be present in the analysis and, for land use change to have occurred over the time period examined (2000–2022) – see Suppl. material 1: table S1 for a full list of land-cover types removed from the study.

Clustering

Fuzzy spatial clustering was used to investigate the degree to which pixels belonged to multiple groups. This was implemented by using the geocmeans package in R, version 0.3.3 (Gelb 2021, 2023). Fuzzy clustering was used because it allowed the inclusion of an 'unclassified' group where the probability of the pixel being in any identified group was low (< 0.45). This approach has been used in similar studies and is often appropriate ecologically due to the continuous nature of vegetation transitions, particularly in dry tropical regions (Torello-Raventos et al. 2013; Feilhauer et al. 2021). In addition, a fuzzy clustering approach aligns with the

IUCN GET theoretically, as the GET acknowledges partial membership of locations to multiple EFGs. Spatial clustering allows consideration of the spatial proximity between observations, improving the suitability of clustering for spatial data (Cai et al. 2007; Zhao et al. 2013; Gelb and Apparicio 2021). Clustering was done using the 'robust' function in the geocmeans package, which improves the ability of the clustering to identify groups of different sizes, shapes and density. This is done by normalising the distance between the observations and the centre of the groups (Tsai and Lin 2011). We used the silhouette index, explained inertia and Xie-Beni index to assess cluster quality, select the optimum number of clusters (k) and the parameters of the clustering algorithm (Gelb and Apparicio 2021). The algorithm parameters were: m, which controls the degree of fuzziness, β, which controls the speed of convergence and classification crispness, α, which represents the weight of space in the analysis and the moving window size (Gelb and Apparicio 2021). Selecting the most appropriate number of clusters – and other parameter values – is subjective and here we need to balance multiple indices and parameters. We selected the parameter values sequentially, optimising them one by one. The parameters k and m were first selected, based on classifications using the standard fuzzy c-means algorithm. We aimed to: i) maximise the silhouette index, ii) find the k at which increases in explained inertia were marginal and iii) minimise the Xie-Beni index. Then we used the selected values of k and m in the generalised fuzzy c-means algorithm to explore cluster quality metrics for a range of values of β, selecting the value of β, based on the minimised combined rank of the Xie-Beni and Silhouette indices. Finally, we optimised the value of α and the moving window size for spatial generalised fuzzy c-means clustering, again aiming to minimise the combined rank of the Xie-Beni and Silhouette Index.

Random forest models to quantify the determinants of vegetation groups

To understand the determinants of each cluster, hypothesised explanatory variables were selected that have previously been used to explain biome distribution in NE Brazil (e.g. Silva de Miranda et al. (2018); Moonlight et al. (2020); Cardoso et al. (2021)). These were: the climatic variables Mean Annual Temperature (MAT), temperature seasonality (standard deviation of monthly temperature * 100), Mean Annual Precipitation (MAP), precipitation seasonality (coefficient of variation of monthly precipitation), precipitation in the wettest and driest months and precipitation in the warmest quarter (Fick and Hijmans 2017); edaphic factors (soil water capacity, Cation Exchange Capacity (CEC), soil sand content, soil organic carbon content (SOC), soil pH and soil depth, Height Above Nearest Drainage (HAND) (Yamazaki et al. 2019; Poggio et al. 2021), geology, elevation (Yamazaki et al. 2017); and measures of human disturbance (population density, road density and length of time to nearest city (Meijer et al. 2018; Weiss et al. 2018; WorldPop 2018). Climate data were sourced from WorldClim (Fick and Hijmans 2017) and soil data from SoilGrids (Poggio et al. 2021). Although non-natural land cover was removed from the clustering, human-disturbance variables were included in the determinants analysis to account for the fact that intensity of disturbance is experienced as a gradient, rather than a dichotomy. In addition, these measures of human disturbance were included as a check for residual impacts not removed by the removal of non-natural land cover classes. The predictor data were cleaned and processed in the same manner as the vegetation attributes. To check for correlation in the predictors, which can complicate interpretation of random forest outputs, a Pearson's correlation test was carried out. Where $r > 0.8$, one of the correlated predictors was removed, based on which variable was considered more mechanistically meaningful. Data sources and justification for their inclusion are found in Suppl. material 1: table S2 and the correlation matrix from the Pearson's correlation test is found in Suppl. material 1: table S3.

The IUCN GET identifies five types of ecological drivers of biome distribution, three of which are encapsulated in our random forest analysis as variables ecologically relevant to NE Brazil (resource drivers, ambient environment and human activities). Given that only a few drivers shape the properties of ecosystems (Keith et al. 2020b), it is reasonable to select specific drivers for this analysis focused upon dry tropical ecosystems rather than a global analysis. Resource drivers include those related to the availability of water, carbon and nutrients. Ambient environment factors include climate variables and are important for modifying resource availability. Thirdly, human activities are included in the random forest classification. Although disturbance regime is another driver highlighted by the IUCN GET, natural disturbance was not included in the random forest analysis (note that fire was used in the clustering methods as a proxy for the presence of grass, which is difficult to measure using remote sensing techniques).

Random forest classification (Breiman 2001) was carried out using the remaining (non-correlated) explanatory variables with group identity as the response variable. The 'randomForest' package in R was used for this analysis (Breiman et al. 2024). We used a random 70:30 train:test split, with model accuracy assessed by calculating the percentage of the pixels in the test group whose vegetation group was correctly predicted by the random forest model. Partial dependence plots were used to illustrate the partial effect of each predictor on the log odds of a pixel being in each group, when all other predictors are held at their mean.

Naming the outputs of the unsupervised classification

We reviewed the literature to identify the vegetation types that we expected the classification to detect. In order to describe the vegetation groups found by the clustering, these structural descriptions of differences between vegetation groups were cross referenced with additional sources of information: i) previous descriptions of the geographical distribution of the vegetation types compared to our groups (Fig. 1A), ii) the mean and spread of the soil, climate and human disturbance variables used in the random forest analysis (Suppl. material 1: fig. S3, tables S5–S7), iii) quantitative descriptions of the vegetation properties of each group (Fig. 2, Suppl. material 1: table S4) and iv) floristic information (Suppl. material 1: table S8). We used data from Silva de Miranda et al. (2018) to match floristic groups in the region to our groups, based on spatial co-location. Silva de Miranda et al. (2018) used hierarchical clustering to assign 4,103 sites across lowland South America and neighbouring subtropical areas into biomes, based upon tree species composition (Silva de Miranda et al. 2018). The floristic groups corresponded to: Amazon Forest, Atlantic Forest, savanna, seasonally dry tropical forest (SDTF) and chaco. All of these groups are present in our study region, except chaco vegetation. Sites with an assigned floristic group were joined to the structural map using qgis (QGIS Team 2021) and the percentage of each floristic group present in each vegetation group which we identified was calculated.

Comparison to the IUCN Global Ecosystem Typology

The structural groups identified in the clustering were compared to the descriptions of EFGs in the IUCN GET. We focused on ‘T1.1 Tropical subtropical lowland rain forests’, ‘T1.2 Tropical subtropical dry forests and thickets’, ‘T3.1 Seasonally dry tropical shrublands’ and ‘T4.2 Pyric tussock savannas’. We considered the ecological equivalence between our vegetation groups and the IUCN EFG descriptions, with a particular focus upon the links between vegetation structure and ecosystem processes emphasised in each EFG. Furthermore, we considered the results of IBGE’s analysis of EFGs in Brazil (IBGE 2021). Comparisons were also made to other vegetation and biome maps, including MapBiomas (MapBiomas 2022) and the Terrestrial Ecoregions of the World datasets (Olson et al. 2001); these comparisons can be found in the Suppl. material 1 (Suppl. material 1: figs S4, S5).

Results

Clustering output

The silhouette index, explained inertia and Xie-Beni indices gave varying results for the optimum values of *k*, the number of clusters (Suppl. material 1: fig. S1). The Xie-Beni index suggested that five or six clusters is the optimum, depending on the value of *m*. Contrastingly, the silhouette index suggested seven or nine clusters, slightly depending on the value of *m*. The explained inertia values improved as *k* increased, although there were diminishing returns beyond $k = 7$ or 8 clusters, depending on the value of *m*. As such, it was not possible to fulfil all three of our conditions for selecting the optimum value of *k* and *m*. In balancing the different criteria, we considered $k = 7$ as being the best value and focus our analyses and discussion on seven groups. This set of seven includes multiple caatinga vegetation groups besides the expected divergent vegetation groups (namely, rain forest and cerrado). Based on the silhouette score suggesting seven or nine clusters, we compared the clustering results for $k = 7$ and $k = 9$ in Suppl. material 1: appendix S1 and found that the change in *k* does not alter our main arguments, except that the definition of groups of caatinga vegetation is sensitive to the number of caatinga-type clusters. After selecting $k = 7$, we selected $m = 1.4$, $\beta = 0.6$, $\alpha = 0.7$ and a 3 × 3 pixel moving window, following sequential optimisation which minimised the combined ranks of the silhouette and Xie-Beni indices (Suppl. material 1; Fig. 1).

We also conducted the clustering analysis using a hard clustering approach, but found little difference between fuzzy and hard clustering results. In addition, we also tested whether using PCAs of the vegetation attributes as the clustering input altered our results and found that it did not. Details of these analyses are found in Suppl. material 1: appendix S2.

The seven groups are mapped in Fig. 1A (and individually in Suppl. material 1: fig. S2), with the vegetation attributes summarised in Fig. 2. In the section, Naming the groups and linking to previous vegetation descriptions, we name the groups, based on our interpretation of the literature and, for simplicity, we use these names henceforth. The groups are named scrubby caatinga, hyper-seasonal caatinga, heterogeneous caatinga, transition caatinga, arboreal caatinga, cerrado and rain forest. The seven groups are primarily distinguished by the amount of the vegetation in terms of canopy height, biomass and LAI. The presence of fire is the main defining factor separating cerrado from the other vegetation groups (Fig. 2). Most vegetation groups are identified by a particularly high value of at least one vegetation attribute, for example, rain forest has high biomass, LAI and canopy height and heterogeneous caatinga has high biomass heterogeneity. Only 0.24% of the clustered area had a probability < 0.45 of being in any group and was grouped as unclassified.

Determinants of vegetation type

Random forest classification was highly accurate at predicting cluster membership, based on the predictors listed in section Random forest models to quantify the determinants of vegetation groups. The random forest model had an accuracy value of 81% for the held-out ‘test’ data set. The per group error rate from the random forest confusion matrix was below 24% for all groups, except the unclassified group, which had an error rate of 95%. Based on the importance values (which quantify the mean decrease in Gini coefficient when that variable is randomised; Fig. 3), the

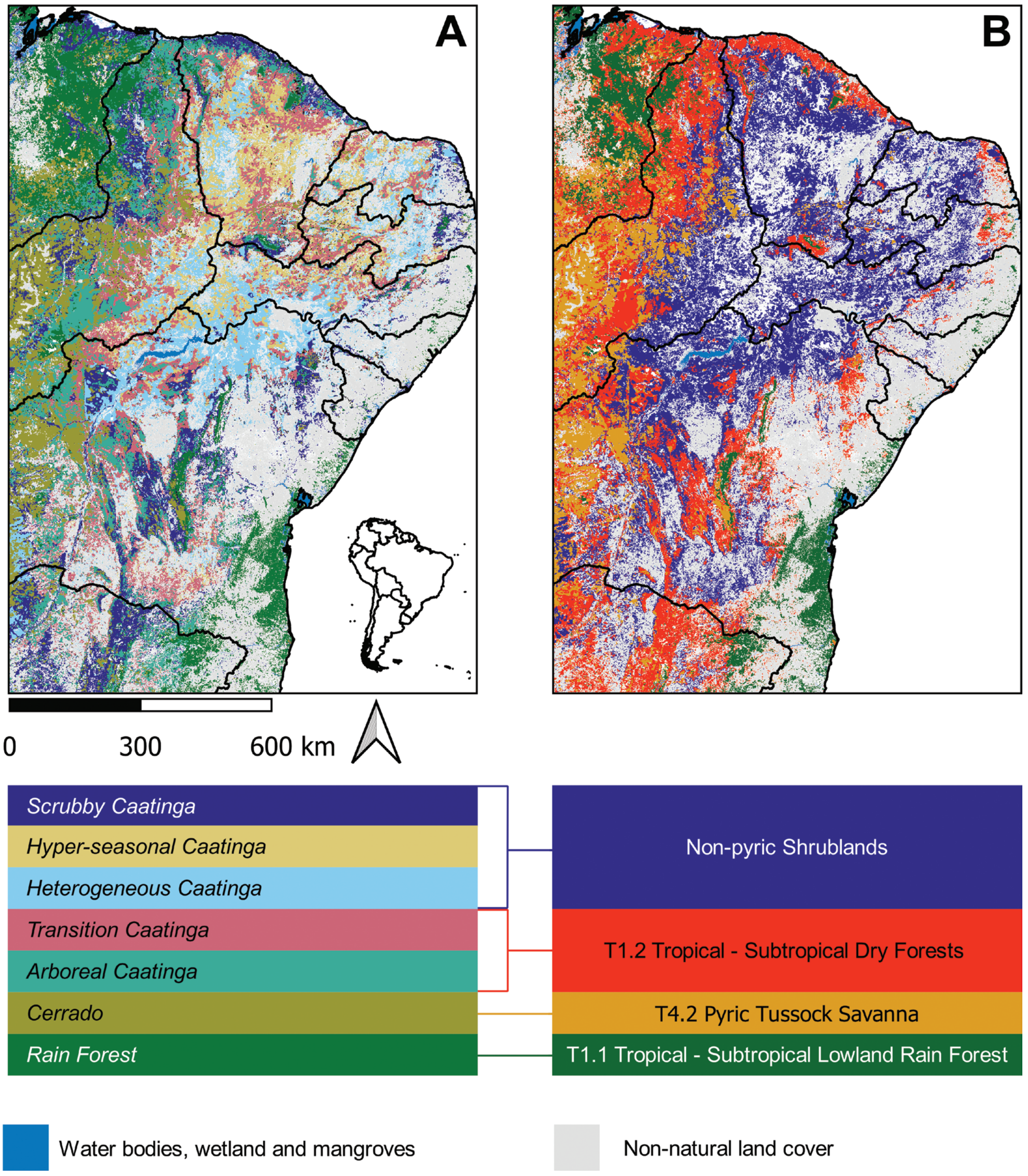

Figure 1. Comparison of the vegetation structural groups identified by c-means clustering of vegetation attributes (**A**) with the IUCN Global Ecosystem Typology 2.0 Ecosystem Functional Groups in NE Brazil (**B**). The maps of the EFGs in **B** is produced by the aggregation of the structural groups in **A**, as per the lower panel, rather than the GET indicative maps. Scrubby caatinga, heterogeneous caatinga and hyper-seasonal caatinga represent an additional EFG not yet described by the IUCN GET which we term 'non-pyric shrublands' as they consist of vegetation that does not burn, but which structurally resembles a shrubland or short-statured forest rather than a tall canopy, closed forest.

most important predictor of group membership was soil pH (mean decrease in Gini coefficient = 65,000). Following this were climate and soil variables (47,000 to 17,000) with travel time to nearest city the most important human use variable (27,000). The least important variables included all the other human disturbance related variables, soil cation exchange capacity, soil depth, Height Above Nearest Drainage, soil water content and geology. Surprisingly, geology was the least important variable by some distance – and repeating the analysis without soil variables (which might be somewhat correlated) gave the same result.

The partial dependence plots (PDPs; Fig. 4) for the three most important explanatory variables showed substantial, mostly non-linear effects. Overall, rain forest and arboreal caatinga groups show similar trends, being located in wetter areas. For rain forest, there is a strong pH and precipitation seasonality threshold, being more likely to be located on soil with pH < 5.5 and in areas with precipitation seasonality < 90. The caatinga groups are located in drier areas, with high soil pH, with heterogeneous and hyper-seasonal caatinga, in particular, having a strong affinity with for soil pH > 5.8 and hyper-seasonal caatinga an affinity for precipitation seasonality > 95. The rain forest and hyper-seasonal caatinga groups showed opposite associations with all three of the most important environmental variables.

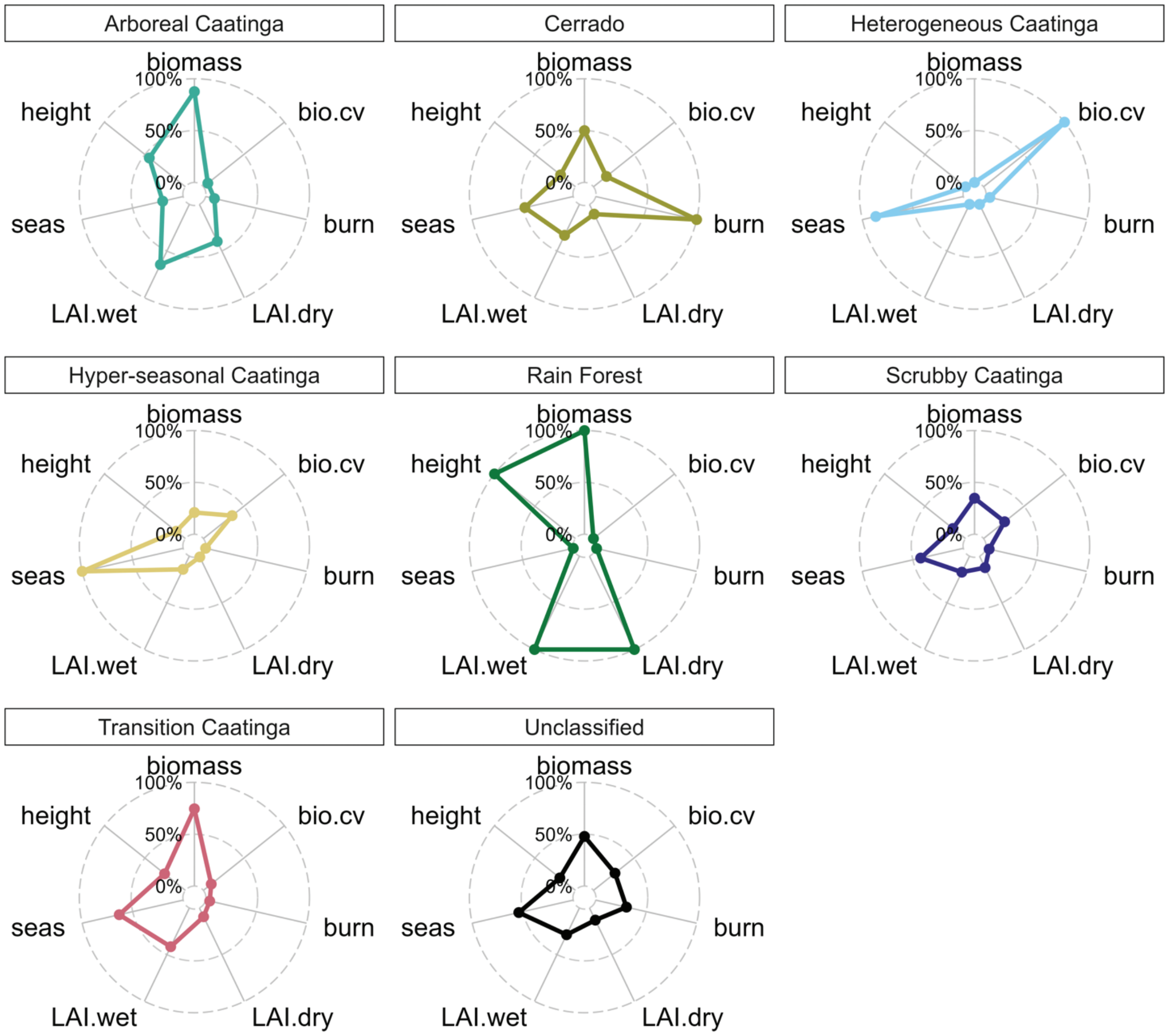

Figure 2. Attributes of the vegetation in each group in NE Brazil. The scaled mean vegetation attribute for each vegetation group is shown, as a percentage of the maximum of each vegetation attribute. The vegetation attributes were (clockwise from top) aboveground woody biomass, aboveground woody biomass heterogeneity (bio.cv), fire frequency (burn), mean Leaf area index (LAI) in the dry season (LAI.dry), mean LAI in the wet season (LAI.wet), NDVI seasonality (seas) and canopy height (height).

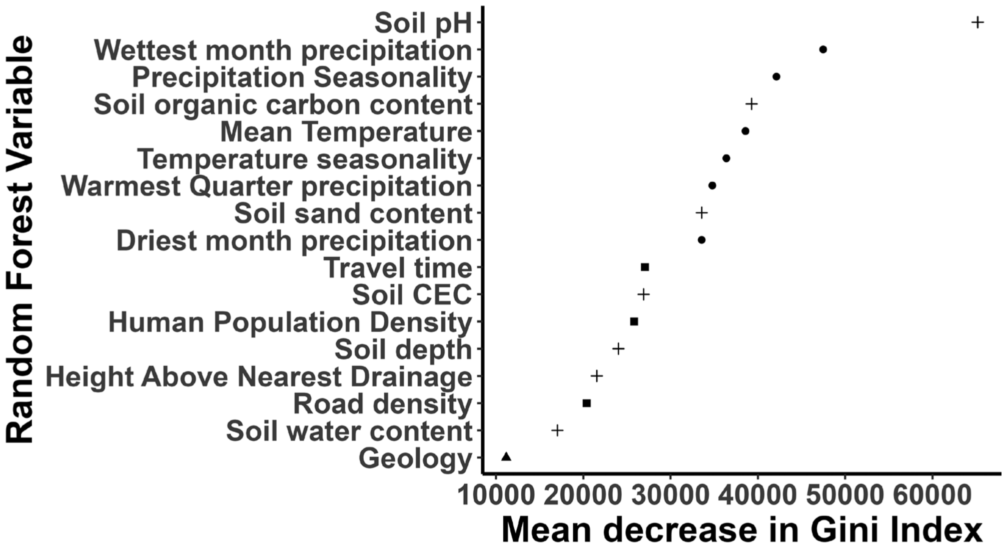

Figure 3. Variable importance plot illustrating the importance of explanatory variables in determining the structural groups. Values indicate the mean decrease in Gini Index after removal of each variable from the random forest classification. Shapes indicate the type of variable; cross = soil variables, circle = climatic variables, square = human-related variables, triangle = geology.

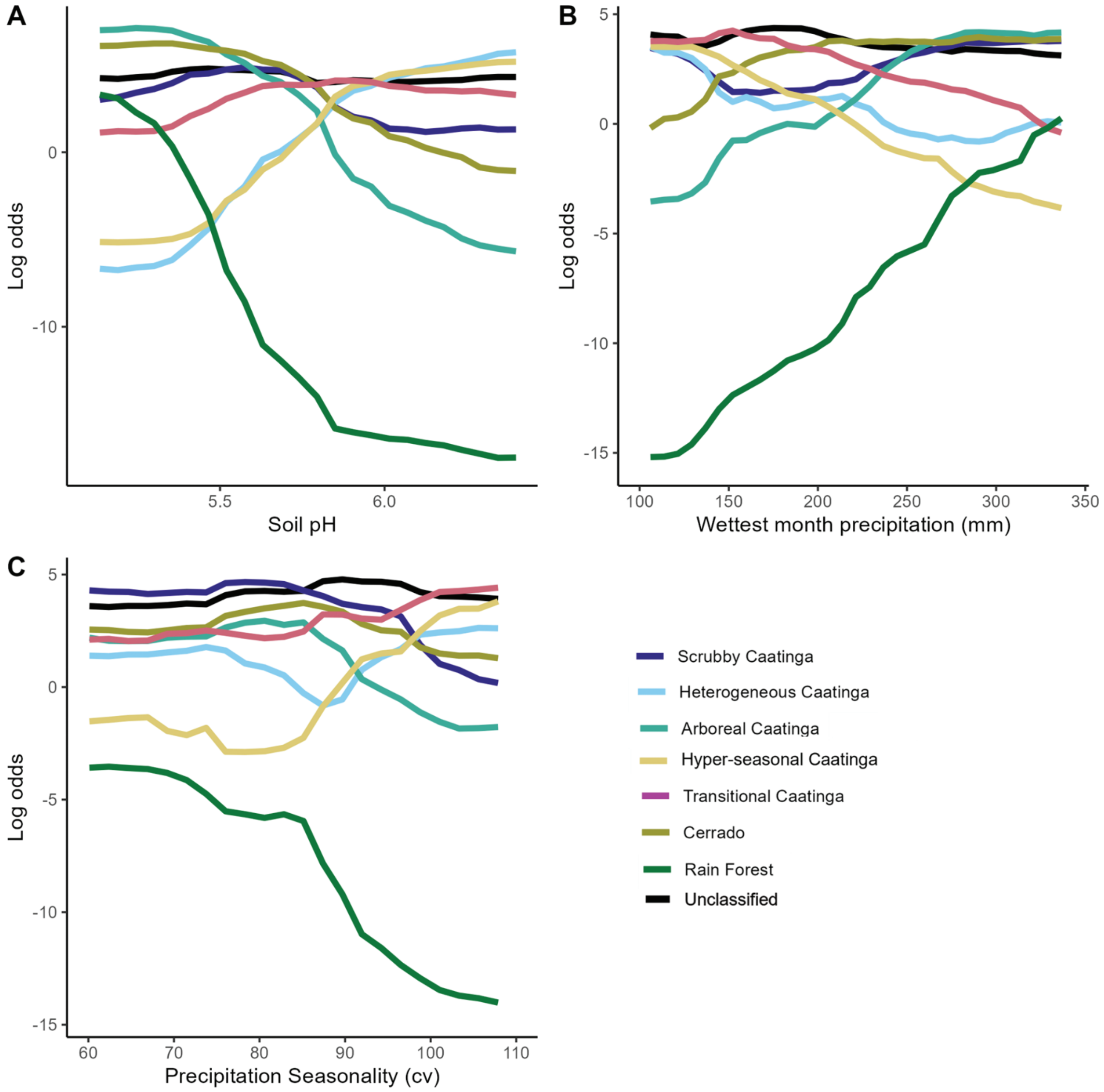

Figure 4. Partial dependence plots showing the effect of the three most important variables predicating vegetation type in northeast Brazil. **A.** Soil pH; **B.** Precipitation in the wettest month; **C.** Precipitation seasonality (coefficient of variation of monthly precipitation). The y axis shows the log odds of a pixel being in each vegetation type, given varying levels of the focal predictor variable, with all other predictors held at the mean of the dataset.

The hyper-seasonal caatinga group generally was more sensitive to environmental variables compared to heterogeneous caatinga, although they exhibit similar trends. This could suggest the hyper-seasonal caatinga group has stricter environmental requirements and is less likely to be found in unfavourable conditions.

Naming the groups and linking to previous vegetation descriptions

Our review of the literature identified nine vegetation groups we would expect to find in this region: rain forest, cerrado savanna, restinga, campos rupestres, cerradão and four groups of caatinga vegetation: crystalline, sedimentary, karst and arboreal (Fig. 5). Arboreal caatinga is forest-like, but crystalline, sedimentary and karst caatinga have distinct forms, determined by their underlying geology (Queiroz et al. 2017; Fernandes et al. 2022). Crystalline caatinga is the most typical caatinga phytophysiognomy, with deciduous, spinescent trees, a high diversity of herbaceous climbers and non-woody plants and few grasses (Quieroz et al. 2017). Crystalline caatinga is commonly found in the Sertaneja Depression, with shallow, fertile, stony soils and flat to slightly hilly topography (Queiroz 2006; Moro et al. 2016; Queiroz et al. 2017). Sedimentary caatinga – which has a more scrub-like formation – is found in Mesozoic sedimentary basins, with sandy, deep, oligotrophic soils (Rocha et al. 2004; Queiroz 2006; Moro et al. 2016; Queiroz et al. 2017) which may have a greater capacity to retain water. Thirdly, karstic caatinga is found on karst outcrops in small islands within sedimentary basins and has the richest soils in the region, distinct physiognomies and many endemic species (Queiroz et al. 2017; Fernandes et al. 2020).

We found the following links between the previous descriptions of vegetation in NE Brazil and our vegetation structural groups. One group clearly represented rainforest (of both the Atlantic and Amazon Regions). Cerrado and campos rupestres were represented together in the cerrado group and the remaining five groups represented different formations of caatinga. These caatinga groups were named arboreal caatinga, scrubby caatinga, transition caatinga, hyper-seasonal caatinga and heterogeneous caatinga. Arboreal caatinga is equivalent to the caatinga type of the same name as described in literature. Scrubby caatinga is equivalent to sedimentary caatinga and restinga. The hyper-seasonal caatinga and heterogeneous caatinga may collectively comprise the crystalline caatinga, given their geographical distribution, characteristics of high seasonality and shallow soils with high soil pH, which are typical of crystalline caatinga (Quieroz et al. 2017). Finally, transition caatinga seems to be a type of caatinga that is transitional between the more typical caatinga vegetation groups and arboreal caatinga. We did not identify a structural group corresponding to cerradão or karst caatinga vegetation.

Descriptions of the vegetation structural groups

Detailed descriptions of the vegetation groups can be found in Suppl. material 1: appendix S3; below we briefly outline each group.

Scrubby caatinga

What we term scrubby caatinga broadly matches the description and geographic distribution of sedimentary caatinga and restinga in past work (Fernandes et al. (2022), see Suppl. material 1, Fig. A3). Scrubby caatinga has intermediate biomass (28 ± 13 Mg/ha) with a low canopy height (3.5 ± 1.2 m) that is slightly taller than the hyper-seasonal (3.0 ± 1.0 m) and heterogeneous caatingas (1.7 ± 1.1 m). This could suggest a low shrubby layer contributes most biomass in this vegetation group. Restinga, or coastal woodland, is structurally variable and influenced by its sandy soils and location along the coast, but often has a shrubby structure and can be an extension of Atlantic Forest, Cerrado or Caatinga (Santos-Filho et al. 2015). Fernandes et al. (2022) summarise sedimentary caatinga as 'open scrub occupying patchily distributed nutrient-poor sandy soils'.

Hyper-seasonal caatinga

This group was identified as a distinct type of caatinga in this analysis and as part of the crystalline caatinga when compared to literature. Hyper-seasonal caatinga was the most frequent caatinga vegetation group found in the study region, comprising 21% of the clustered area. Hyper-seasonal caatinga had spatial distribution in the central north-east of the study area, within the core Caatinga Region (Cardoso et al. 2021). This group has the highest NDVI seasonality (1.00 ± 0.13) (Fig. 2). Hyper-seasonal caatinga is much more likely to be found where precipitation in the wettest month is low (170 ± 49 mm) and where precipitation seasonality is above the threshold of 95 – this is the group with the most seasonal rainfall (Suppl. material 1: table S5; Fig. 4).

Heterogeneous caatinga

This group aligns as a subset of the crystalline caatinga, along with hyper-seasonal caatinga. We describe this group as heterogeneous caatinga because it has low, but

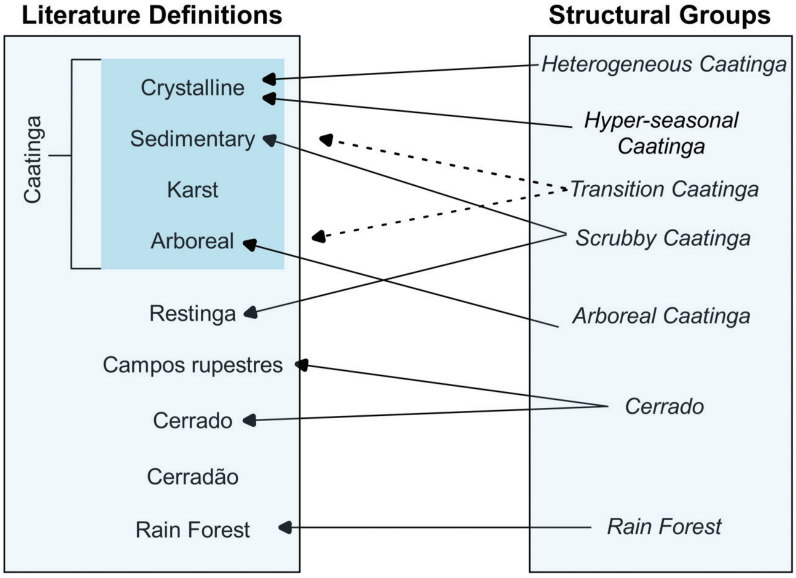

Figure 5. Comparison of vegetation groups in NE Brazil according to the literature and structural groups identified in this work by unsupervised classification of vegetation attributes. Solid lines indicate clearly identified parallels between the literature definition and our structural groups and dotted lines identify more tenuous relationships between the structural groups and literature definitions.

spatially heterogeneous biomass (biomass cv: 1.4 ± 0.5), likely due to intermittent large trees or patchy tree distribution within the 1 km^2 grids used. This may suggest that this group represents human-disturbed caatinga. It is located alongside areas masked as 'non-natural' land cover. As human disturbance also exists to some degree in 'natural' land covers, heterogeneous caatinga may represent a human – disturbed form of caatinga.

Arboreal caatinga

Arboreal caatinga is often found in areas of transition between biomes. Structurally, it could be described as 'tall caatinga', with a canopy of intermediate height between other caatingas and rain forest (Andrade-Lima 1981; Santos et al. 2012). However, its definition is contested, described as both a transition between cerrado and caatinga (Veloso et al. 1991; Santos et al. 2012) or alternatively between rain forest and caatinga (Oliveira-Filho et al. 2006; IBGE 2008; Castro Oliveira et al. 2019). Overall, we find our arboreal caatinga group to have structural characteristics intermediate between the rain forest and the other caatinga groups, having the second highest values after rain forest for biomass (65 ± 14 Mg/ha), LAI_{dry} (2.1 ± 0.7), LAI_{wet} (3.4 ± 0.7) and canopy height (6.8 ± 2.0 m). The geographical distribution of arboreal caatinga is between the rain forest and core caatinga; we, therefore, suggest that it is transitional between rain forest and caatinga, as opposed to cerrado and caatinga.

Transition caatinga

The fifth group is similar to arboreal caatinga as it is forest-like, having an intermediate biomass (56 ± 14 Mg/ha) and biomass heterogeneity (0.3 ± 0.1). It also has taller canopy height (4.7 ± 1.0 m), similar to arboreal caatinga and cerrado, as opposed to the other caatinga groups. However, its NDVI seasonality (0.7 ± 0.2) is high, being just slightly less than hyper-seasonal caatinga (1.0 ± 0.1) (Fig. 2). This suggests it is a dry forest group. It does not correspond to any of the caatinga vegetation types in the literature and we regard it as transitional between the arboreal caatinga and the more scrubby caatinga groups.

Cerrado

The cerrado is the main extent of savanna in Latin America, found towards the south and west of the study region. This vegetation group was identified as cerrado primarily because it is the only group which experiences substantial fire, suggesting the presence of a flammable grassy layer. Pixels experienced fire on average 1.3 ± 0.4 times from 2001–2021. The group is distributed in the west of the study region, with a significant patch in central Bahia, within the Chapada Diamantina (Fig. 1A). This is as expected for the cerrado biome, corresponding to the Campo Maior transition between the cerrado and caatinga (Barros and Castro 2006) and campos rupestres vegetation in the Chapada Diamantina (Fig. 1A).

Rain forest

This group has high values of all the vegetation abundance attributes (biomass of 74 ± 13 Mg/ha and canopy height of 13 ± 3.5 m), a mostly evergreen canopy with the lowest NDVI seasonality (0.15 ± 0.05) (Fig. 2). It corresponds mostly to the Atlantic Forest on the coast, with some patches in the northwest of the study area, within the main block of arboreal caatinga, corresponding to the eastern edge of the Amazon Forest. In NE Brazil, the Atlantic Forest is a tropical rain forest with tall evergreen or semi-deciduous trees, having an intermediate degree of aridity between more southern Atlantic Forest and the interior of the Caatinga (Moro et al. 2015). It transitions towards caatinga vegetation via the agreste ecotone (Oliveira-Filho and Fontes 2000). In addition, the distribution of rain forest displays some patches throughout the region; these likely correspond to Brejos de Altitude, which typically occur due to orographic rainfall, for example, in the highlands around Araripe and the Chapada Diamantina (Fig. 1A).

Unclassified

A small number of pixels were not aligned to any structural group; 0.24% of the clustered area. These pixels displayed medium values of the vegetation attributes and low value for fire (Suppl. material 1: table S4), with generally quite large values of standard deviation for the vegetation attributes. The unclassified pixels were spatially spread out across the region and, therefore, they are not likely to represent a missing vegetation group and are not discussed further.

Relating the vegetation groups to the IUCN GET

The vegetation groups were assigned to the IUCN GET EFG groups present in the study region by comparing the descriptions of the EFGs (Keith et al. 2020b) and the vegetation descriptions above.

Rain Forest

This group corresponds to EFG 'T1.1 Tropical and Subtropical Lowland Rainforest', due to its similarities in terms of ecological traits to the IUCN description – such as high biomass and LAI and absence of grasses – and the drivers including less seasonal precipitation and moist soil (Keith et al. (2020a), Fig. 1B).

Cerrado

This group corresponds to the ‘T4.2 Pyric tussock savannas’ EFG, which is described as dominated by C4 grasses, with variable tree cover and sub-decadal fire regimes (Lehmann et al. 2020). The MAP of 1030 ± 210 mm fits within the IUCN GET’s definition of 650–1500 mm (Lehmann et al. 2020). In addition, some of the cerrado structural group may correspond to campos rupestres, which is described as ‘T3.1 Seasonally dry tropical shrublands’. Campos rupestres (rupestrian grasslands) are a type of rocky grassland comprising herbaceous and shrubby vegetation. The northern portion of the core campos rupestres is located in the highlands of the southern caatinga, including the Chapada Diamantina (in Bahia State) (Conceição et al. 2016). Campos rupestres are floristically and functionally related to cerrado vegetation; several traits of rupestrian species demonstrate fire tolerance (Figueira et al. 2016).

Dry forest and shrublands

Arboreal caatinga and transition caatinga have a forest-like structure, with a higher biomass and taller canopy than the three other caatinga groups we found. Therefore, we have placed these groups in the ‘T1.2 Tropical and subtropical dry forests and thickets’ EFG (Pennington et al. 2020). Our data-driven descriptions are a good match for the EFG description of closed canopy forest, with seasonally high LAI and deciduous or semi-deciduous phenology and absence of grasses due to canopy shading (Pennington et al. 2020).

In contrast, scrubby, hyper-seasonal and heterogeneous caatinga do not easily align to an EFG within the GET. They structurally align with the GET description of ‘T3.1 Seasonally dry tropical shrublands’ (low open forests less than 6 m tall, with spatially heterogeneous vegetation (Keith and Russell-Smith 2020)), but, because they do not burn, they functionally align with the GET-defined dry forests and so would be placed in ‘T1.2 Tropical and subtropical dry forests and thickets’ (Fig. 2; Suppl. material 1: table S4). On balance we consider that their fit with T1.2 is poor. These caatinga groups share the functional characteristics of T1.2, especially the absence of fire (Pennington et al. 2020) due to limited ground fuels (and are, therefore, often included in a widely defined “seasonally dry tropical forest” biome, for example, by Pennington et al. (2020)), but do not structurally align with T1.2, as they have a low canopy height and biomass, unlike forests and thickets.

Discussion

We find seven structural vegetation groups, which overall match well with previous descriptions of the vegetation of the region. Broadly, these mappable, locally relevant vegetation groups can be classified within three IUCN EFGs or biomes: T1.2 Tropical subtropical dry forests and thickets, T4.2 Pyric tussock savannas and T1.1 Tropical-subtropical lowland rainforests. However, we also find three vegetation groups that are similar to each other, but not well aligned with the GET. The attributes of these three groups may indicate a new IUCN GET EFG category is required. The distribution of all the vegetation groups is well explained by our hypothesised drivers, with an accuracy of 81% for predicting the test dataset. Overall vegetation group appears mostly to be determined by soil pH, with only a secondary role for climate, highlighting how these biomes do not fit with a climatically deterministic view of vegetation classification (e.g. the classic Holdridge Life Zone system (Holdridge 1967)).

Seven structural vegetation groups in northeast Brazil

The clustering method identifies seven ecologically meaningful groups with different physiognomies. These structural groups align with previously recognised vegetation groups in the region – categorised using a variety of methods including floristic classifications (Queiroz 2006; Moro et al. 2016; Queiroz et al. 2017) (Fig. 5). Although we do not identify a single group corresponding to the previously described crystalline caatinga, our hyper-seasonal caatinga and heterogeneous caatinga together may comprise this formation. This suggests that our work may have generated more subdivisions within the crystalline caatinga, due to the emphasis on vegetation structure as opposed to floristic composition and evolutionary history.

The number of groups identified is suitable for capturing the heterogeneity of the region without being overly complex for interpretation at a regional scale. For further discussion of the effectiveness of our method, see Suppl. material 1: appendix S4 for an example of regions of topographic complexity; the Chapada Diamantina and Serra do Araripe. We used an unsupervised clustering approach, enabling us to categorise the vegetation as quantified by remote sensing products, rather than based upon preconceived opinions on vegetation in the region. However, there are limitations to this method, including the difficulties associated with labelling resulting groups. This interpretation may introduce biases in a similar manner to supervised classifications. Furthermore, quantifying the accuracy of an unsupervised approach to check whether the labelled groups match existing ecosystems requires extensive field effort; this may be a particular challenge in human-modified areas, which add complexity within an already complicated region. For example, we suggest the heterogeneous caatinga group may, in part, represent human-modified caatinga or may result from high levels of encroachment of invasive species (Nogueira et al. 2019). However, verifying this would take considerable on-the-ground effort. Finally, groups identified by unsupervised classification may not align with previous classification systems, as we found here for some caatinga groups, which may reduce the applicability of the research to some users. For example, we did not locate a group aligned to karst caatinga. However, we have made considerable efforts to relate our retrieved groups to previously described vegetation, in order to accurately place them within existing frameworks from literature.

Soil and rainfall determine vegetation type

Random forest classification showed that soil pH was the most important determinant of vegetation structural groups. Precipitation in the wettest month and precipitation seasonality were the next most important variables. These results broadly agree with the argument that climate is not the main determinant of vegetation type in NE Brazil and align with the findings of several authors who describe edaphic factors as important drivers of vegetation in NE Brazil (Murphy and Bowman 2012; Terra et al. 2018; Castro Oliveira et al. 2019; Maia et al. 2020; Souza et al. 2020).

The importance of soil pH as a determinant of vegetation type, may be due to its relationship with soil fertility. Hyper-seasonal caatinga and heterogeneous caatinga (collectively comprising crystalline caatinga) were associated with higher soil pH (Fig. 4), which agrees with previous findings of crystalline caatinga vegetation having higher soil pH (Castro Oliveira et al. 2019). On the other hand, rain forest and arboreal caatinga (to a lesser extent), are more likely to be found in areas of lower pH (Fig. 4). The shapes of the pH partial dependence plot for rain forest and hyper-seasonal caatinga suggests there is an abrupt threshold pH value for these structural groups, either side of which one is unlikely to be found. Other structural groups do not have such a strong association with a particular pH value (Fig. 4) and it is likely that different environmental variables determine their presence. However, it is also possible that pH may not be the key driver; other, correlated but unmeasured, environmental variables may also play a role. In particular, the use of climate variables to model pH in the SoilGrids product used here (Poggio et al. 2021) makes causality hard to infer. Furthermore, previous work has demonstrated uncertainty in SoilGrids variables in dry ecosystems (Cramer et al. 2019; Dandabathula et al. 2022), particularly for chemical as opposed to physical soil properties (Miller et al. 2024). Although some previous studies have found climatic, fire and anthropogenic variables to be of greater importance than soil in determining vegetation type, for example, in tropical Africa, (Bond et al. 2005; Greve et al. 2011; Pausas and Ribeiro 2017), others acknowledge the importance of including soil properties in studies on vegetation distribution – particularly in savanna ecosystems (Campo-Bescós et al. 2013; Arruda et al. 2015; Arruda et al. 2017; Oliveira et al. 2021).

Variables describing precipitation were the most important climatic determinants of vegetation structural group. This makes sense in a dry region where water availability is important for plants and where elevational variation is relatively limited. The IUCN EFG descriptions for the groups present in the region all describe water availability as an important driver for these groups (Keith et al. 2020b; Keith and Russell-Smith 2020; Lehmann et al. 2020; Pennington et al. 2020), which matches the importance of the precipitation regime in the random forest results. In addition, precipitation regime interacts with edaphic and other environmental variables. Maia et al. (2020) found a negative impact of precipitation seasonality and a positive impact of precipitation in the driest quarter on tree species richness in Caatinga and Cerrado sites. Both of these effects were mediated by soil sand content, being stronger in soils with less sand. Interactions between environmental variables increases the complexity in understanding determinants of vegetation structure. Hyper-seasonal caatinga has a particularly strong relationship with precipitation (Fig. 4), being much more likely to be found in areas with wettest month precipitation below 200 mm and precipitation seasonality above 95. Finally, it is important to note that these drivers display interactions including feedback loops, which makes their interpretation in structuring vegetation groups complex (see Suppl. material 1: appendix S2, p. 6 of Keith et al. (2022) for further discussion of interactions).

A difficulty in the analysis is that SoilGrids data do not include some relevant variables, including phosphorus and aluminium content which are known to be important determinants of tree species composition (Bueno et al. 2018). For example, although correlated to soil pH, aluminium content is considered a key feature differentiating caatinga and cerrado soils (Castro Oliveira et al. 2019) and, consequently, would be particularly relevant for this study. Furthermore, a caveat in using SoilGrids is that these modelled products utilise a number of covariate datasets, including climate variables, vegetation indices such as NDVI and raw MODIS bands (Poggio et al. 2021). These data are also utilised in our vegetation structural attributes, which means that they are not entirely independent, leading to some potential circularity in the results of the random forest classification. As such, further work would be required using regional sampling and modelling to improve the accuracy of this analysis, as carried out by Cramer et al. (2019) for the Greater Cape Region, South Africa.

Advancing the IUCN Global Ecosystem Typology: a missing EFG?

We have identified both coherence and challenges with the existing IUCN's GET EFGs, with the key issues surrounding the complexities of classifying dry forest, thicket and shrubland. We find that the transition caatinga and arboreal caatinga groups align well with the 'T1.2 Tropical subtropical dry forests & thickets' EFG, a finding supported by IBGE (2021), which demonstrated full equivalency of 'forested caatinga' to T1.2 (IBGE 2021). However, the placement of scrubby caatinga, hyper-seasonal caatinga and heterogeneous caatinga into existing EFGs is not clear. These groups might be considered to fall under either 'T3.1 Seasonally Dry Tropical Shrublands' or 'T1.2 Tropical-subtropical dry forests and thickets', but neither are a complete fit. This finding of non-alignment is also in agreement with IBGE (2021) which suggested that the caatinga vegetation groups have only partial equivalence to multiple EFGs (IBGE (2021) considered the caatinga partially equivalent to 'T1.2 Tropical-subtropical Dry Forests and Thickets', 'T5.2 Thorny deserts and semi-deserts' and 'T3.1 Seasonally dry tropical shrublands').

In the IUCN GET, dry forest and thickets are described as closed canopy forest, with seasonally high LAI and deciduous or semi-deciduous phenology and absence of grasses (Pennington et al. 2020). In contrast, T3.1 shrublands are low, open forests, shrublands and shrubby grasslands, which are sometimes evergreen and have a canopy height below 6 m. C4 grasses may be co-dominant, but not continuous and there are recurrent fires (Keith and Russell-Smith 2020). Scrubby caatinga, hyper-seasonal caatinga and heterogeneous caatinga are more or less shrubby in terms of remotely-sensed vegetation attributes (Fig. 6) and fit the main IUCN GET structural description of T3.1. However, according to our remotely-sensed data, scrubby caatinga, hyper-seasonal caatinga and heterogeneous caatinga do not burn, which is a key ecosystem process of T3.1. This lack of fire means they are, in some ways, a better fit for 'T1.2 Dry forests and thickets', but, in turn, classifying these three caatinga groups as dry forest seems inappropriate as they are distinct in height, biomass, LAI etc. from the more forest-like arboreal and transition caatingas (Suppl material 1: appendix S3; Fig. 6). Whilst it is possible that there are misrepresentations in the remote-sensing data or that, recently, fire has been suppressed in these vegetation types, on balance, we think the difficulty in alignment with the GET is because it is missing an EFG. This conclusion agrees with IBGE (2021), which suggests a new EFG be added to the GET to accommodate the predominant vegetation formations in the caatinga.

Campos rupestres is a vegetation type in eastern Brazil that does align closely with EFG T3.1 (shrublands). This vegetation type has a different floristic and functional composition from the vegetation in our three caatinga groups. Most importantly, plant species in campos rupestres show many fire-resistance and fire-tolerance traits (Figueira et al. 2016) and there is floristic overlap with fire-prone savanna vegetation in the Cerrado. In contrast, many species, which are characteristic of the caatinga, do not feature fire adaptations, notably members of the Cactaceae family and other stem succulent species, such as the large Malvaceous tree *Cavallinesia umbellata* (Pennington et al. 2009; Oliveira-Filho et al. 2013; Queiroz et al. 2017). Furthermore, plants typical of caatinga vegetation often have thorns and are deciduous, with leaves having a low LMA, whereas campos rupestres plants are often evergreen, unarmed, with a high LMA (Queiroz et al. 2017; Mariano et al. 2021). Finally, the caatinga woody flora is more similar to that of tall dry forests scattered through the Cerrado and in the Chiquitania dry forest region of Bolivia, as opposed to campos rupestres (Silva de Miranda et al. 2018).

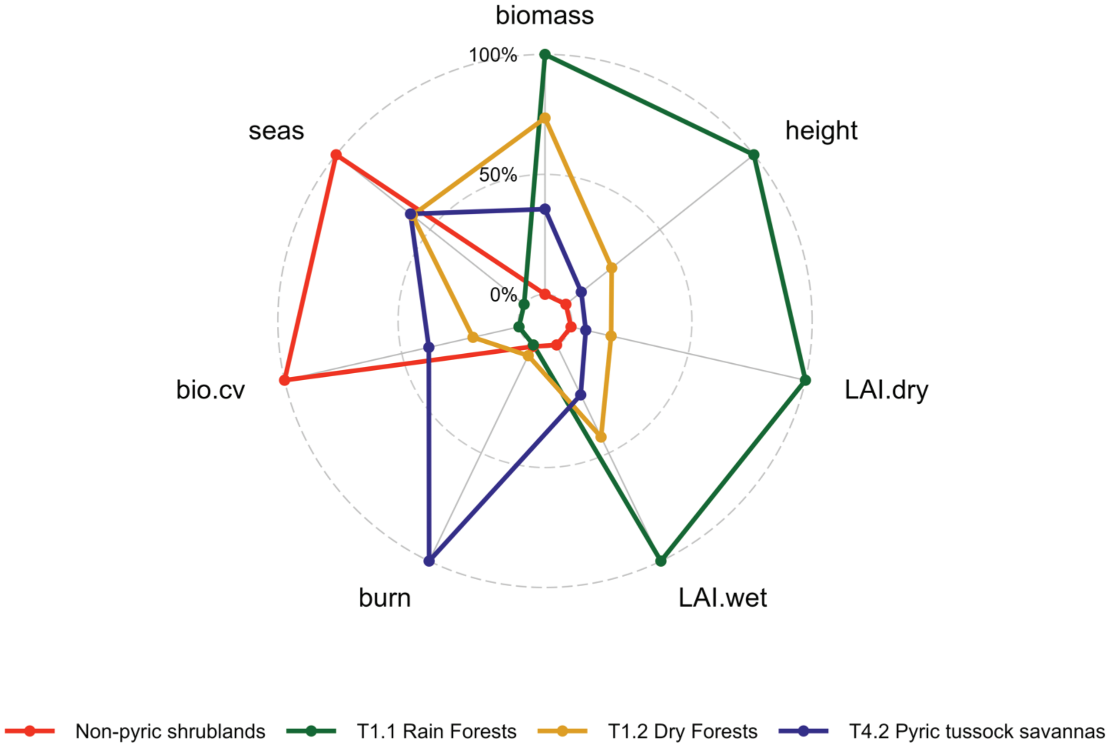

Figure 6. Scaled mean vegetation structural attributes for the IUCN Global Ecosystem Typology's Ecosystem Functional Groups in NE Brazil. Axes show the percentage for that structural vegetation group of the maximum data point for each vegetation attribute. The vegetation attributes were (clockwise from top) aboveground woody biomass, canopy height (height), mean Leaf area index (LAI) in the dry season, mean Leaf area index (LAI) in the wet season, burn count (burn), aboveground woody biomass heterogeneity (bio.cv) and NDVI seasonality (seas). Scrubby caatinga, heterogeneous caatinga and hyper-seasonal caatinga are placed in the Non-pyric shrublands group (blue). Arboreal caatinga and transition caatinga are placed in T1.2 Tropical-subtropical Dry Forests EFG (red) and the cerrado and rain forest structural groups in T4.2 Pyric tussock savannas (yellow) and T1.1 Tropical-subtropical Lowland rain forests (green) EFGs, respectively.

Given the discussion above, we suggest a new EFG should be incorporated into the GET to describe scrubby caatinga, hyper-seasonal caatinga and heterogeneous caatinga. However, it is beyond the scope of this paper to definitively determine the name of this new EFG and its location within the higher biome level of the Typology, given the regional nature of this study in comparison to the international scope of the IUCN Typology. Solely from our NE Brazil perspective, we suggest that the new EFG should be 'T1.5 Tropical-subtropical non-pyric thickets and shrublands'. This is because of the lack of fire and the short stature, but closed canopy of our caatinga groups (Fig. 2; Dong et al., in review). This new EFG would correspond to the 'T1 Tropical-subtropical forests' biome and not the current T3.1 shrublands, as the latter are defined as open ecosystems (Keith et al. 2020a).

The new EFG which we propose would occupy the drier end of the seasonally dry tropical forest spectrum (*sensu* Pennington et al. (2000)) and be functionally distinct from 'T3.1. Seasonally dry tropical pyric shrublands'. It would be analogous to the "succulent biome" *sensu* Schrire et al. (2005), describing a non-fire-adapted, succulent-rich, grass-poor biome, mapped by Ringelberg et al. (2020) using distribution of stem succulents. This splitting of dry forests from thickets affirms a division suggested by other authors (Schrire et al. 2005; Oliveira-Filho et al. 2013; Ringelberg et al. 2020; IBGE 2021) and provides an alternative to the broad definition of dry forest suggested by Murphy and Lugo (1986) and Pennington et al. (2000), which was adopted in the current definition of the T1.2 EFG. Our work suggests that splitting T1.2 into two EFGs is necessary to accurately describe vegetation in NE Brazil.

Is the IUCN GET mappable?

Overall, we find that the remote sensing-driven approach which we develop here can be used to identify structural groups in a complex region. Using an understanding of the drivers of EFG distribution including disturbance regime, soil and climatic factors, the structural groups from an unsupervised clustering can largely be grouped into EFGs as described by the IUCN GET. Therefore, this approach is suitable for fulfilling the fifth criterion of the GET (Keith et al. 2020b, 2022), that of providing spatially explicit maps, based upon an understanding of structure and vegetation ecology, using remote sensing. By going one level deeper into the thematic legend to map structural groups within EFGs, we have addressed the trade-offs between hierarchical levels in the IUCN GET – increased local applicability and realism at the local level, while methodologically and ecologically linking with the global EFGs. It is useful to understand highly complex regions with multiple interdigitated biomes and, as such, starting at the unsupervised structural level allows detailed information on the structural heterogeneity within EFGs to be obtained.

This work, therefore, adds to the growing body of work implementing the IUCN GET in a range of ecosystems (Murray et al. 2019; Karger et al. 2021). In addition, it extends the use of NE Brazil as a model system for biome delineation and mapping and demonstrates improvement in the ability of remote sensing since the efforts of Beuchle et al. (2015). Further work could aim to incorporate other strands of information into biome mapping in order to develop more ecologically meaningful maps. This would align with other parts of the 'Ecosystem Properties' method of defining biomes described by Keith et al. (2020b). For example, this could include the distribution of important functional traits such as CAM photosynthesis, C4 grasses, leaves with drip tips, deciduousness and spinescence (Conradi et al. 2020).

Conclusion

There is ongoing work within the IUCN assessing risks to ecosystems, through the Red List of Ecosystems, to help conservation prioritisation (Keith et al. 2015). Such work requires accurate ecosystem mapping to be effective (Keith et al. 2015). Consequently, our work operationalising the mapping of EFGs is crucial for IUCN-style approaches to conservation. By using an unsupervised classification of vegetation expression as the starting point, this work reduces reliance upon expert opinion, aligning with the arguments of Higgins et al. (2016) and Conradi et al. (2020) for more objective vegetation mapping and allows consideration of the defining physiognomic features of EFGs and the ecological differences between them. The quantitative and hierarchical approach to describing groups, EFGs and biomes which is implemented here might provide utility for conservation at different scales. For example, the GET's EFGs can be related to Level 2 in the MapBiomas classification in Brazil and structural groups as a finer level within this. MapBiomas is increasingly utilised for legislative processes in Brazil (Vidal and Allen 2023). As such, the ability to spatially relate both EFGs and vegetation structural groups to legislative processes through a hierarchical classification increases the applicability of academic biome mapping. Furthermore, considering all biomes together in a spatially explicit framework can enhance conservation efforts (Lewis et al. 2023) and will be particularly important in areas where biomes are highly interdigitated, such as NE Brazil.

Acknowledgements

This research has been supported by the Natural Environment Research Council (NERC) through a SENSE CDT studentship (grant no. NE/T00939X/1) and SECO: Resolving the Current and Future Carbon Dynamics of the Dry Tropics (grant no. NE/T01279X/1). DR was supported by FAPESP (grants #2017/17380-1 and #2022/02323-0). Thanks to Samuel Bowers for his help in writing Google Earth Engine code used for developing the seasonality and burn count data layers. This publication has been prepared using European Union's Copernicus Land Monitoring Service information; https://land.copernicus.eu/en/products/vegetation/leaf-area-index-v2-0-1km. We also thank the editor and three reviewers for their valuable comments which resulted in a much improved analysis and manuscript.

Author contributions

Lucy Wells: Conceptualisation, Methodology, Analysis, Interpretation, Writing – original draft, review and editing, Visualisation. Kyle Dexter: Conceptualisation, Methodology, Interpretation, Writing – review and editing. Toby Pennington: Conceptualisation, Interpretation, Writing – review and editing. Ítalo Coutinho: Fieldwork, Writing – review. Desiree Ramos: Fieldwork, Writing – review. Oliver Phillips: Conceptualisation, Writing – review. Tim Baker: Conceptualisation. Casey Ryan: Conceptualisation, Methodology, Analysis, Interpretation, Writing – review and editing.

Data accessibility statement

The data is shared via the University of Edinburgh Open Access DataShare platform, with the DOI of: https://doi.org/10.7488/ds/7879.

References

Andrade-Lima D (1981) The caatingas dominium. Brazilian Journal of Botany 4: 149–153.

Antongiovanni M, Venticinque EM, Matsumoto M, Fonseca CR (2020) Chronic anthropogenic disturbance on Caatinga dry forest fragments. Journal of Applied Ecology 57: 2064–2074. https://doi.org/10.1111/1365-2664.13686

Arruda DM, Schaefer CEGR, Corrêa GR, Rodrigues PMS, Duque-Brasil R, Ferreira JRWG, Oliveira-Filho AT (2015) Landforms and soil attributes determine the vegetation structure in the Brazilian semiarid. Folia Geobotanica 50: 175–184. https://doi.org/10.1007/s12224-015-9221-0

Arruda DM, Fernandes-Filho EI, Solar RRC, Schaefer CEGR (2017) Combining climatic and soil properties better predicts covers of Brazilian biomes. The Science of Nature 104: 32. https://doi.org/10.1007/s00114-017-1456-6

Barros JS, Castro AAJF (2006) Compartimentação geoambiental no complexo de campo maior, PI: uma área de tensão ecológica. Interações (Campo Grande) 8. https://doi.org/10.1590/S1518-70122006000200013

Beuchle R, Grecchi RC, Shimabukuro YE, Seliger R, Eva HD, Sano E, Achard F (2015) Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. Applied Geography 58: 116–127. https://doi.org/10.1016/j.apgeog.2015.01.017

Bond WJ (2005) Large parts of the world are brown or black: A different view on the ‘Green World’ hypothesis. Journal of Vegetation Science 16: 261–266. https://doi.org/10.1111/j.1654-1103.2005.tb02364.x

Bond WJ, Woodward FI, Midgley GF (2005) The global distribution of ecosystems in a world without fire. New Phytologist 165: 525–538. https://doi.org/10.1111/j.1469-8137.2004.01252.x

Breiman L (2001) Random forests. Machine Learning 45: 5–32. https://doi.org/10.1023/A:1010933404324

Breiman L, Cutler A, Liaw A, Wiener M (2024) Breiman and Cutlers Random Forests for Classification and Regression. R package version 4.7-1.1.

Bueno ML, Dexter KG, Pennington RT, Pontara V, Neves DM, Ratter JA, Oliveira-Filho ATde (2018) The environmental triangle of the Cerrado Domain: Ecological factors driving shifts in tree species composition between forests and savannas. Journal of Ecology 106: 2109–2120. https://doi.org/10.1111/1365-2745.12969

Cai W, Chen S, Zhang D (2007) Fast and robust fuzzy *c*-means clustering algorithms incorporating local information for image segmentation. Pattern Recognition 40: 825–838. https://doi.org/10.1016/j.patcog.2006.07.011

Campo-Bescós MA, Muñoz-Carpena R, Kaplan DA, Southworth J, Zhu L, Waylen PR (2013) Beyond precipitation: Physiographic gradients dictate the relative importance of environmental drivers on savanna vegetation. PLOS ONE 8. https://doi.org/10.1371/journal.pone.0072348

Conceição AA, Rapini A, do Carmo FF, Brito JC, Silva GA, Neves SPS, Jacobi C (2016) Rupestrian Grassland Vegetation, Diversity, and Origin. In: Fernandes G (Ed.) Ecology and Conservation of Mountaintop grasslands in Brazil, 105–127. Springer, Cham. https://doi.org/10.1007/978-3-319-29808-5_6

Cardoso D, Moonlight PW, Ramos G, Oatley G, Dudley C, Gagnon E, Queiroz LPd, Pennington RT, Särkinen TE (2021) Defining biologically meaningful biomes through floristic, functional, and phylogenetic data. Frontiers in Ecology and Evolution 9: 723558. https://doi.org/10.3389/fevo.2021.723558

Castanho AD, Coe M, Andrade EM, Walker W, Baccini A, Campos DA, Farina M (2020) A close look at above ground biomass of a large and heterogeneous Seasonally Dry Tropical Forest –Caatinga in North East of Brazil. Anais da Academia Brasileira de Ciências 92: e20190282. https://doi.org/10.1590/0001-3765202020190282

Castro Oliveira Gde, Francelino MR, Arruda DM, Fernandes-Filho EI, Schaefer CEGR (2019) Climate and soils at the Brazilian semiarid and the forest-Caatinga problem: new insights and implications for conservation. Environmental Research Letters 14: 104007. https://doi.org/10.1088/1748-9326/ab3d7b

Coelho AJP, Magnago LFS, Matos FAR, Mota NM, Diniz ÉS, Meira-Neto JAA (2020) Effects of anthropogenic disturbances on biodiversity and biomass stock of Cerrado, the Brazilian savanna. Biodiversity and Conservation 29: 3151–3168. https://doi.org/10.1007/s10531-020-02013-6

Conceição AA, Pirani JR, Meirelles ST (2007) Floristics, structure and soil of insular vegetation in four quartzite-sandstone outcrops of “Chapada Diamantina”, Northeast Brazil. Revista Brasileira de Botânica 30: 641–656. https://doi.org/10.1590/S0100-84042007000400009

Conradi T, Slingsby JA, Midgley GF, Nottebrock H, Schweiger AH, Higgins SI (2020) An operational definition of the biome for global change research. New Phytologist 227: 1294–1306. https://doi.org/10.1111/nph.16580

Cramer MD, Wootton LM, Mazijk R van Verboom GA (2019) New regionally modelled soil layers improve prediction of vegetation type relative to that based on global soil models. Diversity and Distributions 25: 1736–1750. https://doi.org/10.1111/ddi.12973

Dandabathula G, Salunkhe S, Bera A, Ghosh K, Hari R, Biradar P, Chirala K, Gaur M (2022) Validation of SoilGrids 2.0 in an Arid Region of

India using In Situ Measurements. European Journal of Environment and Earth Sciences 3: 49–58. https://doi.org/10.24018/ejgeo.2022.3.6.356

Dantas VdL, Hirota M, Oliveira RS, Pausas JG (2016) Disturbance maintains alternative biome states. Ecology Letters 19: 12–19. https://doi.org/10.1111/ele.12537

De Lima CCU, Nolasco MC (2015) Chapada Diamantina: A Remarkable Landscape Dominated by Mountains and Plateaus. Landscapes and Landforms of Brazil. In: Vieira BC, Salgado AAR, Santos LJC (Ed.) Series Title: World Geomorphological Landscapes. Dordrecht: Springer Netherlands, 211–220. [ISBN: 978-94-017-8022-3 978-94-017-8023-0] https://doi.org/10.1007/978-94-017-8023-0_19

De Vasconcelos SF, De Araújo FS, Lopes AV (2010) Phenology and dispersal modes of wood species in the Carrasco, a tropical deciduous shrubland in the Brazilian semiarid. Biodiversity and Conservation 19: 2263–2289. https://doi.org/10.1007/s10531-010-9841-x

Dexter KG, Pennington RT, Oliveira-Filho AT, Bueno ML, Silva de Miranda PL, Neves DM (2018) Inserting tropical dry forests into the discussion on biome transitions in the tropics. Frontiers in Ecology and Evolution 6: 104. https://doi.org/10.3389/fevo.2018.00104

Dias LCC, Moschini LE, Trevisan DP (2017) A Influência das Atividades Antrópicas na Paisagem da Área de Proteção Ambiental Estadual do Rio Pandeiros, MG - Brasil. Fronteiras: Journal of Social, Technological and Environmental Science 6: 85. https://doi.org/10.21664/2238-8869.2017v6i2.p85-105

Didan K (2021) MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V061 NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD13A1.061

Dinerstein E, Olson D, Joshi A, Vynne C, Burgess NB, Wikramanayake E, Hahn N, Palminteri S, Hedao P, Noss R, Hansen M, Locke H, Ellis EC, Jones B, Barber CV, Hayes R, Kormos C, Martin V, Crist E, Sechrest W, Price L, Baillie JE, Weeden D, Suckling K, Davis C, Sizer N, Moore R, Thau D, Birch T, Potapov P, Turubanova S, Tyukavina A, de Souza N, Pintea L, Brito JC, Llewellyn OA, Miller AG, Patzelt A, Ghazanfar SA, Timberlake J, Klöser H, Shennan-Farpón Y, Kindt R, Barnekow Lillesø J, van Breugel P, Graudal L, Voge M, Al-Shammari KF, Saleem M (2017) An ecoregion-based approach to protecting half the terrestrial realm. BioScience 67: 534–545. https://doi.org/10.1093/biosci/bix014

DRYFLOR (2016) Plant diversity patterns in neotropical dry forests and their conservation implications. Science 353: 1383–1387. https://doi.org/10.1126/science.aaf5080

Feilhauer H, Zlinszky A, Kania A, Foody GM, Doktor D, Lausch A, Schmidtlein S (2021) Let your maps be fuzzy! — Class probabilities and floristic gradients as alternatives to crisp mapping for remote sensing of vegetation. Remote Sensing in Ecology and Conservation 7: 292–305. https://doi.org/10.1002/rse2.188

Fernandes MF, Cardoso D, De Queiroz LP (2020) An updated plant checklist of the Brazilian Caatinga seasonally dry forests and woodlands reveals high species richness and endemism. Journal of Arid Environments 174: 104079. https://doi.org/10.1016/j.jaridenv.2019.104079

Fernandes MF, Cardoso D, Pennington RT, Queiroz LP (2022) The origins and historical assembly of the Brazilian Caatinga seasonally dry tropical forests. Frontiers in Ecology and Evolution 10: 723286. https://doi.org/10.3389/fevo.2022.723286

Fick S, Hijmans RJ (2017) WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37: 4302–4315. https://doi.org/10.1002/joc.5086

Figueira JEG, Ribeiro KT, Ribeiro MC, Jacobi CM, França H, Neves AC, Conceição AA, Mourão FA, Souza JM, Miranda CAK (2016) Fire in Rupestrian Grasslands: Plant Response and Management. In: Fernandes G (Ed.) Ecology and Conservation of Mountaintop grasslands in Brazil. Springer, Cham, 415–448. https://doi.org/10.1007/978-3-319-29808-5_18

Funch RR, Harley RM, Funch LS (2009) Mapping and evaluation of the state of conservation of the vegetation in and surrounding the Chapada Diamantina National Park, NE Brazil. Biota Neotropica 9: 21–30. https://doi.org/10.1590/S1676-06032009000200001

Gelb J (2021) Package geocmeans LAEQ. https://github.com/JeremyGelb/geocmeans

Gelb J (2023) geocmeans: An R package for spatial fuzzy c-means. Journal of Open Source Software 8: 5259. https://doi.org/10.21105/joss.05259

Gelb J, Apparicio P (2021) Apport de la classification floue c-means spatiale en géographie : essai de taxinomie socio-résidentielle et environnementale à Lyon. Cybergeo: European Journal of Geography. Publisher: CNRS-UMR Géographie-cités 8504. https://doi.org/10.4000/cybergeo.36414

Giglio L, Justice C, Boschetti L, Roy D (2021) MODIS/Terra+Aqua Burned Area Monthly L3 Global 500m SIN Grid V061 NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MCD64A1.061

Greve M, Lykke AM, Blach-Overgaard A, Svenning JC (2011) Environmental and anthropogenic determinants of vegetation distribution across Africa. Global Ecology and Biogeography 20: 661–674. https://doi.org/10.1111/j.1466-8238.2011.00666.x

Higgins SI, Buitenwerf R, Moncrieff GR (2016) Defining functional biomes and monitoring their change globally. Global Change Biology 22: 3583–3593. https://doi.org/10.1111/gcb.13367

Higgins SI, Conradi T, Kruger LM, O'Hara RB, Slingsby JA (2023) Limited climatic space for alternative ecosystem states in Africa. Science 380: 1038–1042. https://doi.org/10.1126/science.add5190

Hijmans RJ (2022) _raster: Geographic Data Analysis and Modeling_. R package version 3.5-29. https://CRAN.R-project.org/package=raster

Hijmans R (2024) terra: Spatial Data Analysis. R package version 1.7-82. https://rspatial.github.io/terra/

Holdridge LR (1967) Life Zone Ecology. Tropical Science Center, San Jose, Costa Rica.

IBGE (1993) Mapa de vegetação do Brasil. Rio de Janeiro.

IBGE (2008) Mapa da área de aplicação da Lei n° 11,428 de 2006, Decreto n° 6.660, de 21 de novembro de 2008. Brasília, Brazil. https://geoftp.ibge.gov.br/informacoes_ambientais/estudos_ambientais/biomas/mapas/lei11428_mata_atlantica.pdf

IBGE (2021) Classificação internacional dos ecossistemas no Brasil: comparação entre as classes de vegetação e de cobertura e uso da terra do IBGE e os grupos funcionais de ecossistemas da IUCN / IBGE, Coordenação de Recursos Naturais e Estudos Ambientais. Rio de Janeiro, 73 pp. [Textos para discussão. Diretoria de Geociências, ISSN 1806-4531; n. 5]

Karger DN, Kessler M, Lehnert M, Jetz W (2021) Limited protection and ongoing loss of tropical cloud forest biodiversity and ecosystems worldwide. Nature Ecology & Evolution 5: 854–862. https://doi.org/10.1038/s41559-021-01450-y

Keith DA, Russell-Smith J (2020) T3.1 Seasonally dry tropical shrublands. The IUCN Global Typology 2.0: Descriptive profiles for biomes and ecosystem functional groups. Gland, Switzerland: IUCN, 52. https://doi.org/10.2305/IUCN.CH.2020.13.en

Keith DA, Rodríguez JP, Brooks TM, Burgman MA, Barrow EG, Bland L, Comer PJ, Franklin J, Link J, McCarthy MA, Miller RA, Murray NJ, Nel J, Nicholson E, Oliveira-Miranda MA, Regan TJ, Rodríguez-Clark KM, Rouget M, Spalding MD (2015) The IUCN Red List of Ecosystems: Motivations, Challenges, and Applications. Conservation Letters 8: 214–226. https://doi.org/10.1111/conl.12167

Keith DA, Young KR, Corlett RT (2020a) T1.1 Tropical subtropical lowland rainforests. The IUCN Global Typology 2.0: Descriptive profiles for biomes and ecosystem functional groups. Gland, Switzerland: IUCN, 39. https://doi.org/10.2305/IUCN.CH.2020.13.en

Keith DA, Ferrer-Paris JR, Nicholson E, Kingsford RT (2020b) IUCN Global Ecosystem Typology 2.0: Descriptive profiles for biomes and ecosystem functional groups. Tech. rep. Gland, Switzerland: IUCN, 192. https://doi.org/10.2305/IUCN.CH.2020.13.en

Keith DA, Ferrer-Paris JR, Nicholson E, Bishop MJ, Polidoro BA, Ramirez-Llodra E, Tozer MG, Nel JL, Mac Nally R, Gregr EJ, Watermeyer KE, Essl F, Faber-Langendoen D, Franklin J, Lehmann CER, Etter A, Roux DJ, Stark JS, Rowland JA, Brummitt NA, Fernandez-Arcaya UC, Suthers IM, Wiser SK, Donohue I, Jackson LJ, Pennington RT, Iliffe TM, Gerovasileiou V, Giller P, Robson BJ, Pettorelli N, Andrade A, Lindgaard A, Tahvanainen T, Terauds A, Chadwick MA, Murray NJ, Moat J, Pliscoff P, Kingsford RT (2022) A function-based typology for Earth's ecosystems. Nature 610: 513–518. https://doi.org/10.1038/s41586-022-05318-4

Klink CA, Machado RB (2005) Conservation of the Brazilian Cerrado. Conservation Biology 19: 707–713. https://doi.org/10.1111/j.1523-1739.2005.00702.x

Lehmann CER, Etter A, Young KR, Keith DA (2020) T4.2 Pyric Tussock Savannas. The IUCN Global Ecosystem Typology 2.0: Descriptive profiles for biomes and ecosystem functional groups. Gland, Switzerland: IUCN, 57.

Lewis K, Barros FDV, Moonlight PW, Hill TC, Oliveira RS, Schmidt IB, Sampaio AB, Pennington RT, Rowland L (2023) Identifying hotspots for ecosystem restoration across heterogeneous tropical savanna-dominated regions. Philosophical Transactions of the Royal Society B: Biological Sciences 378: 20210075. https://doi.org/10.1098/rstb.2021.0075

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2022) cluster: Cluster Analysis Basics and Extensions. R package version 2.1.3.

Maia VA, De Souza CR, De Aguiar-Campos N, Alencar Fagundes AC, Santos ABM, Pires de Paula G, Ferreira Santos P, Bento Silva W, de Oliveira Menino GC, dos Santos RM (2020) Interactions between climate and soil shape tree community assembly and above-ground woody biomass of tropical dry forests. Forest Ecology and Management 474: 118348. https://doi.org/10.1016/j.foreco.2020.118348

MapBiomas Project (2022) MapBiomas Project – Collection 6 of the Annual Land Use and Land Cover Maps of Brazil.

Mariano E, Gomes TF, Lins SRM, Abdalla-Filho AL, Soltangheisi A, Araújo Almeida RF, Augusto FG, Canisares LP, Chaves SSF, Costa CFG, Diniz-Reis TR, Galera LA, Martinez MG, Morais MC, Perez EB, Reis LC, Simon CP, Mardegan SF, Domingues TF, Miatto RC, Oliveira RS, Reis CRG, Nardoto GB, Kattge J, Martinelli LA (2021) LT-Brazil: A database of leaf traits across biomes and vegetation types in Brazil. Global Ecology and Biogeography 30(11): 2136–2146. https://doi.org/10.1111/geb.13381

Meijer JR, Huijbregts MAJ, Schotten KCGJ, Schipper AM (2018) Global patterns of current and future road infrastructure. Environmental Research Letters 13: 064006. https://doi.org/10.1088/1748-9326/aabd42

Melo FP (2017) The Socio-Ecology of the Caatinga: Understanding How Natural Resource Use Shapes an Ecosystem. In: Silva SM, Leal IR, Tabarelli M (Ed.) Caatinga The Largest Tropical Dry Forest Region in South America. Springer: Switzerland, 369–383. https://doi.org/10.1007/978-3-319-68339-3_14

Miller T, Blackwood CB, Case AL (2024) Assessing the utility of SoilGrids250 for biogeographic inference of plant populations. Ecology and Evolution 14: e10986. https://doi.org/10.1002/ece3.10986

Moncrieff GR, Lehmann CER, Schnitzler J, Gambiza J, Hiernaux P, Ryan CM, Shackleton CM, Williams RJ, Higgins SI (2014) Contrasting architecture of key African and Australian savanna tree taxa drives intercontinental structural divergence: Savanna tree architecture. Global Ecology and Biogeography 23: 1235–1244. https://doi.org/10.1111/geb.12205

Moncrieff GR, Bond WJ, Higgins SI (2016) Revising the biome concept for understanding and predicting global change impacts. Journal of Biogeography 43: 863–873. https://doi.org/10.1111/jbi.12701

Moonlight PW, Silva De Miranda PL, Cardoso D, Dexter KG, Oliveira-Filho AT, Pennington RT, Ramos G, Särkinen TE (2020) The strengths and weaknesses of species distribution models in biome delimitation. Global Ecology and Biogeography 29: 1770–1784. https://doi.org/10.1111/geb.13149

Morales N, Assine ML (2015) Chapada do Araripe: A Highland Oasis Incrusted into the Semiarid region of North-eastern Brazil. Landscapes and Landforms of Brazil. Springer, Dordrecht, 231–242. [ISBN: 2213-2090] https://doi.org/10.1007/978-94-017-8023-0_21

Moro MF, Macedo MB, Moura-Fé MMD, Castro ASF, Costa RCD (2015) Vegetação, unidades fitoecológicas e diversidade paisagística do estado do Ceará. Rodriguésia 66: 717–743. https://doi.org/10.1590/2175-7860201566305

Moro MF, Lughadha NE, De Araújo FS, Martins FR (2016) A phytogeographical metaanalysis of the semiarid Caatinga domain in Brazil. The Botanical Review 82: 91–148. https://doi.org/10.1007/s12229-016-9164-z

Mota NM, Gastauer M, Carrión JF, Meira-Neto JAA (2022) Roads as conduits of functional and phylogenetic degradation in Caatinga. Tropical Ecology 64: 72–85. https://doi.org/10.1007/s42965-022-00245-x

Mucina L (2019) Biome: evolution of a crucial ecological and biogeographical concept. New Phytologist 222: 97–114. https://doi.org/10.1111/nph.15609

Murphy PG, Lugo AE (1986) Ecology of tropical dry forest. Annual Review of Ecology and Systematics 17: 67–88. https://doi.org/10.1146/annurev.es.17.110186.000435

Murphy BP, Bowman DM (2012) What controls the distribution of tropical forest and savanna?: Tropical forest and savanna distribution. Ecology Letters 15: 748–758. https://doi.org/10.1111/j.1461-0248.2012.01771.x

Murray NJ, Phinn SR, DeWitt M, Ferrari R, Johnston R, Lyons MB, Clinton N, Thau D, Fuller RA (2019) The global distribution and trajectory of tidal flats. Nature 565: 222–225. https://doi.org/10.1038/s41586-018-0805-8

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB da Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403: 853–858. https://doi.org/10.1038/35002501

Neves DM, Dexter KG, Pennington RT, Valente ASM, Bueno ML, Eisenlohr PV, Fontes MAL, Miranda PLS, Moreira SN, Rezende VL, Saiter FZ, Oliveira-Filho AT (2017) Dissecting a biodiversity hotspot: the importance of environmentally marginal habitats in the Atlantic Forest Domain of South America. Diversity and Distributions 23: 898–909. https://doi.org/10.1111/ddi.12581

Nicholson E, Watermeyer KE, Rowland JA, Sato CF, Stevenson SL, Andrade A, Brooks TM, Burgess ND, Cheng S, Grantham HS, Hill AL, Keith DA, Maron M, Metzke D, Murray NJ, Nelson CR, Obura D, Plumptre A, Skowno AL, Watson JEM (2021) Scientific foundations for an ecosystem goal, milestones and indicators for the post-2020 global biodiversity framework. Nature Ecology and Evolution 5: 1338–1349. https://doi.org/10.1038/s41559-021-01538-5

Nogueira FC, Pagotto MA, Aragão JRV, Pagotto MA, Aragão JRV, Roig FA, Ribeiro AS, Lisi CS (2019) The hydrological performance of *Prosopis juliflora* (Sw.) growth as an invasive alien tree species in the semiarid tropics of north-eastern Brazil. Biological Invasions 21: 2561–2575. https://doi.org/10.1007/s10530-019-01994-y

Ocón JP, Ibanez T, Franklin J, Pau S, Keppel G, Rivas-Torres G, Shin ME, Gillespie TW (2021) Global tropical dry forest extent and cover: A comparative study of bioclimatic definitions using two climatic data sets. PLOS ONE 16. https://doi.org/10.1371/journal.pone.0252063

Oliveira GdC, Arruda DM, Fernandes Filho EI, Veloso GV, Francelino MR, Schaefer CEGR (2021) Soil predictors are crucial for modelling vegetation distribution and its responses to climate change. Science of The Total Environment 780: 146680. https://doi.org/10.1016/j.scitotenv.2021.146680

Oliveira-Filho AT, Fontes MAL (2000) Patterns of Floristic Differentiation among Atlantic Forests in Southeastern Brazil. Biotropica 32: 793–810. https://doi.org/10.1111/j.1744-7429.2000.tb00619.x

Oliveira-Filho AT, Jarenkow J, Rodal MJN (2006) Floristic relationships of seasonally dry forests of eastern South America based on tree species distribution patterns. Neotropical savannas and dry forests: Plant diversity, biogeography and conservation. Taylor and Francis, Boca Raton, 159–192. https://doi.org/10.1201/9781420004496-7

Oliveira-Filho A, Cardoso D, Schrire B, Lewis G, Pennington R, Brummer T, Rotella J, Lavin M (2013) Stability structures tropical woody plant diversity more than seasonality: Insights into the ecology of high legume-succulent-plant biodiversity. South African Journal of Botany 89: 42–57. https://doi.org/10.1016/j.sajb.2013.06.010

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D-Amico JA, Itoua I, Strand HE, Morrison JC, Loucks CJ, Allnutt TF, Ricketts TH, Kura Y, Lamoreux JF, Wettengel WW, Hedao P, Kassem KR (2001) Terrestrial ecoregions of the world: A new map of life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51: 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Parr CL, Lehmann CE, Bond WJ, Hoffmann WA, Andersen AN (2014) Tropical grassy biomes: misunderstood, neglected, and under threat. Trends in Ecology & Evolution 29: 205–213. https://doi.org/10.1016/j.tree.2014.02.004

Pausas JG, Ribeiro E (2017) Fire and plant diversity at the global scale. Global Ecology and Biogeography 26: 889–897. https://doi.org/10.1111/geb.12596

Pennington RT, Prado DE, Pendry CA (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. Journal of Biogeography 27: 261–273. https://doi.org/10.1046/j.1365-2699.2000.00397.x

Pennington RT, Lavin M, Oliveira-Filho A (2009) Woody plant diversity, evolution, and ecology in the tropics: Perspectives from seasonally dry tropical forests. Annual Review of Ecology, Evolution, and Systematics 40: 437–457. https://doi.org/10.1146/annurev.ecolsys.110308.120327

Pennington RT, Lehmann CE, Rowland LM (2018) Tropical savannas and dry forests. Current Biology 28: R541–R545. https://doi.org/10.1016/j.cub.2018.03.014

Pennington RT, Franklin J, Brummitt NA, Etter A, Young KR, Corlett RT, Keith DA (2020) T1.2 Tropical subtropical dry forests and thickets. The IUCN Global Typology 2.0: Descriptive profiles for biomes and ecosystem functional groups. IUCN, Gland, Switzerland, 40.

Poggio L, De Sousa LM, Batjes NH, Heuvelink GBM, Kempen B, Ribeiro E, Rossiter D (2021) SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7: 217–240. https://doi.org/10.5194/soil-7-217-2021

Pörtner HO, Roberts DC, Tignor M, Poloczanska K, Mintenbeck K, Alegría A, Craig M, Langsdorf S, Löschke S, Möller V, Okem A, Rama B (2022) Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Tech. rep. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Potapov P, Li X, Hernandez-Serna A, Tyukavina A, Hansen MC, Kommareddy A, Pickens A, Turubanova S, Tang H, Edibaldo Silva C, Armston J, Dubayah R, Blair JB, Hofton M (2021) Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment 253: 112165. https://doi.org/10.1016/j.rse.2020.112165

Queiroz LP (2006) The Brazilian Caatinga: phytogeographical patterns inferred from distribution data of the Leguminosae. In: Pennington RT, Ratter JA (Eds) Neotropical Savannas and Seasonally Dry Forests: Plant Diversity, Biogeography and Conservation. CRC Press, London, UK, 121–158. https://doi.org/10.1201/9781420004496-6

Queiroz LP, Cardoso D, Fernandes MF, Moro MF (2017) Diversity and evolution of flowering plants of the Caatinga Domain. Caatinga: the Largest Tropical Dry Forest Region in South America. Cham: Springer International Publishing, 23–63. https://doi.org/10.1007/978-3-319-68339-3_2

Ramos MB, Diniz FC, Almeida HAd, Almeida GRd, Pinto AS, Meave JA, Lopes SdF (2020) The role of edaphic factors on plant species richness and diversity along altitudinal gradients in the Brazilian semi-arid region. Journal of Tropical Ecology 36: 199–212. https://doi.org/10.1017/S0266467420000115

Ratter J (1997) The Brazilian Cerrado Vegetation and Threats to its Biodiversity. Annals of Botany 80: 223–230. https://doi.org/10.1006/anbo.1997.0469

Ribeiro EMS, Arroyo-Rodríguez V, Santos BA, Tabarelli M, Leal IR (2015) Chronic anthropogenic disturbance drives the biological impoverishment of the Brazilian Caatinga vegetation. Journal of Applied Ecology 52: 611–620. https://doi.org/10.1111/1365-2664.12420

Ribeiro EMS, Santos BA, Arroyo-Rodríguez V, Tabarelli M, Souza G, Leal IR (2016) Phylogenetic impoverishment of plant communities following chronic human disturbances in the Brazilian Caatinga. Ecology 97: 1583–1592. https://doi.org/10.1890/15-1122.1

Ringelberg JJ, Zimmermann NE, Weeks A, Lavin M, Hughes CE (2020) Biomes as evolutionary arenas: Convergence and conservatism in the transcontinental succulent biome. Global Ecology and Biogeography 29: 1100–1113. https://doi.org/10.1111/geb.13089

Rito KF, Arroyo-Rodríguez V, Queiroz RT, Leal IR, Tabarelli M (2017) Precipitation mediates the effect of human disturbance on the Brazilian Caatinga vegetation. Journal of Ecology 105: 828–838. https://doi.org/10.1111/1365-2745.12712

Rivero-Villar A, De La Peña-Domene M, Rodríguez-Tapia G, Giardina CP, Campo J (2022) A Pantropical Overview of Soils across Tropical Dry Forest Ecoregions. Sustainability 14: 6803. https://doi.org/10.3390/su14116803

Rocha PLBD, Queiroz LPD, Pirani JR (2004) Plant species and habitat structure in a sand dune field in the Brazilian Caatinga: a homogeneous habitat harbouring an endemic biota. Revista Brasileira de Botânica 27: 739–755. https://doi.org/10.1590/S0100-84042004000400013

Santoro M, Cartus O, Mermoz S, Bouvet A, Le Toan T, Carvalhais N, Rozendaal D, Herold M, Avitabile V, Quegan S, Carreiras J, Rauste Y, Balzter H, Schmullius C, Seifert FM (2018) GlobBiomass global above-ground biomass and growing stock volume datasets. http://globbiomass.org/products/global-mapping

Santos RM, Oliveira-Filho AT, Eisenlohr PV, Queiroz LP, Cardoso DBOS, Rodal MJN (2012) Identity and relationships of the Arboreal Caatinga among other floristic units of seasonally dry tropical forests (SDTFs) of north-eastern and Central Brazil: Biogeography of Brazilian Seasonally Dry Tropical Forests. Ecology and Evolution 2: 409–428. https://doi.org/10.1002/ece3.91

Santos-Filho FS, Bezerra de Almeida Jr E, dos Reis Santos Soares CJ, Zickel CJ (2015) Flora and woody vegetation structure in an insular area of restinga in Brazil. International Journal of Ecology and Environmental Sciences 41: 147–160.

Scarano FR (2009) Plant communities at the periphery of the Atlantic rain forest: Rare-species bias and its risks for conservation. Biological Conservation 142: 1201–1208. https://doi.org/10.1016/j.biocon.2009.02.027

Schimper AFW (1903) Plant geography upon a physiological basis. Clarendon Press, Oxford, UK. https://doi.org/10.5962/bhl.title.122577

Schrire M, Lavin M, Lewis GP (2005) Global distribution patterns of the Leguminosae: Insights from recent phylogenies. Biologiske Skrifter 55: 375–422.

Schulz C, Koch R, Cierjacks A, Kleinschmit B (2017) Land change and loss of landscape diversity at the Caatinga phytogeographical domain – Analysis of pattern-process relationships with MODIS land cover products (2001–2012). Journal of Arid Environments 136: 54–74. https://doi.org/10.1016/j.jaridenv.2016.10.004

Silva AC, Souza AF (2022) Spatial structure of the Caatinga woody flora: abundance patterns have environmental, Pleistocene, and indigenous drivers. Anais da Academia Brasileira de Ciências 94: e20211019. https://doi.org/10.1590/0001-3765202220211019

Silva JMCd, Leal IR, Tabarelli M [Eds] (2017) Caatinga. Cham: Springer International Publishing. [ISBN: 978-3-319-68338-6 978-3-319-68339-3]

Silva de Miranda PL, Oliveira-Filho AT, Pennington RT, Neves DM, Baker TR, Dexter KG (2018) Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America. Global Ecology and Biogeography 27: 899–912. https://doi.org/10.1111/geb.12749

Simon MF, Pennington T (2012) Evidence for adaptation to fire regimes in the tropical savannas of the Brazilian Cerrado. International Journal of Plant Sciences 173: 711–723. https://doi.org/10.1086/665973

Souza CM, Z Shimbo J, Rosa MR, Parente LL, Alencar AA, Rudorff BFT, Hasenack H, Matsumoto M, Ferreira LG, Souza-Filho PWM, de Oliveira SW, Rocha WF, Fonseca AV, Marques CB, Diniz CG, Costa D, Monteiro D, Rosa ER, Vélez-Marton E, Wever EJ, Lenti FEB, Paternost FF, Pareyn FGC, Siqueira JV, Viera JL, Ferreira Neto LC, Saraiva MM, Sales MH, Salgado MPG, Vasconcelos R, Galano S, Mesquita VV, Azevedo T (2020) Reconstructing three decades of land use and land cover changes in Brazilian biomes with landsat archive and earth engine. Remote Sensing 12: 2735. https://doi.org/10.3390/rs12172735

Souza Jr C, Azevedo T (2017) MapBiomas General “Handbook”.

Team QD (2021) QGIS Software.

Terra MDCNS, Santos RMD, Prado Júnior JAD, De Mello JM, Soares Scolforo JR, Leite Fontes MA, Schiavini I, Aparecida dos Reis A, Bueno IT, Fernando Silva Magnago L, ter Steege H (2018) Water availability drives gradients of tree diversity, structure and functional traits in the Atlantic–Cerrado–Caatinga transition, Brazil. Journal of Plant Ecology 11: 803–814. https://doi.org/10.1093/jpe/rty017

Torello-Raventos M, Feldpausch TR, Veenendaal E, Schrodt F, Saiz G, Domingues TF, Djagbletey G, Ford A, Kemp J, Marimon BS, Marimon Junior BH, Lenza E, Taedoumg H, Villarroel D, Schwarz M, Quesada CA, Yoko Ishida F, Nardoto GB, Affum-Baffoe K, Arroyo L, Bowman DMJS, Compaore H, Davies K, Diallo A, Fyllas NM, Gilpin M, Hien F, Johnson M, Killeen TJ, Metcalfe D, Miranda HS, Steininger M, Thomson J, Sykora K, Mougin E, Hiernaux P, Bird MI, Grace J, Lewis SL, Phillips OL, Lloyd J (2013) On the delineation of tropical vegetation types with an emphasis on forest/savanna transitions. Plant Ecology & Diversity 6: 101–137. https://doi.org/10.1080/17550874.2012.762812

Tsai DM, Lin CC (2011) Fuzzy C-means based clustering for linearly and nonlinearly separable data. Pattern Recognition 44: 1750–1760. https://doi.org/10.1016/j.patcog.2011.02.009

Veloso HP, Rangel Filho ALR, Lima JCA (1991) Classificação da vegetação brasileira, adaptada a um sistema universal. Rio de Janeiro: Ministério da Economia, Fazenda e Planejamento, Fundação Instituto Brasileiro de Geografia e Estatística, Diretoria de Geociências, Departamento de Recursos Naturais e Estudos Ambientais. [ISBN: 978-85-240-0384-4]

Vidal F, Allen J (2023) MapBiomas External Evaluation and Impact Assessment. Tech. rep. São Paulo.

Weiss DJ, Nelson A, Gibson HS, Temperley W, Peedell S, Lieber A, Hancher M, Poyart E, Belchior S, Fullman N, Mappin B, Dalrymple U, Rozier J, Lucas TCD, Howes RE, Tusting LS, Kang SY, Cameron E, Bisanzio D, Battle KE, Bhatt, S, Gething PW (2018) A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature 553: 333–336. https://doi.org/10.1038/nature25181

Whittaker RH (1970) Communities and Ecosystems. Macmillan, New York, NY, USA.

WorldPop (2018) Global 1km Population. https://doi.org/10.5258/SOTON/WP00647

Yamazaki D, Ikeshima D, Tawatari R, Yamaguchi T, O'Loughlin F, Neal JC, Sampson CC, Kanae S, Bates PD (2017) A high accuracy map of global terrain elevations. Geophysical Research Letters 44: 5844–5853. https://doi.org/10.1002/2017GL072874

Yamazaki D, Ikeshima D, Sosa J, Bates PD, Allen GH, Pavelsky TM (2019) MERIT hydro a high resolution global hydrography map based on latest topography dataset. Water Resources Research 55: 5053–5073. https://doi.org/10.1029/2019WR024873

Zhao F, Jiao L, Liu H (2013) Kernel generalized fuzzy c-means clustering with spatial information for image segmentation. Digital Signal Processing 23: 184–199. https://doi.org/10.1016/j.dsp.2012.09.016

Supplementary materials

Supplementary material 1

Suppl. appendices S1–S4, figures S1–S5, tables S1–S8 (.docx)

Link: https://doi.org/10.21425/fob.18.145498.suppl1