

Capitalizing on a wealth of spatial information: Improving biogeographic regionalization through the use of spatial clustering

Peng Gao*, John A. Kupfer

Department of Geography, University of South Carolina, 709 Bull Street, Columbia SC, 29208, USA



ARTICLE INFO

Keywords:

Biogeographic regionalization
Angola
Mammals
Spatially constrained clustering

ABSTRACT

Biogeographic regionalization, the categorization of geographical areas on the basis of their biotas, provides a valuable approach for understanding biogeographical and ecological patterns and processes and serves as a valuable tool in conservation management practices. Contemporary, quantitative approaches for delineating and mapping biogeographic regions that make use of increasingly available species occurrence data have typically been conducted using clustering methods that do not consider spatial information of sample sites during the aggregation processes. This shortcoming has led to challenges in identifying spatial patterns and interpreting the underlying ecological factors responsible for these patterns. To address the shortcomings of non-spatial clustering methods and to highlight the value of utilizing spatial information during regionalization, we conducted biogeographic regionalization on a dataset of observed mammalian species locations in Angola using a non-spatial clustering method (Ward's clustering) and a spatial clustering method (Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning). When compared to results from the non-spatial method, biogeographic regions delineated by the spatial clustering method were more closely associated with distinct climatic conditions, had greater concurrence to accepted ecoregions, and were more strongly associated with species assemblages within these regions. We argue that the spatial clustering method makes regions more meaningful and interpretable; as a result, biogeographic regions identified by our approach could facilitate prioritizing conservation plans, developing natural resources management strategies, and reducing data complexity for spatial representation and ecological interpretation of species distribution.

1. Introduction

Biogeographers have long been interested in identifying and mapping regions that are distinct with respect to biotic structure or composition as a means for exploring and better understanding the factors responsible for shaping species distributions (Ficetola, Mazel, & Thuiller, 2017). Various tools and approaches associated with geographical regionalization have since provided analytical frameworks for a range of biogeographical, ecological, and environmental applications. For example, biogeographic regionalization has been used at scales ranging from the entire globe down to geographic regions, nations, and landscapes to prioritize conservation plans and options (Kier et al., 2005; Olson et al., 2001) and to aid in identifying areas in which similar or specific natural resources management strategies could be applied (Bernert, Eilers, Sullivan, Freemark, & Ribic, 1997; Hobbs & McIntyre, 2005). Regionalization is also a powerful tool that can be used to reduce data complexity for spatial representation and facilitate ecological interpretation (Kupfer, Gao, & Guo, 2012; Long, Nelson, & Wulder, 2010).

The practice of mapping biogeographic regions based on aspects of species composition or physiognomy dates back more than a century (Huggett, 2002; Wallace, 1876). Earlier approaches, based largely on researchers' knowledge of species distributions, were generally subjective and non-replicable, sometimes leading to disagreement among different parties (e.g., African biogeographic regions: Chapin, 1923; Cox, 2001; Dasmann, 1972; Diamond & Hamilton, 1980; Williams, de Klerk, & Crowe, 1999). Recently, the importance of quantitative, replicable and transparent regionalization methods has been recognized (Mackey, Berry, & Brown, 2008), and biogeographers have begun actively exploring and utilizing a variety of quantitative techniques to solve regionalization-related problems (e.g., Hattab et al., 2015).

Morrone (2018) has provided a useful overview of important concepts and methods involved in biogeographic regionalization. Generally speaking, one set of approaches to regionalizing biogeographic data is rooted in principles associated with biogeographic homology and seeks to identify spatial-temporal elements with a mutual biogeographic history (Morrone, 2001). An underlying assumption is that the resultant

* Corresponding author. Tel.: +001 803 777 6746; fax: +001 803 777 4972.

E-mail addresses: gaop@mailbox.sc.edu (P. Gao), kupfer@mailbox.sc.edu (J.A. Kupfer).

<https://doi.org/10.1016/j.apgeog.2018.08.002>

Received 24 January 2018; Received in revised form 1 August 2018; Accepted 12 August 2018

Available online 28 August 2018

0143-6228/© 2018 Elsevier Ltd. All rights reserved.

biogeographic regions are ‘natural’ or monophyletic, that is, the included biota possess a shared historical relationship or evolutionary history (Ebach & Parenti, 2015; Escalante, 2017). Such approaches are rooted in traditions associated with evolutionary biogeography, which emphasizes taxonomic groups and the importance of past biogeographic events such as dispersal, evolution and vicariance, and extinction. In terms of biogeographic regionalization, approaches focused on homology emphasize the identification of areas of endemism, commonly beginning with an analysis of the geographic distribution of individual species and leading to the hierarchical categorization of geographical areas with distinct sets of endemic taxa (Escalante, Morrone, & Rodríguez-Tapia, 2013; Morrone, 2018).

Other approaches, such as the ones used in this paper, seek instead to identify regions containing sets (communities) of species characteristic of that region and distinct from sets of species in other regions. Grounded in concepts from the field of ecological biogeography and utilizing tools common to multivariate ecological analysis (e.g., clustering), such approaches emphasize more contemporary time scales and ecological constraints on species patterns. The assumption is that clusters of species with similar niches form groups with similar morphological and physiological characteristics and shared common ecological roles (Crisci, Sala, Katinas, & Posadas, 2006); the goal of regionalization is thus to identify spatial units that effectively capture those groups based on a stated criterion such as within-vs. between-region species similarity. Biogeographic regions from this perspective are more akin to ecoregions, but with an explicit focus solely on biotic elements (e.g., species occurrences). Kreft and Jetz (2010) demonstrated the usefulness of these types of quantitative regionalization methods by evaluating the performance of nine clustering methods used to delineate regions based on the global distribution of mammalian fauna.

As Crisci et al. (2006) note, biogeographic approaches associated with historical- and ecological biogeography employ different concepts that are frequently explored in the literature but are rarely integrated. In this paper, we utilize clustering-based techniques because such approaches are more familiar to many applied biogeographers and ecologists and because our interests center on contemporary management applications that emphasize an understanding of community-level species patterns.

Despite the range of regionalization applications and concerted efforts aimed at exploring, utilizing, developing, and testing regionalization techniques, the important role of spatial adjacency or proximity has not been adequately considered in biogeographic regionalization procedures, even though spatial relationships can greatly influence ecological processes that shape biogeographic patterns and thus are of interest in designing effective networks of protected areas (e.g., Delmelle, Desjardins, & Deng, 2017; Gao, Kupfer, Guo, & Lei, 2013). Prevailing regionalization methods adopted by biogeographers, including *k*-means (e.g., Heikinheimo, Fortelius, Eronen, & Mannila, 2007; Mateo, Vanderpoorten, Munoz, Laenen, & Desamore, 2013; Razavi & Coulibaly, 2013; Xu et al., 2014), unweighted pair-group method using arithmetic averages algorithm/Average Linkage clustering (UPGMA/ALK) (e.g., Bradshaw, Colville, & Linder, 2015; Dapporto, Ciolli, Dennis, Fox, & Shreeve, 2015; Hattab et al., 2015; Kreft & Jetz, 2010), principal component analysis (e.g., Zhang, Wu, Wang, Yuan, & Zhao, 2011), and Ward's clustering (e.g., Rodrigues, Figueira, Vaz Pinto, Araújo, & Beja, 2015; Wohlgemuth, 1996), are all clustering or dimension reduction methods that do not consider spatial information during the aggregation processes. We use the term ‘non-spatial’ for such methods.

Though non-spatial techniques can be used for the purpose of regionalization by aggregating spatial units without using spatial information, they do not strictly follow the principles of regionalization (Huang, Fan, Li, & Wang, 2013), and the use of non-spatial approaches may impose challenges on interpreting spatial patterns and hinder the understanding of underlying ecological processes. Conversely, Andrew,

Wulder, and Coops (2011) found that regionalization methods that consider spatial constraints outperformed non-spatial methods in capturing patterns of butterfly community composition and species affinities that are strongly structured by space. Rickbeil et al. (2014) similarly found that spatially contiguous regionalization methods were superior to non-contiguous ones for delineating bird communities. Incorporating spatial contiguity into non-spatial clustering methods such as ALK also facilitated characterizing forest patterns and interpreting underlying ecological processes in different regions of the continental United States (Kupfer et al., 2012). In contrast, non-spatial clustering methods failed to distinguish small-scale forest formations and resulted in increased similarities of grid cells across borders of various animal regions in the regionalization of European biota (Rueda, Rodríguez, & Hawkins, 2010).

Biogeographic regionalization of species patterns employs data on species presence or abundance that are often collected from a limited number of accessible sites due to economic and logistical constraints. Sample sites also often have spatial bias because of (in)accessibility to potential sample locations. Given the inadequate consideration of spatial information in the procedure of biogeographic regionalization, approaches to utilizing spatial information from *discrete* and usually *unevenly distributed* samples when conducting regionalization on species distribution is even more underexplored. For instance, Rodrigues et al. (2015) had to compromise on the optimal number of regions when they interpreted the regionalization patterns of Angolan mammals because regions delineated by their non-spatial clustering method lost spatial coherence. To address the shortcomings of non-spatial clustering methods and to highlight the value of utilizing spatial information during regionalization, we present a regionalization method that incorporates spatial information from species sample points and compare results from our method against those obtained using non-spatial clustering methods.

2. Materials and methods

2.1. Study areas and species datasets

We used the same dataset of mammalian occurrences that was analyzed by Rodrigues et al. (2015) using a non-spatial clustering method to produce a biogeographical regionalization of Angola. The Republic of Angola is located in subtropical southern Africa and covers an area of ca. 1,246,700 km², with elevations that range from 0 to 2,620 m above sea level. Topography ranges from the arid coastal lowland, which is characterized by low plains and terraces, to hills and mountains that rise from the coast into a great inland escarpment, to a large area of high plains of dry savanna extending east from the escarpment, to rain forest in the north (Hance, 1975; Rodrigues et al., 2015). Generally, the climate includes a dry and hot season from May to October, a transitional season with some rain from November to January, and a rainy season from February to April, with coastal areas being affected by the cool Benguela Current. The Angolan terrestrial mammalian fauna includes ca. 275 species, including a number of species of conservation interest (Kuedikuenda & Xavier, 2009; Rodrigues et al., 2015).

The dataset included 9,880 occurrence records for 140 rodent, ungulate, and carnivore species that were collected from the literature of Crawford-Cabral (1989), Crawford-Cabral (1998), Crawford-Cabral and Simoes (1988), and Crawford-Cabral and Veríssimo (2005). The sampled points were aggregated to grids at a resolution of ¼ degree (approximately 25°25 km). The northwestern enclave of Cabinda was not included in the analyses, because it is discontinuous with the rest of Angola, and aquatic species were excluded. Grid cells that contained records for fewer than five species were excluded to avoid potential bias (Rodrigues et al., 2015), resulting in a total of 457 grid cells.

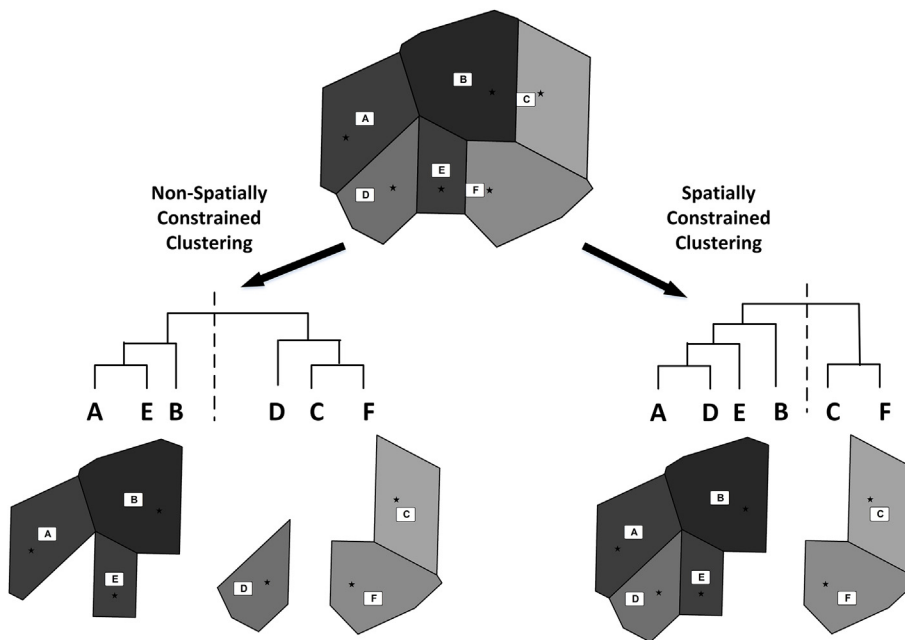


Fig. 1. A comparison of conventional non-spatially constrained clustering and spatially constrained clustering. Stars represent the centroids of sampled grid cells and polygons are Thiessen polygons that contain the centroids. Grey shading contrasts between polygons stand for the Simpson dissimilarity index (β_{sim}) between them. Non-spatially constrained clustering produces two clusters, one of which contains polygons (C, D, and E) that are spatially disjoint. In contrast, the two clusters produced by the spatially constrained clustering form two spatially contiguous regions.

2.2. Data analysis

Our regionalizations sought to find sets of grid cells (i.e., regions) that had more similar species composition within the region when compared to other such regions. This was done using both non-spatial and spatially-constrained clustering methods. The difference between these two approaches is whether the spatial adjacency of grid cells was used in the clustering process. Both approaches begin by aggregating units with less species dissimilarity into groups. We used Ward's clustering to perform the biogeographical regionalization on the Angolan mammal dataset so that our results would be consistent with those published in [Rodrigues et al. \(2015\)](#). Unlike the non-spatial methods, our method, REDCAP (Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning; [Guo \(2008\)](#)), ensures that the identified clusters form spatially continuous regions by enforcing a spatial contiguity constraint during the clustering procedure ([Fig. 1](#)).

Clustering began at the scale of individual grid cells on the basis of species occurrences. Dissimilarity in composition between all pairs of grid cells was quantified using the Simpson dissimilarity index (β_{sim}), which has the advantage of being independent to species richness gradients and not taking double absences into account ([Koleff, Gaston, & Lennon, 2003](#); [Kreft & Jetz, 2010](#)). The centroids of sampled grid cells were converted to Thiessen polygons, each of which contained one and only one cell. The area of each polygon represents the locality that is closer to the sampled cell in that polygon than to other cells. Thiessen polygons were chosen because they offer a simple approach for visualizing spatial variation in sampling intensity, with larger polygons suggesting sparser sampling ([Schulman, Toivonen, & Ruokolainen, 2007](#); [Vale & Jenkins, 2012](#)). Two cells were considered to be contiguous when the Thiessen polygons containing them were adjacent.

Initially, REDCAP considers each cell as an individual cluster. It then iteratively merges pairs of clusters that are spatially contiguous and have the lowest dissimilarity, which is determined as the averaged β_{sim} of each pair of cells across the two clusters to be merged. This follows the same clustering procedure as the conventional ALK except that REDCAP requires that the pair of clusters to be merged must be spatially contiguous. A spatially contiguous dendrogram is constructed once all the grid cells are merged. REDCAP then partitions the spatially contiguous dendrogram into the desired number of sub-trees assigned by a user, each of which represents a spatially contiguous region ([Fig. 1](#)). At each cutting of the dendrogram, REDCAP minimizes the sum

of within-region heterogeneity (H) for all regions. For a single region having n grid cells, H is defined as:

$$H = \sum_i^n \beta_{sim(i)} \quad (1)$$

where $\beta_{sim(i)}$ is the averaged β_{sim} of cell i to the other $(n-1)$ cells within the region.

To determine the statistically-optimal number of regions, we used the L method, which finds the 'elbow' in the plot of total H against the number of regions delineated by REDCAP ([Salvador & Chan, 2004](#)). However, we also examined regionalizations produced using other numbers of regions, which helps to avoid potential pitfalls associated with relying on a single, statistically-derived number of regions (cf. [Vilhena & Antonelli, 2015](#)) and can help to inform a better understanding of the hierarchical nature of biogeographic regions.

2.3. Evaluation

There is no single 'true' regionalization against which we can validate and compare the biogeographic regionalizations produced by REDCAP and the non-spatial clustering method. Therefore, evaluation of the methods was based on three criteria. The first criterion was the degree to which biogeographic regions were homogeneous with respect to climatic conditions. This criterion assumes that broad-scale regions delineated on the basis of species similarity should reflect the effect of environmental constraints known to structure species assemblages, represented here by climate. Correspondence between biogeographic regions and climatic patterns is also an important criterion for meshing biogeographic regionalizations with more broadly focused ecoregions.

We obtained data for 19 variables depicting aspects of temperature and precipitation from the WorldClim v.2.2 dataset ([Hijmans, Cameron, Parra, Jones, & Jarvis, 2005](#)), and conducted a Principal Component Analysis (PCA) to reduce collinearity of the variables for each study area ([Dormann et al., 2013](#)). We then identified individual climatic variables associated with each significant PCA axis and examined the differences of their distributions across regions identified by the spatial and non-spatial clustering methods. The Earth Mover's Distance (EMD), which has been used to examine differences in rainfall patterns ([Van den Berg, Vandenberghe, De Baets, & Verhoest, 2011](#)), was adopted to quantify the distribution differences of climate variables across regions. Intuitively, given two distributions to be compared, one can be seen as a

mass of earth spread in space, the other as a collection of holes in that same space. EMD measures the least effort of moving ‘dust’ both vertically and horizontally to fill the ‘holes’ (i.e., transforming one distribution into the other, Van den Berg et al. (2011)). Formally, the EMD of two histograms P and Q is defined as

$$EMD(P, Q) = \frac{\min \sum_{i,j} f_{ij} d_{ij}}{\left(\sum_{i,j} f_{ij} \right)} \quad (2)$$

where f_{ij} is the amount that needs to be transported from bin i to bin j to transform P to Q and d_{ij} is the ground distance between bin i and bin j in the histograms (Xu & Zhang, 2013). A larger EMD suggests a greater difference between two distributions. The R package ‘emd’ was used to calculate EMD (Urbanek & Rubner, 2012), and additional background on issues associated with calculating EMD based on the transportation problem can be found in Rubner, Tomasi, and Guibas (1998) and Van den Berg et al. (2011). To facilitate comparisons of EMD among regions, the R package ‘corplot’ was used to visualize the EMD matrix (Wei & Simko, 2016). Our assumption is that a ‘more effective’ biogeographic regionalization would result in regions with more consistent and distinct climatic patterns.

The second criterion involved concurrence between mapped biogeographic regions and previously identified ecoregions used for conservation and management purposes. Ecoregions are recognized as broad areas united by a similar climate, topography, and geology, which leads to a geographically-distinct assemblage of species and natural communities (e.g., Bailey, 2014). While biogeographic regions are typically defined based solely on biotic factors without reference to the abiotic factors responsible for those patterns, a large degree of agreement between biogeographic regions and ecoregions typically occurs because the abiotic factors included in ecoregional definitions are often those most responsible for dictating broad-scale biogeographic patterns (e.g., Lomolino, Riddle, Brown, & Brown, 2006). Here, we used ecoregions accepted by the World Wide Fund for Nature (WWF) (Olson et al., 2001), which recognizes fourteen terrestrial ecoregions in Angola.

To examine correspondence between ecoregions and biogeographic regions delineated in this study, we extracted the derived biogeographic region and the WWF ecoregion for each grid cell using ArcGIS. For each biogeographic region, we tallied the number of cells classified as belonging to each ecoregion (e.g., for Biogeographic Region 1: the number of cells classified as Angolan miombo woodlands, Angolan montane forest-grassland mosaic, Namibian savanna woodlands, and so on). Similarly, for each ecoregion, we tallied the number of cells classified as belonging to each biogeographic region (e.g., for Angolan miombo woodlands: the number of cells classified as Biogeographic Region 1, Region 2, Region 3, and so on). A perfect correspondence would result in all cells from each biogeographic region being associated with a single ecoregion and vice versa. Perfect correspondence is highly unlikely, so we assumed that a more effective biogeographic regionalization would be demonstrated by greater correspondence between biogeographic regions and ecoregions as well as a lower number of biogeographic regions (or ecoregions) per ecoregion (or biogeographic region).

The third criterion assessed how strongly species characterize a region and discriminate differences in species assemblages among regions. While our method is not specifically focused on endemics, our assumption is that a more useful biogeographic regionalization will be characterized by more species that have a stronger association with individual regions, rather than being spread across many regions. Conceptually, this approach follows that of Procheş and Ramdhani (2012), who examined patterns of characteristic taxa in analytically-defined global zoogeographical clusters. In this study, we conducted indicator species analysis using a modified form of the Indicator Value

Method (IVM) (*sensu* Dufrêne & Legendre, 1997 Cáceres & Legendre, 2009) for regions delineated by REDCAP and non-spatial clustering methods.

Indicator species, defined as species that are used as ecological indicators of pre-defined community or habitat types, are usually determined using an analysis of the relationship between the observed species' presence-absence (or abundance values) in a set of sampled cells and a classification of the same cell (Cáceres & Legendre, 2009). Here, the cell classifications were the biogeographic regions defined by REDCAP and the non-spatial clustering method. Because our data consisted of species presences (rather than abundances), the species indicator values were calculated as the product of two factors: 1) the relative frequency of the species in the target region divided by the sum of relative frequencies over all regions, and 2) the relative frequency of occurrence of the species inside the target region. The former factor measures the probability that a sample cell belongs to the target site region given that the species has been found, while the latter factor is the probability that the species is present when the cell belongs to the target cell region (Cáceres & Legendre, 2009). After identifying significant indicator species ($p < 0.05$) for each biogeographic region in a classification, we divided the number of indicators by the total number of species in each region. The analysis was conducted using the R package ‘indicspecies’ (Cáceres & Jansen, 2016).

3. Results

Based on the L method using REDCAP results, the optimal number of biogeographic regions for Angolan mammals was six; consequently, all analyses were performed at this level of detail for the biogeographic regionalization. Visually, the regions delineated using the non-spatial approach (Fig. 2a) and REDCAP (Fig. 2b) differed in the greater continuity of regions produced using the latter approach, which results from enforcing spatial contiguity during the clustering process. In contrast, regions produced by the non-spatial method were patchier with greater interspersed regions. Because REDCAP produces spatially-constrained hierarchical regions, it is possible to examine clusters produced at any level of aggregation and to examine the biotic similarities among clusters (Figs. S1 and S2 in supplementary materials).

In the first evaluation step, we selected four variables based on the PCA results to investigate climatic variability within the biogeographic regions: annual precipitation, precipitation of the warmest quarter, mean temperature of the driest quarter, and maximum temperature of the warmest month. These variables mirrored those selected by Rodrigues et al. (2015). In most cases, the six biogeographic regions delineated by REDCAP were marked by: 1) climatic conditions that were unique to that region and distinct from other regions (as indicated by the darker color and larger size of the dots representing larger EMD in Fig. 3), and 2) a narrower range of values for one or more variables than those for the regions delineated using the non-spatial approach (the range of climate variables in a region compared to the ranges in other regions or the entire study area in Fig. 3). For example, Region 4, which was located in the arid southwestern corner of Angola, had low values for all four variables (low regional maxima compared to the maximum values of the study area in Fig. 3). However, the range of values for the four variables in REDCAP generated regions was much smaller when compared to the corresponding region produced by non-spatial Ward's clustering. For example, the minimum values of variables were the same in Region 4 generated by REDCAP and Ward's clustering (annual precipitation: 43 mm; mean temperature of the driest quarter: 28 mm; mean temperature of the driest quarter: 15.9 °C; maximum temperature of the warmest month: 25.0 °C), but REDCAP Region 4 had lower regional maxima: annual precipitation (285 mm vs. 926 mm); precipitation of the warmest quarter (188 mm vs. 470 mm); mean temperature of the driest quarter (20.3 °C vs. 22.4 °C); and maximum temperature of the warmest month (29.4 °C vs. 34.0 °C) (Figs. 3 and 4).

As another example, Region 6, which occupied the southeastern

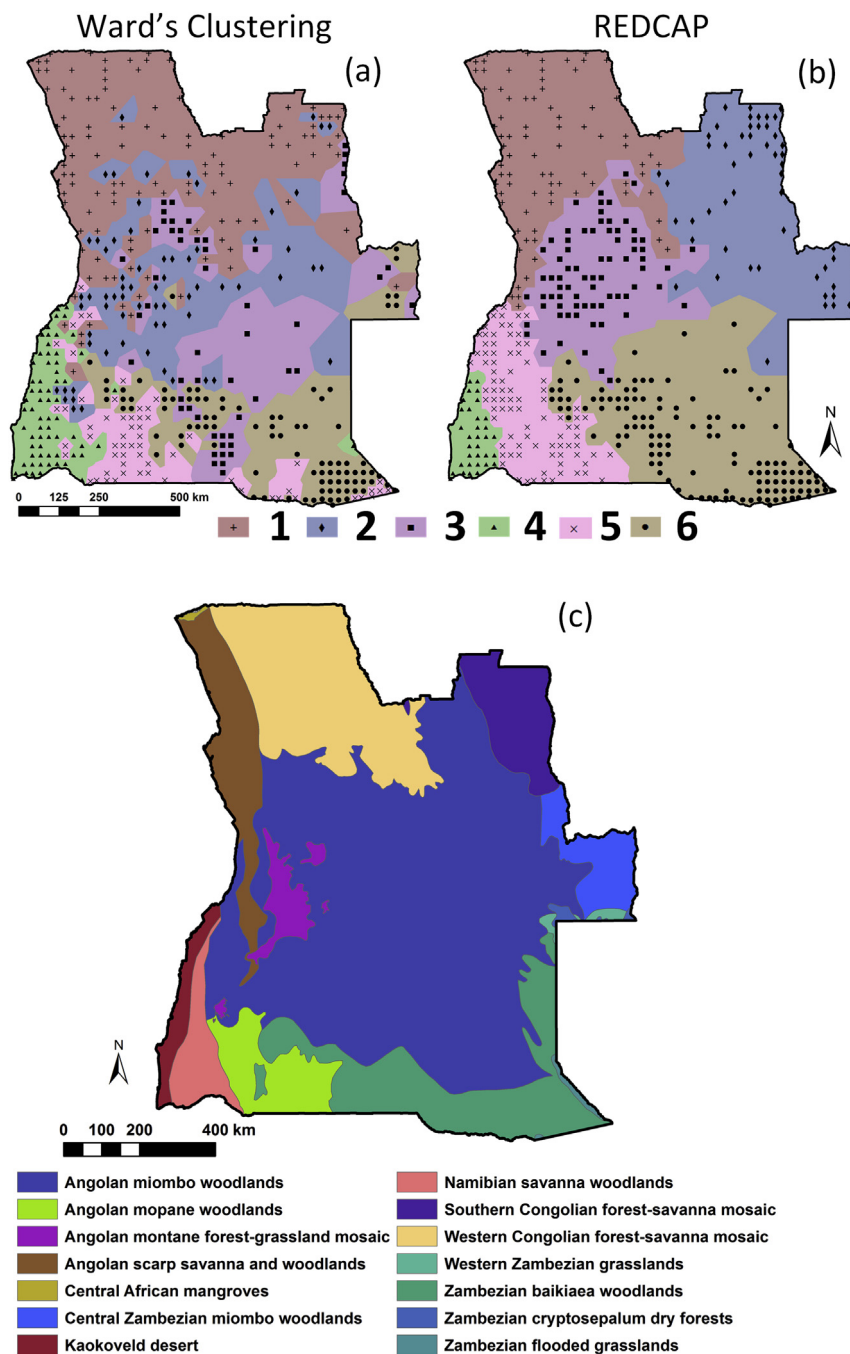


Fig. 2. Biogeographic regions for Angolan mammals delimited by (a) Ward's clustering and (b) Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP); (c) previously delineated ecoregions accepted by the World Wide Fund for Nature (WWF) (Olson et al., 2001).

quarter of Angola, was characterized by low mean temperatures in the driest quarter and high maximum temperature in the warmest month and, like Region 4, had far less variation in climatic values when compared to its non-spatial counterpart. Regions 2 and 3 both had high annual precipitation but were distinguished by high maximum temperature of the warmest month. These regions were also much more clearly differentiated (climatically) by REDCAP than the non-spatial Ward's method. While Regions 1 and 5 covered wide ranges of the four climate variables and showed the least distinction between regions created by the two methods, our results supported the hypothesis that the use of spatially-constrained clustering would produce biogeographic regions that were more homogeneous with respect to climatic conditions than those produced using non-spatial clustering.

For our second evaluation criterion, biogeographic regions

produced by both methods showed some correspondence to the WWF ecoregions (Fig. 2c), but also marked areas of differentiation, especially for regions defined using Ward's clustering. The regions produced by REDCAP typically differed from the WWF ecoregions for one of two reasons. First, the REDCAP regions sometimes comprised aggregations of two or more smaller ecoregions. This was the case for Region 4, which coincided with the Kaokoveld desert and the southern half of the Namibian savanna woodlands ecoregions, as well as Region 1, which covered three WWF ecoregions (Angolan scarp savanna and woodlands, Western Congolian forest-savanna mosaic, Central African mangrove). When such areas were aggregated, there was considerable overlap between the ecoregions and the biogeographic regions. In other cases, the biogeographic regions identified finer divisions within more extensive ecoregions. Examples of this pattern included Regions 2, 3, and 6,

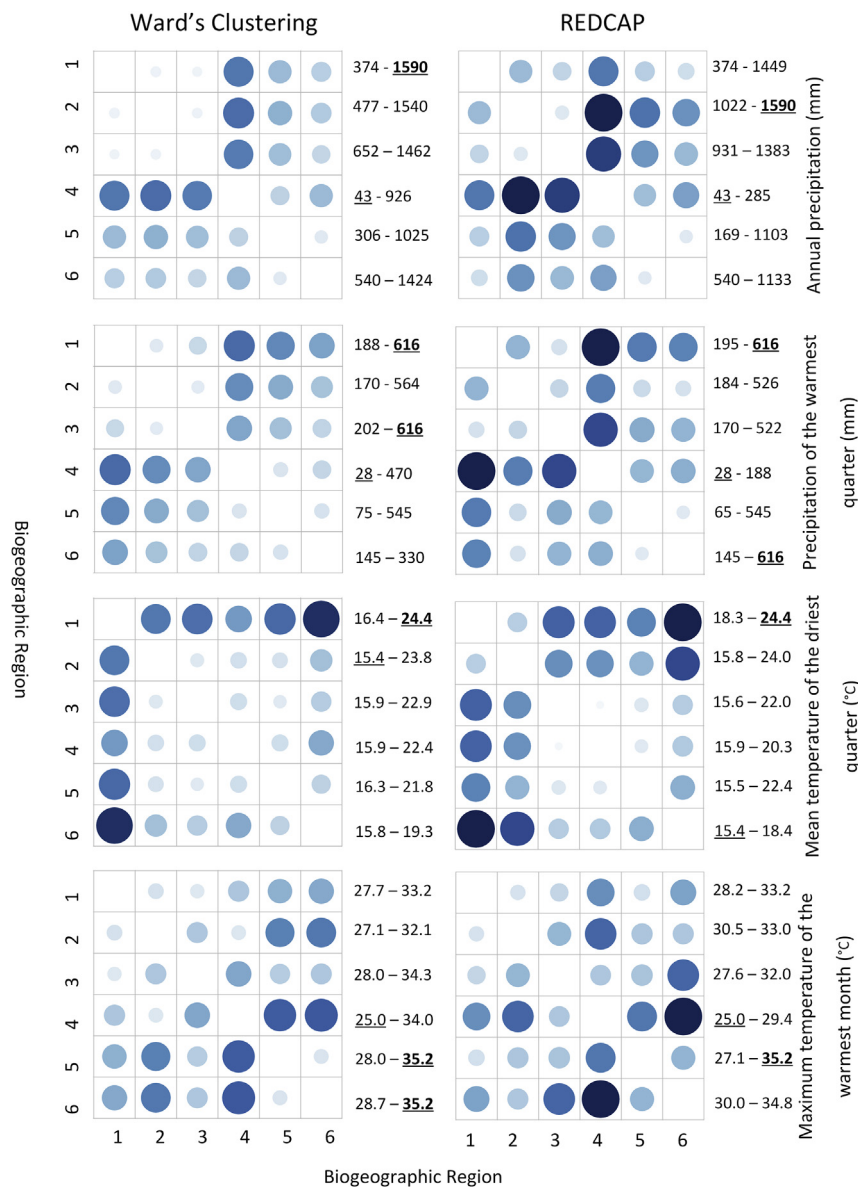


Fig. 3. Distinctness of four climatic variables between pairwise biogeographic regions for Angolan mammals delineated by Ward's clustering and Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP). Darker and larger dots represent greater differences quantified by the Earth Mover's Distance (EMD). Values in each row are the range of the variable in that region. Underlined numbers are minimum values of the entire study area. Underlined and bolded numbers are maximum values of the entire study area.

which split the Angolan Miombo woodlands that comprise more than half of the area in Angola but span a range of climatic and geomorphic conditions. Finally, in contrast to the regions produced by REDCAP, regions identified solely by Ward's clustering lost spatial coherence and were broken into spatially disjoint pieces spread across Angola, although Regions 1, 4, 5, and 6 had some approximation to regions detected by REDCAP.

Quantitatively, grid cells within each REDCAP-generated biogeographic region were associated with fewer ecoregions (mean = 4.3) than those for regions delineated by non-spatial clustering (mean = 6.0) (*analyzing down columns in Table 1*). All cells assigned to Region 4 in the REDCAP classification, for example, were from the Kaokoveld desert or Namibian savanna woodlands ecoregions, whereas 26% of the cells in Region 4 in the non-spatial classification came from four other ecoregions. Similarly, nearly all (95.4%) of the cells assigned to REDCAP Region 6 were from the Angolan Miombo Woodlands or Zambebian baikiaea woodlands ecoregions vs. only 86% of the cells in Region 6 in the non-spatial classification, with the remaining locations

coming from four other ecoregions. A comparable pattern was observed in which the number of grid cells from each ecoregion fell into fewer biogeographic regions as delineated by REDCAP (mean = 2.08) vs. non-spatial Ward's clustering (mean = 2.92) (*analyzing across rows in Table 1*). In fact, cells from five ecoregions (Angolan mopane woodlands, Central African mangroves, Central Zambebian miombo woodlands, Southern Congolian forest-savanna mosaic, Zambebian flooded grasslands) were associated with a single biogeographic region in the REDCAP regionalization, with an additional three ecoregions split between only two adjacent regions. These findings support our expectations that spatially-constrained clustering would delineate biogeographic regions with greater concurrence to accepted ecoregions.

The final criterion involved the identification of indicator species and their fidelity to specific biogeographic regions. For the six regions produced by Ward's clustering, we identified 99 indicator species, but 40 were associated with a single biogeographic region while three regions had 10 or fewer indicator species (*Table 2*). The percentage of all species in individual biogeographic regions that were identified as

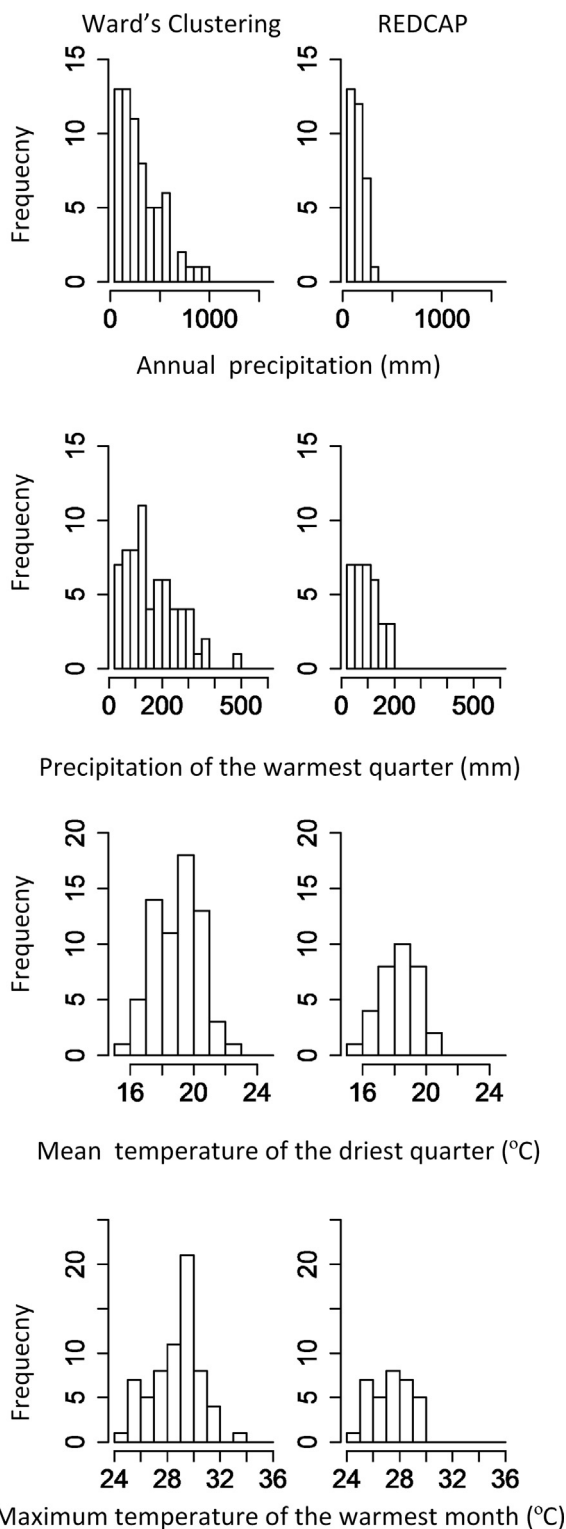


Fig. 4. Histograms of four climatic variables in Region 4 delineated by Ward's clustering and Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP) for Angolan mammals.

indicator species for that region ranged from 8.7–37.0%. Biogeographic regions produced by REDCAP had a similar total number of indicator species (92), but they were more evenly spread across the regions, with five of the six REDCAP regions exhibiting a greater percentage of indicator species per region than those identified through Ward's clustering. The results thus suggest that species had a stronger association with biogeographic regions produced by spatially-constrained

clustering (REDCAP) than by a non-spatial approach.

Regions 2 and 3 provide examples of how differences in the continuity of regions translated into differences in the fidelity of specific indicator species. The species with the highest indicator values for REDCAP Regions 2 and 3 were *Kobus vardonii* and *Myomyscus angolensis*, respectively (Fig. 5a and b). By its nature, REDCAP identified spatially contiguous regions, which in turn had spatially-concentrated indicator species with relatively few observed occurrences falling outside of their associated biogeographic regions. Further, *K. vardonii* is an endangered species threatened by expanding human settlements and cultivation (IUCN, 2016; Kingdon & Hoffmann, 2013) while *M. angolensis* is a regional endemic found in western Angola, primarily on the Angolan Plateau (Crawford-Cabral, 1998); the contiguous nature of regions could help land managers and scientists in identifying effective conservation areas for these (and related) species. In contrast, the patchy distribution of Regions 2 and 3 delimited by Ward's clustering was most highly characterized by *Lemniscomys griselda* and *Kobus leche* (Fig. 5c and d), both of which are widespread species of least concern (IUCN, 2016).

4. Discussion

The identification and delineation of biogeographic regions remains an active area of ongoing research, with advancements taking place along two fairly distinct biogeographic threads (e.g., Linder et al., 2012; Morrone, 2018; Whittaker, Riddle, Hawkins, & Ladle, 2013). The development of new methodologies, technologies, and approaches for analyzing large species databases coupled with greater accessibility to species location data through online databases such as the Global Biodiversity Information Facility database (<http://www.gbif.org/>) or the U.S. Geological Survey's BISON database (Biodiversity Information Serving Our Nation: <https://bison.usgs.gov>) facilitate more quantitative approaches to biogeographic regionalization. Studies such as this one provide insights that can improve the quality, interpretability, and ecological relevance of the regions derived using more quantitative approaches. In particular, the growing wealth of spatially-referenced species data, which is also used in species distribution modeling (e.g., Brotons, Thuiller, Araujo, & Hirzel, 2004; Hill et al., 2017; Phillips & Dudik, 2008; Zaniwski, Lehmann, & Overton, 2002), coupled with improved tools for analyzing them, represents a valuable potential resource for quantitative approaches to biogeographic regionalization, much of which is not adequately utilized by non-spatial clustering-based approaches.

While traditional clustering and classification methods are widely used and can yield spatially coherent clusters (He, Kreft, Gao, Wang, & Jiang, 2017), our results underscore that regions delineated by the spatial clustering method (REDCAP) may be more interpretable because the spatial coherence in regions is better maintained during the clustering procedure. The non-spatial clustering method yielded biogeographic regions that were challenging to interpret, especially when considering more than a very small number of regions. For example, the optimal number of biogeographic regions delineated for the Angolan mammal dataset by Rodrigues et al. (2015) using non-spatial Ward's clustering was twelve, but they were unable to interpret the spatial patterns of mammal distributions for more than four regions because clusters of grid cells started losing spatial coherence as the number of regions increased. Romeiras, Figueira, Duarte, Beja, and Darbyshire (2014) similarly noticed that clusters of Angolan timber species produced by non-spatial clustering methods lacked spatial consistency beyond a few basic regions and suggested that spatial coherence needed to be considered in determination of the optimal number of clusters for interpretation. REDCAP, as a spatial clustering method, has the advantage of maintaining spatial coherence, which made the biogeographic regions in both case studies more meaningful, a finding echoing those of Divíšek, Storch, Zelený, and Culek (2016).

In addition to producing biogeographic regions with greater spatial

Table 1

Number of grid cells in each combination of World Wide Fund for Nature (WWF) ecoregions and biogeographic regions identified by Ward's clustering or Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP) for Angolan mammal data.

World Wide Fund for Nature (WWF) Ecoregions	Ward							REDCAP						
	Region ID						Number of Biogeographic Regions	Region ID						Number of Biogeographic Regions
	1	2	3	4	5	6		1	2	3	4	5	6	
Angolan miombo woodlands	31	40	37	11	13	38	6	15	15	53		29	58	5
Angolan montane forest-grassland mosaic	5	12					2			15		2		2
Angolan mopane woodlands		1		4	23	1	4					29		1
Angolan scarp savanna and woodlands	28	5			1	1	4	30		3		2		3
Central African mangroves	2						1	2						1
Central Zambezan miombo woodlands	2		2			6	3		10					1
Kaokoveld desert				18			1				9	9		2
Namibian savanna woodlands				31	2		2			24	9			2
Southern Congolian forest-savanna mosaic	16	3	5				3	1	23					1
Western Congolian forest-savanna mosaic	32	5					2	32	2	3				3
Zambezan baikiaea woodlands		1	10	1	22	42	5		1			8	67	3
Zambezan flooded grasslands				1		5	2						6	1
Number of Ecoregions	7	7	3	6	5	6		5	5	4	2	7	3	

Table 2

Numbers and percentages of indicator species in biogeographic regions delineated by non-spatial clustering methods (Ward's clustering) and Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP) for Angolan mammals.

Region ID	Ward's Clustering		REDCAP	
	Number of Indicator Species	Percent (%)	Number of Indicator Species	Percent (%)
1	9	8.9	11	14.5
2	40	37.0	18	24.0
3	11	14.3	19	22.1
4	23	33.8	21	45.7
5	10	13.5	12	14.0
6	6	8.7	11	14.3

coherence, spatially-constrained methods show promise in being able to better identify transitions between areas with distinct biotic characteristics. Ecoregions, by definition, are large areas within which local ecosystems recur in fairly predictable fashions on similar sites; consequently, efforts to delineate transitions (places where those patterns 'change') have been an important aspect of mapping ecoregions (Bailey, 2004). Transition zones are similarly of interest to biogeographic regionalizations, although it has been previously noted that non-spatial, clustering-based approaches to bioregionalization have an inherent limitation in dealing with such transitions (Kreft & Jetz, 2013). For this dataset, REDCAP produced spatially-coherent regions with fewer boundaries whose locations corresponded better to recognized ecoregion boundaries. Further, because of its hierarchical nature and ability to allow users to interactively alter and rapidly display coarser or more detailed regionalizations, REDCAP can serve as a tool for exploring locations of biotic transition zones.

With respect to the evaluation criteria, REDCAP was better able to identify regions with more distinct climatic conditions that had narrower ranges than those delineated by Ward's clustering, as would be expected of an effective regionalization (Bailey, 2004). This trend was more pronounced in the regions where spatial coherence was poorly maintained by non-spatial clustering methods, such as Regions 2 and 3 (the second and third row/column in matrices in Fig. 3). Space is often strongly related to ecological and biogeographical patterns, in part because of spatial structure inherent in many environmental gradients in the form of spatial autocorrelation but also because of its influence on spatially-contagious processes such as dispersal. It is well known that space and location play a key role in producing climatic gradients,

which in turn shapes biotic patterns. By incorporating spatial information during clustering, REDCAP more effectively distinguished the characteristics of climate variables in different biogeographic regions. Since climate is one of the key factors that defines ecoregions, it is reasonable that biogeographic regions created by the spatial clustering had a better correspondence with accepted ecoregions. The distinct climate conditions also strongly influenced the distribution of mammals among these regions. Having more distinct climate variables, the regions detected by the spatial clustering method resulted in a higher percentage of indicator species in most regions (Table 2).

Finally, biogeographic regions are ultimately meant to capture and simplify important patterns in species distributions and assemblages, leading some to conceptualize them as 'operational species pools' that provide insights into the physical and biological forces that have shaped the geographical organization of biodiversity (Carstensen, Lessard, Holt, Borregaard, & Rahbek, 2013). Consequently, as noted by Vasconcelos, Prado, da Silva, and Haddad (2014: p1), "Regionalization schemes thus provide spatially explicit frameworks for answering many basic and applied questions in historical and ecological biogeography, evolutionary biology, systematics, and conservation". In some circumstances, biogeographic regionalizations could help to suggest unique or distinct locations for the conservation of species of concern. Compared to ecoregions, which are usually defined by similar conditions of climate, topography, geology and soils, species-based biogeographic regionalization can identify species-specific regions that have great conservation concerns (Romeiras et al., 2014). This advantage was enhanced by the spatial clustering in this study (e.g., indicator species in REDCAP region 2 and 3: Fig. 5).

In the context of biogeographic homology, biogeographic regions delineated by clustering methods such as those employed here are 'artificial' (*sensu* Ebach & Parenti, 2015) in that they are statistical constructs that may or may not be correspond to 'natural' biotic areas associated with the evolutionary history of the species involved, particularly endemics. The imposition of adjacency as an aggregation constraint could be criticized as adding a further element of artificiality in that regions with different historical and ecological features could be statistically aggregated before those with taxa having closer historic or evolutionary ties. From the perspective of ecological biogeography, however, results of this study underscore the potential value of such an approach, as long as the regions that are defined are also evaluated with respect to their characteristic species and for their relationships to known environmental constraints. Where further research is needed is in determining and assessing the tradeoffs between: 1) aggregating to a smaller number of regions, which may be more interpretable and

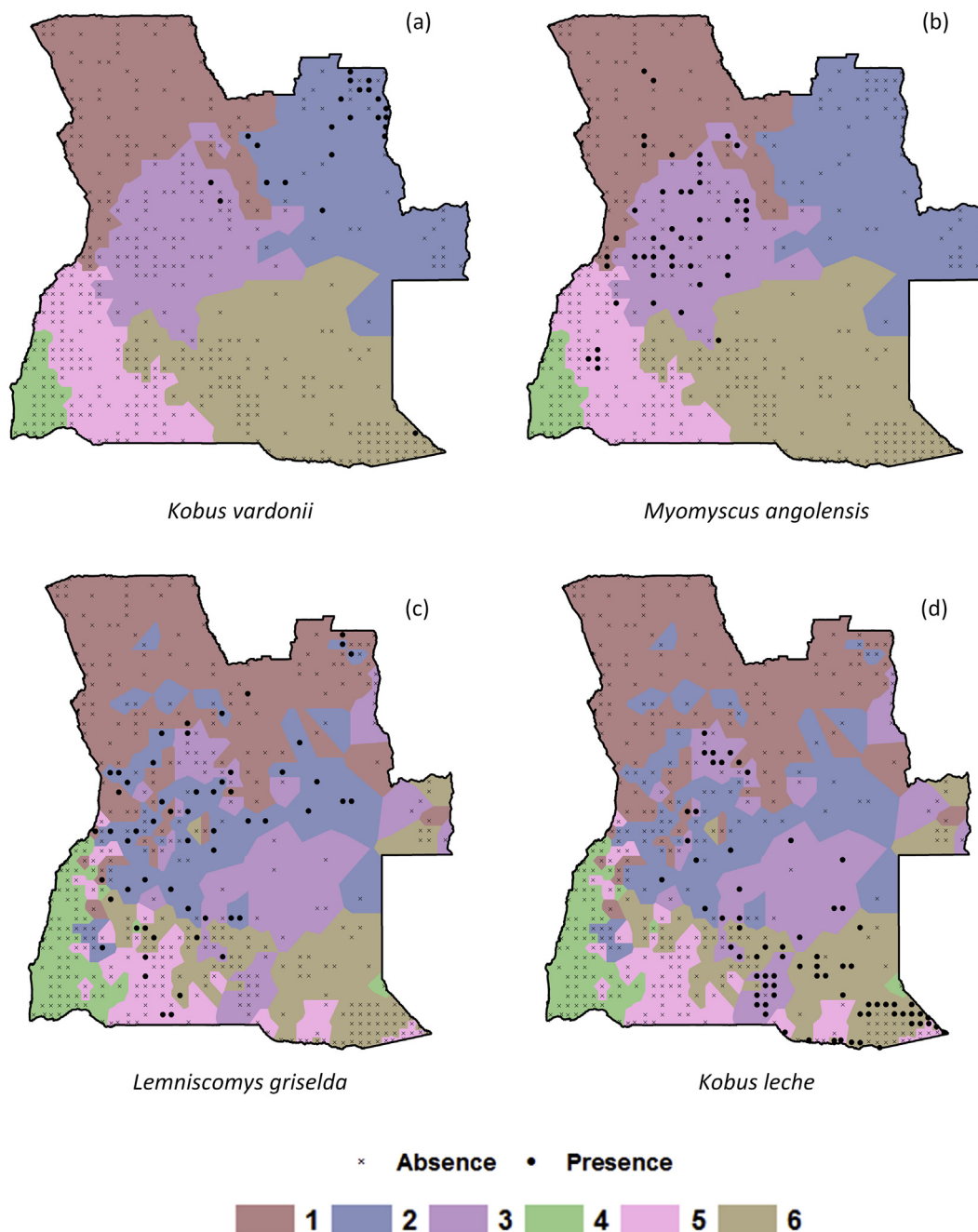


Fig. 5. Indicator species with the highest indicator values for Biogeographic Regions 2 and 3 in the Angolan study area as delineated by (a–b) Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP) and (c–d) Ward's clustering. Black dots represent documented species locations.

spatially coherent but will, by default, include areas of greater community heterogeneity and perhaps incorporate discontinuous regions that are better left unmerged, and 2) using a larger number of regions that are more internally consistent but are spatially-disjunct and perhaps more difficult to interpret.

5. Conclusion

This study was motivated by recent interest and advancements in quantitative approaches to identifying and delineating biogeographic regions. Given the burgeoning interest in this area, we compared biogeographic regions developed using a spatial clustering method, REDCAP, with those identified by a non-spatial clustering method. The

biogeographic regions delineated by the spatial clustering method had more distinct patterns of climatic variables and stronger association with accepted ecoregions and species assemblages due to the incorporation of spatial information during clustering. The spatial clustering method also delineated biogeographic regions that were more interpretable and had valuable ecological implications, which could facilitate the prioritization of conservation plans, the development of natural resources management strategies, and the reduction of data complexity for spatial representation and ecological interpretation of species distributions.

Declarations of interest

None.

Acknowledgements

The authors would especially like to thank Patrícia Rodrigues and Pedro Beja, for the generous offer of the Angolan mammal dataset used in their published work. We also appreciate the constructive criticism and suggestions of two anonymous reviewers, which greatly improved this manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.apgeog.2018.08.002>.

References

- Andrew, M. E., Wulder, M. A., & Coops, N. C. (2011). How do butterflies define ecosystems? A comparison of ecological regionalization schemes. *Biological Conservation*, 144(5), 1409–1418.
- Bailey, R. G. (2004). Identifying ecoregion boundaries. *Environmental Management*, 34, S14–S26.
- Bailey, R. G. (2014). *Ecoregions: The ecosystem geography of the oceans and continents*. Berlin: Springer Science & Business Media.
- Bernert, J. A., Eilers, J. M., Sullivan, T. J., Freemark, K. E., & Ribic, C. (1997). A quantitative method for delineating regions: An example for the western Corn Belt plains ecoregion of the USA. *Environmental Management*, 21(3), 405–420.
- Bradshaw, P. L., Colville, J. F., & Linder, H. P. (2015). Optimising regionalisation techniques: Identifying centres of endemism in the extraordinarily endemic-rich cape floristic region. *PLoS One*, 10(7), e0132538.
- Brotons, L., Thuiller, W., Araujo, M. B., & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27(4), 437–448.
- Cáceres, M. D., & Jansen, F. (2016). *indicspecies: Relationship between species and groups of sites*. R package version 1.7.6. Available at: <https://cran.r-project.org/web/packages/indicspecies/index.html>, Accessed date: 24 May 2018.
- Cáceres, M. D., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, 90(12), 3566–3574.
- Carstensen, D. W., Lessard, J. P., Holt, B. G., Borregaard, M. K., & Rahbek, C. (2013). Introducing the biogeographic species pool. *Ecography*, 36(12), 1310–1318.
- Chapin, J. P. (1923). Ecological aspects of bird distribution in tropical Africa. *American Naturalist*, 57, 106–125.
- Cox, C. B. (2001). The biogeographic regions reconsidered. *Journal of Biogeography*, 28(4), 511–523.
- Crawford-Cabral, J. (1989). Distributional data and notes on Angolan carnivores (Mammalia: Carnivora). 1—small and median-sized carnivores. *Garcia de Orta, Séries de Zoologie, Lisboa*, 14(2), 3–27.
- Crawford-Cabral, J. (1998). *The Angolan rodents of the superfamily Muroidea: An account on their distribution*. Lisbon (Portugal): Instituto de Investigação Científica Tropical.
- Crawford-Cabral, J., & Simoes, A. P. (1988). Distributional data and notes on Angolan carnivores (Mammalia: Carnivora). 2—Larger species. *Garcia de Orta (Portugal)*, 14, 3–27.
- Crawford-Cabral, J., & Veríssimo, L. N. (2005). The ungulate fauna of Angola: Systematic list, distribution maps, database report. *Ministério da Ciência, Tecnologia e Ensino Superior*. Lisboa, Portugal: Instituto de Investigação Científica Tropical.
- Crisci, J. V., Sala, O. E., Katinas, L., & Posadas, P. (2006). Bridging historical and ecological approaches in biogeography. *Australian Systematic Botany*, 19(1), 1–10.
- Dapporto, L., Ciolli, G., Dennis, R. L. H., Fox, R., & Shreeve, T. G. (2015). A new procedure for extrapolating turnover regionalization at mid-small spatial scales, tested on British butterflies. *Methods in Ecology and Evolution*, 6(11), 1287–1297.
- Dasmann, R. F. (1972). Towards a system for classifying natural regions of the world and their representation by national parks and reserves. *Biological Conservation*, 4(4), 247–255.
- Delmelle, E., Desjardins, M. R., & Deng, J. (2017). Designing spatially cohesive nature reserves with backup coverage. *International Journal of Geographical Information Science*, 31(12), 2505–2523.
- Diamond, A. W., & Hamilton, A. C. (1980). The distribution of forest passerine birds and quaternary climatic change in tropical Africa. *Journal of Zoology*, 191(Jul), 379–402.
- Divíšek, J., Storch, D., Zelený, D., & Culek, M. (2016). Towards the spatial coherence of biogeographical regionalizations at subcontinental and landscape scales. *Journal of Biogeography*, 43(12), 2489–2501.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46.
- Dufrène, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, 67(3), 345–366.
- Ebach, M. C., & Parenti, L. R. (2015). The dichotomy of the modern bioregionalization revival. *Journal of Biogeography*, 42(10), 1801–1808.
- Escalante, T. (2017). A natural regionalization of the world based on primary biogeographic homology of terrestrial mammals. *Biological Journal of the Linnean Society*, 120(2), 349–362.
- Escalante, T., Morrone, J. J., & Rodriguez-Tapia, G. (2013). Biogeographic regions of North American mammals based on endemism. *Biological Journal of the Linnean Society*, 110(3), 485–499.
- Ficetola, G. F., Mazel, F., & Thuiller, W. (2017). Global determinants of zoogeographical boundaries. *Nature Ecology & Evolution*, 1, 0089.
- Gao, P., Kupfer, J. A., Guo, D., & Lei, T. L. (2013). Identifying functionally connected habitat compartments with a novel regionalization technique. *Landscape Ecology*, 28(10), 1949–1959.
- Guo, D. (2008). Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). *International Journal of Geographical Information Science*, 22(7), 801–823.
- Hance, W. A. (1975). *Geography of modern Africa*. New York: Columbia University Press.
- Hattab, T., Albouy, C., Lasram, F. B., Le Loc'h, F., Guilhaumon, F., & Leprieux, F. (2015). A biogeographical regionalization of coastal Mediterranean fishes. *Journal of Biogeography*, 42(7), 1336–1348.
- Heikinheimo, H., Fortelius, M., Eronen, J., & Mannila, H. (2007). Biogeography of European land mammals shows environmentally distinct and spatially coherent clusters. *Journal of Biogeography*, 34(6), 1053–1064.
- He, J. K., Kreft, H., Gao, E. H., Wang, Z. C., & Jiang, H. S. (2017). Patterns and drivers of zoogeographical regions of terrestrial vertebrates in China. *Journal of Biogeography*, 44(5), 1172–1184.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978.
- Hill, L., Hector, A., Hemery, G., Smart, S., Tanadini, M., & Brown, N. (2017). Abundance distributions for tree species in Great Britain: Atwo-stage approach to modeling abundance using species distribution modeling and random forest. *Ecology and Evolution*, 7(4), 1043–1056.
- Hobbs, R. J., & McIntyre, S. (2005). Categorizing Australian landscapes as an aid to assessing the generality of landscape management guidelines. *Global Ecology and Biogeography*, 14(1), 1–15.
- Huang, B. R., Fan, T., Li, Y. M., & Wang, Y. (2013). Division scheme for environmental management regionalization in China. *Environmental Management*, 52(1), 289–307.
- Huggett, R. (2002). *Biogeographical regions*. Chichester: John Wiley & Sons Ltd.
- IUCN. (2016). *The IUCN red list of threatened species*. Available at: www.iucnredlist.org, 2016-3, Accessed date: 24 May 2018.
- Kier, G., Mutke, J., Dinerstein, E., Ricketts, T. H., Küper, W., Kreft, H., et al. (2005). Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, 32(7), 1107–1116.
- Kingdon, J., & Hoffmann, M. (2013). *Mammals of Africa volume VI pigs, hippopotamuses, chevrota, giraffes, deer and bovids*. London: Bloomsbury Publishing.
- Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, 72(3), 367–382.
- Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, 37(11), 2029–2053.
- Kreft, H., & Jetz, W. (2013). Comment on “an update of wallace's zoogeographic regions of the world”. *Science*, 341(6144).
- Kuedikuenda, S., & Xavier, M. N. (2009). *Framework report on Angola's biodiversity*. Luanda: Republic of Angola, Ministry of Environment.
- Kupfer, J. A., Gao, P., & Guo, D. (2012). Regionalization of forest pattern metrics for the continental United States using contiguity constrained clustering and partitioning. *Ecological Informatics*, 9, 11–18.
- Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldsa, J., & Rahbek, C. (2012). The partitioning of Africa: Statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography*, 39(7), 1189–1205.
- Lomolino, M. V., Riddle, B. R., Brown, J. H., & Brown, J. H. (2006). *Biogeography*. MA: Sinauer Associates Sunderland.
- Long, J., Nelson, T., & Wulder, M. (2010). Regionalization of landscape pattern indices using multivariate cluster analysis. *Environmental Management*, 46(1), 134–142.
- Mackey, B. G., Berry, S. L., & Brown, T. (2008). Reconciling approaches to biogeographical regionalization: A systematic and generic framework examined with a case study of the Australian continent. *Journal of Biogeography*, 35(2), 213–229.
- Mateo, R. G., Vanderpoorten, A., Munoz, J., Laenen, B., & Desamore, A. (2013). Modeling species distributions from heterogeneous data for the biogeographic regionalization of the European bryophyte flora. *PLoS One*, 8(2), e55648.
- Morrone, J. J. (2001). Homology, biogeography and areas of endemism. *Diversity and Distributions*, 7, 297–300.
- Morrone, J. J. (2018). The spectre of biogeographical regionalization. *Journal of Biogeography*, 45(2), 282–288.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *Bioscience*, 51(11), 933–938.
- Phillips, S. J., & Dudik, M. (2008). Modeling of species distributions with maxent: New extensions and a comprehensive evaluation. *Ecography*, 31(2), 161–175.
- Proches, S., & Ramdhani, S. (2012). The world's zoogeographical regions confirmed by cross-taxon analyses. *Bioscience*, 62(3), 260–270.
- Razavi, T., & Coulibaly, P. (2013). Classification of Ontario watersheds based on physical attributes and streamflow series. *Journal of Hydrology*, 493, 81–94.
- Rickbeil, G. J. M., Coops, N. C., Andrew, M. E., Bolton, D. K., Mahony, N., & Nelson, T. A. (2014). Assessing conservation regionalization schemes: Employing a beta diversity metric to test the environmental surrogacy approach. *Diversity and Distributions*, 20(5), 503–514.
- Rodrigues, P., Figueira, R., Vaz Pinto, P., Araújo, M. B., & Beja, P. (2015). A

- biogeographical regionalization of Angolan mammals. *Mammal Review*, 45(2), 103–116.
- Romeiras, M. M., Figueira, R., Duarte, M. C., Beja, P., & Darbyshire, I. (2014). Documenting biogeographical patterns of African timber species using herbarium records: A conservation perspective based on native trees from Angola. *PLoS One*, 9(7), e0103403.
- Rubner, Y., Tomasi, C., & Guibas, L. J. (1998). A metric for distributions with applications to image databases. *Computer vision, 1998. Sixth international conference on* (pp. 59–66). IEEE.
- Rueda, M., Rodriguez, M. A., & Hawkins, B. A. (2010). Towards a biogeographic regionalization of the European biota. *Journal of Biogeography*, 37(11), 2067–2076.
- Salvador, S., & Chan, P. (2004). Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. *16th IEEE international conference on tools with artificial intelligence* (pp. 576–584). IEEE.
- Schulman, L., Toivonen, T., & Ruokolainen, K. (2007). Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. *Journal of Biogeography*, 34(8), 1388–1399.
- Urbanek, S., & Rubner, Y. (2012). emdist: Earth mover's distance. Available at: <https://cran.r-project.org/web/packages/emdist/index.html>, Accessed date: 24 May 2018.
- Vale, M. M., & Jenkins, C. N. (2012). Across-taxa incongruence in patterns of collecting bias. *Journal of Biogeography*, 39(9), 1744–1748.
- Van den Berg, M., Vandenberghe, S., De Baets, B., & Verhoest, N. (2011). Copula-based downscaling of spatial rainfall: A proof of concept. *Hydrology and Earth System Sciences*, 15(5), 1445.
- Vasconcelos, T. S., Prado, V. H. M., da Silva, F. R., & Haddad, C. F. B. (2014). Biogeographic distribution patterns and their correlates in the diverse frog fauna of the Atlantic forest hotspot. *PLoS One*, 9(8), e0104130.
- Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, 6, e7848.
- Wallace, A. R. (1876). *The geographical distribution of animals: With a study of the relations of living and extinct faunas as elucidating the past changes of the earth's surface*. London: Macmillan.
- Wei, T., & Simko, V. (2016). corrplot: Visualization of a correlation matrix. Available at: <https://cran.r-project.org/web/packages/corrplot/index.html>, Accessed date: 24 May 2018.
- Whittaker, R. J., Riddle, B. R., Hawkins, B. A., & Ladle, R. J. (2013). The geographical distribution of life and the problem of regionalization: 100 years after Alfred Russel Wallace. *Journal of Biogeography*, 40(12), 2209–2214.
- Williams, P. H., de Klerk, H. M., & Crowe, T. M. (1999). Interpreting biogeographical boundaries among Afrotropical birds: Spatial patterns in richness gradients and species replacement. *Journal of Biogeography*, 26(3), 459–474.
- Wohlgemuth, T. (1996). A biogeographical division of Switzerland based on a floristic approach. *Botanica Helvetica*, 106(2), 227–254.
- Xu, C., Sheng, S., Chi, T., Yang, X. J., An, S. Q., & Liu, M. S. (2014). Developing a quantitative landscape regionalization framework integrating driving factors and response attributes of landscapes. *Landscape and Ecological Engineering*, 10(2), 295–307.
- Xu, E. Q., & Zhang, H. Q. (2013). Spatially-explicit sensitivity analysis for land suitability evaluation. *Applied Geography*, 45, 1–9.
- Zaniewski, A. E., Lehmann, A., & Overton, J. M. C. (2002). Predicting species spatial distributions using presence-only data: A case study of native New Zealand ferns. *Ecological Modelling*, 157(2–3), 261–280.
- Zhang, Q. F., Wu, F. Q., Wang, L., Yuan, L. F., & Zhao, L. S. (2011). Application of PCA integrated with CA and GIS in eco-economic regionalization of Chinese Loess Plateau. *Ecological Economics*, 70(6), 1051–1056.