Received: 2 February 2022 | Accepted: 5 April 2022

DOI: 10.1111/2041-210X.13869

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Selection of sampling sites for biodiversity inventory: Effects of environmental and geographical considerations

Claudia Nuñez-Penichet[1] | Marlon E. Cobos[1] | Jorge Soberón[1] | Tomer Gueta[2] | Narayani Barve[3] | Vijay Barve[3,4] | Adolfo G. Navarro-Sigüenza[5] | A. Townsend Peterson[1]

[1]Biodiversity Institute and Department of Ecology & Evolutionary Biology, University of Kansas, Lawrence, KS, USA

[2]Faculty of Civil and Environmental Engineering, Technion Israel Institute of Technology, Haifa, Israel

[3]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

[4]Department of Entomology, Purdue University, West Lafayette, IN, USA

[5]Museo de Zoología, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico

Correspondence
Marlon E. Cobos
Email: manubio13@gmail.com

Funding information
2020 Google Summer of Code program; Deal Scholarship Fund, Entomology Endowment Committee and Department of Ecology and Evolutionary Biology, University of Kansas; John M. Deal Entomology Scholarship Fund, Entomology Endowment Committee and Department of Ecology and Evolutionary Biology, University of Kansas; National Science Foundation, Grant/Award Number: IIA-1920946; R. H. Beamer Scholarship Fund, Entomology Endowment Committee and Department of Ecology and Evolutionary Biology, University of Kansas

Handling Editor: Nick Isaac

Abstract

1. Biodiversity inventory is among the major challenges for conservation biology in the face of global change. Species exist in two spaces that are linked in the so-called Hutchinsonian Duality: distributions in geographical space and ecological niches in environmental space. We explore implications of using distinct methods to select locations for biodiversity inventories, based on this idea of two-space distributions.
2. We combined empirical and statistical methods to facilitate selecting localities for biodiversity inventory based on either or both of geographical and environmental considerations. These approaches were applied to select sites for inventory in four example countries. For one of our examples, we tested how effective distinct methods were in sampling biodiversity.
3. Random and geographically uniform selections are generally biased towards the most common environments in the regions; selections aiming for uniform sampling of environments are concentrated spatially in areas of high heterogeneity in geographical context. Considering disparate geographical distributions of environments helped to cover geographical areas more broadly when selections were environmentally uniform. Generally, sets of sites selected considering environmental conditions perform better in sampling known biodiversity in regions of interest.
4. Our results underline the benefits of considering environmental and geographical conditions when selecting sites on the effectiveness of resulting inventories. Our tools, implemented in the R package BIOSURVEY, will help researchers to design biodiversity survey systems taking into account the Hutchinsonian Duality and the crucial considerations that it suggests.

KEYWORDS
biodiversity indices, environmental conditions, geographical distance, Hutchinsonian duality, sampling bias, species diversity

1 | INTRODUCTION

Biodiversity inventory represents an important task in conservation biology, given increasing threats related to global change processes (Margules & Pressey, 2000; Sarkar et al., 2006). Although biodiversity can be estimated across multiple dimensions, knowledge of where each species is distributed, and which species occur together in a region (taxon diversity; Cardoso et al., 2014) is crucial to the design and implementation of effective conservation strategies (Conroy et al., 2011; Hurlbert & Jetz, 2007). However, inventorying biodiversity is a challenging undertaking, requiring significant resources, such that efficient and optimal design of survey and inventory efforts is particularly important (Colwell et al., 1994; Peterson & Slade, 1998).

To date, most efforts aimed at selection of areas for survey and inventory efforts have been based on proxies and approximations (Hill et al., 2005); only a few efforts have been made to optimize sampling in both geographical and environmental spaces (D'Antraccoli et al., 2020; Funk et al., 2005; Hortal & Lobo, 2005; Medina et al., 2013; Velásquez-Tibatá, 2019). As a result, most biodiversity patterns derived from inventories include different types of biases (Oliveira et al., 2016; Sastre & Lobo, 2009; Yang et al., 2013), which could be prevented if more comprehensive considerations are taken when planning systems for inventory. Biodiversity inventory is challenging because it requires significant resources (Balmford & Gaston, 1999), not to mention time and effort that are often unavailable. Inventory strategies need to consider available resources and logistics, but also geographical and environmental conditions across the region of interest (Morrison et al., 2008). Geographical conditions are commonly considered when planning these strategies because considerations of distance, accessibility and survey coverage are evident when seeing geographical representations of the areas to be sampled. Environmental conditions, however, are less visible, and too often are neglected when planning biodiversity inventory efforts (Hortal et al., 2015). This focus on geography over environment is nonetheless in contrast to most results from the field of distributional ecology, in which ecological niches drive major features of species' presences across a region (Soberón & Peterson, 2005). The complex relationships between geographical locations and environmental conditions underlie the Hutchinsonian Duality (Colwell & Rangel, 2009), and are therefore of critical importance when planning for systems for biodiversity survey.

Defining which areas to sample to optimize survey and inventory efforts is crucial to detecting and documenting more species with less effort and expense, thereby obtaining a more complete list of the species across a region (Soberón & Llorente, 1993). This efficiency is paramount in biodiversity inventory endeavours (Eckblad, 1991; Gotelli & Colwell, 2001), and considering both geographical and environmental spaces could certainly improve the efficacy of these efforts (Hirzel & Guisan, 2002). An early implementation of these ideas on biodiversity survey was presented by Austin and Heyligers (1989), who proposed a method by which to select sampling transects considering classes of environmental conditions across an area. Hortal and Lobo (2005) proposed another approach using a rule-step site-allocation procedure, based partially on Faith and Walker's 'ED' criterion (a framework linking species data and environmental information to explore underlying environmental variation related to a biological pattern; Faith, 2003; Faith & Walker, 1996). Using similar considerations, Funk et al. (2005) employed a method to complement survey systems by selecting sampling localities based on a survey-gap analysis (see also Medina et al., 2013). These methods require certain knowledge of the biodiversity in the region such that application in areas where biodiversity data are scarce could be difficult. More recently, D'Antraccoli et al. (2020) proposed an approach that combines considerations of geographical and environmental distances to select areas for sampling based on considerations of both dimensions. Although no previous knowledge of sampling in the area was required to select sampling localities, the authors demonstrated that the considerations made effectively led to more species being sampled.

Here, we present a review of conceptual frameworks, and from them derive new methodological approaches to design survey and inventory efforts. Our methods are designed to require simple input data to select sites efficiently for biodiversity surveys; previous knowledge of sampling effort in the region of interest can be used, but is not required. Four approaches to site selection are explored, represented and tested, one of which is designed specifically to consider the duality of geographical and environmental spaces across the region of interest. To allow researchers to apply the approaches presented here to any region of interest, we developed software tools and have made them available in the R package 'BIOSURVEY' (Nuñez-Penichet et al., 2021a).

2 | MATERIALS AND METHODS

2.1 | General description

We selected four contrasting countries (Philippines, Mexico, Rwanda and Uruguay, roughly in order of decreasing geographical and environmental heterogeneity) to explore different survey site-selection approaches. We used geographical and environmental information derived from spatial polygons and raster environmental layers. Given availability of high-quality distributional data, we used Mexico as our primary example with which to test the efficacy of our approaches, and to illustrate further analyses that consider reduced areas in the region of interest (e.g. areas with primary habitat), and use of localities selected a priori (e.g. existing well-surveyed sites) in algorithms for site selection.

All analyses described (except for some spatial processes described below) were performed in R 4.0.5 (R Core Team, 2021). Data, code and guidelines to reproduce all analyses and plots are available at https://doi.org/10.6084/m9.figshare.14700819.

2.2 | Data

We used polygons summarizing the spatial boundaries of Mexico, Philippines, Rwanda and Uruguay, obtained from the Natural Earth database (https://www.naturalearthdata.com/). For Mexico, we excluded the two westernmost islands (Clarion Island, Socorro Island) to match the area across which information on species' distributions was available. We also used a mask layer summarizing areas of natural vegetation cover in Mexico, obtained by selecting categories corresponding to natural land-cover types from the layer of land use and vegetation (INEGI, 2016), obtained from the geodata portal of the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO; http://www.conabio.gob.mx/informacion/gis/). This mask helps to exclude areas that are not relevant for analyses as they do not hold natural vegetation (note that this feature may or may not be desirable in a given analysis, depending on the goals of that analysis). To reduce computational time and considering viability and integrity of small habitat patches, we removed polygons from this mask with areas <25 km^2, and simplified the remaining polygons using the algorithm 'Bend simplify' (tolerance 5 km) in ArcMap 10.5.1. For Mexico, we also explored incorporating information on a set of six localities for which existing biodiversity inventories were already relatively complete (hereafter called 'preselected sites'; Table S1), a situation that we expect will frequently be the case in biodiversity inventory planning.

To represent environmental conditions across the regions of interest, we used the bioclimatic variables from the WorldClim database 1.4, at a spatial resolution of 2.5′ (~4.5 km; https://www.worldclim.org/data/v1.4/worldclim14.html; Hijmans et al., 2005). Variables that combine information of temperature and precipitation (i.e. mean temperature of wettest quarter, mean temperature of driest quarter, precipitation of warmest quarter and precipitation of coldest quarter) were excluded, as they are known to present spatial artefacts not corresponding to known discontinuities across geography (Escobar et al., 2014).

For analyses and tests of the efficacy of our prioritization approaches, we used expert-curated species distribution model outputs for two target groups, birds and 'herps' (amphibians and reptiles combined). These distributional summaries were derived from species distribution models (SDMs) constructed at 30″ (~1 km) spatial resolution, and curated by experts in each of the groups (Flores Villela & Ochoa Ochoa, 2010; Navarro-Sigüenza & Gordillo-Martínez, 2018). The process of curation of SDM results consisted of species-by-species inspection of each SDM output for each species. Experts in the distributions of each group of Mexican species inspected each map in concert, considering the distribution of known occurrences, as well as local topography and other geographical features. Based on these inspections, SDM outputs were edited to produce a relatively conservative view of the likely geographical distribution of each species (i.e. the equivalent of the occupied distributional area, and not the potential distributional area). This set of distributional information is considered to be authoritative, and is as close to a summary of actual distributions of species in the country as is available. Bird data were in raster format, with values of 1 (suitable) and 0 (unsuitable), whereas herp data were in GeoJSON format, with suitable areas represented as polygons. Bird data were provided by two of the authors (ATP and AGNS), who developed the datasets; herp data layers were obtained from the geodata portal of CONABIO (where the bird data are also available). We resampled bird data layers to a resolution of 2.5′ to reduce computational demands (aggregation used the modal value of cells involved). We used the modal value to assure that the resulting resampled layer will be binary, while also avoiding overestimation of the area in which the species is most likely to be present. Species of birds for which ranges include Mexico only during non-breeding (winter) periods were excluded from analysis.

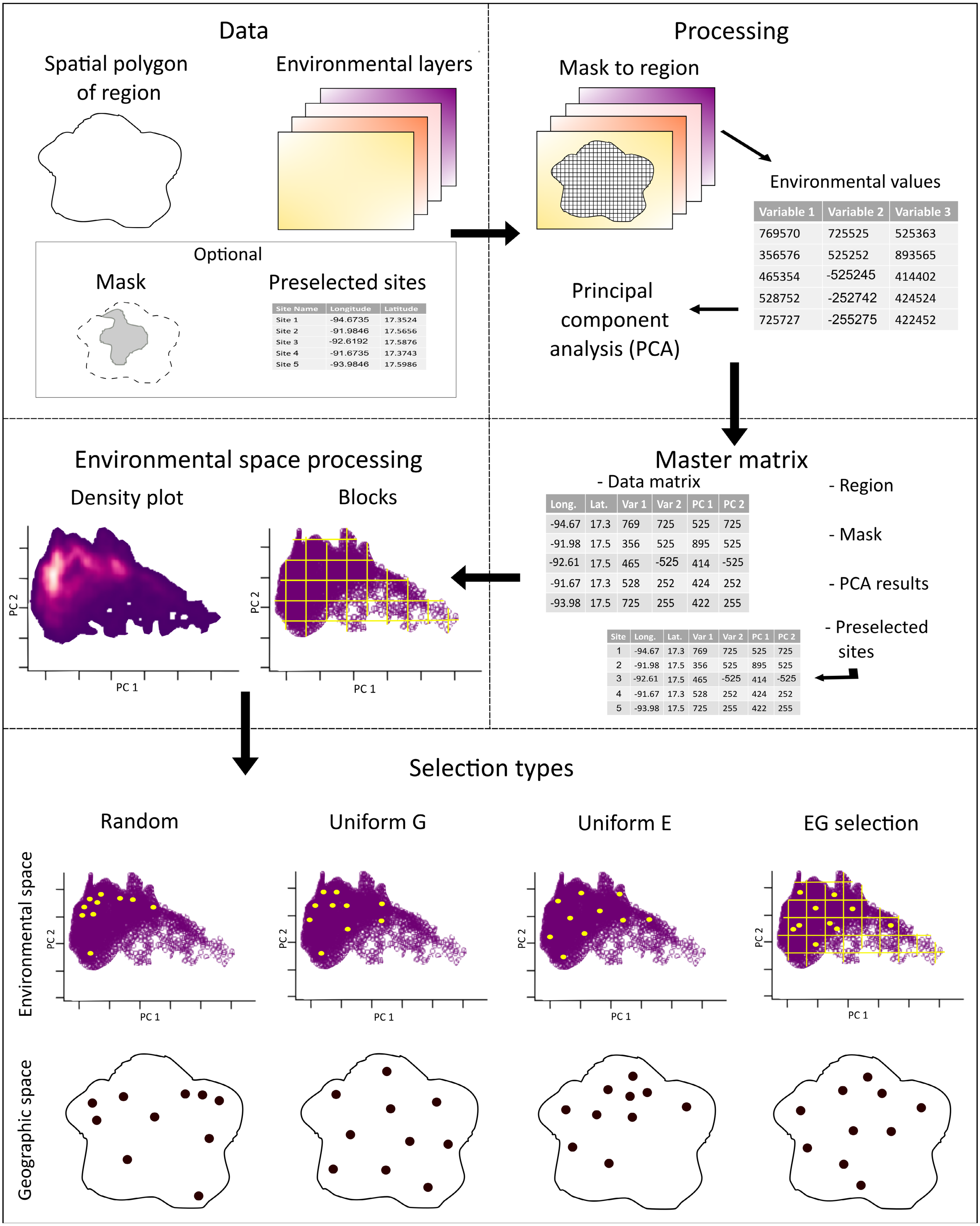
2.3 | Pre-processing

To prepare data for analyses of sampling site selection and testing (Figure 1; Figure S1), we started by masking the raster bioclimatic variables using the polygons for each of the four countries of interest. For Mexico, we also used a shapefile summarizing areas with natural vegetation to mask the raster data layers further (Figure 2). After that, a principal components analysis (PCA) was done using the values of the layers masked to each country (Figures S2–S6). Geographical coordinates derived from raster layers, the environmental values associated, and the first two principal components were put together in a single matrix (master matrix) to be used in later analyses. For Mexico, the values of the environmental layers and the first two principal components (PCs) were also extracted to the six preselected sites. Additionally, we separated the environmental space into blocks, this space defined in the two-dimensional space of the two first PCs. These blocks were delimited using a grid of equal-sized cells, aiming for 25 rows × 25 columns between the minimum and maximum of each of the two environmental dimensions (see Figures S7–S8). These blocks figure in one of the analyses in which we were selecting regions of environmental space uniformly (see Considering environment and geography in selections).

We prepared four presence–absence matrices (PAMs; Arita et al., 2011) for Mexico: two based on bird distributions and two based on herp distribution data, and with and without masking to areas of natural vegetation (Figures S1, S9 and S10). These PAMs will be used later to test effectiveness of sampling site selections using derived biodiversity indices (Soberón et al., 2021; Soberón & Cavner, 2015).

2.4 | Selection of sites for biodiversity surveys

We used distinct approaches for sampling site selection to explore implications of stratification in geographical (G) and environmental (E) dimensions in biodiversity inventory design (Figure 1). G comprises the spatial arrangement of coordinates derived from the



Site Name	Longitude	Latitude
Site 1	-94.6735	17.3524
Site 2	-91.9846	17.5656
Site 3	-92.6192	17.5876
Site 4	-91.6735	17.3743
Site 5	-93.9846	17.5986

Variable 1	Variable 2	Variable 3
769570	725525	525363
356576	525252	893565
465354	-525245	414402
528752	-252742	424524
725727	-255275	422452

Long.	Lat.	Var 1	Var 2	PC 1	PC 2
-94.67	17.3	769	725	525	725
-91.98	17.5	356	525	895	525
-92.61	17.5	465	-525	414	-525
-91.67	17.3	528	252	424	252
-93.98	17.5	725	255	422	255

Site	Long.	Lat.	Var 1	Var 2	PC 1	PC 2
1	-94.67	17.3	769	725	525	725
2	-91.98	17.5	356	525	895	525
3	-92.61	17.5	465	-525	414	-525
4	-91.67	17.3	528	252	424	252
5	-93.98	17.5	725	255	422	255

FIGURE 1 Schematic workflow of the analysis executed to prepare data and select sampling sites using four approaches. E = environmental space; G = geographical space; EG = combined environmental and geographical spaces

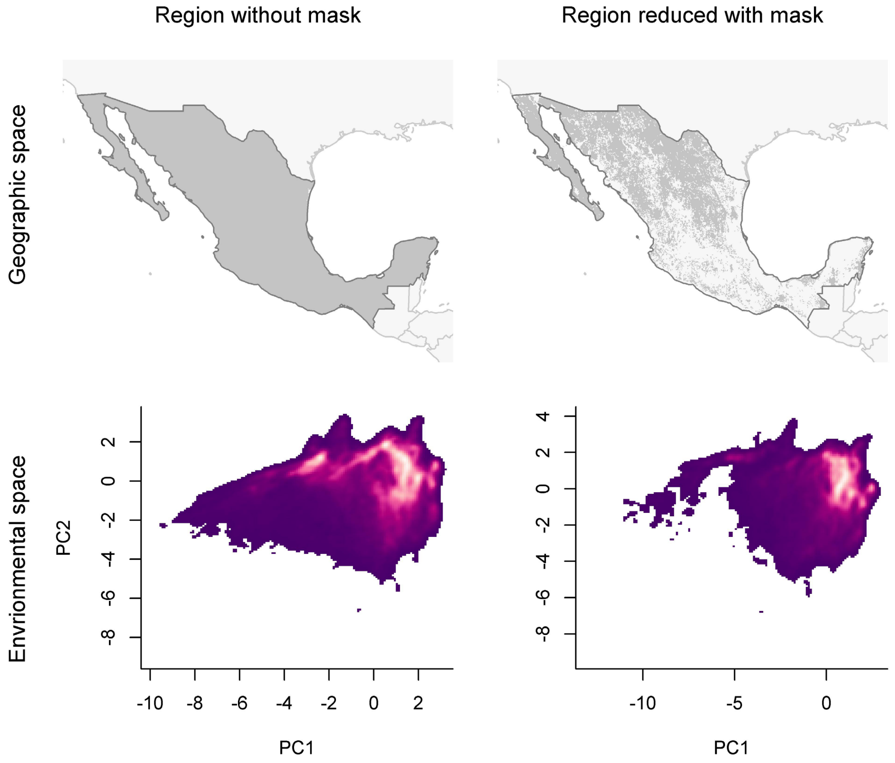


FIGURE 2 Representation of Mexico in geographical (top) and environmental spaces (bottom), for the full extent of the country (left) and the country masked to places presenting natural vegetation (right). The first two principal components (PC) of layers representing bioclimatic conditions in Mexico are used to visualize two major dimensions of the full environmental space

raster layers reduced to the region of interest. E is represented by the two PCs deriving from the PCA performed on the bioclimatic variables. Our four approaches for site selection were as follows: (a) a spatially random selection of points across the area of interest; (b) a selection of sites aiming to achieve uniform coverage of the geographical region of interest; (c) a selection of sites aiming to achieve uniformity in a two-dimensional environmental space represented by the two PCs; and (d) a selection to achieve uniformity in environmental space, but accounting for geographical clustering of environmental regions.

All four approaches to selecting sampling sites were applied to the four countries; these processes were applied four times to Mexico, to explore implications of masking to natural-vegetation areas and of inclusion of a priori well-inventoried sites. In all, in our explorations, 30 sampling sites were selected in Mexico, 24 in the Philippines and Uruguay, and 20 in Rwanda. The various procedures that we employed to explore these methods are described in greater detail in the paragraphs that follow.

2.4.1 | Random selection

Random selection of sites can be achieved easily by picking randomly from the entire set of points available in the region. Considering the geographical context, under this approach, every point in geography has the same probability to be selected. However, seen from an environmental point of view, classes of environmental conditions (e.g. dry or wet, or cold or warm) that are more common in the region of interest will be selected with higher probability, and rarer sets of conditions will often be left out of the sampling plan (Figure 1). We performed these analyses by selecting 100 sets of random points and filtering them to keep the set that has the maximum median geographical distance (MMGD) among points. This step means that the 'random' points are selected with an aim for spatial overdispersion; however, this filter is not likely to derive in sites that cover the region as uniformly as those from the method that aims explicitly for uniformity. For Mexico, an extra set of analyses assessed all 100 sets, to understand the variability that can arise from this type of selection as regards correspondence to known distributions of species (birds, herps). A random selection of one of the 100 sets of points from this last example would be an alternative way to keep a purely random set of selected sites.

2.4.2 | Uniformity in geographical space

In this approach, we selected points such that they were overdispersed across geography, covering the region of interest as evenly as is possible (Figure 1). We used all available points in the master matrix as a base, and thinned the mass of points with increasing geographical distances until we obtained the desired number of points (see Thinning process for details). Geographical distances were measured after projecting the points with an azimuthal equidistant projection centred on the centroid of the region of interest. This process of projection makes the values measured approximate geographical distances, and potential bias increases with distance from the centroid. Because the order in which points are selected affects the final set of points, replicate analyses result in different sets of points. We performed 10 replicates, and used the one that had the MMGD among points.

2.4.3 | Uniformity in environmental space

The principle of this type of selection is similar to the previous one, in the sense that points are selected based on distances and a thinning routine. However, here, the goal of this sampling is to select points evenly considering the environmental conditions present across the region of interest (Figure 1). Euclidean distances are measured in a space represented by two environmental axes, in this case, the two first PCs obtained from bioclimatic layers. Again, this analysis can be produced with replicates that result in distinct outcomes, so we used 10 replicates and selected the set with the MMGD among points.

2.4.4 | Considering environment and geography in selections

To perform a selection that combines considerations of environmental and geographical characteristics in the region of interest (termed 'EG selections'), we used a multistep procedure. First, we selected a predefined number of blocks (see Pre-processing) configured to maximize uniformity in environmental space (see Thinning processing). Once environmental-space blocks were selected, the geographical pattern of all points falling in each environmental-space block was explored to detect whether these points were grouped in one or more geographical-space clusters. To this end, we measured geographical distances among a random sample of the points in the block; whenever multimodality was detected, based on a unimodality test (Hartigan, 1985), a clustered pattern was assumed. Clusters were then hierarchically assigned based on the distance between largest modes in the distribution of all geographical distances. For blocks with clustered geographical patterns, the two largest clusters (i.e. those including more points) were identified, and one point was selected from each (Figures 1; Figure S11)—we selected points from each cluster as those closest to the centroid of each group of points in environmental space. This process was repeated for all blocks selected. The final number of sites selected can, therefore, be larger than the initial number of blocks defined, if geographical clusters are numerous within environmental-space blocks. The reasoning behind using more than one point per environmental block, if several clusters are detected, is that similar environments can be found in distant, disjunct areas, which will often host distinct biotic communities. The sets of blocks selected at the beginning of this approach can be different if the process is replicated. We used 10 replicates, and selected the replicate with the set of points that had the MMGD in environmental space among points.

2.4.5 | Thinning process

This process is performed in our three methods aiming to create sets of overdispersed points in geographical or environmental spaces. The cloud of points (all of them for uniformity in E or G, or only block centroids in EG selections) is explored to identify which points are too close given a threshold distance. Once groups of too-close points are identified, only one point of each group is retained. The number of points remaining is counted; if there are more points than is needed, the threshold distance is increased; otherwise, it is decreased. The value to be added to or subtracted from the threshold distance is adjusted if at a certain point it is not possible to reach the number of points needed after thinning. These processes are repeated until the desired number of points is reached. To find points that are closer than the threshold distance, we used the R package SPATSTAT.GEOM (Baddeley et al., 2016; Baddeley & Turner, 2005). As multiple distances need to be explored to obtain the number of points required, the analyses are performed in a conditional loop. As every time this algorithm is run, the points used to start measuring distances can change, multiple answers can be obtained if the entire process is repeated. We programmed this routine such that replicates can be performed when selecting points for sampling.

2.4.6 | Using preselected sites in selections

All for approaches to site selection were described in the paragraphs above as if preselected sites were not considered. However, in many or most regions, valuable information derived from previous sampling efforts exists already for some areas. Such information may be complete enough that researchers would wish to include those areas in the set of sites selected, to take advantage of the already-existing information.

To include preselected sites when points are selected randomly, we randomly choose a number of points equal to the total number required minus the number of preselected sites. After that, random sites and preselected sites were combined. When sampling sites were selected to achieve uniformity in geographical or environmental spaces, all points closer than a certain distance from the selected sites (in the corresponding space) were excluded before the process of selection began. Exclusion of points was done based on environmental blocks when environment and geography were considered together in selections. Excluding points around preselected sites guarantees that all sites selected using the approaches described above meet the requirement that they be distributed uniformly, but also are distanced enough from preselected sites to maintain that uniformity.

However, as preselected sites must be included and do not necessarily follow the distance criterion, they may or may not be as uniformly distributed as if they were not included. The distance used to exclude points that are too close to preselected sites is selected using a multistep thinning process that results in the desired total number of points when filtering all points in the region of interest.

2.5 | Testing effectiveness of selections

We emphasize that testing the effectiveness of our site-selection approaches can be done only in the relatively rare cases in which reliable information about species' geographical distributions is available. In this case, we compared the completeness of inventories among distinct sets of sampling sites selected, based on the near-unique distributional summaries available for all birds and herps in Mexico. To perform these comparisons, we associated the sites selected to the cells of the PAMs that overlapped geographically. This step allowed us to obtain subsets of the PAMs for each set of sites selected.

Using this information, we created plots of pairwise comparisons of species accumulation curves (Soberón & Llorente, 1993) derived from the subsets of the PAMs corresponding to each of the sets of sites selected. The number of species and shape of these curves were inspected to understand which site-selection approaches resulted in more complete inventories. We also calculated dissimilarities (Jaccard indices; Faith et al., 1987) among communities of species sampled using distinct sets of sites to explore patterns among results from distinct site-selection approaches. All of these analyses are presented for the example of birds and herps from Mexico, with and without masking to natural areas and with and without preselected sites.

3 | RESULTS

3.1 | Initial results

Using the two first principal components allowed us to summarize the variance from environmental conditions across the regions of interest (Mexico 74% with no mask, 72% with mask; Philippines 67%; Rwanda 94%; and Uruguay 77%). For Mexico, the two-dimensional cloud of environmental conditions had an oblong shape, with density focused at higher values of both principal components (Figure 2). This shape shifted slightly when Mexico was masked to areas with native vegetation (Figure 2). The other three countries also had odd shapes, with gaps, infoldings and strings of outlying points (Figures S2–S6). Distinct numbers of environmental blocks were obtained despite initial grids having the same number of rows and columns, owing to lack of representation of environments in some of the initial grid cells (Mexico: no mask 311, with mask 293; Philippines 298; Rwanda 160; Uruguay 222; Figures S7–S8).

The PAMs created for the whole of Mexico had 3448 and 3038 cells for the country masked to natural vegetation, respectively. Highest bird species richness was in the southeastern parts of the country (lowlands to medium elevations), whereas highest herp richness was associated with tropical montane areas of the country (Figures S9 and S10). Maximum values of species richness for birds were 491 (no mask) and 489 (with mask), whereas values of 207 (no mask) and 203 (with mask) were found for herps.

3.2 | Selected sites

Each of the approaches resulted in different sets of sites selected, although relatively similar distributions of points were observed only in environmental space for random selections and geographical uniformity. Some similarity of selections was also noted between approaches aiming for environmental uniformity and the EG selections that consider both spaces (Figures 3 and 4; Figures S12–S17). In general, all random selections were biased towards the most common environments across the region (e.g. compare the upper-left panel of Figure 3 with the lower panels in Figure 2), although the geographical position of points was not biased towards any particular area. Sites selected aiming for uniformity in geography showed the desired geographical pattern in all examples, but were biased towards the environments most common in the region.

Selections aiming for uniformity in environmental space covered the set of environments present in the regions of interest uniformly. However, this type of selection turned out to be rather biased in geographical space. That is, records were highly clumped geographically, in areas of high environmental heterogeneity. The pattern of sites selected with EG selections resembled that of sites selected aiming for uniformity in environmental space, although some areas of this space were represented by two points instead of just one. In geography, EG selections still looked somewhat biased towards highly environmentally heterogeneous areas, but more broadly distributed compared to uniformity in environmental space (Figure 3; Figures S12–S17).

3.3 | Effectiveness of selected sites

In general, bird species were represented in our selections more completely and efficiently than herps (Figures 5; Figures S18 and S19). Relating our site-selection results to observed distributional patterns of species for Mexico, random selections of points generally showed the worst performance in terms of total number of species sampled and how efficiently the species were sampled. That is, random selections consistently required sampling more sites to recover similar numbers of species than the other approaches (Figure 5; Figures S18 and S19). Sites selected seeking for uniformity in geography performed slightly better than random selections in terms of effectiveness. EG selections and selections aiming for uniformity in environmental space generally showed the best performance in our tests, for both birds and herps.

We assessed the effects of different individual random selections of sites, such that the 100 random sets of sites were compared with results of other selection approaches. These analyses showed that only a few of the random-site sets performed comparably to the other approaches, most again performing poorly, with fewer species represented and more samples required (Figure S20). However, selecting such best-performing random selections is not feasible if high-quality testing data are not available, which is the case for most situations in which these methods will be applied. Considering community dissimilarities, sites selected randomly and those aiming for uniformity in G were usually more similar to each other than to sites selected with partial or full consideration of representation of environments (Figures S21 and S22).

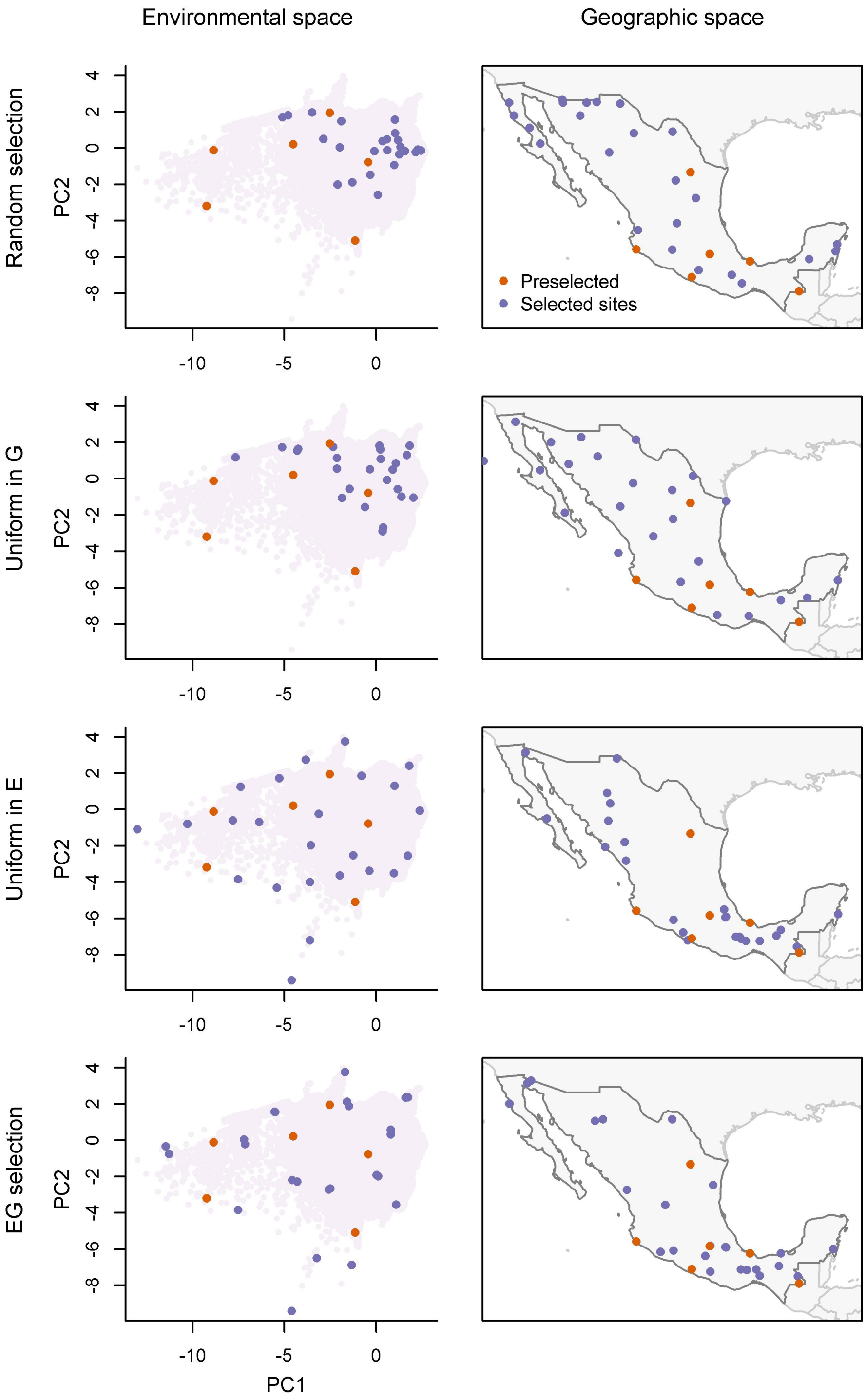


FIGURE 3 Representation of 30 sampling sites selected based on four different approaches for sites across Mexico. Sampling localities were selected considering the mask to natural vegetation cover for the country, and with sites selected a priori based on their importance for biodiversity monitoring. E = environmental space; G = geographical space; EG = environmental and geographical spaces; PC = principal component

3.4 | Effects of mask and preselected sites

The use of a mask to restrict our analyses in Mexico reduced the geographical area over which analyses were performed considerably (by 48%), which also translated into a reduction of the environmental space under analysis (Figure 2). This change had two major effects: (a) computational time required for execution of the site-selection analyses decreased when using the mask. Perhaps more importantly, (b) the sets of points available for selection were restricted to more relevant areas for at least some biodiversity surveys (Figure 3; Figures S12 and S14). The effect of using a mask in terms of efficiency in representing species could be observed in all of our results; this effect is more clearly noticeable when comparing species curves obtained with random selections and other methods with or without the mask (Figure 5; Figures S18 and S19). Regarding the effect of masking the region of interest on the PAMs obtained for the two taxa, cells with maximum values of richness were affected slightly, but the general pattern was similar (Figures S9 and S10).

Preselected sites affected results of selections positively, at least in the Mexican example that we explored in detail (i.e. if those sites were

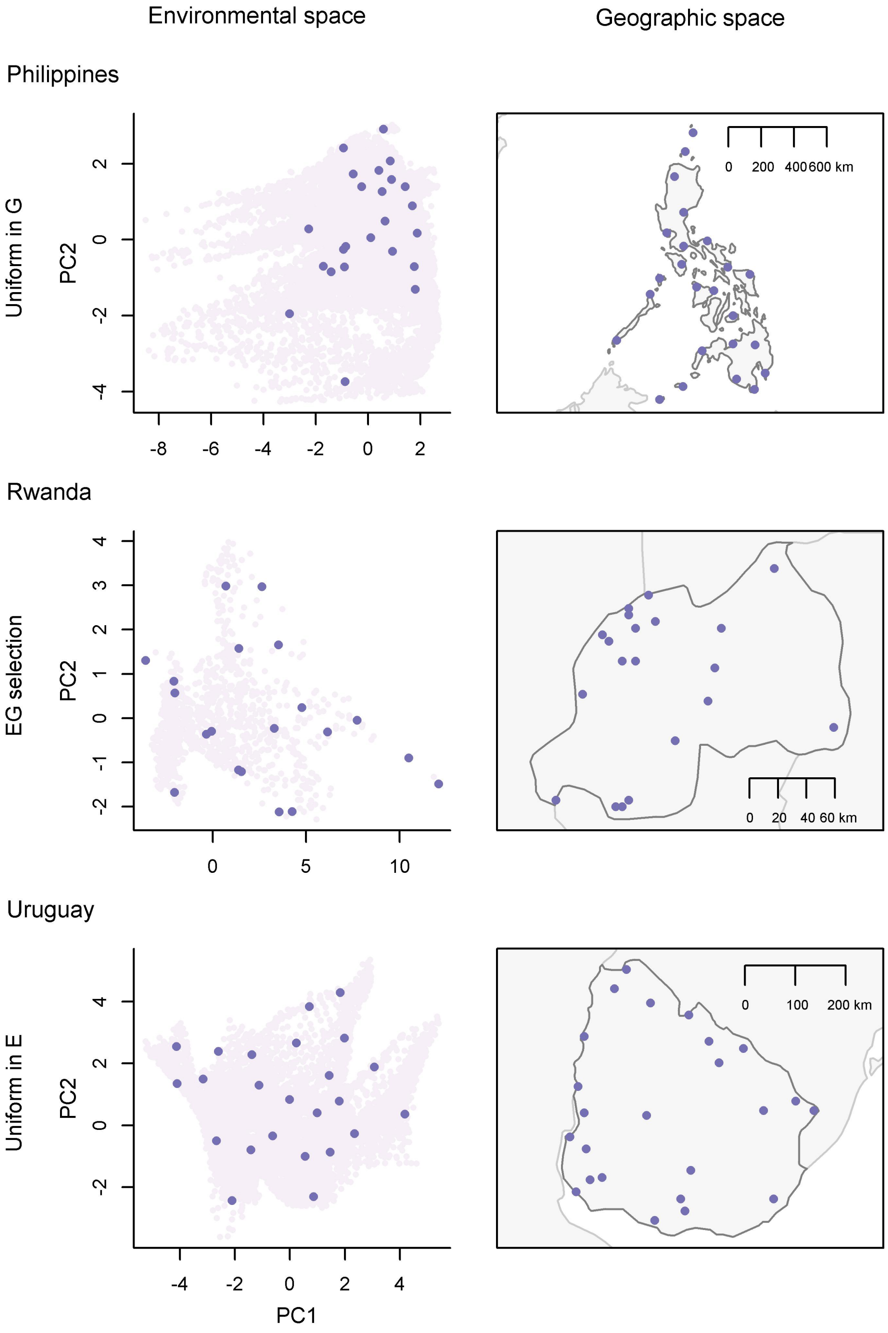


FIGURE 4 Examples of sampling sites selected for Philippines (*N* = 24), Rwanda (*N* = 20) and Uruguay (*N* = 24) represented in environmental (left) and geographical (right) spaces. Note that distinct methods for site selection are shown for each country. Purple dots represent sites selected; E = environmental space; G = geographical space; EG = environmental and geographical spaces; PC = principal component. See other results in Figures S15–S17

highly similar to one another, it could affect selection efficiency negatively). When selections did not consider environmental conditions, sites selected a priori extended the coverage of environmental space; when geography did not play a major role in selections, preselected sites allowed site-selection results to cover some areas that were not considered otherwise (Figure 3; Figures S12 and S14). Results viewed in terms of effectiveness were perhaps where the effects of preselection of sites could be seen most clearly. The effects of using preselected sites or not in random selections were clear despite lower performance in general compared to other approaches (Figures 5; Figures S18 and S19). The use of preselected sites greatly improved the ability of random selections to sample more species with fewer sampling sites, especially for amphibians and reptiles (Figure 5), again at least in the Mexican example.

4 | DISCUSSION

All four site-selection approaches explored in this study resulted in sets of selected sites that met our design expectations in terms of distribution in geographical and/or environmental spaces. Although our maximum median geographical distance (MMGD) approach optimized all of our selections, differences arising from using distinct approaches were evident. Perhaps the major takeaway lesson is that explicit exploration of existing environmental conditions can give a better idea of the challenge of designing a system of sites that allows efficient biodiversity inventory and monitoring (D'Antraccoli et al., 2020; Hortal & Lobo, 2005; Velásquez-Tibatá, 2019).

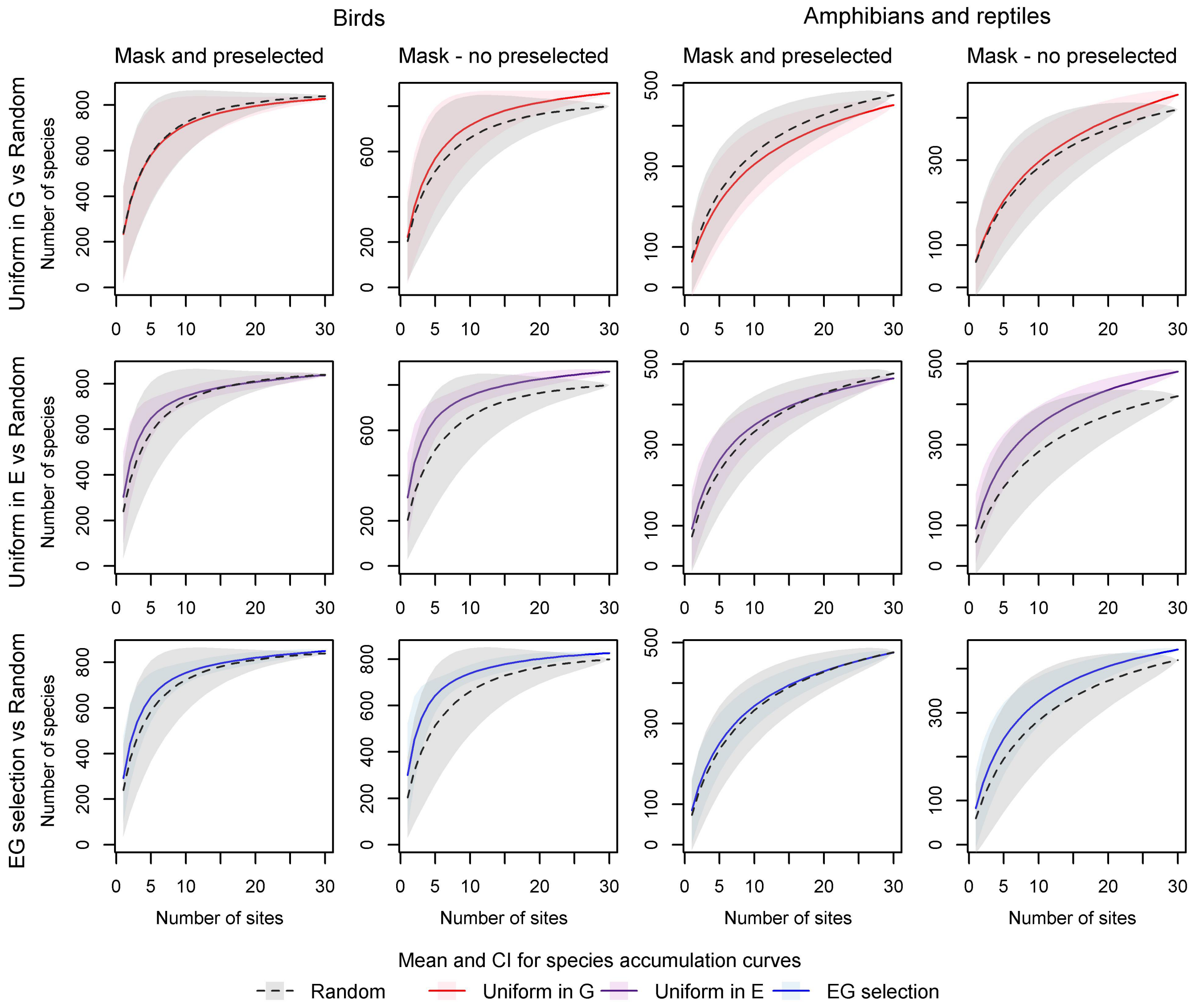


FIGURE 5 Comparison of species accumulation curves derived from presences and absences of bird and amphibian and reptile species in sampling sites selected using four distinct approaches for Mexico masked to areas holding natural vegetation. Results with and without preselected sites are shown. E = environmental space; G = geographical space; EG = environmental and geographical spaces; CI = confidence intervals

This observation is precisely why bringing ideas from the Hutchinsonian Duality (Colwell & Rangel, 2009) into consideration is helpful in site-selection exercises. Because species' geographical distributions and consequent biodiversity patterns are related closely to geographical configurations of environmental conditions, considering such conditions in planning biodiversity surveys is logical. However, geographical patterns must not be ignored, as biogeographical barriers may also be important drivers of patterns of species' distributions in geographically complex areas (e.g. the Philippines). Although we do not see a scenario under which random selection of survey sites would be preferable over other approaches, researchers should explore options and make decisions based on their knowledge of the region of interest, the variables used to represent environmental conditions, and the biology and biogeography of the taxa of interest.

Although tests of site-selection effectiveness were performed only for Mexico, for lack of detailed species distributional information in other regions, they illustrated how well different site selections would be able to sample regional biodiversity. Some approaches performed better than others, but again, choice of selection approach should not be based on this example, but rather on the specific conditions of each region and taxon under study. Our intuition is that biogeographically complex regions (e.g. Philippines) will be sampled better by site-selection approaches that consider geography (perhaps in tandem with environments), but that less complex regions (e.g. Uruguay, Rwanda) will emphasize the importance that environmental variation is considered. In cases in which high-quality information is available about species' distributions, the tests applied in the example of Mexico offer a quantitative way to choose an approach. This type of data can also be used to explore the effects of using more or fewer points in the set of sampling sites to be selected. Our results of effectiveness show that, depending on the taxon of interest, using more points may be necessary (see herp results; Figure 5; Figure S19), although this parameter in the planning of biodiversity inventory efforts may be restricted by resources available.

We also found that the use of preselected sites, corresponding to sites already well inventoried, and a mask to restrict analysis to sites that are interesting and/or accessible, has positive effects on the effectiveness of the set of selected sites. These ideas have been explored in previous studies (Funk et al., 2005; Hortal & Lobo, 2005; Medina et al., 2013), in which definition of areas suitable or unsuitable for surveys, and use of existing information from previous surveys play critical roles in implementation of methods for sampling site selection (see also Gillespie et al., 2017; Hoffmann et al., 2019; Tessarolo et al., 2021; Xu et al., 2017). In the example presented, our mask was used to focus analyses in areas with natural vegetation in Mexico. However, many other factors that could limit surveys or favour selections can be considered when preparing a mask, for instance, considering accessibility (e.g. distance to roads) and excluding developed areas or other sites that are not relevant in sampling the taxon of interest.

Preselected sites are perforce included as part of the final set of sites obtained from any of the site-selection approaches; for this reason, they alter selection of additional sites. An a priori selection of sites to be included as part of final sets of sampling sites is, therefore, an important task and needs to be done based on appropriate considerations. Depending on the taxon of interest and availability of data, researchers could benefit greatly from this option to improve how sites are selected. Final selections will consider not only the environmental and/or geographical conditions across the region of interest, but also another filter that relates to the existing knowledge of biodiversity in the area (Peterson et al., 2016).

The idea of considering both geographical and environmental dimensions in sampling site selection has been explored previously in the development of tools that facilitate this task. The main considerations explored in our approaches are shared among some of these previous proposals, especially those of Hortal and Lobo (2005), Funk et al. (2005) and Medina et al. (2013). That is, a region of interest is explored in both relevant spaces, areas that are relevant for exploration are delimited, and previous survey data are used to understand where the survey gaps are located. Consideration of knowledge derived from previous sampling efforts is also present in a more recent development by Velásquez-Tibatá (2019), in which this information is combined with environmental data to identify environmental regions underrepresented in existing inventories (see also Tessarolo et al., 2021). More recently, D'Antraccoli et al. (2020) proposed an approach to search for environmental distance-optimized random points, a simple approach that also seeks for sites that could lead to better surveys. The approaches that we have explored and implemented here differ from previous approaches in various aspects. First, although we also allowed the use of sites selected a priori and a mask for the region of interest when performing sampling site selections, these inputs are not required, so our methods can be extended to areas where these data are not readily available. Second, the analyses used to select sites than sample comprehensively in environmental space or in both spaces are different; our implementations are based on thinning procedures that are faster than other statistical approaches and similarly effective. Finally, the high level of automation reached in the implementation of our tools in the R package BIOSURVEY, makes the application of the methods presented here easier for researchers interested in selecting sampling sites in different regions of the planet.

The methods explored here were applied to examples of relatively large regions, but they can be applied to smaller regions (e.g. provinces or river basins). However, we would expect that biogeographical limitations will be of reduced importance on smaller spatial extents, and environmental considerations will be increasingly relevant. Also, the environmental variables required to characterize conditions across such smaller regions should be selected accordingly (Storch et al., 2007). For instance, instead of climatic datasets that may not be relevant because they do not show major or pronounced variation across smaller regions (Peterson & Soberón, 2018), environmental data related to habitat, vegetation characteristics or substrate, often derived from remote sensing data, could be more appropriate.

One important process in performing some of the analysis is the filtering (thinning) of points based on geographical distances. As this process is executed over points converted to a geographical projection that suits distance calculations (Azimuthal Equidistant projection), a limitation may arise for very large areas. Because we summarize environmental dimensions in the region of interest using a PCA, the initial set of environmental layers should not be categorical or discrete. Considering that some environmental factors of interest may be represented in categories, this limitation may be important in some applications. Using a mask to restrict analysis to certain categories of environments may offer an option to deal with this obstacle.

We consider a sampling site as each independent unit of the set of localities for sampling selected using the approaches presented. Depending on the taxa of interest and other aspects that determine sampling effort directly (e.g. human resources, periodicity of sampling and sampling coverage), different survey (sampling) methods can be used (Cutko, 2009; Hill et al., 2005; Morrison et al., 2008). Importantly, the answers that can be obtained with these tools can be considered initial options and could be explored and refined in greater depth when defining final sets of sites for survey. For instance, as the geographical coordinates of selected sites derive from raster layers, actual geographical locations of sites could be modified to consider local characteristics of each area that facilitate accessibility and/or feasibility of sampling depending on the taxa of interest, resources and the methods to be used.

ACKNOWLEDGEMENTS

We thank the Comisión Nacional para el Uso y Conocimiento de la Biodiversidad (CONABIO), especially Raul Sierra, for providing us with the distributional data for Mexican herps. C.N.-P. was supported by the John M. Deal Entomology Scholarship Fund, the R. H. Beamer Scholarship Fund and the Deal Scholarship Fund, from the Entomology

Endowment Committee and the Department of Ecology and Evolutionary Biology, University of Kansas. The BIOSURVEY package was initially developed by CNP during the 2020 Google Summer of Code program (R Organization). C.N.-P., M.E.C. and A.T.P. were partially supported by a grant from the National Science Foundation (IIA-1920946).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

C.N.-P., M.E.C., J.S. and A.T.P. participated in method design; C.N.-P., M.E.C., J.S., T.G., N.B., V.B. and A.T.P. participated in software design; A.T.P. and A.G.N.-S. created and provided data for testing methods; C.N.-P. and M.E.C. executed the analyses. All authors participated in the manuscript preparation.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13869.

DATA AVAILABILITY STATEMENT

Data, code and a more detailed description of methods needed to reproduce all analyses are available at https://doi.org/10.6084/m9.figshare.14700819 Nuñez-Penichet et al. (2021b).

ORCID

Claudia Nuñez-Penichet https://orcid.org/0000-0001-7442-8593
Marlon E. Cobos https://orcid.org/0000-0002-2611-1767
Jorge Soberón https://orcid.org/0000-0003-2160-4148
Tomer Gueta https://orcid.org/0000-0003-1557-8596
Narayani Barve https://orcid.org/0000-0002-7893-8774
Vijay Barve https://orcid.org/0000-0002-4852-2567
Adolfo G. Navarro-Sigüenza https://orcid.org/0000-0003-2652-7719
A. Townsend Peterson https://orcid.org/0000-0003-0243-2379

REFERENCES

Arita, H. T., Christen, A., Rodríguez, P., & Soberón, J. (2011). The presence-absence matrix reloaded: The use and interpretation of range-diversity plots. *Global Ecology and Biogeography*, *21*(2), 282–292. https://doi.org/10.1111/j.1466-8238.2011.00662.x

Austin, M. P., & Heyligers, P. C. (1989). Vegetation survey design for conservation: Gradsect sampling of forests in North-Eastern New South Wales. *Biological Conservation*, *50*(1), 13–32. https://doi.org/10.1016/0006-3207(89)90003-7

Baddeley, A., Rubak, E., & Turner, R. (2016). *Spatial point patterns: Methodology and applications with R*. CRC Press.

Baddeley, A., & Turner, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, *12*(1), 1–42. https://doi.org/10.18637/jss.v012.i06

Balmford, A., & Gaston, K. J. (1999). Why biodiversity surveys are good value. *Nature*, *398*(6724), 204–205. https://doi.org/10.1038/18339

Cardoso, P., Rigal, F., Borges, P. A. V., & Carvalho, J. C. (2014). A new frontier in biodiversity inventory: A proposal for estimators of phylogenetic and functional diversity. *Methods in Ecology and Evolution*, *5*(5), 452–461. https://doi.org/10.1111/2041-210X.12173

Colwell, R. K., Coddington, J. A., & Hawksworth, D. L. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London B*, *345*(1311), 101–118. https://doi.org/10.1098/rstb.1994.0091

Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(2), 19651–19658. https://doi.org/10.1073/pnas.0901650106

Conroy, M. J., Runge, M. C., Nichols, J. D., Stodola, K. W., & Cooper, R. J. (2011). Conservation in the face of climate change: The roles of alternative models, monitoring, and adaptation in confronting and reducing uncertainty. *Biological Conservation*, *144*(4), 1204–1213. https://doi.org/10.1016/j.biocon.2010.10.019

Cutko, A. (2009). *Biodiversity inventory of natural lands: A how-to manual for foresters and biologists*. NatureServe.

D'Antraccoli, M., Bacaro, G., Tordoni, E., Bedini, G., & Peruzzi, L. (2020). More species, less effort: Designing and comparing sampling strategies to draft optimised floristic inventories. *Perspectives in Plant Ecology, Evolution and Systematics*, *45*, 125547. https://doi.org/10.1016/j.ppees.2020.125547

Eckblad, J. W. (1991). How many samples should be taken? *Bioscience*, *41*(5), 346–348. https://doi.org/10.2307/1311590

Escobar, L. E., Lira-Noriega, A., Medina-Vogel, G., & Peterson, A. T. (2014). Potential for spread of the white-nose fungus (*Pseudogymnoascus destructans*) in the Americas: Use of maxent and NicheA to assure strict model transference. *Geospatial Health*, *9*(1), 221–229. https://doi.org/10.4081/gh.2014.19

Faith, D. P. (2003). Environmental diversity (ED) as surrogate information for species-level biodiversity. *Ecography*, *26*(3), 374–379. https://doi.org/10.1034/j.1600-0587.2003.03300.x

Faith, D. P., Minchin, P. R., & Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, *69*(1), 57–68. https://doi.org/10.1007/BF00038687

Faith, D. P., & Walker, P. A. (1996). Environmental diversity: On the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodiversity and Conservation*, *5*(4), 399–415. https://doi.org/10.1007/BF00056387

Flores Villela, O., & Ochoa Ochoa, L. (2010). *Áreas potenciales de distribución y GAP análisis de la herpetofauna de México* (p. 23). Universidad Nacional Autónoma de México. Facultad de Ciencias. Informe final SNIB-CONABIO proyecto No. DS009.

Funk, V. A., Richardson, K. S., & Ferrier, S. (2005). Survey-gap analysis in expeditionary research: Where do we go from here? *Biological Journal of the Linnean Society*, *85*(4), 549–567. https://doi.org/10.1111/j.1095-8312.2005.00520.x

Gillespie, M. A. K., Baude, M., Biesmeijer, J., Boatman, N., Budge, G. E., Crowe, A., Memmott, J., Morton, R. D., Pietravalle, S., Potts, S. G., Senapathi, D., Smart, S. M., & Kunin, W. E. (2017). A method for the objective selection of landscape-scale study regions and sites at the national level. *Methods in Ecology and Evolution*, *8*(11), 1468–1476. https://doi.org/10.1111/2041-210X.12779

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*(4), 379–391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Hartigan, P. M. (1985). Computation of the dip statistic to test for unimodality. *Journal of the Royal Statistical Society C*, *34*(3), 320–325. https://doi.org/10.2307/2347485

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. https://doi.org/10.1002/joc.1276

Hill, D., Fasham, M., Tucker, G., Shewry, M., & Shaw, P. (2005). *Handbook of biodiversity methods: Survey, evaluation and monitoring*. Cambridge University Press.

Hirzel, A., & Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, *157*(2), 331–341. https://doi.org/10.1016/S0304-3800(02)00203-X

Hoffmann, S., Steiner, L., Schweiger, A. H., Chiarucci, A., & Beierkuhnlein, C. (2019). Optimizing sampling effort and information content of biodiversity surveys: A case study of alpine grassland. *Ecological Informatics*, *51*, 112–120. https://doi.org/10.1016/j.ecoinf.2019.03.003

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hortal, J., & Lobo, J. M. (2005). An ED-based protocol for optimal sampling of biodiversity. *Biodiversity and Conservation*, *14*(12), 2913–2947. https://doi.org/10.1007/s10531-004-0224-z

Hurlbert, A. H., & Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13384–13389. https://doi.org/10.1073/pnas.0704469104

INEGI. (2016). *Conjunto de Datos Vectoriales de Uso de Suelo y Vegetación* (1st ed.) [Map]. Instituto Nacional de Estadística y Geografía. Retrieved from http://www.conabio.gob.mx/informacion/metadata/gis/usv250s6gw.xml?_httpcache=yes&_xsl=/db/metadata/xsl/fgdc_html.xsl&_indent=no

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*(6783), 243–253. https://doi.org/10.1038/35012251

Medina, N. G., Lara, F., Mazimpaka, V., & Hortal, J. (2013). Designing bryophyte surveys for an optimal coverage of diversity gradients. *Biodiversity and Conservation*, *22*(13), 3121–3139. https://doi.org/10.1007/s10531-013-0574-5

Morrison, M. L., Block, W. M., Strickland, M. D., Collier, B. A., & Peterson, M. J. (Eds.) (2008). Sample survey strategies. In *Wildlife study design* (pp. 137–197). Springer. https://doi.org/10.1007/978-0-387-75528-1_4

Navarro-Sigüenza, A. G., & Gordillo-Martínez, A. (2018). *Mapas de distribución de las aves terrestres nativas de Mesoamérica* (p. 30). Universidad Nacional Autónoma de México. Facultad de Ciencias. Informe final SNIB-CONABIO, proyecto No. JM071.

Nuñez-Penichet, C., Cobos, M. E., Soberón, J., Gueta, T., Barve, N., Barve, V., Navarro-Sigüenza, A. G., & Peterson, A. T. (2021a). *Biosurvey: Tools for biological survey planning*. R package. Retrieved from https://CRAN.R-project.org/package=biosurvey

Nuñez-Penichet, C., Cobos, M. E., Soberón, J., Gueta, T., Barve, N., Barve, V., Navarro-Sigüenza, A. G., & Peterson, A. T. (2021b). Data from: Supplementary material: Selection of sampling sites for biodiversity inventory: Effects of environmental and geographic considerations. *figshare*, https://doi.org/10.6084/m9.figshare.14700819

Oliveira, U., Paglia, A. P., Brescovit, A. D., de Carvalho, C. J. B., Silva, D. P., Rezende, D. T., Leite, F. S. F., Batista, J. A. N., Barbosa, J. P. P. P., Stehmann, J. R., Ascher, J. S., de Vasconcelos, M. F., Marco, P. D., Löwenberg-Neto, P., Dias, P. G., Ferro, V. G., & Santos, A. J. (2016). The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. *Diversity and Distributions*, *22*(12), 1232–1244. https://doi.org/10.1111/ddi.12489

Peterson, A. T., Navarro-Sigüenza, A. G., & Martínez-Meyer, E. (2016). Digital accessible knowledge and well-inventoried sites for birds in Mexico: Baseline sites for measuring faunistic change. *PeerJ*, *4*, e2362. https://doi.org/10.7717/peerj.2362

Peterson, A. T., & Slade, N. (1998). Extrapolating inventory results into biodiversity estimates and the importance of stopping rules. *Diversity and Distributions*, *4*(3), 95–105. https://doi.org/10.1046/j.1365-2699.1998.00021.x

Peterson, A. T., & Soberón, J. (2018). Essential biodiversity variables are not global. *Biodiversity and Conservation*, *27*(5), 1277–1288. https://doi.org/10.1007/s10531-017-1479-5

R Core Team. (2021). *R: A language and environment for statistical computing* (4.0.5). R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Sarkar, S., Pressey, R. L., Faith, D. P., Margules, C. R., Fuller, T., Stoms, D. M., Moffett, A., Wilson, K. A., Williams, K. J., Williams, P. H., & Andelman, S. (2006). Biodiversity conservation planning tools: Present status and challenges for the future. *Annual Review of Environment and Resources*, *31*(1), 123–159. https://doi.org/10.1146/annurev.energy.31.042606.085844

Sastre, P., & Lobo, J. M. (2009). Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation*, *142*(2), 462–467. https://doi.org/10.1016/j.biocon.2008.11.002

Soberón, J., & Cavner, J. (2015). Indices of biodiversity pattern based on presence-absence matrices: A GIS implementation. *Biodiversity Informatics*, *10*, 22–34. https://doi.org/10.17161/bi.v10i0.4801

Soberón, J., Cobos, M. E., & Nuñez-Penichet, C. (2021). Visualizing species richness and site similarity from presence-absence matrices. *Biodiversity Informatics*, *16*, 20–27. https://doi.org/10.17161/bi.v16i1.14782

Soberón, J., & Llorente, J. (1993). The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, *7*(3), 480–488. https://doi.org/10.1046/j.1523-1739.1993.07030480.x

Soberón, J., & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, *2*, 1–10. https://doi.org/10.17161/bi.v2i0.4

Storch, D., Marquet, P., & Brown, J. (Eds.). (2007). *Scaling biodiversity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511814938

Tessarolo, G., Ladle, R. J., Lobo, J. M., Rangel, T. F., & Hortal, J. (2021). Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. *Ecography*, *44*(12), 1743–1755. https://doi.org/10.1111/ecog.05793

Velásquez-Tibatá, J. (2019). *WhereNext: Biological survey recommending system based on general dissimilarity modeling*. R package. Retrieved from https://github.com/jivelasquezt/WhereNext-Pkg/

Xu, H., Cao, M., Wu, Y., Cai, L., Cao, Y., Ding, H., Cui, P., Wu, J., Wang, Z., Le, Z., Lu, X., Liu, L., & Li, J. (2017). Optimized monitoring sites for detection of biodiversity trends in China. *Biodiversity and Conservation*, *26*(8), 1959–1971. https://doi.org/10.1007/s10531-017-1339-3

Yang, W., Ma, K., & Kreft, H. (2013). Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *Journal of Biogeography*, *40*(8), 1415–1426. https://doi.org/10.1111/jbi.12108

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Nuñez-Penichet, C., Cobos, M. E., Soberón, J., Gueta, T., Barve, N., Barve, V., Navarro-Sigüenza, A. G., Peterson, A. T. (2022). Selection of sampling sites for biodiversity inventory: Effects of environmental and geographical considerations. *Methods in Ecology and Evolution*, 13, 1595–1607. https://doi.org/10.1111/2041-210X.13869