Biodiversity Informatics, 20, 2026, pp. 89-101

DATA CURATION ON SPECIES OCCURRENCE RECORDS PLACED IN UNRECOGNIZED DISTRIBUTIONS SHAPES OUR UNDERSTANDING OF PLANT BIOCLIMATIC NICHES

Cristina Ronquillo[*,1], Juliana Stropp[2], Joaquín Hortal[1]

[1] *Dept. Biogeography & Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain*
[2] *Department of Biogeography, Trier University, Trier, Germany*

Abstract. The limitations, biases and uncertainties associated with occurrence records often limit their utility to describe species' realised bioclimatic niches. We assess changes in the perception of the realised niche of 156,500 terrestrial vascular plant species when excluding records that fall out of their recognized distribution (either native or introduced) as documented by their biogeographical status in Plants of the World Online. Filtering records following the species' geographical distribution recognised by experts results in a distortion of the representation of their niches, in some cases reducing the bioclimatic space occupied by more than half. This reduction affected between 20-40% of the species across orders, and raised up to 68% for Osmundales and Pinales or 83% in Equisetales. While most of the species evaluated (96%) only presented records in native botanical countries, about 90% of the species with introduced records expanded their realised niche beyond native climatic conditions. This calls for caution when selecting occurrence records from public repositories to conduct massive analyses on terrestrial plants. Validating these niche discrepancies derived from the curation process will allow drawing robust conclusions about species' responses to environmental changes. An effort that would also require the mobilization of occurrence data from under-sampled regions, ultimately mitigating both the Wallacean and Hutchinsonian shortfalls.

Key words: Biogeographical status, digitally accessible information, occurrence curation, realised bioclimatic niche, species occurrence records, species distributions, vascular plants.

Introduction

The availability of large volumes of biodiversity data with spatial information allows integrating a massive number of species occurrences from across the globe to address key questions in biogeography and global change. These macroecological analyses strongly depend on the available data, which often have limited geographical coverage (Hortal et al., 2007; Meyer et al., 2015, 2016; Stropp et al., 2016). One of the core sources of occurrence records is the Global Biodiversity Information Facility (GBIF), constructed from a vast amount of heterogeneous data sources, which lead to the potential inclusion of geographical gaps, errors or mismatches in the information across different data dimensions such as geographical, temporal or taxonomical (Boakes et al., 2010; Feeley & Silman, 2011; Meyer et al., 2016; Serra-Diaz et al., 2017). This generates a discrepancy between the realised and extant knowledge of biodiversity, causing a series of shortfalls that affect the different aspects of biodiversity (Hortal et al., 2015).

Our limited knowledge about the geographical distribution of species (i.e. Wallacean shortfall; Lomolino 2004; Hortal et al., 2015) manifests in multiple, interconnected ways, being one of the most pervasive ones the limited number and distribution of the species occurrence records available in public repositories. Moreover, these species occurrence records often show mismatches their known geographical distribution ranges (Hortal, 2008; Zizka et al., 2019; Arlé et al., 2021; Ronquillo et al., 2024). Distribution maps such as those provided by the Plants of the

World Online (POWO, 2024), European Forest Genetic Resources Programme (EUFORGEN, 2025) or the Red List of Threatened Species database (IUCN, 2022) use expert knowledge to depict species' distribution ranges, labelling also their biogeographical status in each territory (i.e. whether a record is native or introduced) (Svenning & Skov, 2004; Arlé et al., 2021). Biodiversity databases often contain occurrence records located distant and outside of the species recognised native and introduced geographic ranges (Chapman, 2005; Zizka et al., 2019).

The origin of these 'geographical outliers' is diverse, as they can be due to incorrect or outdated distributional maps (including changes in the names of administrative units over time), misidentifications or taxonomical mismatches, misapplication of location information, delays in updating natural and human-assisted changes in the geographic distributions, non-self-sustaining populations, or even ornamental species (Sax et al., 2013; Bocsi et al., 2016). Besides this uncertainty, occurrence data typically lack information on their biogeographical status, which prevents users from differentiating between records located in or out of the recognised geographical distribution of the species (Serra-Diaz et al., 2017; Arlé et al., 2021). The detection of these 'outliers', both in the geographic or environmental space (i.e. 'geographic outliers' and 'environmental outliers' respectively), needs to be considered in the curation workflow, labelling errors or suspicious records (see Chapman 2005; Hijmans et al., 2005; Serra-Diaz et al., 2017; Ronquillo et al., 2024). The importance of this step is endorsed by dedicated functions included in R packages (Vandepitte et al., 2015; Zizka et al., 2019; Arlé et al., 2021).

Species' environmental preferences and tolerances can be inferred from the available information provided by species occurrences records (Bocsi et al., 2016; Soberón et al., 2017) due to the direct relationship between climate and geographic space (i.e. Hutchinson's duality; Colwell & Rangel, 2009). Any geographic information can be mapped in the bioclimatic space establishing the geographic extent of a climatic condition (i.e. bioclimatic area; Coelho et al., 2023). However, evaluating the climatic suitability of species using records with uncertainty or even potential errors could compromise the reliability of such ecogeographical inferences (Hortal et al., 2008, 2012; Feeley & Silman, 2010; Lobo et al., 2010; Maldonado et al., 2015; Bocsi et al., 2016; Hughes et al., 2021b), thereby perpetuating the Hutchinsonian shortfall (i.e. the lack of knowledge on species' tolerances to abiotic conditions; Hortal et al., 2015). In particular, misidentified or incorrectly located records may lead to an incorrect definition of the climatic suitability of the species, and consequently poor modelling and predictions of future climate scenarios or conservation or management actions (Arlé et al., 2021). Whereas niche inferences using properly identified records located out of their geographical distribution would alter the known climatic limitations of species if they occupy non-overlapping environmental conditions (Broennimann et al., 2012). Distinguishing uncertain from inaccurate records is critical to assess the dynamics of geographic ranges, as both species' niches and distributions may shift through time, expanding beyond their native ranges (Araujo 2005; Bocsi et al., 2016). This is further complicated by the fact that some apparent expansions may be due to non-reported invasions, sink populations, and ornamental or cultivated specimens (Bocsi et al., 2016; Meyer et al., 2016).

Biodiversity data curation protocols are often effective to detect these geographical outliers, which are typically filtered out. However, discarding these occurrence records may result in losing potentially valuable information on the species' bioclimatic niche (Maldonado et al., 2015), an impact that has not yet been quantified. The main objective of this work is to analyse and quantify the importance of species geographical outliers in the representativeness of their realised bioclimatic niches for all terrestrial vascular plants, assessing also whether the eventual rates of niche misrepresentation differ between orders. We do so by taking into account both their geographical distributions and their recognised biogeographical status (native and introduced) in the administrative units where occurrence records are located. Specifically, we categorise different configurations of the climatic space of each species based on whether the occurrence records are placed within their geographical distribution at the botanical country level. In addition, we assess whether and to what extent outliers modify the observed niches and the main climate types occupied by each species.

Material & Methods

Records downloading and pre-processing

We downloaded all records of terrestrial Tracheophyta occurrences from GBIF via *rgbif* (Chamberlain, 2017). We included records with information of coordinates and no geospatial issues labelled ('hasGeospatialIssue'), and discarded 'machine observations', 'fossil' and 'living' specimens (the latter to avoid including records from botanical gardens). This process was done separately for each taxonomic order due to computational restrictions (see Supporting Information Table S1; GBIF.org, 2023). For each dataset, we validated and filtered data following some of the steps included in a guide for curating species

occurrence records (Ronquillo et al., 2024). The geographical check discarded records with latitude and/or longitude equal to zero or the exact same value in both fields. We kept only records whose coordinates presented at least one decimal digit. We discarded records with coordinates coinciding with capital cities and country centroids, records located near research institutions, and those located in the ocean. Then, we kept only records placed in the exact same country assigned by collectors. For the taxonomic validation, we kept records identified at species and subspecies level or varieties. The information included in the 'scientificName' field was then harmonised following the World Checklist of Vascular Plants (WCVP, Govaerts et al., 2021) using *rWCVP* R package v.1.2 (Brown et al., 2023) (see the taxonomical harmonisation output and the final checklist with the species names selected in Ronquillo, 2024). This non-supervised method led us to establish a strict threshold keeping only those records whose species names are 'Accepted' by WCVP (version 10) or the corresponding species name accepted if it was a synonym and a match type 'Exact (with author)' or 'Fuzzy (phonetic)' with match similarity higher than 0.95. No temporal filtering was applied due to the limited data available for many species that could lead us without large reductions of information. We finally discarded duplicate information based on the coordinate values and WCVP accepted species name (see Supporting Information Table S1 for occurrence numbers before and after the pre-processing).

Each validated and non-duplicated record was classified as native or introduced according to the geographical match between its location and the expert-based distribution maps from Plants of the World Online (POWO, 2024) at botanical countries level 3 (*rWCVP,* Brown et al., 2023). POWO defines a 'native' botanical country as those countries or country-like territories (for large countries such as Brazil, Canada, China, Russia or the United States of America) where the plants have been present since before the last ice age or arrived by natural colonisation afterwards (POWO, 2024). 'Introduced' countries are defined as those where the plant species were accidentally or deliberately introduced by humans, including hybrids but not the cultivated/planted ones with no self-reproduction. We classified records that were located outside of both native and introduced territories in POWO species distribution maps as 'geographical outliers'. Species with less than 10 non-duplicated points (less than five of them in their native or introduced range) were not considered for the outlier comparison analyses ('not Evaluated').

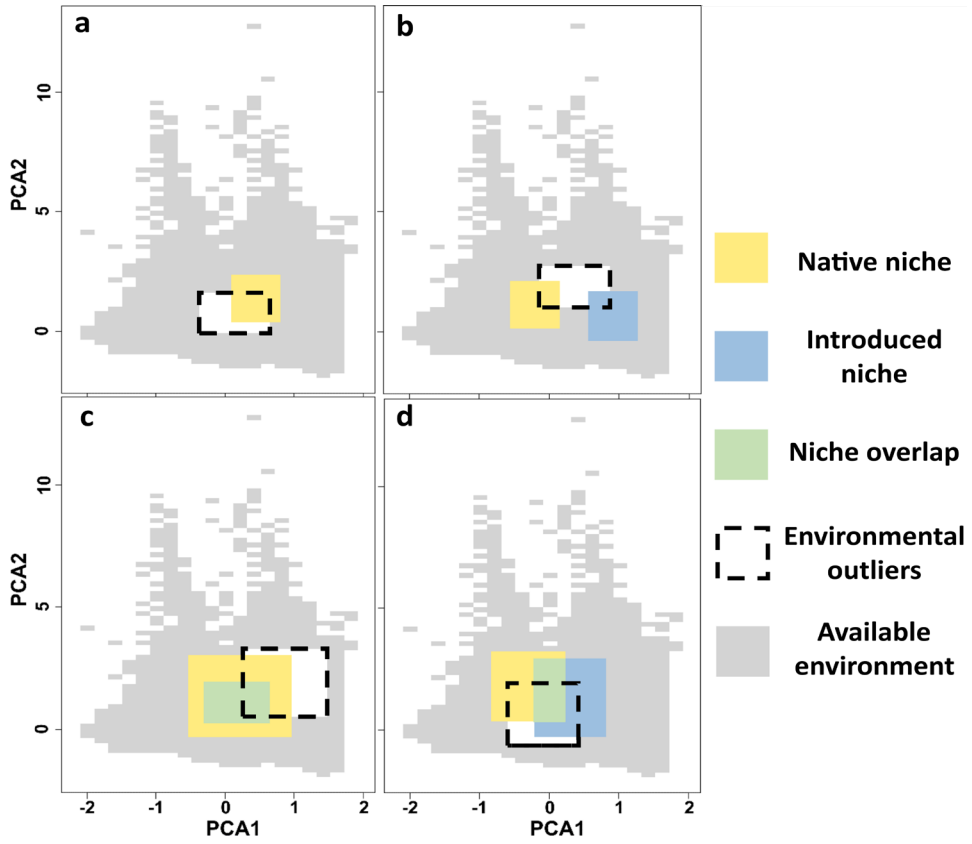
We also identified the predominant type of climate occupied by each species based on the five main classes of the Köppen-Geiger classification (arid, tropical, temperate, cold and polar). We overlapped all selected records with the global map of current climate types at 1 km resolution (Beck et al., 2018, 2023). Then, we repeated the overlap using only the subset of the records placed in their native area of distribution. Species with more than 40% of their records in two different climate classes were assigned to both categories.

Representativeness of the realised bioclimatic niche

For this study, we calculated the realised bioclimatic niches of species by using ordination techniques to recreate the multidimensional space derived from the occupied geographic space (Broennimann et al., 2012; Varela et al., 2014; Sobral-Souza et al., 2021; Coelho et al., 2023). We used Principal Components Analysis (PCA) to summarise the multivariate bioclimatic space for all terrestrial areas of the world defined by the 19 bioclimatic variables and the aridity index variable from CHELSA (version 2.1; Karger et al., 2018; Brun et al., 2022) at 50 x 50 km grid cell resolution (geographic coordinate system). By reducing this information, we selected the first two axes of the PCA representing 0.74 of the cumulative variability. Then, the bioclimatic space defined by these two PCA axes was binned by dividing these values into equal-area cells within the PCA space. These bins correspond to different 'climate domains' that provide a classification of the entire bioclimatic space available (Guisan et al., 2014). We conducted this binning at two different scales. One divided the two axes into 0.2 x 0.2 cells (n = 613) as a fine resolution that can capture more differences between the bioclimatic domains used by each species, and the other divided the two axes into 0.5 x 0.5 cells (n = 138) to provide a coarser description of species' bioclimatic niches. Each cell represents a unique set of climate conditions (climate type) observed in one or several geographic locations (Colwell & Rangel, 2009, Coelho et al., 2023). Thus, we assigned each species occurrence record to its corresponding cell of the multivariate bioclimatic space based on the values of PCA1 and PCA2 in its geographic location (see Ronquillo et al., 2020 and Sobral-Souza et al., 2021 for similar approaches).

We defined which cells of the bioclimatic space are covered by each species, discriminating between cells hosting records within the native distribution, and records located in introduced countries. For the geographical outliers, we identified the corresponding cells in the bioclimatic space only when they were not already hosting records placed within the native or introduced geographical distribution (here 'environmental outliers'). Then, we established four hypothetical configurations based on the spatial combination of native and introduced cells (that

Figure 1. Conceptual framework followed to define the four hypothetical configurations of species' realised niche based on the biogeographical status of the occurrence records. (a) Native niche; (b) native niche does not overlap the introduced niche; (c) introduced niche is included into the native niche; (d) introduced niche expand the native niche (partial or totally). Note that environmental outliers are represented in all configurations and may 1) not overlap, 2) overlap partially or 3) overlap completely with native/introduced maps in the environmental space.



may or may not include environmental outliers). These hypothetical configurations follow the terms in Guisan et al. (2014), as species (a) whose records occupy native niche or are environmental outliers (Fig. 1a); (b) whose native niche does not overlap with the introduced niche (i.e. niche expansion) (Fig. 1b); (c) whose records are located in the introduced geographical space occupying climatic conditions also available in the native niche (i.e. niche stability) (Fig. 1c); and (d) whose native and introduced records partially overlap of the same available environment (i.e. niche expansion) (Fig. 1d). Besides, there is another case in which the data available in GBIF only allow to define the introduced niche of the species, due to the insufficient number of records placed in the native range (< 10 records).

All analyses were performed in R Studio 2024.09.0+375 "Cranberry Hibiscus" (R Core Team, 2021). The main R packages used were *rSDM* (Rodriguez-Sanchez, 2024), *sf* (Pebesma & Bivand, 2023) and *terra* (Hijmans, 2024).

Results

We downloaded 351,360,932 records from GBIF. After data curation and filtering, the number of validated records decreased to 164,441,270 (see Supporting Information Table S1). These records accounted for the occurrence of 276,090 accepted Tracheophyta species from 84 accepted orders. Our data included all the orders of Lycopodiopsida (Isoetales, Lycopodiales and Selaginalles in PPG I, 2016), vascular ferns (Cyathales, Equisetales, Gleicheniales, Hymenophyllales, Marattiales, Ophioglossales, Osmundales, Polypodiales, Psilotales, Salviniales and Schizaeales in Nitta et al., 2022), as well as Gymnosperm orders (see Yang et al., 2022) being Araucariales and Cupressales included in Pinales. For the angiosperm orders, we included all accepted orders included in The Angiosperm Phylogeny Group (2016).

The classification of each species in their corresponding climate space configuration led us to discard 119,590 species (43.32%), which were not evaluated because they presented less than 10 different valid and unique occur-

rence records. More than half of the species were not assessed for Arecales, Asparagales, Canellales, Cucurbitales, Gunnerales, Isoetales, Marattiales, Pandanales, Piperales and Zingiberales (Fig. 2). Up to 54% of the species considered (150,200) only occupied native administrative units (see example in Fig. 3a), being also the majority for each order except for the species-poor Ginkgoales (1 spp) and Acorales (2 spp) orders (Fig. 2). There were 408 species that presented non-overlapping niches for their native and introduced geographical distributions (see example in Fig. 3b). The number of species that had their introduced niche represented within the native niche was 708 (see example in Fig. 3c), while 5,024 species expanded their introduced bioclimatic niche beyond the bioclimatic native niche (see example in Fig. 3d). Finally, for 185 species, there were not enough records available in GBIF whose coordinates were located within any native botanical country. Then, their bioclimatic niche was defined only using records from their introduced distribution (see Supporting Information Figure S1). In addition, no species show a configuration with exactly the same native and introduced cells covering the species bioclimatic niche.

The use of different grid resolution also led us to obtain different bioclimatic niches when evaluating species individually (Supporting Information Figure S2). Particularly for those species with records distributed

Figure 2. Proportion of species by order in each niche configuration. (a) Native (red), (b) native niche does not overlap the introduced niche (yellow), (c) the introduced niche is included into the native niche (orange), (d) both native and introduced niches partially overlap (purple). Species with less than 10 records were not classified (grey).

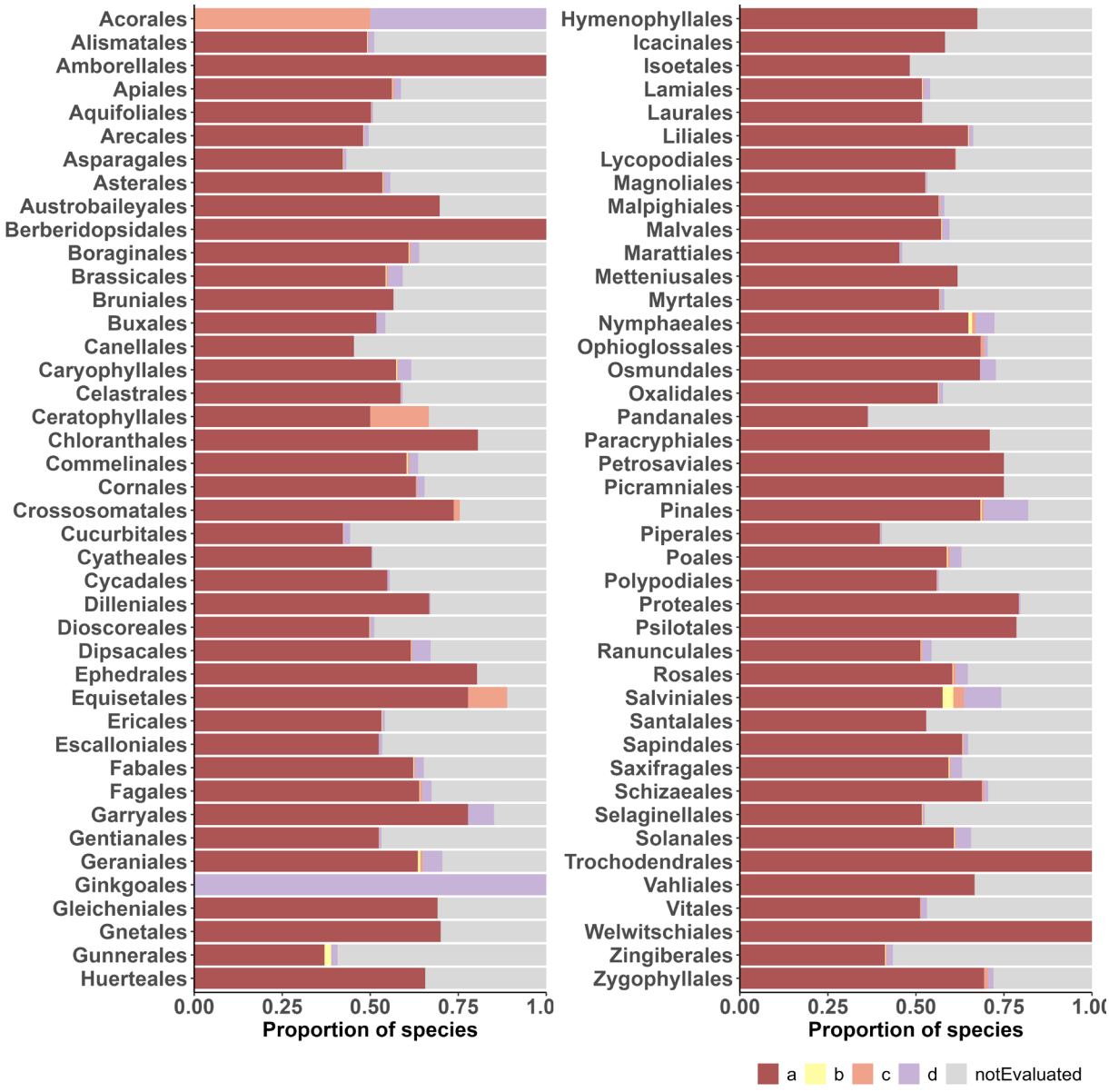


Figure 3. Examples of each hypothesised species' niche configurations at 0.2 x 0.2 resolution and maps (Mollwide projection) based on the presence of species occurrences in their native (yellow) or introduced (blue) botanical countries; the grey area corresponds to the available environment. (a) Native (*Sorbaria grandiflora*); (b) native niche does not overlap with the introduced niche (*Stauntonia hexaphylla*); (c) introduced niche is included into the native niche (*Cordia dentata*); (d) both native and introduced niches partially overlap (*Linaria vulgaris*).

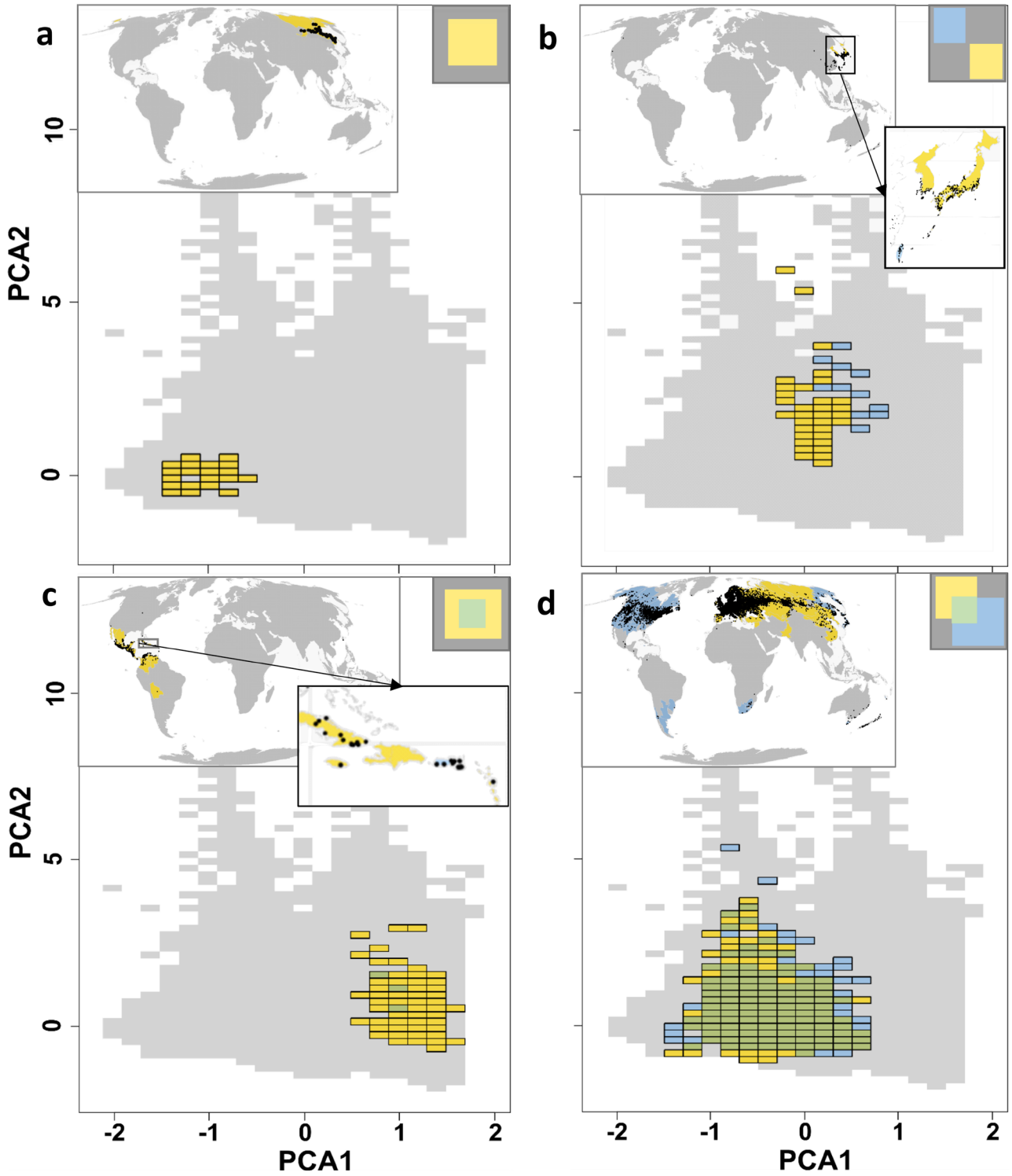
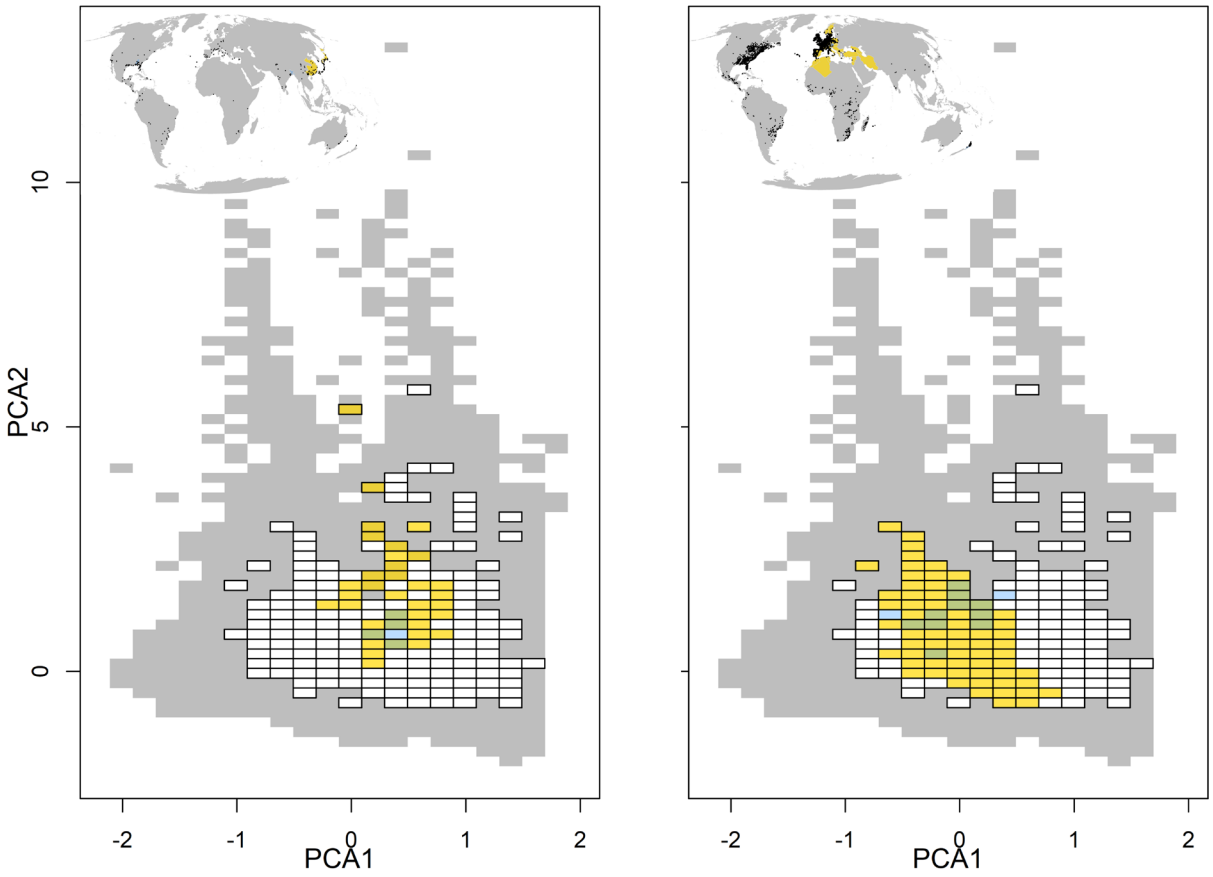


Figure 4. Examples of the niche representation of species occurrence records (Left panel *Cycas revoluta;* Right panel *Osmunda regalis*) if only those located in their native (yellow) and introduced range (blue) were selected or if the range outliers were included (white cells). Graphs at 0.2 x 0.2 resolution and maps use Mollwide projection.



along introduced ranges that occupy part of the niche near the native one. However, the finest resolution gave us more detail of potential differences in the occupied climate. By order, the proportion of the available environment represented by cells with species occurrences was similar at both resolutions (see Supporting Information Table S2), yet, 0.5 x 0.5 cells cover a bigger proportion. The largest changes between resolutions were observed in Gunnerales, Pandanales and Paracryphiales, occupying respectively 13.7%, 12.8% and 15.7% more of the available environment than at 0.2 x 0.2 resolution.

The proportion of cells discarded was significantly greater after excluding both the environmental outliers and introduced niche cells than when excluding only the introduced cells (Wilcoxon signed-rank test *p-values* < 0.05, see Supporting Information Table S3). This occurred at both resolutions for most orders (see Supporting Information Figure S3 and S4 for both cell resolutions). However, it was not significant (*p-values* > 0.05) for Acorales, Amborellales, Berberidopsidales, Ceratophyllales, Ginkgoales, Petrosaviales, Trochodendrales, Vahliales and Welwitschiales because all these orders have small numbers of species (between 1 and 6 each; Supporting Information Table S4). In general, the proportion of species with environmental outliers (cells in the available environment defined only by geographical outliers) presented high variability by order (Supporting Information Table S4). Nevertheless, it is noteworthy that they represented 83% of the species from Equisetales and 68% from Osmundales and Pinales. Indeed, while less than 3% of the species had more than half of their realised niche defined by geographical outliers (see examples in Fig. 4), for Pinales and Cycadales this value increased to 7% and 10% of their species, respectively (Supporting Information Table S4).

Most of the species occupied regions of temperate climate, a trend also observed across the configurations of species' bioclimatic niche (Supporting Information Figure S5 and S6). However, this was not the case for the 119,590 species 'non-evaluated' for bioclimatic analyses (<10 records), which were mostly located in tropical areas, albeit

closely followed by the temperate climate class (Supporting Information Figure S7). We observed that between c. 90% of species were distributed in the same climate class of Köppen-Geiger after the application of a filter to keep only records within the native range (Supporting Information Figure S6a and Table S5). The same pattern was observed in species whose introduced niche is included within the native one. However, for species whose introduced niche does not overlap or expanded the native one, the inclusion of records out of the native range may produce more than 10% of changes from arid, cold or tropical climate towards temperate (Supporting Information Figure S6b,c,d). This pattern was also detected for species which information were mainly assessed using records from the introduced ranges and range outliers. By using only one to nine records located in the native range, a high proportion of species were transferred from their native climate classes to all directions (Supporting Information Fig. S1c and Table S6). Even so, most of these species had no native information available in GBIF and thus were not classified in any Köppen-Geiger climate class.

Discussion

Our results show that applying geographical filters to occurrence records based on the biogeographical status of species (either native or introduced) affects the representativeness of their realised bioclimatic niche –or at least the portion of it that can be described with currently available distributional data. The main factor driving change in the size and shape of the bioclimatic niche occupied by each species was the inclusion of geographical outliers – records placed in botanical countries unrecognised in the species distribution ranges described in POWO (2024). We showed that filtering out these outliers from GBIF data leads to important decreases in the representativeness of bioclimatic niches for most orders. For some species excluding these records that may even correspond to more than half of the cells included. This affected up to 10% of the species of Cycadales or 7% of Pinales, albeit for most orders that proportion was only 1-3% of the species evaluated.

Robust ecological research demands the mitigation of potential errors that could introduce noise and bias into broad-scale analyses. This requires a rigorous and adequately justified data selection process. The implementation of specific data curation protocols influences the quantity of usable occurrences and, in turn, the spatial representation of biodiversity across different territories. Researchers must therefore adopt the analytical strategy that best suits the particular aims of each study, whether these involve delineating species distributions, constructing ecological niche models, characterizing niche attributes, inferring community structure across environmental gradients, or evaluating knowledge gaps. Each of these research lines entails distinct objectives and taxonomic groups, which demand the application of diverse processing steps and filters. Recent contributions (Ronquillo et al., 2023, 2024; Tessarolo et al., 2021) emphasize the necessity of implementing reproducible data curation protocols and ensuring transparent communication of biases and uncertainties. Upholding data quality and methodological soundness is fundamental to strengthening the robustness and interpretability of ecological findings.

Discrepancies between data from GBIF and expert-based maps have been reported previously (Hughes et al., 2021b) and in this study, we further highlight the importance of evaluating the application of filters to records placed outside the recognised distributions. It is possible that these occurrences correspond to taxonomic misidentifications, ornamental cultivars or sink populations, rather than viable wild populations that provide a fairer representation of the species' geographical distribution and the occupied environment due to niche tolerance in the range edges (see Sax et al., 2013). However, filtering out geographical outliers by default could lead to omission errors if they correspond to populations well established in the region. Distributional maps are sometimes too imprecise, not updated, or they do not define properly the actual distribution of the species in edge countries (or even countries placed far from the delimited distribution published). In this situation, filtering out geographical outliers would hamper our ability to obtain a correct estimation of species niches (Bush et al., 2018, Sax et al., 2013). Also, some poorly known groups such as bryophytes within plants, or many invertebrates within animals, do not have proper repositories or sources of well-defined distributional polygons which do not allow to evaluate this dimension of the curation process. While the methodology presented here has this limitation, we encourage to evaluate environmental outliers to detect mistakes in occurrence repositories. Indeed, some alternatives ways to evaluate 'environmental outliers' have already been proposed (Vandepitte et al., 2015; Robertson et al., 2016; Arlé et al., 2021; Ronquillo et al., 2024).

Nonetheless, our results prove that there is an important proportion of records along Tracheophyta in GBIF that require a careful examination of the quality and uncertainty associated. We advocate for the use of alternative sources of distributional information for local or regional studies, such as national atlases or more specific maps (e.g., EUFORGEN; BiodiversityMapping.org; Anthos in www.anthos.es). Note that without knowing the actual biogeo-

graphical status of these geographical outliers, it is not possible to assess whether the increase in the realised bioclimatic niche is due to previously undetected native areas, or naturalised areas (Soberón, 2007). However, the bioclimatic niches described by outliers may largely correspond to missing portions of their realised niches included in the species' fundamental niche. It is therefore necessary to integrate all field observations into the geographical range of species using expert knowledge (Hughes et al., 2021a). By doing so, we can derive consistent biodiversity patterns from reliable species diversity maps (Hughes et al., 2021b). To control this problem, researchers may consider labelling the biogeographical status of occurrence records for specific sets of data. They can filter records placed within either native or non-native distributions, improving data accuracy or even detecting invasion processes to obtain a better definition of niche changes (Meyer et al., 2016; Arlé et al., 2021; Ronquillo et al., 2024). Using only the native range occurrences can be a poor proxy for climatic tolerance (Bocsi et al., 2016). Yet, it might be argued that without this distinction of biogeographical status, the results of macroecological analyses and conservation assessments can be biased by the inclusion of non-natives within native species ranges (Meyer et al., 2016).

The majority of species across all orders have been recorded only in their native botanical countries. In contrast, our study includes 6,325 species with occurrences in non-native areas, and c. 90% of them exhibit expanded bioclimatic niches beyond the native realised niche. While niche shifts to naturalised and invaded areas have been documented for other groups (e.g., Silva et al., 2016; Zhu et al., 2017; Pili et al., 2020), they were considered rare for terrestrial plants (Petitpierre et al., 2012). This has important implications for assessing uncertainties in model projections to invaded ranges that are climatically distinct. An accurate definition of species' realised niche is key for forecasting invasions under climate change scenarios (Qiao et al., 2017). By adding the introduced ranges, we are including areas without the native range's biotic constraints (Broennimann & Guisan, 2008). Thus, users should carefully select occurrence records that capture species' requirements to properly define niche attributes such as stability, drift or expansion (Guisan et al., 2014; Zhu et al., 2017; Pili et al., 2020). It is also possible that range outliers occupy climatic conditions comparable to those prevailing within their confirmed geographical range. In such instances, applying filters based on inferred biogeographic status would not alter the delineation of species niches nor influence downstream climatic interpretations, including those arising from Species Distribution Models (SDMs) or Ecological Niche Models (ENMs).

Our results may be limited by using coarse sources of geographical distribution on species such as POWO. However, this type of information provides a reasonable tool for representing large-scale biodiversity patterns, as they synthesise extensive data of species presence or absence within botanical (or zoological) countries (e.g., Svenning & Skov, 2004; König et al., 2019; Coelho et al., 2023; Divieso et al., 2024). For many of the species we analysed, numerous records fall outside the recognised botanical countries, but alternative expert-maps (such as the IUCN) would be even stricter when discarding geographical outliers from repositories. Some limitations and mismatches between distributional maps and records are the outdated information, the influence of political boundaries (missing records that are located near assigned regions) or biases associated with overoptimistic representations of species distributions attributed to the whole territory (Hughes et al., 2021b; Lemoine & Svenning, 2022).

One of the main restrictions of this study was that a large proportion of species could not be evaluated, as more than 43% of the Tracheophyta did not meet the threshold of 10 occurrences per species. This percentage was roughly similar for all the orders considered. We implemented for this particular study a conservative threshold of 10 records to ensure the robustness and reliability of the conclusions derived from our specific methodological workflow. Even though the large number of discarded species, we were able to evaluate this filter and their scenarios for 276,090 species. The curation process and the criteria defined by the authors varies with the ecological question, group of taxa, methods and analyses selected. As highlighted previously, the limited effort devoted to plant inventories in vast areas of the globe hampers the accurate definition of the species distributions, thus leaving the Wallacean shortfall yet unresolved (Lomolino, 2004; Whittaker et al., 2005; Meyer et al., 2016; Hortal et al., 2015; Cornwell et al., 2019). This geographic data deficiency necessarily constrains the assessment of realized bioclimatic niches, thereby failing to address the Hutchinsonian shortfall (Hortal et al., 2015; Meyer et al., 2015). We encourage researchers to use of alternative tools when the number of records is too low as '*leave one out*', '*plug-and-play*' modelling or *environmental-range* models (see Maitner et al., 2026 for small (<20) sample sizes; Drake, 2014 and Drake & Richards, 2018 - S4DM 'Small Sample Size Species Distribution Modeling' or flexsdm and the thinning functions 'occfilt' in Velazco et al., 2022).

Our results corroborate previously documented uneven patterns of data availability –and their related geographic biases– in GBIF, showing a disproportionate representation of temperate climate regions. This geographic

bias is not random; it reflects the concentration of biodiversity data mobilization efforts in wealthy, developed nations, primarily in Europe and North America, where historical funding, political stability, and institutional research capacity have enabled the long-term accumulation of digitized occurrence records, while tropical and biodiverse regions remain comparatively underrepresented (Hughes et al., 2021a; Chapman et al., 2024). It is then not surprising that most species with limited information available (i.e. 'not evaluated') were located in the tropical climate associated to under-sampled and less-known areas of the world (Feeley & Silman, 2010; Hughes et al., 2021a), an example of how the Wallacean shortfall in general, and sampling biases in particular, affect differentially certain habitats and ecoregions. We strongly suggest that future efforts of surveying and digitization of natural history collections take place in these areas to improve the definition of past, present and future species ranges, which will ameliorate the assessment of realised niches and potential niche shifts. Moreover, the absence of a temporal filter and the limitation of at least 10 occurrences to calculate the bioclimatic space highlights the importance of a proper curation process of public data. In this study, we opted not to apply a temporal filter due to the substantial reduction in data availability for many species with limited records, which could compromise the robustness of our analyses. Nevertheless, we acknowledge that failing to match the temporal window of climatic data (here CHELSA window – 1980-2010) may influence the representation of bioclimatic niches (Supporting Information Fig. S8). Our findings underscore the trade-off between data availability and temporal matching with environmental data, and we encourage users of biodiversity information to consider temporal filtering when feasible, as highlighted in recent works (Tessarolo et al., 2017; Ronquillo et al., 2023). Importantly, while the ideal scenario would incorporate temporal alignment, the computational demands and data limitations for thousands of species make such an approach impractical for the scope of this study.

Conclusions

The results of this study advise that labelling geographical outliers and evaluating their influence on species bioclimatic niche are crucial steps for assessing the reliability of niche estimates. The mismatch between distribution map information and occurrence records affects to a substantial proportion of public data (see also Hortal, 2008). Thus, we demonstrate the existence and level of a potential source of distortion when defining the realised bioclimatic niche of the species based on this issue and the incomplete and biased data currently available. This may be considered together with other ecological assumptions in broad-scale analyses. Identifying these outliers can help detecting dispersion of alien species, taxonomic misidentifications or unrecognised geographical areas (Vandepitte et al., 2015). Developing a thoughtful and well-informed process of filtering species records is key to reduce data-driven uncertainty, particularly for analyses that will derive climatic responses of the species (e.g., SDMs, ENMs).

Comprehending the geographical dynamics of species distributions is essential for macroecological and biogeographical studies. By understanding both spatial distributions and the structure of the biases and gaps that make up the related knowledge shortfalls, we will be able to reconstruct the large-scale patterns of biodiversity in a fair way. To achieve this, we still need a series of methodologies that allow us to evaluate the quality and reliability of the available biodiversity data. Only through a cautious assessment of the representativeness of the biodiversity information available, we will provide fair representations of ecological processes, robust models and projections.

Acknowledgments

This work was funded by Spanish AEI coordinated projects SCENIC (grant PID2019-106840GB-C21) and NICED (grant PID2022-140985NB-C21) funded by MCIN/AEI/10.13039/501100011033. Author contributions − CR: Conceptualisation (equal); data curation (lead); formal analysis (lead); methodology, software and visualisation (lead); writing – original draft (lead); writing – review and editing (equal). JS: Conceptualisation (equal); formal analysis, validation and visualisation (supporting); writing – original draft (supporting); writing – review and editing (equal). JH: Conceptualisation (equal); formal analysis, methodology and visualisation (supporting); funding acquisition and project administration (lead); writing – original draft (supporting); writing – review and editing (equal).

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article exist.

Data Accessibility

Our Figshare repository (Ronquillo, 2024) contains all R codes used for processing, analyses, and figure creation, along with the validated occurrences datasets, taxonomically harmonized datasets for each group, and a complete list of accepted species names with their corresponding GBIF-indexed terms. The downloaded occurrence datasets are accessible via links to the Global Biodiversity Information Facility (GBIF.org, 2023) provided in the Supporting Information Table S1.

References

APG [Angiosperm Phylogeny Group]. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc, 181(1):1-20. https://doi.org/10.1111/boj.12385

Araújo, M.B. 2005. Equilibrium of species' distributions with climate. Ecography, 28(5):693-695. https://www.jstor.org/stable/3683759

Arlé, E., Zizka, A., Keil, P., Winter, M., Essl, F., Knight, T., Jetz, W., Kreft, H., Weigelt, P., Meyer, C. 2021. bRacatus: a method to estimate the accuracy and biogeographical status of georeferenced biological data. Methods Ecol Evol, 12(9):1609-1619. https://doi.org/10.1111/2041-210X.13629

Beck, H.E., McVicar, T.R., Vergopolan, N., Berg, A., Lutsko, N.J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A.I.J.M., Miralles, D.G. 2023. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. Sci Data, 10:724. https://doi.org/10.1038/s41597-023-02549-6

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data, 5:180214. https://doi.org/10.1038/sdata.2018.214

Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-Qing, D., Clark, N.E., O'Connor, K., Mace, G.M. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol, 8(6):e1000385. https://doi.org/10.1371/journal.pbio.1000385

Bocsi, T., Allen, J.M., Bellemare, J., Kartesz, J., Nishino, M., Bradley, B.A. 2016. Plants' native distributions do not reflect climatic tolerance. Divers Distrib, 22:615-624. https://doi.org/10.1111/ddi.12432

Broennimann, O., Guisan, A. 2008. Predicting current and future biological invasions: both native and invaded ranges matter. Biol Lett, 4(5):585-589. https://doi.org/10.1098/rsbl.2008.0254

Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz, N.G., Thuiller, W., Fortin, M.J., Randin, C., Zimmermann, N.E., Graham, C.H., Guisan, A. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. Glob Ecol Biogeogr, 21(4):481-497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Brown, M.J.M., Walker, B.E., Black, N., Govaerts, R.H.A., Ondo, I., Turner, R., Lughadha, E.N. 2023. rWCVP: a companion R package for the World Checklist of Vascular Plants. New Phytol, 240(4):1355-1365. https://doi.org/10.1111/nph.18919

Brun, P., Zimmermann, N.E., Hari, C., Pellissier, L., Karger, D.N. 2022. Data from: CHELSA-BIOCLIM+ a novel set of global climate-related predictors at kilometre resolution. EnviDat, https://doi.org/10.16904/envidat.332

Bush, A., Catullo, R.A., Mokany, K., Thornhill, A.H., Miller, J.T., Ferrier, S. 2018. Truncation of thermal tolerance niches among Australian plants. Glob Ecol Biogeogr, 27(1):22-31. https://doi.org/10.1111/geb.12637

Chamberlain, S. 2017. rgbif: interface to the Global Biodiversity Information Facility API. R package version 3.7.8. https://CRAN.R-project.org/package=rgbif

Chapman, A.D. 2005. Principles and methods of data cleaning - primary species and species occurrence data, version 1.0. Copenhagen: Global Biodiversity Information Facility.

Coelho, M.T.P., Barreto, E., Rangel, T.F., Diniz-Filho, J.A.F., Wüest, R.O., Bach, W., Skeels, A., McFadden, I.R., Roberts, D.W., Pellissier, L., Zimmermann, N.E., Graham, C.H. 2023. The geography of climate and the global patterns of species diversity. Nature, 622:537-544. https://doi.org/10.1038/s41586-023-06577-5

Colwell, R.K., Rangel, T.F. 2009. Hutchinson's duality: the once and future niche. Proc Natl Acad Sci USA, 106(Suppl 2):19651-19658.

Cornwell, W.K., Pearse, W.D., Dalrymple, R.L., Zanne, A.E. 2019. What we (don't) know about global plant diversity. Ecography, 42(11):1819-1831. https://doi.org/10.1111/ecog.04481

Drake, J.M. 2014. Ensemble algorithms for ecological niche modeling from presence-background and presence-only data. Ecosphere, 5(6):76. http://dx.doi.org/10.1890/ES13-00202.1

Drake, J.M., Richards, R.L. 2018. Estimating environmental suitability. Ecosphere, 9(9): e02373. 10.1002/ecs2.2373

Divieso, R., Pie, M.R., Hortal, J. 2024. On the macroecology of rarity and vulnerability to extinction in terrestrial mammals. Biol Conserv, 296:110673. https://doi.org/10.1016/j.biocon.2024.110673

EUFORGEN [European Forest Genetic Resources Programme]. 2025. EUFORGEN website. https://www.euforgen.org/

Feeley, K.J., Silman, M.R. 2010. Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering. J Biogeogr, 37(4):733-740. https://doi.org/10.1111/j.1365-2699.2009.02240.x

Feeley, K.J., Silman, M.R. 2011. The data void in modelling current and future distributions of tropical species. Glob Chang Biol, 17(1):626-630. https://doi.org/10.1111/j.1365-2486.2010.02239.x

GBIF.org. 2023. GBIF Occurrence Download Accessed from R via rgbif (https://github.com/ropensci/rgbif) on 2023-12-(04-22). See Table S1 for individual links.

Chapman, M., Goldstein, B.R., Schell, C.J., Brashares, J.S., Carter, N.H., Ellis-Soto, D., Faxon, H.O., Goldstein, J.E., Halpern, B.S., Longdon, J., Norman, K.E.A., O'Rourke, D., Scoville, C., Xu, L., Boettiger, C. (2024). Biodiversity monitoring for a just planetary future. Science, 383(6678): 34-36. https://doi.org/10.1126/science.adh8874

Govaerts, R., Lughadha, E.N., Black, N., Turner, R., Paton, A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. Sci Data, 8:215. https://doi.org/10.1038/s41597-021-00997-6

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., Kueffer, C. 2014. Unifying niche shift studies: insights from biological invasions. Trends Ecol Evol, 29(5):260-269. https://doi.org/10.1016/j.tree.2014.02.009

Hijmans, R.J. 2024. terra: spatial data analysis. R package version 1.8-73. https://github.com/rspatial/terra

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. Int J Climatol, 25(15):1965-1978. https://doi.org/10.1002/joc.1276

Hortal, J. 2008. Uncertainty and the measurement of terrestrial biodiversity gradients. J Biogeogr, 35(8):1355-1356. https://doi.org/10.1111/j.1365-2699.2008.01955.x

Hortal, J., Bello, F. de, Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M., Ladle, R.J. 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annu Rev Ecol Evol Syst, 46: 523-549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M., Baselga, A. 2008. Historical bias in biodiversity inventories affects the observed realized niche of the species. Oikos, 117(6):847-858. https://doi.org/10.1111/j.0030-1299.2008.16434.x

Hortal, J., Lobo, J.M., Jiménez-Valverde, A. 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. Conserv Biol, 21(3):853-863. https://doi.org/10.1111/j.1523-1739.2007.00686.x

Hortal, J., Lobo, J.M., Jiménez-Valverde, A. 2012. Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. Nat Conserv, 10(2):108-118. http://dx.doi.org/10.4322/natcon.2012.029

Hughes, A.C., Orr, M.C., Ma, K., Costello, M.J., Waller, J., Provoost, P., Yang, Q., Zhu, C., Qiao, H. 2021a. Sampling biases shape our view of the natural world. Ecography, 44(9):1259-1269. https://doi.org/10.1111/ecog.05926

Hughes, A.C., Orr, M.C., Yang, Q., Qiao, H.. 2021b. Effectively and accurately mapping global biodiversity patterns for different regions and taxa. Glob Ecol Biogeogr, 30(7):1375-1388. https://doi.org/10.1111/geb.13304

IUCN [International Union for Conservation of Nature]. 2022. The IUCN Red List of Threatened Species. https://www.iucnredlist.org

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M. 2018. Data from: climatologies at high resolution for the earth's land surface areas. EnviDat. Version 2.1. https://doi.org/10.16904/envidat.228.v2.1

König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J., Kreft, H. 2019. Biodiversity data integration - the significance of data resolution and domain. PLoS Biol, 17(3):e3000183. https://doi.org/10.1371/journal.pbio.3000183

Lemoine, R.T., Svenning, J.C. 2022. Nativeness is not binary—a graduated terminology for native and non-native species in the Anthropocene. Restor Ecol, 30(5):e13636. https://doi.org/10.1111/rec.13636

Lobo, J.M., Jiménez-Valverde, A., Hortal, J. 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography, 33(1):103-114. https://doi.org/10.1111/j.1600-0587.2009.06039.x

Lomolino, M.V. 2004. Conservation biogeography. In: Lomolino MV, Heaney LR, editors. Frontiers of biogeography: new directions in the geography of nature. Sunderland (MA): Sinauer Associates. p. 293-296.

Maitner, B.S., Richards, R.L., Carlson, B.S., Drake, J.M., Merow, C. 2026. Flexible methods for species distribution modeling with small samples. Ecography, e08112. https://doi.org/10.1002/ecog.08112

Maldonado, C., Molina, C.I., Zizka, A., Persson, C., Taylor, C.M., Albán, J., Chilquillo, E., Rønsted, N., Antonelli, A. 2015. Estimating species diversity and distribution in the era of big data: to what extent can we trust public databases? Glob Ecol Biogeogr, 24(8):973-984. https://doi.org/10.1111/geb.12326

Meyer, C., Kreft, H., Guralnick, R., Jetz, W. 2015. Global priorities for an effective information basis of biodiversity distributions. Nat Commun, 6:8221. https://doi.org/10.1038/ncomms9221

Meyer, C., Weigelt, P., Kreft, H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecol Lett, 19(8):992-1006. https://doi.org/10.1111/ele.12624

Nitta, J.H., Schuettpelz, E., Ramírez-Barahona, S., Iwasaki, W. 2022. An open and continuously updated fern tree of life. Front Plant Sci, 13:909768. https://doi.org/10.3389/fpls.2022.909768

Pebesma, E., Bivand, R. 2023. Spatial data science: with applications in R. Boca Raton (FL): Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., Guisan, A. 2012. Climatic niche shifts are rare among terrestrial plant invaders. Science, 335 (6074):1344-1348. https://doi.org/10.1126/science.1215933

Pili, A.N., Tingley, R., Sy, E.Y., Diesmos, M.L.L., Diesmos, A.C. 2020. Niche shifts and environmental non-equilibrium undermine the usefulness of ecological niche models for invasion risk assessments. Sci Rep, 10:7972. https://doi.org/10.1038/s41598-020-64568-2

POWO [Plants of the World Online]. 2024. Royal Botanic Gardens, Kew. http://www.plantsoftheworldonline.org/

PPG I [The Pteridophyte Phylogeny Group]. 2016. A community-derived classification for extant lycophytes and ferns. J Syst Evol, 54(6):563-603. https://doi.org/10.1111/jse.12229

Qiao, H., Escobar, L.E., Peterson, A.T. 2017. Accessible areas in ecological niche comparisons of invasive species: recognized but still overlooked. Sci Rep, 7:1213. https://doi.org/10.1038/s41598-017-01313-2

R Core Team. 2021. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. https://www.R-project.org/

Robertson, M.P., Visser, V. Hui, C. 2016. Biogeo: an R package for assessing and improving data quality of occurrence

record datasets. Ecography, 39:394-401. https://doi.org/10.1111/ecog.02118

Rodriguez-Sanchez, F. 2024. rSDM: species distribution and niche modelling in R. R package version 0.4.

Ronquillo, C., Alves-Martins, F., Mazimpaka, V., Sobral-Souza, T., Vilela-Silva, B.G., Medina, N.G., Hortal, J. 2020. Assessing spatial and temporal biases and gaps in the publicly available distributional information of Iberian mosses. Biodivers Data J, 8:e53474. https://doi.org/10.3897/BDJ.8.e53474

Ronquillo, C., Stropp, J., Medina, N.G., Hortal, J. 2023. Exploring the impact of data curation criteria on the observed geographical distribution of mosses. Ecol Evol. 13(12):e10786. https://doi.org/10.1002/ece3.10786

Ronquillo, C., Stropp, J., Hortal, J. 2024. OCCUR Shiny application: a user-friendly guide for curating species occurrence records. Methods Ecol Evol, 15(5):816-823. https://doi.org/10.1111/2041-210X.14271

Ronquillo, C. 2024. Data and R Code of Data curation on species occurrence records placed in unrecognized distributions shapes our understanding of plant bioclimatic niches (Version 9). Figshare. https://doi.org/10.6084/m9.figshare.26021467.v9

Sax, D.F., Early, R., Bellemare, J. 2013. Niche syndromes, species extinction risks, and management under climate change. Trends Ecol Evol, 28(9):517-523. https://doi.org/10.1016/j.tree.2013.05.010

Serra-Diaz, J.M., Enquist, B.J., Maitner, B., Merow, C., Svenning, J.C. 2017. Big data of tree species distributions: how big and how good? For Ecosyst, 4:30. https://doi.org/10.1186/s40663-017-0120-0

Silva, D.P., Vilela, B., Buzatto, B.A., Moczek, A.P., Hortal, J. 2016. Contextualized niche shifts upon independent invasions by the dung beetle Onthophagus taurus. Biol Invasions, 18(11):3137-3148. https://doi.org/10.1007/s10530-016-1204-4

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecol Lett, 10(12):1115-1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Soberón, J., Osorio-Olvera, L., Peterson, T. 2017. Diferencias conceptuales entre modelación de nichos y modelación de áreas de distribución. Rev Mex Biodivers, 88(2):437-441. http://dx.doi.org/10.1016/j.rmb.2017.03.011

Sobral-Souza, T., Stropp, J., Santos, J., Prasniewski, V.M., Szinwelski, N., Vilela, B., Freitas, A.V.L., Ribeiro, M.C., Hortal, J. 2021. Knowledge gaps hamper understanding the relationship between fragmentation and biodiversity loss: the case of Atlantic Forest fruit-feeding butterflies. PeerJ, 9:e11673. https://doi.org/10.7717/peerj.11673

Stropp, J., Ladle, R.J., Malhado, A.C.M., Hortal, J., Gaffuri, J., Temperley, W., Skøien, J.O., Mayaux, P. 2016. Mapping ignorance: 300 years of collecting flowering plants in Africa. Glob Ecol Biogeogr, 25(9):1085-1096. https://doi.org/10.1111/geb.12468

Svenning, J.C., Skov, F. 2004. Limited filling of the potential range in European tree species. Ecol Lett, 7(7):565-573. https://doi.org/10.1111/j.1461-0248.2004.00614.x

Tessarolo, G., Ladle, R., Rangel, T., Hortal, J. 2017. Temporal degradation of data limits biodiversity research. Ecol Evol, 7:6863–6870. https://doi.org/10.1002/ece3.3259

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

Vandepitte, L., Bosch, S., Tyberghein, L., Waumans, F., Vanhoorne, B., Hernandez, F., De Clerck, O., Mees, J. 2015. Fishing for data and sorting the catch: assessing the data quality, completeness and fitness for use of data in marine biogeographic databases. Database (Oxford). 2015:bau125. https://doi.org/10.1093/database/bau125

Varela, S., Anderson, R.P., García-Valdés, R., Fernández-González, F. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography. 37 (11):1084-1091. https://doi.org/10.1111/j.1600-0587.2013.00441.x

Velazco, S.J., Rose, M.B., de Andrade, A.F., Minoli, I., Franklin, J. 2022. flexsdm: An r package for supporting a comprehensive and flexible species distribution modelling workflow. Methods Ecol Evol. 13: 1661–1669. https://doi.org/10.1111/2041-210X.13874

Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M., Willis, K.J. 2005. Conservation biogeography: assessment and prospect. Divers Distrib. 11(1):3-23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Yang, Y., Ferguson, D.K., Liu, B., Mao, K., Gao, L., Zhang, S., Wan, T., Rushforth, K., Zhang, Z. 2022. Recent advances on phylogenomics of gymnosperms and a new classification. Plant Divers. 44(4):340-350. https://doi.org/10.1016/j.pld.2022.05.003

Zhu, G., Gariepy, T.D., Haye, T., Bu, W. 2017. Patterns of niche filling and expansion across the invaded ranges of Halyomorpha halys in North America and Europe. J Pest Sci. 90(4):1045-1057. https://doi.org/10.1007/s10340-016-0786-z

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesoon, S., Wengström, N., Zizka, V., Antonelli, A. 2019. CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. Methods Ecol Evol. 10(5):744-751. https://doi.org/10.1111/2041-210X.13152