



## OPEN ACCESS

### EDITED BY

Mikolo Yobo Christian,  
National Center for Scientific and  
Technical Research (CENAREST), Gabon

### REVIEWED BY

Charles Marsh,  
National University of Singapore,  
Singapore  
Bishal Majhi,  
Govind Ballabh Pant National Institute of  
Himalayan Environment and Sustainable  
Development, India

### \*CORRESPONDENCE

Brice B. Hanberry  
✉ brice.hanberry@usda.gov

RECEIVED 26 March 2026

REVISED 17 July 2026

ACCEPTED 23 July 2026

PUBLISHED 19 August 2026

### CITATION

Hanberry BB (2026) Generating data  
ranges for species with merged kernels  
and aggregated polygons from  
georeferenced records.  
*Front. Conserv. Sci.* 7:1839361.  
doi: 10.3389/fcsc.2026.1839361

### COPYRIGHT

© 2026 Hanberry. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License](#)  
(CC BY). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance  
with accepted academic practice. No  
use, distribution or reproduction is  
permitted which does not comply with  
these terms.

# Generating data ranges for species with merged kernels and aggregated polygons from georeferenced records

Brice B. Hanberry \*

USDA Forest Service, Rocky Mountain Research Station, Rapid City, SD, United States

Range maps are time-consuming for experts to produce, but georeferenced records offer an alternative foundation for range mapping. To test data ranges from records for abundant species, I compared data ranges for 3,960 species globally to expert ranges, which are subjective approximations with errors. I produced kernels that contained 99% of georeferenced records, contributing high-certainty range areas limited to sampling extents. To extend kernels, I merged kernels with aggregated polygons of records, which provided undersampled areas. Selecting one of six different aggregation distances ranging from 200 to 5,000 km for complex images was based on supervised classification, guided by proximity to kernels and resemblance to International Union for Conservation of Nature (IUCN) expert ranges. Maximizing overlap with expert ranges using intersection metrics resulted in similar scores to supervised image classification. Using supervised classification, median intersection of data ranges with expert ranges (relative to expert range areas) varied by taxa, from 86% for mammals, 91% for anurans and reptiles, 95% for birds, and 93% for plants. As a baseline, median intersection percentage was 93% for IUCN ranges with another source of expert range maps (Mammal Diversity Database). Data range areas were larger by factors (median) of 1.2 to 1.4 for animal taxa and 2.0 for plants (driven by non-native plant locations) than IUCN ranges, resulting in the median intersection relative to total union area of 59% to 66% for animals and 41% for plants. For IUCN ranges with  $\geq 95\%$  intersection with georeferenced records, data ranges were greater in area by a factor of 1.1 (median), but for IUCN ranges with  $< 80\%$  intersection with georeferenced records, data ranges were greater by a factor of 2.2. Because of range complexity, simple rulesets based on information from records and kernels are unlikely to provide consistent classification of aggregation distances, and nonlinear modeling is the next development phase for mapping of species without ranges. Data ranges, which overlapped with most of the expert ranges and 99% of records, can supply provisional range maps, with selection of aggregation distances based on expertise or modeled distance, to assist conservation planning.

### KEYWORDS

anurans, birds, data-driven, IUCN, mammalian diversity database, mammals, plants, reptiles

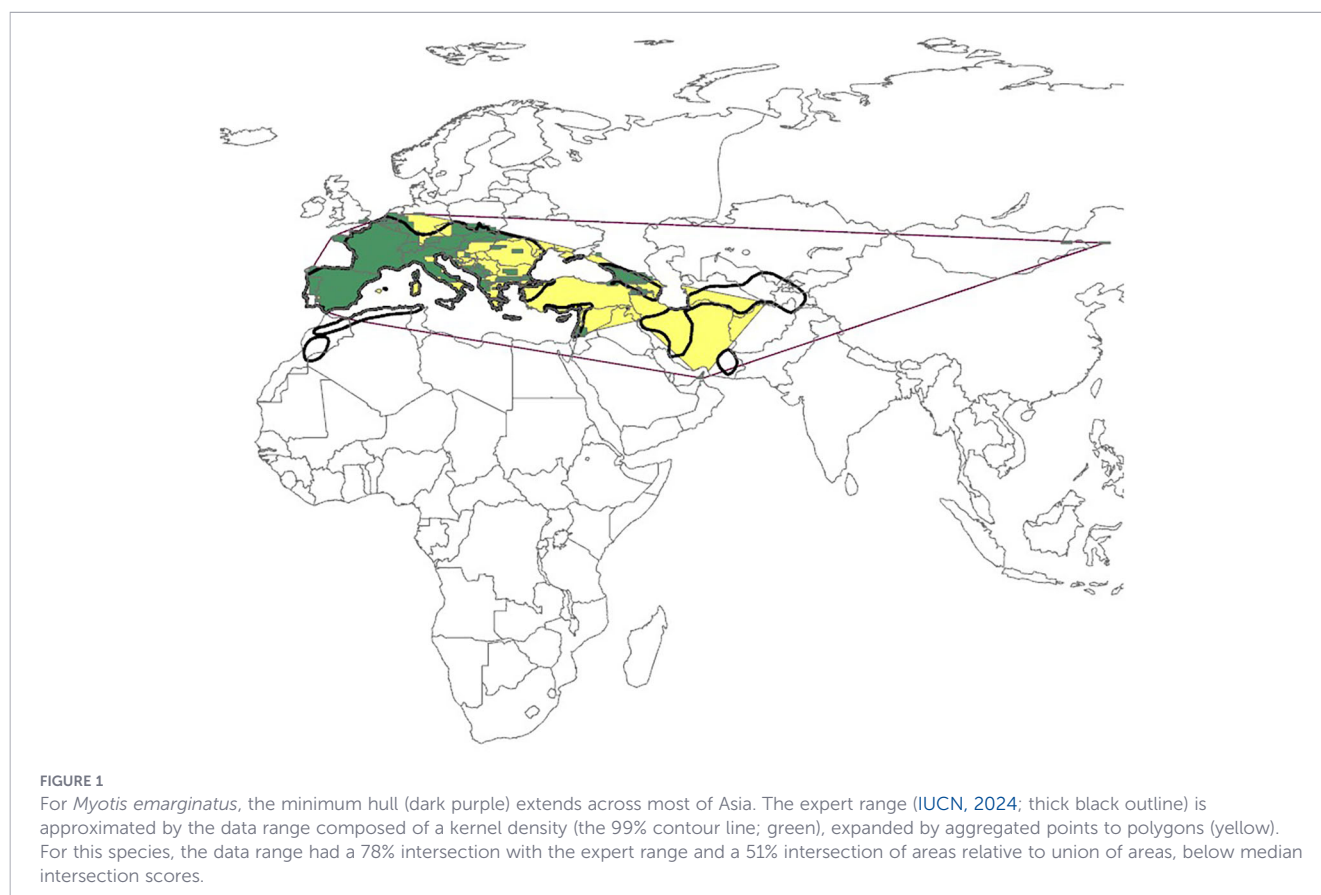
# 1 Introduction

Range maps are subjective approximations of species ranges based on the best knowledge available to experts whereas georeferenced records are accurate locations of species occurrences but limited by sampling. Range maps are basic units to understand species and species diversity, resulting in numerous applications for conservation, research, and monitoring (Cantú-Salazar and Gaston, 2013). Ranges offer a coarse representation of the maximum extent of potential occurrence for species, by connecting the outermost locations known to encompass the species during a time interval (Cantú-Salazar and Gaston, 2013; Jetz et al., 2012). Customarily, experts have defined range maps based on compiled knowledge about species with progression in published range maps over time (Fourcade, 2016; Marsh et al., 2022). The process of range map development is seemingly unique by expert group and species, with undocumented details but extensive time commitment to develop (Fourcade, 2016; Marsh et al., 2022). Range maps contain information about species understood by experts through field experience, literature review, species accounts, species ecology, and available ecosystems and resources, which may be absent in databases of compiled georeferenced records. Specifically, although georeferenced records are highly accurate (Zizka et al., 2020; Hanberry, 2025a), georeferenced records usually undersample species ranges, depending on factors such as range area, accessibility of ranges to observers, and survey intensity by geographic regions (Cantú-Salazar and Gaston, 2013; Hanberry, 2026a). Expert-driven range maps offer useful approximations of measured reality, but

range maps developed by experts are subjective, representing undocumented time intervals and containing various errors, such as approximately 25% of range maps include less than 80% of georeferenced records (Hanberry, 2025b).

Data-driven range maps based solely on georeferenced records have been attempted by modeling minimum hulls, which represent the smallest single polygons to enclose the species records, similarly to initial development of expert-driven ranges (Gaston and Fuller, 2009). Hulls can encircle a specified percentage of the records, delimiting different extents. Nevertheless, hulls generally encompass too large an area due to poorly fitting simple polygons (Figure 1; Palacio et al., 2021). Hulls sometimes are modified for better fit to complex shapes with cut-out holes and sides, for example, by applying interpolation (Palacio et al., 2021). After modifying hull shapes by environmental masks of climate and elevation, Aronsson et al. (2024) calculated 52% overlap in the combined area (i.e., the intersected areas divided by the union of areas) between data-driven ranges and expert ranges for 7,385 non-marine bird species (resident or short-distance migrants with spatially connected breeding and wintering ranges). Therefore, Aronsson et al. (2024) concluded that data-driven methods could not replicate expert-driven range maps.

However, options to create data-driven ranges include more than minimum hulls amended for better fit (Figure 1). Kernels (mathematical functions) can estimate densities of coordinates with a smoothing parameter to distribute locations in space, generating greater areas depending on point clustering and the influence of number of points over a surrounding area (Chacón and Duong, 2018).



Kernel densities are often used to display home ranges of individuals, as opposed to species ranges, but kernel species range estimates are possible (Worton, 1987). Indeed, home range delineation appears to have followed a similar progression as perhaps species ranges will in delineation: from expert-drawn minimum polygons, to modeled minimum hulls, to modifications to improve fit of minimum hulls, and eventually arriving at kernel density estimation (Worton, 1987). Nonetheless, while very fitted in shape to sampled areas, providing high-certainty areas for species ranges, kernel densities remain restricted in coverage by undersampled species distributions. Likewise, representative species distribution models do not model beyond conditions provided by georeferenced records, the sampled areas of species ranges (Hanberry, 2026a; Hanberry, 2026b).

Aggregating points into polygons can provide comprehensive and complex coverage, encompassing as many multifarious polygons as necessary, without the oversimplification of a single polygon of limited shape. As a feature, a parameter for aggregating is specification of the distance within which points will be considered part of the same cluster, resulting in tightly fitted complex shapes to increasingly wide coverage (ESRI, 2025). Indeed, with aggregated points to polygons, the kernels may not be necessary to provide the species sampled areas because the aggregated polygons cover the georeferenced records. However, kernels represent known, sampled areas of occupancy, which also assists selection of aggregation distance. Aggregated polygons at different distances can be generated, merged with kernels, and then one aggregation by distance can be selected for the best fit, in terms of shape, connectedness, and coverage relative to known, certain georeferenced records. Even if expert range maps are available to guide selection of best fit for coverage beyond the extent of georeferenced records (i.e., the kernel), expert ranges are subjective and contain errors (Hanberry, 2025b).

Whereas range maps previously have been generated by expert teams, one approach for ranges is to start with minimum hulls, or large polygons to connect the outermost locations known to contain the species, and then reduce the polygons for better fit (Aronsson et al., 2024). Alternatively, my overall objective was to develop and validate a standardized methodology for mapping species ranges from georeferenced records by combining kernel density estimation with aggregated polygons to produce distribution maps comparable to expert-derived reference maps. Specific objectives were (1) to generate sampled areas of species ranges from kernel density estimates encompassing approximately 99% of georeferenced occurrence records and then extend kernel-based ranges using aggregated polygons and determine the optimal aggregation distance, (2) to compare selection of the best-fitting range map through supervised classification with automated approaches based on the Jaccard and Tversky indices, (3) to evaluate the agreement between the generated range maps and International Union for Conservation of Nature (IUCN) expert range maps for 3,960 species and also to compare another set of expert range maps from the Mammal Diversity Database (MDD; Marsh et al., 2022) to establish the variation among expert range maps and as an independent validation of the data range maps, and (4) to identify the factors explaining differences in area between data-driven and expert-derived range maps. The process to generate data-driven range

maps will enable provisional range mapping for species lacking expert maps and support future predictive modeling of species ranges.

## 2 Methods

### 2.1 Expert range maps for comparison with data-driven range maps

Range maps of terrestrial mammal, anuran, squamate reptile, and plant species were from IUCN (2024), for extant ranges of available species. Range maps for bird species, resident-only extant ranges, were from BirdLife International (2024), which share range formats with IUCN. For a baseline contrast, expert range maps for mammal species were obtained from the MDD (Marsh et al., 2022), if they matched IUCN species names and were from the most abundant orders (Artiodactyla, Carnivora, Chiroptera, Diprotodontia, Eulipotyphla or Soricomorpha, Lagomorpha, and Rodentia).

### 2.2 Georeferenced records of species

Georeferenced records occurred during 1990 to 2024 with basic processing (Hanberry, 2025a) for unique terrestrial coordinates, and basis of record of preserved specimens, human observations, or observation [Global Biodiversity Information Facility (GBIF), 2024; GBIF occurrence downloads: <https://doi.org/10.15468/dl.ewt6xj>; <https://doi.org/10.15468/dl.9kafm2>; <https://doi.org/10.15468/dl.u8exef>; <https://doi.org/10.15468/dl.w2krf2>; <https://doi.org/10.15468/dl.nzcwcd>; <https://doi.org/10.15468/dl.k6k2cg>; <https://doi.org/10.15468/dl.cqsusc>; <https://doi.org/10.15468/dl.neehbu>; <https://doi.org/10.15468/dl.nr9yek>; <https://doi.org/10.15468/dl.6bkr63>; <https://doi.org/10.15468/dl.92s77y>; <https://doi.org/10.15468/dl.uaskaz>; <https://doi.org/10.15468/dl.jhu36x>; <https://doi.org/10.15468/dl.6fehdK>]. Location distance uncertainties were filtered to unknown distance uncertainties alone or combined with <1 km for bird species (due to the preponderance of unknown distance uncertainties), <10 km for plants (immobile), and <36 km for mammals, squamate reptiles, and anurans (these location distance uncertainties are suitable for ranges; Hanberry, 2025b). To have robust samples, I retained species that had at least 400 georeferenced records. To ensure that range map creation was not covered by species kernels alone without the need for aggregated polygons, I selected species that had IUCN ranges  $\geq 10,000$  km<sup>2</sup>. In total, 732 mammal species, 367 anuran species, 528 reptile species, 1,504 bird species, and 829 plant species were represented (Supplementary Table 1). While approximately 5% to 10% of scientific names were different between IUCN and GBIF, less than 5% of IUCN scientific names included ranges for both an accepted species name and synonym, according to GBIF. In these cases, I merged IUCN ranges of the subspecies (according to GBIF) to match GBIF. However, taxonomies are complex and not always compatible, such that complete reconciliation may not have occurred, and an older species classification may be employed (due to novel “systematic syntheses”; Franz and Sterner, 2018). As for

sampling bias, or clustered records, greater sampling generally is beneficial for range mapping to confirm that species occur at those locations and avoid error. For the kernels, 1% of records were excluded, with the possibility that areas of concentrated sampling prevented registration of less sampled areas (clustered samples can be a small issue for species distribution models because modeling algorithms will concentrate the models according to sample concentration; Hanberry, 2025a). However, the aggregated polygons provide another level of redundancy to record areas missed by kernels for these data-driven ranges.

## 2.3 Generation of kernel densities from georeferenced records

Many choices for generating spatial two-dimensional kernel densities exist. In the Comprehensive R Archive Network (R Core Team, 2025), either numerous packages are specific for kernel density estimation or kernel density estimation is one of the many functions; additionally, a variety of different parameter arguments are available. The package *ks* (Duong, 2007) generates well-fitted, complex polygons, which can be extracted at 99% contour lines to encircle approximately 99% of georeferenced records, and ease of use without specifications (i.e., due to suitable default specifications, including bandwidth) or generating errors. The 99% threshold will contain most records and filter out any outliers that may contain potential error due to species misidentification outside of ranges or vagrant individuals, and extending kernels beyond primary sampled areas (Hanberry, 2025a). Bias-variance trade-offs can result in a range of oversmoothed and undersmoothed kernels (Fieberg, 2007). Smaller sample sizes typically result in larger, smoother, more generalized kernels (due to increased bandwidths automatically selected by the *ks* package to avoid modeling random patterns; Fieberg, 2007). Other parameter changes also create more generalized or specific shapes. Extracting contour lines to raster layers at a resolution of 0.3° (i.e., approximately 33 km at the equator, approximately matching the distance uncertainty of the coordinates and within the range of values for macroecological studies; Hurlbert and Jetz, 2007; Hanberry, 2026a), instead of to vector files, maximized coverage, while still representing isolated areas, and with fast processing time. The raster layers did result in stairstep edges, which can be smoothed post-modeling, but the kernels were retained as grid cell shapes to aid visual comparisons in this study and identify the high-certainty range areas of the kernel from the lower-certainty range areas of the aggregated polygons. A rare issue is that package *ks* can return an empty global extent when a few georeferenced records are distant from the majority of records. This problem does not appear resolvable by adjusting the bandwidth within the package. A simple contingency for empty outputs is to supply the package *MASS* (Venables and Ripley, 2002), which applies a coarser default bandwidth, as a replacement. As an additional related consideration, for species that occur on multiple continents, I generated kernels for each continent (minimum samples size of 25), to account for uneven sampling intensity that otherwise may result in omission of less well-sampled continents, and then combined kernels (Supplementary File 2). The same rare issue of patchy

records that occur throughout the continents may result in kernels that are too coarse, covering large extents or even continents relative to the aggregated polygons. For cases (i.e., 5 species out of 3,960) where a suitable kernel output is intractable, the aggregated polygons alone can be used, guided by georeferenced records, which can be buffered.

## 2.4 Generation of aggregated polygons from georeferenced records

Options for aggregating points into a range of polygon shapes and sizes appear more limited, to my knowledge, although a multitude of clustering packages are available. Therefore, I used ArcGIS Pro (ESRI, Redlands, CA, USA), which has an aggregate points function in the cartography tools (ESRI, 2025). This function creates polygons around clusters of three or more points (i.e., necessary to create a polygon rather than a line) within the specified aggregation distance. For this method, small samples or thinned samples may not be able to aggregate points in undersampled locations outside of the kernel, resulting in lack of coverage. For most operations, including this one, ArcGIS Pro works better with projected data layers. I projected the georeferenced records into Eckert IV, a global area-preserving projection (Snyder, 1987). After examining aggregated polygons by 14 distances ranging from 200 to 5,000 km for a pilot test of 50 species, I selected six distances of 200, 250, 500, 1,000, 2,000, and 5,000 km that generally produced both noticeable separation between outputs and two adjacent distances that were similar enough to be potential options for ranges, with one more fitted in shape and one providing greater coverage.

## 2.5 Classification of aggregation distances for merged kernels and polygons and comparison to expert range maps

After merging the kernels with six different distances of aggregated polygons with IUCN ranges, I selected the distance that best represented the information from high-certainty range areas of the sampled extents of the kernels, maximized overlap with IUCN ranges, and minimized areas outside of the IUCN ranges. Selection of the best-fitting range map is structurally analogous to supervised classification, which is common for images that often are too complex for unsupervised classification using scores, simple thresholds, or other automatic approaches such as clustering (Cheng et al., 2020). For mammal species, I illustrated a contrasting automated approach to time-consuming supervised classification. I used the intersection over union score (Jaccard Index; Gilbert, 1884) and Tversky score (a weighted version to penalize commission error with standard values of  $\alpha = 0.7$ ,  $\beta = 0.3$ ; Tversky, 1977; Yeung et al., 2022) to maximize intersection with IUCN range maps. I also examined constant aggregation distances and increased the supervised classification by one distance (e.g., 200 to 250 km, with the exception of the maximum distance of 5,000 km). For supervised classification of all species, I calculated area of intersection between the data ranges and IUCN ranges, relative to area of IUCN ranges, and the intersection over union (i.e., the shared intersection area combined with area outside of the intersection), resulting in two

scores. Spatially comparing data range maps, or species distribution models, to range maps is the most rigorous assessment of spatial representation (Hanberry, 2026b). As a baseline contrast and independent validation, I also intersected IUCN range maps with MDD range maps and calculated intersection scores, relative to area of IUCN ranges, and likewise, data range maps with MDD range maps and calculated intersection scores, relative to area of MDD ranges. Additionally, I calculated the uncertainty ratio of the area of the low-certainty aggregated polygons, excluding intersection with high-certainty kernels, relative to the total data range area. However, the low-certainty areas represent data-poor locations where there are few georeferenced records and provide the critical range that is missing in the kernels.

## 2.6 Factors that influence range areas

In addition, I determined factors that could influence data range areas relative to IUCN range areas. Variables included omission errors in IUCN range areas, using percentage of georeferenced records that overlapped with IUCN ranges (Hanberry, 2025b), divided into four bins that approximated quartiles, with the thresholds at 80%, 90%, and 95% intersection between georeferenced records and range maps (that is, for 25% of species range maps, intersection with georeferenced records was below the lower quartile of 80%). Another variable was range expansion beyond native ranges. I merged the species with the Global Register of Introduced and Invasive Species (GRIIS; Pagad et al., 2022) to determine whether species had a known non-native or invasive (i.e., documented disservices) status. The status of species that did not match with the GRIIS was presumed primarily native, albeit potential taxonomic incompatibilities may have prevented matching. A third factor was range contraction, based on an IUCN Red List category of critically endangered, endangered, vulnerable, or near threatened. A fourth factor was the influence of the aggregation distance. That is, perhaps smaller or larger aggregation areas diverged more from IUCN range areas.

## 3 Results

### 3.1 Supervised classification compared to automated or standard selection of aggregation distances

For illustration of supervised classification compared to an automated approach to fit aggregation distances to species for abundantly record mammal species, scores relative to the expert range maps and areas of the merged kernels and aggregated polygons were similar, whether maximizing the intersection over union score and Tversky score or with classified supervision. Median intersection over union scores were 0.61, 0.62, and 0.59, respectively, for maximizing intersection over union score, Tversky score, and classified supervision. Median Tversky scores were 0.75, 0.74, and 0.73, respectively, for maximizing intersection over union score, Tversky score, and classified supervision. Median areas were 2,352,307, 2,473,168, and 2,691,180 km<sup>2</sup> for intersection over union

score, Tversky score, and classified supervision, respectively. However, maximizing intersection with expert range maps may have no support from georeferenced records, resulting in potential commission error (see examples in Figure 2).

Relative to varying species-specific distances, for one constant aggregation distance, in terms of scores and range areas, the default distance of 1,000 km most resembled range maps, reducing the median intersection over union score to 0.50 and Tversky score to 0.67, with a similar median area of 2,685,812 km<sup>2</sup> (Figure 3). A default distance of 200 and 250 km reduced the median intersection over union score to 0.41 and Tversky score to 0.62, with median areas of 1,416,061 and 1,471,224 km<sup>2</sup>, respectively. A default distance of 500 km reduced the median intersection over union score to 0.45 and Tversky score to 0.65, and median area to 1,946,931 km<sup>2</sup>. A default distance of 2,000 km reduced the median intersection over union score to 0.50 and Tversky score to 0.67, and also increased median area to 3,558,890 km<sup>2</sup>. A default distance of 5,000 km reduced the median intersection over union score to 0.49 and Tversky score to 0.60, and also increased median area to 4,361,224 km<sup>2</sup>. Relative to aggregation distances fitted to species by supervised classification, increasing the supervised classification by one class reduced the median intersection over union score to 0.55 and Tversky score to 0.66, and also increased median area to 3,426,663 km<sup>2</sup>.

### 3.2 Comparison of data-driven range maps, selected by supervised classification, and expert range maps

For data range maps with the aggregation distance selected by supervised classification, for all abundantly recorded species, area of IUCN expert range maps and data-driven range maps had a linear relationship ( $R^2 = 0.66$ ; Figure 4), and intersection of data ranges relative to area of IUCN range map was high. Median intersection in area was 85.8% for 732 mammal species, 91.3% for 367 anuran species, 91.2% for 528 reptile species, 94.5% for 1,504 bird species, and 92.7% for 829 plant species (Figure 5). There was not a linear relationship between the number of georeferenced records and area of data-driven range maps ( $R^2 = 0.07$ ).

For a baseline reference of variation among expert range maps and independent validation of the data ranges, median intersection between IUCN range maps and MDD range maps, also derived by experts, was 93.1% for 464 mammal species and the median intersection between data range maps and MDD range maps was 84.7%. The intersection for this subset of mammal species between IUCN range maps and data range maps was 87.2%. Overall, areas were similar for IUCN range maps and MDD range maps.

Nevertheless, most of the data ranges covered greater areas than the IUCN range maps, by factors of 1.2 (median) for mammal species, 1.3 for anurans, 1.2 for reptiles, 1.4 for birds, and 1.9 for plants. Given 3,960 species, greater areas of data range maps were due to a variety of reasons, primarily the influence of the sampled area of the georeferenced records. Percentage overlap of IUCN ranges with the georeferenced records affected the relative areas similarly for all taxa, albeit with different effect sizes (Figure 6). For IUCN ranges with  $\geq 95\%$  intersection with georeferenced records, data range maps were greater in area by a factor of 1.1 (median)

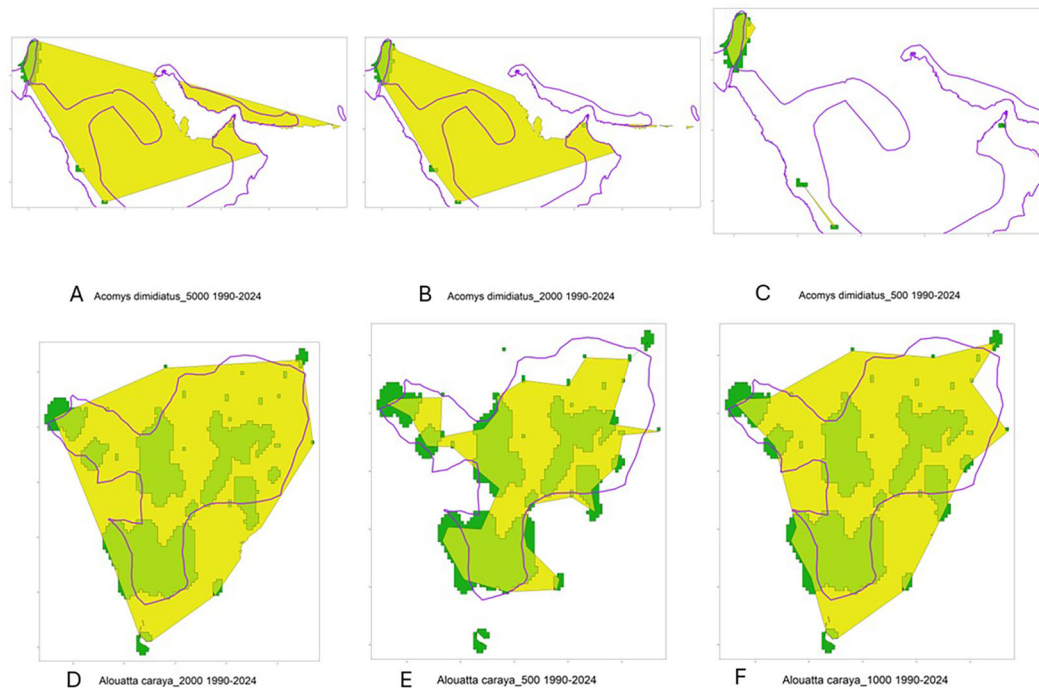


FIGURE 2

Mammal species of *Acomys dimidiatus*, maximized to fit the expert range map (outline) by the intersection over union and Tversky score with an aggregation distance of 5,000 km [yellow, merged with green kernel; (A)], supervised classification with an aggregation distance of 2,000 km, with liberal connection of the kernel (green) by the aggregated polygons (yellow), guided by the expert range map (B) and the conservative alternative option of an aggregation distance of 500 km, a moderate standard distance [(C), very similar to aggregation distance of 1,000 km]. Mammal species of *Alouatta caraya*, maximized to fit the expert range map (outline) by the intersection over union score with an aggregation distance of 2,000 km [yellow, merged with green kernel; (D)], maximized to fit the expert range maps by the Tversky score with an aggregation distance of 500 km [yellow, merged with green kernel; (E)], and supervised classification with an aggregation distance of 1,000 km [yellow, merged with green kernel; (F)].

than IUCN range maps. For IUCN ranges with  $\geq 80\%$  but  $< 90\%$  intersection with georeferenced records, data range maps were greater by a factor of 1.5. For IUCN ranges with  $< 80\%$  intersection with georeferenced records, data range maps were greater by a factor of 2.2. Because of greater areas of data ranges, area of intersection relative to area of union of data ranges and IUCN ranges had reduced scores (Figure 5). Median values were 59.3% for

mammals, 64.1% for anurans, 66.4% for reptiles, 62.4% for birds, and 41.1% for plants. For example, *Myotis emarginatus* (Figure 1) had 78% intersection with the area of the IUCN range and 51% intersection of areas relative to union of areas. Indeed, thousands of examples are available (see the Data availability statement). Overall, mammals and birds exhibited both the largest median data ranges ( $> 2.1$  million  $\text{km}^2$ ) and the greatest median uncertain range ratios of

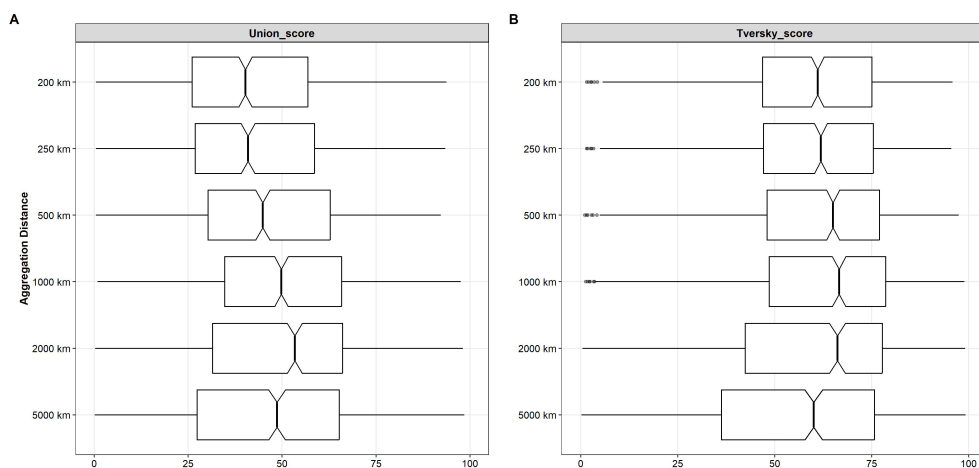


FIGURE 3

Intersection scores of intersected area of IUCN expert ranges and data ranges at different aggregation distances relative to the union of IUCN and data range areas (i.e., shared area; union score (A) and weighted union score, the Tversky score (B)).

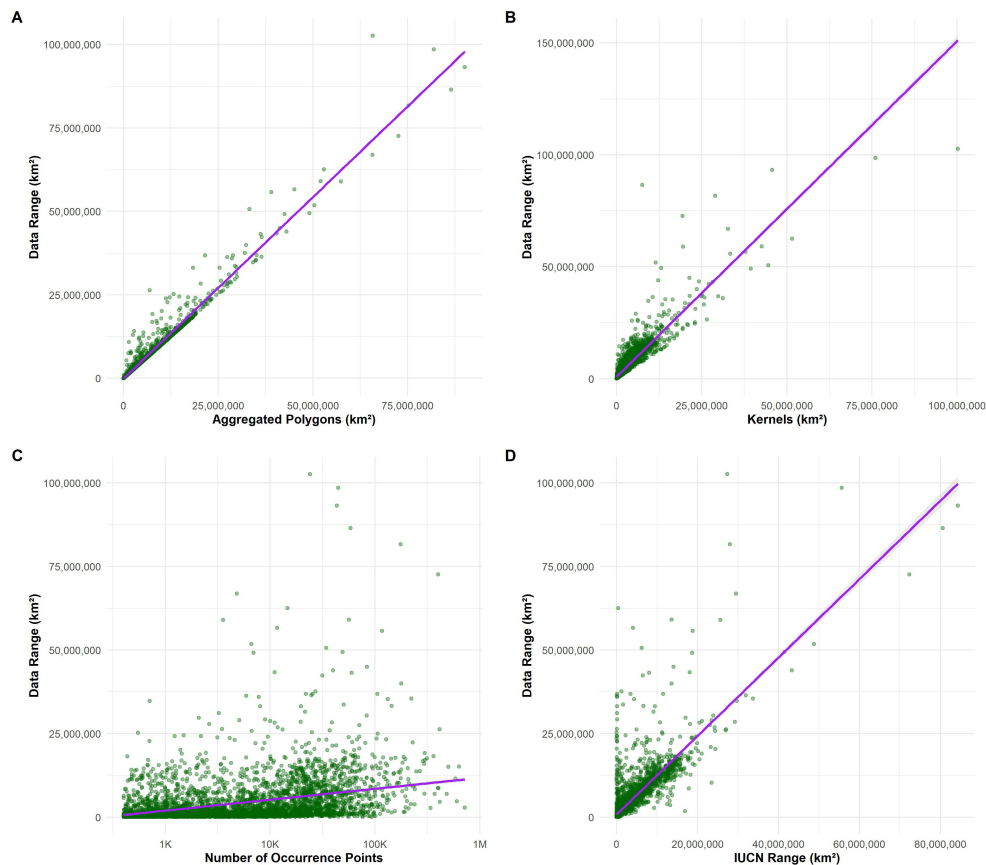


FIGURE 4

Area of data ranges (one aggregation distance per species selected by supervised classification) compared to (A) area of aggregated polygons at the selected distance, (B) area of kernels, (C) number of georeferenced records (logarithmic scale but labeled with untransformed number of records), and (D) area of IUCN expert ranges. Purple lines represent linear model fit.

46.0% and 43.7%, respectively (i.e., percentage area where the aggregated polygons extended high-certainty kernels). Reptiles had uncertain range ratios of 39.1%, plants had ratios of 36.6% (a small median range size of 633,466 km<sup>2</sup> but a large mean range area of 4,043,456 km<sup>2</sup>; see below for non-native species), and anurans had ratios of 35.9%.

### 3.3 Other factors that influence range areas

Non-native or invasive status was identified for 843 species, increasing data range areas particularly for plants, for which data range maps were greater in area than IUCN ranges by a factor of 5.2 for species with invasive status and 2.6 for species with non-native status from a baseline of 1.5 for species with unknown status (Figure 7; unknown status was presumed primarily native). Contrarily, for 345 species with an IUCN Red List status of critically endangered, endangered, vulnerable, or near threatened, data range maps covered greater areas than species without a listing status. However, this was not the trend for reptile species. One explanation for greater areas of data range maps relative to IUCN range maps for decreasing species is related to recent rapid declines, such as for Australian species (Figure 8). Another trend generally was increased ratios of areas for data ranges relative to IUCN ranges as the aggregation distance decreased (i.e., smaller areas in data ranges

covered greater areas relative to IUCN range maps; Figure 9). Nonetheless, range maps from plant species did not follow this trend, as aggregation distances of 1,000 and 2,000 km had the greatest divergence in ratios of areas.

## 4 Discussion

### 4.1 Key findings

Range maps provide fundamental information about species, with applications for management, conservation, and research, such as species distribution modeling (Gaston and Fuller, 2009). As an alternative to expert range maps and extreme commission error from data-driven convex hulls, the method developed here builds out ranges from kernels, which are high-certainty range areas from georeferenced records, using aggregated polygons, to produce at least provisional products from georeferenced records, for abundantly recorded species with large ranges. The distance for aggregated polygons was selected by supervised classification using the kernels as support for connecting and extending samples area, along with expert range maps. To be clear, expert range maps are subjective, for undocumented time intervals, and not perfectly located, with errors (Hanberry, 2025b). Data-driven range maps

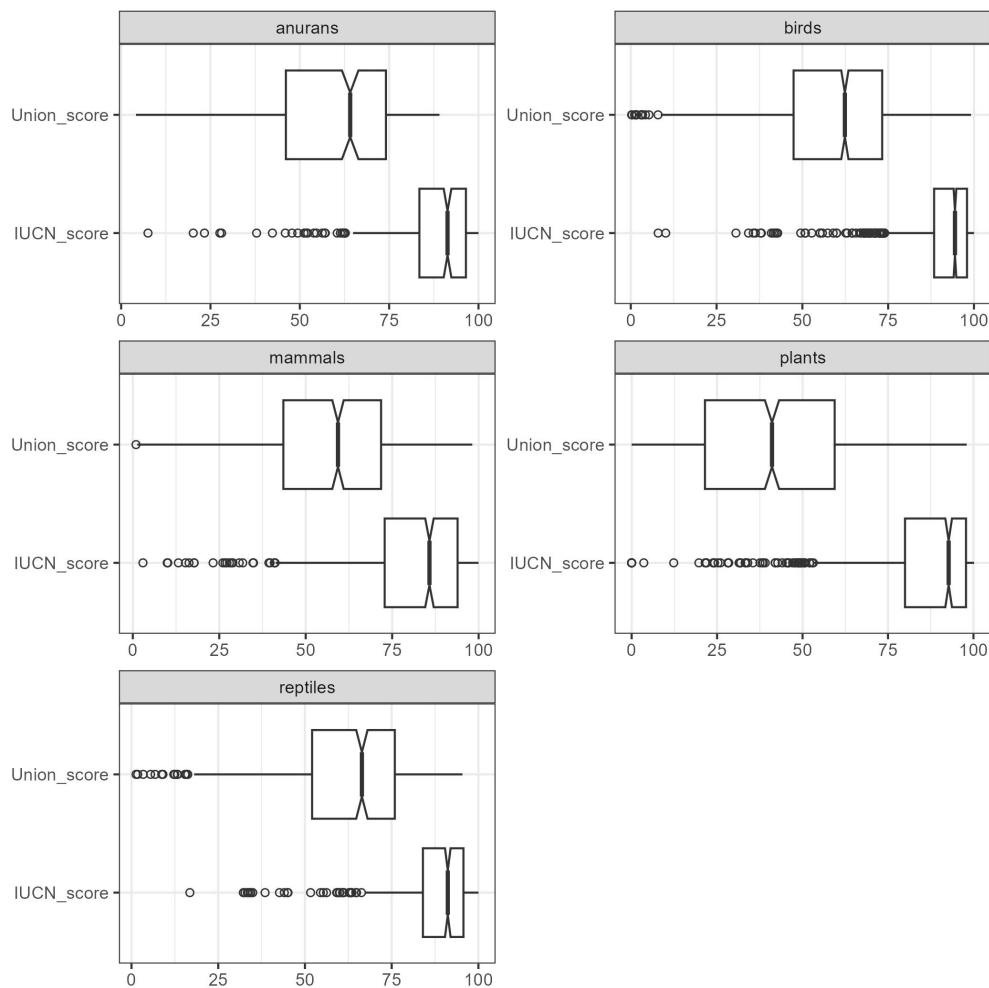


FIGURE 5

Intersection scores of intersected area of IUCN expert ranges and (final selected) data ranges relative to IUCN range area (IUCN score) and relative to the union of IUCN and data range areas (i.e., shared area; union score), for five taxa.

for abundantly recorded species covered most IUCN expert range maps (IUCN, 2024), with some mildly moderate omission error, with similar results for another, independent set of expert range maps. That is, 75% of data ranges intersected at least 73% of mammal species ranges, 84% of anuran and reptile species ranges, 88% of (resident) bird species ranges, and 93% of plant species ranges from IUCN. Data ranges covered greater areas, encompassing nearly all (i.e., 99%) georeferenced records as well as most of the area of IUCN range maps. The additional extent of data range maps resulted in intersection to union (combined area) percentages (median) of 62% for bird species, which was greater than the 52% value reported by Aronsson et al. (2024), albeit for different bird species. Aronsson et al. (2024) concluded that data-driven methods could not reproduce expert-driven range maps. However, for the method documented here, intersection over union scores seemingly were reasonable due to omission errors in expert range maps, indicated by lack of intersection with georeferenced records, and range expansion by non-native species and range contraction by declining species. Ranges produced by kernels (high-certainty range areas of the sampled extents) combined with aggregated polygons (lower-certainty range areas of the undersampled extents) appeared

successful, with greater range areas contributed by the aggregated polygons providing better coverage and smaller range areas providing better fitted shapes. Both the scores and representativeness of the data ranges indicated that data range maps generally approximated most of the expert ranges, while accounting for omitted areas in range maps. Because species ranges are complex, no simple threshold or rule adequately determined aggregation distances, as even maximizing overlap with expert ranges did not appear to provide the best fit. Instead, complex classification problems will require modeling solutions, as the next phase of development in future research.

## 4.2 Comparison of data ranges to expert ranges

Intersection of the data ranges (i.e., final merged kernel with aggregated polygons at one distance selected by supervised classification per species) with expert ranges varied by taxa, with median intersection percentages ranging from 86% for mammal species to 95% for bird species between data and expert ranges, relative to areas of expert ranges. On the whole, data ranges best matched bird

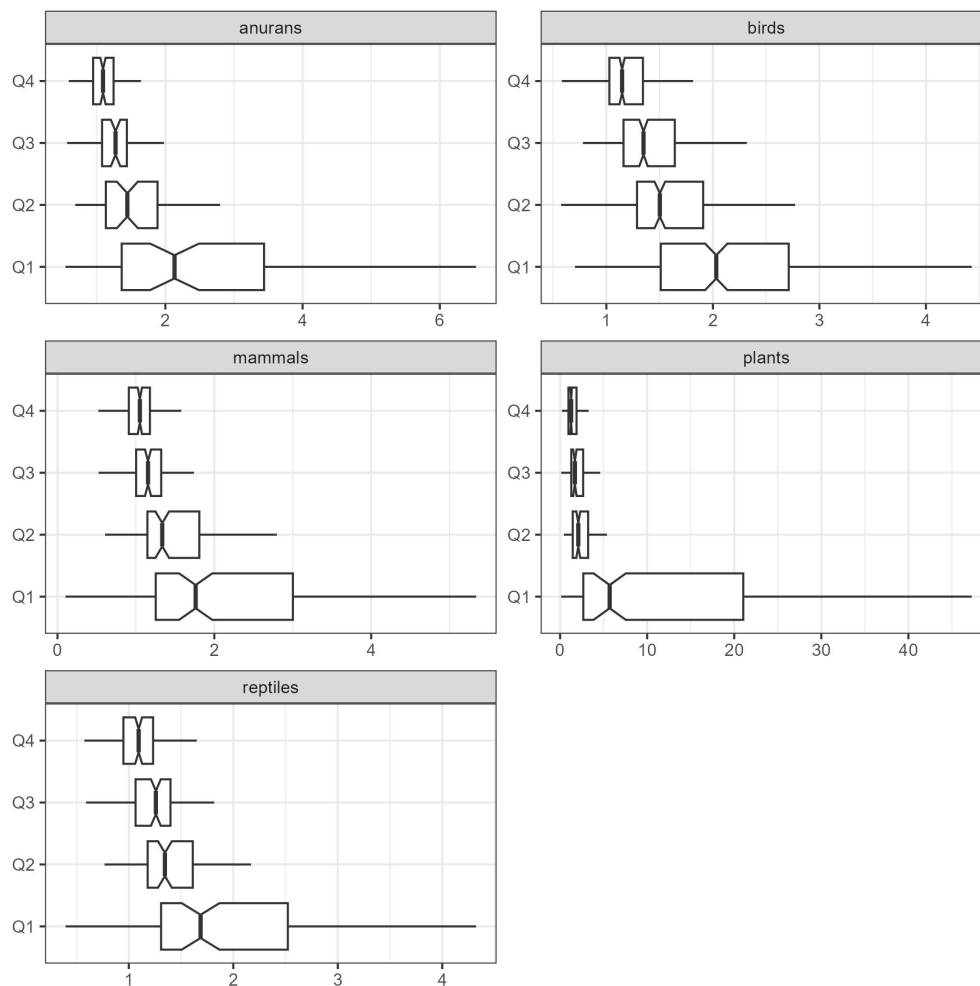


FIGURE 6

Ratios of (final selected) data range areas relative to IUCN expert range areas by four bins that approximated quartiles, with the thresholds at 80% (Q1 and Q2), 90% (Q2 and Q3), and 95% (Q3 and Q4), for percentage of georeferenced records that overlapped with IUCN ranges, or omission errors, for five taxa.

species ranges, taking into account also the union area, or area beyond IUCN ranges, and given that bird ranges also generally covered the greatest areas. Moreover, data ranges visually appeared to most easily fit expert range maps of bird species. Bird species ranges developed by experts perhaps may be most accurate in terms of fitting the shape of species distributions, with limited commission and omission error, due to the visibility of birds in the field and widespread human effort in observing birds (Hurlbert and Jetz, 2007). Conversely, ranges of mammal species were more challenging to replicate. This was reflected in the intersection scores and greater difficulty in selecting aggregation distances that best matched expert mammal ranges while upholding fidelity to the georeferenced records. Median intersection percentages between data ranges and expert ranges from mammal species were less than intersection percentages (93% median) of IUCN expert range maps with MDD expert range maps (Marsh et al., 2022). The MDD expert range maps were comparable in size with IUCN range maps, resulting in greater median intersection over union scores, compared with data range and IUCN expert range maps intersection over union scores. Data range areas were greater than IUCN range

areas, resulting in decreased intersection over union area scores, varying from 41% for plants to 66% for reptiles (median values).

However, overall decreased scores appeared to indicate reduction in omission errors, rather than gain in commission errors, in data ranges relative to expert ranges. Georeferenced records supply known species observations, and the data range maps encircled approximately 99% of records, due to 99% contour lines of the kernel species range estimates. Data range areas increased relative to decreased percentage of intersection with georeferenced records in expert ranges. Data ranges were greater in area by a factor of 1.1 (median) than IUCN range maps that intersected with  $\geq 95\%$  georeferenced records. In contrast, data range maps were greater by a factor of 2.2 than IUCN ranges that intersection with  $< 80\%$  georeferenced records. Expert range maps may not reflect some lesser-known range extents, such as for the woodland vole (*Microtus pinetorum*) in Canada (Canadian Wildlife Service, 2015). By design, IUCN range maps adhere to a precautionary principle to avoid inclusion of areas with uncertain or low probabilities of species occurrence (IUCN Standards and Petitions Committee, 2024).

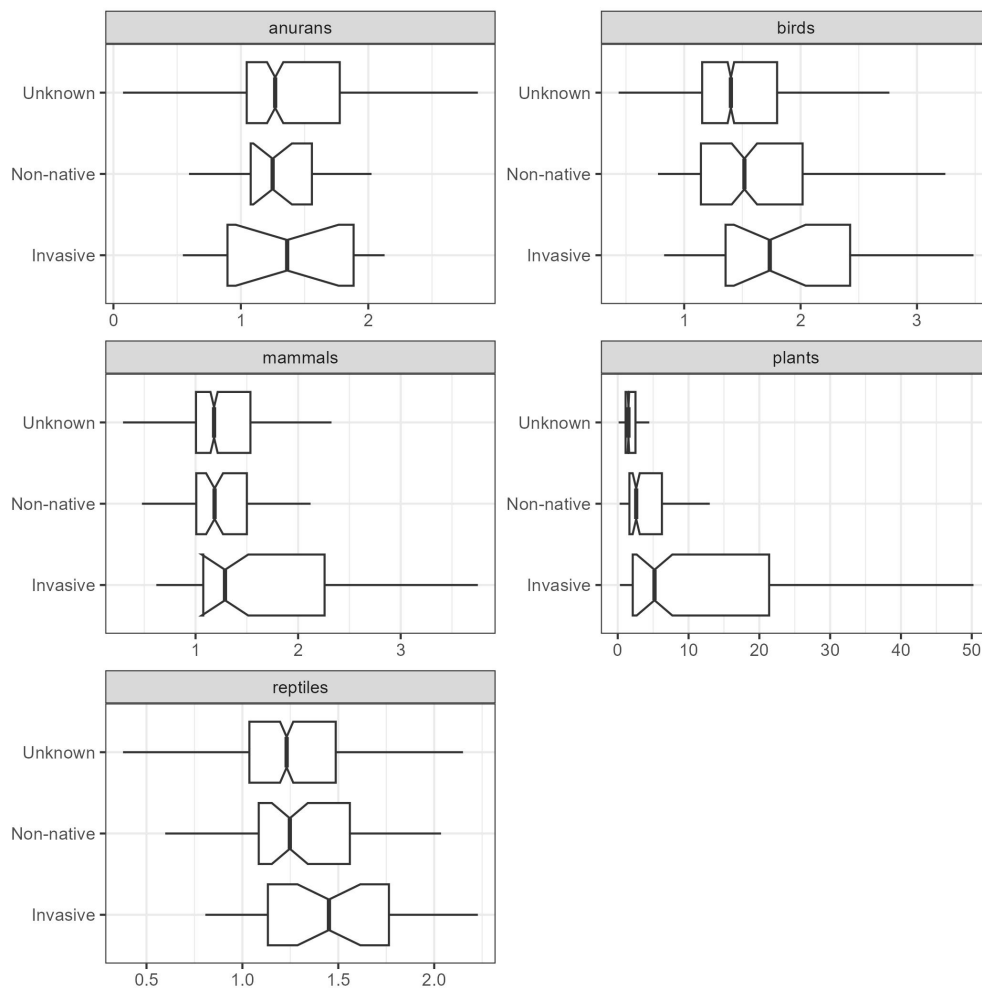


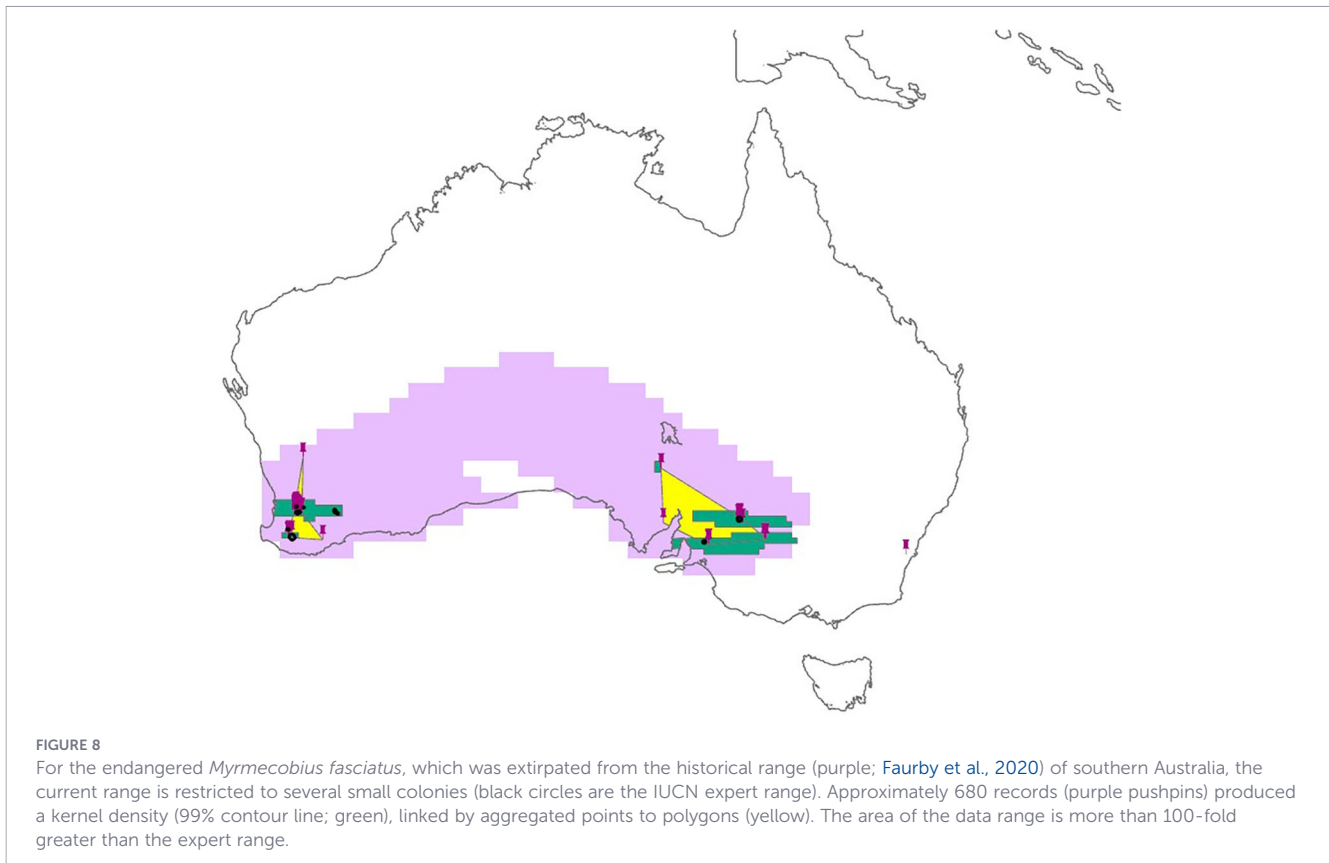
FIGURE 7

Ratios of (final selected) data range areas relative to IUCN expert range areas by status (Pagad et al., 2022) of invasive (documented ecosystem disservices), non-native, and presumed native (unknown), for five taxa.

Data ranges also encompassed range expansion, which amplified the overall trend of greater areas for data ranges than expert range maps. Expansion by non-native species was demonstrated particularly by plant species identified as non-native or invasive (i.e., with documented disservices; Pagad et al., 2022), which appeared to be the driving factor for larger plant areas in data-driven ranges. Furthermore, although species that did not have a non-native or invasive status (Pagad et al., 2022) were primarily within native ranges, without expansion, this may not be the case for some species. Species ranges are dynamic and specifically may be undergoing state transitions due to land use and disturbance changes (Hanberry, 2022). When examining mammal species closely, expert range maps for mammals did not cover recent expansions by mammals, which data ranges encompassed. Expansions represented both non-native species in non-native ranges (e.g., *Rattus norvegicus*) and also apparently by native species expanding beyond range maps within a country, most likely not transported by human vectors. For example, the data ranges incorporated recent rapid expansion of coyotes (*Canis latrans*) in Newfoundland, Canada, which would update the expert range maps (Blake, 2006). Likewise, opossums (*Didelphis virginiana*) are now

present in North and South Dakota, United States (Walsh et al., 2017). The nine-banded armadillo (*Dasypus novemcinctus*) is also moving northwards in the United States (Taulman and Robbins, 1996). Recent expansions may not necessarily be identified as expansion by non-native species, or may be a combination of both (e.g., *Procyon lotor*, *Sciurus carolinensis*, and *Sciurus niger*). The distinction between non-native species and native species sometimes is not definable or documentable.

Conversely, a third factor that influenced divergence in range areas for some species was recent range contractions. The specified interval for the data ranges was from georeferenced records since the year 1990, with rapid accumulation of records from community naturalists (e.g., 25% of mammal records were from iNaturalist (2024), which contributed less than 1% of records during the decade of the 1990s and exceeded 50% of records during 2023, when the number of observations was greater by a factor of 20 than during the year 1990). Following how IUCN employs range maps, I identified species potentially undergoing range contraction based on the IUCN Red List categories of critically endangered, endangered, vulnerable, or near threatened, although species of least concern may also be undergoing range contractions (Hanberry, 2021).



While the time frame is undocumented, expert range maps may not represent longer-standing ranges (i.e., georeferenced records since 1990) before recent rapid declines, as in Australia (Figure 6). The area of the data range for the endangered numbat (*Myrmecobius fasciatus*) was more than 100-fold greater than the IUCN range, which is limited to several small colonies. Both range maps are expressions over time for the species, which was extirpated from an even greater historical range (Faurby et al., 2018; Faurby et al., 2020). The data ranges were generally greater for species categorized of concern rather than least concern, except for reptiles, relative to areas of expert range maps.

Relatedly, the specified distance for aggregating points to polygons affected data range areas for species compared to expert range areas. In general, for species with data ranges that had smaller aggregation distances of 200 and 250 km, and thus smaller aggregated polygons and smaller ranges, area was greater by a factor of 1.5 to 1.6 (median) than IUCN range maps, whereas species with data ranges that had aggregation distances of 2,000 and 5,000 km were about equal in area with IUCN range maps. This was a result of intentional classification being more conservative as species ranges became larger and more widespread. For applications that require known internal range structure and porosity (Jetz et al., 2008; Cantú-Salazar and Gaston, 2013), with minimal commission error, smaller aggregations at distances less than 200 km may be desirable options. The data ranges with specified aggregation distances are constructed to provide potential external boundaries, with the embedded kernels available to isolate internal structure. Kernels supply internal structure for ranges, based on known occurrences, but will be incomplete due to sampling limitations

(Cantú-Salazar and Gaston, 2013). Indeed, the aggregated polygons extended range areas, contributing (median) 36% to 46% of data range area, varying by taxa.

### 4.3 Method development and application of data ranges

An advantage of the aggregated polygons is that ranges can be built out from core areas, rather than requiring contraction from limits that are too expansive, as occurs with minimum hulls. Even with 99% of records, minimum hull polygons typically are not well-fitted to distributions (Figure 1). Starting the range build from shapes based on minimum hulls, unlike kernels, is not likely to be successful without experts, as determined by Aronsson et al. (2024), or efficient for experts (Sanderson et al., 2002). For species without range maps, maps at moderate aggregation distances can be provisional or starting points for ranges, with selection and modification based on expert guides, and also classifications of aggregation distances can be applied by modeling. Incorporating expert validation could improve the ecological realism and reduce errors of the resulting data ranges.

Depending on user-specified distances, aggregation generates polygons that vary from fitted, small shapes to wide coverage. There is not one correct or best aggregation distance, because generally true ranges are unknown, but some distances are more reasonable than others if compared to expert opinion of IUCN ranges and compared to the kernels of georeferenced records. I used supervised classification rather than a metric (e.g., area of intersection relative to union) to achieve balance with IUCN ranges, but foremost

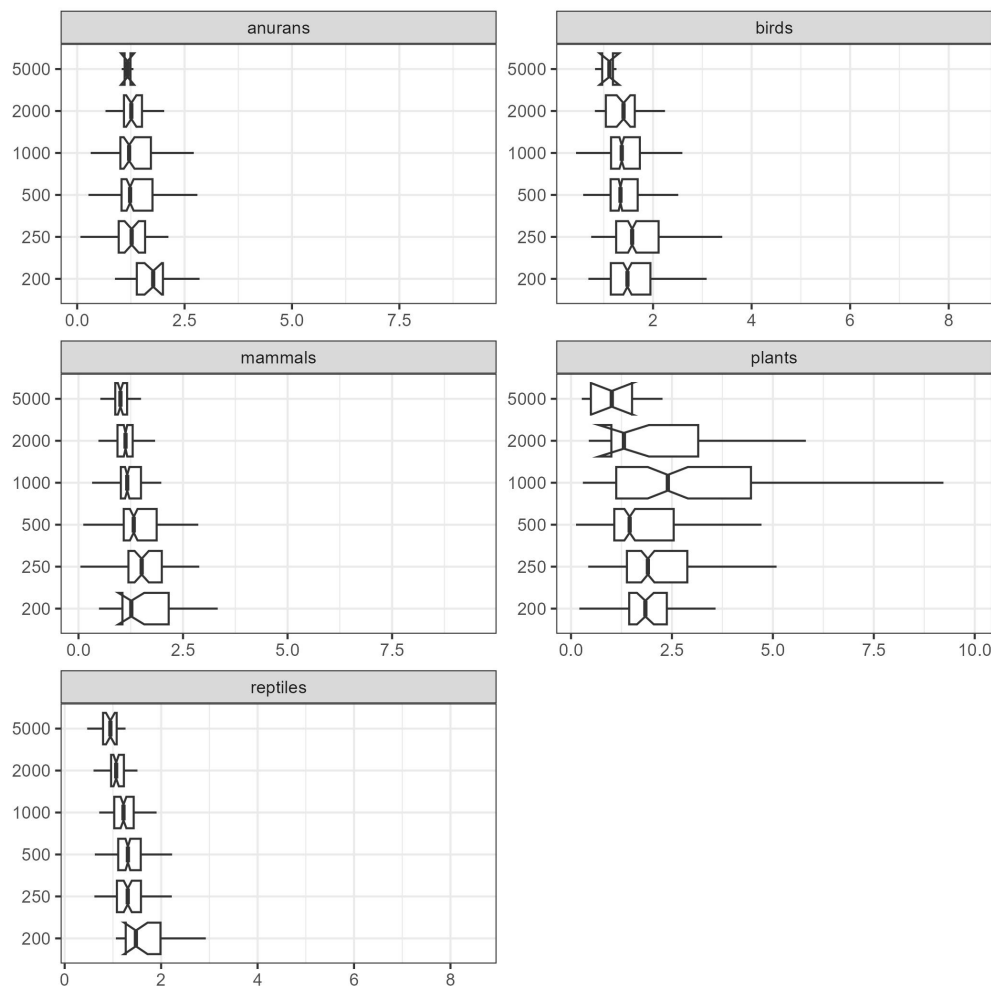


FIGURE 9  
Ratios of (final selected) data range areas relative to IUCN ranges as the aggregation distance (one per species).

respected the integrity and contribution of the kernels, which represent the known sampled area of georeferenced records. For example, if numerous patches from the kernels of georeferenced records occurred beyond IUCN ranges, connecting the patches through aggregation may be a more reasonable depiction of current species ranges than maximizing the match with IUCN ranges. Conversely, there may not be support from georeferenced records for maximizing intersection with expert range maps (Figure 2), which means that maximizing scores, while near instantaneous time-wise, may not be the best option and of course are not an option for species without range maps. Commission error is a greater problem than omission error for range maps, and while greater penalties can be applied to the parameter values of the Tversky score (i.e.,  $\alpha = 0.96$ ,  $\beta = 0.04$ ) to reduce potential omission error from aggregation distances of 5,000 km, this results in almost all other aggregation distances being 200 km. Because supervised classification is time-intensive, perhaps a score could be used to assign initial aggregation distance and then evaluated. As evidenced by issues when directly maximizing overlap of data ranges with expert ranges, I was not able to develop a ruleset, after trying combinations of scores and thresholds for three spatial layers of kernels, aggregated polygons, and expert range maps and

additional information such as the number of georeferenced records. The need for supervised classification indicated a first step for a modeled solution. Images are complex, range maps are imperfect (a quarter of range maps may contain less than 80% of georeferenced records; Hanberry, 2025b), and most species do not have range maps, a situation that ultimately requires reliance on georeferenced records and derived products alone for data range generation.

Range maps reasonably outline most but not 100% of recorded individuals. The georeferenced records for some species may contain a few errors or outliers that are not consistent with the bulk of georeferenced records (apparent in Figure 8; Hanberry, 2025a). As an example of an error, Dall sheep (*Ovis dalli*) are not known in the continental United States, but one georeferenced record occurred in Utah and would become part of the kernels with maximum inclusion. This Dall sheep record perhaps is due to similarity resulting from hybridization with bighorn sheep (*Ovis canadensis*), or a “Texas Dall sheep”, a hybrid bred for hunting purposes (Santos et al., 2021). Outlying records may correctly identify species that have not been transported by human vectors, such as wandering juveniles or individuals accidentally outside of normal ranges due to wind events. Regardless of reasonable

representation, if all records are included in kernel densities, some species will have range areas that are too spacious even for well-fitted kernels. Inclusion of 99% of records is a compromise to remove any inconsistent outliers, such as vagrant individuals, and not lose critical information about sampled areas.

Future research can refine aspects of data-driven development of ranges beyond the next phase to determine an aggregation distance through predictive modeling of classified samples. Overall, I only generated ranges for species with large samples and ranges (i.e., 400 georeferenced records and range areas  $\geq 10,000$  km<sup>2</sup>), to avoid risk of errors and limited range coverage for species with small samples and unnecessary mapping of range maps completely contained within kernels for species with small ranges. Data-driven ranges from species with small samples and small ranges can be examined as thoroughly as possible. Small samples indicate that sampling is incomplete, but sample size effects depend on both range size and sample coverage, that is, the distribution of samples. Small samples for kernels result in larger extents (due to increased bandwidth to compensate for variance and avoid small sample errors; [Fieberg, 2007](#)). For the application of coarse range maps, small samples for kernels mean more generalized representations of sampled area, albeit with some slight risk of introducing error. For the aggregated polygons, which extend the kernels into undersampled areas based on clusters of three points (necessary to create a shape), the potential for complete range mapping decreases in probability with small samples and thinned samples that reduce clusters (as in species distribution models; [Gábor et al., 2020](#)). While not presented, few mammal species with small ranges met the minimum threshold of robust sample sizes, of which only two mammal species had range maps that contained  $\geq 80\%$  of georeferenced records (median = 51% of georeferenced records). Therefore, both sampling and range maps may need improvement before small ranges can be thoroughly assessed. In any event, if georeferenced records do not occur beyond the kernel, then aggregated polygons also do not extend beyond the kernel (i.e., the same aggregation of points occurs for each aggregation distance). Most species, excepting migratory birds, have one expert-driven range map, but some species have variations based on seasonality or other factors, which could be developed in data-driven range maps based on information contained in the georeferenced records. Additional future development could examine kernel options and optimize range generation within one platform.

## 5 Conclusions

Georeferenced records offer locational information as a foundation to expedite range map creation. Data range maps, based on kernel densities and aggregated polygons, contained 99% of georeferenced records and covered most of expert range maps, with intersection increasing as expert range maps intersected with

georeferenced records. Data range maps at minimum could accelerate the development of range maps by teams of experts in a time-intensive process. In the absence of any accessible range maps, data range maps may be beneficial, at least as an initial starting product with progression over time. Data range maps can complement existing expert-based approaches and contribute to improved range mapping frameworks.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

## Author contributions

BH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by the U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, and the U.S. Department of Agriculture, Forest Service, with funding from the Infrastructure Investment and Jobs Act of 2021.

## Acknowledgments

I thank the Global Biodiversity Information Facility, International Union for Conservation of Nature, BirdLife International, and contributors to the range maps and georeferenced observations.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author BH declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or

claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Author disclaimer

The findings and conclusions in this publication are those of the author and should not be construed to represent any official USDA or U.S. Government determination or policy.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcsc.2026.1839361/full#supplementary-material>.

## References

- Aronsson, H., Zizka, A., Antonelli, A., and Faurby, S. (2024). Expert-based range maps cannot be replicated using data-driven methods but macroecological conclusions arising from them can. *J. Biogeogr.* 51, 1549–1559. doi: 10.1111/jbi.14847
- BirdLife International. (2024). *Bird Species Distribution Maps of the World. Version 2023.1*. Available online at: <http://datazone.birdlife.org/species/requestdis> (Accessed March 10, 2025).
- Blake, J. (2006). Coyotes in insular Newfoundland. Available online at: <https://www.gov.nl.ca/ffa/files/publications-wildlife-51f40aed01.pdf> (Accessed March 1, 2025).
- Canadian Wildlife Service (2015). *Management Plan for the Woodland Vole (Microtus pinetorum) in Canada [Proposed] Species at Risk Act Management Plan Series* (Ottawa: Canadian Wildlife Service), iv–18.
- Cantú-Salazar, L., and Gaston, K. J. (2013). Species richness and representation in protected areas of the Western hemisphere: discrepancies between checklists and range maps. *Divers. Distrib.* 19, 782–793. doi: 10.1111/ddi.12034
- Chacón, J. E., and Duong, T. (2018). *Multivariate Kernel Smoothing and its Applications* (New York, USA: Chapman and Hall/CRC).
- Cheng, G., Xie, X., Han, J., Guo, L., and Xia, G. S. (2020). Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13, 3735–3756. doi: 10.1109/jstars.2020.3005403
- Duong, T. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *J. Stat. Software* 21, 1–16. doi: 10.18637/jss.v021.i07
- ESRI (2025). *Aggregate Points (Standard Feature Analysis)*. Available online at: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/cartography/aggregate-points.htm> (Accessed February 21, 2025).
- Faurby, S., Davis, M., Pedersen, R. Ø., Schowaneck, S. D., Antonelli, A., and Svenning, J. C. (2018). PHYLACINE 1.2: the phylogenetic atlas of mammal macroecology. *Ecology* 99, 2626. doi: 10.1002/ecy.2443
- Faurby, S., Davis, M., Pedersen, R. Ø., Schowaneck, S. D., Jarvie, S., Antonelli, A., et al. (2020). PHYLACINE 1.2: an update to the phylogenetic atlas of mammal macroecology. doi: 10.5281/zenodo.3690867
- Fieberg, J. (2007). Kernel density estimators of home range: smoothing and the autocorrelation red herring. *Ecology* 88, 1059–1066. doi: 10.1890/06-0930
- Fourcade, Y. (2016). Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. *Ecol. Inf.* 36, 8–14. doi: 10.1016/j.ecoinf.2016.09.002
- Franz, N. M., and Sterner, B. W. (2018). To increase trust, change the social design behind aggregated biodiversity data. *Database* 2018, bax100. doi: 10.1093/database/bax100
- Gábor, L., Moudrý, V., Barták, V., and Lecours, V. (2020). How do species and data characteristics affect species distribution models and when to use environmental filtering? *Int. J. Geogr. Inf. Sci.* 34, 1567–1584.
- Gaston, K. J., and Fuller, R. A. (2009). The sizes of species' geographic ranges. *J. Appl. Ecol.* 46, 1–9. doi: 10.1111/j.1365-2664.2008.01596.x
- Gilbert, G. K. (1884). Finley's tornado predictions. *Am. Meteorological J.* 1, 166–172.
- Global Biodiversity Information Facility [GBIF] (2024). *Free and Open Access to Biodiversity Data*. Available online at: [www.gbif.org](http://www.gbif.org) (Accessed January 2, 2025).
- Hanberry, B. B. (2021). Forest disturbance types and current analogs for historical disturbance-independent forests. *Land* 10, 136. doi: 10.3390/land10020136
- Hanberry, B. B. (2022). Confronting the issue of invasive native tree species due to land use change in the eastern United States. *Land* 11, 161. doi: 10.3390/land11020161
- Hanberry, B. B. (2025a). Clarifying influences of sampling bias (concentration) and locational errors (uncertainties) on precision or generality of species distribution models. *Land* 14, 1620. doi: 10.3390/land14081620
- Hanberry, B. B. (2025b). Validating ranges, georeferenced records, and species distribution models, for vertebrate and plant species globally. *Acta Oecologica* 129, 104137. doi: 10.1016/j.actao.2025.104137
- Hanberry, B. B. (2026a). Beyond thinning: Equalized sampling based on global sampling intensity gradients for species distribution models. *Front. Ecol. Evol.* 14, 1831501. doi: 10.3389/fevo.2026.1831501
- Hanberry, B. B. (2026b). Model bigger to avoid unrepresentative species distribution models from truncated spatial extents, with the modulating influence of prevalence. *Environ. Challenges* 24, 101551. doi: 10.1016/j.envc.2026.101551
- Hurlbert, A. H., and Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proc. Natl. Acad. Sci.* 104, 13384–13389. doi: 10.1073/pnas.0704469104
- iNaturalist (2024). *A Community for Naturalists Ć Inaturalist*. Available online at: <https://www.inaturalist.org/> (Accessed September 30, 2024).
- IUCN (2024). *The IUCN Red List of Threatened Species. Spatial Data Download*. Available online at: <https://www.iucnredlist.org/resources/spatial-data-download> (Accessed March 25, 2025).
- IUCN Standards and Petitions Committee (2024). *Guidelines for using the IUCN red list categories and criteria*. Available online at: <https://www.iucnredlist.org/documents/RedListGuidelines.pdf> (Accessed March 27, 2025).
- Jetz, W., McPherson, J. M., and Guralnick, R. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends Ecol. Evol.* 27, 151–159. doi: 10.1016/j.tree.2011.09.007
- Jetz, W., Sekercioglu, C. H., and Watson, J. E. (2008). Ecological correlates and conservation implications of overestimating species geographic ranges. *Conserv. Biol.* 22, 110–119. doi: 10.1111/j.1523-1739.2007.00847.x
- Marsh, C. J., Sica, Y. V., Burgin, C. J., Dorman, W. A., Anderson, R. C., del Toro Mijares, I., et al. (2022). Expert range maps of global mammal distributions harmonised to three taxonomic authorities. *J. Biogeogr.* 49, 979–992. doi: 10.1111/jbi.14330
- Pagad, S., Bisset, S., Genovesi, P., Groom, Q., Hirsch, T., Jetz, W., et al. (2022). Country compendium of the global register of introduced and invasive species. *Sci. Data* 9, 391. doi: 10.1038/s41597-022-01514-z
- Palacio, R. D., Negret, J., Velásquez-Tibatá, J., and Jacobson, A. (2021). A data-driven geospatial workflow to map species distributions for conservation assessments. *Divers. Distrib.* 27, 2559–2570. doi: 10.1111/ddi.13424

- R Core Team (2025). *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing).
- Sanderson, E. W., Redford, K. H., Chetkiewicz, C. L. B., Medellin, R. A., Rabinowitz, A. R., Robinson, J. G., et al. (2002). Planning to save a species: the jaguar as a model. *Conserv. Biol.* 16, 58–72. doi: 10.1046/j.1523-1739.2002.00352.x
- Santos, S. H., Peery, R. M., Miller, J. M., Dao, A., Lyu, F. H., Li, X., et al. (2021). Ancient hybridization patterns between bighorn and thinhorn sheep. *Mol. Ecol.* 30, 6273–6288. doi: 10.1111/mec.16136
- Snyder, J. P. (1987). *Map Projections – a Working Manual* (Denver, Colorado, USA: USGS).
- Taulman, J. F., and Robbins, L. W. (1996). Recent range expansion and distributional limits of the nine-banded armadillo (*Dasypus novemcinctus*) in the United States. *J. Biogeogr.* 23, 635–648. doi: 10.1111/j.1365-2699.1996.tb00024.x
- Tversky, A. (1977). Features of similarity. *psychol. Rev.* 84, 327–352. doi: 10.1016/b978-1-4832-1446-7.50025-x
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics With s, Fourth Edition* (New York: Springer). Available online at: <https://www.stats.ox.ac.uk/pub/MASS4/> (Accessed July 10, 2026).
- Walsh, L. L., Seabloom, R., and Thompson, C. W. (2017). Range extension of the Virginia opossum (*Didelphis virginiana*) in North Dakota. *Prairie Naturalist.* 49 (2).
- Worton, B. J. (1987). A review of models of home range for animal movement. *Ecol. Modell.* 38, 277–298. doi: 10.1016/0304-3800(87)90101-3
- Yeung, M., Sala, E., Schönlieb, C. B., and Rundo, L. (2022). Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graphics* 95, 102026. doi: 10.1016/j.compmedimag.2021.102026
- Zizka, A., Carvalho, F. A., Calvente, A., Baez-Lizarazo, M. R., Cabral, A., Coelho, J. F. R., et al. (2020). No one-size-fits-all solution to clean GBIF. *PeerJ* 8, e9916. doi: 10.7717/peerj.9916