

BIODYN: A TOOL FOR THE SPATIOTEMPORAL ANALYSIS OF INTRA-SPECIFIC GEOGRAPHIC DATA

TIAGO A. HERRADOR^{1*†}, CARLA A. REATI^{1*†}, PABLO YAIR HUAIS¹, LUIS OSORIO-OLVERA^{2,3},
RUSBY G. CONTRERAS-DÍAZ^{3,4}, SUSANA I. PELUC¹, JAVIER NORI^{1*}

¹ Instituto de Diversidad y Ecología Animal (IDEA-CONICET), Facultad de Ciencias Exactas, Físicas y Naturales (FCEFN), Universidad Nacional de Córdoba, Córdoba, Argentina.

² Instituto de Ecología, Universidad Nacional Autónoma de México (UNAM), Ciudad de México, México.

³ Laboratorio Nacional de Biología del Cambio Climático. Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Ciudad de México, México.

⁴ Escuela Nacional de Estudios Superiores Unidad Mérida, Universidad Nacional Autónoma de México (UNAM), Mérida, México.

Abstract. The rapid growth of large biodiversity databases has profoundly expanded the availability of species occurrence data across space and time. However, this increase in data volume has not been accompanied by equivalent gains in knowledge, largely due to persistent biases, gaps, and heterogeneity in sampling effort, temporal resolution, and environmental representation. These limitations contribute to shortfalls in biogeographical knowledge, particularly in understanding spatiotemporal patterns at the intra-specific level. Here, we present BioDyn, an R package designed to provide a reproducible workflow for spatiotemporal exploration of species occurrence data, with a focus on interannual variation and intra-specific patterns. The package integrates key pre-modeling steps related to organization, inspection, and segmentation of occurrence data, construction of temporal presence-absence datasets, environmental variable enrichment, and exploratory visualization across space and time. Through a transparent, modular and flexible framework, users are enabled to assess data coverage, evaluate sampling biases, and explore spatial and temporal patterns before formal modeling. In doing so, it provides a practical tool for studying spatial dynamics, seasonal variation, and migratory systems across large and heterogeneous biodiversity datasets.

Key words: species occurrence data; presence-absence data; environmental data enrichment; spatial segmentation; sampling bias; intraspecific dynamics; sampling effort; pre-modeling framework

INTRODUCTION

Although scientific progress is intrinsically linked to identifying knowledge gaps and uncertainties, naturalists and biogeographers have often been more concerned with organizing, classifying, and delimiting existing biological knowledge than with explicitly identifying and quantifying what remains unknown (e.g., Wallace, 1876; Whittaker et al., 2005). This contrast has become particularly evident in the era of Big Data, where large digital infrastructures have profoundly transformed scientific disciplines by enabling the aggregation, standardization, and global dissemination of information (Kitchin, 2014; García-Roselló et al., 2023). In biogeography and ecology, global biodiversity repositories such as GBIF, eBird, and VertNet now provide access to billions of georeferenced

species records, representing an unprecedented advance in data availability. These resources have revolutionized our capacity to describe species distributions (Peterson et al., 2011; Zurell et al., 2020), generated large-scale ecological and biogeographical hypotheses, and supported applied research in fields ranging from biodiversity conservation to epidemiology and public health, including the study of vector-borne diseases and climate-driven health risks (Peterson, 2006; Messina et al., 2019; Nori et al., 2014).

However, as the use of large-scale biodiversity repositories has expanded, there has been growing recognition that the availability of vast amounts of data does not necessarily translate into comprehensive, representative, or unbiased knowledge (Nori et al., 2022; Cordier et al., 2021). This recognition has increased attention on identi-

[†] These authors contributed equally to this work.

* Corresponding author: javiernori@gmail.com

fyng and quantifying what we do not know, as a necessary strategy to guide research efforts and improve ecological and biogeographical inference (Hortal et al., 2015; Nori et al., 2023). In biogeography, these limitations were formally conceptualized as major knowledge shortfalls that constrain our understanding of biodiversity patterns and processes (Hortal et al., 2015). Among them, the Linnaean shortfall, for instance, reflects the discrepancy between the number of species that actually exist and those that have been formally described, highlighting persistent gaps in taxonomic knowledge (Brown & Lomolino, 1998; Cardoso et al., 2011). In contrast, the Wallacean shortfall describes the biases and uncertainties in our understanding of species' geographic distributions, even for well-known taxa, thereby limiting our ability to accurately characterize spatial patterns (Hortal et al., 2015; Lomolino et al., 2004).

At the core of biogeography, species distributions emerge from the interaction of biotic and abiotic factors operating across space and time, together with the evolutionary history of each taxon (Soberón and Peterson, 2005; Girón-Juárez et al., 2025). Our understanding of these distributional patterns is often constrained by incomplete knowledge of where and when species occur, a central consequence of major biodiversity knowledge shortfalls. This problem becomes particularly relevant when species distributions are analyzed as dynamic spatiotemporal phenomena rather than as static geographic patterns, potentially leading to incomplete or biased representations of seasonal dynamics, environmental associations, and population-level variation. Addressing these biodiversity knowledge shortfalls are essential for accurately characterizing distributional patterns and for disentangling the processes that shape them. Understanding the distribution of a species requires not only knowledge of where it occurs in geographic space but also how its distribution varies over time and how this spatiotemporal variation relates to environmental conditions, life-history requirements, and population structure. While spatial patterns of species distributions have been extensively studied, the temporal dimension has received comparatively less attention, particularly at the intraspecific level, where seasonal dynamics, population-level variation, and migratory behaviors play a central role.

In this context, migration represents a core ecological process, as it explicitly links temporal variability with spatial patterns of occurrence. Rather than conforming to a strict dichotomy between residency and seasonal movement, it encompasses flexible strategies shaped by environmental heterogeneity and behavioral plasticity (Lovette and Fitzpatrick, 2016; Newton, 2023). In fact, a single species may exhibit contrasting migratory strategies

among populations, even within geographically proximate areas (Vergara-Tabares & Rojas, 2016; Jahn et al., 2020; Hernández-Hernández et al., 2022). Despite their importance, detailed characterizations of these space-time dynamics remain limited for many taxa. This results in substantial information deficits regarding the environments effectively used by species throughout their annual cycles, as well as those absent or poorly represented in available data (Jahn et al., 2020).

Precisely because migration integrates spatial, temporal, and environmental dimensions, its analysis requires the treatment of multiple data types. However, although the availability of species occurrence data has increased through citizen science platforms, monitoring programs, and repositories (Contreras-Díaz et al., 2023), their integrated analysis remains methodologically challenging. Identifying differentiated migratory strategies involves multiple stages such as data cleaning, spatial segmentation, or visualization, often distributed across several tools and software platforms. Additionally, linking occurrences with environmental variables requires further data extraction and processing steps. This fragmentation hampers the systematic assessment of how environmental requirements vary across populations and seasons, and constrains the explicit detection of environmental information gaps (Nakazawa et al., 2004; Berthold et al., 2013; Jahn et al., 2020). To fill this gap, we introduce BioDyn, an R package that integrates these analytical stages into a reproducible and flexible workflow for the spatiotemporal exploration of species occurrence data. BioDyn supports data import and standardization, spatial and seasonal structuring of observations, user-defined spatial segmentation, extraction of environmental covariates, and exploratory visualization to assess spatiotemporal patterns and environmental drivers before formal modeling.

BioDyn

BioDyn is an R package designed to support reproducible workflows for the exploration of temporally structured species occurrence data in ecology and biogeography, with a particular focus on seasonal and interannual variation at the intraspecific level. The package was conceived as a diagnostic tool intended to support stages prior to formal modeling and statistical analyses. It integrates key pre-modeling steps such as data filtering and standardization, spatial structuring, environmental variable enrichment, and visualization.

Although the workflow is presented as a sequence of analytical stages, the architecture is flexible and supports non-sequential execution. Each function operates on clearly defined inputs and returns explicit and exportable

outputs, allowing users to pause, branch, or repeat steps as needed. In addition, its transparent framework enables users to customize the workflow according to their needs, facilitating the inspection of data structure and the exploration of spatiotemporal hypotheses.

The growing availability of data derived from online databases, citizen science platforms, and monitoring programs (e.g., GBIF, BirdLife, iNaturalist) has greatly expanded the spatial and temporal coverage of biodiversity information. However, these data are often affected by strong heterogeneity in sampling effort, temporal resolution, and environmental representation. In this context, BioDyn facilitates the organization of occurrence data across spatial, seasonal, and environmental dimensions, while supporting the identification of biases in environmental and distributional information. The package also supports the identification of distributional information deficits associated with major biogeographical knowledge shortfalls, particularly those related to species distributions across seasons and geographic space.

In terms of scope, BioDyn operates on occurrence-based observations, including geographic coordinates (latitude and longitude) and monthly temporal information. It supports both presence-only and presence-absence datasets derived from structured sampling protocols, such as eBird checklists, if records provided are georeferenced and temporally indexed. When required, BioDyn also supports the integration of user-defined spatial units, polygon-based regionalizations, and raster-based environmental data. In addition, BioDyn includes extensive input validation to ensure robust handling of heterogeneous biodiversity datasets.

Architecture and workflow

BioDyn offers a modular workflow that transforms raw occurrence records into spatially and environmentally enriched datasets for spatiotemporal analyses, with components that can be applied either sequentially or independently. The workflow comprises six main stages (Fig. 1): (i) compilation of user inputs; (ii) data ingestion and eBird-specific filtering; (iii) spatial filtering and segmentation; (iv) seasonal enrichment; (v) environmental enrichment; and (vi-vii) diagnostic visualization.

Compilation of user inputs.—Before running the workflow, users must prepare the required input files, including occurrence records (sightings and checklists), ecoregion layers, and environmental raster variables (e.g., CHELSA climate data or digital elevation models; Farr et al., 2007; Jarvis et al., 2008; Karger et al., 2017). Additionally, users may include a polygon-based layer representing geographic units or segments, useful for grouping occur-

rence records according to ecological, spatial, or biogeographical user-defined criteria.

Data ingestion and eBird-specific filtering.—The workflow in R begins with importing all data files. The user must provide the file paths to the compiled records of presence data, checklists, ecoregions, or other files of interest. Presence-absence datasets can be constructed using the eBird checklist file via the `build_data()` function, a fundamental requirement for a wide range of ecological and biogeographic analyses. Constructing this table allows an explicit distinction between true absences and unsampled locations, which is important to avoid confounding biological patterns with gaps in sampling coverage.

Optional filtering operations can subsequently be applied to eBird-specific sampling and effort variables using the function `filter_ebird()`. By implementing these thresholds as explicit and configurable steps within the workflow, BioDyn enables users to align with established best practices for processing citizen science data (Johnston et al., 2021) while maintaining reproducibility. The resulting presence-absence matrices provide a basis for subsequent environmental enrichment, spatial aggregation, and exploratory analyses.

Spatial filtering and segmentation.—This step incorporates interactive, user-defined spatial filtering and segmentation. Rather than relying exclusively on predefined spatial units or fully automated clustering approaches, BioDyn allows users to define spatial segments based on visual inspection and ecological criteria to explore potentially distinct seasonal dynamics, environmental associations, and migration strategies, enabling the explicit testing of hypotheses about intraspecific spatial structure.

The function `filter_ecoregion()` enables filtering based on externally defined polygon layers. At the same time, `filter_radius()` applies a distance-based spatial filter by retaining data points within a user-defined spatial radius around presence records to construct proximity-based spatial subsets.

Using the `create_segments()` function, users can create spatial segments of interest by drawing polygons or rectangular extents directly on an interactive map. This allows users to explore hypotheses about population structure, or exploratory partitions informed by domain expertise. Once the desired polygons have been drawn, users must explicitly press the *save* button within the interactive map interface to export the geometries as a GeoJSON file. The `filter_pol()` function then retains observations falling within the defined polygons and excludes all others, returning a statistical summary and several diagnostic visual elements.

Seasonal enrichment.—Migratory and partially migratory species may exhibit substantial variation in their



BioDyn workflow

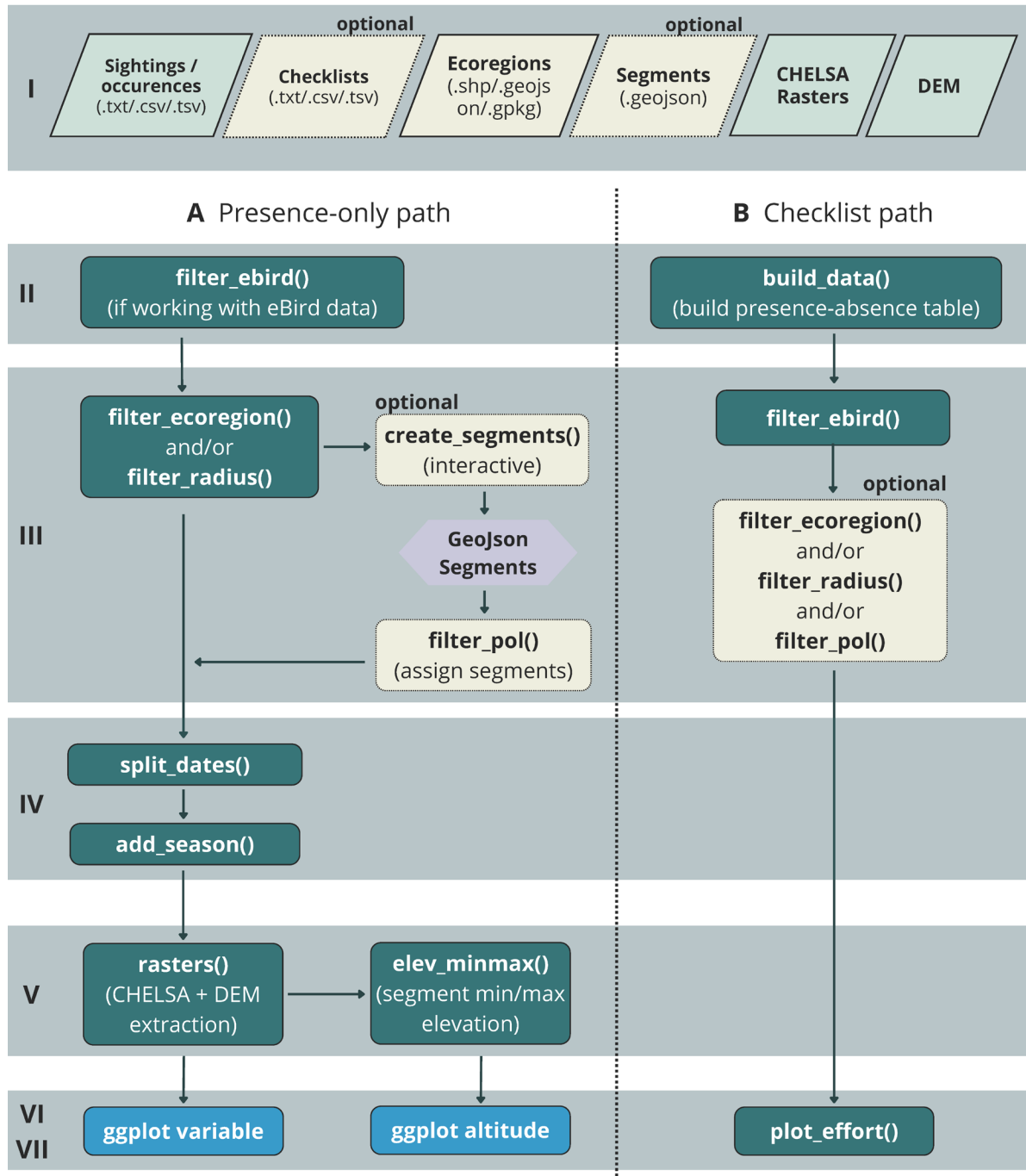


Figure 1. Schematic representation of the BioDyn workflow. The diagram illustrates the sequential data flow from user inputs (top) to final plotting and visual analysis (dark teal boxes). The workflow is structured into six primary phases: (I) compilation of user inputs; (II) data ingestion and eBird-specific filtering via `filter_ebird()` and `build_data()`; (III) spatial filtering and segmentation, offering alternative pathways through `filter_ecoregion()` or `filter_radius()`, with optional interactive refinement using `create_segments()` and `filter_pol()`; (IV) seasonal enrichment, where observations are annotated with seasonal data (`split_dates()` and `add_season()`); (V) environmental enrichment, where observations are annotated with raster values (`rasters()`) and computation of topographic constraints via `elev_minmax()`; and (VI-VII) diagnostic visualization using plots from `ggplot2` package.

spatial distribution and environmental associations across seasons. Decomposing raw date records into discrete temporal components and assigning categorical seasons allows the detection of intraspecific seasonal dynamics and phenological shifts.

Once observations have been standardized and, eventually, segmented, the workflow proceeds with record enrichment by restructuring temporal data into seasonal cohorts using the `split_dates()` and `add_season()` functions, enabling the assessment of seasonal and phenological shifts.

Environmental enrichment.—Enriching occurrence records with these variables allows the understanding of under what conditions a species occurs. The `rasters()` function integrates environmental variables (e.g. Worldclim layers; Fick & Hijmans, 2017) into the data frame. BioDyn supports the extraction of climatic variables and topographic information from gridded datasets, enabling the integration of environmental context directly at the observation level. For topographic characterization, point-based extraction is complemented by computing segment-level summaries using the `elev_minmax()` function.

Exploratory data analysis and visualization.—To support the inspection of data structure, seasonal spatial patterns, and environmental distributions across multiple stages, the workflow provides a set of functions and pre-defined ggplot2 graphics (Wickham, 2016).

Visual outputs include sampling effort (`plot_effort()` function); density curves of the latitudinal distribution of records; bar plot of monthly proportion of records; boxplots of latitudinal distribution by season; density curves of environmental gradient distributions of records; boxplots of altitudinal distribution by season; and boxplots of the altitudinal distribution by season, all segments displayed in a single plot. These visualizations are automatically generated as integrated outputs within the workflow. Besides, interactive maps complement static plots by linking spatial segments, seasons, observation points, and metadata within an exploratory environment. These maps serve as a rapid diagnostic tool, enabling a comparison of environmental conditions across seasons and spatial segments.

Software availability

BioDyn is available for installation in the R environment from the Zenodo repository at <https://doi.org/10.5281/zenodo.19055001>. The repository provides the source package (*BioDyn_0.1.0.tar.gz*), which can be installed in R using `install.packages("path/to/BioDyn_0.1.0.tar.gz", repos = NULL, type = "source")`, and an example dataset (*Example_nrufiventris.rar*). This .rar file

must be extracted into the working directory, as it contains the datasets used in the tutorial examples, including sample occurrence records (*Occurrences.txt*), sampling event data (*Checklists.txt*), environmental raster layers (CHELSA climatic variables and elevation), and ecoregion shapefiles (*Ecoregion.shp* and associated files) required to reproduce the analytical workflow implemented in the package.

EXAMPLE APPLICATION

To illustrate the workflow and utility of BioDyn, we use the chocolate-vented tyrant (*Neoxolmis rufiventris*) as a case study species. This bird is a large, ground-dwelling tyrannid associated with open and semi-open steppe habitats, mainly below 500 m elevation, characterized by grasslands with scattered shrubs or low bushes. Its distribution ranges from southeastern Brazil to Tierra del Fuego, Argentina. The species is an austral migrant: during the austral winter, it concentrates in the Pampean region in the northern part of its range, whereas in summer it moves south to the Patagonian steppe, where it breeds (Farnsworth and Langham, 2025).

A comprehensive workflow example, featuring step-by-step code and explanatory comments, is available within the documentation section of the package repository. These resources, together with the necessary supplementary data, can be accessed after downloading and installing the package.

The workflow of BioDyn begins by assigning the routes to the files: raw eBird occurrences (`sightings_file`), sampling event data (`checklist_file`) covering the Argentine extent, polygon layers (`eco_path`), a digital elevation model (`dem`), and environmental variables (`raster_list`). BioDyn reads heterogeneous occurrence data sources (e.g., GBIF, BirdLife, or personal datasets) and supports multiple file formats and text encodings. Spatial boundaries can be provided as polygon vector files (.shp, .geojson, .json, .gpkg) or as sf objects, while environmental predictors must be raster layers compatible with the terra package (Hijmans et al., 2026). This standardized structure ensures consistency for spatial segmentation, environmental enrichment, and exploratory analyses.

At this step, occurrence records are automatically standardized, and geographic coordinates are validated and coerced to numeric values, allowing the removal of invalid locations.

1. LOAD DATA AND/OR BUILD A PRESENCE-ABSENCE TABLE

The first step uses the `fread()` function to load sightings data and, when sampling-event metadata is available, the `build_data()` function to load and integrate both sightings and absence data. This function constructs a unified

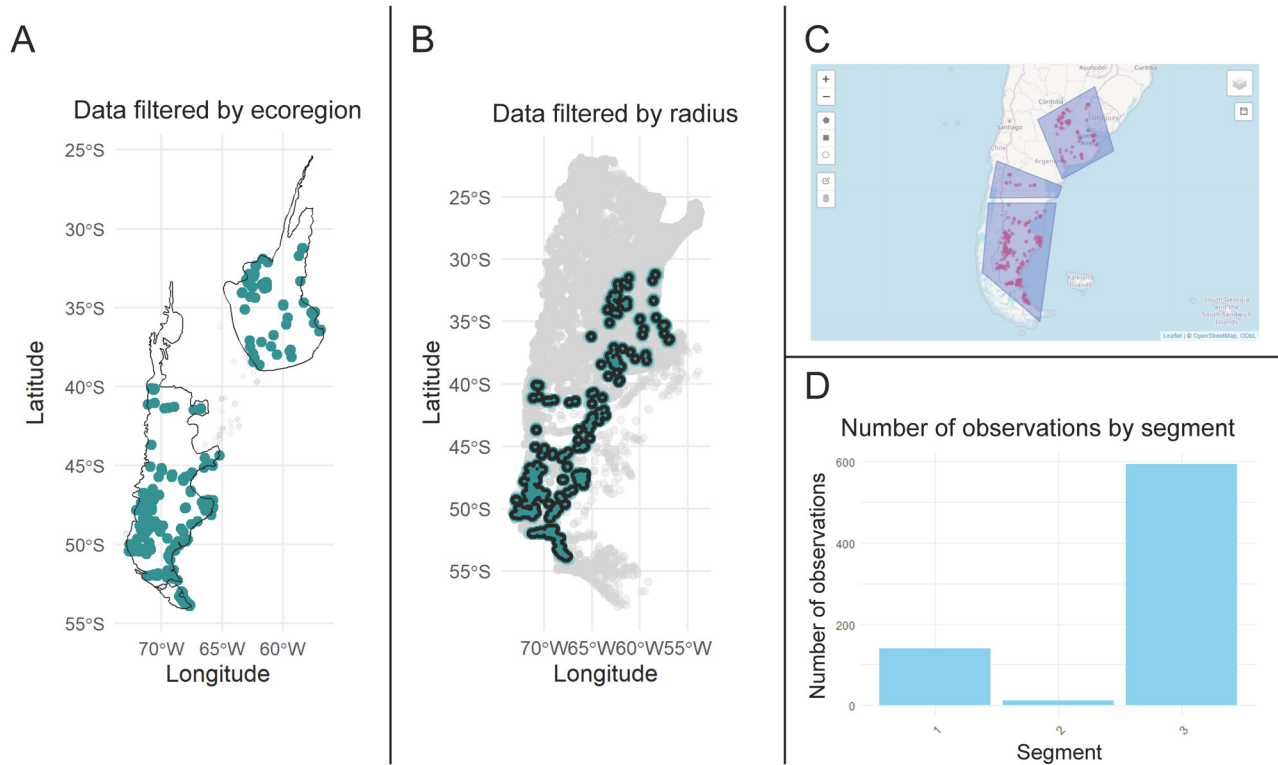


Figure 2. Spatial filtering and segmentation procedures. (A) Static map produced by `filter_ecoregion()`, showing records retained after restricting the dataset to the selected ecoregional boundaries. (B) Static map generated by `filter_radius()`, illustrating proximity-based filtering of checklists within a 20 km radius around presence centroids. In (A) and (B), gray dots represent absence records excluded by the filtering criteria, light teal dots indicate retained absence records within the defined spatial extent, and dark teal dots correspond to presence records. (C) Three spatial segments interactively delineated using `create_segments()` (blue polygons) based on literature and exploratory inspection. The resulting GeoJSON file enables subsequent region-specific analyses within a reproducible framework. (D) Output generated by `filter_pol()` after applying the user-defined spatial segments, showing the number of presence records retained within each segment.

presence-absence table by standardizing the dataset, linking valid sightings with sampling effort information, and enabling the explicit distinction between inferred absences and records missing due to incomplete or unavailable sampling information. At this stage, column separators are adapted to match the format of the input.txt files.

Because absence data is often missing, BioDyn can also be run using presence-only data. In such cases, only the `build_data()` and `plot_effort()` functions (see Section 3.6) are affected, as they specifically rely on inferred absences and sampling-effort information, while the remaining package functionalities can operate using presence records alone.

2. FILTER BY EBIRD EFFORT

To mitigate sampling bias, we applied the `filter_ebird()` function to both presence and checklist data. In this second step, we removed duplicate records and limited sampling effort by retaining only presences with a maximum travel distance of 10 km, no more than ten observers, and a duration of up to 300 minutes. We also restricted

the protocols to “Stationary” and “Traveling.” Following this filtering step with *N. rufiventris* data, 814 records were retained from the original data table of 3,329 records, and 280,958 records were retained from the original checklists table of 1,010,756 records of Argentina.

These thresholds follow the guidelines recommended by Johnston et al. (2021) for processing eBird data, which suggest limiting travel distance, number of observers, and duration.

3. FILTER BY ECOREGION, RADIUS, AND CUSTOM SEGMENTS

To illustrate alternative filtering strategies, we applied three complementary approaches. First, we used `filter_ecoregion()` to restrict the dataset to the selected Argentine ecoregions, with a 0 km buffer zone. This excluded 69 presence records and 180,528 checklists from the Andean, from outside the focal area, and from the middle of the distribution (Fig. 2a). Second, we generated a proximity-based subset using `filter_radius()`, retaining checklists located within a 20 km radius of presence records (Fig. 2b). Both functions were customized by editing the input

file, the buffer distance around the polygon, and the radius distance. They return a data table that can be directly used in subsequent steps, and generate a map-based visualization of the filtered dataset, which can be rendered either as a static map for documentation and reporting purposes or as an interactive map to support exploratory spatial analysis and user-guided inspection of the filtering results.

Third, using `create_segments()`, observation points are displayed within an interactive mapping interface at global or regional scales. To facilitate the workflow demonstration, we delineated three spatially distinct segments (northern, central, and southern; Fig. 2c). The resulting geometries are exported as GeoJSON objects and re-integrated into the workflow with `filter_pol()`, a function that assigns observations to the user-defined segments, excluding those falling outside, and returns a statistical summary along with five visual elements: an interactive map with all observation points, an interactive map with all observation points differentiated by colored polygons, a bar plot representing the number of observations in each segment (Fig. 2d), a dot plot illustrating the spatial distribution of observations by segment, and an observation-level data frame (.csv file) augmented with an additional categorical column indicating the spatial segment to which each point belongs.

4. EXTRACT DATE AND SEASON

Once spatially structured, the dataset underwent temporal and environmental enrichment. The function `split_dates()` decomposes a single date column into separate DAY, MONTH, and YEAR fields, facilitating temporal filtering, aggregation, and subsequent seasonal analyses. Building on this structure, `add_season()` assigns a categorical season variable based on the temporal information and appends a new column indicating the calendar-based season in which each observation was made, according to the hemisphere assigned through a function argument.

5. EXTRACT RASTER DATA

Concurrently, high-resolution environmental variables were extracted at point locations using `rasters()` and `elev_minmax()`. The `rasters()` function involves spatial operations that explicitly account for coordinate reference system mismatches and incomplete raster metadata. Observation coordinates are transformed as needed to match raster projections, ensuring spatial consistency across data sources. Missing raster values and extraction failures are handled explicitly, preventing the silent propagation of errors into downstream analyses. On the other hand, the `elev_minmax()` function computes the topographic con-

straints for each spatial segment by extracting the minimum and maximum elevations from a digital elevation model (DEM) file. These summaries enable direct comparison between the elevations available within each spatial segment and the observed elevational range of species presence.

This process annotated observations with CHELSA V.2.1 bioclimatic variables (Karger et al., 2017), specifically Annual Mean Temperature (bio01) and Annual Precipitation (bio12), as well as elevation data derived from a 100 m DEM. The workflow generates two complementary outputs: (i) a summary reporting the main aggregated results, including counts, ranges, or descriptive statistics of the extracted variables; and (ii) an observation-level data frame (.csv file), in which the newly extracted information for each observation is organized into columns.

6. PLOT EFFORT

By applying the `plot_effort()` function, we can distinguish between total observations (complete checklists with and without presences) and presence records alone. This function was specifically designed to evaluate sampling effort along latitudinal gradients, allowing comparisons of the relative seasonal contribution of records at each degree of latitude. Although this visualization focuses on latitudinal variation, the broader BioDyn workflow also integrates environmental characterization, elevational structure, spatial segmentation, and seasonal grouping of occurrence records. By visualizing presences alongside complete checklists, users can distinguish biological patterns from effort-driven artifacts, helping to identify sampling biases and avoid misinterpreting distributional gaps as ecological absences before formal modeling.

In this example, checklist numbers are relatively homogeneous across both latitudes and seasons (Fig. 3a). This pattern suggests relatively consistent sampling effort throughout the year, reducing the likelihood that the absence of presence records observed in the presence observations plot (Fig. 3a, right panel) across seasons is solely attributable to uneven effort. In this context, a sampling artifact would be if, for example, the absence of *N. rufiventris* records at southern latitudes during winter would correlate to a lack of checklists in that region during that season. Since checklist coverage is relatively homogeneous, the observed seasonal distribution shift can be more confidently interpreted as a biological signal reflecting genuine migratory displacement. At present, `plot_effort()` is limited to latitudinal visualization because BioDyn is primarily focused on exploring seasonal and migratory distributional dynamics.

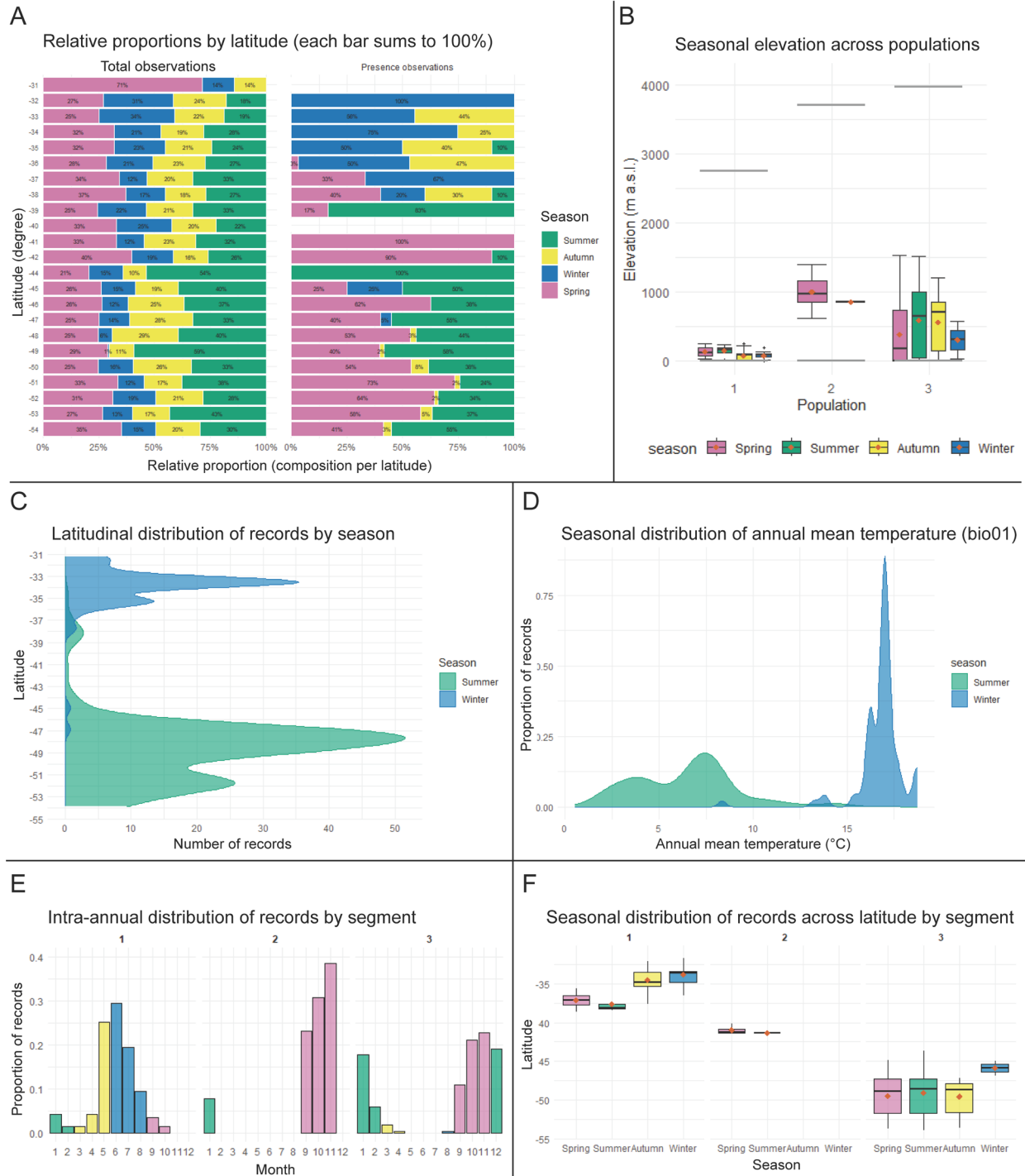


Figure 3. Exploratory visualizations of spatiotemporal and environmental structure generated using `plot_effort()` function and pre-defined `ggplot2` graphics. (A) Latitudinal composition of sampling effort and presence records across seasons generated with `plot_effort()`. Stacked bar plots show the relative seasonal proportions of records, with each bar representing a 1° latitudinal band (e.g., $-32^{\circ} = -32^{\circ}00'$ to $-32^{\circ}59'$; each bar sums to 100%). (Left) Total observations, including complete checklists with and without presences, illustrate the seasonal distribution of sampling effort along the latitudinal gradient. (Right) Presence records only, highlighting the seasonal composition of species occurrences across latitudes. (B) Seasonally stratified boxplots of observed elevation by segment. Horizontal lines indicate the absolute topographic limits (minimum and maximum elevation) of each spatial unit as calculated with `elev_minmax()`. (C) Latitudinal density curves showing the seasonal distribution of presence records, highlighting spatial segregation between summer and winter segments along the north–south gradient. (D) Density plots of annual mean temperature (bio01) associated with observations in each season, revealing seasonal shifts in the climatic conditions occupied by the species. (E) Monthly bar plots of record proportions within each spatial segment (1 = northern, 2 = central, 3 = southern), illustrating seasonal clustering and differences in phenological timing among segments. (F) Seasonal latitudinal distribution of records within each spatial segment. Boxplots summarize the latitude of presence records by season, allowing comparison of seasonal shifts in geographic position among segments.

7. VISUALIZATION OF SPATIOTEMPORAL STRUCTURE

Example outputs include seasonally stratified boxplots of observed altitudes overlaid with topographic constraints for each segment (Fig. 3b); density curves illustrating the degree of distributional overlap and differentiation among seasonal segments across the latitudinal gradient (Fig. 3c); density curves showing the distribution of annual mean temperature values used by the species in different seasons (Fig. 3d); monthly bar plots describing the seasonal distribution of records within each spatial segment (Fig. 3e); and seasonal latitudinal distribution of records within each spatial segment (Fig. 3f).

This visualization revealed that the species exhibits a clear latitudinal shift between seasons, occupying northern latitudes during winter and southern latitudes during summer, consistent with previously documented austral migratory behavior. Although these patterns are broadly consistent with the known migratory behavior of the species, this example illustrates how BioDyn enables the simultaneous exploration of seasonal geographic shifts, environmental associations, elevational structure, and sampling effort within a unified workflow. Annual mean temperature (*bio01*) distributions showed limited seasonal overlap, indicating distinct environmental conditions between breeding and non-breeding periods, but given that the species occupies different latitudes across seasons this pattern should be interpreted with caution. Elevational patterns further highlighted segment-level differences, with southern segments occupying broader and higher elevational ranges during the breeding season.

Together, these outputs illustrate how BioDyn enables the structured characterization of seasonal distributional shifts and associated environmental variation within a reproducible workflow.

DISCUSSION

In the midst of the Big Data era, and with the scientific community increasingly aware of the need to systematically organize biodiversity information to address major knowledge shortfalls (Kitchin, 2014; Zurell et al., 2020; García-Roselló et al., 2023), we introduce BioDyn. This package provides a structured and reproducible framework for organizing species occurrence data across spatial, temporal, and environmental dimensions. By integrating heterogeneous occurrence records with spatial segmentation, seasonal stratification, environmental enrichment, and pre-configured visual outputs within a unified workflow, BioDyn enables explicit assessment of sampling biases, environmental representation, and intraspecific seasonal and spatial variation. In addition, consolidating these pre-modeling and exploratory analytical steps within a

single framework facilitates the detection of environmental information gaps, uneven sampling coverage, and distributional biases that could otherwise remain unnoticed when analyses are fragmented across multiple tools and workflows.

The package is particularly useful for studies focused on intraspecific spatial–temporal variation, such as migratory systems, partial migration, seasonal niche shifts, and population-level environmental differentiation (e.g., Jahn et al., 2020; Newton, 2023). By enabling user-guided spatial segmentation and seasonal stratification, BioDyn incorporates ecological reasoning into exploratory analyses while maintaining analytical flexibility and facilitating the generation of qualitative hypotheses about population structure and seasonal dynamics. In particular, the `build_data()` function allows users to distinguish between true absences and unsampled locations through the integration of occurrence and checklist information, helping to avoid confounding biological patterns with gaps in sampling coverage. In addition, the `plot_effort()` function provides an intuitive graphical representation of sampling effort across latitudinal and seasonal gradients, facilitating the identification of effort-driven artifacts and improving the interpretation of apparent distributional gaps prior to downstream analyses. Together, these functions contribute to a more cautious interpretation of observed distributional patterns by helping distinguish biological signals from artifacts derived from uneven sampling intensity—a critical issue when working with citizen science data (Johnston et al., 2021; Contreras-Díaz et al., 2023).

BioDyn also promotes a shift from occurrence datasets as static point clouds toward recognizing them as structured, multi-dimensional data systems in which space, seasonality, environmental conditions, and sampling effort are integrated. This represents a necessary step toward understanding the complex processes shaping species' distributions, in line with process-based frameworks such as the BAM conceptual model (Soberón & Peterson, 2005). Considering the growing volume of data, BioDyn facilitates the identification and visualization of distributional, temporal, environmental, and sampling-information gaps associated with major biogeographical shortfalls, particularly those related to incomplete and biased distributional knowledge (Lomolino et al., 2004; Hortal et al., 2015). Likewise, the integration of environmental enrichment and seasonal exploratory analyses may help identify gaps related to species' environmental requirements and niche limits, indirectly contributing to the Hutchinsonian shortfall. More indirectly, BioDyn may also contribute to addressing aspects of the Prestonian shortfall, insofar as structured spatiotemporal occurrence and sampling-effort

data can provide useful proxies for population dynamics despite not incorporating direct demographic or abundance information (Hortal et al., 2015).

Despite these strengths, BioDyn presents opportunities for improvement. For instance, the package operates strictly within the domain of occurrence-based data and does not incorporate demographic information, movement trajectories, abundance estimates, or genetic structure. Consequently, inferences regarding population differentiation or migratory connectivity remain indirect and should be interpreted with caution. Likewise, although the interactive segmentation component provides flexibility and ecological interpretability, it also introduces a degree of subjectivity when spatial segments are defined through expert criteria. Future developments could strengthen BioDyn's integration within broader ecological workflows by establishing standardized interfaces with species distribution, occupancy, or niche modeling frameworks, as well as by incorporating quantitative bias metrics to complement the current exploratory visual diagnostics.

Within this context, BioDyn occupies a complementary but distinct analytical space relative to other R packages for biodiversity data processing and spatial exploration, such as *rgbif* (Chamberlain et al., 2026), *auk* (Strimas-Mackey et al., 2026a), and *sf* (Pebesma & Bivand, 2023), which primarily focus on occurrence data acquisition, filtering, and spatial data exploration workflows. In contrast, BioDyn extends these approaches by explicitly structuring occurrence records within a spatiotemporal framework centered on intra-specific seasonality, and population-level segmentation, and exploratory analyses of dynamic distributions. BioDyn also complements modeling-oriented frameworks, such as *ebirdst* (Strimas-Mackey et al., 2026b). While *ebirdst* provides high-resolution estimates of abundance, occurrence, and population trends through large-scale modeling approaches, BioDyn is designed to work directly with raw occurrence data, organizing them to support intraspecific and population-level exploration prior to formal modeling.

With the current fast expansion of biodiversity data availability (García-Roselló et al., 2023; Contreras-Díaz et al., 2023), most macroecological and biogeographical tools continue to operate at the species level (Peterson et al., 2011; Zurell et al., 2020), often overlooking essential processes such as intraspecific movements and seasonal population differentiation (Jahn et al., 2020; Newton, 2023). BioDyn came to occupy a particularly relevant position within this analytical landscape by explicitly structuring occurrence data to capture spatial and seasonal dynamics at the population level. By integrating exploratory analyses of sampling effort, environmental representation, and seasonal variation within a unified workflow, the

package also facilitates the identification of information gaps and potential biases affecting the interpretation of intra-specific distributional patterns.

REFERENCES

- Brown, J. H., and Lomolino, M. V. 1998. *Biogeography*. 2nd Ed. Sinauer Associates, Sunderland, Massachusetts.
- Cardoso, P., Erwin, T. L., Borges, P. A., and New, T. R. 2011. The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, 144, 2647–2655.
- Chamberlain, S., Barve, V., McGlinn, D., Oldoni, D., Desmet, P., Geffert, L., and Ram, K. 2026. *rgbif*: Interface to the Global Biodiversity Information Facility API. R package version 3.8.5.¹
- Contreras-Díaz, R. G., Nori, J., Chiappa-Carrara, X., Peterson, A. T., Soberón, J., and Osorio-Olvera, L. 2023. Well-intentioned initiatives hinder understanding biodiversity conservation: Cloaked iNaturalist information for threatened species. *Biological Conservation*, 282, 110042.²
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D. 2007. The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45, RG2004.³
- Farnsworth, A., and Langham, G. 2025. Chocolate-vented Tyrant (*Neoxolmis rufiventris*), version 1.1. In *Birds of the World* (J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, Eds.). Cornell Lab of Ornithology, Ithaca, NY, USA.⁴
- Fick, S. E., and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302–4315.
- García-Roselló, E., González-Dacosta, J., and Lobo, J. M. 2023. The biased distribution of existing information on biodiversity hinders its use in conservation, and we need an integrative approach to act urgently. *Biological Conservation*, 283, 110118.
- Girón-Juárez, D. M., Contreras-Díaz, R. G., Medellín, R. A., Rojas-Soto, O., Díaz-Martínez, S., Nori, J., Martín, G., and Osorio-Olvera, L. 2025. Beyond climate: Anthropogenic pressures reshape vampire bat distributions across the Neotropics. *Biological Conservation*, 311, 111443.

¹ <https://CRAN.R-project.org/package=rgbif>

² <https://doi.org/10.1016/j.biocon.2023.110042>

³ <https://doi.org/10.1029/2005RG000183>

⁴ <https://doi.org/10.2173/bow.chvttyr2.01.1>

- Hernández-Hernández, N., Mota-Vargas, C., Lara, C., and Rojas-Soto, O. 2022. Seasonal distribution of the broad-tailed hummingbird (*Selasphorus platycercus*): a climatic approach. *Zoological Studies*, 61, 23.⁵
- Hijmans, R., Brown, A., and Barbosa, M. 2026. terra: Spatial Data Analysis. R package version 1.8-97.⁶
- Jahn, A. E., Cueto, V. R., Fontana, C. S., Guaraldo, A. C., Levey, D. J., Marra, P. P., and Ryder, T. B. 2020. Bird migration within the Neotropics. *The Auk*, 137, ukaa033.⁷
- Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E. 2008. Hole-filled SRTM for the Globe Version 4. CGIAR-CSI SRTM 90m Database.⁸
- Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Ruiz Gutierrez, V., Robinson, O. J., Miller, E. T., Fink, D., Wood, C., Davies, I., and Kelling, S. 2021. Analytical guidelines to increase the value of community science data: an example using eBird data to estimate species distributions. *Diversity and Distributions*, 27, 1265–1277.
- Karger, D. N., Conrad, O., Böhrer, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, P., and Kessler, M. 2017. Climatologies at high resolution for the Earth land surface areas. *Scientific Data*, 4, 170122.⁹
- Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage Publications Ltd, Thousand Oaks, California.
- Lomolino, M. V. 2004. Conservation biogeography. In *Frontiers of Biogeography: New Directions in the Geography of Nature* (M. V. Lomolino and L. R. Heaney, Eds.), pp. 293–296. Sinauer Associates, Sunderland, Massachusetts.
- Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U. G., Wint, G. W., Ray, S. E., Hay, S. I., Pigott, D. M., Shearer, F. M., Johnson, K., Earl, L., Marczak, L. B., Shirude, S., Davis Weaver, N., Gilbert, M., Velayudhan, R., Jones, P., Jaenisch, T., Scott, T. W., Reiner, R. C., and Hay, S. I. 2019. The current and future global distribution and population at risk of dengue. *Nature Microbiology*, 4, 1508–1515.
- Nori, J., Carrasco, P. A., and Leynaud, G. C. 2014. Venomous snakes and climate change: ophidism as a dynamic problem. *Climatic Change*, 122, 67–80.
- Pebesma, E., and Bivand, R. 2023. *Spatial Data Science: With Applications in R*. Chapman and Hall/CRC.¹⁰
- Peterson, A. T. 2006. Ecologic niche modeling and spatial patterns of disease transmission. *Emerging Infectious Diseases*, 12, 1822–1826.
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., and Araújo, M. B. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press.
- Soberón, J., and Peterson, A. T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 689–698.
- Strimas-Mackey, M., Ligocki, S., Auer, T., and Fink, D. 2026b. ebirdst: Access and Analyze eBird Status and Trends Data Products. R package version 3.2023.2.¹¹
- Strimas-Mackey, M., Miller, E., and Hochachka, W. 2026a. auk: eBird Data Extraction and Processing in R. R package version 0.9.2.¹²
- Wallace, A. R. 1876. *The Geographical Distribution of Animals*. Harper & Brothers, New York.
- Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., and Willis, K. J. 2005. Conservation biogeography: assessment and prospect. *Diversity and Distributions*, 11, 3–23.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Merow, C., Maitner, B., Schmatz, D. R., Schröder, B., Ritter, C., and Zimmermann, N. E. 2020. A standard protocol for reporting species distribution models. *Ecography*, 43, 1261–1277.

⁵ <https://doi.org/10.6620/ZS.2022.61-23>⁶ <https://github.com/rspatial/terra>⁷ <https://doi.org/10.1093/auk/ukaa033>⁸ <http://srtm.csi.cgiar.org>⁹ <https://doi.org/10.1038/sdata.2017.122>¹⁰ <https://doi.org/10.1201/9780429459016>¹¹ <https://ebird.github.io/ebirdst/>¹² <https://cornelllabofornithology.github.io/auk/>