Nature Conservation

Nature Conservation 58: 11–30 (2025)
DOI: 10.3897/natureconservation.58.140644

Data Paper

MIAU: An analysis-ready dataset on presence-only and presence-absence data of Neotropical carnivores (Mammalia, Carnivora) from 2000 to 2021

Florencia Grattarola[1], Kateřina Tschernosterová[1], Petr Keil[1]

1 *Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Praha - Suchdol, 16500, Czech Republic*
Corresponding author: Florencia Grattarola (flograttarola@gmail.com)

Abstract

In the last decade, databases of records of species observed at the same location at different points in time over large spatial extents have been made available. Unfortunately, these sources are scarce in regions such as Latin America. We present a dataset of 60,179 point occurrences (i.e. presence-only data, PO) and 45,468 camera-trap survey records (i.e. presence-absence data, PA) for 63 species of carnivores of the Neotropical Region from 2000 to 2021. We collated the data from various sources, including 64 newly-digitised bibliographic references. We cleaned, taxonomically harmonised and standardised the data following the Darwin Core and Humboldt Core standards and present them here as csv files. We have also made these data fit for analyses by aggregating the data into two time periods (time1: 2000–2013 and time2: 2014–2021), with PO grid cell counts of 100 × 100 km and PA polygons of varying size, presented as geopackage files. These data can be used for large-scale species distribution models, calculation of population trends, extinction risk analyses and educational purposes.

Key words: Camera trap, data deficiency, Latin America, point occurrence, species distribution models

OPEN ACCESS

Academic editor: William Magnusson
Received: 31 October 2024
Accepted: 4 December 2024
Published: 20 January 2025

ZooBank: https://zoobank.org/B5133377-E0AF-4956-B12D-46B15B7B11A9

Citation: Grattarola F, Tschernosterová K, Keil P (2025) MIAU: An analysis-ready dataset on presence-only and presence-absence data of Neotropical carnivores (Mammalia, Carnivora) from 2000 to 2021. Nature Conservation 58: 11–30. https://doi.org/10.3897/natureconservation.58.140644

Introduction

To understand and monitor global biodiversity change over time, we need data on species distributions spanning long time periods and large spatial areas (Yoccoz et al. 2001; Schmeller et al. 2017; Wüest et al. 2020). In the last decade, several databases holding these types of data have been developed. For example, BioTIME (Dornelas et al. 2018), the Living Planet Index (WWF 2022) and Breeding Birds Surveys from North America (Ziolkowski et al. 2022) or the UK (Heywood et al. 2024). Unfortunately, these sources are mostly biased towards Europe and the US, while they are practically non-existent in regions such as Latin America (Meyer et al. 2016).

To study changes in the distribution of continental-wide species in such data-scarce areas, there are two options. First, we can gather data from scratch, but this is challenging at large scales. Alternatively, we can rescue and collate multiple already-published sources and digitise, clean, harmonise and standardise them for reuse (Griffin 2015). This is the approach we followed to study the range dynamics of Neotropical carnivores. With the data, we first assessed the temporal change in the geographic distribution of the jaguarundi, *Herpailurus yagouaroundi*, through a novel modelling approach that integrates presence-only and presence-absence data (Grattarola et al. 2023). Later, we used similar data for modelling multiple carnivore species and calculated hotspots of change, i.e. areas where range contractions and expansions are accumulating more (Grattarola et al. 2024). The data we gathered can also be used for different new purposes.

An important lesson from these analyses and also from other authors (e.g. Elith et al. 2020) is the value of having different types of data systematically revised and accessible, particularly with presence-only and presence-absence data organised in a way that allows them to be used together effectively. This is important, particularly in relatively under-studied regions where each of the data is relatively scarce, but together they form a basis for robust analyses; specifically, the presence-only data provide better geographical coverage, while the presence-absence data can be used to calibrate predictions to a probability scale (Guillera-Arroita et al. 2015). Another (painful) lesson is how much time one must spend on the process of gathering, cleaning and preparing the data for the purpose of statistical analyses. We believe that the scientific community would benefit from having access to a dataset that has already undergone such treatment.

Here, we present MIAU, a ready-to-use, cleaned, continental-wide dataset on presences-only and presences-absences of Neotropical carnivores. We expect the dataset to be useful for many purposes, such as large-scale species distribution models, calculation of population trends and extinction risk analyses, ultimately providing information for large-scale conservation of one of the most charismatic taxonomic groups in the Neotropics.

Methods

We compiled data from different sources (Table 1), for 63 species of carnivores from the Neotropical Region (Table 2). We included records from the GBIF database (GBIF.org 2024), records extracted from a data paper (Nagy-Reis et al. 2020) and from 64 digitised literature sources (see Suppl. material 1 for a complete list). This last is one of the most important contributions of our dataset. We have mobilised 10,684 records from literature and made them publicly available here for the first time. We cleaned, taxonomically harmonised (according to Mammal Diversity Database 2022) and standardised all records following the Darwin Core (https://dwc.tdwg.org/; Wieczorek et al. 2012) and Humboldt

Table 1. Data sources in our dataset, including the source type, number of datasets involved, data type, number of species and number of records they span.

Source	Source type	Datasets involved	Data type	Number of species	Number of records
GBIF.org (2024)	online database	434	presence-only	59	56,413
Nagy-Reis et al. (2020)	data paper	105	presence-only	31	3,766
Nagy-Reis et al. (2020)	data paper	207	presence-absence	45	34,784
Literature sources processed in this study (Suppl. material 1)	literature	64	presence-absence	40	10,684

Table 2. List of species covered by our dataset, including family and the number of presence-only (PO) and presence-absence (PA) records (only reported presences).

Species	Family	Number of PO records	Number of PA records
Atelocynus microtis	Canidae	41	290
Canis latrans	Canidae	2288	89
Canis lupus	Canidae	241	0
Cerdocyon thous	Canidae	4332	4065
Chrysocyon brachyurus	Canidae	575	480
Lycalopex culpaeus	Canidae	8561	41
Lycalopex fulvipes	Canidae	72	0
Lycalopex grisea	Canidae	80	11
Lycalopex gymnocerca	Canidae	2087	508
Lycalopex sechurae	Canidae	13	34
Lycalopex vetula	Canidae	91	127
Speothos venaticus	Canidae	60	64
Urocyon cinereoargenteus	Canidae	2868	204
Vulpes macrotis	Canidae	132	0
Herpailurus yagouaroundi	Felidae	1171	787
Leopardus colocola	Felidae	183	1
Leopardus geoffroyi	Felidae	302	249
Leopardus guigna	Felidae	509	6
Leopardus guttulus	Felidae	34	742
Leopardus jacobitus	Felidae	1	0
Leopardus pajeros	Felidae	8	0
Leopardus pardalis	Felidae	3371	4242
Leopardus tigrinus	Felidae	379	224
Leopardus wiedii	Felidae	471	918
Lynx rufus	Felidae	1055	56
Panthera onca	Felidae	1230	2074
Puma concolor	Felidae	3999	2478
Conepatus chinga	Mephitidae	629	189
Conepatus leuconotus	Mephitidae	406	71
Conepatus semistriatus	Mephitidae	578	380
Mephitis macroura	Mephitidae	569	28
Mephitis mephitis	Mephitidae	210	0
Spilogale angustifrons	Mephitidae	161	6
Spilogale gracilis	Mephitidae	106	11
Spilogale pygmaea	Mephitidae	16	3
Eira barbara	Mustelidae	2957	2223
Galictis cuja	Mustelidae	374	101
Galictis vittata	Mustelidae	196	44
Lontra canadensis	Mustelidae	1	0
Lontra longicaudis	Mustelidae	959	86
Lontra provocax	Mustelidae	81	1
Neogale felipei	Mustelidae	1	0
Neogale frenata	Mustelidae	251	26
Pteronura brasiliensis	Mustelidae	272	22
Taxidea taxus	Mustelidae	154	0
Bassaricyon alleni	Procyonidae	32	1
Bassaricyon gabbii	Procyonidae	53	0
Bassaricyon medius	Procyonidae	11	0
Bassaricyon neblina	Procyonidae	26	0
Bassariscus astutus	Procyonidae	1113	40
Bassariscus sumichrasti	Procyonidae	63	1

Species	Family	Number of PO records	Number of PA records
Nasua narica	Procyonidae	5127	303
Nasua nasua	Procyonidae	2620	2908
Nasua olivacea	Procyonidae	275	3
Potos flavus	Procyonidae	815	17
Procyon cancrivorus	Procyonidae	2681	1661
Procyon lotor	Procyonidae	2098	120
Procyon pygmaeus	Procyonidae	1	0
Tremarctos ornatus	Ursidae	2753	26
Ursus americanus	Ursidae	436	22

Core (https://eco.tdwg.org/; Guralnick et al. 2018) standards. For a complete list of columns/terms and definitions, see the metadata files ('Data availability: underlying data'). In total, our dataset contains 105,647 records (Table 1). The data and code used to process the data of this manuscript can be accessed at https://github.com/bienflorencia/MIAU (Grattarola et al. 2024).

This dataset was generated to study the range dynamics of eight Neotropical carnivores (Grattarola et al. 2023, 2024).

Presence-absence (PA) data

We extracted the PA data from two main data sources. The first was the database of Nagy-Reis et al. (2020) (207 surveys). The second was our own manual extraction of data from primary published sources such as camera-trap surveys (64 surveys). Details on both data sources follow:

The Nagy-Reis et al. (2020) database of Neotropical carnivores collates records from different heterogeneous sources (e.g. researchers, governmental agencies, non-governmental organisations and private consultants) and methods (e.g. camera trapping, museum collections, roadkill, line transect and opportunistic records). From this database, we kept the data generated by surveys using camera traps (with detection and non-detection values), geographic coordinates, information about the study sampling area, with starting and ending month and year of the study and reported sampling effort (i.e. the number of active camera-trap days). We calculated 'temporalSpan' as the difference in days from 'end date of the study - start date of the study', assuming that the studies started the first day of the month and ended the last day of the month, as the start and end day were not recorded in the data. We considered the size of the study area as either the reported area for studies at the level of 'UNIT' or the latitude/longitude precision in metres for the studies at the sampling level of 'AREA' (see the metadata in Nagy-Reis et al. (2020) for more details on these definitions). Finally, we standardised column names and harmonised the taxonomy.

For the literature data extraction, we explored 262 potential sources and kept 64 (see 'Data availability: source data') that included studies in the Neotropics using camera traps that were performed from the 2000s onwards, reported all surveyed species and stated the sampling effort and the study area. We excluded studies that were exclusively focused on arboreal species, reported only some focal species and discarded others, and used a combination of sampling methods for which the effort in camera-trap days to detect a species was not possible to extract or calculate. We excluded further 22 studies for being duplicated sources and did not digitise 18 studies that were located in areas where we already had sufficient data. For all studies, we report the presence/absence of the species under 'presence'. For those that included an abundance metric, we report it under the 'abundance' column and report the abundance units used in 'abundanceUnits' (e.g. NOIR: number of individual records, RAI: relative abundance index - number of records per trap effort, AI/month: abundance index per month). Digitisation of the literature data represented a huge challenge as the different sources reported the spatial information, sampling effort and sampling period of the studies in very heterogeneous and incomplete ways and many times, they did not provide the primary data, but aggregated information. Therefore, as we often had to estimate or calculate these values, we report the origin of the information about effort, area of study and time span in specific columns. The column 'areaOrigin' refers to whether the area was given in the article, estimated from information provided in the article or calculated by manually georeferencing the study area or extracting the information from WDPA (UNEP-WCMC 2022) if it was a protected area. When the coordinates of individual camera traps were available, but no study area was provided, we considered the area as 10 m^2 (circle with a radius of 1.784 m) per camera (i.e. the estimated effective area of a camera). When the survey area was provided as a transect (either mentioning the centroid and length or with a map), we estimated the study area as 'transect length * 3.568 m' (diameter of effective camera trap area). The column 'effortOrigin' refers to whether the sampling effort was given in the article or calculated from the information provided. When the effort was not given, we calculated it either by multiplying the 'number of camera traps' by the 'sampling effort in days' for each individual camera or as the 'number of camera traps' by the 'temporal span in days', depending on which information was provided in the article. The column 'temporalSpanOrigin' refers to whether the temporal span of the study was either given in the article or calculated (as 'end date of the study - start date of the study').

Presence-only (PO) data

We extracted the PO data from two main data sources, the GBIF's (2024) database (56,413 records) and the Nagy-Reis et al.'s (2020) database (3,766 records). Details follow:

We extracted occurrence records from the GBIF database (GBIF.org 2024), the largest open data aggregator of primary biodiversity data. We downloaded all records available for the 71 species in the Neotropics using 'rgbif' (Chamberlain et al. 2024), considering records with geographic coordinates and no spatial issues, that were not fossil records nor living specimens and which establishment means was not managed, introduced, naturalised or invasive. These data were then further cleaned by removing records with coordinate uncertainty greater than 25,000 metres and filtering those records that belonged to *Canis lupus familiaris* or *Canis familiaris* (domestic dog). We also used 'CoordinateCleaner' (Zizka et al. 2019) to remove records within 2 kilometres of country and capital centroids (i.e. a common georeference generalisation), within 2 km of zoo and herbaria and records that fell outside the landmass and outliers (i.e. records located more than 1000 km to all other records of a species). Finally, we discarded duplicates considering independent records as species recorded on different dates and latitude and longitude coordinates.

For the records that had a sampling period as the collection/observation date, we considered the first day of the survey as the event date. From the total 209,277 records downloaded we ended up keeping 56,413 (27%). The largest exclusion of records was done when considering duplicates, yet many records also lacked critical information in terms of taxon, time and location (Peterson et al. 2018). The largest data provider of the final data is the iNaturalist citizen-science platform (47.1%).

We complemented the GBIF data with records from the Nagy-Reis et al. (2020) database. We kept records that reported count data (i.e. not data on detection/non-detections) and that had been collected using one of the following methods: 'line transect', 'active searching', 'roadkill', 'museum', 'opportunistic'. As the day of the event date was not stated for these data, we considered the 1st of the month to be the day of the event and discarded those that did not have an event date. Finally, we removed duplicates considering independent records as species recorded on different dates and latitude and longitude coordinates.

Analysis-ready data

The main goal of creating this analysis-ready data was to use it in the studies Grattarola et al. (2023, 2024) to model the temporal dynamics of the carnivores in the Neotropics (i.e. how their geographic ranges vary over time). With this purpose, we generated a dataset for each data type (PO and PA) (Table 3) based on different spatial features (i.e. grid cells of presence-only counts and presence-absence polygons) (Fig. 1) and splitting the data into two time periods, time1: 2000–2013 and time2: 2014–2021. This temporal division was the minimum possible to run our statistical models efficiently, given the low number of data points we had for some species, but it was also chosen to be able to represent, on average, 50% of the data in each period for our temporal change analyses. See more details in Grattarola et al. (2023) and Grattarola et al. (2024).

The PO data consist of 100 × 100 km grid cells with counts for each species in the two time periods. We provide the code to generate the grid-cells, which can be adapted to any other preferred size and temporal extension.

The PA data consists of polygons of varying sizes of aggregated camera-trap studies per time period. To create them, we generated a buffer polygon for each survey using the latitude and longitude of the survey as centroid and the study area as buffer. Then, all overlapping polygons were combined and absences were generated for each species in those polygons where the species was not recorded. For each polygon at each time period, we calculated the total surface area, timespan and the effort in camera-trap days and the aggregated presence for each species. We also provide the code to generate the polygons to split the data at any other preferred temporal split (e.g. biannual or every five years).

Table 3. Summary of the analysis-ready data, including the data type, spatial features and spatial and temporal resolution.

Data type	Spatial features	Spatial resolution	Temporal resolution
Presence-only (PO)	2,265 grid cells with counts per species	100 × 100 km	2 time periods (2000 to 2013 and 2014 to 2021)
Presence-absence (PA)	565 polygons of presences/absences values per each species	Varying sizes	2 time periods (2000 to 2013 and 2014 to 2021)

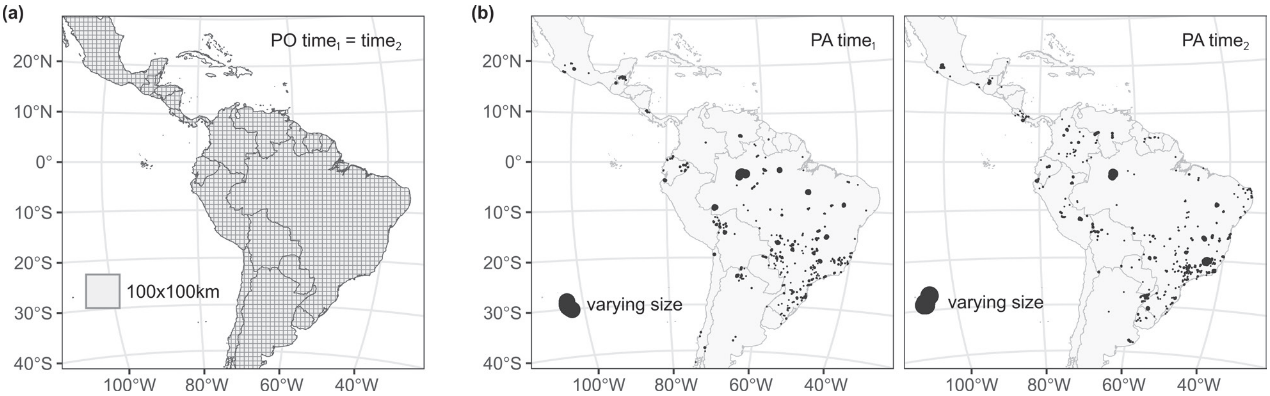

Figure 1. Spatial features (geometries) of the analysis-ready **a** presence-only (100 × 100 km grid cells) and **b** presence-absence data (aggregated polygons of varying sizes). PA polygons were buffered by 20 km to improve visibility.

Dataset coverage and limitations

Geographic coverage

The data cover the entire Neotropical Region, spanning 26 countries from Central to South America (Fig. 2). For the presence-only data, the country with the highest number of records is Mexico (n = 15,409), followed by Colombia (n = 12,780) and Chile (n = 10,632), while the countries with the least number of records are Venezuela (n = 31) and Guyana (n = 29). If we consider the density of records per area, then Costa Rica is the best-covered country (58.6 records /1,000 km^2), followed by Chile (15.6 records /1,000 km^2), while the poorest covered country is Venezuela (0.04 records /1,000 km^2) (Suppl. material 2). In the case of the presence-absence data, the country with the most records is Brazil (n = 17,341 detected presences), followed by Bolivia (n = 1,858) and Argentina (n = 1,365), while those with the least number of detected presences are Honduras and Belize (n = 5 each). Considering the density of surveys per area, then again, Costa Rica is the best-covered country (0.157 surveys/1,000 km^2), while Venezuela and Chile are the worst-covered (0.001 surveys/1,000 km^2) (Suppl. material 2).

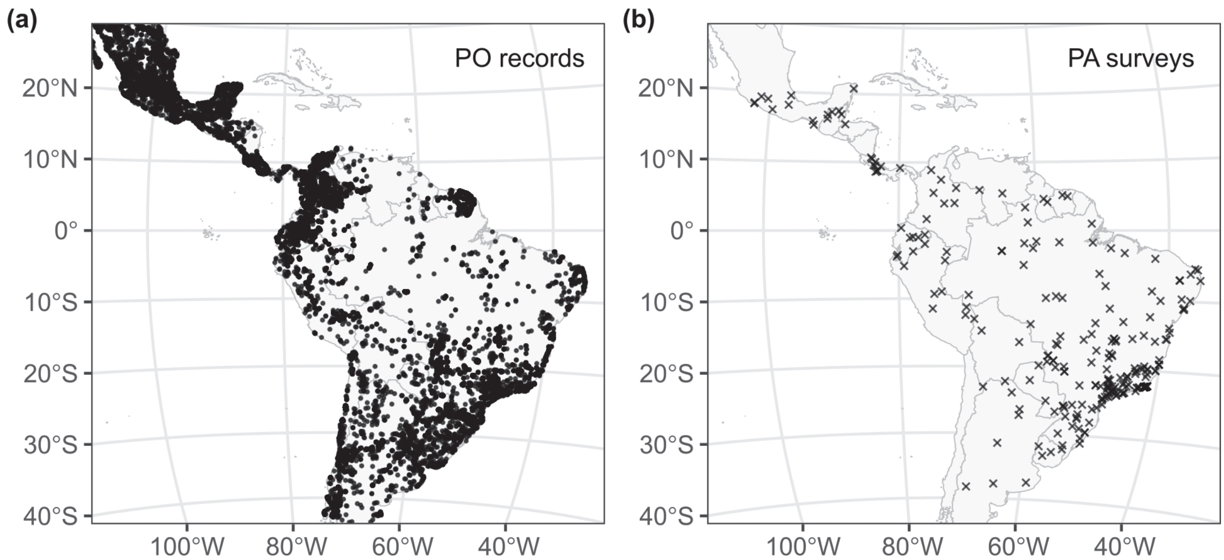

Figure 2. Distribution of the **a** presence-only (PO) occurrence records (n = 60,179) and **b** presence-absence (PA) surveys (n = 271) of carnivore species from the Neotropics.

Taxonomic coverage

The dataset includes a total of 63 species of carnivores native to the Neotropics (Fig. 3); the PO data spans 60 species, while the PA data includes 49 species. Species names were harmonised to follow the Mammal Diversity Database of the American Society of Mammalogists (Mammal Diversity Database 2022). According to this database, there are 71 carnivore species in the region (12 Canidae, 13 Felidae, 3 Mephitidae, 6 Mustelidae, 8 Procyonidae and 1 Ursidae); thus, our dataset spans 88.7% of the species recorded in the Neotropics (Table 2).

The following species are not included in our dataset: *Leopardus braccatus*, *L. fasciatus* and *L. garleppi* (from the *Leopardus colocola* complex), *L. narinensis* and *L. emiliae* (from the *Leopardus tigrinus* complex), *Spilogale interrupta*, *S. leucoparia* and *S. yucatanensis* and *Enhydra lutris*. The following species are poorly covered by our dataset (only a few records are included): *Leopardus pajeros* and *L. colocola* (from the *Leopardus colocola* complex), *Leopardus jacobita*, *Lyncodon patagonicus*, *Lontra canadensis*, *Neogale felipei* and *N. africana*. This is because most of these species are distributed either in Mexico or Argentina (northern and southernmost countries of the Neotropics) and they are not exclusive to the Neotropics or abundant in this region. Other species, such as those in the pampas cat (*Leopardus colocola*) species complex and the tiger cat (*Leopardus tigrinus*) species complex, have gone through several recent taxonomic changes and rediscoveries (Nascimento et al. 2020; Astorquiza et al. 2023; Lescroart et al. 2023) that make the correct assignment of the geographic distribution of each record a challenge.

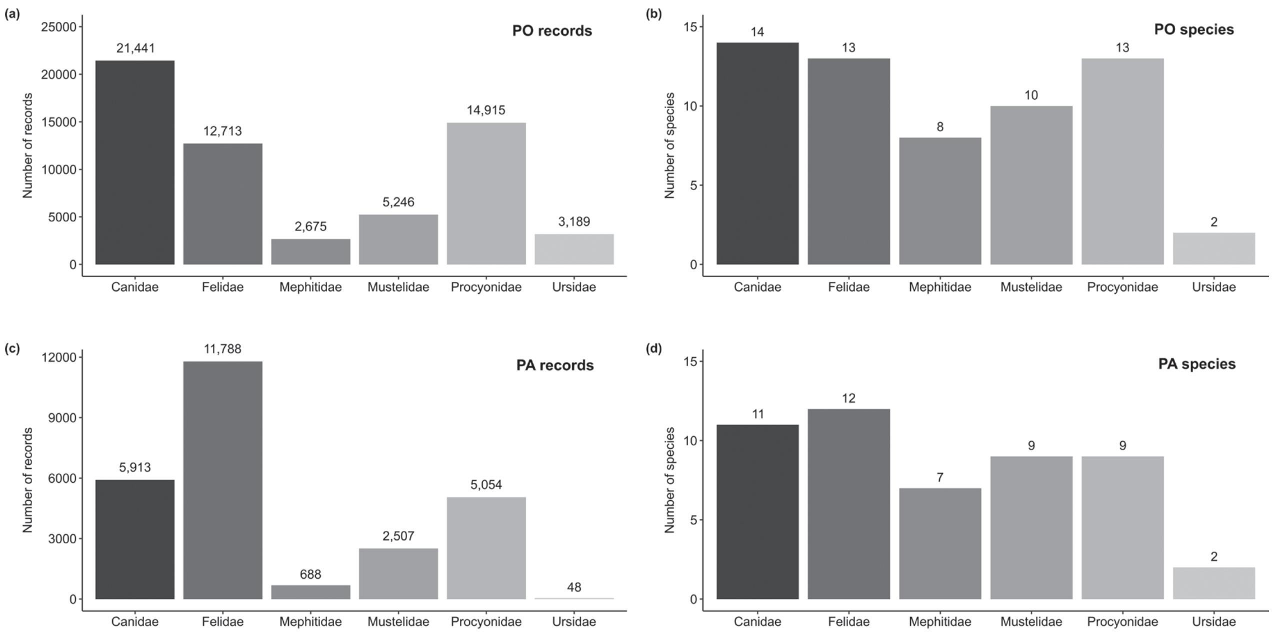

Figure 3. Number of records (**a**) and species (**b**) for the presence-only (PO) data and number of records (**c**) and species (d) for the presence-absence (PA) data in our dataset. Shown in greyscale are the carnivores' family.

Temporal coverage

The data have been observed from 01-01-2000 to 31-12-2021, with a more intense effort over the last five years (Fig. 4).

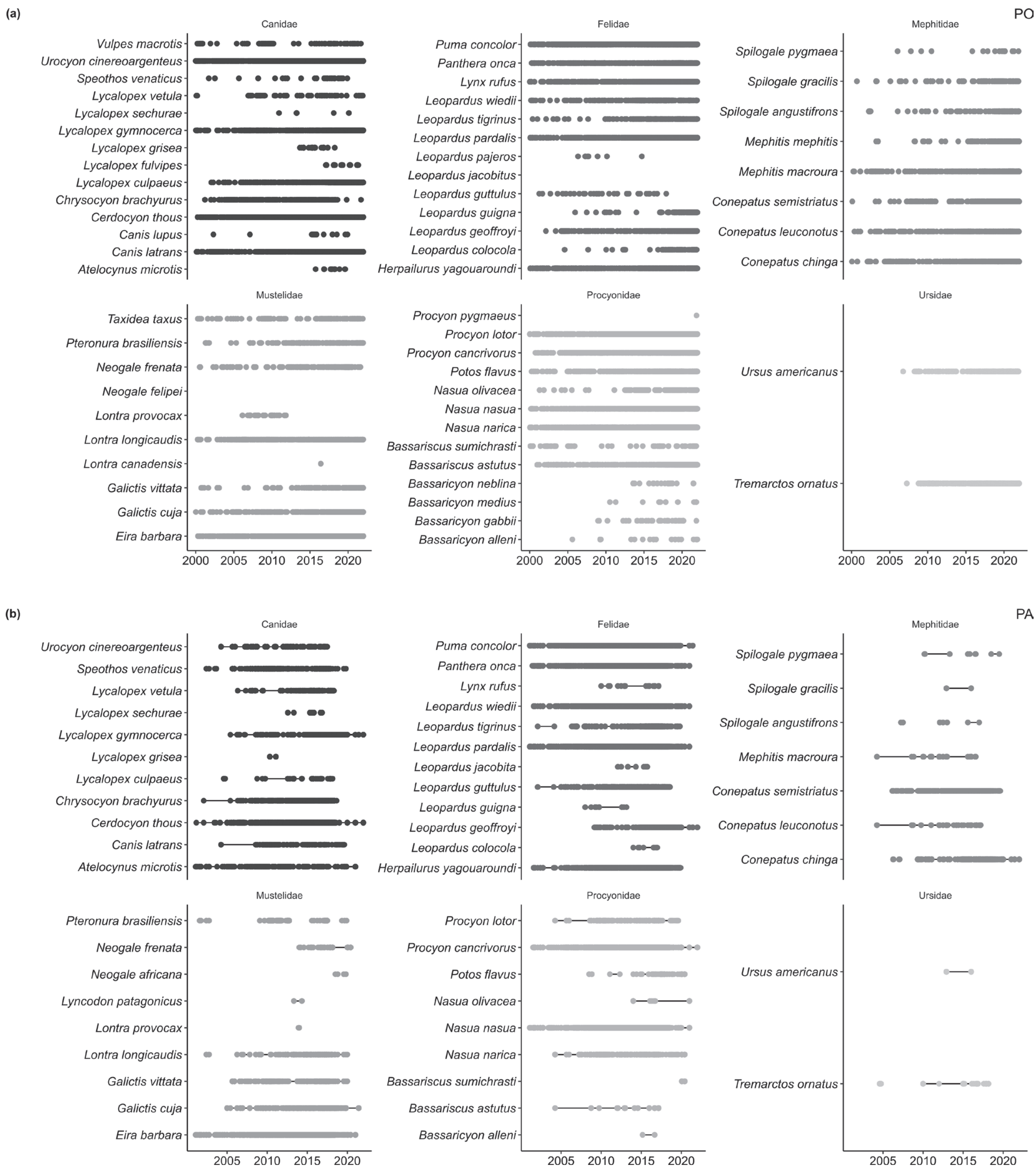

Figure 4. Number of (a) presence-only (PO) and (b) presence-absence (PA) records of carnivore species from the Neotropics over the years. For the PA data, the start and end of the survey is displayed as a segment. Shown in greyscale are the carnivores' family.

Limitations

Geographic

Some countries are poorly covered in our dataset (i.e. have fewer records than expected), while others are well-covered (Fig. 1, Suppl. material 2). These differences mostly arise because countries in Latin America have different data-sharing capacities (with Brazil and Colombia at the top, De Vega et al. 2020; Soberón 2022) and citizen-science levels of participation (Ortega-Álvarez and Casas 2022), yet the disparity does not necessarily reflect real differences in species richness or abundance per country. Due to this, but also because of the limited access to some areas of the continent, certain regions/biomes are likewise unevenly covered by our dataset (e.g. Amazon, Fig. 2). Thus, we recommend data users be cautious and account for this uneven sampling effort between countries and regions when using the data. An example of how to account for this is what we did in Grattarola et al. (2023) and Grattarola et al. (2024). We expected that highly accessible grid cells would have more point records than inaccessible grid cells and that differences would also vary amongst countries. Thus, we included the country of origin of the record and the degree of accessibility to the area in the observation process of our integrated species distribution model. If accounting for sampling bias is unfeasible, users could exclude sites with sampling effort below a set threshold to ensure adequate coverage for posterior analyses. However, how to address data gaps will depend on the specific research question (Bowler et al. 2024).

Temporal

Most PO records come from the second time period. This is characteristic of the data made available on GBIF (i.e. an artefact) and not a real difference in species abundance over time. In Grattarola et al. (2024), we found that the number of records between 2000–2013 and 2014–2021, using all the data available in GBIF for the eight studied species, was on average 27% higher in the second period. Thus, we recommend data users be cautious and account for this increasing sampling effort over the years. An example can be to include the ratio of the number of records between periods in the model, as we did in Grattarola et al. (2024). More solutions for data gaps can be explored at Bowler et al. (2024).

Taxonomic

Although we cover 87.7% of the species recorded in Neotropical countries (63 out of 71), for those species that are not exclusively distributed there (i.e. they are primarily distributed in the Nearctic Region, Mammal Diversity Database 2022), the data may not be representative of their entire distribution, but only their distribution in this region. Thus, for the following species, we caution users to consider our data insufficient for single species distribution analyses: *Canis latrans*, *Canis lupus*, *Vulpes macrotis*, *Lynx rufus*, *Puma concolor*, *Conepatus leuconotus*, *Mephitis macroura*, *M. mephitis*, *Enhydra lutris*, *Spilogale angustifrons*, *S. gracilis*, *S. interrupta*, *S. leucoparia*, *S. pygmaea*, *Lontra canadensis*, *Taxidea taxus*, *Bassariscus astutus*, *Procyon lotor* and *Ursus americanus*. Many of the studies we digitised from literature were conducted in forest habitats; thus, grassland and riverine species (such as *Lycalopex gymnocerca* and *Lontra longicaudis*) may also be under-represented in our data. For several rare species, our dataset has very few records; thus, the same previous reasoning applies. These are: *Procyon pygmaeus* and *Spilogale pygmaea* endemic from Mexico, *Lycalopex vetula* and *Leopardus emiliae* endemic from Brazil, *Lycalopex fulvipes* and *Leopardus colocola* endemic from Chile and *Leopardus narinensis* endemic from Colombia. Despite we collated the literature data following the same strategy with which we filtered the Nagy-Reis data paper, in the case of the literature data, we also did not consider data that only reported a few species from the total found in the survey (e.g. only the jaguar was reported). As we could not know the taxonomic scope of the surveys in the Nagy-Reis data paper, an unintended bias could arise from this.

Future directions

As the project has ended, we do not plan to update the dataset soon. However, with our data structure description, detailed data cleaning and standardisation workflow and code available, we encourage future users to update the dataset as needed.

Data availability

1) Source data: the downloaded/digitised data sources

- 64 literature sources. See the files 'literature_all_references.ods' for a complete list and 'literature_digitised_references.bib' for the BibTeX bibliographical database. See also Suppl. material 1.
- GBIF.org. (2024). 'Occurrence Download Neotropical Carnivores'. https://doi.org/10.15468/dl.67zvau.
- Nagy-Reis et al. (2020). 'NEOTROPICAL CARNIVORES: A Data Set on Carnivore Distribution in the Neotropics'. Ecology 101(11): e03128. https://doi.org/10.1002/ecy.3128.

2) Underlying data: the data we generated

Tables

- `data_PO.csv`: a csv file with the cleaned, standardised and harmonised presence-only data.
- `metadata_PO.csv`: a csv file with the column names, standard terms (e.g. Darwin Core or Humboldt Core) and definitions for the presence-only data.
- `data_PA.csv`: a csv file with the cleaned, standardised and harmonised presence-absence data.
- `metadata_PA.csv`: a csv file with the column names, standard terms (e.g. Darwin Core or Humboldt Core) and definitions for the presence-absence data.
- `carnivores.csv`: a csv file with the carnivore species' list extracted from the Mammal Diversity Database (2022), including the family, taxon key from GBIF and IUCN conservation status.

Spatial files

- `PO.gpk`: a geopackage with 2 layers; `time1`: a multi polygon sf file with 2,265 grid cells of 100 × 100 km resolution with counts per species in the temporal period from 2000 to 2013 and `time2`: a multi polygon sf file with 2,265 grid cells of 100 × 100 km resolution with counts per species in the temporal period from 2014 to 2021. Projection: Lambert azimuthal equal-area projection; centre latitude 0°S and centre longitude 73.125°W.

- `PA.gpk`: a geopackage with 2 layers; `time1`: a multi polygon sf file with 565 varying size polygons of presences/absences values for each species, area of the polygon and sampling effort in days in the temporal period from 2000 to 2013 and `time2`: a multi polygon sf file with 1013 varying size polygons of presences/absences for each species in the temporal period from 2014 to 2021, with the area of the polygon and sampling effort in days. Projection: Lambert azimuthal equal-area projection; centre latitude 0°S and centre longitude 73.125°W.
- `latam.gpk`: a geopackage with 4 layers; `countries`: a multi polygon sf file for all 27 Latin American countries, `countries_land` a multi polygon sf file for the 21 landmass countries of Latin America (excluding islands), `latam` a single polygon that combines all the landmass countries of Latin America and `latam_grids` a multi polygon sf file with 2,265 grid cells of 100 × 100 km resolution and `latam` as extension. Projection: Lambert azimuthal equal-area projection; centre latitude 0°S and centre longitude 73.125°W.

Other files

- 'literature_digitised_references.bib': BibTeX bibliographical database file with the 64 literature references digitised and included in our database.
- 'literature_digitised_references.csv': a csv file with the 64 literature references digitised and included in our database (see also Suppl. material 1).
- 'literature_all_references.ods': an open-source spreadsheet file with literature references (title and DOI or URL) including 4 sheets; `articles_EXCLUDED` articles that did not fulfil our assumptions and were excluded (reasons are reported in the column notes), `articles_DUPLICATED`: articles that were found in the reference lists of other datasets already digitised (e.g. Nagy-Reis et al. 2020), `articles_DIGITISED` articles that were digitised and included in the data and `articles_TO_PROCESS`: articles that fulfil our assumptions, but were not digitised.

Citation: Grattarola F, Tschernosterová K, Keil P (2024) MIAU: An analysis-ready dataset on presence-only and presence-absence data of Neotropical carnivores (Mammalia, Carnivora) from 2000 to 2021; Zenodo; https://doi.org/10.5281/zenodo.14278694. [Dataset].

If you use our underlying data, please cite the source data as well.

3) Extended Data: the code we used to process the data.

Quarto files

- `sources_species_and_countries.qmd`: an overview of the different sources, carnivore species and countries considered in the study.
- `presence-absence.qmd`: an overview of the presence-absence records in the database, including the geographic, taxonomic and temporal coverage of the data.
- `presence-only.qmd`: an overview of the presence-only records in the database, including the geographic, taxonomic and temporal coverage of the data.
- `analysis_ready_data.qmd`: a full descriptive code to reproduce the generation of `PO.gpk` and `PA.gpk`.

Citation: Grattarola F, Tschernosterová K, Keil P (2024) MIAU: An analysis-ready dataset on presence-only and presence-absence data of Neotropical carnivores (Mammalia, Carnivora) from 2000 to 2021; Zenodo; https://doi.org/10.5281/zenodo.14278694. [Code].

Acknowledgements

Thanks to Diego Alejandro Torres (Universidad de Caldas, Colombia), Marcelo Magioli (Instituto Pró-Carnívoros and Instituto Chico Mendes de Conservação da Biodiversidade, Brazil), Daniel Renison (Universidad Nacional de Córdoba, Argentina), Alexandra Cravino (Universidad de la República, Uruguay), Paul E. Ouboter (Institute for Neotropical Wildlife and Environmental Studies, Suriname) and María Florencia Aranguren (Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina) for providing extra information and useful comments on their records published in the literature.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

This work was funded by the European Union (ERC, BEAST, 101044740). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

FG conceptualised the work, designed the methodology, implemented the computer code, did the data standardisation and validation, prepared the data visualisation, supervised the data digitisation work and wrote the original draft. Literature data collation and digitisation was led by KT. FG and PK did the project administration. PK acquired funding. FG, KT and PK reviewed and edited the manuscript. The authors' contributions to the scholarly output followed the 'Contributor Roles Taxonomy' (CRediT; https://credit.niso.org/).

Author ORCIDs

Florencia Grattarola https://orcid.org/0000-0001-8282-5732
Kateřina Tschernosterová https://orcid.org/0009-0002-8097-8836
Petr Keil https://orcid.org/0000-0003-3017-1858

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Astorquiza JM, Noguera-Urbano EA, Cabrera-Ojeda C, Cepeda-Quilindo B, González-Maya JF, Eizirik E, Bonilla-Sánchez A, Buitrago DL, Pulido-Santacruz P, Ramírez-Chaves HE (2023) Distribution of the northern pampas cat, Leopardus garleppi, in northern South America, confirmation of its presence in Colombia and genetic analysis of a controversial record from the country. Mammalia 87(6): 606–614. https://doi.org/10.1515/mammalia-2022-0114

Bowler DE, Boyd RJ, Callaghan CT, Robinson RA, Isaac NJB, Pocock MJO (2024) Treating gaps and biases in biodiversity data as a missing data problem. Biological Reviews of the Cambridge Philosophical Society, brv.13127. https://doi.org/10.1111/brv.13127

Chamberlain S, Barve V, Mcglinn D, Oldoni D, Desmet P, Geffert L, Ram K (2024) rgbif: Interface to the Global Biodiversity Information Facility. R package version 3.8.0. https://CRAN.R-project.org/package=rgbif

De Vega JJ, Davey RP, Duitama J, Escobar D, Cristancho-Ardila MA, Etherington GJ, Minotto A, Arenas-Suarez NE, Pineda-Cardenas JD, Correa-Alvarez J, Camargo Rodriguez AV, Haerty W, Mallarino-Robayo JP, Barreto-Hernandez E, Muñoz-Torres M, Fernandez-Fuentes N, Di Palma F (2020) Colombia's cyberinfrastructure for biodiversity: Building data infrastructure in emerging countries to foster socioeconomic growth. Plants, People, Planet 2(3): 229–236. https://doi.org/10.1002/ppp3.10086

Dornelas M, Antão Laura H, Moyes F, Bates Amanda E, Magurran AE, Adam D, Akhmetzhanova AA, Appeltans W, Arcos JM, Arnold H, Ayyappan N, Badihi G, Baird AH, Barbosa M, Barreto TE, Bässler C, Bellgrove A, Belmaker J, Benedetti-Cecchi L, Bett BJ, Bjorkman AD, Błażewicz M, Blowes SA, Bloch CP, Bonebrake TC, Boyd S, Bradford M, Brooks AJ, Brown JH, Bruelheide H, Budy P, Carvalho F, Castañeda-Moya E, Chen CA, Chamblee JF, Chase TJ, Collier LS, Collinge SK, Condit R, Cooper EJ, Cornelissen JH C, Cotano U, Crow SK, Damasceno G, Davies CH, Davis RA, Day FP, Degraer S, Doherty TS, Dunn TE, Durigan G, Duffy JE, Edelist D, Edgar GJ, Elahi R, Elmendorf SC, Enemar A, Ernest SKM, Escribano R, Estiarte M, Evans BS, Fan T-Y, Farah FT, Fernandes LL, Farneda FZ, Fidelis A, Fitt R, Fosaa AM, Franco GADC, Frank GE, Fraser WR, García H, Gatti RC, Givan O, Gorgone-Barbosa E, Gould WA, Gries C, Grossman GD, Gutierréz JR, Hale S, Harmon ME, Harte J, Haskins G, Henshaw DL, Hermanutz L, Hidalgo P, Higuchi P, Hoey A, Van Hoey G, Hofgaard A, Holeck K, Hollister RD, Holmes R, Hoogenboom M, Hsieh C-h, Hubbell SP, Huettmann F, Huffard CL, Hurlbert AH, Ivanauskas NM, Janík D, Jandt U, Jażdżewska A, Johannessen T, Johnstone J, Jones J, Jones FAM, Kang J, Kartawijaya T, Keeley EC, Kelt DA, Kinnear R, Klanderud K, Knutsen H, Koenig CC, Kortz AR, Král K, Kuhnz LA, Kuo CY, Kushner DJ, Laguionie-Marchais C, Lancaster LT, Lee CM, Lefcheck JS, Lévesque E, Lightfoot D, Lloret F, Lloyd JD, López-Baucells A, Louzao M, Madin JS, Magnússon B, Malamud S, Matthews I, McFarland KP, McGill B, McKnight D, McLarney WO, Meador J, Meserve PL, Metcalfe DJ, Meyer CFJ, Michelsen A, Milchakova N, Moens T, Moland E, Moore J, Moreira CM, Müller J, Murphy G, Myers-Smith IH, Myster RW, Naumov A, Neat F, Nelson JA, Nelson MP, Newton SF, Norden N, Oliver JC, Olsen EM, Onipchenko VG, Pabis K, Pabst RJ, Paquette A, Pardede S, Paterson DM, Pélissier R, Peñuelas J, Pérez-Matus A, Pizarro O, Pomati F, Post E, Prins HHT, Priscu JC, Provoost P, Prudic KL, Pulliainen E, Ramesh BR, Ramos OM, Rassweiler A, Rebelo JE, Reed DC, Reich PB, Remillard SM, Richardson AJ, Richardson JP, van Rijn I, Rocha R, Rivera-Monroy VH, Rixen C, Robinson KP, Rodrigues RR, de Cerqueira Rossa-Feres D, Rudstam L, Ruhl H, Ruz CS, Sampaio EM, Rybicki N, Rypel A, Sal S, Salgado B, Santos FAM, Savassi-Coutinho AP, Scanga S, Schmidt J, Schooley R, Setiawan F, Kwang-Shao T, Shaver GR, Sherman S, Sherry TW, Siciński J, Sievers C, da Silva AC, da Silva FR, Silveira FL, Slingsby J, Smart T, Snell SJ, Soudzilovskaia NA, Souza G G, Souza FM, Souza VC, Stallings CD, Stanforth R, Stanley EH, Sterza JM, Stevens M, Stuart-Smith R, Suarez YR, Supp S, Tamashiro JY, Tarigan S, Thiede GP, Thorn S, Tolvanen A, Toniato MTZ, Totland Ø, Twilley RR, Vaitkus G, Valdivia N, Vallejo MI, Valone TJ, Colen CV, Vanaverbeke J, Venturoli F, Verheye HM, Vianna M, Vieira RP, Vrška T, Vu CQ, Vu LV, Waide RB, Waldock C, Watts D, Webb S, Wesołowski T, White EP, Widdicombe CE, Wilgers D, Williams R, Williams SB, Williamson M, Willig MR, Willis TJ, Wipf S, Woods KD, Woehler EJ, Zawada K, Zettler ML (2018) BioTIME: A database of biodiversity time series for the Anthropocene. Global Ecology and Biogeography 27(7): 760–786. https://doi.org/10.1111/geb.12729

Elith J, Graham C, Valavi R, Abegg M, Bruce C, Ferrier S, Ford A, Guisan A, Hijmans RJ, Huettmann F, Lohmann L, Loiselle B, Moritz C, Overton J, Peterson AT, Phillips S, Richardson K, Williams S, Wiser SK, Wohlgemuth T, Zimmermann NE (2020) Presence-only and Presence-absence Data for Comparing Species Distribution Modeling Methods. Biodiversity Informatics 15(2): 69–80. https://doi.org/10.17161/bi.v15i2.13384

GBIF.org (2024) Occurrence Download Neotropical Carnivores. https://doi.org/10.15468/dl.67zvau

Grattarola F, Tschernosterová K, Keil P (2024) bienflorencia/MIAU. https://doi.org/10.5281/zenodo.14278694

Grattarola F, Bowler DE, Keil P (2023) Integrating presence-only and presence–absence data to model changes in species geographic ranges: An example in the Neotropics. Journal of Biogeography 50(9): 1561–1575. https://doi.org/10.1111/jbi.14622

Grattarola F, Tschernosterová K, Keil P (2024) A continental-wide decline of occupancy and diversity in five Neotropical carnivores. Global Ecology and Conservation 55: e03226. https://doi.org/10.1016/j.gecco.2024.e03226

Griffin RE (2015) When are Old Data New Data? GeoResJ 6: 92–97. https://doi.org/10.1016/j.grj.2015.02.004

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA (2015) Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography 24(3): 276–292. https://doi.org/10.1111/geb.12268

Guralnick R, Walls R, Jetz W (2018) Humboldt Core–toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. Ecography 41(5): 713–725. https://doi.org/10.1111/ecog.02942

Heywood JJN, Massimino D, Balmer DE, Kelly L, Marion S, Noble DG, Pearce-Higgins JW, White DM, Woodcock P, Wotton S, Gillings S (2024) The Breeding Bird Survey 2023. British Trust for Ornithology, Thetford. https://www.bto.org/our-science/publications/breeding-bird-survey-report/breeding-bird-survey-2023 [July 8, 2024]

Lescroart J, Bonilla-Sánchez A, Napolitano C, Buitrago-Torres DL, Ramírez-Chaves HE, Pulido-Santacruz P, Murphy WJ, Svardal H, Eizirik E (2023) Extensive Phylogenomic Discordance and the Complex Evolutionary History of the Neotropical Cat Genus Leopardus. Molecular Biology and Evolution 40(12): msad255. https://doi.org/10.1093/molbev/msad255

Mammal Diversity Database (2022) Mammal Diversity Database. https://doi.org/10.5281/zenodo.5945626 [March 27, 2022]

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

Nagy-Reis M, Oshima JE de F, Kanda CZ, Palmeira FBL, de Melo FR, Morato RG, Bonjorne L, Magioli M, Leuchtenberger C, Rohe F, Lemos FG, Martello F, Alves-Eigenheer M, da Silva RA, Silveira dos Santos J, Priante CF, Bernardo R, Rogeri P, Assis JC, Gaspar LP, Tonetti VR, Trinca CT, Ribeiro AS, Bocchiglieri A, Hass A, Canteri A, Chiarello AG, Paglia AP, Pereira AA, de Souza AC, Gatica A, Medeiro AZ, Eriksson A, Costa AN, González-Gallina A, Yanosky AA, Jesus de la Cruz A, Bertassoni A, Bager A, Bovo AAA, Cravino Mol A, Bezerra AMR, Percequillo A, Vogliotti A, Costa Lopes AM, Keuroghlian A, Zúñiga Hartley AC, Devlin AL, de Paula A, García-Olaechea A, Sánchez A, Aquino ACMM, Srbek-Araujo AC, Ochoa AC, Tomazzoni AC, Lacerda ACR, Bacellar AEF, Campelo AKN, Herrera Victoria AM, Paschoal AMO, Potrich AP, Gomes APN, Olímpio APM, Cunha Costa AR, Jácomo ATA, Calaça AM, Jesus AS, de Barros Barban A, Feijó A, Pagoto A, Rolim AC, Hermann AP, Souza ASMC, Chein Alonso A, Monteiro A, Mendonça AF, Luza AL, Moura ALB, da Silva ALF, Lanna AM, Antunes AP, Nunes AV, Dechner A, Carvalho AS, Novaro AJ, Scabin AB, Gatti A, Nobre AB, Montanarin A, Deffaci ÂC, de Albuquerque ACF, Mangione AM, Pinto AMS, Mendes Pontes AR, Bertoldi AT, Calouro AM, Fernandes A, Ferreira AN, Ferreguetti AC, Rosa ALM, Banhos A, Francisco BSS, Cezila BA, Beisiegel BM, de Thoisy B, Ingberman B, Neves BS, Pereira-Silva B, Bertagni de Camargo B, Andrade BS, Santos BS, Leles B, Torres Parahyba Campos BA, Kubiak BB, França BRA, Saranholi BH, Pereira Mendes C, Cantagallo Devids C, Pianca C, Rodrigues C, Islas CA, de Lima CA, de Lima CR, Gestich CC, Tedesco CD, De Angelo C, Fonseca C, Hass C, Peres CA, Kasper CB, Durigan CC, Fragoso CE, Verona CE, Rocha CFD, Salvador CH, Vieira CL, Ruiz CEB, Cheida CC, Sartor CC, Espinosa CC, Fieker CZ, Braga C, Sánchez-Lalinde C, Machado CIC, Cronemberger C, Luna CL, Del Vechio C, Bernardo CSS, Hurtado CM, Lopes CM, da Rosa CA, Cinta CC, Costa CG, Zárate-Castañeda CP, Novaes CL, Jenkins CN, Seixas CS, Martin C, Zaniratto CP, López-Fuerte CF, da Cunha CJ, De-Carvalho CB, Chávez C, Santos CC, Polli DJ, Buscariol D, Carreira DC, Galiano D, Thornton D, Ferraz DS, Lamattina D, Moreno DJ, Moreira DO, Farias DA, Barros-Battesti DM, Tavares DC, Costa Braga D, Gaspar DA, Friedeberg D, Astúa D, Silva DA, Viana DC, Lizcano DJ, Varela DM, Loretto D, Gräbin DM, Eaton DP, Machado da Silva D, Dias DM, Camara EMVC, Barbier E, Chávez-González E, Rocha EC, Lima ES, Carrano E, Eizirik E, Nakano-Oliveira E, Rigacci ED, Santos EM, Venticinque EM, Alexandrino ER, Abreu Ribeiro E, Setz E, Rocha ECLD, Carvalho Jr EAR,

Rechenberg E, Fraga EC, Mendonça EN, D'Bastiani E, Isasi-Catalá E, Guijosa-Guadarrama E, Ramalho EE, González E, Hasui É, Saito EN, Fischer E, Aguiar EF, Rocha ES, Martínez Nambo ED, de la Peña-Cuéllar E, Castro ÉP, de Freitas EB, Pedó E, Rocha FL, Girardi F, Pereira FA, Soares FAM, Roque FO, Díaz-Santos FG, Patiu FM, do Nascimento FO, Keesen Ferreira F, Diaz-Santos F, Moreli Fantacini F, Pedrosa F, Pessoa da Silva F, Velez-Garcia F, Gomes FBR, Guedes da Silva F, Michalski F, de Azevedo FC, de Barros FC, Santos FS, Abra FD, Ramalho FP, Hatano FM, Anaguano-Yancha F, Gonçalves F, Pedroni F, Passos FC, Jacinavicius FC, Bonfim FCG, Puertas FH, Contreras-Moreno FM, Tortato FR, Santos FM, Chaves FG, Tirelli FP, Vilas Boas FE, Rodrigues FHG, Ubaid FK, Grotta-Neto F, Palomares F, Souza FL, Costa FE, França FGR, Ramírez Pinto F, Aguiar GL, Hofmann GS, Heliodoro G, Duarte GT, Ribeiro de Andrade G, Beca G, Zapata-Ríos G, Giné GAF, Powell GVN, Wilson Fernandes G, Forero-Medina G, Melo GL, Santana GG, Ciocheti G, Alves GB, Souto GHBO, Villarroel GJ, Porfirio GEO, Batista GO, Behling GM, Ayala Crespo GM, Mourão GM, Rezende GZ, Toledo GAC, Herrera HM, Alves Prado H, Bergallo HG, Secco H, Rajão H, Roig HL, Concone HVB, Duarte H, Ermenegildo H, Ferreira Paulino Neto H, Quigley H, Lemos HM, Cabral H, Fernandes-Ferreira H, del Castillo HF, Ribeiro IK, Coelho IP, Franceschi IC, Melo I, Oliveira-Bevan I, Mourthe I, Bernardi I, de la Torre JA, Marinho-Filho J, Martinez J, Palacios Perez JX, Pérez-Torres J, Bubadué J, Silveira JR, Seibert JB, Oliveira JF, Assis JR, De la Maza J, Hinojosa J, Metzger JP, Thompson JJ, Svenning J-C, Gouvea JA, Souza JRD, Pincheira-Ulbrich J, Nodari JZ, Miranda J, Zecchini Gebin JC, Giovanelli JGR, Rossi Junior JL, Pandini Favoretti JP, Villani JP, Just JPG, Souza-Alves JP, Costa JF, Rocha J, Polisar J, Sponchiado J, Cherem JJ, Marinho JR, Ziegler J, Cordeiro J, de Sousa e Silva Júnior J, Rodriguez-Pulido JA, Chaves dos Santos JC, dos Reis Júnior JC, Mantovani JE, Moreira Ramírez JF, Sarasola JH, Cartes JL, Duarte JMB, Longo JM, Dantas JO, Venancio JO, de Matos JR, Pires JSR, Hawes JE, Santos JG, Ruiz-Esparza J, Martínez Lanfranco JA, Rudolf JC, Charre-Medellin JF, Zanón-Martínez JI, Peña-Mondragón JL, Campos Krauer JM, Arrabal JP, Beduschi J, Ilha J, Mata JC, Bonanomi J, Jordao J, de Almeida-Rocha JM, Pereira-Ribeiro J, Zanoni JB, Bogoni JA, Chacón Pacheco JJ, Contreras Palma KM, Strier KB, Rodriguez Castro KG, Didier K, Schuchmann KL, Chávez-Congrains K, Burs K, Ferraz KMPMB, Juarez KM, Flesher K, Morais KDR, Lautenschlager L, Grossel LA, Dahmer LC, de Almeida LR, Fornitano L, Barbosa LNB, Bailey LL, Barreto LN, Villalba LM, Magalhães LM, Cullen Jr L, Marques L, Marques Costa L, Silveira L, Moreira LS, Sartorello L, Oliveira LC, Gomes LP, Aguiar LS, da Silva LH, Mendonça LS, Valenzuela LA, Benavalli L, Dias LCS, Munhoes LP, Catenacci L, Rampim LE, de Paula LM, Nascimento LA, Gonçalves da Silva L, Quintilham L, Ramis Segura L, Perillo LN, Rezende LR, Martínez Retta L, Rojas LNS, Guimarães LN, Araújo L, Zago da Silva L, Querido LCA, Verdade LM, Perera-Romero LE, Carvalho-Leite LJ, Hufnagel L, Rezende Bernardo LR, Oliveira LF, Oliveira Santos LGR, Lyra LH, Borges LHM, Severo MM, Benchimol M, Quatrocchi MG, Martins MZA, Rodrigues M, Penteado MJF, Figuerêdo Duarte Moraes M, Oliveira MA, Lima MGM, Pônzio MC, Cervini M, da Silva M, Passamani M, Villegas MA, dos Santos Junior MA, Yamane MH, Jardim MMA, Leite de Oliveira M, Silveira M, Tortato MA, Figueiredo MSL, Vieira MV, Sekiama ML, Andrade da Silva MA, Nuñez MB, Siviero MB, Carrizo MC, Barros MC, Barros MAS, do Rosário MCF, Peñuela Mora MC, Fleytas Jover MC, Morandi MEF, Huerta ME, Fernandes MEA, Viscarra Siñani ME, Iezzi ME, Ramos Pereira MJ, Gomez Vinassa ML, Lorini ML, Jorge MLSP, Morini MS, Guenther M, Landis MB, Vale MM, Xavier MS, Tavares MS, Kaizer M, Velilla M, Bergel MM, Hartmann MT, Lima da Silva M, Rivero M, Salles Munerato M, Xavier da Silva M, Zanin M, Marques MI, Haberfeld M, Di Bitetti

MS, Bowler M, Galliez M, Ortiz-Moreno ML, Buschiazzo M, Montes MA, Alvarez MR, Melo-Dias M, Reis MG, Corrêa MRJ, Tobler MW, Gompper ME, Nunez-Regueiro M, Brandão Vecchi M, Graipel ME, Godoi MN, Moura MO, Konzen MQ, Pardo MV, Beltrão MG, Mongelli M, Almeida MO, Gilmore MP, Schutte M, Faria MB, Luiz MR, de Paula M, Hidalgo-Mihart MG, Perilli MLL, Freitas-Junior MC, da Silva MP, Denkiewicz NM, Torres NM, Olifiers N, De Lima NDS, de Albuquerque NM, Canassa NF, de Almeida Curi NH, Prestes NP, Falconi N, Gurgel-Filho NM, Pasqualotto N, Cáceres NC, Peroni N, de la Sancha NU, Zanella N, Monroy-Vilchis O, Pays O, Arimoro OA, Ribeiro OS, Villalva P, Gonçalves PR, Santos PM, Brennand P, Rocha P, Akkawi P, Cruz P, Ferreira PM, Prist PR, Martin PS, Arroyo-Gerala P, Auricchio P, Hartmann PA, Antas PTZ, Camargo PHSA, Marinho PH, Ruffino PHP, Prado PI, Martins PW, Cordeiro-Estrela P, Luna P, Sarmento P, Faria Peres PH, Galetti Jr PM, de Castilho PV, Renaud P-C, Scarascia PO, Cobra PDPA, Lombardi PM, Bessa R, Reyna-Hurtado R, de Souza RCC, Hoogesteijn RJ, Alves RSC, Romagna RS, Silva RL, de Oliveira R, Beltrão-Mendes R, Alencar RM, Coutinho R, da Silva RC, Caribé Grando RLSC, Matos RG, Araujo RS, Pedroso RF, Durães RMN, Ribeiro RLA, Chagas R, Miotto R, Twardowsky Ramalho Bonikowski R, Muylaert RL, Pagotto RV, Hilário RR, Faria RT, Bassini-Silva R, Sampaio R, Sartorello R, Pires RA, Hatakeyama R, Bianchi RC, Buitenwerf R, Wallace R, Paolino RM, Fusco-Costa R, Trovati RG, Tomasi RJ, Espíndola Hack RO, Magalhães RA, Nobrega RAA, Nobre RA, Massara RL, Fróes RM, Araújo RPC, León Pérez RR, Jorge RSP, de Paula RC, Martins R, da Cunha RGT, Costa R, Alves RRN, Garcia-Anleu R, Santos Almeida RP, Cueva Loachamín RD, Andrade RS, Juárez R, Bordallo SU, Guaragni SA, Carrillo-Percastegui SE, Seber S, Astete S, Hartz SM, Espinosa S, Álvarez Solas S, Ramos Lima S, Silvestre SM, Machado SAS, Keuroghlian-Eaton S, Albanesi S, Costa SA, Bazilio S, Mendes SL, Althoff SL, Pinheiro SD, Napiwoski SJ, Fernández Ramirez S, Talamoni SA, Age SG, Pereira TC, Moreira TC, Trigo TC, Gondim TMS, Karlovic TC, Cavalcante T, Maccarini T, Rodrigues TF, de Camargo e Timo TP, Monterrubio TC, Piovezan U, Cavarzere V, Towns V, Onofrio VC, Oliveira VB, Araújo VC, Melo VL, Kanaan VT, Iwakami V, Vale V, Picinatto Filho V, Alberici V, Bastazini VAG, Orsini VS, Braz VS, Rojas Bonzi VB, Guedes Layme VM, Gaboardi VTR, Rocha VJ, Martins WP, Tomas WM, Hannibal W, Dáttilo W, Silva WR, Endo W, Bercê W, Bravata de la Cruz Y, Ribeiro YGG, Galetti M, Ribeiro MC (2020) NEOTROPICAL CARNIVORES: A data set on carnivore distribution in the Neotropics. Ecology 101(11): e03128. https://doi.org/10.1002/ecy.3128

Nascimento FOD, Cheng J, Feijó A (2020) Taxonomic revision of the pampas cat Leopardus colocola complex (Carnivora: Felidae): an integrative approach. Zoological Journal of the Linnean Society 191(2): 575–611. https://doi.org/10.1093/zoolinnean/zlaa043

Ortega-Álvarez R, Casas A (2022) Public participation in biodiversity research across Latin America: Dissecting an emerging topic in the Neotropics. Environmental Science & Policy 137: 143–151. https://doi.org/10.1016/j.envsci.2022.08.016

Peterson AT, Asase A, Canhos DAL, de Souza S, Wieczorek J (2018) Data Leakage and Loss in Biodiversity Informatics. Biodiversity Data Journal 6: e26826. https://doi.org/10.3897/BDJ.6.e26826

Schmeller DS, Böhm M, Arvanitidis C, Barber-Meyer S, Brummitt N, Chandler M, Chatzinikolaou E, Costello MJ, Ding H, García-Moreno J, Gill M, Haase P, Jones M, Juillard R, Magnusson WE, Martin CS, McGeoch M, Mihoub J-B, Pettorelli N, Proença V, Peng C, Regan E, Schmiedel U, Simaika JP, Weatherdon L, Waterman C, Xu H, Belnap J (2017) Building capacity in biodiversity monitoring at the global scale. Biodiversity and Conservation 26(12): 2765–2790. https://doi.org/10.1007/s10531-017-1388-7

Soberón J (2022) Biodiversity Informatics for Public Policy. The case of CONABIO in Mexico. Biodiversity Informatics 17. https://doi.org/10.17161/bi.v17i.18270 [June 9, 2023]

UNEP-WCMC (2022) Protected areas map of the world. www.protectedplanet.net

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. https://doi.org/10.1371/journal.pone.0029715

Wüest RO, Zimmermann NE, Zurell D, Alexander JM, Fritz SA, Hof C, Kreft H, Normand S, Cabral JS, Szekely E, Thuiller W, Wikelski M, Karger DN (2020) Macroecology in the age of Big Data – Where to go from here? Journal of Biogeography 47(1): 1–12. https://doi.org/10.1111/jbi.13633

WWF (2022) Living Planet Report 2022 – Building a nature-positive society. https://livingplanet.panda.org/en-GB/ [July 8, 2024]

Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. Trends in Ecology & Evolution 16(8): 446–453. https://doi.org/10.1016/S0169-5347(01)02205-4

Ziolkowski D, Lutmerding M, Aponte VI, Hudson M-AR (2022) 2022 Release - North American Breeding Bird Survey Dataset (1966–2021). https://doi.org/10.5066/P97WAZE5

Zizka A, Silvestro D, Andermann T, Azevedo J, Ritter CD, Edler D, Farooq H, Herdean A, Ariza M, Scharn R, Svanteson S, Wengstrom N, Zizka V, Antonelli A (2019) CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution 7(5): 744–751. https://doi.org/10.1111/2041-210X.13152

Supplementary material 1

Complete list of the 64 digitised literature sources

Authors: Florencia Grattarola, Kateřina Tschernosterová, Petr Keil

Data type: csv

Explanation note: See https://doi.org/10.5281/zenodo.14278694 for a BibTeX with the bibliographical database file.

Link: https://doi.org/10.3897/natureconservation.58.140644.suppl1

Supplementary material 2

Supplementary information

Authors: Florencia Grattarola, Kateřina Tschernosterová, Petr Keil

Data type: docx

Explanation note: **table S2.** Density of presence-only (PO) and presence-absence (PA) surveys per country. **figure S1.** Number of presence-only (PO) and presence-absence (PA) records per country and density of records and surveys per area (1/1,000 km2).

Link: https://doi.org/10.3897/natureconservation.58.140644.suppl2