

Integrating presence-only and presence-absence data to model changes in species geographic ranges: An example of yaguarundí in Latin America

Florencia Grattarola^{1,*}, Diana E. Bowler², Petr Keil¹

¹ Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Praha – Suchbátka, 16000, Czech Republic

² UK Centre for Ecology & Hydrology, Benson Ln, Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB, United Kingdom

* Correspondence: flograttarola@gmail.com

Abstract

Anthropogenic changes such as land use and climate change affect species' geographic ranges, causing range shifts, contractions, or expansions. However, data on range dynamics are insufficient, heterogeneous, and spatially and temporally biased in most regions. Integrated species distribution models (IDMs) offer a solution as they can complement good quality presence-absence data with opportunistically collected presence-only data, simultaneously accounting for heterogeneous sampling effort. However, these methods have seen limited use in the estimation of temporal change of geographic ranges and are not yet widespread as they have a steep learning curve. Here we present a generalisable model and case example to ease their adoption. Using data on presence-absence and presence-only on the yaguarundí (*Herpailurus yagouaroundi*), we modelled the species distribution at two time periods (2000-2013 and 2014-2021) using a Bayesian model based on Poisson point process in JAGS. Our model integrates the different types of data while accounting for varying sampling effort and spatial effect. We predicted the species range at the two time periods and quantified their changes. We found that between the two time periods, the yaguarundí has contracted its southern and northern range limits towards the equator, but expanded its area of distribution over the entire species' range. Also, our results show that modelled geographic range (either pre or post) is not entirely consistent with the current expert range map from IUCN. Our modelling approach provides a working example with the potential to address data gaps and biases in other taxa and regions. Given the increasing number of incidental data being generated by community-derived initiatives in Latin America, IDMs can become a valuable source for species distribution modelling in the region. To our knowledge, this is the first application of the IDM approach with temporal dimension and over the entire species' geographic range.

Keywords: species distribution models, Poisson point process, MCMC, autocorrelation, sampling bias, range size, data-poor regions, community-science records, camera-trap surveys

Introduction

Mapping the temporal change (kinetics) of species' geographic ranges in today's changing world (Cardinale et al., 2012; Urban, 2015) is a critical task for biogeography. To describe the dynamics of entire geographic ranges of species, we need both data over large and often heterogeneous regions, sometimes across entire continents or even the world, as well as data collected over a long-time period (Yoccoz et al., 2001). Despite increasing access to open data (Wüest et al., 2020), they are still sparse and spatially and temporally biased (Boakes et al., 2010; Maldonado et al., 2015; Meyer et al., 2016; Shirey et al., 2021). Moreover, the available data rarely come from a single large and standardised sampling effort (Ondei et al., 2018), but instead comprise a mix of local surveys that used different sampling methods (e.g., camera-traps, eDNA, or acoustic data loggers; Deiner et al., 2017; Gibb et al., 2019; Steenweg et al., 2017, as well as incidental occurrence records (e.g., derived from museum specimen collections and citizen-science records; Chandler et al., 2017; Osawa, 2019). To solve the global challenges without any further delay, we must seek to improve species' distribution models with the heterogeneous data that are available (Heberling et al., 2021).

Integrated species distribution modelling (hereafter IDMs) comprises a recently developed family of parametric species distribution models that combine ecological information within multiple data types that were typically collected by different survey approaches (Isaac et al., 2020; Kéry & Royle, 2021; Miller et al., 2019). IDMs capitalise on each data type's strengths, i.e., standardised surveys can provide information on the local abundance of species but often only at a relatively small number of sites, while opportunistic occurrence records can cover larger geographic/environmental spaces and can inform on range boundaries. IDMs are usually hierarchical models that explicitly model the sampling process of each dataset to account for limitations including imperfect detection and sampling bias (Fithian et al., 2015; Fletcher & Fortin, 2018; K. Pacifici et al., 2017), as well as varying effort and area of surveys (Keil & Chase, 2019). A characteristic of most IDMs is that they assume a common underlying spatial point process that determines the spatial locations of individuals of a species (Dorazio, 2014; Fletcher Jr. et al., 2019; Miller et al., 2019). Following this assumption, parameters affecting the intensity (or density) of the resulting point pattern (e.g., land cover or climate) are estimated using the joint likelihood for all included data types (Fletcher Jr. et al., 2019). Most applications so far fit IDMs in a Bayesian framework (van de Schoot et al., 2021), which also helps propagate the uncertainties associated with each data type into the predictions and parameter estimates.

Studies have already shown the many advantages of model-based data integration. First, the increased sample size from making use of diverse data streams tends to increase the precision of parameter estimates (Farr et al., 2021; Zipkin et al., 2017) and the accuracy of the predictions (Zulian et al., 2021). Second, combining structured or semi-structured data with unstructured data helps to factor out spatial biases in the latter, and consequently helps to make better use of data streams coming from opportunistic citizen science events (Dorazio, 2014; Zulian et al., 2021). Finally, the greater geographic coverage achieved by data integration may lead to better sampling of environmental gradients and hence improve the accuracy and precision of the estimated effects of covariates such as land cover and climate (Bowler et al., 2019; Chevalier et al., 2021).

Despite the great promise of IDMs, applications of them are still limited. Studies have used IDMs to address a wide range of data integration problems (e.g., (Martino et al., 2021; Rose et al., 2020; Schank et al., 2017; Zulian et al., 2021), but they have mostly been used over local (Farr et al., 2021) or nationwide (Hertzog et al., 2021) extents, and at fine grains, but not to model entire geographic ranges of species over coarse grains. As an exception, Zulian et al., (2021) used data integration to model the full geographic distribution of a parrot species endemic to the tropical South American Atlantic Forest. Further, with some exceptions (Doser et al., 2022; Hertzog et al., 2021; Pagel et al., 2014), IDMs have not been used to model temporal change of distributions, although this could be their obvious application, given the scarcity of temporally replicated data. Finally, IDMs can appear complex, with a lack of user-friendly tools available; thus, their implementation can be challenging, particularly for inexperienced users. Hence, the full potential of IDMs has yet to be realized and made accessible and there is a need for models that balance pragmatism and realism for combining the data typically available for large-scale distribution models.

Here, we introduce an IDM that addresses these shortcomings and models the temporal dynamics of entire species' geographic ranges by integrating two common data types: presence-only observations (e.g., as available in GBIF) and presence-absence surveys (e.g., from systematic surveys, in our case from camera traps). The model also accounts for common data problems such as local and regional variation in sampling effort and unequal area of surveys. The model can predict the temporal change of geographic distributions, change in range size, and the associated uncertainty, at any spatial resolution. Lastly, one of our important aims is to lower the learning threshold of IDMs for new users.

The American tropics (the Neotropics) have been identified among the most important hotspots of biodiversity in the world (Antonelli et al., 2018; Morrone, 2017). At the same time, they are one of the areas where biodiversity is declining at higher rates (IPBES, 2019). Unfortunately, data challenges are particularly pronounced in this region (Meyer et al., 2016), thus, the range dynamics of many species that occur there remain unknown. The available data are scattered and heterogeneous, typically coming from countries such as Colombia, Brazil, and Mexico and mainly from the birds' group. As a test case for our IDM, we chose the jaguarundi (*Herpailurus yagouaroundi*) (Figure 1), which has a large distribution across Latin America but knowledge of it has been limited by the data. Evidence shows that carnivore species such as the jaguarundi, have been recently varying their geographic distribution, most often noted around range edges (Grattarola et al., 2016; Lombardi et al., 2022; Luengos Vidal et al., 2017), and their abundance, over their entire distribution range (Caso et al., 2015). However, whether these changes are a product of previous lack of monitoring efforts in the region or due to the expansion or contraction of this species' range over time has not been quantitatively studied. Here, we develop an IDM to fill this knowledge gap. Moreover, we use our study design to provide a clear working example with R code, which can easily be copied, extended, and applied to model the range dynamics of other species.



Figure 1. Individuals of jaguarundí (*Herpailurus yagouaroundi*) displaying the main two-coat colour variants. Left observed in Mexico by albamaya (CC-BY-NC) and right in Argentina by hhulsberg (CC-BY-NC). Photos from iNaturalist.org.

Material and methods

The data

Occurrence records (i.e., **presence-only data**; Figure 2) were downloaded from GBIF (GBIF.org, 2021), filtering all records from Neotropical carnivores with geographic coordinates and with no spatial issues. Jaguarundí data were subset by removing records with coordinate precision smaller than three decimal places (i.e., 0.001), and coordinate uncertainty greater than 25,000 meters. Since we aimed to compare the distributional change in time, we divided the data into two time periods (pre: 2000-2013 and post: 2014-2021), which were chosen since most of the data were collected from 2000 onwards and on average each period represented 50% of the data (presence-absence and presence-only). For each time period, we mapped the data to 100 x 100 km resolution grid-cells (Lambert azimuthal equal-area projection; centre latitude 0° S and centre longitude 73.125° W) covering the entire Neotropical region (i.e., from Mexico to the south of Argentina) - 100 km was chosen as a compromise between sufficiently coarse for computational efficiency and sufficiently fine to produce useful descriptions of a species' range at a continental scale. In total, there were 261 occurrence records for the first time period and 242 records for the second period.

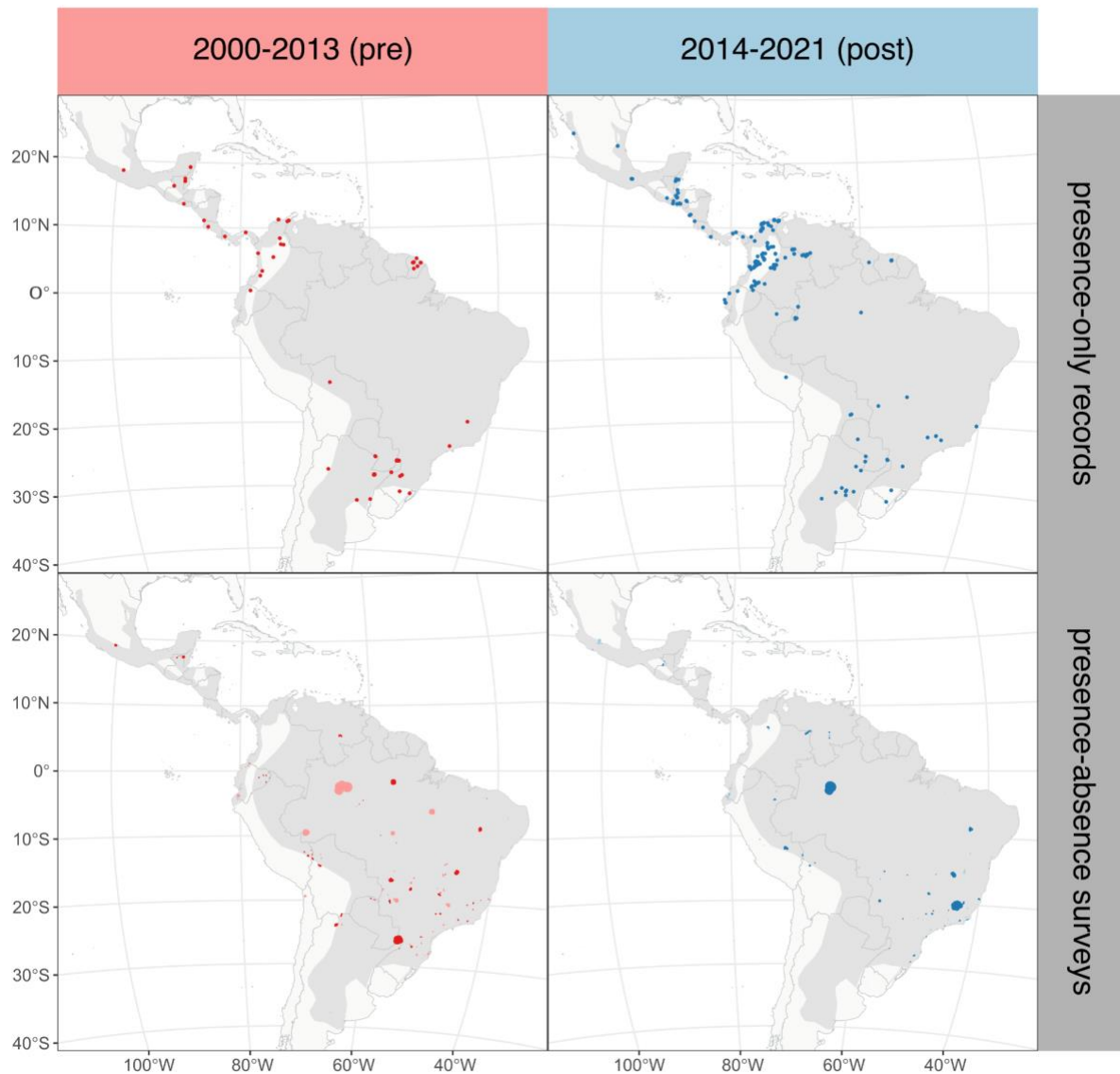


Figure 2: Distribution of jaguarundi data from the two time periods: from 2000 to 2013 (pre) and from 2014 to 2021 (post). The top row shows occurrence records from GBIF.org (2021), and the bottom, camera-trap surveys from Nagy-Reis et al. (2020) with presences in a darker colour and absences in a lighter tone (i.e., blobs in which some carnivore species were reported, but not the jaguarundi). The geographic range distribution of the jaguarundi according to IUCN is shown in shaded grey for all maps (IUCN, 2022).

Presence-absence data (Figure 2) were extracted from Nagy-Reis et al. (2020), a database of neotropical carnivores records. Data were cleaned by retaining surveys using camera traps (with detection and non-detection values), with geographic coordinates, with information about the study sampling area, with starting and ending month and year of the study, and with reported sampling effort (i.e., the number of active camera trap days). For each survey, a buffer polygon (hereafter ‘blob’) was created using the latitude and longitude as centroid and either the study area or the lat/long precision for the studies at the sampling level of “area” as the expected area of the blob (see Nagy-Reis et al. 2020 for details).

Absences were generated from locations where there was a camera-trap study and the yaguarundí was not recorded. We used data from 8,346 surveys for our study period: 4,303 for the first period and 4,043 for the second. The yaguarundí was recorded in 614 of the surveys. Overlapping surveys for each period were then combined in blobs (488 for the first period and 480 for the second) and for each one, we calculated the total surface area, the time span of the records, and the effort in camera trap days.

Thinning variables. The real-world occurrences of the yaguarundí can be thought of as a point pattern (Baddeley et al., 2015), which is then sampled such that only some points are observed (thinned) and end up in GBIF, and ultimately in the presence-only dataset. To adjust the presence-only data for sampling effort (i.e., thinning) in each 100 x 100 km grid cell, we used data on accessibility from urban areas based on travel time (Weiss et al., 2020). Based on many past studies (e.g., Geldmann et al., 2016), we expected that highly accessible grid cells will have more point records than inaccessible grid cells. In addition, for each grid cell, we also included the country of origin to account for differences among countries in data-sharing capacities and citizen-science levels of engagement.

Environmental covariates. The yaguarundí has been reported to occur mainly in lowland areas (up to 3,200 m) (Caso et al., 2015) and in a variety of habitats, from dense tropical rainforest to open grassland, although in open areas it prefers patches of thick cover (Macdonald & Loveridge, 2010). To model its distribution, we chose a set of environmental covariates and assessed their relative importance for the species. For each grid cells in the 100 x 100 km grid, and for each blob, we extracted the 19 bioclimatic variables and elevation (SRTM) from WorldClim V2.1 (Fick & Hijmans, 2017), land cover at 500m/yearly resolution (MCD12Q1) (Friedl & Sulla-Menashe, 2019), net primary production (NPP) 500m/yearly resolution (M*D17A3HGF) (Running & Zhao, 2019), percentage of tree cover and percentage of non-tree vegetation 250m/yearly resolution (MOD44B) (DiMiceli et al., 2015), from NASA MODIS Terra.

Covariate selection. Using all these covariates would lead to convergence problems due to collinearities, so we narrowed down the scope of the covariates for the final integrated model. Yet, doing a formal stepwise variable selection in the IDM setting is challenging because of computationally intensive MCMC sampling; similarly, Bayesian variable selection (O'Hara & Sillanpää, 2009) led to convergence problems. We thus manually selected a subset of covariates using (i) published descriptions of yaguarundí's habitat requirements (Caso et al., 2015; Macdonald & Loveridge, 2010), (ii) Pearson correlations among covariates (we aimed at minimising them), (iii) findings from simple tree-based machine learning (boosted trees, random forests) with the raw presence/absence as a response, and all the covariates as predictors. We ended up selecting bio7 (temperature annual range: maximum temperature of the warmest month - minimum temperature of the coldest month), bio15 (precipitation seasonality), elevation, and NPP (Net Primary Production) (see Figure S1 in Supporting Information).

Covariate extraction to grids and blobs. Continuous covariate data were matched to the occurrence data by averaging values within the 100 x 100 km grid cells and to the presence-absence data by averaging values within blobs. For the land cover (categorical covariate), we assigned the mode value (i.e. the most common value) for each grid cell and blob. We used the 'rnaturalearth' package (South 2022) to obtain Latin American countries' spatial polygons at a large scale. Spatial data analyses were done using 'sf' (Pebesma et al. 2022)

and 'terra' packages (Hijmans et al. 2022). MODIS data were downloaded using 'MODISsp' (Busetto et al. 2021).

The model

Understanding Bayesian IDMs can be challenging. We thus urge readers to pay close attention to our model's code (https://github.com/bienfloresencia/yaguarundi_IDM). It comes with a glossary of terms, extensive comments, and further explanations, all aimed at providing material that readers can reuse in their own projects.

Approach to integration. The model assumes that yaguarundi's distribution can be described by a continuous point pattern intensity (equations 1-2 below) across Latin America. This intensity is then integrated across blobs to calculate the likelihood of the presence-absence (PA) data (eq. 4) or across 100 x 100 km grid cells to calculate the likelihood of the counts of presence-only (PO) records within the cells (eq. 8). Both likelihoods are then used jointly (Miller et al., 2019) to estimate parameter values.

Modelling time. A key feature of our model is that it allows for different probability of occurrence at each location between the pre- and post- 2013 periods. To understand this, let's point out the parts of the model that are constant in time, and parts that change. We model the following model components to be *constant in time* (identical in pre- and post-2013): (i) the relationship between point process intensity and environmental covariates, (ii) the relationship between accessibility and point process thinning, (iii) the random effects of countries on point process thinning, and (iv) the relationship between sampling effort and presences/absence data. The only components that *change in time* are the autocorrelated spline surfaces, one for pre-, and the other for post-2013. These capture any time-dependent spatial structure in the yaguarundi's occurrence that is not captured by (temporally constant) environmental covariates.

Point pattern intensity. To create the point pattern intensity of yaguarundi occurrences, we used design matrices (\mathbf{X}_{PA} and \mathbf{X}_{PO}) that contained as many columns as the fitted model has parameters; in our case 21: an intercept, environmental covariates (elevation, NPP, bio7, bio15) and the spline bases. When the design matrices are multiplied by the vector of parametric effects (\mathbf{b}), they yield the linear predictors (η_{PA} and η_{PO}), i.e., the expected point pattern intensity, given the values of all explanatory variables in the model:

$$\eta_{PA} = \mathbf{X}_{PA} \times \mathbf{b} \quad (1)$$

$$\eta_{PO} = \mathbf{X}_{PO} \times \mathbf{b} \quad (2)$$

For the environmental variables, we used weakly informative priors $b_i \sim \text{Normal}(0,10)$. Note here that \mathbf{b} is the same in both eqs. 1 and 2; this is the central part of the model which connects the PO and PA data and allows calculation of the joint likelihood.

Smoothing splines. We used thin plate regression splines (Wood, 2003) to model the spatial structure in the distribution that was not accounted for by the environmental covariates. We first generated $k=9$ *spline basis* variables prior to the model fitting, using the `jagam` function from the 'mgcv' package (Wood, 2017). These variables are then part of the \mathbf{X}_{PA} and \mathbf{X}_{PO} matrices, and they have their own corresponding coefficients in the \mathbf{b} vector. These coefficients have their own multivariate normal prior, specified using smoothing

penalty matrices and smoothing parameters; for the sake of simplicity we don't present the complex mathematical definition here; we refer readers to the help of the *jagam* R function. We selected $k=9$ as it was the highest value that still gave good convergence of the MCMC, and also provided sufficiently flexible surfaces to model large-scale geographic range.

Modelling presence-absence data. Letting y_{PAi} refer to the observed presence (1) or absence (0) value in each i -th blob for pre- or post- period, we modelled the blob-specific probability of presence (ψ_i) as a function of the fixed effects of the presence-absence linear predictor (η_{PAi}) and sampling effort ($effort_i$, i.e., number of camera trap days), and the logarithm of the area of each blob in m^2 ($area_{PAi}$) as an offset term.

$$cloglog(\psi_i) = \eta_{PAi} + \log(area_{PAi}) + \beta \times \log(effort_i) \quad (3)$$

where the index i identifies blobs.

The state of this variable follows a Bernoulli distribution with mean ψ_i :

$$y_{PAi} \sim \text{Bernoulli}(\psi_i) \quad (4)$$

Prior distribution of β was $\beta \sim \text{Normal}(0, 10)$.

Modelling presence-only data. We assume that the spatial distribution of individuals may be modelled using a Poisson point process. In our model, the true intensity (i.e., mean number of points per grid-cell) for the species in each grid-cell j is denoted as v_j . We modelled it as a function of the exponential of the presence-only linear predictor (η_{POj}) by the area of each grid-cell in m^2 ($area_{POj}$).

$$v_j = area_{POj} \times \exp^{\eta_{POj}} \quad (5)$$

where j denotes a grid-cell.

To model the thinning of the true intensity, we calculated the cell-specific probability of retaining/observing a point (ρ_{retj}) as a decaying exponential function with a random intercept for each country ($\alpha_{0countryj}$) and a fixed slope for grid-cell accessibility ($\alpha_1 \times acce_j$):

$$P_{retj} = \alpha_{0countryj} \times \exp^{-\alpha_1 \times acce_j} \quad (6)$$

With prior distributions of α_0 and α_1 defined as $\alpha_{0c} \sim \text{Beta}(1, 1)$ where $c \in 1:n_{cntr}$, and $\alpha_1 \sim \text{Gamma}(0.5, 0.05)$.

Finally, we calculated the thinned intensity per grid-cell (λ_j) as the product of the true intensity (v_j) times the probability of retaining a point per grid-cell (ρ_{retj}):

$$\lambda_j = v_j \times \rho_{retj} \quad (7)$$

The state of this variable follows a Poisson distribution with mean λ_j where Y_{PO} is the observed data

$$Y_{POj} \sim \text{Poisson}(\lambda_{ij}) \quad (8)$$

Predictions. To predict the probability of occurrence of the species in the two time periods, with the linear predictor η_{pred} , as

$$\eta_{\text{pred}} = \mathbf{X}_{\text{PO}} \times \mathbf{b} \quad (9)$$

The detection probability ($\rho_{\text{pred}j}$) was modelled for each grid-cell j with area ($\text{area}_{\text{PO}j}$) as an offset term.

$$\text{cloglog}(P_{\text{pred}j}) = \eta_{\text{pred}j} + \log(\text{area}_{\text{PO}j}) \quad (10)$$

Derived Quantities. Finally, as derived outputs of the model, we calculated the area of the species range for the first period (A_{pre}) and the second period (A_{post}),

$$A_{\text{pre}} = \sum P_{\text{pred}j} \text{ where } j \in 1:(n_{\text{PO}}/2) \quad (11)$$

$$A_{\text{pos}} = \sum P_{\text{pred}j} \text{ where } j \in n_{\text{PO}}/2:n_{\text{PO}} \quad (12)$$

and the difference in the area (in number of 100x100km grid-cells) for both time periods (ΔA):

$$\Delta A = A_{\text{post}} - A_{\text{pre}} \quad (13)$$

The model was run in JAGS (Plummer, 2003) with the package 'R2jags' (Su & Yajima, 2020) and using 3 chains, 100,000 iterations per chain, 10,000 burning length and 10 as thinning rate. To check for convergence, we controlled "Rhat" statistics and traceplots (see more on model diagnostics in the GitHub repository) using the 'ggmcmc' package (Fernández-i-Marín, 2016). All analyses were performed in R 4.0.5 (R Core Team, 2021). The model and the model definitions can be accessed at: https://github.com/bienflorescencia/yaguarundi_IDM.

Results and Discussion

We successfully fitted an IDM to study the dynamics of the geographic range of the yaguarundí in Latin America over the last two decades. Good convergence values (Rhat < 1.1) were reached for all model parameters. Data integration enabled us to increase the sample size, the geographic extent, and environmental scope for each time period, taking advantage of the complementary information and sampling locations in different data streams (see Table S4). As the open data revolution continues, and citizen science contributes ever increasing amounts of data, we expect IDMs will become a standard tool for ecologists, provided that the tools become available. To our knowledge, we present the first application of the IDM approach with a temporal dimension and over the entire geographic range of a species.

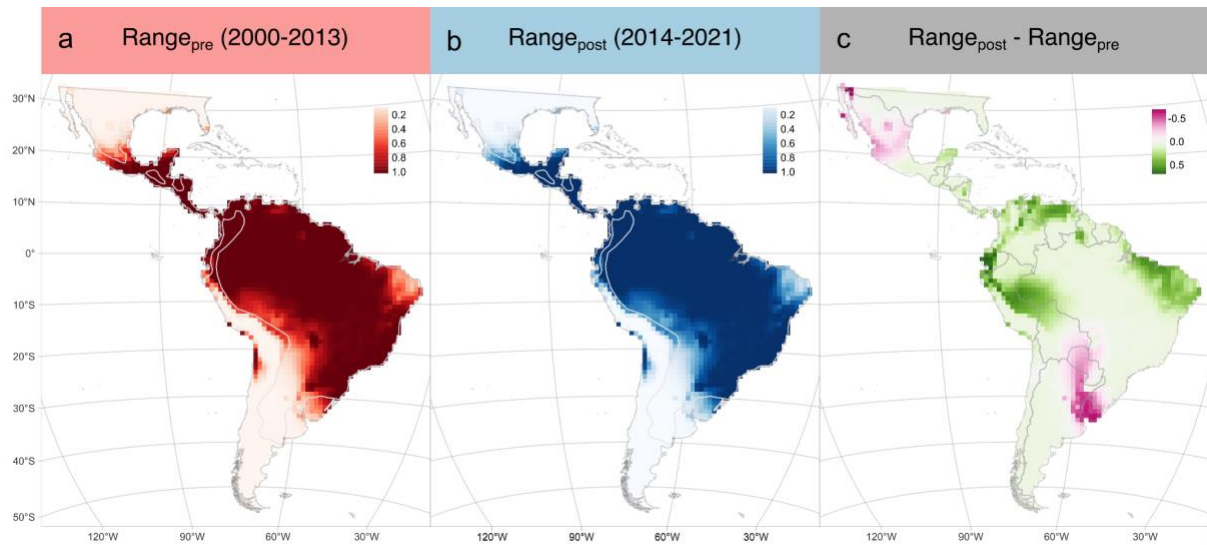


Figure 3: Maps of the yaguraundí's range distribution at the two time periods (a) $\text{Range}_{\text{pre}}$ from 2000 to 2013, (b) $\text{Range}_{\text{post}}$ from 2014 to 2021, and (c) their difference ($\text{Range}_{\text{post}} - \text{Range}_{\text{pre}}$). Shown overlapped in (a) and (b) as a grey outline is the IUCN range map (IUCN, 2022).

Most up-to-date knowledge of the species range: Expert range maps have played important roles in both research and policy by providing information on species distributions where there are data gaps. However, expert range maps are unsurprisingly coarse and infrequently updated, which means that they rapidly become out-of-date and have been less useful for studying range change. Using all the data available (open-access) for the species from 2000 to 2021, our modelled geographic range (pre and post) is not entirely consistent with the current expert range map from IUCN (Figure 3a, b; grey outline). Specifically, in the southern range limit in Argentina and Uruguay and in the Sertão region of north-eastern Brazil, where our model predicted a low probability of occurrence, or in the border of Mexico with Guatemala and the northern Andes where it predicted a high probability. Thus, we have updated the knowledge presented by the IUCN expert's range map.

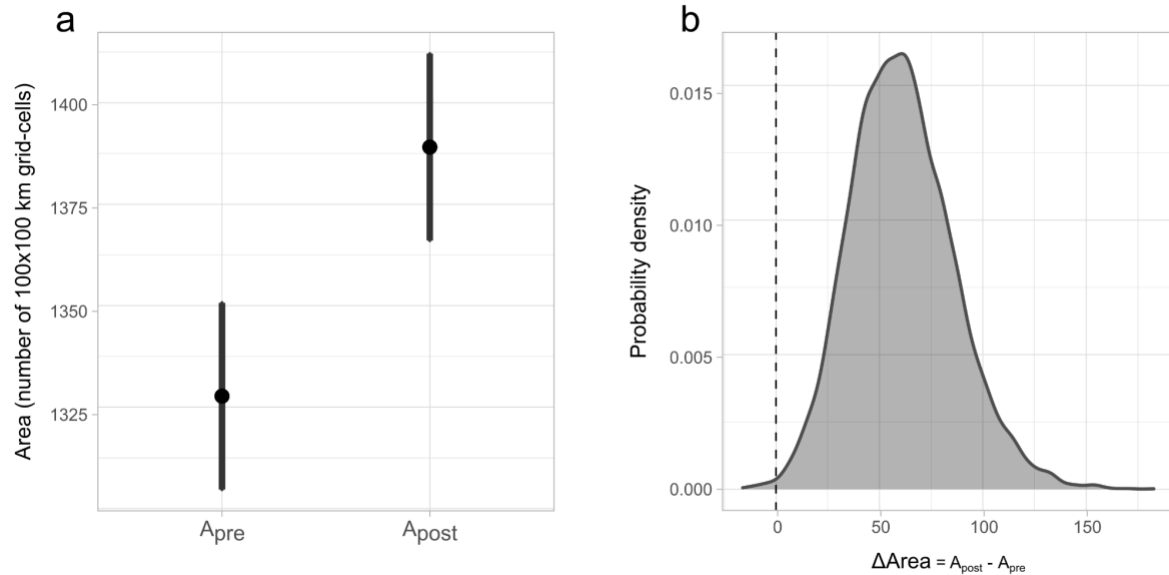


Figure 4 (a) Boxplot of posterior densities of the predicted area in both time periods: A_{pre} from 2000 to 2013 and A_{post} from 2014 to 2021. (b) Posterior distribution of range change ($\Delta Area$), dashed line = no range change.

Temporal change. The main innovation of our IDM is the estimation of temporal change in species' geographic range, which was possible even over a relatively short time span. We found that between the pre- and post-2013 periods, the yaguarundí has contracted its southern and northern range limits towards the equator but expanded its area of distribution over the entire species' range (Figure 3, $Range_{post} - Range_{pre}$; Figure 4). We attribute this to the species being rarer close to its environmental niche limits. Yet, at least in the southern limit, major land conversions have taken place in the recent decades as a result of agriculture expansion, (mainly soybean Baldi & Paruelo, 2008; Song et al., 2021), where the species also occurs at relatively low densities (Giordano, 2016; Luengos Vidal et al., 2017). The yaguarundí can also be shifting its distribution as a response to changes in environmental variables such as increasing temperatures and precipitation anomalies (Magrin et al., 2014), or due to the influence of the distribution of other species, e.g., competitive exclusion with *Leopardus pardalis* due to what is known as “the ocelot effect” (de Oliveira et al., 2010).

However, we note that even though our model can model temporal change of species ranges, it does not directly test causal hypotheses about drivers of the change. This is because we modelled temporal change solely using the different spline surfaces in pre- and post-2013 periods, while the environmental covariates in the model were long-term averages. In a way, our model thus predicts range kinetics (i.e. temporal change) rather than dynamics (i.e. temporal change and its causes). Here we see a clear opportunity for extensions of our IDM to directly assess causal drivers of the change. A simple approach can be to relate the predicted change of P_{pred} to an observed change of environment directly within the model (specifically, in the “predicted quantities” part). Environmental covariates could be decomposed into both the spatial long-term means as well as the temporal anomalies (Oedekoven et al., 2017). A more sophisticated approach could involve velocity

(i.e., magnitude and spatial direction) of both the range and environment (Loarie et al., 2009) or modelling co-occurrence effects (Ovaskainen et al., 2017).

Ecological inference. An advantage of our parametric IDM (over, e.g., a machine learning such as random forest) is that it can also be used for ecological inference. According to the coefficients for the environmental covariates (Figure 5), the yaguarundí prefers productive green areas with good vegetation cover, likely where it can hide (positive effect of NPP in Figure 5); it avoids deserts and semi-arid areas such as the Atacama or the Brazilian Northeast (negative effect of annual temperature range, or “bio7”); it prefers areas with seasonal precipitation, which is most of Latin America except for the deserts, high Andes, and extremely humid Colombian forests (positive effect of precipitation seasonality, or “bio15”). Finally, we found a positive effect of elevation (but it was also weak, relatively to the effect of other covariates, Figure 5). This indicates that the yaguarundí may not be restricted to lowlands, as described by Caso et al., (2015). This is in line with several observations from higher elevations in our data (presence-only records in Figure 2). Hence our model suggests that if an area offers enough vegetation cover and suitable climatic conditions, yaguarundí will likely be present, irrespectively of altitude. This insight would be impossible if a presence-absence approach to species distribution modelling were used, as these data miss the occurrence of yaguarundí in the northern Andes (Figure 2), unlike the presence-only data.

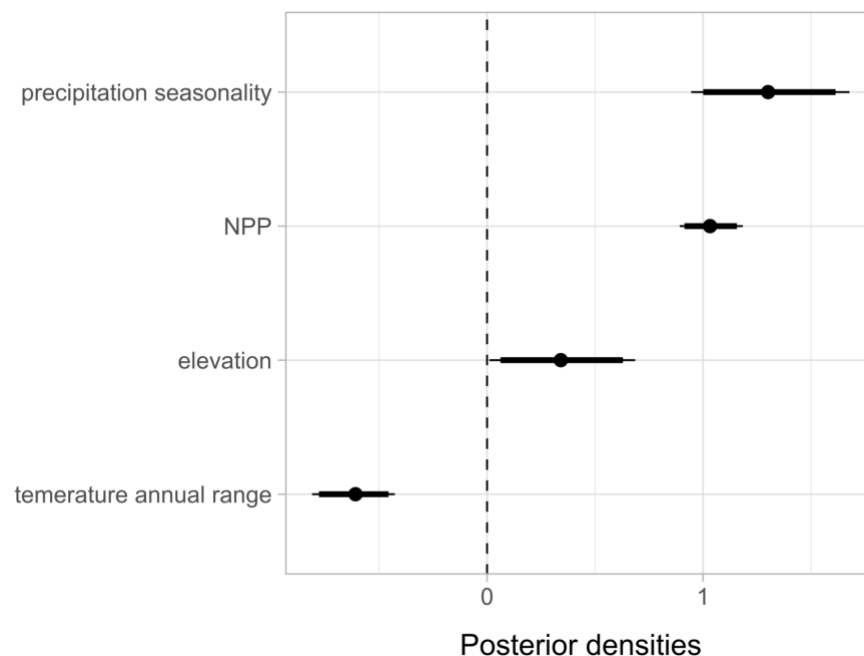


Figure 5. Effect of the environmental covariates on the intensity of the point process. Thick lines represent 90% of the highest posterior densities of the parameters and thin lines represent 95%.

Inference about sampling effort - the thinning function. A notable feature of our model is how we considered the probability of an observation being made if the species was present

(i.e., the thinning process; equation 6). We made it dependent on the grid-cell accessibility since this is a known variable associated with the density of presence-only records and reflects the effect of the number of possible observers and their tendency to observe species near where they live. Moreover, we allowed its effect to vary among countries, allowing for differences in capacities to observe and report species observations. Our results revealed that, for the predicted species range distribution (Figure 3), countries such as Argentina, French Guiana, and Suriname have reasonable levels of records for the species in accessible areas (see Figure S2). However, most countries show low observation-retention probability ($P_{ret} < 0.25$ when accessibility is maximum; see Figure S2), which means that even areas that are easy to reach only get less than 25% of observations. We expected that countries such as Colombia had better levels of sampling effort for the species, as this is one of the countries in Latin America with the highest numbers of records in GBIF (GBIF Secretariat, 2022). However, the low levels of sampling found there can be related to the high abundance (point intensity) of the species in the area. Regardless, some particular regions of the species distribution range, as seen by the uncertainty of the estimates, need more sampling effort (see Figure S3).

Limitations and potential extensions. Like with any other model, our IDM has clear limitations and scope for improvements. First, the model predicts a high probability of yagouarundi's occurrence in Chile around 20°S parallel, even though the species has never been observed in the area; we attribute this to the insufficient flexibility of the spline surface, which should in theory be able to account for the problem; specifically, increasing the number of bases functions (k) could solve this, although it can lead to convergence problems. Second, since it is implemented in JAGS, the model can be extended in various ways. Accounting for imperfect detection (Dorazio, 2014; Koshkina et al., 2017), modelling multiple species in a joint species distribution models (Doser et al., 2022; Ovaskainen & Abrego, 2020), e.g., to share information on the thinning process (Fithian et al., 2015), or accounting for false positives (Kéry & Royle, 2021) are some of the potential extensions that have recently been tested in the IDM framework (Doser et al., 2022), although not over large geographic extents.

Practical applications and challenges. Jumping from classic statistics to full Bayesian statistical inference comes with some hurdles. There are conceptual barriers/obstacles, new terms/definitions and much statistical rethinking McElreath (2020). There are also no pre-made R functions to choose from and users must design and code on their own. This requires a greater knowledge of the inner workings of the models, and their ecological interpretation. Bayesian methods can be computationally intensive; in our case, it took 230 minutes in a 16GB RAM 3.2Ghz 8-core laptop to run MCMC with 100,000 iterations. We used JAGS (an implementation of BUGS; Lunn et al., 2000) to specify and fit the model, since it is more flexible than e.g., INLA (Rue et al., 2009), and also more didactical than e.g., STAN (Stan Development Team, 2022). Spatial splines can be a particular challenge to code for JAGS, but the `jagam` function in the "mgcv" R package is enormously helpful for this. For more advanced users, however, the latter two may be more computationally efficient alternatives. Fortunately, code and data sharing are becoming more and more common. From our experience, designing and implementing an IDM from scratch can take months to a year. Thus, new users would greatly benefit from collaborating with people that have used them before. Hopefully, our analysis, together with the commented code and the data (https://github.com/bienfloresencia/yaguarundi_IDM) will help to overcome these hurdles.

In this respect, we also recommend the excellent tutorials by Kéry (2010) and Kéry & Royle (2021) and McElreath (2020) together with the course (https://github.com/rmcelreath/stat_rethinking_2022).

Outlook

In the tropics and the global South, lack of temporal data has always limited understanding of how species change their geographic ranges through time (Antonelli et al., 2018; Hortal et al., 2015). Most recent studies have focus on the temperate global North (i.e., North America, Europe, and Australia-New Zealand) (Chen et al., 2011). Even studies on the global scale, such as those based on the BioTime (Dornelas et al., 2018) or the Living Planet databases (Loh et al., 2005), have severe data gaps in the tropics. If there are any studies that cover the entire world (M. Pacifici et al., 2020), they rely on static geographic ranges (e.g., IUCN range maps), which in many regions tend to misrepresent actual species distributions (Hughes et al., 2021). Fortunately, unsystematic records from citizen, or community-based, science platforms are a growing source of presence-only data. Global initiatives such as iNaturalist (www.inaturalist.org) have become popular in Latin America, with national nodes in Argentina, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Panama, Mexico and Uruguay, now counting more than 2.6 million research-grade observations (GBIF.org, 2022). These data have been deemed particularly problematic in the estimation of temporal trends of biodiversity and species geographic distributions (Peterson et al., 2011). However, with the type of IDMs that we present here, they can become a potentially valuable source for species distribution modelling in the region.

Data availability statement

All data and extensively commented code used in this paper are openly available at https://github.com/bienflorencia/yaguarundi_IDM under CC-BY licence.

References

- Antonelli, A., Ariza, M., Albert, J., Andermann, T., Azevedo, J., Bacon, C., Faurby, S., Guedes, T., Hoorn, C., Lohmann, L. G., Matos-Maraví, P., Ritter, C. D., Sanmartín, I., Silvestro, D., Tejedor, M., Ter Steege, H., Tuomisto, H., Werneck, F. P., Zizka, A., & Edwards, S. V. (2018). Conceptual and empirical advances in Neotropical biodiversity research. *PeerJ*, 6, e5644. <https://doi.org/10.7717/peerj.5644>
- Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. CRC press.
- Baldi, G., & Paruelo, J. M. (2008). Land-Use and Land Cover Dynamics in South American Temperate Grasslands. *Ecology and Society*, 13(2), 1–20. JSTOR. <https://doi.org/10.5751/ES-02481-130206>
- Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLOS Biology*, 8(6), e1000385. <https://doi.org/10.1371/journal.pbio.1000385>

- Bowler, D. E., Nilsen, E. B., Bischof, R., O'Hara, R. B., Yu, T. T., Oo, T., Aung, M., & Linnell, J. D. C. (2019). Integrating data from different survey types for population monitoring of an endangered species: The case of the Eld's deer. *Scientific Reports*, *9*(1), 7766. <https://doi.org/10.1038/s41598-019-44075-9>
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., & Wardle, D. A. (2012). Biodiversity loss and its impact on humanity. *Nature*, *486*(7401), 59–67. <https://doi.org/10.1038/nature11148>
- Caso, A., de Oliveira, T., & Carvajal, S. (2015). *Herpailurus yagouaroundi*. *The IUCN Red List of Threatened Species 2015: E.T9948A50653167* (p. 13). <https://www.iucnredlist.org/species/pdf/50653167>
- Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, *213*, 280–294. <https://doi.org/10.1016/j.biocon.2016.09.004>
- Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, *333*(6045), 1024–1026. <https://doi.org/10.1126/science.1206432>
- Chevalier, M., Broennimann, O., Cornuault, J., & Guisan, A. (2021). Data integration methods to account for spatial niche truncation effects in regional projections of species distribution. *Ecological Applications*, *31*(7), e02427. <https://doi.org/10.1002/eap.2427>
- de Oliveira, T. G., Tortato, M. A., Silveira, L., Kasper, C. B., Mazim, F. D., Lucherini, M., Jácomo, A. T., Soares, J. B. G., Marques, R. V., & Sunquist, M. (2010). Ocelot ecology and its effect on the small-felid guild in the lowland neotropics. In *The biology and conservation of wild felids* (pp. 559–580). Oxford University Press.
- Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. <https://doi.org/10.1111/mec.14350>
- DiMiceli, C., Carroll, M., Sohlberg, R., Kim, D., Kelly, M., & Townshend, J. (2015). *MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006*. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MOD44B.006>
- Dorazio, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, *23*(12), 1472–1484. <https://doi.org/10.1111/geb.12216>
- Dornelas, M., Antão Laura, H., Moyes, F., Bates Amanda, E., Magurran Anne, E., Adam, D., Akhmetzhanova Asem, A., Appeltans, W., Arcos José, M., Arnold, H., Ayyappan, N., Badihi, G., Baird Andrew, H., Barbosa, M., Barreto Tiago, E., Bässler, C., Bellgrove, A., Belmaker, J., Benedetti-Cecchi, L., ... Zettler Michael, L. (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>
- Doser, J. W., Leuenberger, W., Sillett, T. S., Hallworth, M. T., & Zipkin, E. F. (2022). Integrated community occupancy models: A framework to assess occurrence and biodiversity dynamics using multiple data sources. *Methods in Ecology and Evolution*, *13*(4), 919–932. <https://doi.org/10.1111/2041-210X.13811>
- Farr, M. T., Green, D. S., Holekamp, K. E., & Zipkin, E. F. (2021). Integrating distance sampling and presence-only data to estimate species abundance. *Ecology*, *102*(1),

- e03204. <https://doi.org/10.1002/ecy.3204>
- Fernández-i-Marín, X. (2016). ggmcmc: Analysis of MCMC Samples and Bayesian Inference. *Journal of Statistical Software*, 70(9), 1–20. <https://doi.org/10.18637/jss.v070.i09>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, 6(4), 424–438. <https://doi.org/10.1111/2041-210X.12242>
- Fletcher Jr., R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, 100(6), e02710. <https://doi.org/10.1002/ecy.2710>
- Fletcher, R., & Fortin, M. (2018). *Spatial ecology and conservation modeling*. Springer.
- Friedl, M., & Sulla-Menasse, D. (2019). *MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006*. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MCD12Q1.006>
- GBIF Secretariat. (2022). *GBIF overview slides, April 2022*. The Global Biodiversity Information Facility. <https://www.gbif.org/document/81771>
- GBIF.org. (2021). *Occurrence Download—Carnivores from the Neotropical region*. The Global Biodiversity Information Facility. <https://doi.org/10.15468/DL.3CU474>
- GBIF.org. (2022). *Occurrence Download—INaturalist records from the Latin American network members*. The Global Biodiversity Information Facility. <https://doi.org/10.15468/DL.A66R5F>
- Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C., & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, 22(11), 1139–1149. <https://doi.org/10.1111/ddi.12477>
- Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10(2), 169–185. <https://doi.org/10.1111/2041-210X.13101>
- Giordano, A. J. (2016). Ecology and status of the jaguarundi *Puma yagouaroundi*: A synthesis of existing knowledge. *Mammal Review*, 46(1), 30–43. <https://doi.org/10.1111/mam.12051>
- Grattarola, F., Hernández, D., Duarte, A., Gaucher, L., Perazza, G., González, S., Bergós, L., Chouhy, M., Garay, A., & Carabio, M. (2016). Primer registro de yaguarundi (*Puma yagouaroundi*) (Mammalia: Carnivora: Felidae) en Uruguay, con comentarios sobre monitoreo participativo. *Boletín de La Sociedad Zoológica Del Uruguay*, 25, 85–91.
- Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, 118(6), e2018093118. <https://doi.org/10.1073/pnas.2018093118>
- Hertzog, L. R., Frank, C., Klimek, S., Röder, N., Böhner, H. G. S., & Kamp, J. (2021). Model-based integration of citizen science data from disparate sources increases the precision of bird population trends. *Diversity and Distributions*, 27(6), 1106–1119. <https://doi.org/10.1111/ddi.13259>
- Hortal, J., Bello, F. de, Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J.

- (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 523–549.
<https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Hughes, A. C., Orr, M. C., Yang, Q., & Qiao, H. (2021). Effectively and accurately mapping global biodiversity patterns for different regions and taxa. *Global Ecology and Biogeography*, 30(7), 1375–1388. <https://doi.org/10.1111/geb.13304>
- IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (p. 56). IPBES Secretariat.
<https://doi.org/10.5281/zenodo.3553579>
- Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O. L., Schmucki, R., Simmonds, E. G., & O'Hara, R. B. (2020). Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution*, 35(1), 56–67. <https://doi.org/10.1016/j.tree.2019.08.006>
- IUCN. (2022). *Herpailurus yagouaroundi (spatial data)*. International Union for Conservation of Nature. <https://www.iucnredlist.org/species/9948/50653167>
- Keil, P., & Chase, J. M. (2019). Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution*, 3(3), 390–399.
<https://doi.org/10.1038/s41559-019-0799-0>
- Kéry, M. (2010). *Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses*. Academic Press.
- Kéry, M., & Royle, J. A. (2021). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press.
- Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M., & Stone, L. (2017). Integrated species distribution models: Combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution*, 8(4), 420–430. <https://doi.org/10.1111/2041-210X.12738>
- Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., & Ackerly, D. D. (2009). The velocity of climate change. *Nature*, 462(7276), 1052–1055.
<https://doi.org/10.1038/nature08649>
- Loh, J., Green, R. E., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V., & Randers, J. (2005). The Living Planet Index: Using species population time series to track trends in biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1454), 289–295. <https://doi.org/10.1098/rstb.2004.1584>
- Lombardi, J. V., Haines, A. M., Watts III, G. W., Grassman Jr., L. I., Janečka, J. E., Caso, A., Carvajal, S., Wardle, Z. M., Yamashita, T. J., Stasey, W. C., Branney, A. B., Scognamillo, D. G., Campbell, T. A., Young Jr, J. H., & Tewes, M. E. (2022). Status and distribution of jaguarundi in Texas and Northeastern México: Making the case for extirpation and initiation of recovery in the United States. *Ecology and Evolution*, 12(3), e8642. <https://doi.org/10.1002/ece3.8642>
- Luengos Vidal, E., Guerisoli, M., Caruso, N., & Lucherini, M. (2017). Updating the distribution and population status of Jaguarundi, *Puma yagouaroundi* (É. Geoffroy, 1803) (Mammalia: Carnivora: Felidae), in the southernmost part of its distribution range. *Check List*, 13(4), 75–79. <https://doi.org/10.15560/13.4.75>
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and*

- Computing*, 10(4), 325–337.
- Macdonald, D., & Loveridge, A. (2010). *The biology and conservation of wild felids* (Vol. 2). Oxford University Press.
- Magrin, G., Marengo, J., Boulanger, J., Buckeridge, M., Castellanos, E., Poveda, G., Scarano, F., Vicuña, S., Alfaro, E., & Anhelme, F. (2014). Central and South America in Climate Change 2014: Impacts, Adaptation, and Vulnerability. *Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change* (Eds. Barros, VR et Al.), 1499–1566.
- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Chilquillo, E., Rønsted, N., & Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? *Global Ecology and Biogeography*, 24(8), 973–984. <https://doi.org/10.1111/geb.12326>
- Martino, S., Pace, D. S., Moro, S., Casoli, E., Ventura, D., Frachea, A., Silvestri, M., Arcangeli, A., Giacomini, G., Ardizzone, G., & Jona Lasinio, G. (2021). Integration of presence-only data from several sources: A case study on dolphins' spatial distribution. *Ecography*, 44(10), 1533–1543. <https://doi.org/10.1111/ecog.05843>
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.
- Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, 19(8), 992–1006.
- Miller, D. A. W., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution*, 10(1), 22–37. <https://doi.org/10.1111/2041-210X.13110>
- Morrone, J. J. (2017). *Neotropical Biogeography. Regionalization and Evolution*. CRC Press. <https://doi.org/10.1201/b21824>
- Nagy-Reis, M., Oshima, J. E. de F., Kanda, C. Z., Palmeira, F. B. L., de Melo, F. R., Morato, R. G., Bonjorne, L., Magioli, M., Leuchtenberger, C., Rohe, F., Lemos, F. G., Martello, F., Alves-Eigenheer, M., da Silva, R. A., Silveira dos Santos, J., Priante, C. F., Bernardo, R., Rogeri, P., Assis, J. C., ... Ribeiro, M. C. (2020). NEOTROPICAL CARNIVORES: a data set on carnivore distribution in the Neotropics. *Ecology*, 101(11), e03128. <https://doi.org/10.1002/ecy.3128>
- Oedekoven, C. S., Elston, D. A., Harrison, P. J., Brewer, M. J., Buckland, S. T., Johnston, A., Foster, S., & Pearce-Higgins, J. W. (2017). Attributing changes in the distribution of species abundance to weather variables using the example of British breeding birds. *Methods in Ecology and Evolution*, 8(12), 1690–1702. <https://doi.org/10.1111/2041-210X.12811>
- O'Hara, R. B., & Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis*, 4(1), 85–117. <https://doi.org/10.1214/09-BA403>
- Ondei, S., Brook, B. W., & Buettel, J. C. (2018). Nature's untold stories: An overview on the availability and type of on-line data on long-term biodiversity monitoring. *Biodiversity and Conservation*, 27(11), 2971–2987. <https://doi.org/10.1007/s10531-018-1582-2>
- Osawa, T. (2019). Perspectives on biodiversity informatics for ecology. *Ecological Research*, 34(4), 446–456. <https://doi.org/10.1111/1440-1703.12023>
- Ovaskainen, O., & Abrego, N. (2020). *Joint Species Distribution Modelling: With Applications in R*. Cambridge University Press.

- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20(5), 561–576. <https://doi.org/10.1111/ele.12757>
- Pacifici, K., Reich, B. J., Miller, D. A. W., Gardner, B., Stauffer, G., Singh, S., McKerrow, A., & Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion*. *Ecology*, 98(3), 840–850. <https://doi.org/10.1002/ecy.1710>
- Pacifici, M., Rondinini, C., Rhodes, J. R., Burbidge, A. A., Cristiano, A., Watson, J. E. M., Woinarski, J. C. Z., & Di Marco, M. (2020). Global correlates of range contractions and expansions in terrestrial mammals. *Nature Communications*, 11(1), 2840. <https://doi.org/10.1038/s41467-020-16684-w>
- Pagel, J., Anderson, B. J., O'Hara, R. B., Cramer, W., Fox, R., Jeltsch, F., Roy, D. B., Thomas, C. D., & Schurr, F. M. (2014). Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods in Ecology and Evolution*, 5(8), 751–760. <https://doi.org/10.1111/2041-210X.12221>
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological Niches and Geographic Distributions (MPB-49)*. Princeton University Press. <https://doi.org/10.1515/9781400840670>
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 124(125.10), 1–10.
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rose, J. P., Halstead, B. J., & Fisher, R. N. (2020). Integrating multiple data sources and multi-scale land-cover data to model the distribution of a declining amphibian. *Biological Conservation*, 241, 108374. <https://doi.org/10.1016/j.biocon.2019.108374>
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series b (Statistical Methodology)*, 71(2), 319–392.
- Running, S., & Zhao, M. (2019). *MOD17A3HGF MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500 m SIN Grid V006*. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MOD17A3HGF.006>
- Schank, C. J., Cove, M. V., Kelly, M. J., Mendoza, E., O'Farrill, G., Reyna-Hurtado, R., Meyer, N., Jordan, C. A., González-Maya, J. F., Lizcano, D. J., Moreno, R., Dobbins, M. T., Montalvo, V., Sáenz-Bolaños, C., Jimenez, E. C., Estrada, N., Cruz Díaz, J. C., Saenz, J., Spínola, M., ... Miller, J. A. (2017). Using a novel model approach to assess the distribution and conservation status of the endangered Baird's tapir. *Diversity and Distributions*, 23(12), 1459–1471. <https://doi.org/10.1111/ddi.12631>
- Shirey, V., Belitz, M. W., Barve, V., & Guralnick, R. (2021). A complete inventory of North American butterfly occurrence data: Narrowing data gaps, but increasing bias. *Ecography*, 44(4), 537–547. <https://doi.org/10.1111/ecog.05396>
- Song, X.-P., Hansen, M. C., Potapov, P., Adusei, B., Pickering, J., Adami, M., Lima, A., Zalles, V., Stehman, S. V., Di Bella, C. M., Conde, M. C., Copati, E. J., Fernandes, L. B., Hernandez-Serna, A., Jantz, S. M., Pickens, A. H., Turubanova, S., & Tyukavina, A. (2021). Massive soybean expansion in South America since 2000 and implications

- for conservation. *Nature Sustainability*, 4(9), 784–792.
<https://doi.org/10.1038/s41893-021-00729-z>
- Stan Development Team. (2022). *Stan Modeling Language Users Guide and Reference Manual*. <https://mc-stan.org>
- Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., Townsend, S. E., Carbone, C., Rowcliffe, J. M., Whittington, J., Brodie, J., Royle, J. A., Switalski, A., Clevenger, A. P., Heim, N., & Rich, L. N. (2017). Scaling-up camera traps: Monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, 15(1), 26–34. <https://doi.org/10.1002/fee.1448>
- Su, Y.-S., & Yajima, M. (2020). *R2jags: Using R to Run 'JAGS'*. <https://CRAN.R-project.org/package=R2jags>
- Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, 348(6234), 571–573. <https://doi.org/10.1126/science.aaa4984>
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1), 1. <https://doi.org/10.1038/s43586-020-00001-2>
- Weiss, D. J., Nelson, A., Vargas-Ruiz, C. A., Gligorić, K., Bavadekar, S., Gabrilovich, E., Bertozzi-Villa, A., Rozier, J., Gibson, H. S., Shekel, T., Kamath, C., Lieber, A., Schulman, K., Shao, Y., Qarkaxhija, V., Nandi, A. K., Keddie, S. H., Rumisha, S., Amratia, P., ... Gething, P. W. (2020). Global maps of travel time to healthcare facilities. *Nature Medicine*, 26(12), 1835–1838. <https://doi.org/10.1038/s41591-020-1059-1>
- Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, 65(1), 95–114. <https://doi.org/10.1111/1467-9868.00374>
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>
- Wüest, R. O., Zimmermann, N. E., Zurell, D., Alexander, J. M., Fritz, S. A., Hof, C., Kreft, H., Normand, S., Cabral, J. S., Szekely, E., Thuiller, W., Wikelski, M., & Karger, D. N. (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>
- Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, 16(8), 446–453.
- Zipkin, E. F., Rossman, S., Yackulic, C. B., Wiens, J. D., Thorson, J. T., Davis, R. J., & Grant, E. H. C. (2017). Integrating count and detection–nondetection data to model population dynamics. *Ecology*, 98(6), 1640–1650. <https://doi.org/10.1002/ecy.1831>
- Zulian, V., Miller, D. A. W., & Ferraz, G. (2021). Integrating citizen-science and planned-survey data improves species distribution estimates. *Diversity and Distributions*, 27(12), 2498–2509. <https://doi.org/10.1111/ddi.13416>

Supporting information

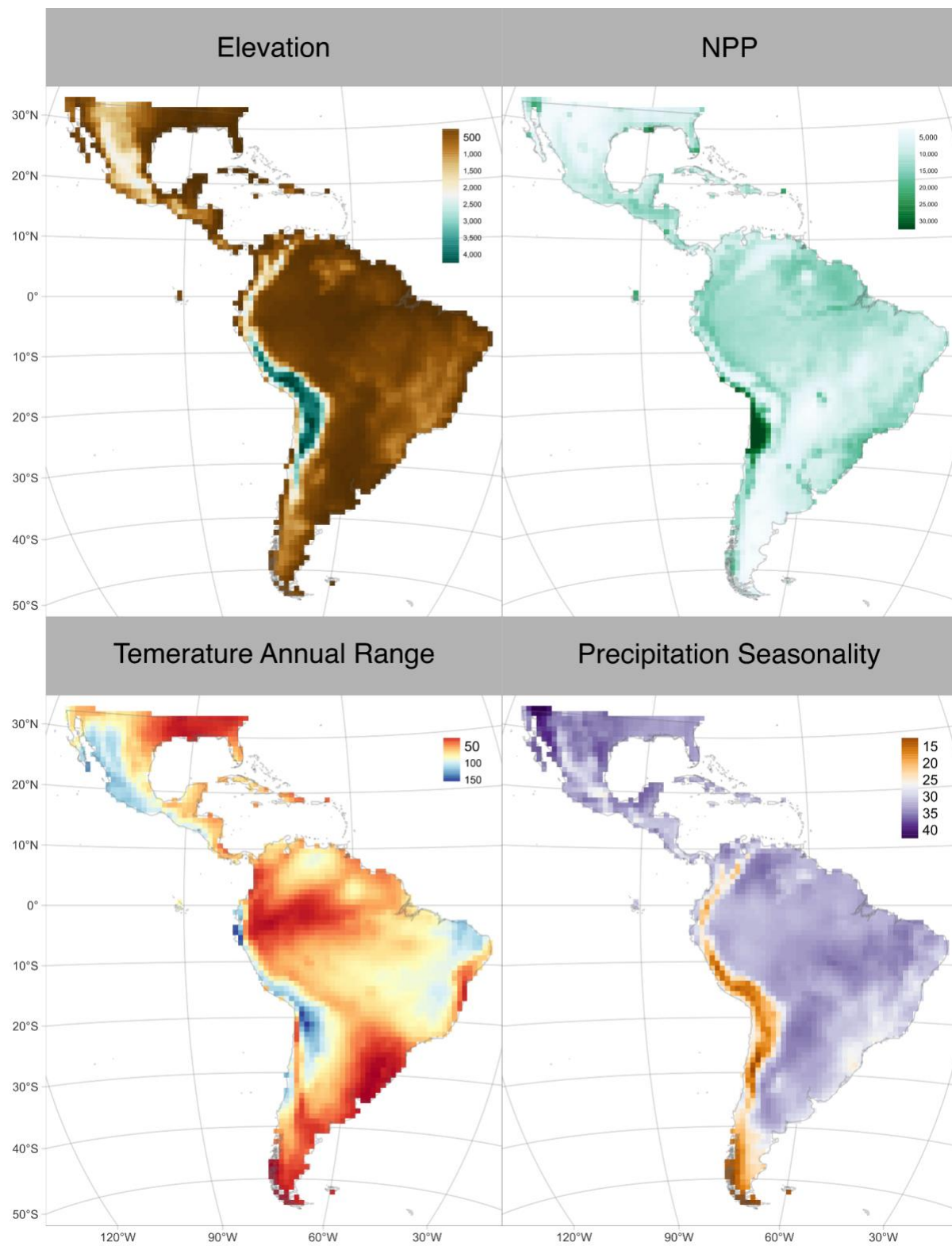


Figure S1. Distribution of the environmental covariates: elevation, NPP, temperature annual range (bio7), and precipitation seasonality (bio15).

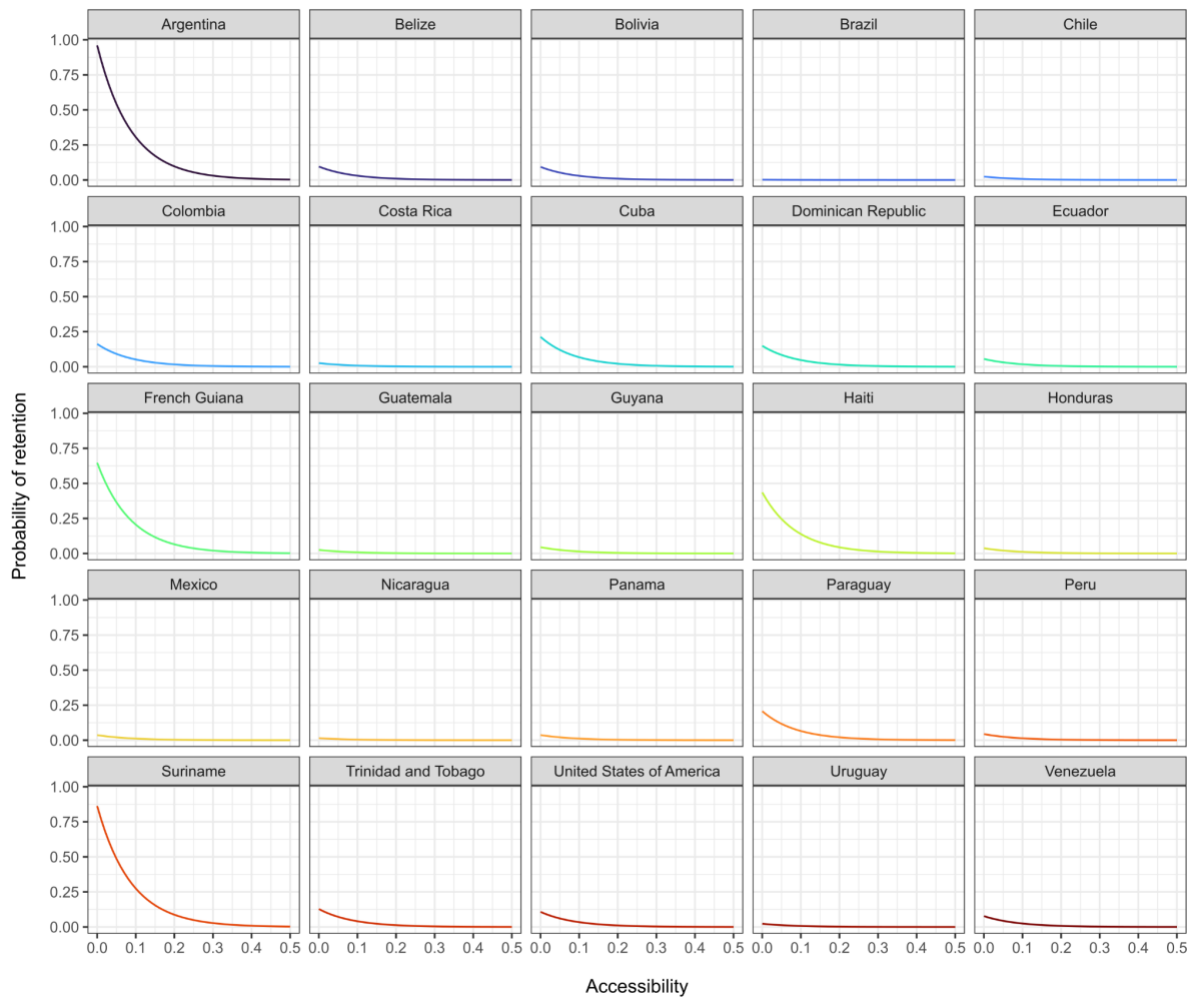


Figure S2. Probability of retention for each country (i.e., the decay of the probability of occurrence as a function of accessibility). Accessibility to a grid-cell is maximum at 0.

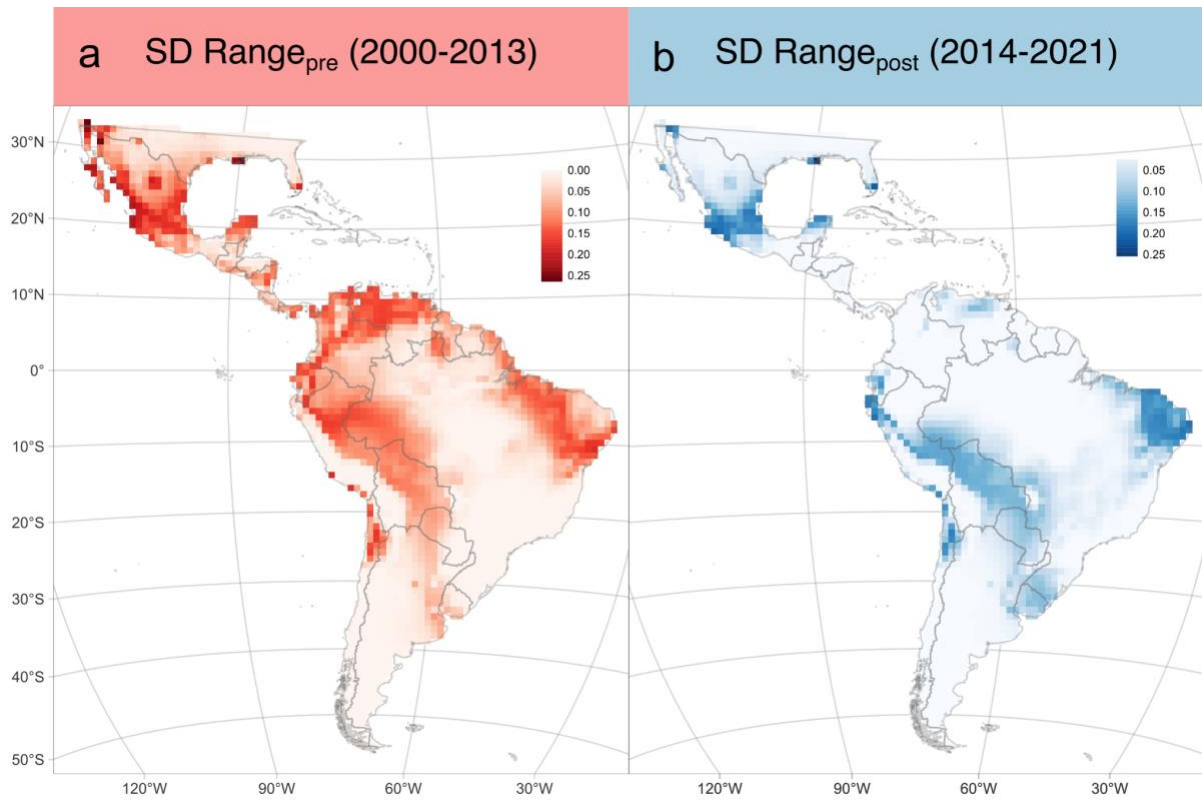


Figure S3. Standard deviation (SD) of the probability of occurrence for both time periods: Range_{pre} from 2000 to 2013 and Range_{post} from 2014 to 2021.

Table S4. The range and interquartile range (IQR) of values for each covariate that is sampled by each dataset.

	presence-absence		presence-only	
	range	IQR	range	IQR
Elevation	4867.706	483.500	4462.538	483.500
bio 1	27.973	625.332	27.797	625.332
bio 2	29.357	5.677	32.532	5.677
bio 3	27.500	4.237	27.235	4.237
bio 4	3949.148	7.009	6400.930	7.009
bio 5	525.345	429.251	720.784	429.251
bio 6	242.290	92.461	340.600	92.461
bio 7	170.527	39.762	154.105	39.762
bio 8	1422.255	28.919	1997.032	28.919
bio 9	745.389	251.934	1207.549	251.934
bio 10	1039.348	140.433	1463.867	140.433
bio 11	1229.406	224.891	1820.896	224.891
bio 12	17.895	215.120	19.223	215.120
bio 13	89.977	2.223	91.201	2.223
bio 14	618.793	7.851	860.984	7.851
bio 15	37.801	131.563	41.966	131.563
bio 16	22.600	4.092	22.770	4.092
bio 17	32.448	6.816	39.095	6.816
bio 18	28.755	2.644	31.390	2.644
bio 19	29.011	3.945	28.786	3.945
NPP	32766.000	6.688	32735.357	6.688
Tree vegetation	200.000	6306.805	112.025	6306.805
Non-tree vegetation	200.000	43.063	124.779	43.063