@flograttarola

ecoevo.social/@flograttarola

Data integration to model species geographic range dynamics

The yaguarundí (*Herpailurus yagouaroundi*) in Latin America as a case study

Florencia Grattarola, Daiana E. Bowler and Petr Keil

MOBI Lab

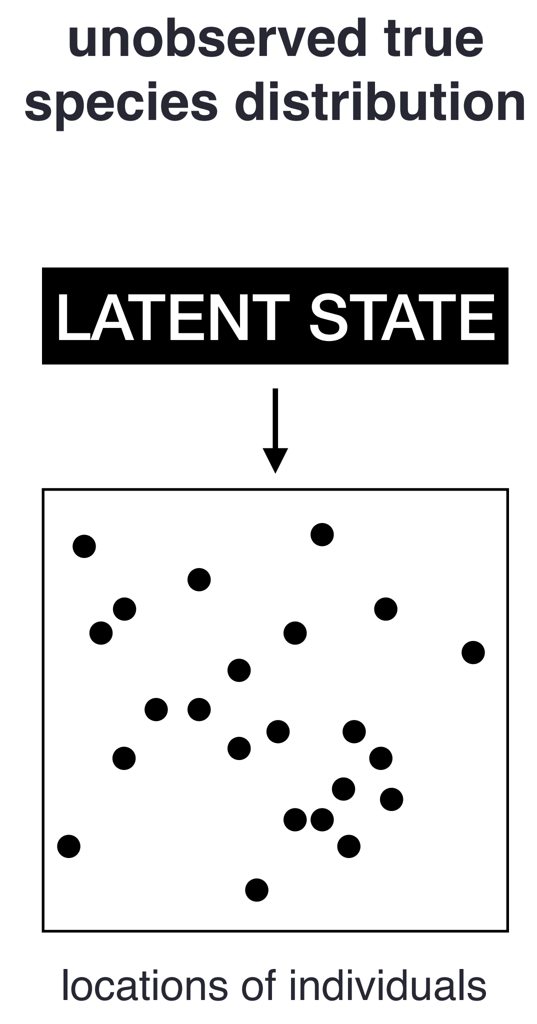
CZU Faculty of Environmental Sciences

BRITISH ECOLOGICAL SOCIETY **Anual Meeting 2022 | S27: Macroecology and Biogeography - Distributions**

IDMs

Integrated species distribution models

- They assume a common underlying spatial point process that determines the spatial locations of individuals of a species.

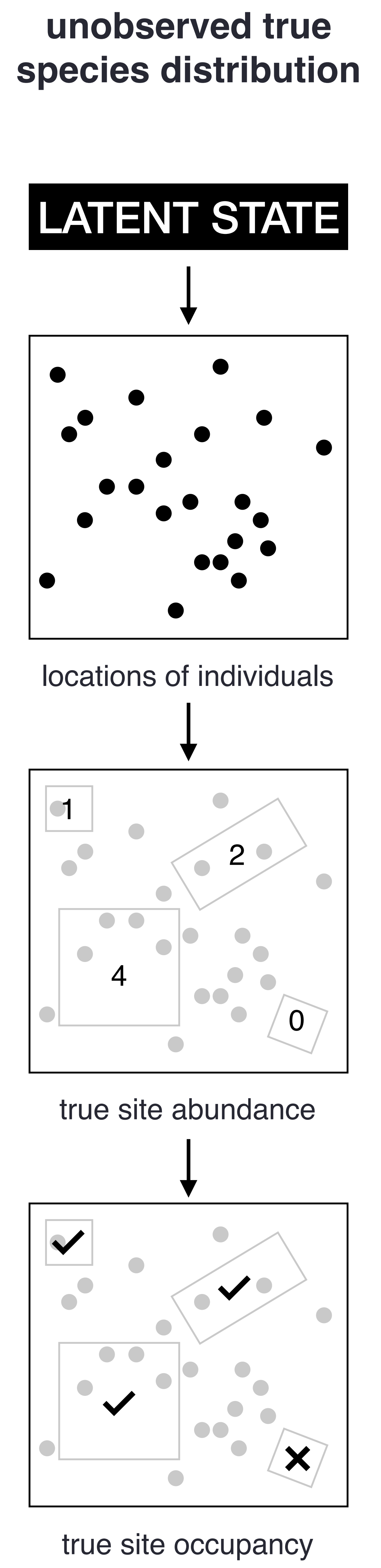


Isaac et al. (2020)

IDMs

Integrated species distribution models

- They assume a common underlying spatial point process that determines the spatial locations of individuals of a species.
- Multiple data emerge from a common set of ecological processes.

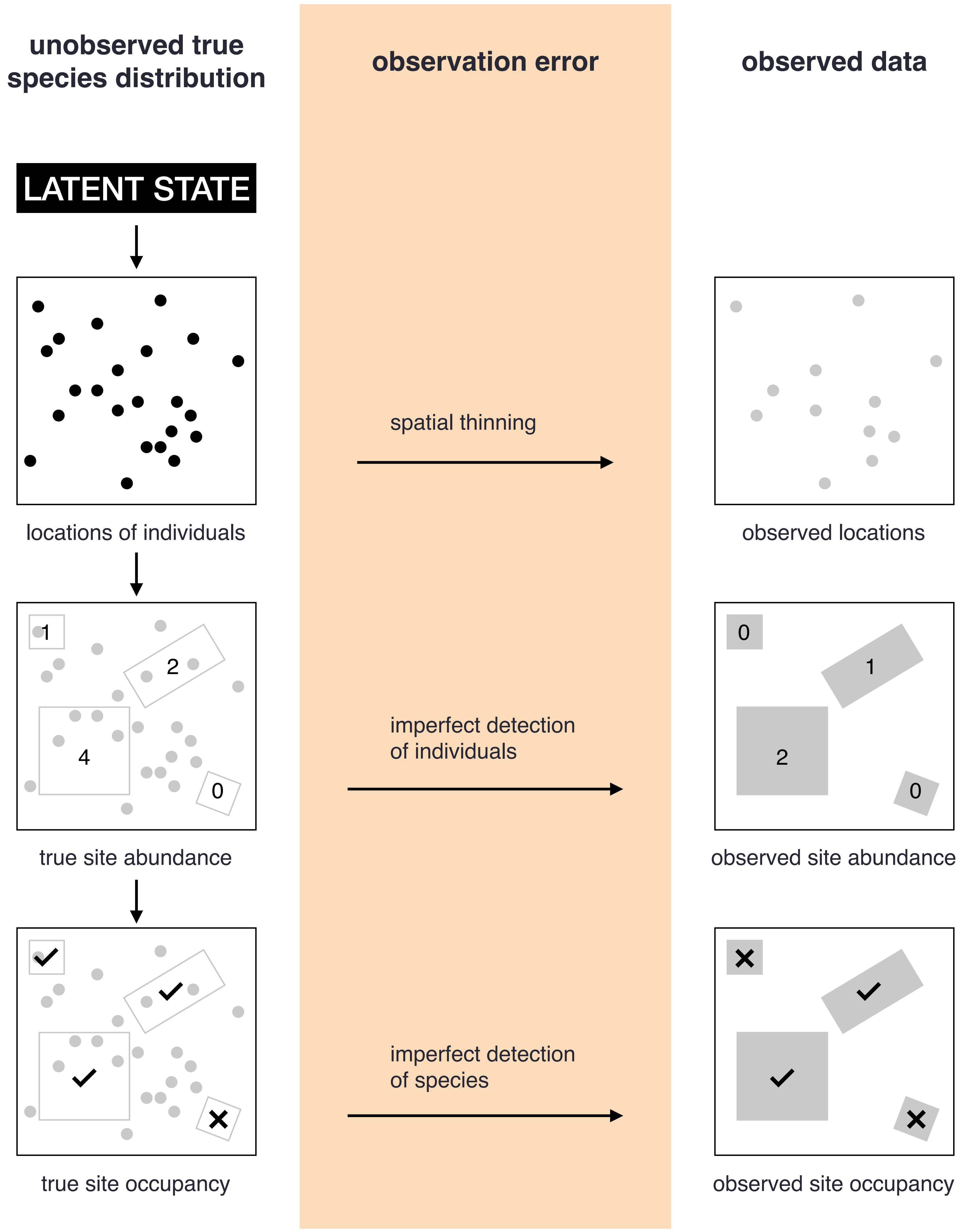


Isaac et al. (2020)

IDMs

Integrated species distribution models

- They assume a common underlying spatial point process that determines the spatial locations of individuals of a species.
- Multiple data emerge from a common set of ecological processes.
- Through different observation processes, we obtain data that are imperfect representations of the truth.



Isaac et al. (2020)

IDMs

Applications

ECOGRAPHY

Research

Integration of presence-only data from several sources: a case study on dolphins' spatial distribution

Sara Martino*, Daniela Silvia Pace*, Stefano Moro, Edoardo Casoli, Daniele Ventura, Alessandro Frachea, Margherita Silvestri, Antonella Arcangeli, Giancarlo Giacomini, Giandomenico Ardizzone and Giovanna Jona Lasinio

Received: 7 April 2021 | Revised: 24 August 2021 | Accepted: 5 September 2021
DOI: 10.1111/ddi.13416

RESEARCH ARTICLE

Diversity and Distributions WILEY

Integrating citizen-science and planned-survey data improves species distribution estimates

Viviane Zulian[1] | David A. W. Miller[2] | Gonçalo Ferraz[1]

Biological Conservation 241 (2020) 108374
Contents lists available at ScienceDirect
Biological Conservation
journal homepage: www.elsevier.com/locate/biocon
ELSEVIER

Integrating multiple data sources and multi-scale land-cover data to model the distribution of a declining amphibian

Jonathan P. Rose[a,*], Brian J. Halstead[a], Robert N. Fisher[b]
[a] U.S. Geological Survey, Western Ecological Research Center, Dixon Field Station, 800 Business Park Dr, Suite D, Dixon, CA 95620, USA
[b] U.S. Geological Survey, Western Ecological Research Center, San Diego Field Station, 4165 Spruance Road, Suite 200, San Diego, CA 92101, USA

DOI: 10.1111/ddi.12631

BIODIVERSITY RESEARCH

WILEY Diversity and Distributions

Using a novel model approach to assess the distribution and conservation status of the endangered Baird's tapir

Cody J. Schank[1,2] | Michael V. Cove[3] | Marcella J. Kelly[4] | Eduardo Mendoza[5] | Georgina O'Farrill[6] | Rafael Reyna-Hurtado[7] | Ninon Meyer[7,8] | Christopher A. Jordan[2,9,10] | Jose F. González-Maya[11] | Diego J. Lizcano[12,13] | Ricardo Moreno[8,14] | Michael T. Dobbins[15] | Victor Montalvo[16] | Carolina Sáenz-Bolaños[16,17] | Eduardo Carillo Jimenez[16] | Nereyda Estrada[18] | Juan Carlos Cruz Díaz[16,17] | Joel Saenz[16] | Manuel Spínola[16] | Andrew Carver[19] | Jessica Fort[19] | Clayton K. Nielsen[19] | Francisco Botello[20,21] | Gilberto Pozo Montuy[22] | Marina Rivero[7,23] | Jesús Antonio de la Torre[23,24] | Esteban Brenes-Mora[25,26] | Oscar Godínez-Gómez[5] | Margot A. Wood[27,28] | Jessica Gilbert[29] | Jennifer A. Miller[1]

SCIENTIFIC REPORTS

OPEN

Integrating data from different survey types for population monitoring of an endangered species: the case of the Eld's deer

ed: 12 April 2018
ed: 8 May 2019
ned online: 23 May 2019

Diana E. Bowler[1], Erlend B. Nilsen[1], Richard Bischof[2], Robert B. O'Hara[3], Thin Thin Yu[4], Tun Oo[5], Myint Aung[5] & John D. C. Linnell[1]

Methods in Ecology and Evolution
British Ecological Society
Methods in Ecology and Evolution 2014, **5**, 751–760 doi: 10.1111/2041-210X.12221

Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records

Jörn Pagel[1,2*], Barbara J. Anderson[3,4], Robert B. O'Hara[5], Wolfgang Cramer[6], Richard Fox[7], Florian Jeltsch[1], David B. Roy[8], Chris D. Thomas[4] and Frank M. Schurr[2,9]

Statistical Report

Ecology, 102(1), 2021, e03204

Integrating distance sampling and presence-only data to estimate species abundance

Matthew T. Farr[1,2,4] David S. Green[1,2,3] Kay E. Holekamp,[1,2] and Elise F. Zipkin[1,2]

Received: 25 November 2020 | Revised: 8 February 2021 | Accepted: 11 February 2021
DOI: 10.1111/ddi.13259

BIODIVERSITY RESEARCH

Diversity and Distributions WILEY

Model-based integration of citizen science data from disparate sources increases the precision of bird population trends

Lionel R. Hertzog[1] | Claudia Frank[2,3] | Sebastian Klimek[1] | Norbert Röder[4] | Hannah G. S. Böhner[4] | Johannes Kamp[2,3]

Neotropical region

Latin America

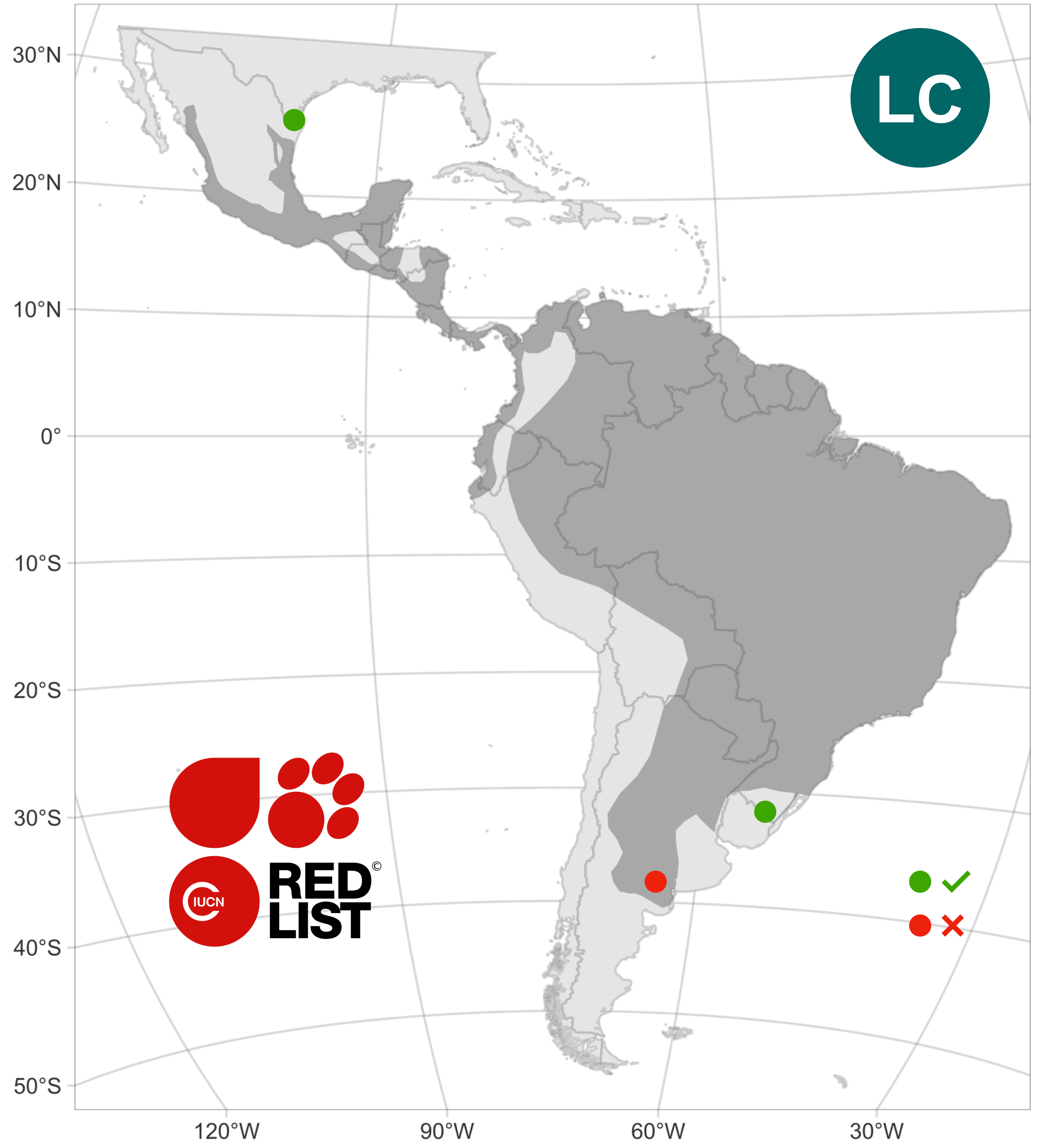
- One of the most important hotspots of biodiversity in the world.

- One of the areas where biodiversity is declining at higher rates.

Morrone (2017)

Yaguarundí

Herpailurus yagouaroundi



IUCN (2022)

Top: observed in Argentina by hhulsberg, and bottom: in Mexico by albamaya (iNaturalist.org)

Goal

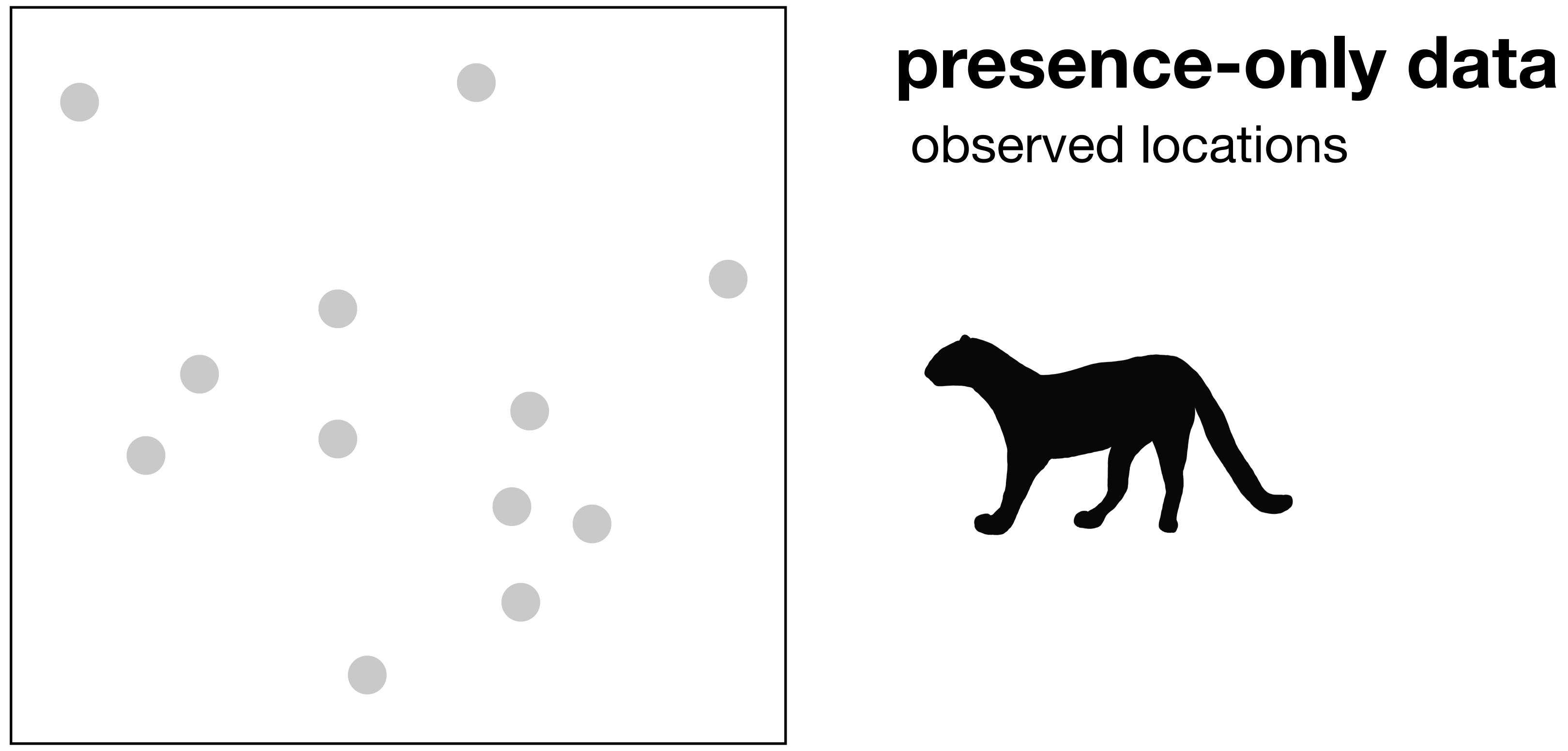
Yaguarundí's range dynamic

- Develop an **integrated species distribution model** (IDM) to model the **temporal dynamics** of the species' entire geographic range (over two time periods).

Methods

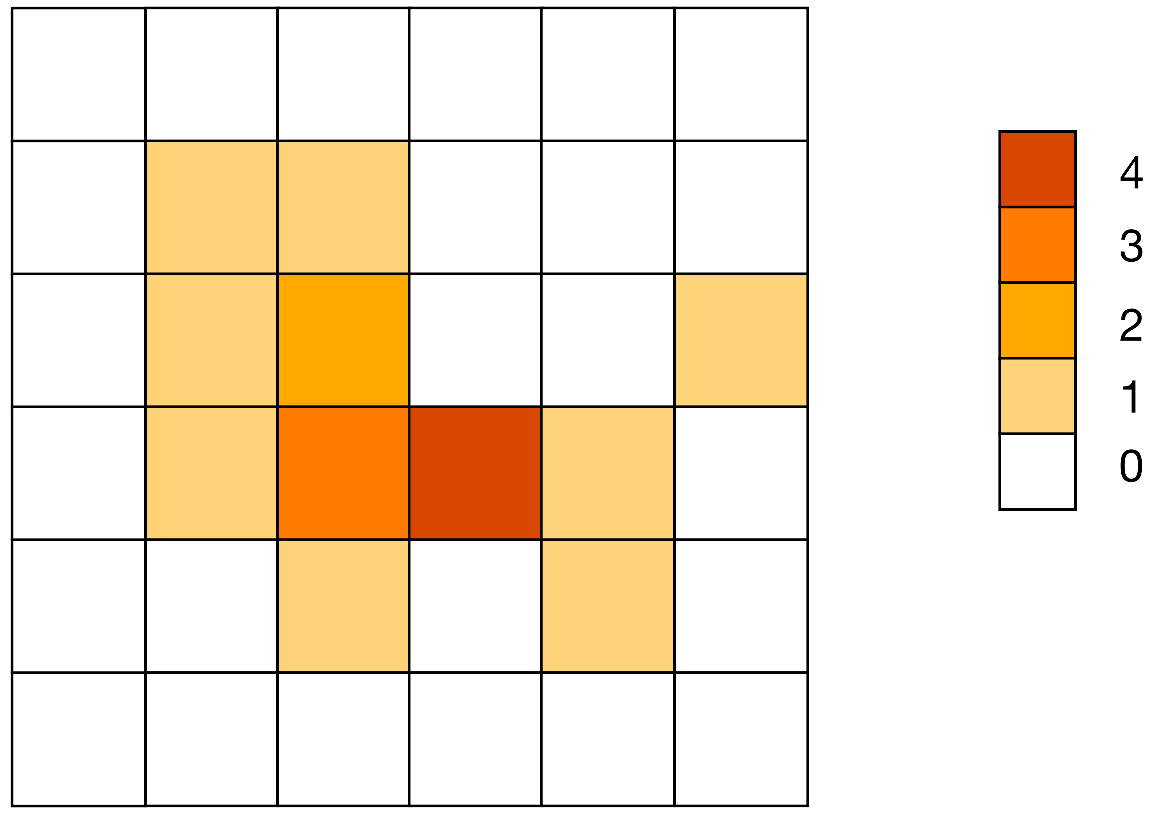
Species data

Yaguarundí (2000-2021)



GBIF.org (2022)

https://doi.org/10.15468/DL.3CU474

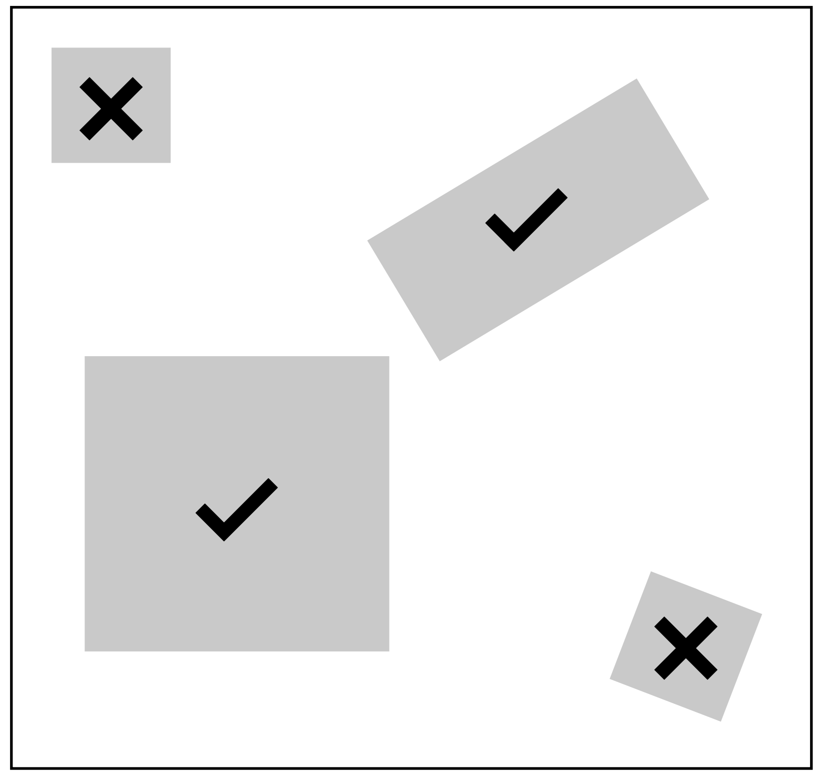


grid of 100x100 km over the study area

- We removed records with a coordinate precision > 0.01 and a coordinate uncertainty > 25km. We also removed duplicated records.

Species data

Yaguarundí and other Neotropical carnivores (2000-2021)

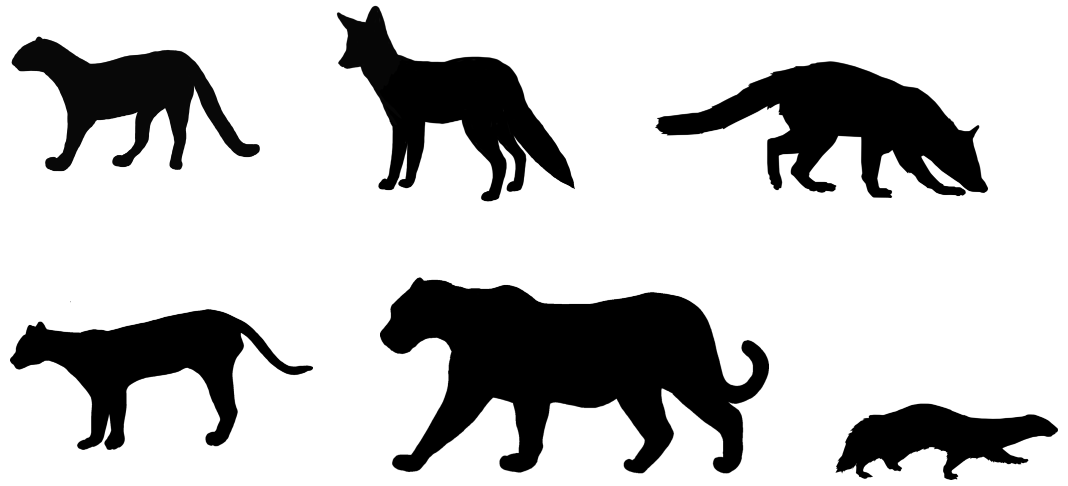


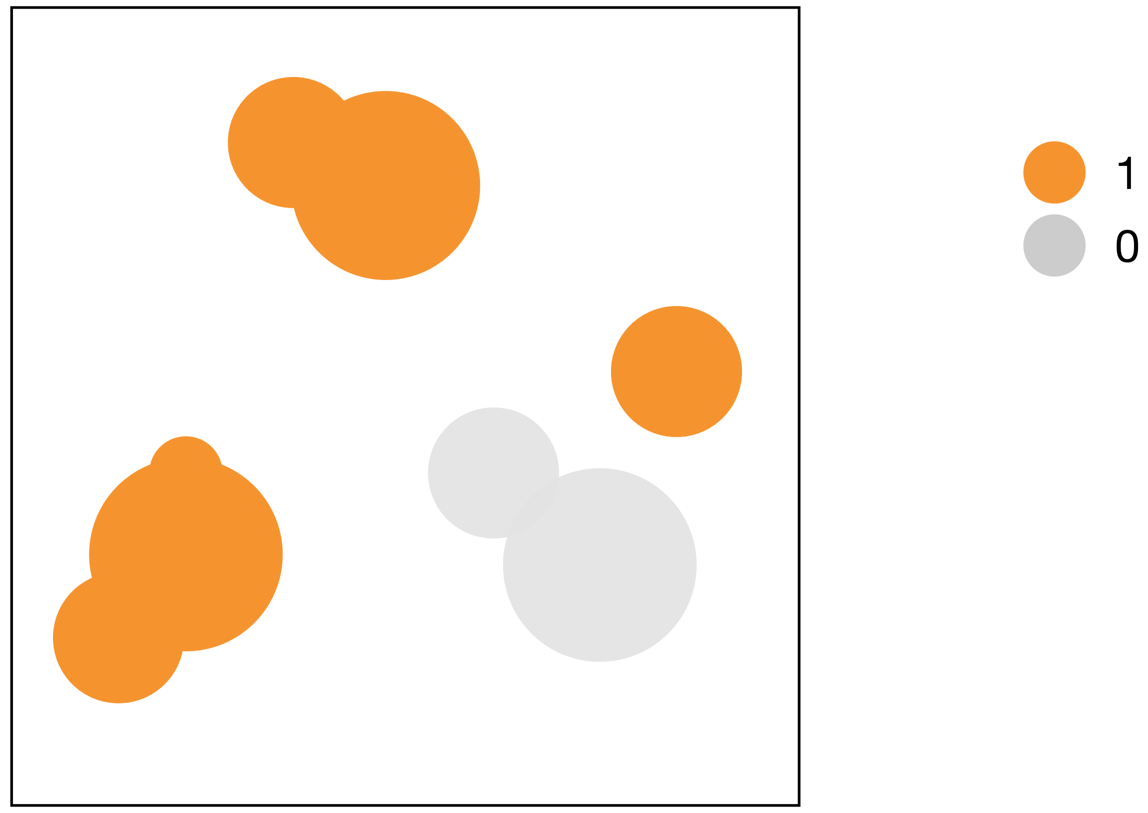
Nagy-Reis et al. (2020)

https://doi.org/10.1002/ecy.3128

presence-absence data

observed site occupancy





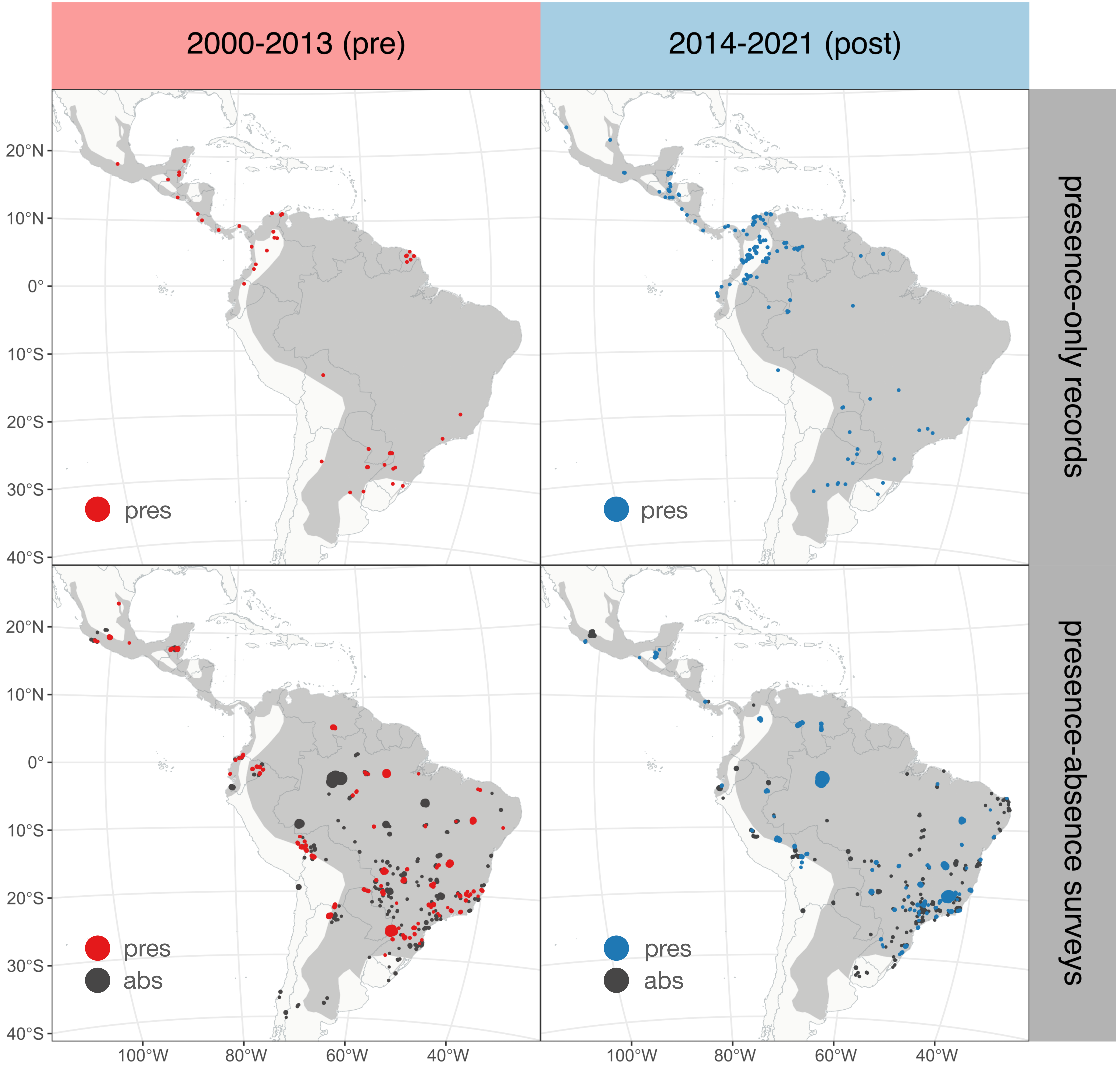
blobs of presence-absence (per time period)

- We kept surveys that used camera traps and had info about the sampling area, the sampling effort, and the temporal span of the study.

Species data

Yaguarundí

- The first period had 196 yaguarundi occurrence records, and the second 234.
- We used data from 8,346 surveys. The yaguarundí was recorded in 614.

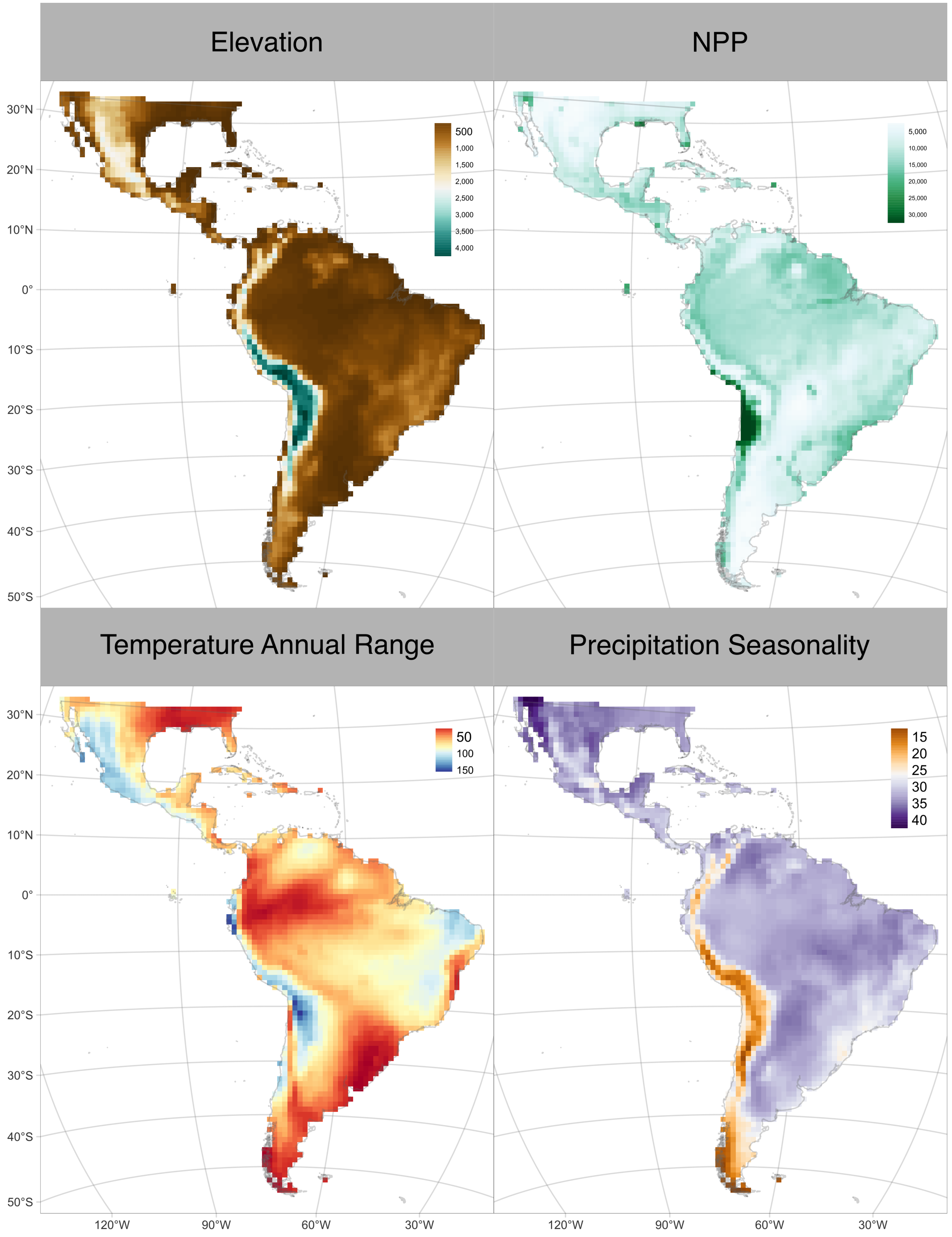


Covariates data

Climate, landcover, vegetation

We used a set of 28 potential covariates:

- Bioclimatic variables (WorldClim V2.1): bio1 - bio19.
- Elevation (WorldClim V2.1 SRTM elevation data)
- Land cover (MODIS - MCD12Q1): urban, barren, water, savanna, wetland, grassland.
- Net Primary Production (NPP) (MODIS - M*D17A3HGF)
- Percentage of Vegetation cover (MODIS TERRA - MOD44B): tree cover, no tree cover, non tree vegetation cover.



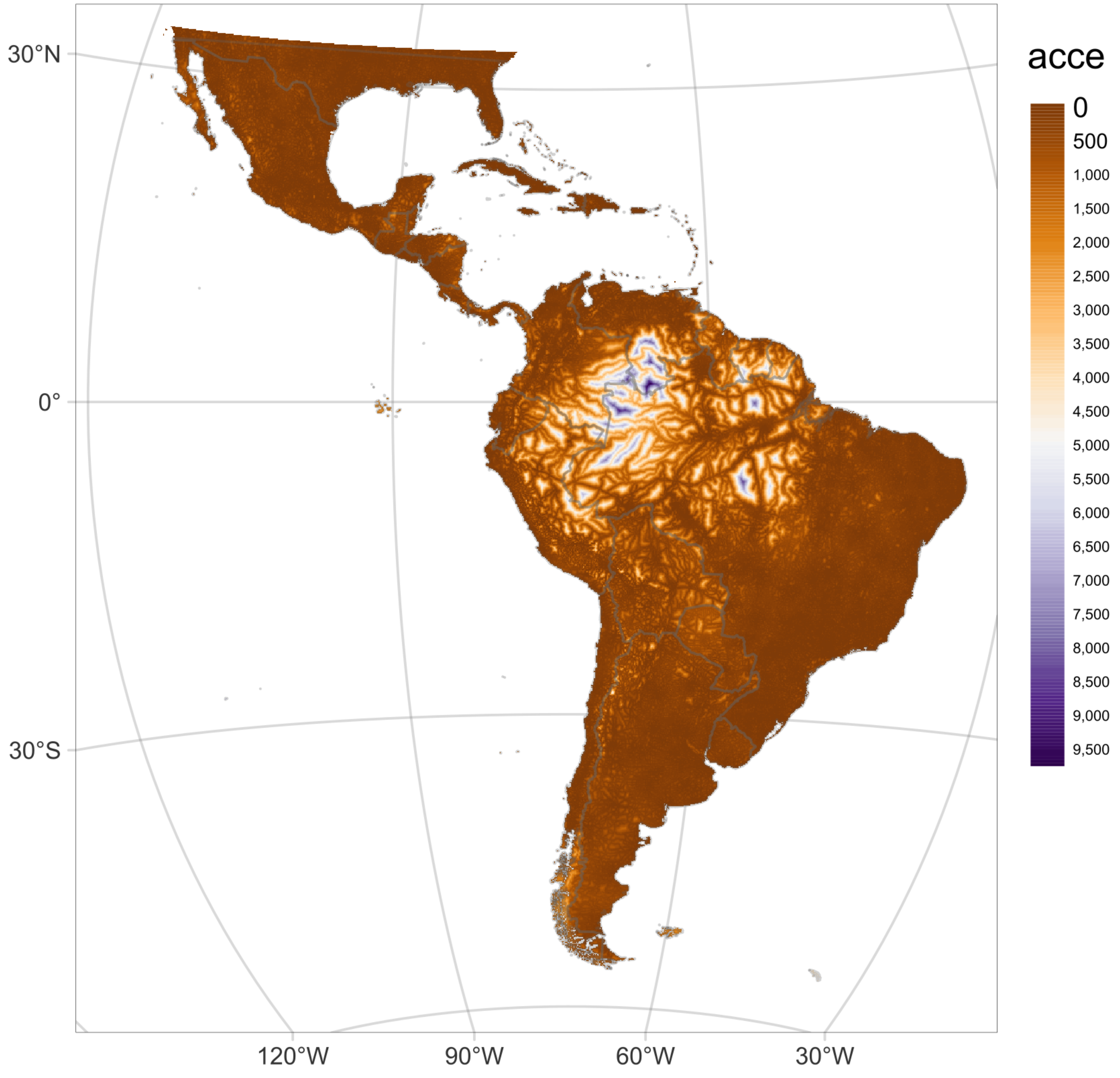
$r < 0.6$

Thinning data

Presence-only observation process

We used:

- **Accessibility**: we expected that highly accessible grid cells would have more point records than inaccessible grid cells.



Weiss et al. (2020)

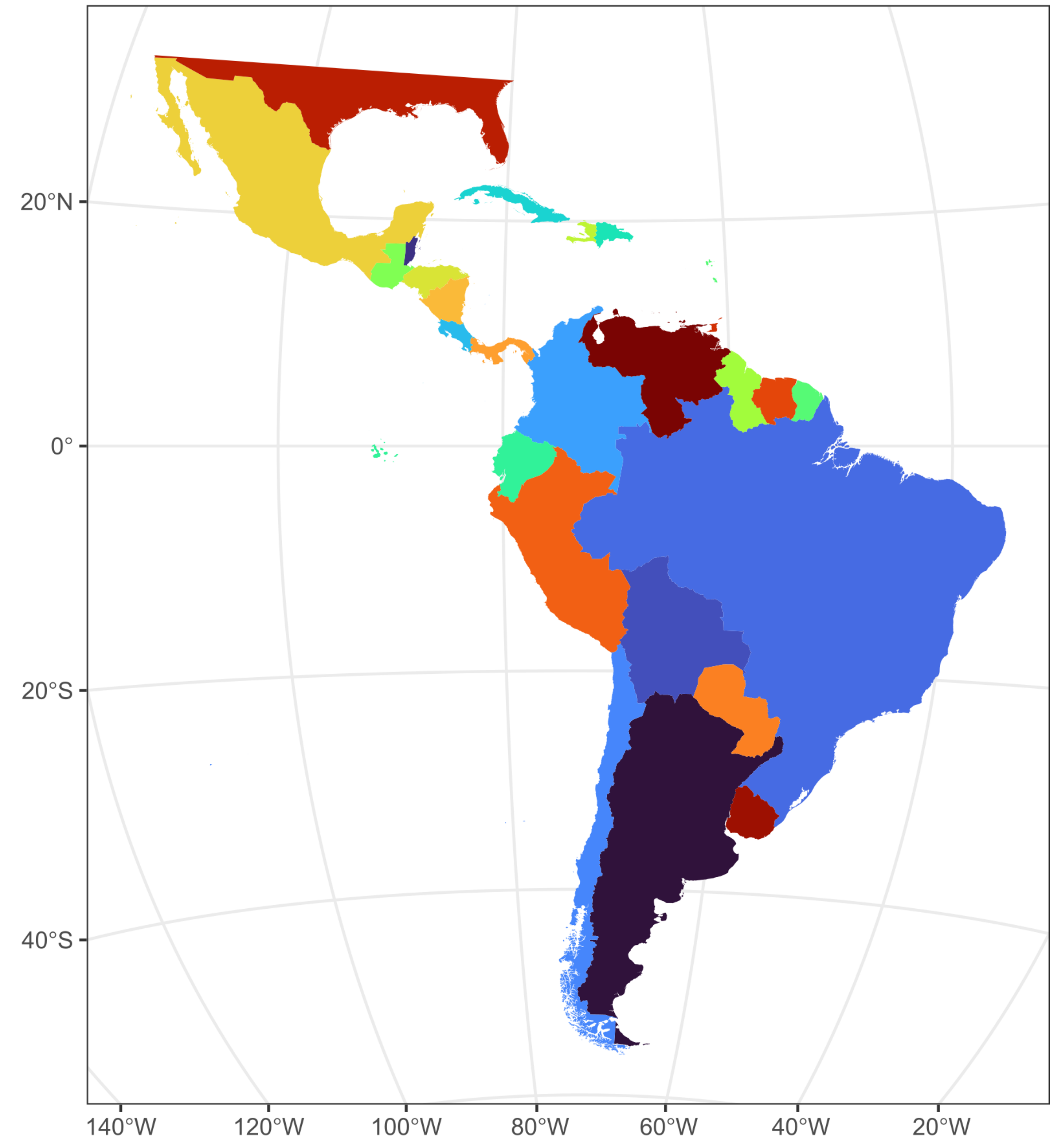
Thinning data

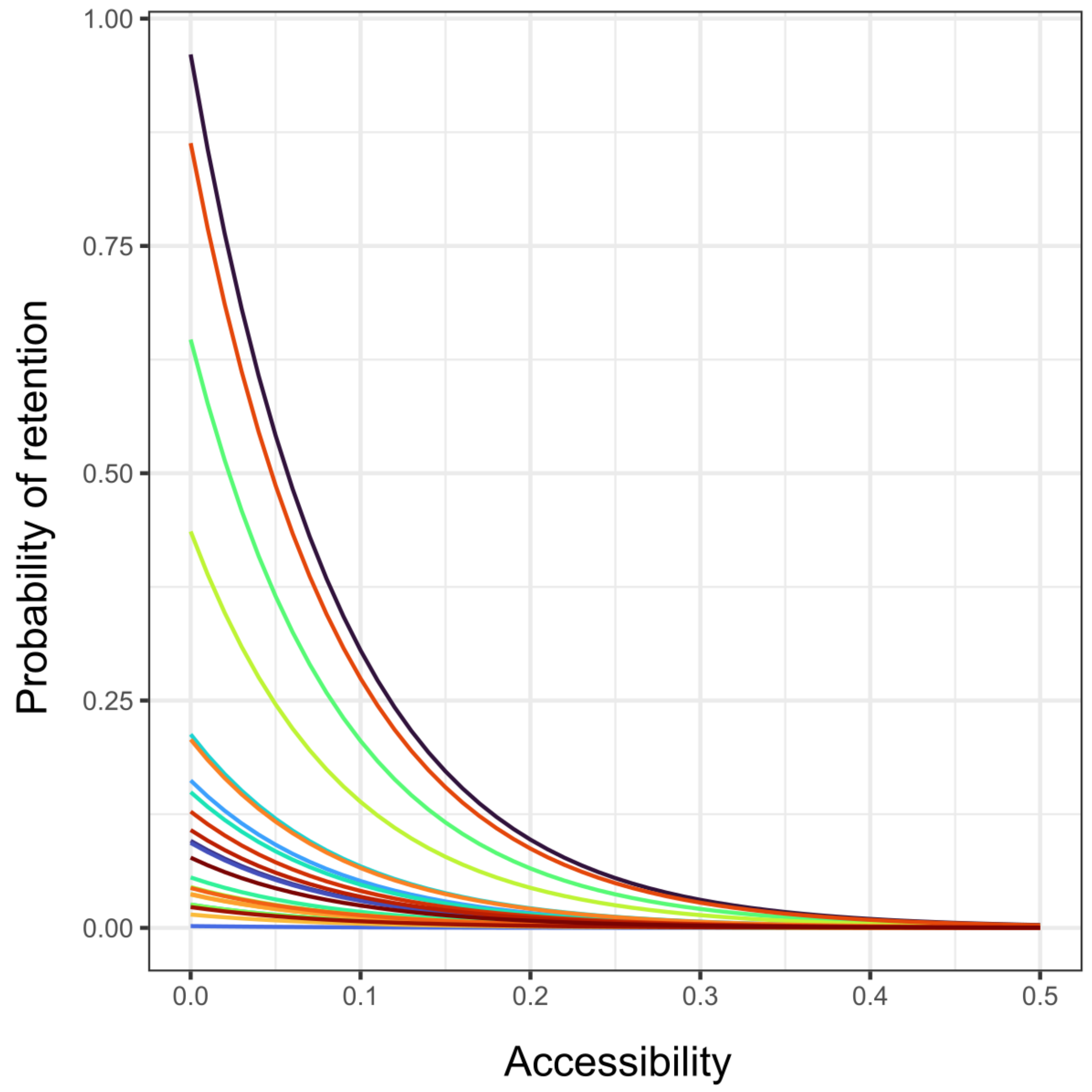
Presence-only observation process

We used:

- **Accessibility**: we expected that highly accessible grid cells would have more point records than inaccessible grid cells.

- **Country of origin**: differences in data-sharing capacities and citizen-science levels of engagement among countries.



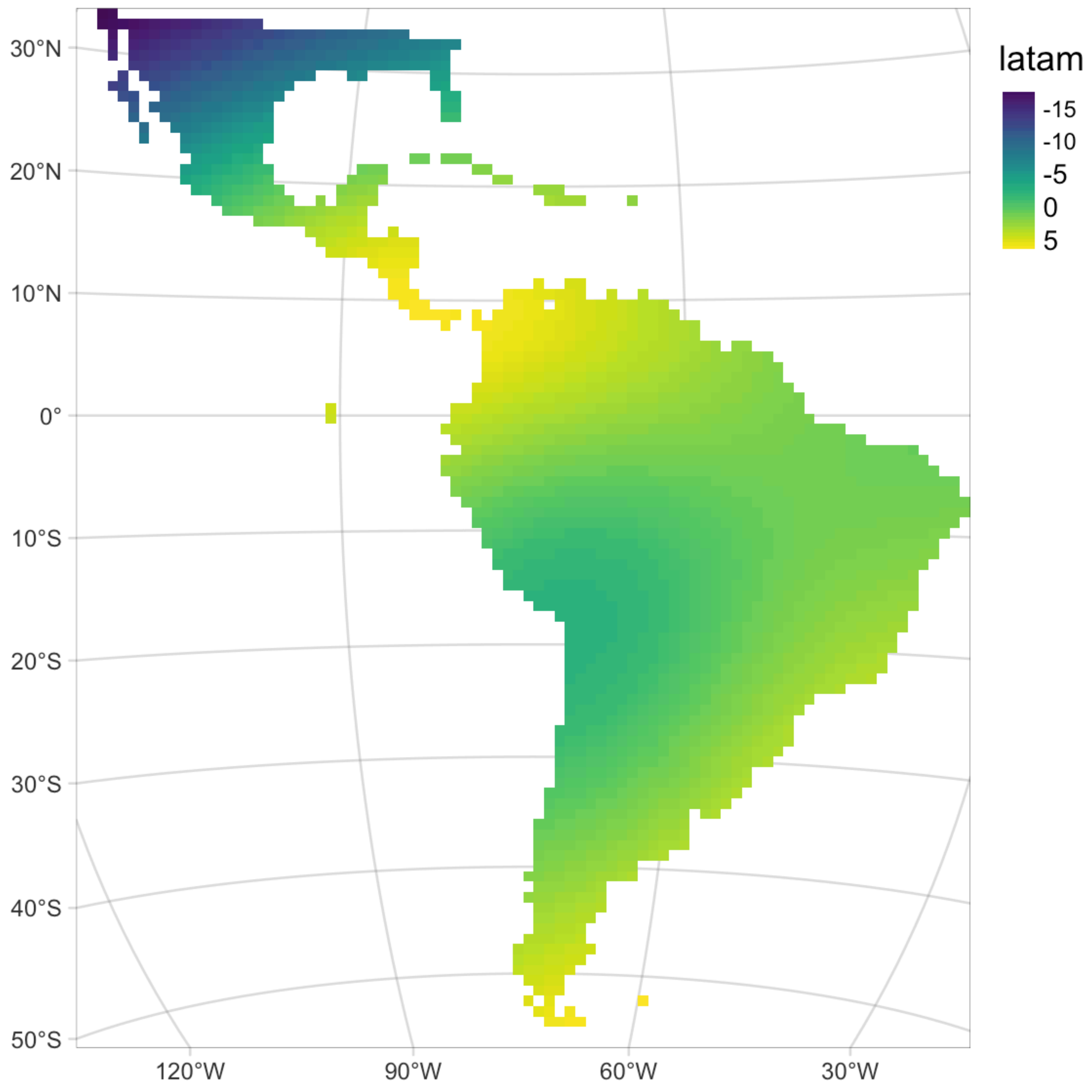


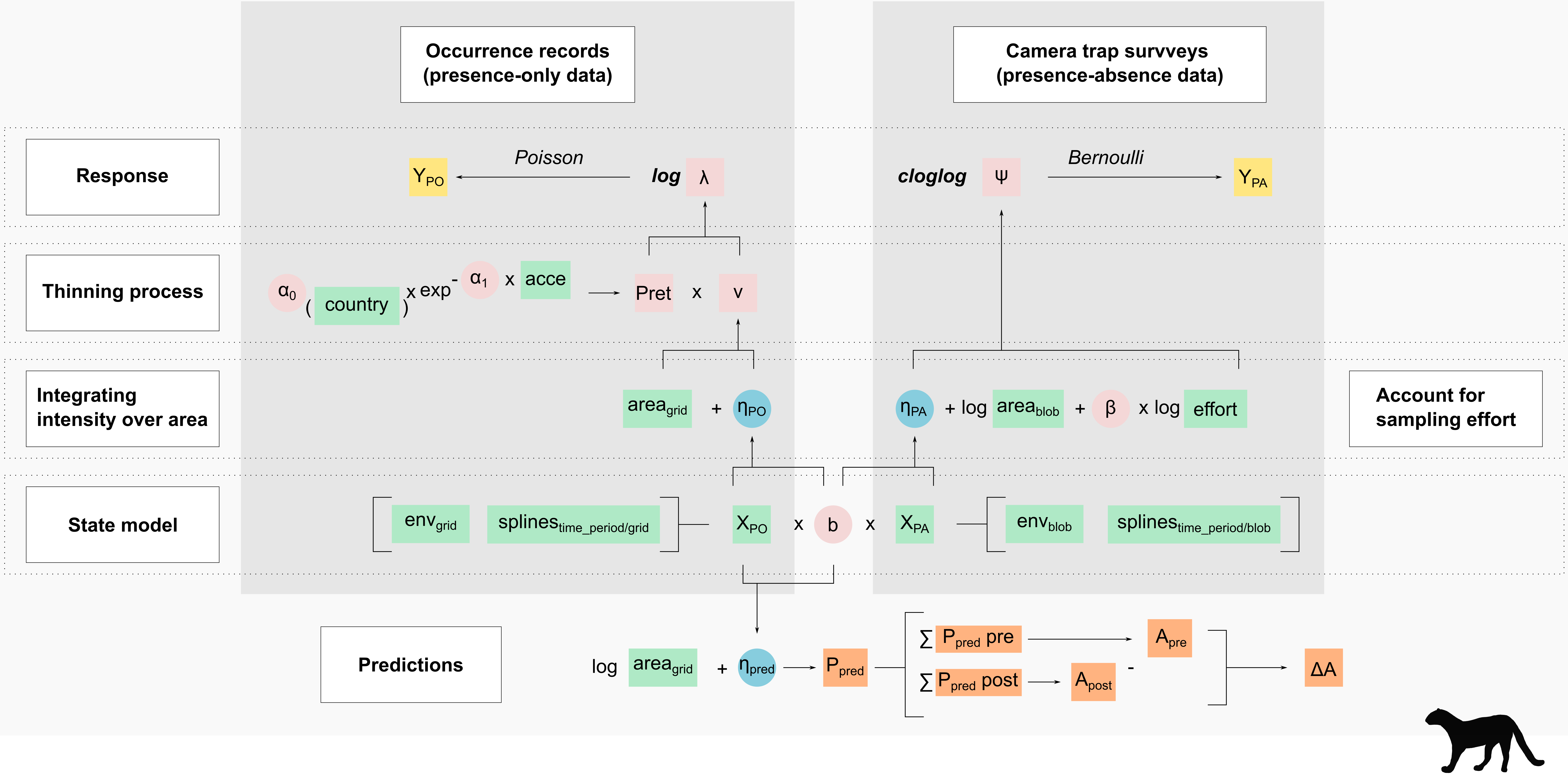
Smoothing splines

Spatial component

- They model the spatial structure in the distribution that is not accounted for by the environmental covariates.

- We used the `jagam` function from the ‘mgcv’ package and *k*=9 spline basis variables.





Occurrence records (presence-only data)
Camera trap survveys (presence-absence data)
Response
Y_{PO}
Poisson
log
λ
cloglog
ψ
Bernoulli
Y_{PA}
Thinning process
α_0
country
x exp -
α_1
x
acce
Pret
x
v
Integrating intensity over area
$area_{grid}$
+
η_{PO}
η_{PA}
+ log
$area_{blob}$
+
β
x log
effort
Account for sampling effort
State model
env_{grid}
$splines_{time\_period/grid}$
X_{PO}
x
b
x
X_{PA}
env_{blob}
$splines_{time\_period/blob}$
Predictions
log
$area_{grid}$
+
η_{pred}
P_{pred}
$\sum P_{pred}$ pre
A_{pre}
$\sum P_{pred}$ post
A_{post}
-
ΔA

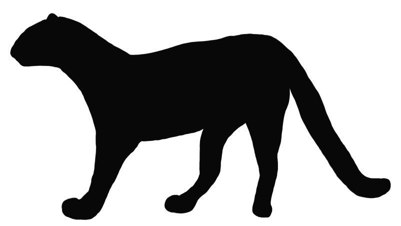
The model

Run

- `jags` function from the ‘R2jags’ package (to call JAGS from R)
    - 3 chains
    - 100,000 iterations per chain
    - 10,000 burning length
    - 10 as thinning rate

97 lines (71 sloc) | 3.01 KB   Raw Blame

```
model
  {
    # PRIORS ---------------------------------------------------

    ## Thinning at locations with complete accessibility in PO data

      # intercept of the decay function for each country of origin.
      # It needs a flat prior between 0 and 1
      for (c in 1:n.cntr)
      {
        alpha0[c] ~ dbeta(1, 1)
      }

      # steepness of the decaying distance-P.ret relationship in PO data
      alpha1 ~ dgamma(0.5, 0.05)

    ## Effect of sampling effort in PA data
      beta ~ dnorm(0, 0.01)

    ## Parametric effects of environment driving the point process intensity
     # (it also includes an intercept)

      for (r in 1:n.par)
      {
        b[r] ~ dnorm(0,0.01)
      }

    ## Splines (imported and adjusted form output of mgcv::jagam)

      ## prior for s(X,Y):as.factor(time)0
      sigma.pre <- S.pre[1:n.spl, 1:n.spl] * gamma[1]  +
            S.pre[1:n.spl, (n.spl + 1):(n.spl * 2)] * gamma[2]
      b[(n.par+1):(n.spl + n.par)] ~ dmnorm(Z[(n.par+1):(n.spl + n.par)], sigma.pre)

      ## prior for s(X,Y):as.factor(time)1
      sigma.post <- S.post[1:n.spl, 1:n.spl] * gamma[3]  +
            S.post[1:n.spl, (n.spl + 1):(n.spl * 2)] * gamma[4]
      b[(n.X - n.spl + 1):(n.X)] ~ dmnorm(Z[(n.X - n.spl + 1):(n.X)], sigma.post)

      ## Priors for smoothing parameter
      for (f in 1:n.fac)
      {
        gamma[f] ~ dgamma(.5,.5)
        rho[f] <- log(gamma[f])
      }

    # LIKELIHOOD ------------------------------------------------

      ## --- Presence-Absence (PA) data ---

       eta.PA <- X.PA %*% b ## linear predictor

       for (i in 1:n.PA)
       {
         # the probability of presence
         cloglog(psi[i]) <- eta.PA[i] + log(area.PA[i]) + beta*log(effort[i])

         # presences and absences come from a Bernoulli distribution
         y.PA[i] ~ dbern(psi[i]*0.9999)
       }

      ## --- Presence-Only (PO) data ---

      eta.PO <- X.PO %*% b  ## linear predictor

      for (j in 1:n.PO)
      {
        # cell-specific probability of retainin (observing) a point is a function of accessibility
        P.ret[j] <- alpha0[country[j]] * exp( (-alpha1) * acce[j])

        # true mean number (nu) of points per cell i is the true intensity multiplied by cell area
        nu[j] <- area.PO[j] * exp(eta.PO[j])

        # thinning: the true lambda
        lambda[j] <- nu[j] * P.ret[j]

        # counts of observed points come from a Poisson distribution
        y.PO[j] ~ dpois(lambda[j])
      }

    # PREDICTIONS -----------------------------------------------

    eta.pred <- X.PO %*% b

    for (j in 1:n.PO)
    {
      # predicted probability of occurrence in grid cell j
      cloglog(P.pred[j]) <- eta.pred[j] + log(area.PO[j])
    }

    # DERIVED QUANTITIES ----------------------------------------

    # area in each time period, and temporal change of area
    A.pre <- sum(P.pred[1:n.PO.half])
    A.post <- sum(P.pred[(n.PO.half+1):n.PO])
    delta.A <- A.post - A.pre
  }
```

The model

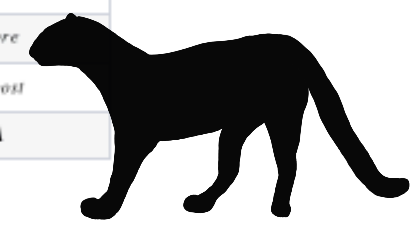
Bayesian IDM

- The data, code, model, outputs and more can be found at: https://github.com/bienflorencia/yaguarundi_IDM

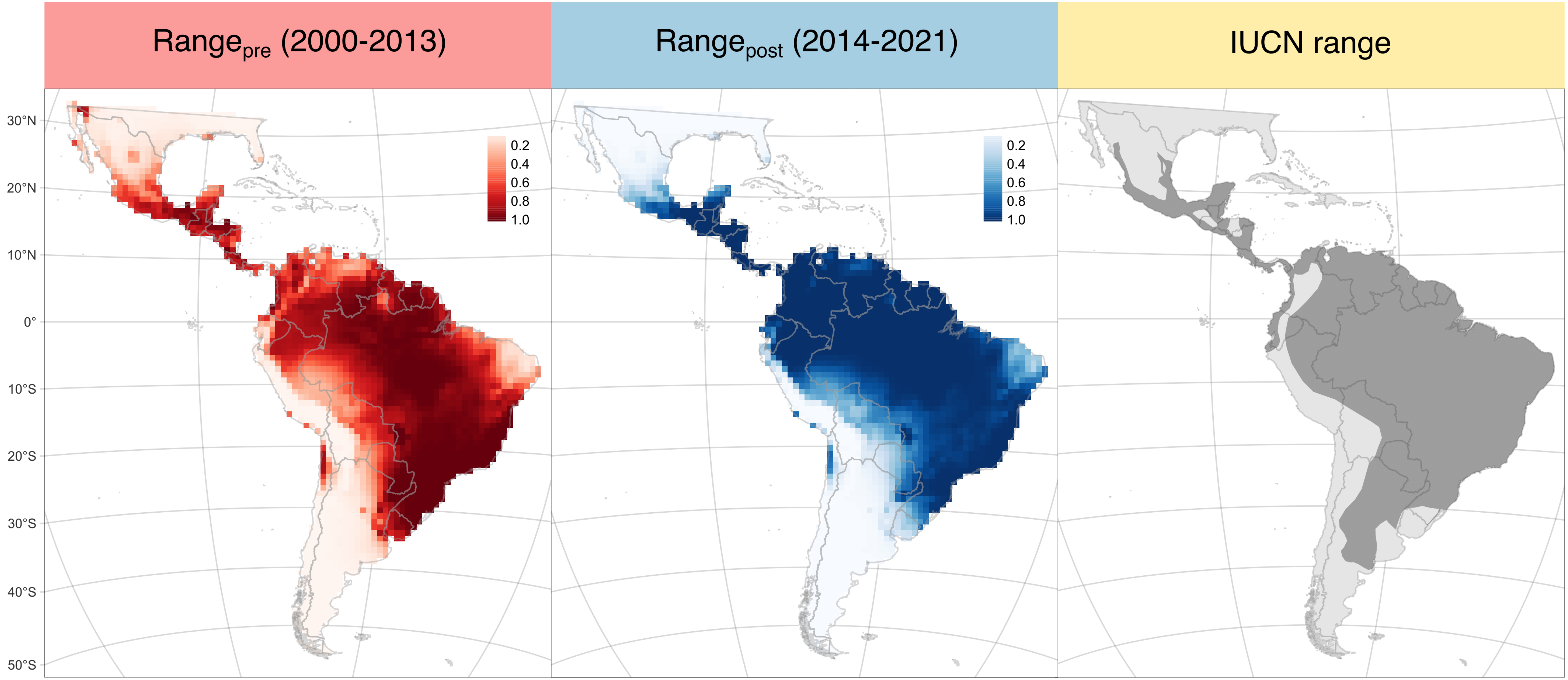
Quick Model References

For more details please see here

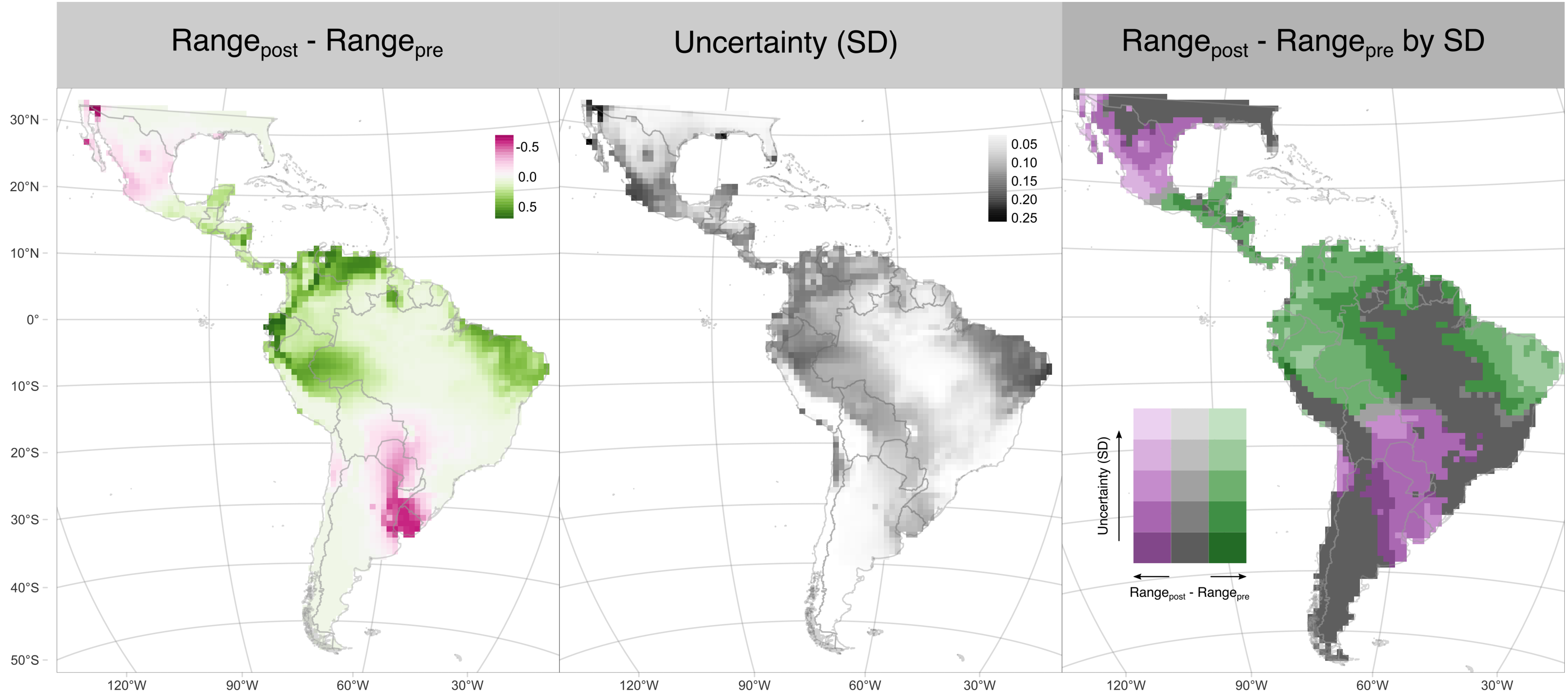
Model term	Definition	Equation notation
n.PA	number of blobs for both time periods (pre and pos)	n_{PA}
i	index identifying blobs	$i \text{ where } i \in 1:n_{PA}$
y.PA[i]	presence (1) or absence (0) value in each i-th blob (overlapping surveys' area), can be for pre- or post- period	y_{PA_i}
X.PA	design matrix including vector of 1s (for intercept) and all the covariates and spline bases for each blob, for both time periods	$\mathbf{X_{PA}}$
area.PA[i]	area of i-th blob in meters for both time periods	$area_{PA_i}$
effort[i]	sampling effort for i-th blob in the given period for both time periods	$effort_{PA_i}$
n.PO	number of grid-cells for both time periods concatenated (pre and pos)	n_{PO}
j	index identifying grid cells	$j \text{ where } j \in j:n_{PO}$
n.PO.half	number of grid-cells for one time period	$n_{PO/2}$
y.PO[j]	count of observed points in j-th grid-cell, can be for pre- or post- period	y_{PO_j}
X.PO	design matrix including vector of 1s (for intercept) and all the covariates and spline bases for each grid-cell for both time periods	$\mathbf{X_{PO}}$
area.PO[j]	area of each grid-cell in meters for both time periods	$area_{PO_j}$
acce[j]	accessibility from urban areas based on travel time for j-th grid-cell for both time periods	$acce_j$
country[j]	country name for j-th grid-cell for both time periods	$country_j$
n.X	total number of columns in X (`X.PA` or `X.PO`)	n_X
n.cntr	total number of countries	n_{cntr}
c	index identifying countries	$c \text{ where } c \in 1:n_{cntr}$
n.par	number of parameters considered (intercept and covariates)	n_{par}
r	index identifying parameters	$r \text{ where } r \in 1:n_{par}$
n.fac	number of factors of time in X (`X.PA` or `X.PO`)	n_{fac}
f	index identifying factors	$f \text{ where } f \in 1:n_{fac}$
n.spl	number of spline bases functions in in X (`X.PA` or `X.PO`)	n_{spl}
S.pre	spline values for the first time period (pre)	S_{pre}
S.post	spline values for the second time period (post)	S_{post}
Z	a vector of zeros (0) of the length of the splines	Z
sigma.pre	variance of splines for the first time period (pre)	σ_{pre}
sigma.post	variance of splines for the second time period (post)	σ_{post}
b	vector of parametric effects of covariates driving the point process intensity (it also includes an intercept)	$b_r \in \mathbf{b}$
alpha0	intercept of the thinning process in the presence-only data	α_0
alhpa1	slope -steepness- of the thinning process in the presence-only data (decaying distance~P.ret relationship)	α_1
beta	coefficient of the effect of sampling effort in the presence-absence data	β
gamma	prior for splines smoothing parameter	γ
eta.PA	linear predictor for presence-absence data	$\boldsymbol{\eta}_{\mathbf{PA}}$
eta.PA[i]	expected presence-absence for the i-th blob	η_{PA_i}
eta.PO	linear predictor for presence-only data	$\boldsymbol{\eta}_{\mathbf{PO}}$
eta.PO[j]	expected count points for the j-th grid-cell	η_{PO_j}
psi[i]	blob-specific probability of presence	ψ_i
P.ret[j]	cell-specific probability of retaining (observing) a point as a function of accessibility and country of origin	P_{ret_j}
nu[j]	true mean number of points per grid-cell (the true intensity)	ν_j
lambda[j]	thinning of the true intensity	λ_j
eta.pred	linear predictor for the predicted probability of occurrence	$\boldsymbol{\eta}_{\mathbf{pred}}$
eta.pred[j]	predicted count points for the j-th grid-cell	η_{pred_j}
P.pred[j]	predicted probability of occurrence for the j-th grid-cell	P_{pred_j}
A.pre	range area in the first time period (pre)	A_{pre}
A.post	range area in the second time period (post)	A_{post}
delta.A	temporal change of range area (post-pre)	ΔA



Results and Discussion



@flograttarola | ecoevo.social/@flograttarola

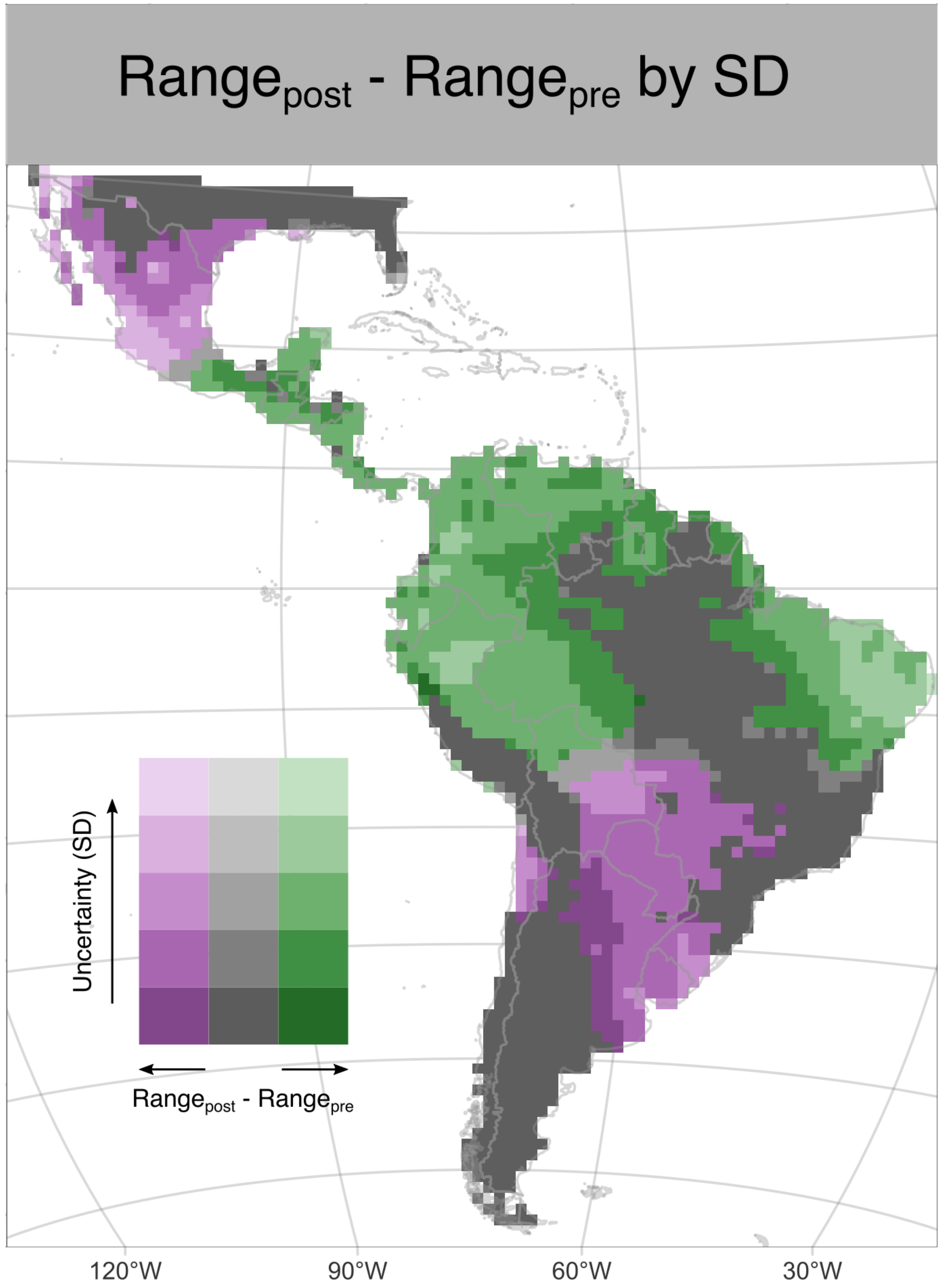


@flograttarola | ecoevo.social/@flograttarola

Species range

Temporal change

- **Retracted** from the southern range limit in Argentina, Uruguay and Paraguay and the northern limit in Mexico.

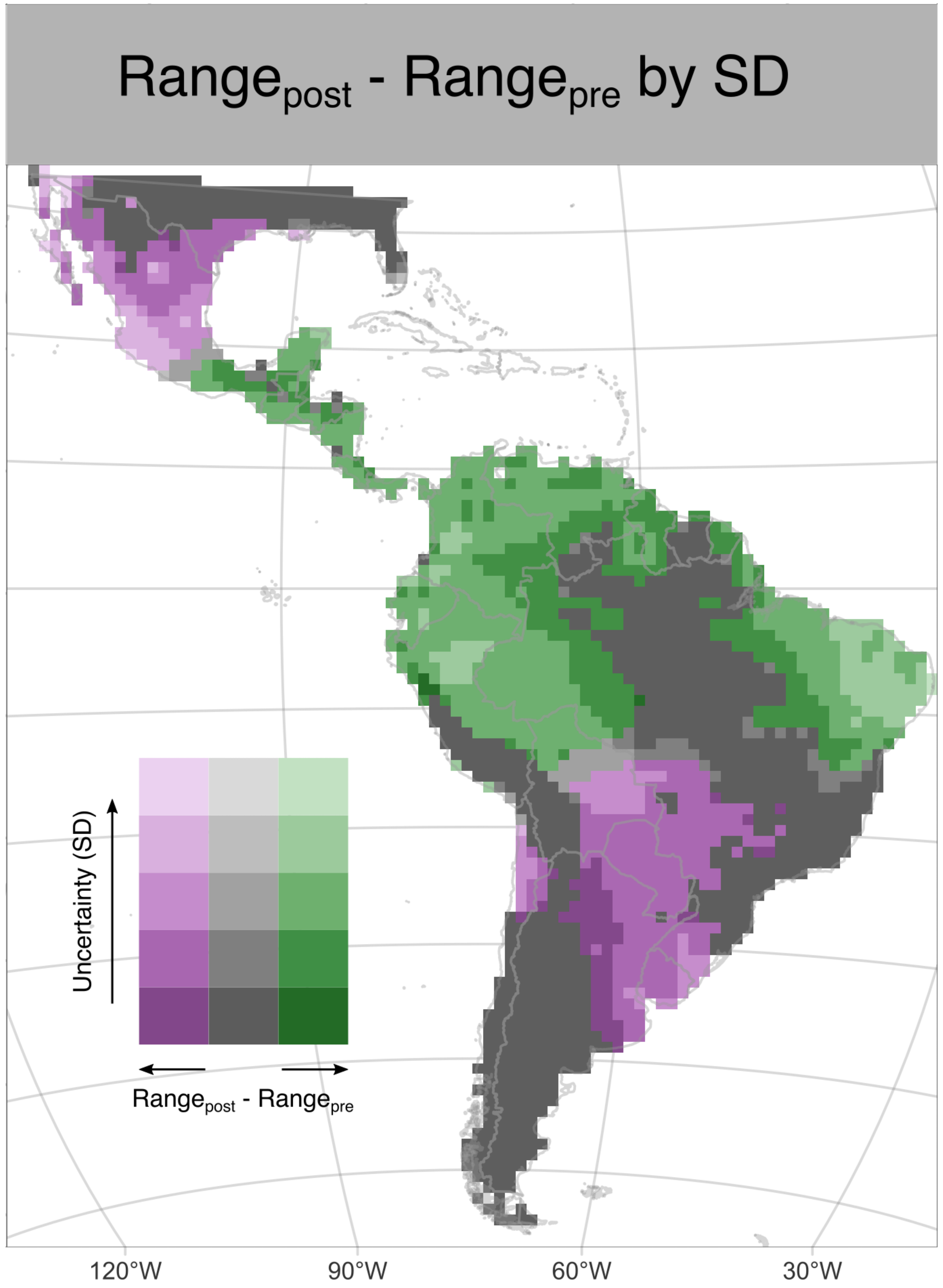


@flograttarola | ecoevo.social/@flograttarola

Species range

Temporal change

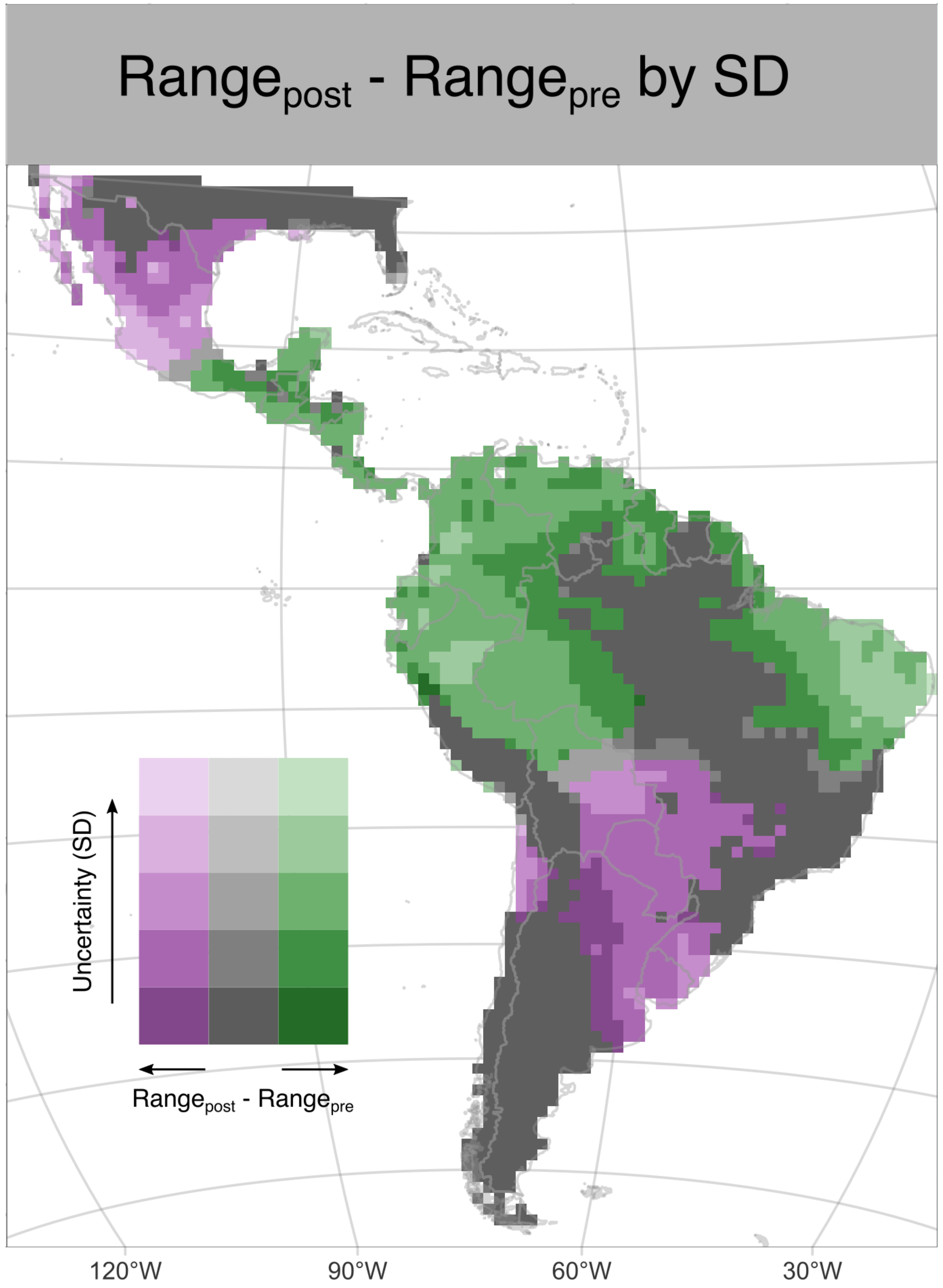
- **Retracted** from the southern range limit in Argentina, Uruguay and Paraguay and the northern limit in Mexico.
- **Maintained** its presence in central and southeast Brazil.



Species range

Temporal change

- **Retracted** from the southern range limit in Argentina, Uruguay and Paraguay and the northern limit in Mexico.
- **Maintained** its presence in central and southeast Brazil.
- **Expanded** at Brazilian and Colombian Amazon, near the Caatinga region of north-eastern Brazil and the border of Mexico with Guatemala.



@flograttarola | ecoevo.social/@flograttarola

Conclusions

Conclusions

IDMs - Temporal change

- Data integration enabled us to increase each period's sample size, geographic extent, and environmental scope.

Observed by alexcg223 (iNaturalist.org)

Conclusions

IDMs - Temporal change

- Data integration enabled us to increase each period's sample size, geographic extent, and environmental scope.
- We were able to estimate the temporal change in the species' geographic range even over a relatively short time span while accounting for sampling bias and spatial autocorrelation.

Observed by alexcg223 (iNaturalist.org)

Conclusions

IDMs - Most up-to-date knowledge

- We have updated the knowledge represented by the IUCN expert's range map.

Observed by ddavilareyes (iNaturalist.org)

Conclusions

IDMs - Most up-to-date knowledge

- We have updated the knowledge represented by the IUCN expert's range map.
- Many global studies rely on these sorts of maps; thus, they need to be more accurate. IDMs can be a solution to improve them.

Observed by ddavilareyes (iNaturalist.org)

What’s next?

What’s next?

Scope for improvements

- **Better predictions** - Should we increase the number of splines? Correct predictions by including experts’ range maps as offsets? (Merow et al. 2016)

What’s next?

Scope for improvements

- **Better predictions** - Should we increase the number of splines? Correct predictions by including experts’ range maps as offsets? (Merow et al. 2016)
- **Spatial component** - How to model spatial autocorrelation in JAGS/NIMBLE: are there any practical solutions other than splines?

What’s next?

Scope for improvements

- **Better predictions** - Should we increase the number of splines? Correct predictions by including experts’ range maps as offsets? (Merow et al. 2016)
- **Spatial component** - How to model spatial autocorrelation in JAGS/NIMBLE: are there any practical solutions other than splines?
- **Changes over time in climatic variables** - How to include the temporality of covariates: are there any products available for Latin America?

What's next?

Scope for improvements

- **Better predictions** - Should we increase the number of splines? Correct predictions by including experts' range maps as offsets? (Merow et al. 2016)
- **Spatial component** - How to model spatial autocorrelation in JAGS/NIMBLE: are there any practical solutions other than splines?
- **Changes over time in climatic variables** - How to include the temporality of covariates: are there any products available for Latin America?
- **Interaction with other species** - How to model co-occurrence effects: can we upgrade our model to model more than one species jointly?

What's next?

Challenge: increase data available in Latin America

- **Engage more community-science users**, i.e., through iNaturalist.

NaturalistaUY

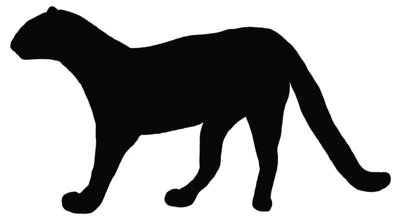
What's next?

Challenge: increase data available in Latin America

- **Engage more community-science users**, i.e., through iNaturalist.
- **Digitise new camera trap studies**, i.e., literature and community science initiatives.

NaturalistaUY

Thanks!



Florencia Grattarola,
Daiana E. Bowler
and Petr Keil

Integrating presence-only and presence-absence data to model changes in species geographic ranges: An example of yaguarundí in Latin America (**2022**) *EcoEvoRxiv*.

https://doi.org/10.32942/osf.io/67c4u

SORTEE EcoEvoRxiv

CZU Faculty of Environmental Sciences

UK Centre for Ecology & Hydrology

Acknowledgement
REES (Research Excellence in Environmental Sciences) grant from the Faculty of Environmental Sciences, Czech University of Life Sciences in Prague.

Credits
Photos by Hugo Hulsberg (CC-BY-NC), albamaya (CC-BY-NC), Christoph Moning (CC-BY), Bernard Foy (CC-BY-NC), Scott Ramos (CC-BY-NC), drewdavis74 (CC-BY-NC), ddavilareyes (CC-BY-NC), alexcg223 (CC-BY-NC) - iNaturalist.
Silhouettes by Gabriela Palomo-Muñoz (CC BY-NC) and Margot Michaud (CC0) - PhyloPic.

flograttarola.com | GitHub github.com/bienflorencia | @flograttarola | ecoevo.social/@flograttarola