



Spatial distribution of the greater rhea, *Rhea americana* (Linnaeus, 1758), in Rio Grande do Sul, southern Brazil: citizen-science data, probabilistic mapping, and comparison with expert knowledge

André L. Luza^{1,2} · Arthur V. Rodrigues^{1,3} · Louisa Mamalis^{1,4} · Viviane Zulian^{1,5}

Received: 3 March 2023 / Revised: 31 July 2023 / Accepted: 5 August 2023 / Published online: 21 August 2023
© The Author(s), under exclusive licence to Sociedade Brasileira de Ornitologia 2023

Abstract

The popularization of citizen-science platforms has increased the amount of data available in a fine spatial and temporal resolution, which can be used to fill distribution knowledge gaps through probabilistic maps. In this study, we gathered expert-based information and used species distribution models to produce two independent maps of the greater rhea (*Rhea americana*, Rheiformes, Rheidae) distribution in the state of Rio Grande do Sul, Brazil. We integrated municipality level detection/non-detection data from five citizen-science datasets into a Bayesian site occupancy model, accounting for false negatives, sampling effort, habitat covariates, and spatial autocorrelation. We addressed whether habitat (grassland and crop field cover, number of rural properties) and spatial autocorrelation explains the realized occurrence of the species and compared model-based and expert-based occurrence maps. The mean estimated percentage of occupied municipalities was 48% (239 out of 497 municipalities), whereas experts declared 21% of the municipalities (103) as occupied by the species. While both mapping approaches showed greater rhea presence in most municipalities of the Pampa biome, they disagreed in the majority of the municipalities in the Atlantic Forest, where more fieldwork must be undertaken. The greater rhea distribution was exclusively explained by the spatial autocorrelation component, suggesting that the species expanded its distribution towards the north of the state, reaching the Atlantic Forest, following deforestation and agriculture expansion.

Keywords Area of occupancy · Bayesian models · Hierarchical modeling · Imperfect detection · Species distribution modeling

Communicated by Marcos P. D. Santos (Associate Editor)

André L. Luza has first authorship

✉ André L. Luza
luza.andre@gmail.com

✉ Viviane Zulian
zulian.vi@gmail.com

¹ Departamento de Ecologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

² Universidade Federal de Santa Maria, Santa Maria, RS, Brazil

³ Research Centre for Ecological Change, Organismal and Evolutionary Biology Research Programme, Helsinki Institute for Life Sciences, University of Helsinki, Helsinki, Finland

⁴ Leverhulme Centre for Anthropocene Biodiversity, University of York, York, England

⁵ Department of Applied Ecology, North Carolina State University, Raleigh, NC, USA

Introduction

Conservation and management planning for a species requires accurate knowledge of its distribution. Historically, a key source of species distributional data has been expert knowledge, which defines species distribution boundaries based on where they are likely to occur (Hurlbert and Jetz 2007; Mainali et al. 2020). Expert-based range maps are generally drawn by minimizing the probability of false negatives: they often avoid declaring absence in truly occupied regions (Mainali et al. 2020). However, the resulting maps lack fine-scale resolution, and they often encompass environmentally unsuitable and unoccupied sites, skewing the actual distribution (Hurlbert and Jetz 2007; Mainali et al. 2020). Experts also may have limited ability to extrapolate their knowledge beyond their region of expertise as they are unable to study other regions (Murray et al. 2009; Hochachka et al. 2012). These factors highlight the need to include

additional sources of data for species distribution mapping.

Research in ecology benefits from people's contribution and engagement, which can contribute towards nature conservation programs (Hochachka et al. 2012). Citizen-science initiatives can contribute to scientific efforts by providing large datasets useful for mapping distributions (Kéry et al. 2013; Zulian et al. 2021) and planning population management and conservation (Stauffer et al. 2018). Information gathered in these initiatives are commonly aggregated into Species Distribution Models (SDMs) (Fletcher et al. 2019) which relate the species observations to relevant environmental and geographic variables (Guillera-Aroita et al. 2015; Kéry 2011; Kéry et al. 2013). They provide an alternative to expert-based range maps as they consider the fine scale nuances of species distributions, as well as combining spatially extensive data from several sources to strengthen inference on species geographical distribution and environmental effects (Altwegg and Nichols 2018; Fletcher et al. 2019; Miller et al. 2019). However, inference can be biased when simply aggregating datasets because many citizen-science data sources are not standardized in terms of location, protocol, and sampling effort (Fletcher et al. 2019; Miller et al. 2019).

Data integration, rather than data aggregation, represents a recent methodological advance on the use of SDMs (Fletcher et al. 2019; Miller et al. 2019). Data integration enables the inclusion and analysis of different, non-standardized datasets while accounting for processes that may influence the precision of the distribution estimates, such as the residual spatial autocorrelation and detection probability (Pacifci et al. 2017; Guélat and Kéry 2018; Zulian et al. 2021). Indeed, SDMs accommodating spatial autocorrelation and imperfect detection often have superior performance and precision to estimate distribution for a focal site and sites lacking data than other models (Guélat and Kéry 2018). In short, spatial autocorrelation refers to similar environment and species occurrence probability between geographically close sites (Legendre 1993; Dormann et al. 2007) and imperfect detection to false negatives being recorded in truly occupied sites (Mackenzie et al. 2002).

Globally, there is an urgent need for effective conservation and management of large-bodied species for two main reasons: (i) they are relatively more prone to extinction than many smaller species (Cooke et al. 2020), and (ii) they generally have large home range sizes whereby its conservation might benefit other species' populations (Walpole and Leader-Williams 2002). The greater rhea (*Rhea americana* Linnaeus, 1758 (Rheiformes, Rheidae)) is a large and flightless bird, which is highly agile and has a large home range size (Sick 2001). Although the greater rhea can be seen foraging in crop fields and artificial pastures with domestic cattle, sheep, and horses, preferred habitats

are grasslands, shrublands, and savannas (Sick 2001; Bellis et al. 2008; Giordano et al. 2010). Even with a geographic range that extends from the northeast to southern Brazil, eastern Bolivia, Paraguay, Uruguay, and part of Argentina, the greater rhea populations are presenting declines throughout its distribution due to hunting (for meat, feathers, skin, and eggs) and habitat loss, having a global conservation status of "Near Threatened" (BirdLife International 2016). Also, its preferred habitats are globally threatened by conversion to croplands and tree plantations (Hoekstra et al. 2005; Veldman et al. 2015). Locally, in the Brazilian state of Rio Grande do Sul (hereafter referred as RS), more than 50% of its vegetation cover originally comprised grasslands, especially within the limits of the Pampa biome (Overbeck et al. 2007; Cordeiro and Hasenack 2009; Hofmann et al. 2018). However, today only half of this vegetation cover remains (Souza Jr et al. 2020) due to high rates of grassland fragmentation and loss (Cordeiro and Hasenack 2009; Vélez-Martin et al. 2015) and insufficient and inefficient conservation measures (Overbeck et al. 2007; Andrade et al. 2019).

In this study, we elicited expert knowledge, and created SDMs with data integration, to produce independent maps of the greater rhea distribution in the 497 municipalities of RS state, Brazil. Our objectives were threefold: (1) to compare different models to define whether habitat and/or spatial autocorrelation explains the realized occurrence of the greater rhea in RS; (2) to compare the map of realized occurrence generated by the models with the map generated by experts; and (3) to compare the disparity between the realized occurrence and expert knowledge between the biomes found in RS (Pampa and Atlantic Forest), in order to find whether most of the species' preferred habitat is inside the Pampa biome. The Pampa biome is a region of flat land with large extents of open grasslands (Boldrini 2009) where an observant animal, attentive to its surroundings, such as the greater rhea, can thrive (Sick 2001; Bellis et al. 2008). Given the species' dispersal ability, conspecific attraction, and habitat preference for grasslands and ability to use crop fields, we expected a strong spatial autocorrelation component on occurrence, as well as a positive effect of grasslands and crop fields and a negative effect of the number of rural properties on greater rhea occurrence in RS municipalities. Nonetheless, we compared this hypothesis with alternative scenarios comprising different combinations of covariates. Another hypothesis was that most disparities regarding the species occurrence generated by the expert knowledge and SDMs would occur inside the expert-drawn maps, as it used to have areas not suitable for the species (Mainali et al. 2020).

Methods

We produced maps of the distribution of the greater rhea in the RS state using two procedures: (1) expert knowledge and (2) site-occupancy models based on citizen-science data.

Since some citizen-science data sources provide only the municipality as the most local unit of species record, we used the boundaries of the 497 state municipalities as discrete sites of analysis.

Study area

We studied the distribution of the greater rhea in the RS, the southernmost and the ninth largest state of Brazil, with an area of 281,707 km² (IBGE 2020). Grasslands and grassland-forest mosaics originally covered most of the southern, southwestern, and northeastern regions of the state (Boldrini 2009). In these grasslands, the dominant land use was livestock, traditionally raised by European and Portuguese colonizers (Ribeiro and de Quadros 2015). A considerable portion of these grasslands were converted to crop fields and tree plantations during the last few decades (Vélez-Martin et al. 2015). Forests dominated in the northern region of the state (along the Uruguay river) and the southern limits of the Atlantic Rainforest (central and eastern RS) (Boldrini 2009), whose conversion to agriculture was more pronounced in the beginning of the twentieth century. Currently, the main land use in the state is crop fields, mainly soybean, rice, and maize (Cordeiro and Hasenack 2009; Hofmann et al. 2018) forming the mosaic of habitats potentially suitable for the Greater Rhea.

Expert-drawn maps

We defined an “expert” as a researcher and professional with sound fieldwork experience, with involvement in bird-watching initiatives, and who has supervised and published research regarding the bird species of RS. We gathered expert knowledge through an application created using the R package *shiny* (Chang et al. 2021). In this application, each expert was presented with two drawing steps, each with different mapped features; these were (1) a map which had municipality boundaries and labels, and (2) a map with the municipality boundaries, labels, and grassland area cover per municipality. The different mapped features were used to improve expert precision when mapping the distribution of the greater rhea by providing more information about the presence of suitable habitat for the species. As polygons were similar between mapped features, we used the polygons drawn over the first map (Supplementary Material Fig. S1.1). We requested information from 24 experts by email between November 2020 and February 2021. We received polygons from eight experts and used them to calculate the proportion of experts declaring the greater rhea in each municipality (hereafter “expert-based occurrence”).

SDM data

We fitted species distribution models, namely site-occupancy models with data integration, to five datasets comprising the detection and non-detection of the greater rhea and associated sampling effort per site (municipality) across several visits (sampling occasions) within a site. The five citizen-science datasets used were eBird (<http://www.ebird.org>), WikiAves (<http://www.wikiaves.com.br>), GBIF (Global Biodiversity Information Facility, <http://www.gbif.org>), VertNet (<http://www.vertnet.org>), and iNaturalist (<http://www.inaturalist.org>). The first two are exclusive for bird sightings. eBird is global and includes data as checklists—where the observer uploads a list of all the species detected in a certain georeferenced location (Sullivan et al. 2009). In contrast, WikiAves is a Brazilian initiative that has photos and sound recordings of birds in an individual detection format, at a municipality level (WikiAves 2019). We gathered data from eBird directly from its website (eBird 2019), and used both complete and incomplete lists because (i) removing incomplete lists could result in the loss of valuable data; (ii) our model took into account sampling effort, so that incomplete lists represent less effort to detect the greater rhea; and (iii) incomplete lists represent a small portion of the eBird dataset ($n=80$ of 5492 lists). We gathered WikiAves data directly from its administrator (WikiAves 2019). GBIF, VertNet, and iNaturalist are global platforms of open data, which provide georeferenced detection for a wider range of taxa. We downloaded bird data (searching for the term “Aves” or “Rio Grande do Sul birds”) from these platforms directly on R through the packages *rgbif*, *spocc*, and *rvertnet*. We used functions of the *CoordinateCleaner* package (Zizka et al. 2019) to clean misplaced coordinates (e.g., centroids, capitals, duplicates, institutions, sea, urban areas, and zero coordinates). All data were temporally bound between January 1, 2008, and December 31, 2018.

Because of its checklist format, eBird is the only dataset that closely follows the classical sampling protocol of replicated visits to sites, optimal data for site-occupancy modeling (see MacKenzie et al. 2002; Mackenzie and Royle 2005). In this case, each list was a visit to a certain municipality and received a detection (1) if the greater rhea was one of the species on the list, and a non-detection (0) if not. The number of lists per municipality varied significantly. Municipalities with protected areas or with certain habitat features that make them “birding hotspots,” such as Lagoa do Peixe National Park (municipalities of Tavares, Mostardas, and São José do Norte) which attracts many migrant birds (and birdwatchers) than other municipalities. This resulted in large differences in the number of checklists per municipality (from 0 to 623 lists, Supplementary Material Fig. S1.2). To minimize processing time and avoid issues regarding missing data and parameter convergence when

estimating effort parameters for the eBird model, we used data from the first 11 checklists per municipality, which was the average number of lists found across all 497 municipalities (Supplementary Material Fig. S1.2). We used the total number of observers of the checklist, the duration, and the distance traveled as effort covariates for eBird data.

WikiAves, GBIF, VertNet, and iNaturalist datasets had no direct measurement of sampling effort associated with the municipalities. The effort associated with GBIF, VertNet, and iNaturalist datasets was the total number of any bird species recorded at the municipality. If the greater rhea was among the detected species, the municipality received a detection (1); otherwise, it received a non-detection (0). The effort associated with the WikiAves dataset was the total number of photos and song recordings per municipality, with any photo or call recording of the greater rhea indicating a detection. All these covariates are reasonable proxies for effort in bird sampling (Miller et al. 2019) (for more details, see Zulian et al. 2021). Even though eBird and GBIF had a similar spatial distribution of detections, each dataset brought different metadata characterizing sampling events, such as the number of observers, sighting duration, and distance traveled, that were useful in the modeling process and contributed to the detection probability estimation.

Site covariates

We used the municipality-level area of grasslands, the area of crop fields, and the number of properties relative to the municipality area as site covariates. We measured the area of grasslands and crop fields per municipality using data from Hofmann et al. (2018). To obtain grassland area per municipality, we gathered the total cover of dry, humid, and secondary (regenerating) grasslands within the state, and then divided it by the municipality area. To obtain the crop fields' area, we first summed the cover of soybean, corn, and wheat and then divided it by each municipality area (Hofmann et al. 2018). To have a proxy on how fragmented the greater rhea habitat is, we gathered data on the number of properties per municipality using data from the Brazilian Census of Agriculture (IBGE 2010) and divided it by municipality area to obtain the number of properties relative to municipality area. All spatial objects were converted into Lambert projection before data extraction and spatial analysis to ensure a proper calculation of municipality and habitat area. The three site covariates were transformed (square root) and then standardized by the mean and standard deviation before the modeling.

Site-level covariate maps showed that grasslands are predominantly concentrated in the southern region of RS, whereas the crop fields are distributed throughout the state

(Supplementary Material Fig. S1.3). The municipalities with the highest proportion of rural properties relative to municipality area are concentrated in the north, northwestern, and central regions of RS (Supplementary Material Fig. S1.3).

Statistical analyses

Species distribution modeling

We used hierarchical species distribution models with data integration of standardized (eBird) and non-standardized (other) citizen-science datasets to estimate the greater rhea distribution (Miller et al. 2019). For the biological part of the model, we denote the true occupancy state z of each municipality i , by:

$$z_i \sim \text{Bernoulli}(\psi_i), \quad (1)$$

where the state of this latent variable z is partially observed and follows a Bernoulli distribution, with mean ψ_i . We denote here z_i as the realized state: 1 when occupied and 0 when not occupied. The z_i is estimated by choosing values of parameters that maximize the likelihood of finding the observed data, per citizen-science dataset (Miller et al. 2019).

The site occupancy probability ψ_i is estimated by considering the influence of ecological processes on the probability of site occupancy (MacKenzie et al. 2002), as follows:

$$\text{logit}(\psi_i) = \beta_0 + \beta_1 * X_1 + \dots + \beta_n * X_n + \delta_i. \quad (2)$$

The ψ_i has its own GLM, with β_0 being the GLM intercept (the average site occupancy probability), $\beta_1 - \beta_n$ the regression coefficients, $X_1 - X_n$ the values of each environmental covariate (see the following "Site covariates"), and δ_i being the conditional autoregressive (CAR) random effect of a focal site i :

$$\delta_i \sim \text{CARnormal} \left(\frac{\sum_{j=1}^m \delta_j}{m}, \frac{\tau^2}{m} \right), \quad (3)$$

where δ_j are the values of CAR random effect for the neighbors of the focal site, and τ^2 is the spatial variance parameter (Besag et al. 1991; Guélat and Kéry 2018). We built a discrete spatial neighborhood of hexagonal cells with either 10, 25, or 50 km, to test for the optimal cell size for neighborhood consideration.

For the observational part of the model, we analyzed detection error by building one model per citizen-science dataset (before integrating them to estimate the z_i) and estimating the detection probability. The general model had the following format:

$$p_j^* = 1 - (1 - p)^{E_j}, \tag{4}$$

where the probability of detecting the greater rhea p_j^* in an occasion j was an exponential function of the sampling effort E_j in each sampling occasion j (Pacifci et al. 2017; Stauffer et al. 2018; Zulian et al. 2021). The sampling effort is estimated from data using a linear model without the intercept, so that E_j is in units of sampling effort. The parameter p was fixed to 0.5, as it is not identifiable according to available data. In this exponential function, p_j^* increases non-linearly with increasing sampling effort up to an asymptote (Stauffer et al. 2018). The E_j is estimated separately for each dataset DS_n —with n varying from 1 to 5: eBird, WikiAves, iNaturalist, GBIF, and VertNet—based on the effort covariates from each dataset ($X1_j$, $X2_j$, and $X3_j$), which varied across the j visits, following a linear equation without intercept:

$$E_j^{DS_n} = \alpha_1 * X1_j + \alpha_2 * X2_j + \alpha_3 * X3_j, \tag{5}$$

Finally, we integrated the biological and observational parts of the data integration model multiplying z_i estimated in Eq. 1 by the p_j^* from Eq. 4. The true state of site occupancy, z_i , was then updated in each multiplication with p_j^* . The observation dataset Y_{ij} became a latent variable from a Bernoulli distribution that contained the parameter $z_i p_j^*$:

$$Y_{ij} \sim \text{Bernoulli} \left(z_i \times p_j^* \right). \tag{6}$$

The parameter $z_i p_j^*$ contained the same information—detection conditional to site occupancy state—which is typical for site-occupancy models (MacKenzie et al. 2002): the species is only observed in a site i , for a sampling occasion j ($y = 1$), if the site is truly occupied by the species, and the species is effectively detected by the observer.

We estimated the parameters of the data integration models using Bayesian inference through which it was possible to update the prior knowledge regarding the model parameters using data to form “posterior” estimates of each model parameter (Ellison 2004). We wrote our models in BUGS language and ran them in WinBUGS through the R package *R2WinBUGS*. We ran the model in three independent chains of 60,000 iterations, burn-in phase of 40,000 iterations, adaptive phase of 30,000 iterations, and thinning each 20 samples, yielding a posterior sample of 3000 iterations. We obtained a value of convergence criterion R-hat lower than 1.1 for all the parameters. All analyses were run in the R programming environment (R Core Team 2021).

Model selection analysis

We built 12 different models depicting different alternative hypotheses to explain the spatial distribution of the greater rhea in RS (Table 1). We selected the best model using the model deviance criterion (D , see the following Eq. 8), for which the lowest value represents the best model. The deviance of each dataset was calculated based on two different sets of data: (i) training and (ii) validation datasets. We used 20% of the data from each dataset (~100 municipalities) for validation and used the remaining 80% for training the model and estimating the parameters. Then, the training and validation datasets were compared, in terms of the squared number of detections, producing a likelihood function for each citizen-science dataset n calculated as:

$$L = y^{\hat{y}} (1 - \hat{y})^{1-y}, \tag{7}$$

where y is the validation dataset, and the \hat{y} dataset was estimated using the training dataset (i.e., derived from the model) (Hooten and Hobbs 2015). Deviance across all

Table 1 Model selection analysis, showing models ranked according to the Deviance Criterion. Models 4, 7, 8, and 9, designated as “Complete,” include all the site covariates and account for different spatial structure. Spatial models (1, 2, and 5) account for different spatial structures and do not include the site covariates. Grassland models (6, 10, 11, and 12) only include this site covariate. The Null model does not include site covariates and spatial structure

Model	Spatial structure	Total deviance	Finite sample size (FSS)	95% credible interval		\hat{R}	
				Lower	Upper		
1	Spatial	50 km	297.67	0.48	0.41	0.56	1.00
2	Spatial	10 km	297.98	0.47	0.40	0.55	1.00
3	Null	None	299.09	0.47	0.40	0.54	1.00
4	Complete	None	300.06	0.47	0.41	0.54	1.00
5	Spatial	25 km	300.67	0.48	0.41	0.56	1.00
6	Grassland	25 km	302.08	0.49	0.42	0.57	1.00
7	Complete	10 km	302.17	0.48	0.41	0.55	1.00
8	Complete	50 km	302.27	0.48	0.41	0.56	1.00
9	Complete	25 km	304.17	0.50	0.42	0.57	1.00
10	Grassland	50 km	312.08	0.48	0.41	0.56	1.00
11	Grassland	None	313.71	0.47	0.40	0.54	1.00
12	Grassland	10 km	315.05	0.47	0.41	0.54	1.00

datasets was calculated as the sum of the likelihood across n citizen-science datasets as:

$$D = 2 - \sum_{n=1}^N \log(L). \quad (8)$$

Spatial distribution

We used the z_i estimated by the best-ranked model to obtain the number of municipalities where the realized occurrence of the greater rhea was equal to 1. This measure is also known as “Finite Sample Size” or FSS and it was calculated using the equation:

$$FSS = \frac{\sum_{i=1}^I z_i}{N}, \quad (9)$$

where z_i is the realized occurrence in each municipality i , and N is the total number of sites ($N=497$ municipalities) (MacKenzie et al. 2006; Royle and Kéry 2007). The greater rhea distribution based on the expert knowledge was obtained by counting the number of municipalities within the expert-drawn polygons where they considered the species to be present.

Comparison of expert knowledge and SDM

To compare the differences between the expert-drawn and the SDM map, we overlaid the expert-based occurrence map with the map of z_i estimated by the best-ranked SDM. Values equal to 0 indicated municipalities where experts and SDM results were the same. Positive values indicate municipalities with greater rhea presence according to experts, whereas negative values indicate municipalities with greater rhea presence according to the SDM. To depict the disparity between expert knowledge and SDM, we counted the number of municipalities in which the difference was higher, lower, or equal to 0. We therefore considered the differences between the number of municipalities with greater rhea occurrence declared only by the experts, only by SDM, and by both. Finally, we used linear models (LMs) to test whether expert-based and SDM-based occurrences were related, and whether the average disparity between experts and the SDM was higher inside than outside the expert-drawn map (expert-based occurrence close to 1 and 0, respectively). In these LMs, the expert-based occurrence was considered as the predictor variable, and the absolute difference between expert- and SDM-based occurrence (i.e., both with a value between 0 and 1) as a response variable. The model parameters were estimated through Bayesian inference, and models were run using the function *brm*, from the R package *brms* (Bürkner 2017) and the Stan software. We used flat priors in four parallel Hamilton Monte-Carlo

Markov chains comprising 2000 iterations each, burn-in (warmup) of 1000 iterations and thinning each iteration, producing 4000 samples of the posterior probability distribution of each model parameter. Average parameter values and associated credible intervals were calculated using these 4000 posterior-distribution samples.

Variation between biomes and mapping approach

To investigate the differences between expert-drawn maps and estimated maps from the SDM per biome, we overlaid expert and SDM maps with the shapefile of the Pampa biome. Municipalities outside the Pampa were categorized as Atlantic Forest. We extracted the values of z_i and expert knowledge ranks (1s or 0s) per biome to check variation of greater rhea occurrence relative to the biome. We inspected this variation using a biplot showing the relationship between SDM-based occurrence z_i and the expert-based occurrence. We further tested for differences in average realized occurrence between biomes using ANOVA with biome, data (expert knowledge or SDM), and the interaction between biome and mapping approach as predictor factors, and the realized occurrence (collated z_i from SDM and 1s and 0s from experts) as response variables. The ANOVA was run using the same settings as the LMs. Post hoc test was done using the *emmeans* function, from the *emmeans* R package (Lenth 2022), in which the marginal median of linear trends (with the associated 95% high probability density (HPD) interval) was estimated to identify the direct effect of the biome: mapping approach interaction on each metric while maintaining constant the isolated effect of biome and approach.

Results

The number of municipalities with samples and greater rhea detections varied among the citizen-science datasets (Fig. 1). WikiAves had the largest spatial coverage and the highest number of detections among the five datasets, accounting for 466 municipalities with samples and 99 with greater rhea detections. The second largest spatial coverage was from GBIF, with 246 municipalities with samples and greater rhea detections in 35 of them. Although eBird had a smaller spatial coverage than GBIF, it had a higher number of municipalities with greater rhea detections (36 municipalities). Finally, VertNet and iNaturalist had the smallest spatial coverage and number of municipalities with detections. VertNet had 91 municipalities with samples and two with greater rhea detections, whereas iNaturalist had 73 municipalities with samples and seven with greater rhea detections (Fig. 1).

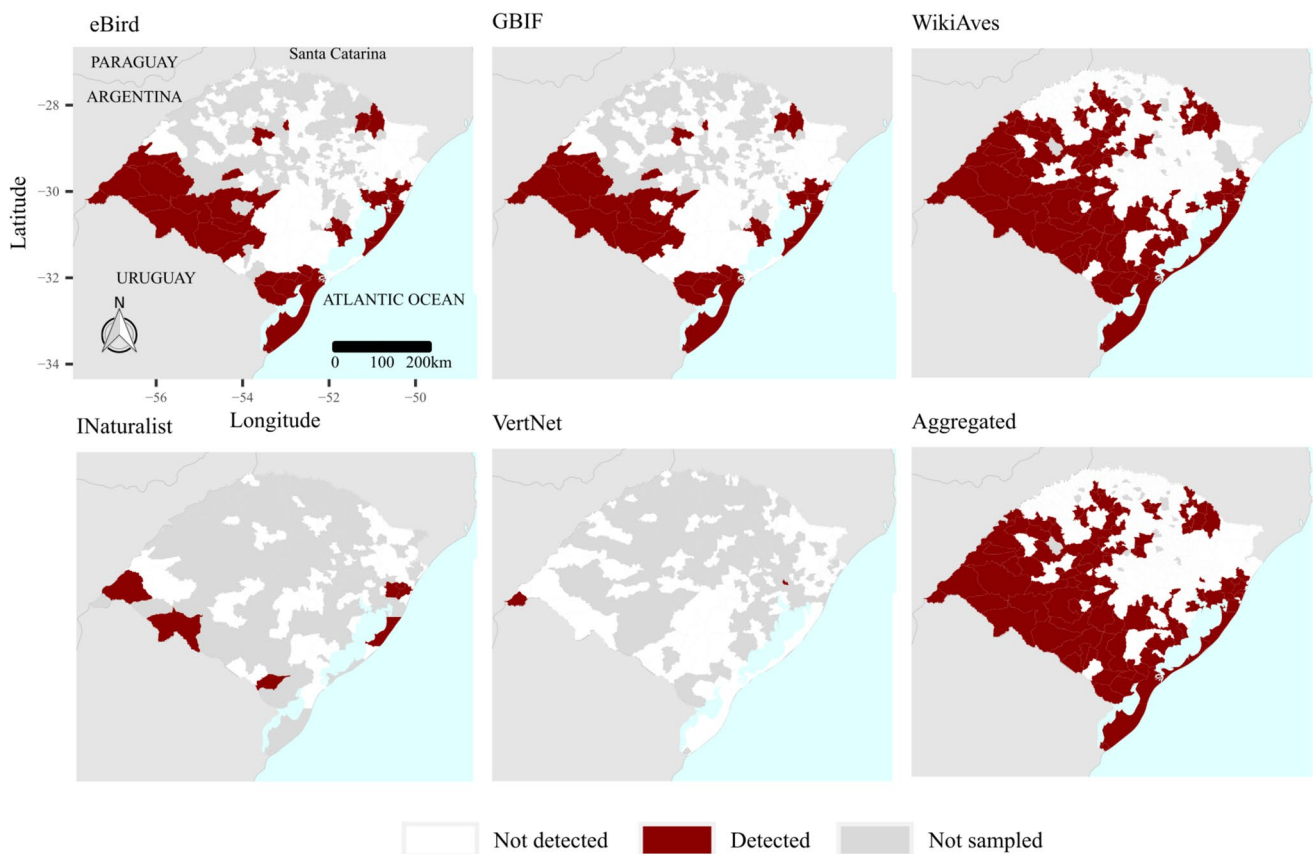


Fig 1 Spatial distribution of sampling and detections of the greater rhea in the municipalities of the RS, Brazil, from the eBird, GBIF, WikiAves, iNaturalist, and VertNet datasets. The greater rhea was detected in the municipalities in red and not detected in municipali-

ties in white. Gray municipalities had no sampling. The bottom-right map shows the aggregated data across the five datasets. The figure was produced in R Programming Environment and edited in Inkscape

Model selection analysis

Model selection analysis based on the deviance criterion showed that the spatial model with a neighborhood cell size of 50 km (model 1, Table 1) had the best predictive performance among the twelve models tested. The spatial model with a neighborhood of 10 km (model 2) was the second best-ranked model and had a slightly worse predictive ability than the first model. The complete models (with all space and site covariate options: models 4, 7, 8, and 9), the models accounting only for site covariates (6, 10, 11, and 12), and the model accounting only for space with a neighborhood of 25 km (model 5) performed slightly worse than the Null model (model 3, Table 1).

Spatial distribution

The estimates of the finite sample size (FSS) were very similar across all models, ranging from 0.47 to 0.50. The mean FSS of the best ranked model was 0.48, with the 95% credible interval (CI) ranging from 0.41 to 0.56 (Table 1), which

translates into a realized occurrence in 239 municipalities (95% CI: 204–278). From now on, we report only the results from the best-ranked model (model 1).

Estimated occurrence based on the integration of the five datasets showed a greater rhea distribution range that includes predominantly the Pampa biome area, comprising municipalities from the southern, western, and coastal regions of RS (Fig. 2a). The estimated occurrence also includes some municipalities within the northeastern and middle plateau of RS. The northern portion of the state had the highest uncertainty in the estimated occurrence, with most of the municipalities showing an average z estimate around 0.5 (Fig. 2a).

Based on eight maps received from experts, the polygon drawn indicates greater rhea presence in most municipalities that are inside the Pampa biome, for example in Santana do Livramento and Alegrete (dark-blue areas in western Pampa), in a portion of the northern municipalities, and in two isolated municipalities in the northeast of RS (Fig. 2b). The sum of expert-based occurrences resulted in 103.13 municipalities (21% out of 497 municipalities) with the presence of the species.

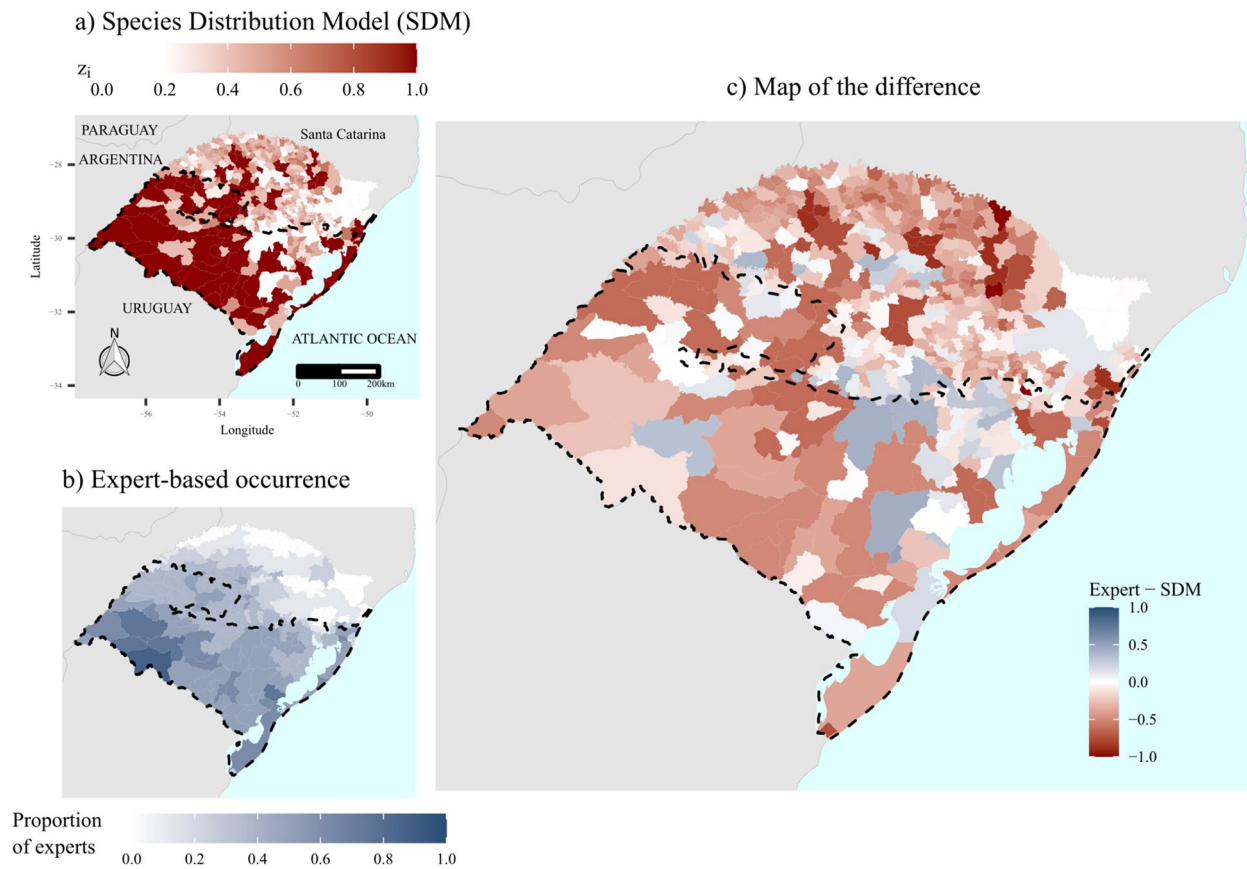


Fig. 2 Geographic distribution of the greater rhea in RS, Brazil. The dashed black line depicts the limits of the Pampa biome. **a** The realized occurrence (average of z_i across the 3000 posterior-distribution samples) estimated from the data integration model, with darker tones representing higher mean z_i and intermediate values representing the municipalities with the highest uncertainty regarding the occurrence. **b** The expert-drawn map, with dark-blue municipalities representing municipalities with a high proportion of experts declaring greater

rhea presence. **c** The disparity between the realized occurrence and the expert knowledge. In **c**, blue colors represent greater rhea occurrence only from the expert drawn map, red colors represent occurrence only from the SDM results, and the white colors represent agreement between the expert drawn map and the SDM result. The figure was produced in R Programming Environment and edited in Inkscape

Comparison of expert knowledge and species distribution modeling (SDM)

Linear models showed a positive relationship between expert-based and SDM-based occurrence (average regression coefficient (β) and 95% credible interval (CI) of β : 0.59; CI: 0.46, 0.72) (Fig. 3a), and a negative, albeit weak, relationship between expert-based occurrence and the absolute difference in expert- and SDM-based occurrence (β : -0.13; CI: -0.23, -0.03) (Fig. 3b). The map of the difference between the expert- and SDM-based occurrence shows that most agreement between the SDM and the expert-drawn map lies within the Pampa biome (Fig. 2c). Most of the municipalities in this region returned a difference close to 0, highlighting that the greater rhea was recorded as present in the region by both the experts and the SDM. While the results from the SDM and from the experts agreed in only four municipalities (in which the difference between expert

and SDM was exactly 0), they disagreed in the majority of the municipalities. Experts recorded the species' occurrence in 78 municipalities (i.e., difference between expert and SDM > 0), and SDM results in 415 municipalities (i.e., difference between expert and SDM < 0).

Variation between biomes and mapping approach

ANOVA with post hoc test showed differences on greater rhea realized occurrence among the levels of the interaction between biome (Pampa or Atlantic Forest) and mapping approaches (SDM or expert aggregated map) (Fig. 3c). The effect of this interaction shows that (1) there was a higher realized occurrence in the municipalities within the Pampa, and (2) experts estimated a lower chance of occurrence in the Atlantic Forest and Pampa than the SDM (Fig. 3c). Experts estimated 13% (lower and higher 95% HPD: 10–15%) of the Atlantic Forest municipalities as occupied by the greater

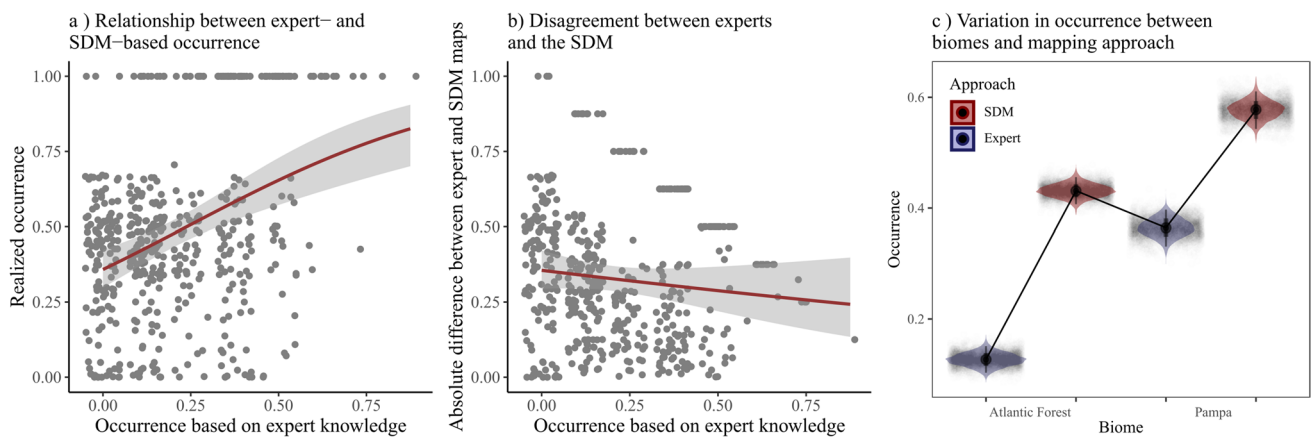


Fig. 3 Plots showing the relationship between expert- and SDM-based occurrence (a), the disagreement between these mapping approaches (b), and the variation in realized occurrence between biomes and mapping approaches (c). Points were subtly jittered in

a and b to improve visualization. Small points in c show the 3000 posterior distribution samples of a Bayesian ANOVA, and the black points depict the median across these samples. The figure was produced in R Programming Environment and edited in Inkscape

rhea, and 37% (HPD: 33–40%) of the Pampa municipalities. The SDM, in its turn, estimated 43% (HPD: 41–46%) of the Atlantic Forest and 58% (HPD: 54–61%) of the Pampa municipalities as occupied by the greater rhea (Fig. 3c).

Discussion

Both expert knowledge and SDM show that the greater rhea is a bird with affinities to open landscapes, and that there is a considerable overlap of its distribution with the area of the Pampa biome. However, expert knowledge and SDMs showed contrasting patterns within the Atlantic Forest biome where the species' occurrence might be more uncertain. We initially expected that most disparity between experts and SDM would occur inside the range depicted by the experts (e.g., Mainali et al. 2020). Our analysis of absolute difference did not support this hypothesis, showing that difference between SDM- and expert-based occurrence was lower in municipalities with occurrence declared by the experts. This result indicates that the experts are aware of the factors determining the spatial distribution of the greater rhea, at least within the grassy physiognomies of the Pampa biome. Regarding the effect of site-level covariates, we found a strong spatial autocorrelation component influencing the spatial distribution of the greater rhea in RS. While a spatial autocorrelation effect was initially expected, we did not expect it to be the only factor influencing the greater rhea distribution in the studied area. We attribute this large spatial effect to the biology of the greater rhea, mainly due to its vagility, large home range size, and strong conspecific attraction during the reproduction period. Finally, the estimated realized occurrence, which corresponds to the area of occupancy (*sensu* IUCN 2001), includes 48% of the state

municipalities, and showed the locations more suitable to greater rhea occurrence in southernmost RS.

Both expert knowledge and SDMs aim to present binary state of species distributions (either occupied or unoccupied) but using different methods (Hurlbert and Jetz 2007; Mainali et al. 2020). We found that the greater rhea distribution generated by the experts and by the SDM agreed on an important point: both resulted in a realized area of occupancy in almost the complete extent of the Pampa biome. The Pampa biome is a region of flat land originally covered by large extents of open grassland (Boldrini 2009). The greater rhea is well adapted for such a habitat with its long neck and legs to escape predators, allowing it to thrive (Bellis et al. 2004). Its long neck enables it to see over large distances and remain vigilant to predators. It is described as a cursorial bird, so it is well adapted for running long distances over open grasslands and can reach speeds of up to 60 km/h (Abourachid and Renous 2000). It has also been described as having a large home range within natural grasslands (11 km²) driven by the availability of its main food source, wild dicots, which tend to be found in low densities within grasslands meaning the birds often have to travel further afield to feed (Bellis et al. 2004).

Nonetheless, the greater rhea distribution maps from the SDM and the expert knowledge differ in the following aspects: (1) the SDM estimated a broader occurrence in the municipalities of highland grasslands (northeastern RS), medium plateau, and close to the Uruguay River, and (2) experts declared occurrence in the southeast and central RS mountain ranges, and in the west portion of the Uruguay river coast. Possible sources of disparity may have arisen due to (i) false negative results from experts in the Atlantic Forest; (ii) false positive results from experts in the Pampa, mainly along the limits between the Pampa and Atlantic

Forest; and (iii) an omission of important variables within the SDMs (see below). First, although the Atlantic Forest is predominantly covered by forests, a habitat not suitable for the species (Souza Jr et al. 2020), it is possible to find natural grassy landscapes and grassland-forest mosaics potentially suitable for the greater rhea in northeastern and middle plateau of RS (Boldrini 2009). Second, expert maps included unsuitable areas not occupied by the species in the eastern Pampa, which might include more forested areas found along the hills in the southernmost limits of the Atlantic Rainforest.

Third, we found a strong effect of the spatial component on greater rhea occurrence. Spatial autocorrelation can be introduced by two different factors (that unfortunately cannot be distinguished based on detection/non-detection data), namely omission of important site covariates, and biotic processes such as dispersal and conspecific attraction (Legendre 1993; Dormann et al. 2007). Important covariates that might vary in space and shape species distribution can be the expansion of agriculture towards northern RS, where large extents of open habitats (crop fields) were created and are used by the greater rhea. Perhaps the species is expanding its range within RS alongside the agriculture expansion, an effect that cannot be detected with static site-occupancy models but was already noticed in RS and in other Brazilian states (Miranda 2021). Also, the species move across large geographic areas when searching for dense and tall grasslands to reproduce (Erize and Villafañe 2017). Such spatiotemporal dynamics are not explained by habitat covariates but might be represented by the spatial autocorrelation component. Therefore, although habitat covariates may not be necessary to explain the geographic distribution of the greater rhea in RS, it is essential to monitor its populations and track changes in its preferred habitat to accurately identify and analyze population dynamics over time.

The conservation of Pampean grasslands can be greatly aided by the definition of a flagship species. This type of species can help increase interest from citizens and governments to support and fund conservation efforts (Simberloff 1998; Walpole and Leader-Williams 2002). When these species are protected, this can in turn ensure the protection of the wider landscape and other species that rely on it (Walpole and Leader-Williams 2002). The greater rhea is an unusual bird due to its morphology and large size, its large home range, and interesting breeding behavior. The species is also classified globally as “Near Threatened” (BirdLife International 2016), predominantly due to the loss of the grassland habitat, illegal hunting (for their feathers, meat, and skin), egg collection, and predation by felids and feral dogs (Sales 2006). Being a species of conservation importance, the greater rhea could be considered as an effective flagship species for the conservation of the Pampa habitat within Brazil.

Using expert knowledge and data integration models based on citizen-science data, which produced maps showing the area of greater rhea occupancy, we highlighted that (i) regions in which both SDM and experts agreed about the greater rhea distribution can be useful and form the core areas for future conservation efforts, and can be targeted by studies considering parameters such as population size and survival; and (ii) regions where the SDM and expert knowledge differed should be the focus of further sampling effort. Acknowledging the possibility of false negatives in observations (MacKenzie et al. 2002; Kéry et al. 2013; Guillera-Arroita et al. 2015; Guélat and Kéry 2018) is crucial for minimizing statistical uncertainty. This cannot be achieved with expert-based range maps or by the aggregation of different datasets. Therefore, the collaboration of experts, ecologists, and citizen scientists in mapping species distribution over space and time is encouraged to produce an informative source of evidence for planning species management and conservation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43388-023-00143-3>.

Acknowledgements We thank Gonçalo Ferraz (UFRGS) and Rafael Dias (UFPEL) for useful insights during the conception of this study. We thank all experts that kindly shared their knowledge with us and Reinaldo Guedes for the help getting the WikiAves dataset.

Author contribution Louisa Mamalis, Viviane Zulian, and André L. Luza conceived the idea of this study. Arthur V. Rodrigues wrote the shiny app to collect expert opinions. André L. Luza and Viviane Zulian curated the data and wrote the models. André L. Luza, Viviane Zulian, and Louisa Mamalis wrote the first version of this manuscript, and all the authors contributed critically to the writing and approved the final manuscript.

Funding A.L.L. received post-doctoral fellowships from the Brazilian National Council for Scientific and Technological Development (CNPq; #153024/2022-4, #164240/2021-7, #151228/2021-3, #152410/2020-1) and Coordination for the Improvement of Higher Education Personnel (CAPES; PDPG-POSDOC, #88887.800011/2022-00).

V.Z. research is funded by the Coordination for the Improvement of Higher Education Personnel (CAPES) and Sea Grant (USA).

A.V.R. research is funded by the Coordination for the Improvement of Higher Education Personnel (CAPES) - Finance Code 001 and Jane and Aatos Erkkö Foundation.

L.M. research is funded by European Commission through the program, Erasmus Mundus Master Course—International Masters in Applied Ecology (EMMC-IMAE) (FPA 2023 – 0224/532524-1-FR-2012-1-ERA MUNDUS-EMMC)

Data availability Data, shinyApp, and R codes supporting the results of this study are available at the GitHub (https://github.com/andreluza/project_Rhea.git) and Zenodo (<https://doi.org/10.5281/zenodo.8110726>).

Declarations

Competing interests The authors declare no competing interests.

References

- Abourachid A, Renous S (2000) Bipedal locomotion in ratites (Paleognathiform): examples of cursorial birds. *Ibis* 142:538–549. <https://doi.org/10.1111/j.1474-919X.2000.tb04455.x>
- Altwegg R, Nichols JD (2018) Occupancy models for citizen-science data. *Methods Ecol Evol* 10:8–21. <https://doi.org/10.1111/2041-210X.13090>
- Andrade BO, Bonilha CL, Overbeck GE, Vélez-Martin E, Rolim RG, Bordignon SAL, Schneider AA, Ely CV, Lucas DB, Garcia ÉN, Santos ED, Torchelsen FP, Vieira MS, Silva Filho PJS, Ferreira PMA, Trevisan R, Hollas R, Campestrini S, Pillar VD, Boldrini II (2019) Classification of South Brazilian grasslands: implications for conservation. *Appl Veg Sci* 22:168–184. <https://doi.org/10.1111/avsc.12413>
- Bellis LM, Martella MB, Navarro JL, Vignolo PE (2004) Home range of greater and lesser rhea in Argentina: relevance to conservation. *Biodivers Conserv* 13:2589–2598. <https://doi.org/10.1007/s10531-004-1086-0>
- Bellis LM, Pidgeon AM, Radeloff VC, St-Louis V, Navarro JL, Martella MB (2008) Modeling habitat suitability for greater rheas based on satellite image texture. *Ecol Appl* 18:1956–1966. <https://doi.org/10.1890/07-0243.1>
- Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. *Ann I Stat Math* 43:1–20. <https://doi.org/10.1007/BF00116466>
- BirdLife International (2016) *Rhea americana*. The IUCN Red List of Threatened Species 2016: e.T22678073A92754472. <https://doi.org/10.2305/IUCN.UK.2016-3.RLTS.T22678073A92754472.en>
- Boldrini II (2009) A flora dos campos do Rio Grande do Sul. In: Pillar VD, Muller SC, Castilhos ZMS, Jacques AVA (eds) *Campos sulinos: Conservação e uso sustentável da biodiversidade*. MMA, Brasília, Brazil, pp 63–77
- Bürkner P-C (2017) brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw* 80:1–28. <https://doi.org/10.18637/jss.v080.i01>
- Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2021) shiny: Web application framework for R. <https://cran.r-project.org/package=shiny>. Accessed 29 January 2018
- Cordeiro JL, Hasenack H (2009) Cobertura vegetal atual do Rio Grande do Sul. In: Pillar VD, Muller SC, Castilhos ZMS, Jacques AVA (eds) *Campos sulinos: Conservação e uso sustentável da biodiversidade*. MMA, Brasília, Brazil, pp 285–299
- Cooke RS, Eigenbrod F, Bates AE (2020) Ecological distinctiveness of birds and mammals at the global scale. *GECCO* 22:e00970. <https://doi.org/10.1016/j.gecco.2020.e00970>
- Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–628. <https://doi.org/10.1111/j.2007.0906-7590.05171.x>
- eBird (2019) eBird: an online database of bird distribution and abundance. <https://ebird.org/home>. Accessed 29 January 2019
- Ellison AM (2004) Bayesian inference in ecology. *Ecol Lett* 7:509–520. <https://doi.org/10.1111/j.1461-0248.2004.00603.x>
- Erize FG, Villafañe IEG (2017) Habitats selected by the endangered greater rhea (*Rhea americana*)—implications for conservation. *Emu* 116:379–386. <https://doi.org/10.1071/MU15124>
- Fletcher RJ, Hefley TJ, Robertson EP, Zuckerberg B, McCleery RA, Dorazio RM (2019) A practical guide for combining data to model species distributions. *Ecology* 100:e02710. <https://doi.org/10.1002/ecy.2710>
- Giordano PF, Navarro JL, Martella MB (2010) Building large-scale spatially explicit models to predict the distribution of suitable habitat patches for the greater rhea (*Rhea americana*), a near-threatened species. *Biol Conserv* 143:357–365. <https://doi.org/10.1016/j.biocon.2009.10.022>
- Guélat J, Kéry M (2018) Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods Ecol Evol* 9:1614–1625. <https://doi.org/10.1111/2041-210X.12983>
- 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: matching distribution models to applications. *Glob Ecol Biogeogr* 24:276–292. <https://doi.org/10.1111/geb.12268>
- Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong W-K, Kelling S (2012) Data-intensive science applied to broad-scale citizen science. *Trends Ecol Evol* 27:130–137. <https://doi.org/10.1016/j.tree.2011.11.006>
- Hoekstra JM, Boucher TM, Ricketts TH, Roberts C (2005) Confronting a biome crisis: global disparities of habitat loss and protection. *Ecol Lett* 8:23–29. <https://doi.org/10.1111/j.1461-0248.2004.00686.x>
- Hofmann GS, Weber EJ, Hasenack H (2018) Uso e cobertura vegetal do Estado do Rio Grande do Sul – situação em 2015. UFRGS, IB, Centro de Ecologia. <http://www.ufrgs.br/labgeo>. Accessed 29 January 2018
- Hooten MB, Hobbs NT (2015) A guide to Bayesian model selection for ecologists. *Ecol Monogr* 85:3–28. <https://doi.org/10.1890/14-0661.1>
- Hurlbert AH, Jetz W (2007) Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *P Natl Acad Sci USA* 104:13384–13389. <https://doi.org/10.1073/pnas.0704469104>
- IBGE (2010) Censo Agropecuário. Instituto Brasileiro de Geografia e Estatística, Brasília, DF, Brazil. <https://censos.ibge.gov.br/>. Accessed 15 June 2020
- IBGE (2020) Áreas Territoriais. Instituto Brasileiro de Geografia e Estatística, Brasília, DF, Brazil. <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html?t=acesso-ao-produto&c=43>. Accessed 15 June 2020
- IUCN (2001) IUCN Red List Categories and Criteria: Version 3.1. IUCN Species Survival Commission. IUCN, Gland, Switzerland and Cambridge, UK
- Kéry M (2011) Towards the modelling of true species distributions: commentary. *J Biogeogr* 38:617–618. <https://doi.org/10.1111/j.1365-2699.2011.02487.x>
- Kéry M, Guillera-Arroita G, Lahoz-Monfort JJ (2013) Analysing and mapping species range dynamics using occupancy models. *J Biogeogr* 40:1463–1474. <https://doi.org/10.1111/jbi.12087>
- Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659–1673. <https://doi.org/10.2307/1939924>
- Lenth RV (2022) emmeans: estimated marginal means, aka least-squares means. R package version 1.7.2. <https://cran.r-project.org/package=emmeans>. Accessed 15 June 2020
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255. [https://doi.org/10.1890/0012-9658\(2002\)083\[2248:esorwd\]2.0.co;2](https://doi.org/10.1890/0012-9658(2002)083[2248:esorwd]2.0.co;2)
- MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE (2006) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press
- Mackenzie DI, Royle JA (2005) Designing occupancy studies: general advice and allocating survey effort: designing occupancy studies. *J Appl Ecol* 42:1105–1114. <https://doi.org/10.1111/j.1365-2664.2005.01098.x>

- Mainali K, Hefley T, Ries L, Fagan W (2020) Matching expert range maps with species distribution model predictions. *Conserv Biol* 34:1292–1304. <https://doi.org/10.1111/cobi.13492>
- Miller DAW, Pacifici K, Sanderlin JS, Reich BJ (2019) The recent past and promising future for data integration methods to estimate species' distributions. *Methods Ecol Evol* 10:22–37. <https://doi.org/10.1111/2041-210X.13110>
- Miranda EB (2021) *Rhea americana* distribution: range expansion and introductions of America's largest bird. In: Mikkola H (ed) *Birds-challenges and opportunities for business, conservation and research*. IntechOpen, London
- Murray JV, Goldizen AW, O'Leary RA, McAlpine CA, Possingham HP, Low-Choy S (2009) How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. *J Appl Ecol* 46:842–851. <https://doi.org/10.1111/j.1365-2664.2009.01671.x>
- Overbeck GE, Müller SC, Fidelis A, Pfadenhauer J, Pillar VD, Blanco CC, Boldrini II, Both R, Forneck ED (2007) Brazil's neglected biome: the South Brazilian Campos. *Perspect Plant Ecol* 9:101–116. <https://doi.org/10.1016/j.ppees.2007.07.005>
- Pacifici K, Reich BJ, Miller DAW, Gardner B, Stauffer G, Singh S, McKerrow A, Collazo JA (2017) Integrating multiple data sources in species distribution modeling: a framework for data fusion. *Ecology* 98:840–850. <https://doi.org/10.1002/ecy.1710>
- R Core Team (2021). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>. Accessed 15 June 2020
- Ribeiro CM, de Quadros FLF (2015) Valor histórico e econômico da pecuária. In: Pillar VD, Lange O (eds) *Os campos do Sul. Rede Campos Sulinos*. UFRGS, Porto Alegre, RS, Brazil, pp 123–132
- Royle JA, Kéry M (2007) A Bayesian state-space formulation of dynamic occupancy models. *Ecology* 88:1813–1823. <https://doi.org/10.1890/06-0669.1>
- Sales J (2006) The rhea, a ratite native to South America. *Avian Poult Biol Rev* 17:105–124. <https://doi.org/10.3184/147020606783438759>
- Sick H (2001) *Ornitologia Brasileira*. Nova Fronteira, Rio de Janeiro
- Simberloff D (1998) Flagships, umbrellas, and keystones: is single-species management passé in the landscape era? *Biol Conserv* 83:247–257. [https://doi.org/10.1016/S0006-3207\(97\)00081-5](https://doi.org/10.1016/S0006-3207(97)00081-5)
- Souza CM Jr, Shimbo JZ, Rosa MR, Parente LL, Alencar AA, Rudor BFT, Hasenack H, Matsumoto M, Ferreira LG, Souza-Filho PWM, Oliveira SW, Rocha WF, Fonseca AV, Marques CB, Diniz CG, Costa D, Monteiro D, Rosa ER, Vélez-Martín E, Azevedo T (2020) Reconstructing three decades of land use and land cover changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sens* 12:2735. <https://doi.org/10.3390/rs12172735>
- Stauffer GE, Miller DAW, Williams LM, Brown J (2018) Ruffed grouse population declines after introduction of West Nile virus. *J Wildlife Manage* 82:165–172. <https://doi.org/10.1002/jwmg.21347>
- Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling S (2009) eBird: a citizen-based bird observation network in the biological sciences. *Biol Conserv* 142:2282–2292. <https://doi.org/10.1016/j.biocon.2009.05.006>
- Veldman JW, Overbeck GE, Negreiros D, Mahy G, Le Stradic S, Fernandes GW, Durigan G, Buisson E, Putz FE, Bond W (2015) Tyranny of trees in grassy biomes. *Science* 347:484–485. <https://doi.org/10.1126/science.347.6221.484-c>
- Vélez-Martín E, Rocha CH, Blanco C, Azambuja BO, Hasenack H, Pillar VD (2015) Conversão e Fragmentação. In: Pillar VD, Lange O (eds) *Os campos do Sul. Rede Campos Sulinos – UFRGS, Porto Alegre, RS, Brazil*, pp 123–132
- Walpole MJ, Leader-Williams N (2002) Tourism and flagship species in conservation. *Biodivers Conserv* 11:543–547. <https://doi.org/10.1023/A:1014864708777>
- WikiAves (2019) WikiAves, a Enciclopédia das Aves do Brasil. <http://www.wikiaves.com.br>. Accessed 15 June 2020
- Zizka A, Silvestro D, Andermann T, Azevedo J, Duarte Ritter C, Edler D, Farooq H, Herdean A, Ariza M, Scharn R, Svantesson S, Wengstrom N, Zizka V, Antonelli A (2019) CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods Ecol Evol* 10:744–751. <https://doi.org/10.1111/2041-210X.13152>
- Zulian V, Miller DAW, Ferraz G (2021) Integrating citizen-science and planned-survey data improves species distribution estimates. *Divers Distrib* 27:2498–2509. <https://doi.org/10.1111/ddi.13416>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.