Invasive potential of the pied crow (Corvus albus) in eastern Brazil: best to eradicate before it spreads

(Supplementary Methods)

Contents:

Table S1 – Pied crow (*Corvus albus*) records in Brazil with their respective coordinates, date of observation, locality and source.

Longitude	Latitude	Data	Locality	Source
-	-	26/07/04	Porto de Santos	Lima & Kamada 2009
46º17'40”W	23º59'11”S	26/03/06	Santos/Guarujá	Olmos & Silva 2007
46º24'05”W	23º52'48”S	20/07/06	Rio Cubatão	Olmos & Silva 2007
-	-	30/03/07	Santo/Guarujá	Olmos & Silva 2007
-	-	28/08/07	Cubatão	Lima & Kamada 2009
46º18'48.08”W	23º55'05.6”S	13/05/08	Estuário Santos	Lima & Kamada 2009
46º19'17.3”W	23º54'54”S	17/06/08	Ilha Barnabé - Santos	Lima & Kamada 2009
46º17'26.27”W	23º59'14.22”S	21/08/08	Santos/Guarujá - Balsa	Lima & Kamada 2009

Supplementary Methods

Spatial filter approach

Spatial filter analyses helps reduce the effect of uneven, or biased occurrence points for a given species (Aiello et al. 2015). We applied Aiello et al. (2015) procedure as implemented in the R package "spThin". Their procedure uses randomization algorithms to return occurrence points based on user supplied minimum distance. We used a minimum distance of 10km and 100 iterations (Figure S1).

Figure S1: Occurrence points

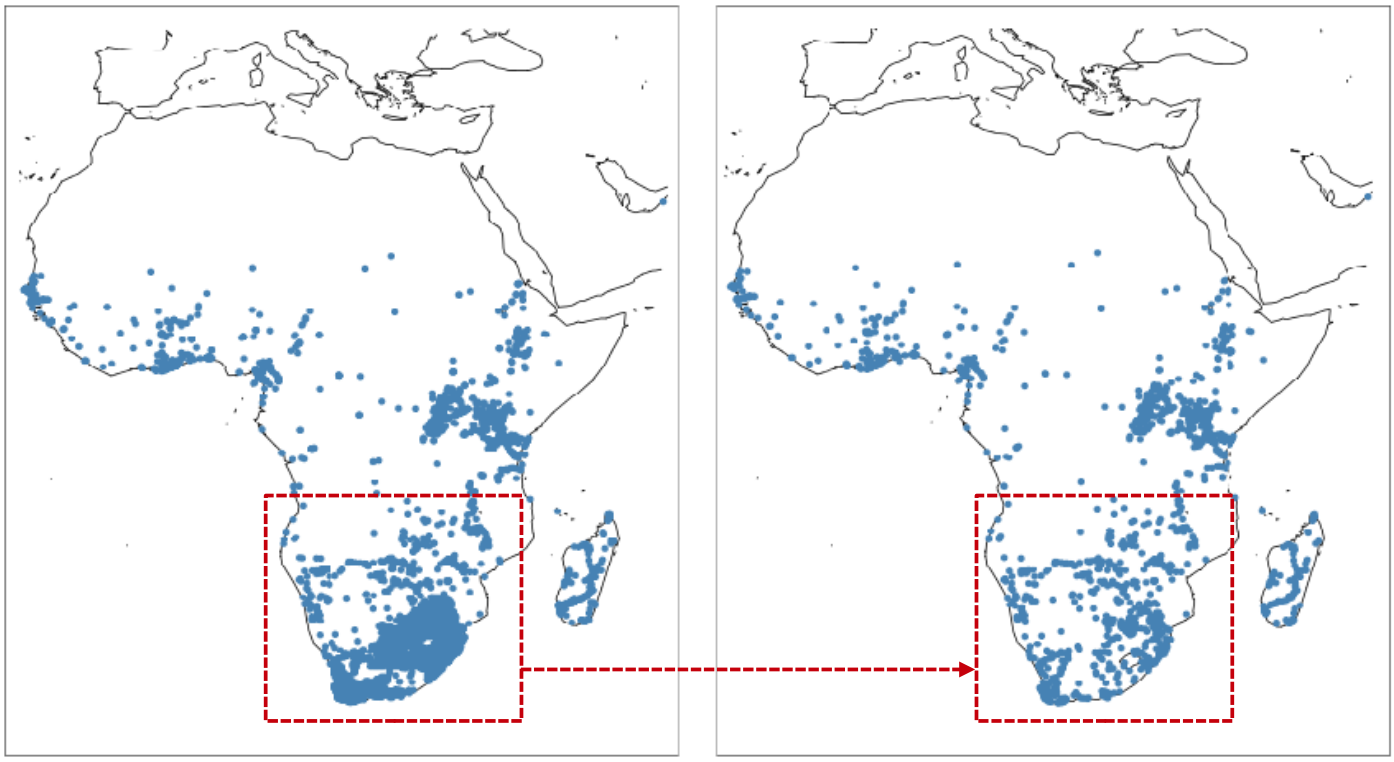

Figure S1: Left, occurrence points before the use of the spatial filter. Right, occurrence points after the use of the spatial filter. Red doted box shows how the spatial filter managed to reduce possible sampling biases.

Selection of Bioclimatic variables

The WorldClim database (Hijman et al. 2005) has 19 bioclimatic variables. However, we interpret that BIO2 = Mean Diurnal Range (mean of monthly (max temp – min temp) and BIO3 = Isothermality ((BIO2/BIO7)* 100) are related to daylight duration, an important condition for species with flight migratory behavior. We also interpret that BIO4 = Temperature Seasonality (standard deviation *100), BIO15 = Precipitation Seasonality (coefficient of variation) are related with sharp defined breeding season. Based on the absence of flight migration behavior and irregularity of the breeding season in *C. albus* (Madge and Juana 2014), we decided to remove these environmental layers. We used Principal Component Analyses (PCA) to identify variables that were correlated (i.e., had similar vector directions; Figure S2) and chose the variable with the highest loading (Table S2). Data used in PCA was obtained by extracting the environmental values from the 16 bioclimatic variables for each of the 1318 data points used to model the potential distribution of *C. albus*. We used the Broken Stick Criteria (Legendre & Legendre 1998) to determine the number of principal components to be retained (Table S3). The Broken Stick criterion assumes that the total variance is divided randomly amongst the various components creating an expected distribution of eigenvalues that follow a broken-stick distribution. Observed eigenvalues are considered interpretable if they exceed eigenvalues generated by the broken-stick model, and components with eigenvalues higher than the Broken Stick criterion were retained. The PCA revealed four clusters (Figure S2). We chose the variable with the highest loading for the clusters identified in red and green (Figure S2). However, this approach could not be used for the yellow and orange clusters (Table S3). In this case we decided to keep the variable that had the highest loading in the first principal component, since this component explained a higher proportion of the variance (Table S2).

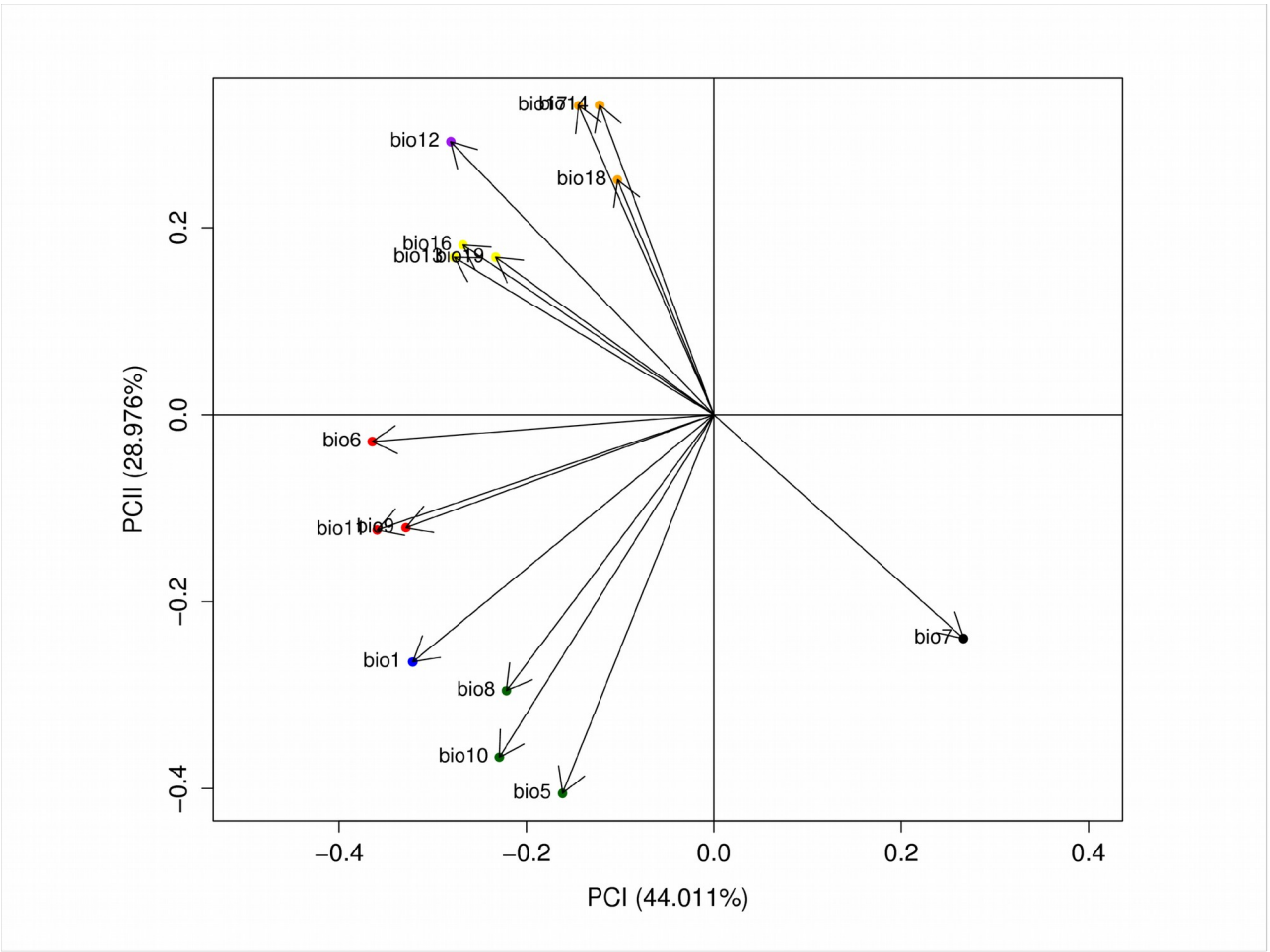

Figure S2: Ordination plot of climate variables. Correlated variables are indicated by same color dots.

Table S2: Loadings of the bioclimatic variables. Numbers in bold indicate highest loadings used for variable selection. In italic the variable retained for modeling the distribution of *C. albus.*

Cluster	Variable	Bioclimatic Variable Names	PC1	PC2
I (Black)	*Bio07*	*Temperature Annual Range*	**0.266**	**-0.239**
II (Green)	*Bio05*	*Max Temperature of Warmest Month*	- 0.161	**-0.405**
	Bio10	Mean Temperature of Warmest Quarter	-0.228	- 0.366
	Bio8	Mean Temperature of Wettest Quarter	-0.221	- 0.295
III (Blue)	*Bio01*	*Annual Mean Temperature*	**-0.321**	**- 0.264**
IV (Red)	*Bio06*	*Min Temperature of Coldest Month*	**- 0.364**	- 0.239
	Bio09	Mean Temperature of Driest Quarter	- 0.328	-0.120
	Bio11	Mean Temperature of Coldest Quarter	-0.228	-0.366
V (Yellow)	*Bio13*	*Precipitation of Wettest Month*	**-0.277**	**0.168**
	Bio16	Precipitation of Wettest Quarter	-0.267	0.181
	Bio19	Precipitation of Coldest Quarter	- 0.232	0.168
VI (Purple)	*Bio12*	*Annual Precipitation*	**-0.280**	**0.291**
VII (Orange)	Bio14	Precipitation of Driest Month	- 0.121	0.330
	Bio17	*Precipitation of Driest Quarter*	**- 0.144**	**0.330**
	Bio18	Precipitation of Warmest Quarter	- 0.102	0.251

Table S3: Component selection using the Broken Stick Criteria. Components with eigenvalues higher than Broken Stick values were retained.

	PC1	PC2	PC3
Broken-Stick	3.318	2.318	1.818
Eigenvalues	6.601	4.346	1.726

Raster layer procedure

All Bioclimatic variables used are from the WorldClim database (Hijmans 2005) at 2.5 arc minutes. 3600 rows and 8640 columns form its extension. However, the anthropogenic environmental layers are at a 30 arc seconds resolution, and therefore, have a different numbers of rows and columns. To deal with the divergence in resolution and extension of the environmental variables in the dataset, we used a bilinear interpolation approach to re-sample the anthropic environmental layers and create a new raster with same extension and resolution as those of the WorldClim dataset. After this procedure, all environmental layers were projected using the WGS84 coordinate system and downscaled to 0.25º cells. We used the mean value for this higher resolution (i.e. we calculated the mean value using the data of the cells that were at a smaller scale). The new dataset was masked to the native African region (20W, 55E, -35S, 45N) and non-native South American region (-85W, -35E, -55S, 12N) and saved as a text file. Finally, to guarantee that all text files are in accordance with the raster dataset, we plotted the grid values from text files against the raster dataset (Figure A and B).

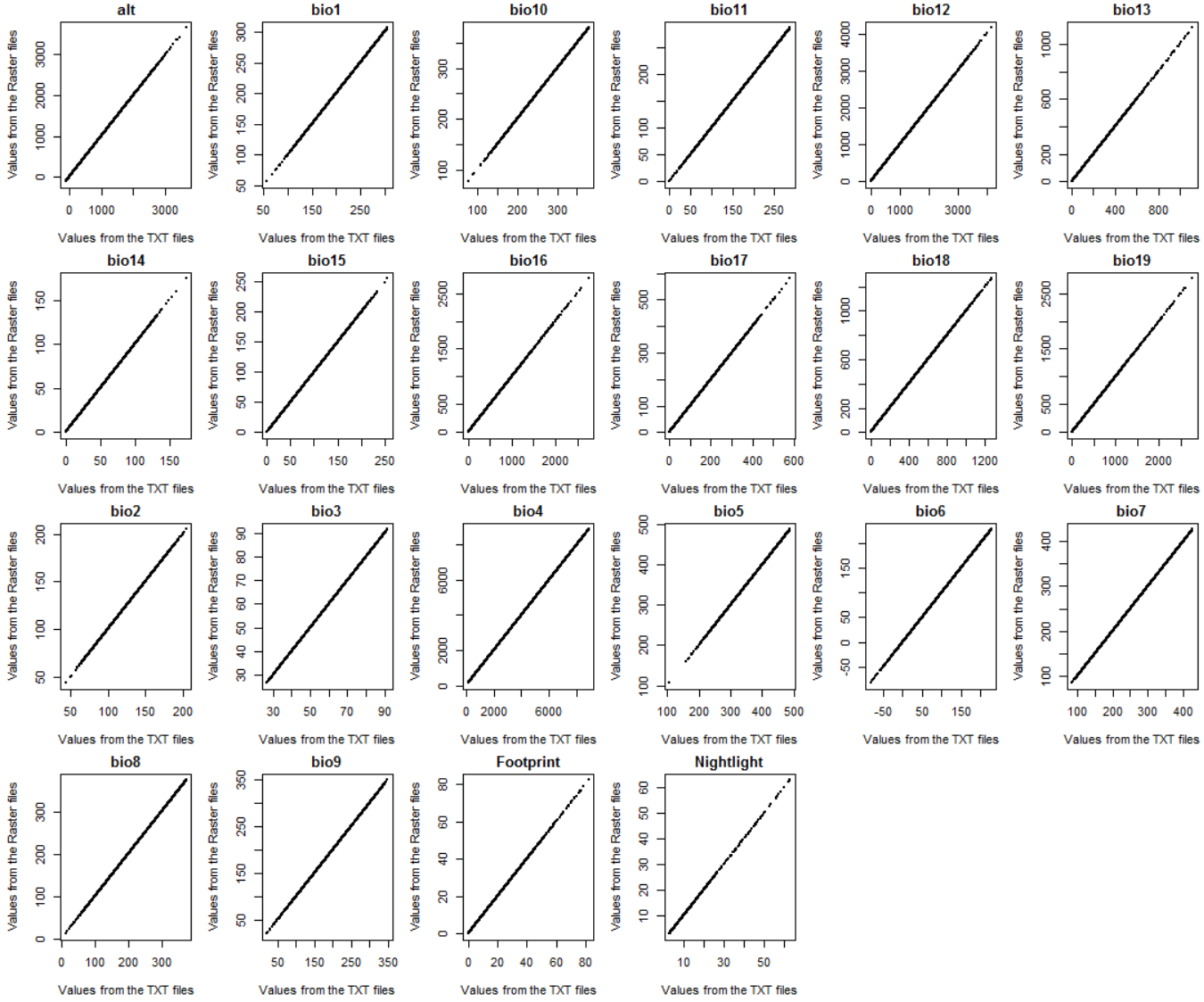

Figure A: correlation between the values extracted from the text and raster files for the native region.

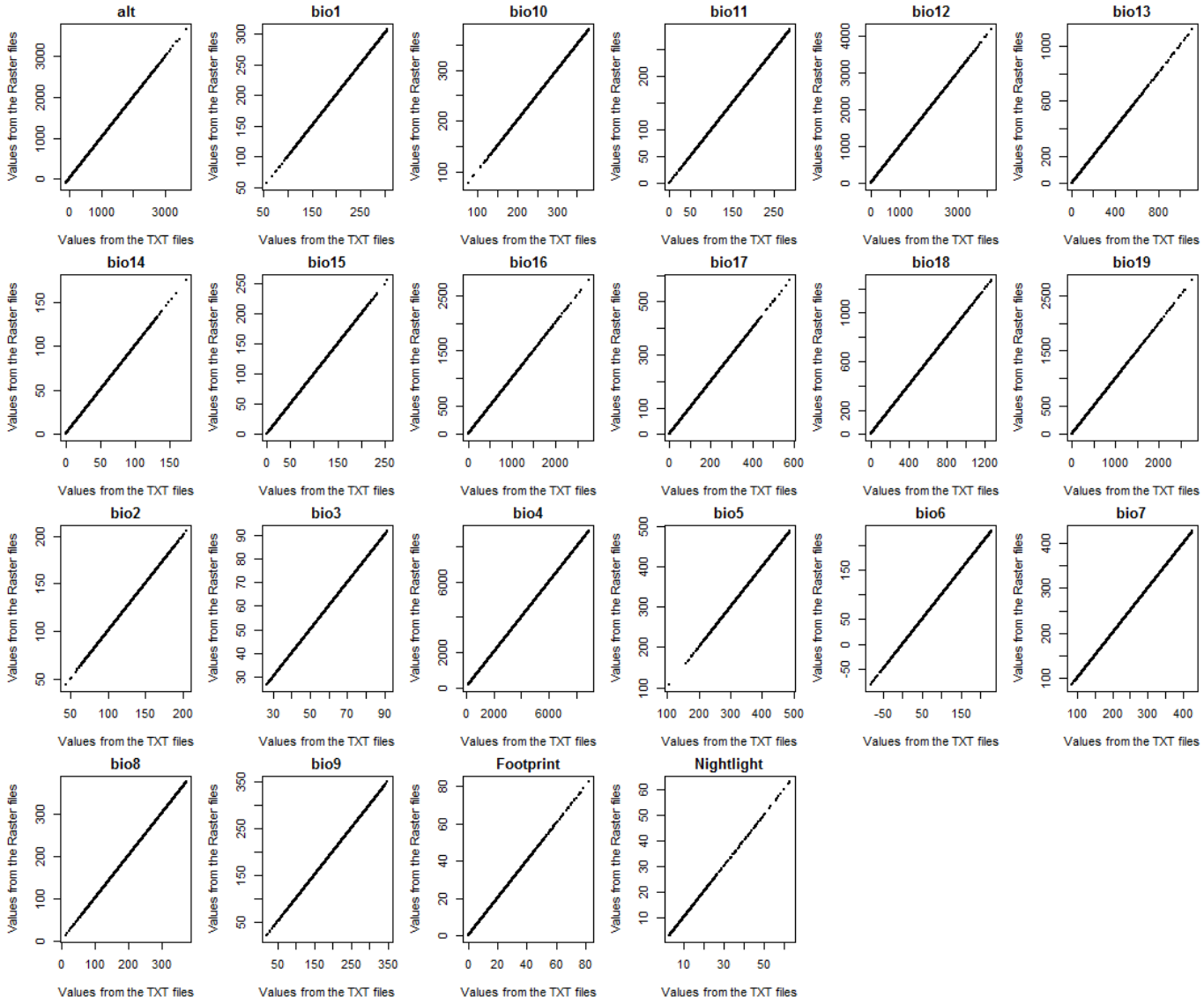

Figure B: correlation between the values extracted from the text and raster files for the non native region.

Figure S3

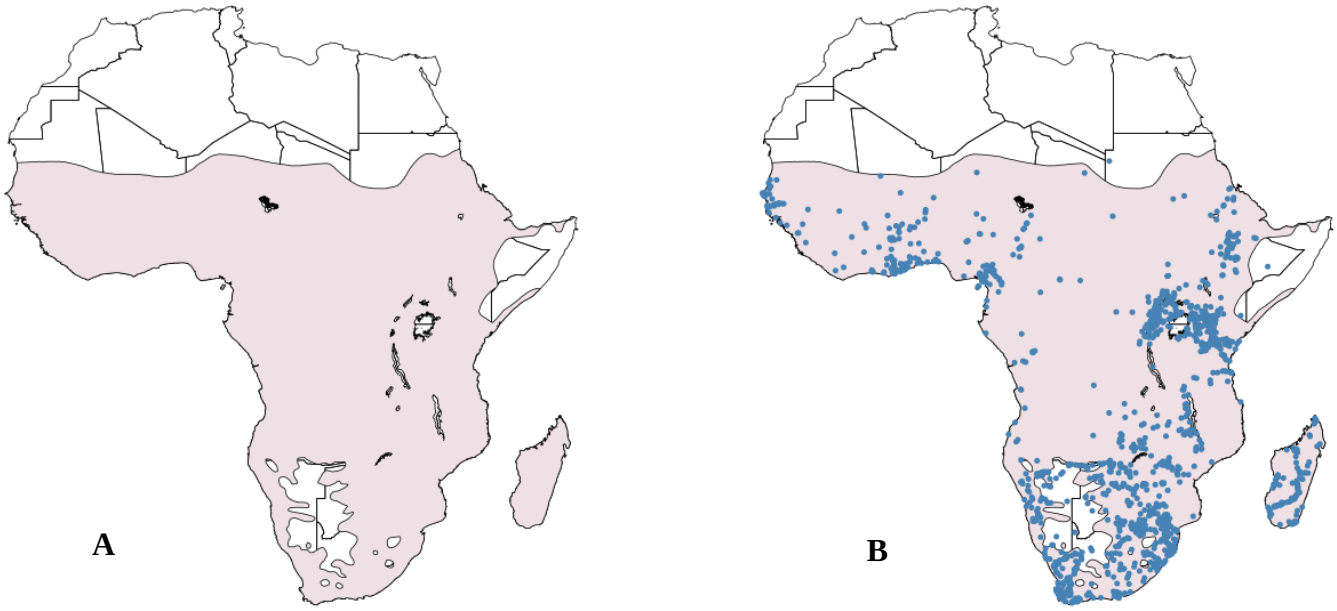

Figure S3: A) Range distribution of *C. albus* according to Bird Life. **B)** Range distribution with final occurrence data points.

Ecological Niche Models and model evaluation

We used the following distribution models: Bioclim (Nix 1986, Booth et al. 2014), Domain (Carpenter et al. 1971), Mahalanobis distance (Farber & Kadmon 2003), Support Vector Machine (Guo & Graham 2005, Drake et al. 2006), and Maxent (Phillips et al. 2004). We choose these distribution models due to the different methodological approaches implemented in their algorithms (Loyola & Rangel, 2000) and the capacity to model presence only data (Franklin et al. 2010). Each niche model was built using 100 iterations. We used mean suitability values obtained from these iterations to generate a final distribution model for each of the five distributional models used. Final models that presented mean pROC values larger than 1 were: Bioclim, Domain, Mahalanobis, Maxent and Support Vector Machine. We then combined the final models from these five modeling procedures to produce a single model. For this procedure we also calculated mean suitability per grid cell. We then, used the sensitivity = specificity threshold to create a binary map for each distribution model. As a final step, we overlapped the binary maps by summing the values in each grid cell, which resulted in six possible values for the predicted presence of *C. albus*: (i) zero, when absence was predicted in the five models; (ii) one, when presence was predicted in only one model; (iii) two, when presence was predicted in two models; (iv) three, when presence was predicted in three models; (v) four, when presence was predicted in four models; and (vi) five, when presence was predicted in all of the models. We considered *C. albus* to be present when grid cells of combined models had a value equal or higher than 2. We chose this value because only two of the models (SVM and Mahalanobis) had very different predictions.

Biome region relative Index approach

To analyze the susceptibility of each biome to the potential occurrence of *C. albus*, we used the biome relative index approach (BRI) as implemented by Ortega-Andrade et al. (2015). This relative index indicates the proportion of suitable area for each biome, which varies from 0 (i.e. no suitable area was predicted for the specific biome) to 1 (suitable area was predicted to occur in a single biome). The index is obtained by dividing the total suitable area (in Km^2) detected by the final binary model by the extent of the biome area (in Km^2) within the binary model (Equation 1). We used GIS tools in the raster package (Hijmans 2015) to estimate the area in Km^2 with cells at a 0.25º resolution. Digital maps of biomes were obtained from Instituto Brasileiro de Geografia e Estatística (IBGE – www.ibge.gov.br).

$$BRI = \frac{Suitablearea\left(Km^2\right)}{Biomearea\left(Km^2\right)}$$

Equation 1: Biome relative index equation

Supplementary Results

Figure results S4

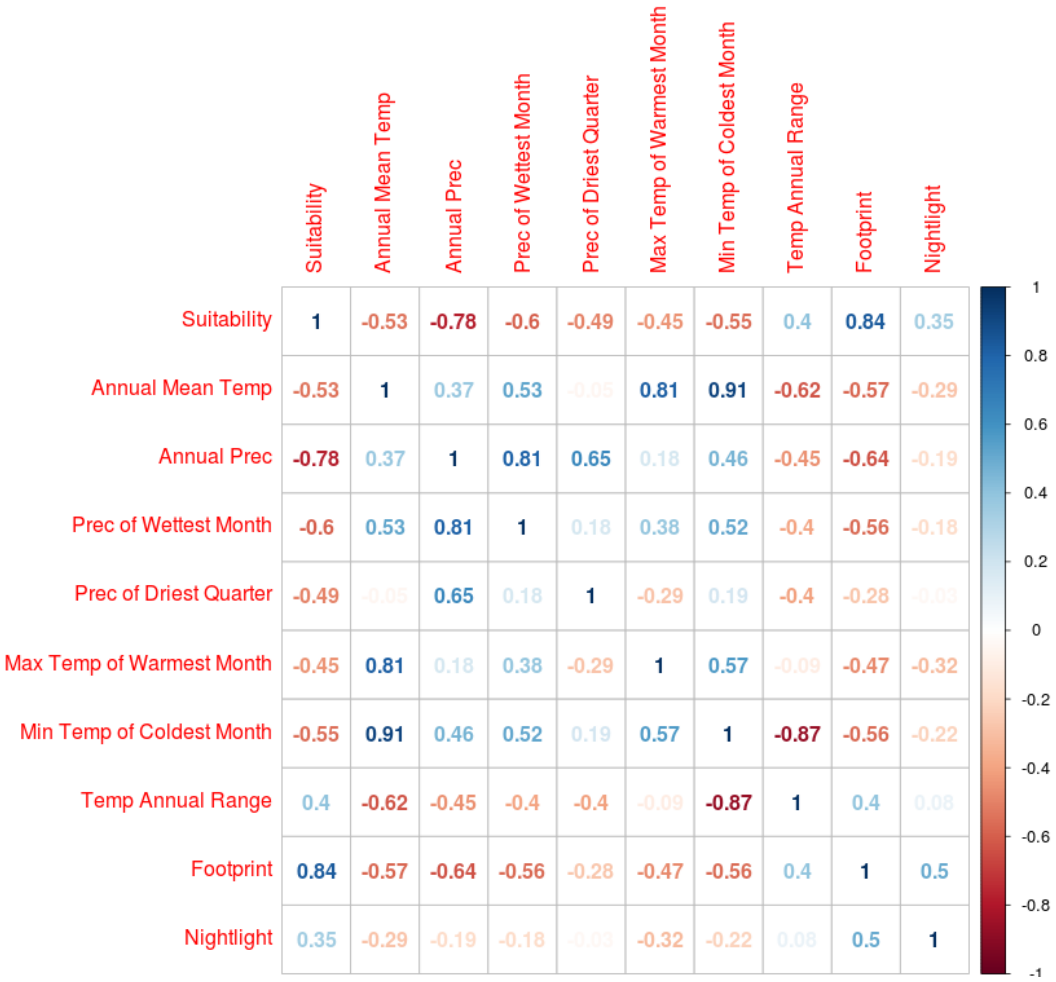

Figure S4: Pairwise correlation values of suitability values obtained by the ensemble approach and the environmental variables used in the modeling procedures.

Figure S5 – Native Model

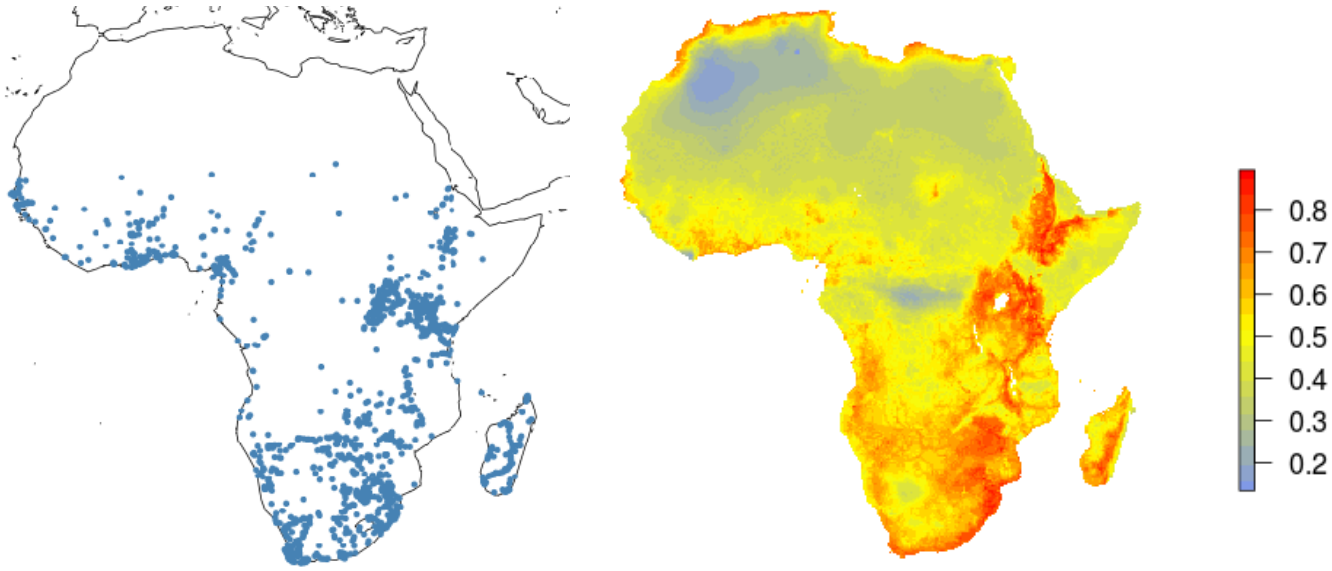

Figure S5: Pied crow (*C. albus*) occurrence records in blue and distribution model for the native range of Africa after the ensemble approach.

Figure results S6

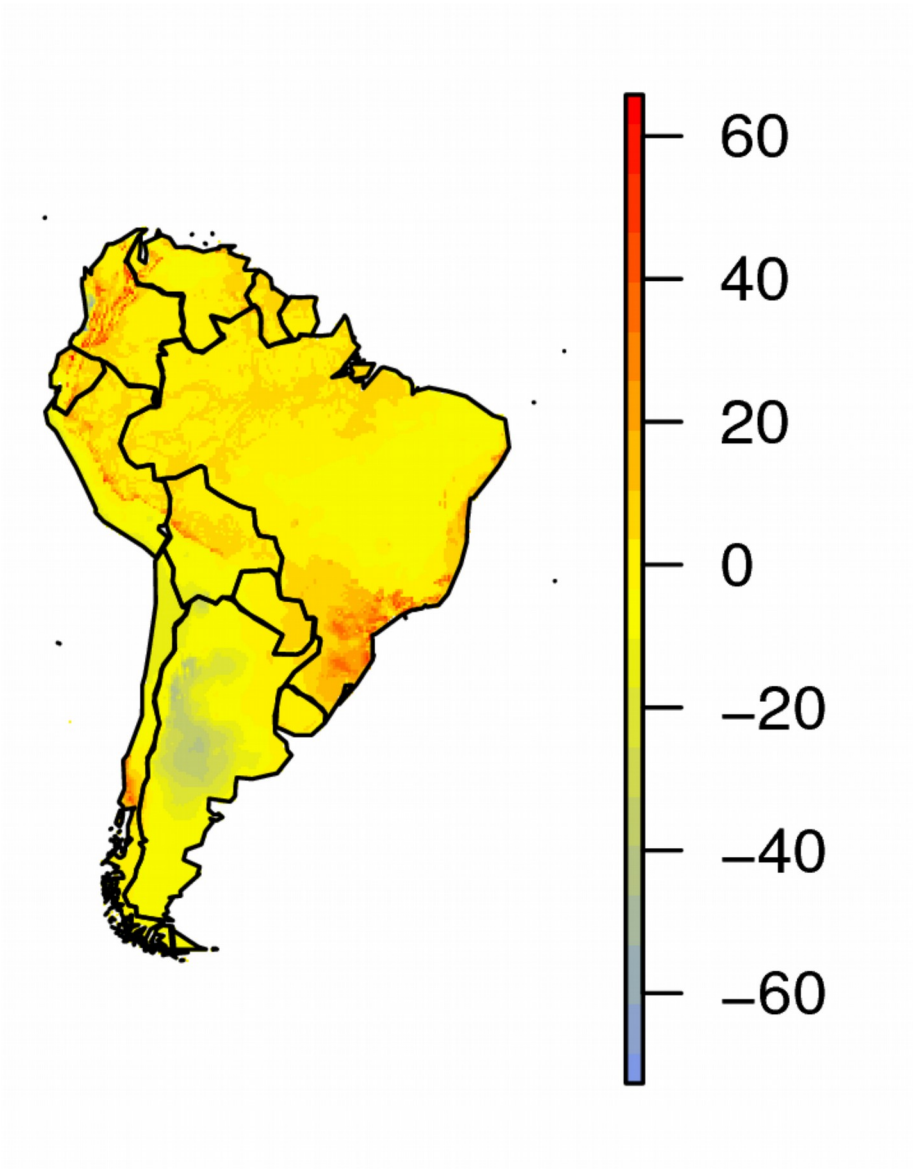

Figure results S6: Multivariate environmental similarity surface (MESS) using training data from the native range of Africa.

Figure results S7

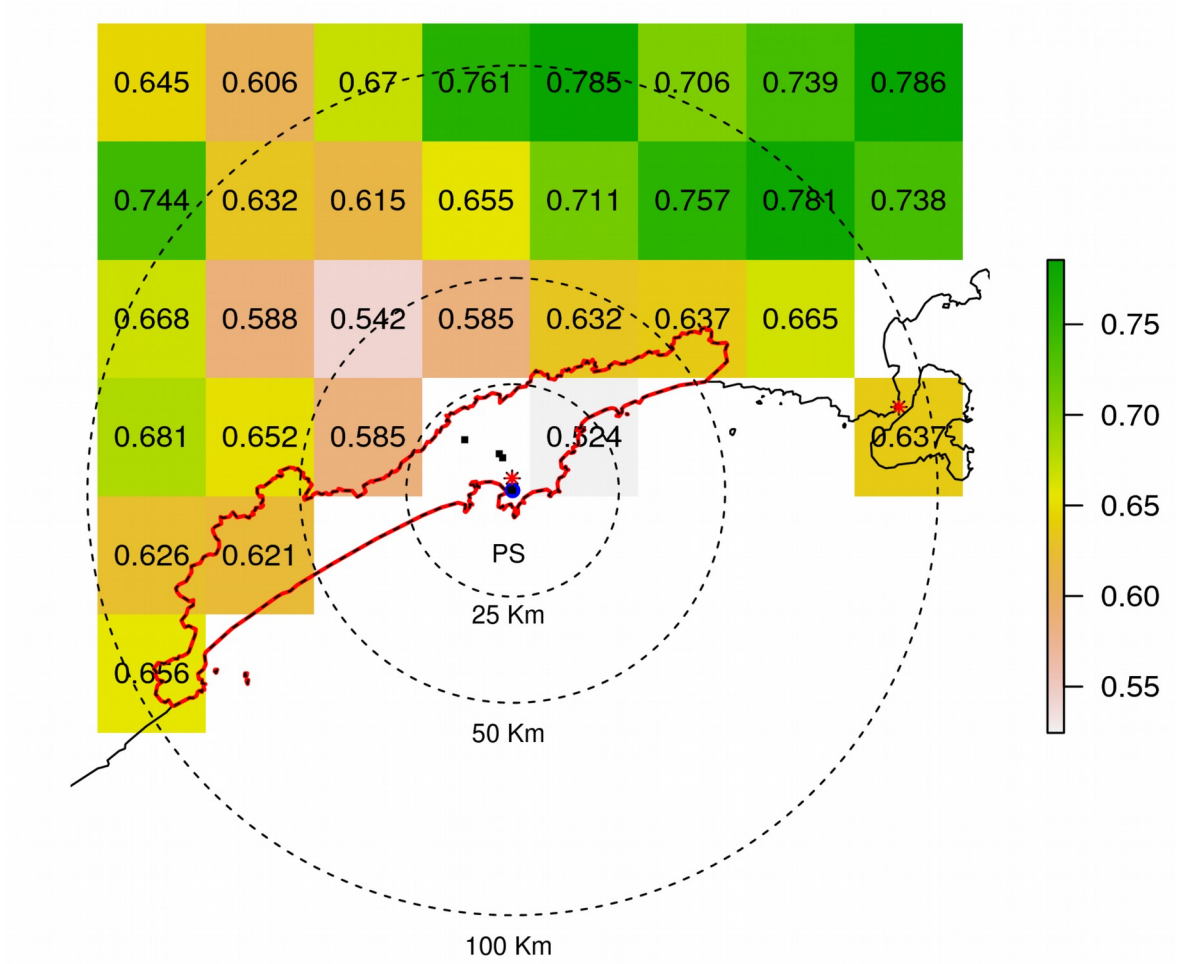

Figure S7: Suitability values for grid cells close to the occurrence records. Because known records for the species are located in a grid cell with non available suitability values, we assumed the nearest neighboring grid cell as its suitability value. Black squares are occurrence sites obtained from Lima & Kamada (2009) and PS is “Porto de Santos”.

Figure results S8

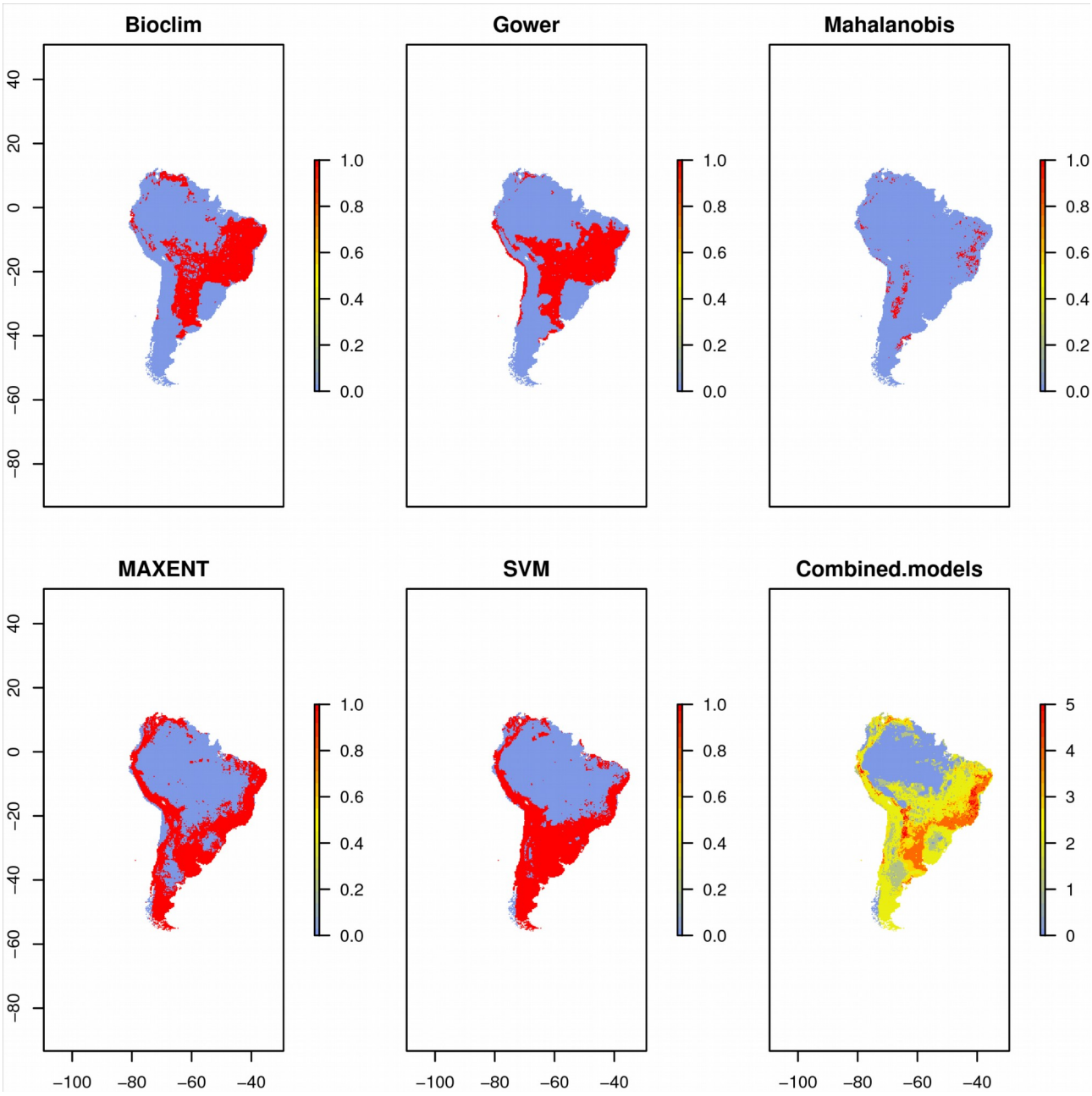

Figure S8: Binary distribution maps for each algorithm. Blue color indicates unsuitable regions, while red colors indicate suitable regions. Combined model is the the sum of all binary model.

References

Aiello-lammens M.E., R.A. Boria, A. Radosavljevic et al., . spThin : an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography, 38, 2015, 541–545. doi:10.1111/ecog.01132

Booth, T.H., H.A. Nix, J.R. Busby et al.,. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Diversity and Distributions, 20, 2014 1-9

Carpenter, G., Gillison, A.N., and Winter, J., Domain - A flexible modeling procedure for mapping potential distributions of plants and animals. Biodivers. Conserv, 1993, 2, 667–680. doi:10.1007/BF00051966

Drake, J.M., Randin, C andGuisan, A., Modelling ecological niches with support vector machines. J. Appl. Ecol. 43, 2006, 424–432. doi:10.1111/j.1365-2664.2006.01141.x

Farber, O. and Kadmon, R.,. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. Ecol. Modell. 160, 2003, 115–130. doi:10.1016/S0304-3800(02)00327-7

Guo, Q., Kelly, M., and Graham, C.H., Support vector machines for predicting distribution of Sudden Oak Death in California. Ecol. Modell. 182, 2005, 75–90. doi:10.1016/j.ecolmodel.2004.07.012

Hijmans R.J., Cameron, S.E., Parra, J.L., et al., Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 2005, 1965–1978. doi:10.1002/joc.1276

Hijmans RJ., raster: Geographic Data Analysis and Modeling. R package version 2.5-8. 2016 https://CRAN.R-project.org/package=raster

Legendre P and Legendre L., Numerical ecology. Amsterdam: Elsevier, 1998

Lima B, and Kamada B., Registros de corvo-bicolor Corvus albus (Passeriformes: Corvidae) em território brasileiro. Atualidades Ornitológicas. 2009

Madge, S. and de Juana, E., . Pied Crow (Corvus albus). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, E. (eds.), 2014, Handbook of the birds of the world Alive. Lynx Edicions, Barcelona.

Nix, H.A., A Biogeographic Analysis of Australian Elapid snakes. In: Longmore, R., Ed., *Atlas of Elapid Snakes of Australia*. *Australian Flora and Fauna Series No*. 7, 1986, Australian Government Publishing Service, Canberra, 4-15.

Ortega-Andrade H.M., Prieto-Torres D.A., Gómez-Lora I., et al., . Ecological and geographical analysis of the distribution of the Mountain Tapir (Tapirus pinchaque) in Ecuador: importance of protected areas in future scenarios of global warming. PLoS ONE 10(3): e0121137, 2015, doi:10.1371/journal.pone.

Phillips, S., Dudík, M., and Schapire, R.,. A maximum entropy approach to species distribution modeling. Proc. Twenty-first . 2004, 655–662. doi:10.1145/1015330.1015412