Species Distribution Modeling for Conservation Purposes

Katia Maria Paschoaletto Micchi de Barros Ferraz*, Silvio Frosini de Barros Ferraz†, Rogerio Cunha de Paula‡, Beatriz Beisiegel§ & Christine Breitenmoser¶

1 Departamento de Ciências Florestais, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo – USP, Piracicaba, Brasil
2 Centro Nacional de Pesquisa e Conservação de Mamíferos Carnívoros, Instituto Chico Mendes de Conservação da Biodiversidade – MMA, Atibaia, SP, Brasil
3 KORA, Muri, Switzerland

Abstract
Species distribution models (SDMs) can be useful for different conservation purposes. We discuss the importance of fitting spatial scale and using current records and relevant predictors aiming conservation. We choose jaguar (Panthera onca) as a target species and Brazil and Atlantic Forest biome as study areas. We tested two different extents (continent and biome) and resolutions (~4 Km and ~1 Km) in Maxent with 186 records and 11 predictors (bioclimatic, elevation, land-use and landscape structure). All models presented satisfactory AUC values (>0.70) and low omission errors (<23%). SDMs were scale-sensitive as the use of reduced extent implied in significant gains to model performance generating more constrained and real predictive distribution maps. Continental-scale models performed poorly in predicting potential current jaguar distribution, but they reached the historic distribution. Specificity increased significantly from coarse to finer-scale models due to the reduction of overprediction. The variability of environmental space (E-space) differed for most of climatic variables between continental and biome-scale and the representation of the E-space by predictors differed significantly (t = 2.42, g.l.= 9, P < 0.05). Refining spatial scale, incorporating landscape variables and improving the quality of biological data are essential for improving model prediction for conservation purposes.

Key words: Conservation Biology, Environmental Space, Jaguar, Landscape, Maxent, Scale.

Introduction
Species distribution models (SDMs) can be useful for different conservation purposes such as predicting current or potential species distribution across landscape (Franklin 2009; Elith et al. 2010), evaluating distribution range and/or environmental suitability for conservation assessments (Anderson & Martínez-Meyer 2004; Chefaoui et al. 2005; Ferraz et al. 2012a, b) and identifying key areas for setting conservation priorities (Thorn et al. 2009). As most of these purposes are directly related to species (and/or ecosystem) conservation, models should provide information at a spatial scale coherent to the management action scale. Accurate estimates of potential current distribution of target species are particularly important and relevant when SDM is used to aid conservation planning (Ferraz et al. 2012a, b). For modeling species over a very short time scale, dynamic and more proximal variables become necessary (Peterson et al. 2011). The model accuracy should be reached by the input of good quality of biological data (current and accurate presence points) and functionally relevant predictor’s variables (Lobo et al. 2008; Elith & Leathwick 2009).

In this paper we tested different spatial scales and predictors leading us to discuss how to improve the quality of models aiming conservation purposes in practice. We emphasize the importance of fitting spatial scale attributes to the question and purpose of study and the incorporation of human aspects and landscape structure variables improving models for decision-making in conservation actions. We choose the jaguar (Panthera onca) as a target species and Brazil and Atlantic Forest biome as study areas since recent reference models (Ferraz et al. 2012a, b), validated by jaguar experts, are available for comparison and discussion. Also, an available database with recent and good quality records provided by the IUCN/SSC Cat Specialist Group became possible to model the potential current species distribution as proposed here. Although jaguar official red list status in Brazil is Vulnerable, a new approach of red listing was recently incorporated by national authorities. Thus, besides
the national evaluation, a biome status has been processed as well. Jaguar has been considered Critically Endangered in the Atlantic Forest because a population reduction of 50-90% was estimated in the past 10-15 years in the largest subpopulations of the Upper Paraná and is suspected at the Coastal Atlantic Forest (Beisiegel et al. 2012).

**Material and Methods**

**Study area**

We fit SDMs for Brazil (10° S and 55° W) and Atlantic Forest biome (3°-31° S and 35°-60° W) (Figure 1). Atlantic Forest, one of the world’s 25 biodiversity hotspots (Myers et al. 2000), has also undergone a huge forest loss, with only 11.7% remaining of the original forest, most of it distributed as small (less than 50 ha) and isolated forest fragments (Ribeiro et al. 2009). The three largest fragments, account for more than 2 million ha (more than 13% of the remaining forest), are located in the Serra do Mar, a 1,500 km long system of mountain ranges and escarpments in Southeastern Brazil. Although nature reserves protect a small amount of the remaining Atlantic forest, the Serra do Mar has 25.2% of its remaining forest under protection (Ribeiro et al. 2009).

**Species data and variable predictors**

We used the Cat SG database provided by the IUCN/SSC Cat Specialist Group (CatSG) (n = 186 for Brazil, n = 54 for Atlantic Forest biome) (Figure 1). This database was compiled by published data of jaguar presence records (live sighting, radio track location, remains, tracks or scat collection) provided by specialists during the workshop (São Francisco de Paula, Brazil, Brazil, 2005), organized by the IUCN/SSC Cat Specialist Group to assess the status and conservation needs of this cat species in the Neotropics. All jaguar occurrence records are temporally related to the landscape information used (see below).

The explanatory variables (predictors) used for modeling comprises environmental landscape variables. The

![Figure 1](image_url). Jaguar database used for modeling (n = 186) and the potential current jaguar distribution in Brazil (in gray) used as a reference model (Ferraz et al. 2012a, b). Limits of the Atlantic Forest biome in black.
environmental ones are: annual mean temperature (bio1), mean diurnal range (bio2), maximum temperature of warmest month (bio5), minimum temperature of coldest month (bio6), annual precipitation (bio12), precipitation of wettest month (bio13), precipitation of driest month (bio14), from Worldclim database (http://www.worldclim.org/); and elevation from SRTM (http://www2.jpl.nasa.gov/srtm/). Landscape variables are land-cover from ESA GlobCover Project (2009, http://nia1.esrin.esa.int/); gradient distance from water from IBGE (2004, http://www.ibge.gov.br/); percent tree cover from Global Land Cover Facility (2000-2001, http://glcf.umiacs.umd.edu/data/vcf/). Environmental variables were available at both ~4 km (coarse) and ~1 km (fine) (cell precision of 0.0416 and 0.0083, respectively) spatial resolutions. Globcover, percent of tree cover and distance from water were resampled from ~1 to 4 km.

**Modeling procedures**

We built models in two different extents (continent and biome) and resolutions (coarse and fine) for two different approaches, using only environmental variables and using environmental variables plus landscape variables (land-use and landscape structure). So, we obtained the following models: 1) Continental-scale model based only on environmental variables at coarse (CSCREnV) and fine resolution (CSFRENv); 2) Continental-scale model based on environmental and landscape variables at coarse (CSCREnVLand) and fine resolution (CSFRENvLand); 3) Biome-scale model based only on environmental variables at coarse (BSCREnV) and fine resolution (BSFRENv); and 4) Biome-scale model based on environmental and landscape variables at coarse (BSCREnVLand) and fine resolution (BSFRENvLand).

We modeled jaguar distribution using Maxent (Phillips & Dudik 2008) using 70% of the species data for training and 30% for testing the models. Data were sampled by bootstrapping of 10 random partitions with replacements (Pearson 2007). All runs were set with random seed, a convergence threshold of 1.0E–5 with 500 iterations, with 10,000 background points. The logistic threshold output format was used and models were “cut-off” by thresholds converting the continuous probability in a binary model. Minimum training presence logistic threshold was used for continental models as others were too restricted and unreal. 10 percentile training presence threshold was used for biome models as the same was used in the reference model. Models performances were evaluated by the AUC value, omission error and the binomial probability (Fielding & Bell 1997; Pearson 2007). Gains in sensitivity and specificity for the prediction in the Atlantic Forest biome were assessed from coarse scale continental models to fine scale biome models. Continental models were cut by 10% percentile training presence threshold and sensitivity and specificity were calculated only for the Atlantic Forest Biome.

The potential current jaguar distribution was not successfully predicted by continental-scale at coarse and fine resolution, neither for Brazil as a whole nor considering only the Atlantic Forest biome boundary (Figure 2a-d), due to the overprediction (commission error). Most of areas considered as unsuitable for jaguar occurrence were predicted as suitable for continental scale models at both spatial resolutions. Although, biome-scale models (Figure 2e-h) presented better performance that continental-scale models predicting potential current jaguar distribution in the Atlantic Forest biome. Biome-scale models were quite similar to the Atlantic Forest reference model in predicting suitable areas for the species occurrence, with reduced commission error (1 – specificity) (Figure S1, see Additional Supporting Information available at www.abeco.org.br). In general the spatial resolution did not influenced model results, but extension scale did. Although model results were similar, gains in sensitivity and specificity were evident. Sensitivity was high (ranging from 0.63 to 0.91) for all models, while specificity increased from continental and coarse to biome and fine models (from 0.30 to 0.80) (Figure S1, in Additional Supporting Information).

The incorporation of landscape variables to the environmental models at continental scale did not alter significantly models results (Figure 2), but it did for biome-scale model at coarse resolution (BSCREnVLand) as commission error was reduced. Besides, this model was the one that had the higher omission error (0.37).

The variability of environmental space (E-space) differed for most of climatic variables (bio1, bio12, bio13, bio2, bio5, bio6) between continental and biome-scale (Figure S2, see Additional Supporting Information available at www.abeco.org.br). Otherwise, elevation, bio14, percentage of forest and distance from water presented the same range of variability. Land cover was the variable that most differed between E-spaces modeled as only some of land-uses classes were common for both. The representation of the E-space by predictors differed significantly (t = 2.42, g.l. = 9, P < 0.05). Records represented 64% of the E-space variability for biome-scale models and 77% for continental-scale models. The frequency distribution of environmental values was
Figure 2. Jaguar distribution models at (a) continental-scale and coarse-resolution with environmental variables (CSCREnv), (b) continental-scale and coarse-resolution with environmental and landscape variables (CRCREnvLand), (c) continental-scale and fine-resolution with environmental variables (CSFREnv), (d) continental-scale and fine-resolution with environmental and landscape variables (CSFREnvLand), (e) biome-scale and coarse-resolution with environmental variables (BSCREnv), (f) biome-scale and coarse-resolution with environmental and landscape variables (BSCREnvLand), (g) biome-scale and fine-resolution with environmental variables (BSFREnv), (h) biome-scale and fine-resolution with environmental and landscape variables (BSFREnvLand).
totally different for all variables, suggesting that the Atlantic Forest biome holds a high environmental variation, but different from the continental one.

Models presented AUC values ranging from 0.708 to 0.846 and low omission errors (<23%) (Table S1). The incorporation of landscape variables lead to an increase in AUC values, except for CSFREnvLand. Models were significant (p < 0.05), except CSFREnv and CSFREnvLand (Table S1, see Additional Supporting Information).

**Discussion**

The use of reduced extent implied in gains to model performance generating more constrained predictive distribution maps. SDMs were scale-sensitive improving significantly its prediction of species occurrence at biome level. Continental-scale models performed poorly in predicting potential current jaguar distribution, but they reached the potential historic jaguar distribution found by Torres et al. (2008, 2012). Otherwise, biome-scale models performed better predicting species current distribution, reaching the potential distribution predicted by the reference model (Ferraz et al. 2012b). Although, model results are quite similar, biome-scale models still overestimate the species distribution in some parts of the biome in part due to the wider records distribution that was used for this modeling.

Several authors have pointed out the importance of scale (extension and resolution) for SDM what should reflect the purpose of the analysis (Elith & Leathwick 2009; Franklin 2010; Peterson et al. 2011). The selection of inappropriate resolution and extent, a critical step in SDM building, can lead to misleading results (Guisan & Thuiller 2005). Vaughan & Ormerod (2003) emphasize that a series of spatial scales should be examined, increasing the understanding of organism-environmental relationships, identifying the most effective scales for predictive modeling and complementing the spatial hierarchies often used in conservation planning. On the contrary to spatial scale, the spatial resolutions used here were not sufficiently different to influence the prediction as also observed by McPherson & Jetz (2007).

Specificity increased significantly from coarse to finer-scale models due to the reduction of overprediction reaching what was expected. The low values of AUC, a threshold-independent measure of model accuracy, denote the true generalist nature of the species distribution obtained in this study (Tsoar et al. 2007; Lobo et al. 2008). The increase of AUC for biome-scale models when compared to continental-scale models is in fact and artifact of the restricted distribution of a species in this region, and not an indicator of good model performance.

It is strongly recommended that the variability of the environmental space (conceptual space defined by the environmental variables to which the species responds) is well represented by the sample (presence points database) used for modeling (Vaughan & Ormerod 2003; Peterson et al. 2011). Accuracy models are inversely related to the species geographic range and environmental tolerance (Hernandez et al. 2006; McPherson & Jetz 2007; Tsoar et al. 2007). So modeling in reduced extension scale (e.g., biome, landscape, regional), such as the approach that has been used for the carnivores National Action Plans in Brazil (Ferraz et al. 2012a, b, c), could be a way of limiting the range of continental variation, and at the same time increasing the representation of the environmental features by sample presence points in the modeling formulation. Also, this may reduce the E-space for pseudo-absences selection in the model building. As Chefauí & Lobo (2008) pointed out the elaboration of reliable simulations on the realized distribution of species unavoidably requires good absence data. This approach seems to be a promising alternative for improving models accuracy at finer-scale level since sampling covers the full range of environmental conditions within the study region. Failure to sample correctly can lead to erroneous organism-relationships, affecting predictive ability and interpretation (Vaughan & Ormerod 2003; Costa et al. 2010). Finally, the use of a higher spatial resolution, compatible with reduced extent, should better represents the environmental variables what is a requirement for using landscape structure variables as a functional predictor of organism-landscape relationships.

Landscape variables implied in gains only for BSCREnvLand as this model was the one closest to the reference model for the current species distribution at biome level. The incorporation of landscape variables lead to an increase in AUC values for most of models suggesting better performance. Probably, the landscape variables did not performed better here as the full range of landscape conditions within the extents considered were well covered by records (what is, in fact, desired). In this case, other landscape variables other than the ones used could better discriminate models with landscape variables from the ones without them. Recent works have showing promising results by the use of proximal variables (e.g., landscape features) in SDM in a very short and recent timescale (Buermann et al. 2008; Franklin 2009; Ferraz et al. 2010, 2012c). Incorporating landscape variables (human aspects and landscape structure) as explanatory variables in SDM framework can lead to a real representation of the current status of the species distribution in a landscape (= realized or current distribution). It is particularly appropriate when current species distribution differ from historical distribution due to the pervasive influence of humans upon the environment. The incorporation of recent land-use maps should improve modeling since it may represent habitat and non-habitat patches. Also, landscape structure variables (e.g., percentage of forest, distance from forest remnants, connectivity) may represent the spatial influence of habitats remnants in the whole landscape.

It is recommended that the choice of predictors in SDM should consider its ecological relevancy to the target species (Elith & Leathwick 2009). Careful selection of predictors
utilizing existing knowledge of physiology and environmental processes would improve the interpretability and evaluation of model results (Austin 2007). At landscape level, the choice should include besides predictors related to ecological requirements, the ones that represent potential threats, risks or changes in landscape patterns as they can also be significant and effective predictors of species distribution. It is more imminent if SDMs are to be used in support of conservation planning. Specifically for the Atlantic Forest biome, the jaguar potential current distribution is mainly restricted to the protected areas as the continuous forest remains is highly suitable for the species (Ferraz et al. 2012a). In this case, percentage of vegetation cover and proximity of forest remains are examples of landscape variables that could improve models accuracy, since jaguars are not strictly restricted to habitat patches remnants, but they are also observed using crop and pasture lands (Cullen Junior et al. 2005; Palmeira et al. 2008; Sollmann et al. 2012).

Finally, as information about potential current species distribution is essential for the implementation of conservation actions, only accurate and recent biological data (instead of data from natural history collections) should be used in SDMs in such conditions avoiding errors in model prediction (e.g. overprediction). As the application of SDM in megadiverse Neotropics presents several challenges, especially because data on species occurrences are scarce (Kamino et al. 2011), validation of model results – by experts as preferred – is extremely recommended as statistical results not always is coherent with the reality.

In conclusion, the spatial resolution did not influence model results, but extension scale did. The ability to model species-environment relations that occur at a finer spatial scale is generally lost at broad-scale resulting in models more related to general climatic range limits of species (Franklin 2009). So continental-scale models could not be more related to general climatic range limits of species scale is generally lost at broad-scale resulting in models species-environment relations that occur at a finer spatial scale is generally lost at broad-scale resulting in models more related to general climatic range limits of species (Franklin 2009). So continental-scale models could not be more related to general climatic range limits of species (Franklin 2009). So continental-scale models could not be more related to general climatic range limits of species. It is more imminent if SDMs are to be used in support of conservation planning. Specifically for the Atlantic Forest biome, the jaguar potential current distribution is mainly restricted to the protected areas as the continuous forest remains is highly suitable for the species (Ferraz et al. 2012a). In this case, percentage of vegetation cover and proximity of forest remains are examples of landscape variables that could improve models accuracy, since jaguars are not strictly restricted to habitat patches remnants, but they are also observed using crop and pasture lands (Cullen Junior et al. 2005; Palmeira et al. 2008; Sollmann et al. 2012).

Finally, as information about potential current species distribution is essential for the implementation of conservation actions, only accurate and recent biological data (instead of data from natural history collections) should be used in SDMs in such conditions avoiding errors in model prediction (e.g. overprediction). As the application of SDM in megadiverse Neotropics presents several challenges, especially because data on species occurrences are scarce (Kamino et al. 2011), validation of model results – by experts as preferred – is extremely recommended as statistical results not always is coherent with the reality.

In conclusion, the spatial resolution did not influence model results, but extension scale did. The ability to model species-environment relations that occur at a finer spatial scale is generally lost at broad-scale resulting in models more related to general climatic range limits of species (Franklin 2009). So continental-scale models could not be more related to general climatic range limits of species (Franklin 2009). So continental-scale models could not be more related to general climatic range limits of species. It is more imminent if SDMs are to be used in support of conservation planning. Specifically for the Atlantic Forest biome, the jaguar potential current distribution is mainly restricted to the protected areas as the continuous forest remains is highly suitable for the species (Ferraz et al. 2012a). In this case, percentage of vegetation cover and proximity of forest remains are examples of landscape variables that could improve models accuracy, since jaguars are not strictly restricted to habitat patches remnants, but they are also observed using crop and pasture lands (Cullen Junior et al. 2005; Palmeira et al. 2008; Sollmann et al. 2012).

Finally, as information about potential current species distribution is essential for the implementation of conservation actions, only accurate and recent biological data (instead of data from natural history collections) should be used in SDMs in such conditions avoiding errors in model prediction (e.g. overprediction). As the application of SDM in megadiverse Neotropics presents several challenges, especially because data on species occurrences are scarce (Kamino et al. 2011), validation of model results – by experts as preferred – is extremely recommended as statistical results not always is coherent with the reality.

In conclusion, the spatial resolution did not influence model results, but extension scale did. The ability to model species-environment relations that occur at a finer spatial scale is generally lost at broad-scale resulting in models more related to general climatic range limits of species (Franklin 2009). So continental-scale models could not be more related to general climatic range limits of species (Franklin 2009). So continental-scale models could not be more related to general climatic range limits of species. It is more imminent if SDMs are to be used in support of conservation planning. Specifically for the Atlantic Forest biome, the jaguar potential current distribution is mainly restricted to the protected areas as the continuous forest remains is highly suitable for the species (Ferraz et al. 2012a). In this case, percentage of vegetation cover and proximity of forest remains are examples of landscape variables that could improve models accuracy, since jaguars are not strictly restricted to habitat patches remnants, but they are also observed using crop and pasture lands (Cullen Junior et al. 2005; Palmeira et al. 2008; Sollmann et al. 2012).

Finally, as information about potential current species distribution is essential for the implementation of conservation actions, only accurate and recent biological data (instead of data from natural history collections) should be used in SDMs in such conditions avoiding errors in model prediction (e.g. overprediction). As the application of SDM in megadiverse Neotropics presents several challenges, especially because data on species occurrences are scarce (Kamino et al. 2011), validation of model results – by experts as preferred – is extremely recommended as statistical results not always is coherent with the reality.


Ribeiro MC et al., 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biological Conservation, 142(6):1141-1153. http://dx.doi.org/10.1016/j.biocon.2009.02.021


Received: October 2012
First Decision: November 2012
Accepted: November 2012