

Ecological niche models, fossil record and the multi-temporal calibration for *Panthera onca* (Linnaeus, 1758) (Mammalia: Felidae)

Ana Karolina Mendes Moreno* and Matheus Souza Lima-Ribeiro

Laboratório de Macroecologia. Universidade Federal de Goiás - UFG. Regional Jataí. BR 364, km 195, nº 3800. Jataí, GO, Brazil (CEP 75801-615). *Email: akarolina_moreno@hotmail.com.

Abstract. Ecological niche modeling (ENM) has been used in large scale to identify locations with suitable environmental conditions for survival of a species in the geographical space. However, if training data do not identify the species' fundamental niche (e.g., in the case of species distribution is not currently in equilibrium with climate), ENMs may not reflect all appropriate locations it can survive, limiting or invalidating its general use in ecology. We predict that adding fossil data to the current species occurrence records, the called multi-temporal calibration approach, may improve the ENMs prediction by modeling a more complete suite of environmental conditions from species' fundamental niche, making them more robust and reliable. To test our prediction, we modeled the potential distribution of *Panthera onca* (Linnaeus, 1758) in New World using occurrence records from three subspecies (*P. o. onca*, *P. o. mesembrina* and *P. o. augusta*) and three time periods (Present, mid-Holocene and Last Glacial Maximum). Five different models were generated, the first with only current occurrence records, three others adding fossil data of *P. o. mesembrina*, *P. o. augusta* and *P. o. onca*, respectively, to the current data, and the last using all fossil and current data sets. Our findings support our prediction and show that fossil data improve the predictive ability of the models, informing complementary environmental conditions of species' fundamental niche and expanding its predicted potential distribution.

Keywords: Equilibrium species-climate, Fundamental niche, Fossil data, Last glacial maximum, Mid-Holocene.

Introduction

Understanding the species distribution in geographic space has currently become an important task in the fields of ecology, biogeography and conservation biology. Characterizing the environmental conditions that are appropriate for survival of the species and mapping these environments in geographic space has been a pronounced strategy to estimate potential distribution of species in different time periods, which is called ecological niche modeling (ENM) (Peterson

et al., 2011; Pearson, 2007). ENM is based on the theory of ecological niche and consist in estimate the multidimensional space (in a Hutchinsonian sense; Hutchinson, 1957) that allows the presence of species with positive population growth (Soberón, 2010). In accordance to Hutchinson (1957), the fundamental niche expresses the entire environmental conditions necessary for the survival and reproduction of individuals. However, the species is unable to occupy all that space due to the barriers to dispersion and biotic interactions, which constrain the locations

Received
August 26, 2015

Accepted
September 20, 2015

Released
December 31, 2015



Open Access
Full Text Article



where the species is in fact found in a given time period and delimit its realized niche.

Among ENMs, the correlative models conceptually estimate the environmental space suitable for the species survival (i.e., the species fundamental niche) by associating species occurrence records with environmental layers expressed in GIS and identify locations on the geographic space that, according to abiotic dimension, are similar to where species are observed (Pearson, 2007). But, depending on the method used to link occurrences to the environmental conditions, as well as factors such as barriers to dispersion and biotic interactions that limit the space of fundamental niche a species really occupy, these models can generate contrasting results and identify different extents of species distribution (Jiménez-Valverde et al., 2008). These effects on model predictions are still stronger when species are not in equilibrium with climate; i.e., the species occurrences do not inform its fundamental niche (Lima-Ribeiro and Diniz-Filho, 2013; Nogués-Bravo, 2009). To solve this problem we suggest adding fossil data to the current species occurrence records, the called multi-temporal calibration approach (Nogués-Bravo, 2009), together with paleoclimatic data. The multi-temporal calibration approach built models with a more complete suite of environmental conditions from species' fundamental niche informed from species data in different time periods (Nogués-Bravo et al., 2008). We predict then that using fossil and paleoclimatic data together with current data (multi-temporal calibration) help us to improve predictions of ENMs and provide a more realistic distribution of the species in geographic space.

To test our prediction, we used the American Jaguar (*Panthera onca* (Linnaeus, 1758)) as a model species. *Panthera onca* is found in a large variety of habitat, like tropical moist forests, xeric shrub lands, tropical dry forests, grasslands and savannas (Zeller, 2007). Its historical occurrence area was from the extreme southwest USA to the southern Argentina (Guggisberg, 1975). Currently its geographic range is declining in the whole America, principally by human activity and

climate change (Zeller, 2007). The subspecies *P. o. onca* is the largest extant feline of the Americas and the only living representative of the genus *Panthera* in the New World (Nowell and Jackson, 1996). Two other subspecies, *P. o. augusta* and *P. o. mesembrina*, known as the North American Jaguar and South American Jaguar, respectively, lived during the Pleistocene and are currently extinct.

To model the niche of *P. onca* by following the multi-model calibration approach, we used fossil data of the three subspecies (*P. o. onca*, *P. o. augusta* and *P. o. mesembrina*) in addition to paleoclimatic data of the mid-Holocene and Last Glacial Maximum (LMG), together with its current occurrence records. The findings supported our prediction and showed that fossil data provide better representation of species niche in all America, with particularities among different methods (algorithms).

Methods

Data of species occurrences

Current occurrence records of *P. onca* (including subspecies *P. o. centralis*, *P. o. goldmani*, *P. o. hernandesii*, *P. o. veraecrucis*, *P. o. peruviana*, *P. o. palustris*, *P. o. arizonensis* and *P. o. onca*) were obtained from online database "Global Biodiversity Information Facility - GBIF" (<http://www.gbif.org>) and "SpeciesLink" (<http://splink.cria.org.br>). Fossil data were obtained from online databases "The Paleobiology Database" (<https://paleobiodb.org>), "Fossilworks" (<http://fossilworks.org>) and "Florida Museum of Natural History" (<https://www.flmnh.ufl.edu/>), besides the grey literature (PhD thesis, reports, meeting annals, etc) and scientific literature (books and articles). We separated the fossil records according to each subspecies, *P. o. augusta*, *P. o. mesembrina* and *P. o. onca*, and then classified them belonging to LGM or mid-Holocene time periods (Figure 1).

All occurrence records were revised for synonymies and mapped on a grid cells with 0.5° resolution (~ 55 x 55 km at the equator). Duplicated points were removed

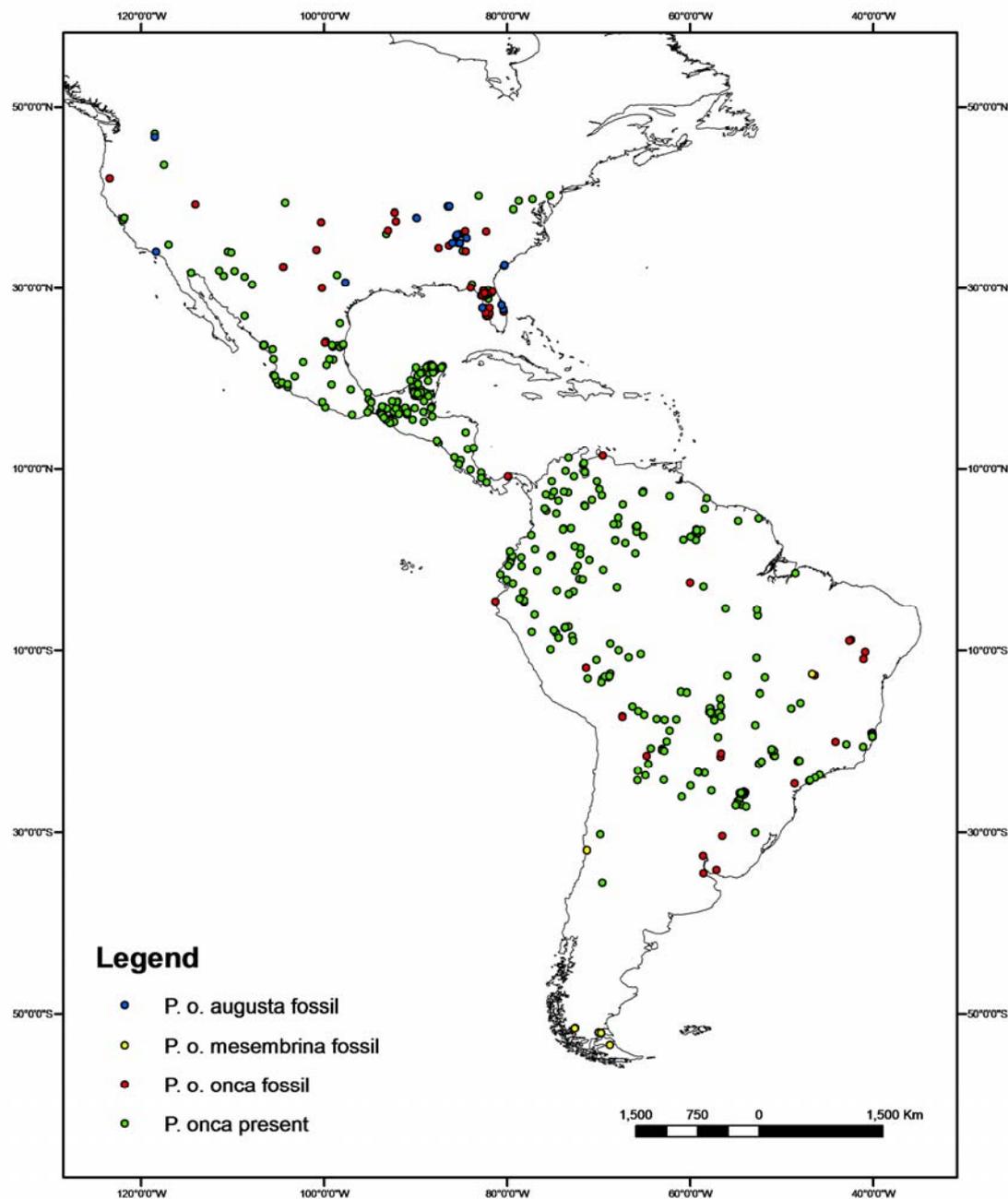


Figure 1. Current and fossil occurrence records of *P. onca* used for niche modeling.

into unique cells, generating a matrix of 292 current presences and 71 fossil records: 52 for *P. o. onca*, 13 for *P. o. augusta* and 6 for *P. o. mesembrina* (Figure 1).

Climate data

The climate data used for construction of ENMs were obtained from the coupled atmosphere-ocean general circulation model (AOGCM) MIROC Braz. J. Biol. Sci., 2015, v. 2, n. 4, p. 309-319.

available at the ecoClimate database (<http://ecoclimate.org>). EcoClimate provides climate simulations for important periods, like Last Glacial Maximum (LGM, 21 ka) and mid-Holocene (6 ka), which were respectively the cooler period (glacial) and warmest period (interglacial) of the last glacial cycle (Lima-Ribeiro et al., 2015).

We downloaded the 19 bioclimatic variables with 0.5° resolution for pre-

industrial (i.e., present), mid-Holocene and LGM. However, many of these bioclimatic variables are correlated, affecting the preference of the species when your niche is modeled and overestimating its distribution in geographic space. To minimize collinearity problems when building the ENMs, the 19 bioclimatic variables were ordered using a factor analysis with Varimax rotation and selected the variables with greater commonality along five factors (see detail in Terribile et al., 2012). The selected variables were: maximum temperature of warmest month, temperature annual range, precipitation of wettest month, precipitation of driest month and precipitation of warmest quarter.

Niche modeling

To estimate the niche of *P. onca* following the multi-temporal calibration approach, five models were built by combining data sets in accordance with current and fossil occurrence records:

- 1 - Only current *P. onca* records
- 2 - Current *P. onca* + fossil *P. o. mesembrina*
- 3 - Current *P. onca* + fossil *P. o. augusta*
- 4 - Current *P. onca* + fossil *P. o. onca*
- 5 - Current *P. onca* + all fossils

We used three algorithms based on presence-only and presence-background methods: Bioclim, Support Vector Machine (SVM) and Maximum Entropy (Maxent). Models were built by linking current and fossil occurrence records with climate conditions from their respective ages and then the potential distribution of *P. onca* was projected for present, mid-Holocene and LGM. All algorithms were run in the R-package “dismo” (Hijmans et al., 2015) using specific functions matching each algorithm, except SVM (run using the function *ksvm* from R-package “kernlab” (Karatzoglou et al., 2004). The Bioclim from R-package *dismo* computes the suitability of a location by comparing the values of environmental variables at any location to a percentile distribution of the values at known locations of occurrence (“training data”, the closer to the 50th

percentile - the median -, the more suitable the location is) instead of predicts binary presences and absences from traditional rectangular envelopes (see Hijmans et al., 2015). The machine-learning methods were fitted using the simplest configurations for classification tasks, like linear kernel function and c-SVC type for SVM (with probabilistic output) and only the linear features to combine climatic predictors in Maxent (with logistic output). Because we are using fossil records to predict potential distribution, simple modeling configurations are recommended (see Varela et al., 2011). All other parameters from SVM (e.g., cost parameter - C) and Maxent (e.g., beta, gamma, and prevalence; see Elith et al., 2006) were used as in default.

Background points were randomly selected throughout the New World grid cells (excluding cells with presences) keeping species prevalence equal to 0.5. This approach was based on recent studies suggesting that the extent of the geographical region in which the background points are taken have important influences for prediction and performance of ENMs (e.g., Thuiller et al, 2004; VanDerWal et al, 2009). We used the entire study area to select background points because there is no reason expected to prevent species dispersal throughout New World (i.e., we considered the entire New World as historically accessible for *P. onca*, see Barve et al., 2011). Thus, selecting background points throughout the New World essentially represents a compromise between generating models that do not generalize well, and do not produce over predictions of distribution areas that ignore important spatial structure associated with finer scale environmental gradients (VanDerWal et al., 2009). Background points were used to calibrate the ENMs based on presence-background observations (SVM and Maxent) and compute confusion matrix to evaluate model performance (see below).

Models were built using 75% of presence as training data and 25% as test data to evaluate them. To avoid biases from data partitioning on model training and evaluation, this procedure was repeated 20 times for each algorithm and data set. Thus,

a total of 300 models were generated (5 data sets * 3 algorithms * 20 repetitions) and evaluated using the True Skilled Statistic (TSS) and Area Under the roc-Curve (AUC) (Allouche et al., 2006). It is important to use more than one metric to evaluate the models' performance, because the statistical adjustments of each algorithm are different and each model quantifies a different aspect of the predictive performance (Elith and Graham, 2009). AUC is a threshold-independent metric and its values vary from 0 to 1; values above of 0.75 we consider that the model has good predictive ability and values below 0.5 indicate bad performances, equivalent to a random model. TSS is a threshold-dependent metric, but not sensitive to species prevalence specified in the modeling process (Allouche et al., 2006), with values ranging between -1 and +1; good models should present TSS values above 0.5 and below this value indicates a model with low predictive capacity.

The value of threshold was chosen for maximizing the sum of specificity and

sensitivity and then applied to both evaluate models and compute binary maps. To evaluate whether the increment of fossil data improves the predictive performance of the models, we computed a factorial Analysis of Variance (ANOVA) using AUC or TSS values as dependent variables and data sets and ENMs as factors. Finally, consensus maps reflecting the species potential distribution were generated for present, mid-Holocene and LGM, indicating the frequency in which grid cells were predicted as species presence.

Results

The values of AUC and TSS show acceptable performances for the models, better than a random classification for example (values often higher than 0.8 for AUC and 0.5 for TSS, except for some Bioclim models). The factorial ANOVA showed a significant difference in both AUC and TSS by adding fossil data to the ENMs (Table 1).

Table 1. Test of significance for TSS and AUC.

Effect	SS _(TSS/AUC)	DF	MS _(TSS/AUC)	F _(TSS/AUC)	P
Dataset	0.13/0.05	4/4	0.03/0.01	7.82/9.0	< 0.001
Method	2.22/0.57	2/2	1.11/0.29	266.46/213.7	< 0.001
Dataset*Method	0.17/0.04	8/8	0.02/0.005	5.08/3.4	< 0.001
Residual	1.19/0.38	285/285	0.004/0.001		

SS: Sum of Squares; DF: Degree of freedom; MS: Mean square. DF and P are equal for TSS and AUC.

The effects of adding fossil records from the five data sets are similar between AUC and TSS (Figure 2). Higher values of both AUC and TSS are presented for the models built with only current occurrences of *P. onca*, with a steady decline when added the fossil data to the current date. The lowest values for both AUC and TSS were observed for models built using all fossil data.

However, when we analyze the interaction between the factors method (algorithms) and data sets we note that Bioclim is the only one affected by adding fossil data (Figure 3). The algorithms Maxent and SVM presented slight differences, not significant, for both AUC and TSS among models built using

different data sets, although AUC is most affected than TSS for Maxent (Figure 3). Bioclim presented the worst performance among the three algorithms, with the lowest values for both AUC and TSS. Maxent and SVM presented higher values, with similar performances.

The effects of fossil data on model prediction were also detected on potential distributions through time (Figure 4). When we insert all fossil data, the predicted distribution of the species expands towards the northern edge in all periods. The species was predicted to occur in almost all New World across present, mid-Holocene and LGM, with slight reduction of potential distribution during the cold phase of LGM (Figure 4).

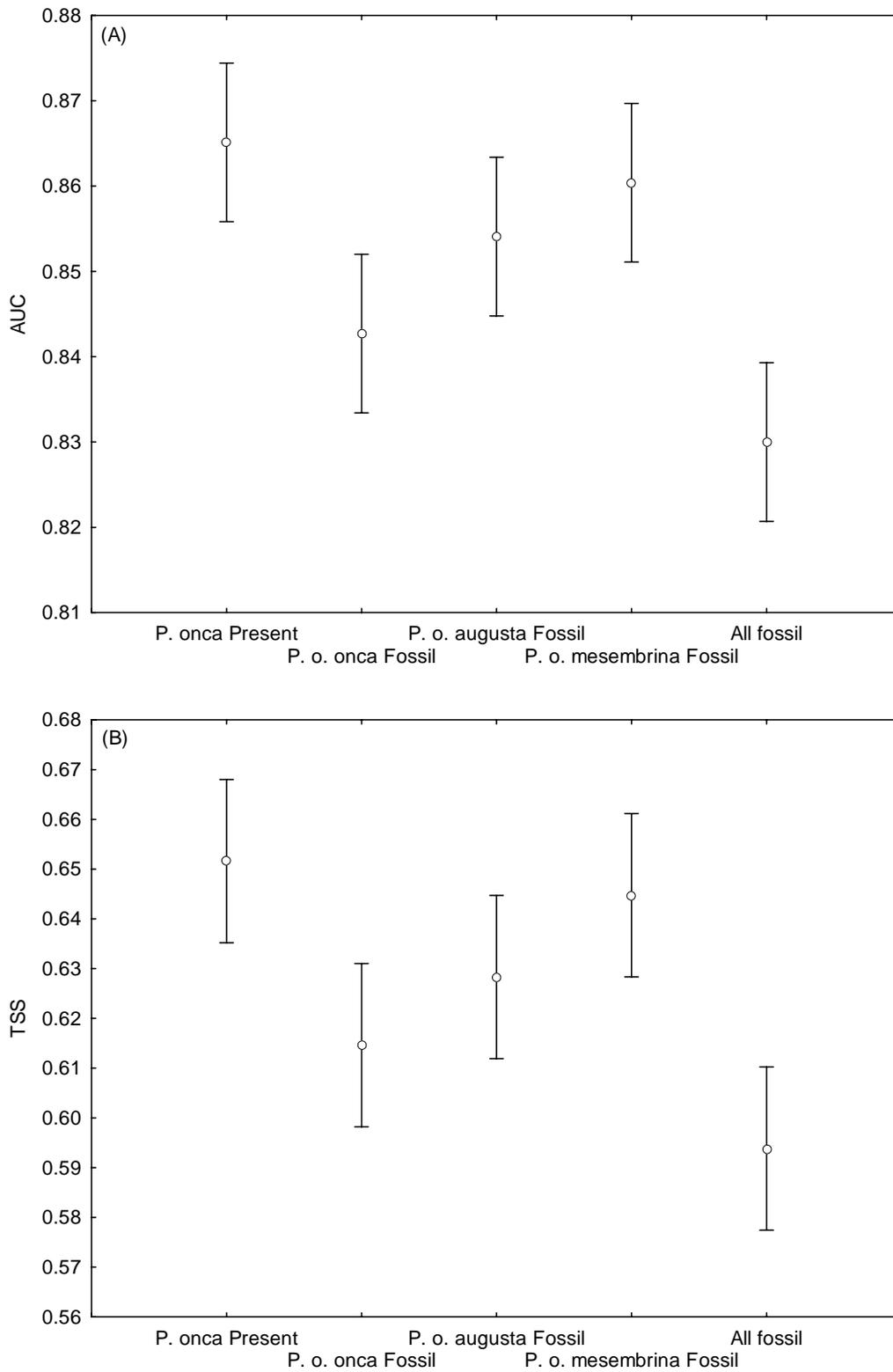


Figure 2. AUC (A) and TSS (B) values for ENMs built with the five data sets.

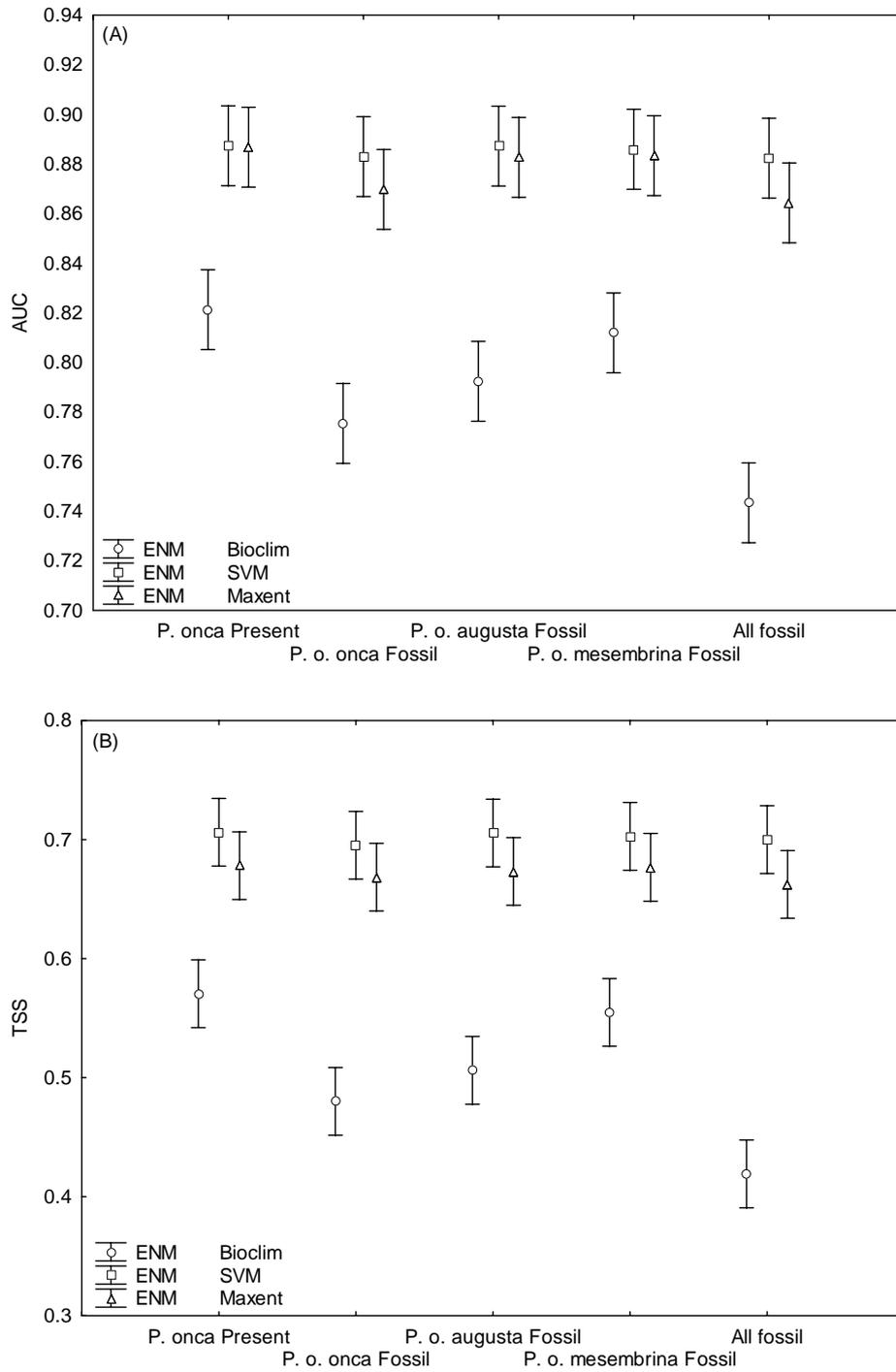


Figure 3. Effects of adding fossil data on ENM performances for three algorithms. (A) AUC, (B) TSS.

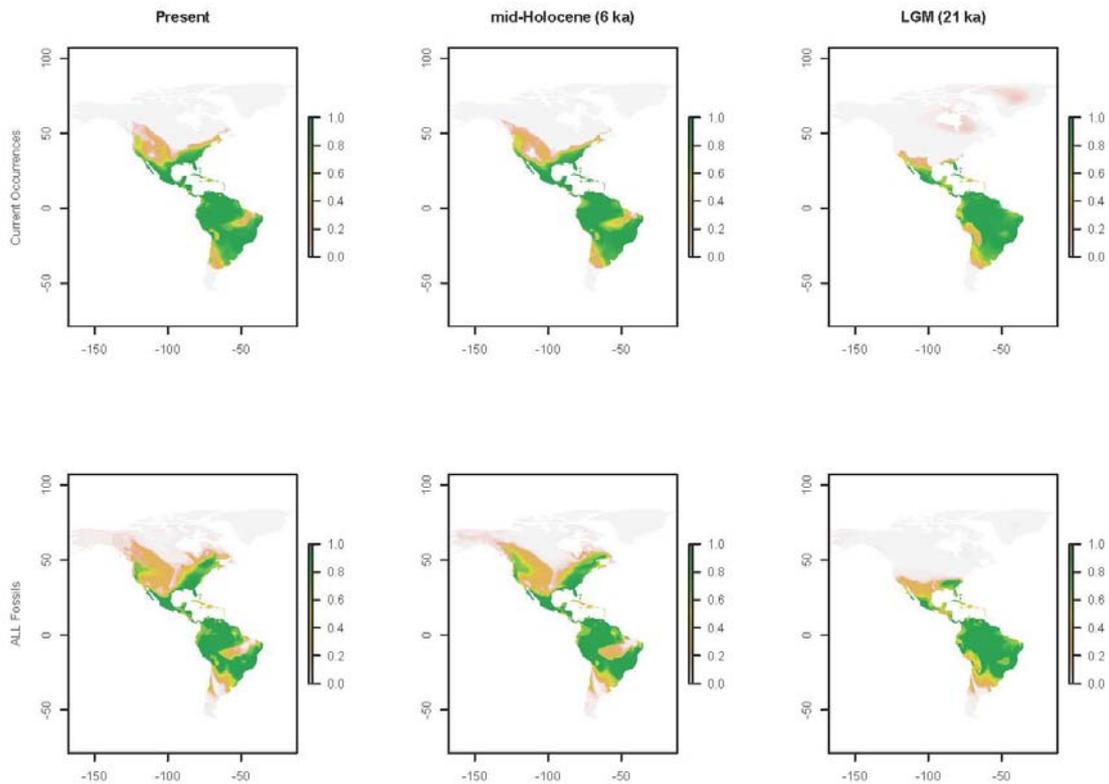


Figure 4. Consensus maps showing the frequency of models predicting species presence in the three time periods. Above, models built with only current data; and below, models with current data + all fossil data.

Discussion

Our findings support our prediction and show clear effects of fossil records on the ENM predictions for *P. onca*. As fossil information becomes greater, the predicted potential distribution expands and reduces the apparent performance of models as measured by AUC and TSS. However, the apparent decrease of model performance is artificially generated by the training and test data used for calibrating and evaluating the models. Because fossil records used as training data inform different climate conditions of species fundamental niche than current occurrences, the models predict suitable locations where the species is not observed from test data (which were drawn only from current data set), increasing the commission error. In other words, the models with fossil data predict presence where the species was not observed, because the increase of the predicted area. Because AUC and TSS are based on

success rates (sensitivity and specificity), their values artificially decrease with addition of fossil data. Fossil data, therefore, improve the actual model predictions by furnishing information closest to the species fundamental niche, regardless of the absolute values of AUC and TSS.

By following this reasoning, we expect that the effects of different data sets increase as better it represent the species fundamental niche, as reported in Figure 5. From the statistical values of AUC and TSS, which gradually decrease with the insertion of data fossil, more the fundamental niche becomes best represented, because a larger amount of fossil data set were able to better predict the most locations climatically favorable for survival of the species than a small fossil data set (Figure 5). When a small amount of fossil data is inserted, as in the case of the insertion of fossil *P. o. mesembrina* and *P. o. augusta*, for example, models predict a

slight different niche than present-only data set. For these two subspecies, the models begin to gradually decrease its performance by predicting new places climatically suitable for the species, although just few places. When a large quantity of fossil data was inserted, as in the case of the insertion

of *P. o. onca*, we note that there is a more significant increase of predicted area and when we insert all fossil data, the fundamental niche of the species increases considerably and expanding even more its predicted potential distribution.

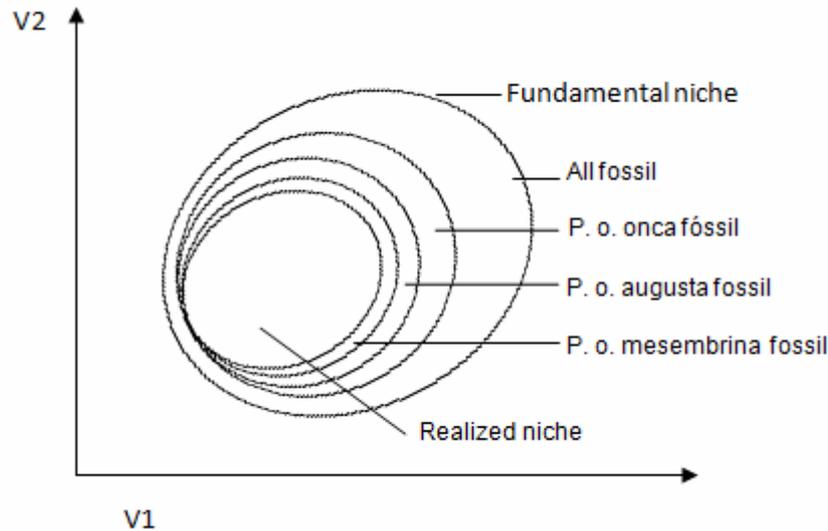


Figure 5. Schematic representation of species fundamental niche and which dimension the fossil data inform for niche modeling. When adding fossil data, particularly all fossil, the fundamental niche of the species became more representative, informing new locations climatically favorable for *P. onca* survival.

Thus, models built using fossil data from the three subspecies better represent the fundamental niche of *P. onca* than its occurrence records at the present-day. Actually, our findings highlight the need for considering occurrence data from different periods to better inform models about species preferences in a multi-temporal calibration approach. Many species are not in equilibrium with the climate currently; i.e., not occupy all the places climatically suitable for its survival and are absent from all areas inadequate (Araújo and Pearson, 2005). By being in disequilibrium with the climate, the niche model using only current data can underestimate the environmental conditions (its fundamental niche) under which species will be able to occur (Dormann, 2007), and partially identify its geographical distribution. Using fossil data and paleoclimatic simulations along with the current occurrence data of the species, what we call multi-temporal calibration (Nogues-Bravo et al., 2008), may avoid

Braz. J. Biol. Sci., 2015, v. 2, n. 4, p. 309-319.

such problems by providing a better estimate of climatically suitable conditions for species survival (i.e., better represent species fundamental niche). On the contrary, when the species is in equilibrium with the climate in present-day, using the multi-temporal calibration approach no longer has effect on model predictions because the species already occupies all the places that are suitable to it, its fundamental niche would already be represented. Thus, because *P. onca* is in disequilibrium with the climate, the use of multi-temporal calibration became important and indispensable to better represent its fundamental niche and predict its potential geographical range in which its populations may survive through time.

Our study shows that using fossil data as a complement of current data, in a multi-temporal calibration approach, improve the ENMs, inform new conditions species can survive, better represent the fundamental niche of species and

contributing to better predict the climatically suitable areas for species occurrence on geographic space.

Acknowledgement

Research projects of MSLR are supported by CNPq (processes 563727/2010-1 and 473811/2013-8) and FAPEG (process 2012/1026.700.1086).

Conflict of interest statement

Authors declare that they have no conflict of interests.

References

- Allouche, O.; Tsoar, A.; Kadmon, A. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). **J. Appl. Ecol.**, v. 43, p. 1223-1232, 2006.
- Araújo, M. B.; Pearson, R. G. Equilibrium of species' distribution with climate. **Ecography**, v. 28, p. 693-695, 2005.
- Barve, N.; Barve, V.; Jiménez-Valverde, A.; Lira-Noriega, A.; Maher, S. P.; Peterson, A. T.; Soberón, J.; Villalobos, F. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. **Ecol. Model.**, v. 222, p. 1810-1819, 2001.
- Dormann, C. F. Promising the future? Global change projections of species distributions. **Basic Appl. Ecol.**, v. 8, p. 387-397, 2007.
- Elith, J.; Graham, C. H.; Anderson, R. P.; Dudík, M.; Ferrier, S.; Guisan, A.; Hijmans, R. J.; Huettmann, F.; Leathwick, J. R.; Lehmann, A.; Li, J.; Lohmann, L. G.; Loiselle, B. A.; Manion, G.; Moritz, C.; Nakamura, M.; Nakazawa, Y.; Overton, J. M.; Peterson, A. T.; Phillips, S. J.; Richardson, K.; Scachetti-Pereira, R.; Schapire, R. E.; Soberón, J.; Williams, S.; Wisz, M. S.; Zimmermann, N. E. Novel methods improve prediction of species' distributions from occurrence data. **Ecography**, v. 29, p. 129-151, 2006.
- Elith, J.; Graham, C. H. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. **Ecography**, v. 32, p. 66-77, 2009.
- Guggisberg, C. A. W. **Wild cats of the World**. New York: Taplinger Publications, 1975.
- Hijmans, R. J.; Phillips, S.; Leathwick, J.; Elith, J. **dismo: Species distribution Modeling**. R package version 1.0-12. 2015. Available from: <<https://cran.r-project.org/web/packages/dismo/>>. Accessed in Jun. 20, 2015.
- Hutchinson, G. E. Concluding remarks. **Cold Spring Harb. Symp. Quant. Biol.**, v. 22, p. 425-427, 1975.
- Jiménez-Valverde, A.; Lobo, J. M.; Hortal, J. Not as good as they seem: the importance of concepts in species distribution modelling. **Diversity Distrib.**, v. 14, p. 885-890, 2008.
- Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab - An S4 package for kernel methods in R. **J. Stat. Softw.**, v. 11, p. 1-20, 2004.
- Lima-Ribeiro, M. S.; Diniz-Filho, J. A. F. **Modelos ecológicos e a extinção da megafauna: clima e homem na América do Sul**. São Carlos: Cubo, 2013.
- Lima-Ribeiro, M. S.; Varela, S.; González-Hernández, J.; Oliveira, G.; Diniz-Filho, J. A. F.; Peterson, A. T.; Terribile, L. C. The EcoClimate Database. Available from: <<http://ecoclimate.org>>. Accessed in Jun. 20, 2015.
- Nogués-Bravo, D. Predicting the past distribution of species climatic niches. **Global Ecol. Biogeogr.**, v. 18, p. 521-531, 2009.
- Nogués-Bravo, D.; Rodríguez, J.; Hortal, J.; Batra, P.; Araújo, M. B. Climate change, humans, and the extinction of the woolly mammoth. **PLoS Biol.**, v. 6, n. 4, e79, 2008. Available from: <<http://www.plosbiology.org/article/doi/10.1371/journal.pbio.0060079&representation=PDF>>. Accessed in Jun. 20, 2015.
- Nowell, K.; Jackson, P. **Wild cats: status survey and conservation action plan**. Gland, Switzerland: Cat Specialist Group, Wild Conservation Union/Species Survival Commission, 1996.
- Pearson, R. G. **Species' distribution modeling for conservation educators and practitioners**. Synthesis. New York: American Museum of Natural History, 2007. Available from: <<http://academic.uprm.edu/~jchinea/UIP-MAPR/refs/modelos/pearson2008.pdf>>. Accessed in Jun. 20, 2015.
- Peterson, A. T.; Pearson, R. G.; Anderson, R. P.; Martínez-Meyer, E.; Nakamura, M.; Araújo, M. B. **Ecological niches and geographic distributions**. Oxford: Princeton University Press, 2011.
- Soberón, J. Niche and area of distribution modeling: a population ecology perspective. **Ecography**, v. 33, p. 159-167, 2010.
- Terribile, L. C.; Lima-Ribeiro, M. S.; Araújo, M. B.; Bizão, N.; Collevatti, R. G.; Dobrovolski, R.; Franco, A. A.; Guilhaumon, F.; Lima, J. S.; Murakami, D. M.; Nabout, J. C.;

Oliveira, G.; Oliveira, L. K.; Rabelo, S. G.; Rangel, T.; Simon, L. M.; Soares, T. N.; Telles, M. P. C.; Diniz-Filho, J. A. F. Areas of climate stability of species ranges in the Brazilian Cerrado: disentangling uncertainties through time. **Nat. Conserv.**, v. 10, p. 152-159, 2012.

Thuiller, W.; Brotons, L.; Araújo, M. B.; Lavorel, S. Effects of restricting environmental range of data to project current and future species distributions. **Ecography**, v. 27, p. 165-172, 2004.

VanderWal, J.; Shoo, L. P.; Graham, C.; Willians, S. E. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? **Ecol. Model.**, v. 22, p. 589-594, 2009.

Varela, S.; Lobo, J. M.; Hortal, J. Using species distribution models in paleobiogeography: a matter of data, predictors and concepts. **Palaeog. Palaeoclima. Palaeocol.**, v. 310, p. 451-463, 2011.

Zeller, K. **Jaguars in the new millennium data set update**: The state of the jaguar in 2006. New York: Wildlife Conservation Society, 2007.