Reassessment of the Extinction Risk of Endemic Species in the Neotropics: How can Modelling Tools Help us?

Luciana Hiromi Yoshino Kamino[1]*, Marinez Ferreira de Siqueira[2],
Andrea Sánchez-Tapia[2] & João Renato Stehmann[1]

[1] *Departamento de Botânica, Universidade Federal de Minas Gerais – UFMG, Belo Horizonte, MG, Brazil*

[2] *Instituto de Pesquisa Jardim Botânico do Rio de Janeiro – JBRJ, Rio de Janeiro, RJ, Brazil*

Abstract

Occurrence data on geographic distribution for most endemic Neotropical plants are scarce and spatially biased, which interferes with decisions on their conservation status. In this study we present a three-step approach applying: 1) Euclidean Environmental Distance models to search for unknown populations; 2) MaxEnt incorporating additional presence data to produce forecasts of the potential distribution and 3) coupling of convex hull methods with SDM. We studied *Petunia mantiqueirensis*, an Atlantic Forest endemic species. This approach enabled us to obtain new presence points, and to improve species distribution models. Finally, we identified and excluded environmental unsuitable areas inside the extent of occurrence, allowing us to reassess the conservation status of the species. These recommendations are important to crucial to establish, in a more assertive way, the extent of occurrence and area of occupancy of the species to determine proper strategies for conservation.

Key words: Euclidean Environmental Distance, MaxEnt, Model Validation, Presence-Only, Small Sample Size, Species Distribution Model, Threatened Species.

Introduction

Endemic and rare species are a priority on conservation plans since they are exclusive to one or few localities and show higher extinction risk when compared to other species (McKinney 1999). Assessing the conservation status of a species is one of the main tools to establish conservation priorities toward the protection of the threatened biological diversity (De Grammont & Cuarón 2006). The main criteria used to evaluate the extinction risk of species have been trends in population size and/or geographical distribution range (Paglia & Fonseca 2009). However, these data are incomplete for most Neotropical species due to the small number of data points and/or spatially biased records, which makes it difficult to decide on the conservation status of species (Kamino *et al.* 2012).

In this scenario, species with very few occurrences in combination with species distribution model (SDMs) methods may be useful for providing a description of the areas climatically similar to the recorded presences, allowing the planning of new field surveys (Kamino *et al.* 2012). These methods are an important tool to make predictions about distribution of poorly known species which are, probably, under risk of extinction (Raxworthy *et al.* 2003). SDMs associate primary data in order to determine suitable environments where species might keep a viable population in a particular moment or a discrete period in time (Jimenez-Valverde *et al.* 2008). This relationship can be mapped to predict a potential geographic distribution. Maps produced may reduce fieldwork costs and time by directing sample efforts to suitable areas and avoiding sampling in unsuitable and/or well sampled areas (Pearson *et al.* 2007).

Paucity of data, spatial inaccuracy, and lack of valid absences are the main reasons identified as shortcomings to the production of useful models for rare and endemic species (Engler *et al.* 2004). Standard sampling methods such as simple or stratified random sampling, based on a simple combination of the main environmental gradients, can be highly inefficient to these species (Guisan *et al.* 2006). This iterative process can be repeated during the same field season or over consecutive seasons, and its success is mainly measured by the number of new occurrences found at each period (Guisan *et al.* 2006). However, sample sizes are most often below the minimum necessary to generate a model. This is the first factor that limits the potential of most species of Neotropical plants to be modeled (Kamino *et al.* 2012).

Here we present a three-step approach to study species with small sample size using a technique based on the environmental distance as a tool for a more efficient sampling

*Send correspondence to: Luciana Hiromi Yoshino Kamino
Departamento de Botânica, Instituto Ciências Biológicas,
Universidade Federal de Minas Gerais – UFMG, CP 486,
CEP 31270-901, Belo Horizonte, MG, Brazil
E-mail: lucianakamino@gmail.com

design to guide new collections. This is accomplished through the use of these distance values in the orientation of searches performed in areas with different environmental similarity values to ensure that the new sampling points are able to confirm if species are poorly collected or rare and/ or endemic in the field. From this new information, it is possible to apply more complex models, such as MaxEnt, aiming now to infer the geographical distribution of species. Allied to this, incorporating a convex polygon technique is important to get an area of occurrence for the species that is much closer to the actual distribution than simply by applying SDMs to identify potentially suitable sites. Improving estimation of the niche size and of the distribution area of the species and information about habitat restriction are very important for conservation purposes.

Material and Methods

Study area and data for modelling

The study area is located on the Mantiqueira mountain range, a central portion of the Atlantic Forest Domain in southeastern Brazil. The region comprises cloud forests, rain forests, semideciduous forests, *Araucaria* mixed forests, and high altitude grasslands on mountain peaks, as well as broad transition areas between these. The floristic composition and distribution patterns of vegetation physiognomies occurring within the region can be associated with rainfall regime, elevation and climatic variations (Oliveira-Filho & Fontes 2000). The climate is mesothermic, markedly seasonal, with cold and dry winters and wet summers. On a fine scale, there is a considerable diversity of pedoenvironments and associated vegetation mosaics, greatly determined by local topography and microenvironmental aspects with high rates of rare, microendemic and endangered species (Benites *et al.* 2007).

Petunia mantiqueirensis T. Ando & Hashim. (Solanaceae) is an annual herbaceous species and its flowers are showy and have a purple, tubulose-funnelform corolla (Figure 1). It blooms from October to January, but fructifies throughout the summer until March, when the plant dies. The species is restricted to the Mantiqueira mountain range, along mixed rain forest boundaries and open formations (Stehmann *et al.* 2009). There is controversy about its conservation status (Fundação Biodiversitas 2010), and this taxon has not yet been assessed for the IUCN Red List. We used eight presence points gathered from herbaria and the species protologue (Ando & Hashimoto 1994), for which the georeferenced points were surveyed.

Figure 1. *Petunia mantiqueirensis* T. Ando & Hashim.: a) view of the habitat, growing on slopes in the forest border, b) detail of the showy corolla; and c) decumbent habit.

We used 19 climate variables obtained from Worldclim (version 1.3, http://www.worldclim.org/); four topographic from USGS Hydro-1K Elevation Derivative Database (http://eros.usgs.gov/) and four layers of Vegetation Index Isolines (EVI Images) composed 16-day from satellite images NASA-MODIS with linear resolution of 0.25 km (https://earthexplorer.usgs.gov/). Environmental data were evaluated in relation to the degree of correlation between them and pairs of variables with correlation values greater than 0.8 were reassessed. A principal components analysis (PCA) was performed and the axis to which each pair was most correlated was determined. Of each pair, the variable with the highest eigenvalue along the axis was retained and the other discarded. We retained annual mean temperature, maximum temperature of warmest month, precipitation of wettest month, precipitation of the driest month, aspect, slope and two EVI Images. All analyses were conducted at the 0.25×0.25 km resolution.

Euclidean Environmental Distance Model

We used a Euclidean Environmental Distance Model (EEDM) to determine areas environmentally similar to the original points based on the distance classification of all pixels in the study area to the centroid niche calculated for the original eight points. EEDMs can be applied to few points (Siqueira *et al*. 2009), generating an environmental similarity map (varying from 0 to 1, where 0 is not similar and 1 totally similar) that can be used as a guide to locate new populations.

We applied the lowest presence training threshold (LPT) on the EEDM map in order to discern suitable and unsuitable areas for *P. mantiqueirensis*. The areas shown as not suitable (pixel values lower than LPT) were classified as zero and the areas considered suitable (pixel values above LPT) kept their similarity values. LPT was chosen according to Pearson *et al*. (2007) because this approach can be ecologically interpreted as identifying pixels predicted as being at least as suitable as those where the species has been recorded and to identify the minimum predicted area possible whilst maintaining zero omission error in the training data set, since all records were previously confirmed in the field.

EEDMs were evaluated using jackknife methods, developed by Pearson *et al*. (2007), which are effective for sample sizes of 25 or less. Each sampling location was removed once from the dataset and a model was built with the remaining data, successively for all points. A significance test was carried out to indicate success or failure of prediction, based on the proportion of the presence-predicted area (P_I) when one location is excluded from the variable (X_I). For each prediction P, P_I and X_I values were computed on the LPT of each generated model to assess predictions.

Field surveys and external test

We conducted field surveys in areas with a high range of environmental similarity, accordingly to the EEDM map (Figure 2), from 2008 through 2009, and concentrated along the species' blooming period. New occurrences were recorded and georeferenced using the GPS. If the species was not found, the area was considered as an absence location.

Presence or absence data gathered during field surveys were used to generate the confusion matrix, which describes the frequency in which presence or absence are correctly and incorrectly predicted (Fielding & Bell 1997). Measures of the omission/commission errors and True Skill Statistic (TSS) were calculated from these field points.

In order to analyze the range of sampling in the study area, a thousand random points in the study area was generated and environmental values for the eight variables under consideration were extracted. The final matrix had environmental values for presences, absences observed in the field and these random points. This matrix was standardized to correct for differences in scale and a PCA was performed, using package vegan (Oksanen *et al*. 2012) in R (R Core Development Team 2012). A multivariate, non-parametric (permutation-based) analysis of variance (Pillar & Orlóci 1996) implemented in Multiv (Pillar 2006) was performed to test for differences in environmental variables between presences, absences and random background points.

Maximum entropy

We initiated the second phase of modeling using MaxEnt (version 3.3.3) to produce a map of potential geographical distribution (Phillips *et al*. 2006). MaxEnt is a robust method and has been shown to perform well with limited samples size in comparison to alternative approaches (e.g. Hernandez *et al*. 2006). MaxEnt uses environmental data from occurrence records, and background samples, in order to estimate the ratio between them. It does this by making an estimate of distribution values for the presence records, which are consistent with occurrence data, choosing the distribution that is closest to the distribution of values for the background. Minimizing distance from background assumes that the species occupies environmental conditions proportionally to their availability in the landscape. Distance from the background is considered as the relative entropy of occurrence data with respect to background (Elith *et al*. 2011). The results obtained are shown as environmental suitability, with values ranging from 0-1. Again LPT and jackknife were applied. Omission/commission errors and TSS were calculated from field points.

Distribution range size

The extent of occurrence (EOO) has been used to calculate and classify species into a category of extinction risk, using a
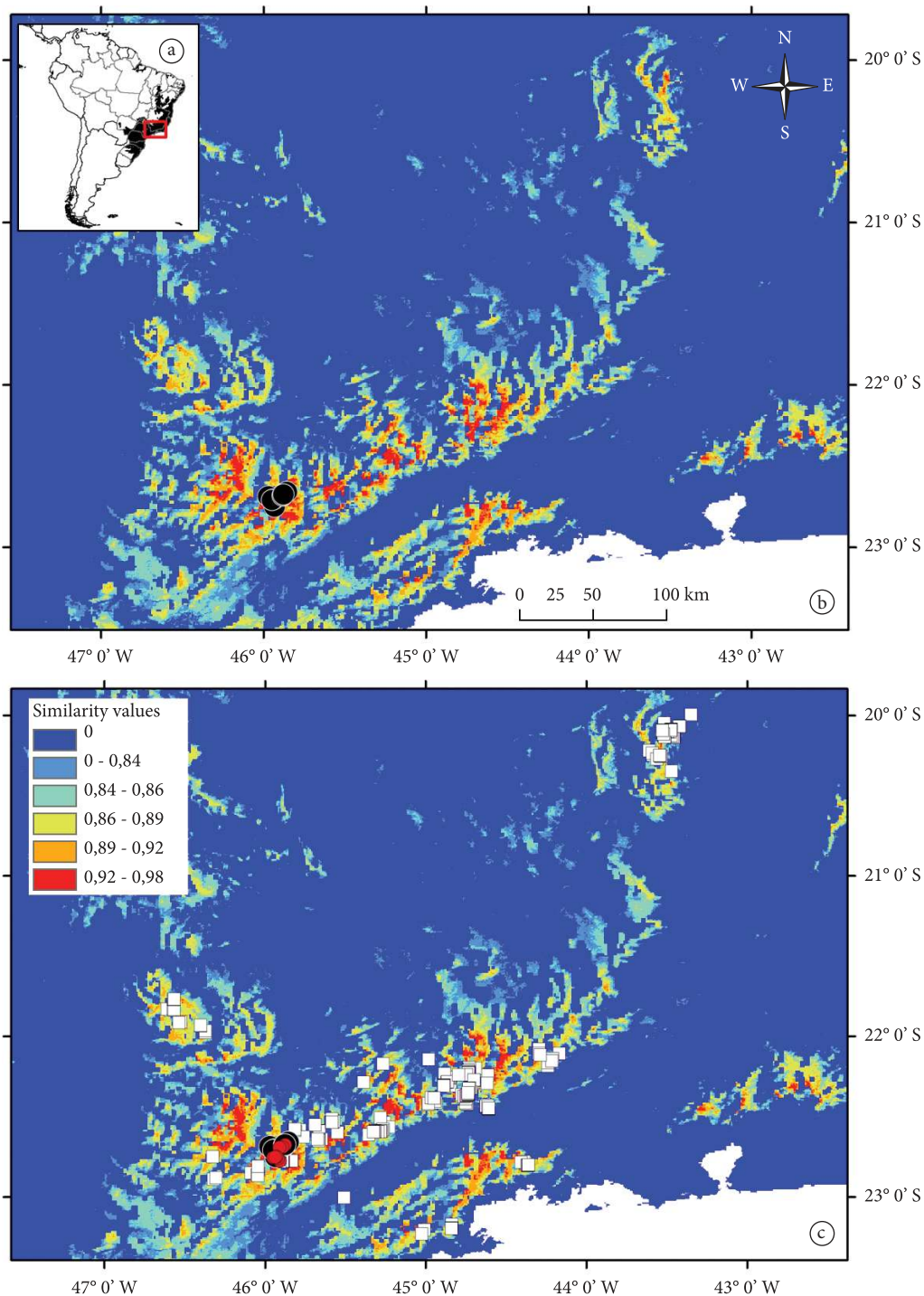
Figure 2. Potential geographical distribution for *Petunia mantiqueirensis* according to Euclidean Environmental Distance. Similarity values above the lowest training threshold are shown from light blue to red, values below this threshold are present in dark blue: (a) Study area highlighting the Atlantic Rain Forest Domain. (b) Environmental similarity map based on eight points of occurrences (black dots). (c) Red dots show localities where new populations were found; white squares are observed absences.

minimum convex polygon (IUCN & Petitions Subcommittee 2011). A second measure was calculated by extracting the area predicted as presence of MaxEnt using LPT inside of polygon, (hereafter we will refer to this as Environmental Suitability in Extent of Occurrence (ES-EOO)). Finally,

the area of occupancy (AOO) was estimated based on counting the number of occupied cells in a uniform grid (250 m × 250 m) covering the entire range of the species, and estimating the total area of all occupied cells (IUCN & Petitions Subcommittee 2011).

Results

EEDM and evaluating predictions

The first predictive model for *P. mantiqueirensis* (Figure 2a) yielded similarity distances of EEDMs in environmental space in the range of 0.545 (±0.007) – 0.989 (±0.003). Jackknife validation indicated high predictive success rate (87%) and significant models ($P < 0.01$). Average values of LPT, P_i e X_i used to evaluate the predictive capability of the models are shown in Table S1, available as Additional Supporting Information available at www.abeco.org.br.

Field surveys and external test

We conducted surveys at 229 sites and found ten new occurrences in the field (Figure 2c). New occurrences were found in areas characterized as similar to those where the species had been observed, maintaining zero omission error in the model. We registered 99 absence points in areas where the presence had been predicted, resulting in

an index of commission error about 45%. Other 120 points were registered as true absences in areas predicted as absence. Results of field validations showed TSS = 0.55. This means the initial data yielded models that predicted the species distribution of test points significantly better than at random. The EEDM predicted potential presence for the species in 5% of the study area. Of the eight test points, seven fell in pixels of predicted presence, and one in pixels of predicted absence (binomial probability, $P < 0.0001$). The unique original point that fell into a predicted absence was within 0.5 km of areas of predicted presence. According to the sampling effort, we can assume that the survey of the potential distribution area was satisfactory, since all sites sampled post-modeling are spread over a range of suitability values (0-0.980).

The PCA (Figure 3) showed that all presence points were concentrated in a small area of environmental space. The two first axis of the PCA explained 39.86% and 25.28% of variation, respectively. The permutation test (Table S2 in the Additional Supporting Information) showed no significant difference between presences and absences ($P = 0.204$) but
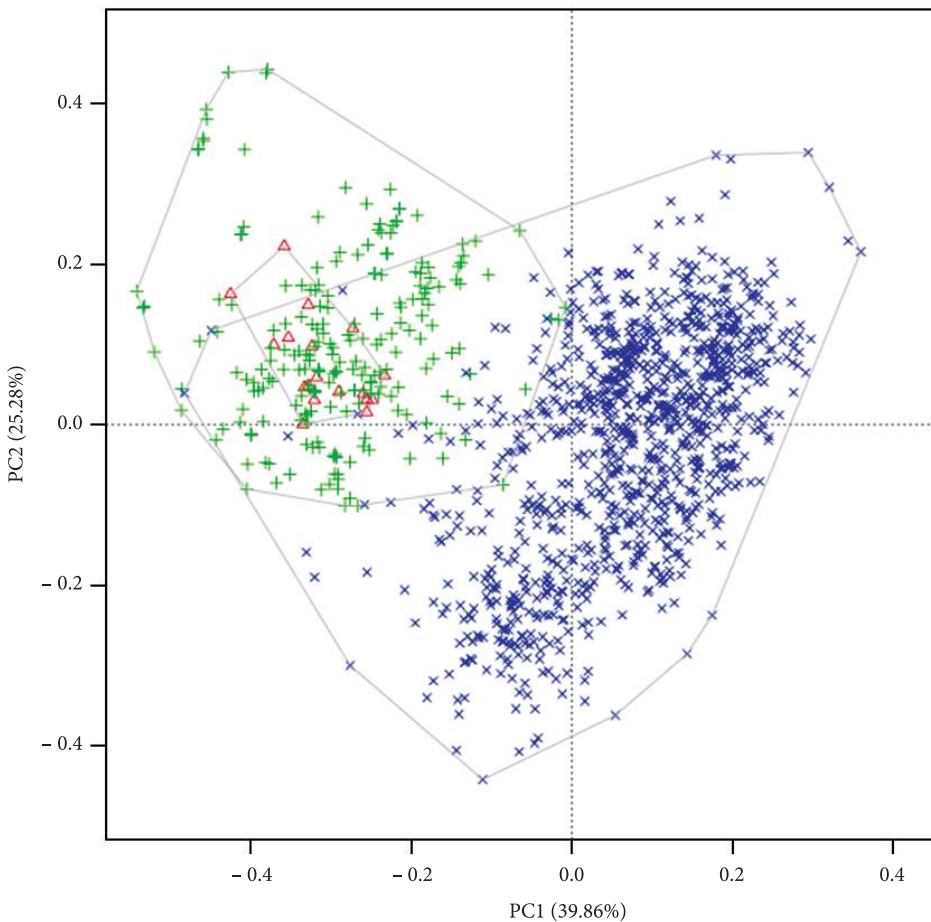

Figure 3. Principal components analysis (PCA) of environmental values associated to presence data (△), absence data (+) and 1,000 points randomly generated in the study area (×).

significant differences between background points and both presences and absences ($P = 0.001$ for both contrasts).

MaxEnt and range size estimate

The model generated by MaxEnt, based on all the points of presence and on the same predictors, showed a reduction of the area predicted for the presence of the species, eliminating several disjunct and continuous areas presented in the EEDM, indicating areas closest to the occurrence points (Figure 4a). Suitability values of MaxEnt in environmental space ranged 0–0.934 (±0.009) across the study area. The result of the jackknife validation showed a success rate of 78% ($P<0.01$) and average values of LPT, Pi e Xi used to
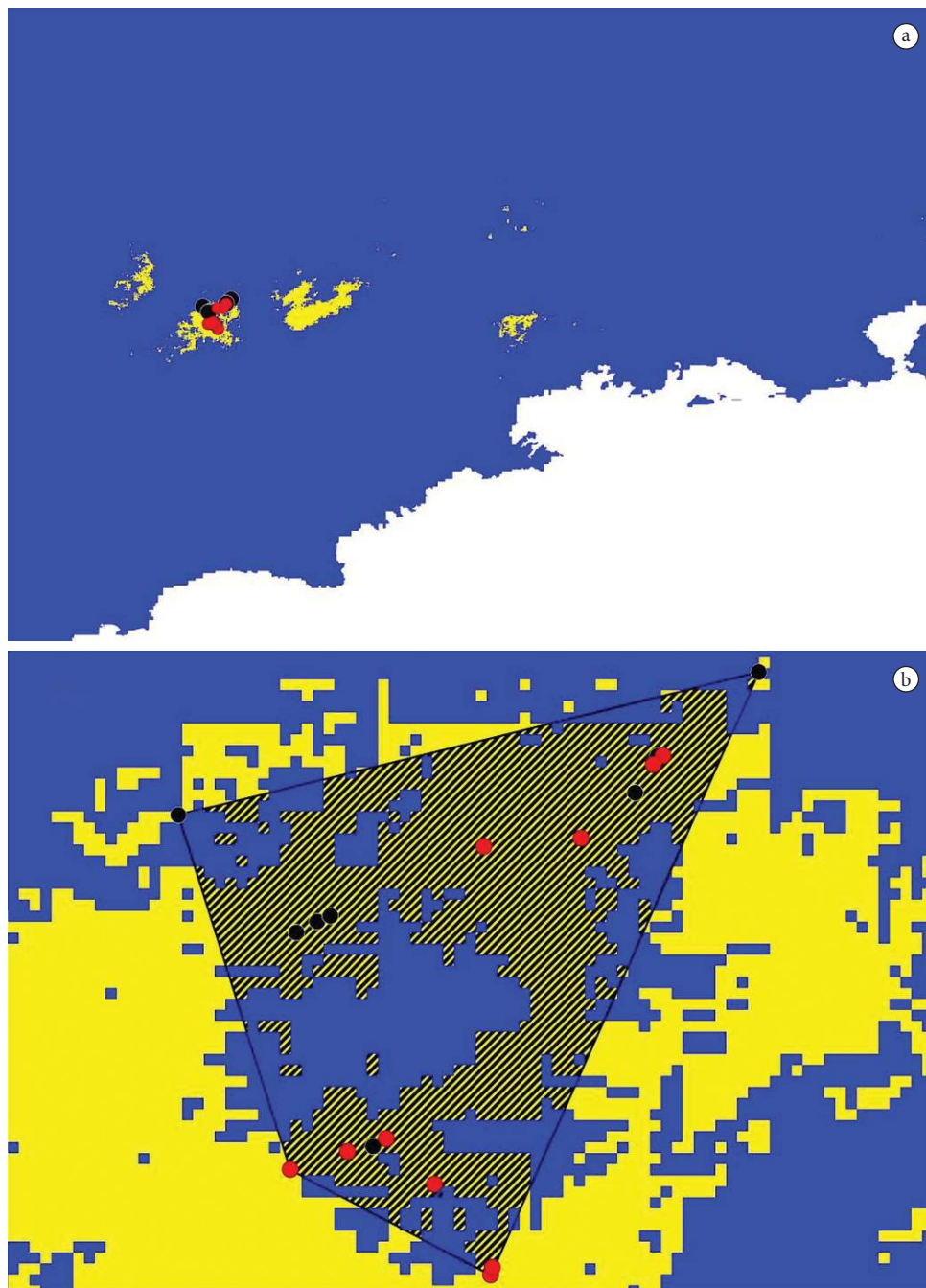

Figure 4. Maps showing potential geographical distribution for *Petunia mantiqueirensis* accordingly with MaxEnt. Suitability values above the lowest predicted threshold (LPT) in yellow values (1068 km^2) below this threshold in blue: (a) Map based on eight points of known occurrences (●) and 10 points of new occurrences (●) for model construction. (b) Detail of the minimum convex polygon including all sites of occurrence to estimate the extent of occurrence area (82 km^2) and the extent of occurrence to area predicted as presence according LPT (▨ 62 km^2).

evaluate the predictive capability of the models are shown on Table S3 (see Additional Supporting Information). Results of field validations showed TSS (0.64), comission error (13.8%) and omission error (22.2%) for MaxEnt models.

EOO was estimated at 82 km^2, ES-EOO was estimated at 62 km^2 (Figure 4b) and AOO was 1.4 km^2.

Discussion

The first goal of the present study was to apply a simple SDM technique to identify potentially suitable unexplored areas of occurrence. In the first phase, prediction maps obtained from distance-based SDM techniques calibrated with eight presence points were used to plan additional surveys in different areas of environmental similarity in order to improve the original data sets and determine the geographical and environmental boundaries of the species. In the following phase, with an improved dataset, we used a complex model (MaxEnt) with additional presence data to produce forecasts of the potential distribution of the species, making the approach suitable to identify new distribution areas and reassess the conservation status of the species.

Recently, similar approaches with consecutive phases using simple (EEDM) and complex models (GARP) have been published independently and also have found new occurrences in cases with poorly collected species (Siqueira *et al.* 2009). Here, our experimental design was useful in directing field survey efforts and allowed us to find additional data of an endemic species.

The final models show a decrease in the area predicted for presence, which was nearly restricted to the currently known distribution of the species. Jimenez-Valverde *et al.* (2008) assert that the use of complex techniques will overfit the presence data and this will result in predicted extents of occurrence that are smaller than those suggested by simpler techniques and thus, a greater number of the true absences in the validation data will be predicted as absences by complex techniques than by the simple ones. The second modelling still shows areas of environmental suitability disjunct from the core area of species occurrence, further studies could clarify the reason why the species is absent in these areas. The combined used of modeling techniques will create feedback loops of increasing data quality and quantity to produce SDMs that may be good enough for conservation (Engler *et al.* 2004).

The second goal of this study was to reevaluate the extinction risk of an endemic species. We identified additional occurrence sites for the species, and suggest that *P. mantiqueirensis* distribution could be classified as one form of rarity according to Rabinowitz (1981): constantly sparse and geographically restricted in a specific habitat. After defining EOO by minimum convex polygon using all presence data, we observed that potential geographic distribution by MaxEnt could indicate unsuitable areas to the species inside the EOO and reestimated its range

size. The ES-EOO allowed us to conclude that the area is environmentally heterogeneous and permitted to exclude unsuitable areas for this species.

With regards to the conservation status, we conclude that *P. mantiqueirensis* should be classified as Critically Endangered. This is justified by the species having localized dispersal ability, narrow habitat specificity, limited habitat availability, severely fragmented, geographic range in <100 km^2 extent of occurrence and <10 km^2 area of occupancy (IUCN & Petitions Subcommittee 2011). Thus, *ex-situ* conservation in germplasm banks is suggested, as well as encouraging the species cultivation taking advantage of its ornamental potential, and the creation of a reserve or the possibility of reintroduction in areas with high environmental suitability within the total area of occurrence.

Concluding Remarks

The geographical distribution of species may be limited to few samples, and models of environmental similarity associated with fieldwork may increase knowledge about current distribution. The use of these distance values in the orientation of searches performed in different environmental similarity values is important to ensure that the new sampling points are not just one environmental duplicate of the calibrate points. In a second phase of experimental design, the suitability map by MaxEnt showed a reduction of predicted area, including areas inside the EOO. This allowed us to infer the potential distribution of presences within a minimum convex polygon, excluding unsuitable areas. These recommendations are important to establish, in a more assertive way, the extent of occurrence and area of occupancy of the species, the main criteria used by IUCN to determine conservation status.

Lastly, an important limitation of this study is to establish if the number of field surveys is sufficient to assess whether the niche is sufficiently sampled. This evaluation is also important to separate those species that are poorly collected and those that are really rare and/or endemic in the field. This separation is crucial to establish the proper strategies for conservation for each case. However, extensive fieldwork involves high costs and considerable time to be made. This situation in a megadiverse country, with high rates of habitat loss, scarce resources and under climate change forward, is a really important limitation. Studies quantifying the minimum number of inventories sufficient to draw the same conclusions, for a wide range of species from different taxonomic groups, would be very important.

Acknowledgements

References

Ando T & Hashimoto G, 1994. A new Brazilian species of *Petunia* (Solanaceae) from the Serra da Mantiqueira. *Brittonia*, 46:340-343. http://dx.doi.org/10.2307/2806919

Benites VM *et al.*, 2007. Soils associated with rock outcrops in the Brazilian mountain ranges Mantiqueira and Espinhaço. *Revista Brasileira de Botânica*, 30:569-577. http://dx.doi.org/10.1590/S0100-84042007000400003

De Grammont PC & Cuarón AD, 2006. An Evaluation of threatened species categorization systems used on the American Continent. *Conservation Biology*, 20(1):14-27. PMid:16909655. http://dx.doi.org/10.1111/j.1523-1739.2006.00352.x

Elith J *et al.*, 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1):43-57. http://dx.doi.org/10.1111/j.1472-4642.2010.00725.x

Engler R, Guisan A & Rechsteiner L, 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, 41:263-274. http://dx.doi.org/10.1111/j.0021-8901.2004.00881.x

Fielding AH & Bell JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1):38-49. http://dx.doi.org/10.1017/S0376892997000088

Fundação Biodiversitas, 2010. *Lista da flora brasileira ameaçada de extinção*. Available from: http://www.biodiversitas.org.br/floraBr.

Guisan A *et al.*, 2006. Using niche-based models to improve the sampling of rare species. *Conservation Biology*, 20(2):501-511. PMid:16903111. http://dx.doi.org/10.1111/j.1523-1739.2006.00354.x

Hernandez PA *et al.*, 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29:773-785. http://dx.doi.org/10.1111/j.0906-7590.2006.04700.x

International Union for Conservation of Nature – IUCN. Standards & Petitions Subcommittee, 2011. *Guidelines for Using the IUCN Red List Categories and Criteria*. Version 9.0. Prepared by the Standards and Petitions Subcommittee. Available from: http://www.iucnredlist.org/documents/RedListGuidelines.pdf.

Jimenez-Valverde A, Lobo JM & Hortal J, 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6):885-890. http://dx.doi.org/10.1111/j.1472-4642.2008.00496.x

Kamino LHY *et al.*, 2012. Challenges and perspectives for species distribution modelling in the Neotropics.

Biology Letters, 8(3):324-326. http://dx.doi.org/10.1098/rsbl.2011.0942

McKinney ML, 1999. High rates of extinction and threat in poorly studied taxa. *Conservation Biology*, 13(6):1273-1281. http://dx.doi.org/10.1046/j.1523-1739.1999.97393.x

Oksanen J *et al.*, 2012. *Vegan*: Community ecology package. R package version 2.0-3. Available from: http://CRAN.R-project.org/package=vegan.

Oliveira-Filho AT & Fontes MAL, 2000. Patterns of floristic differentiation among Atlantic Forests in Southeastern Brazil and the influence of climate. *Biotropica*, 32:793-810.

Paglia AP & Fonseca GAB, 2009. Assessing changes in the conservation status of threatened Brazilian vertebrates. *Biodiversity and Conservation*, 18(13):3563-3577. http://dx.doi.org/10.1007/s10531-009-9660-0

Pearson RG *et al.*, 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34(1):102-117. http://dx.doi.org/10.1111/j.1365-2699.2006.01594.x

Phillips SJ, Anderson RP & Schapire RE, 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190:231-259. http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026

Pillar VD, 2006. *MULTIV Multivariate exploratory analysis, randomization testing and bootstrap resampling*. User's Guide v. 2.4.

Pillar VDP & Orlóci L. 1996. On randomization testing in vegetation science: Multifactor comparisons of releve groups. *Journal of Vegetation Science*, 7:585-592. http://dx.doi.org/10.2307/3236308

R Development Core Team. 2012. *R:* A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available from: http://www.r-project.org/.

Rabinowitz D, 1981. Seven forms of rarity. In: Synge H (ed.). *The Biological aspects of rare plant conservation*. New York: John Wiley and Sons. p. 205-217.

Raxworthy CJ *et al.*, 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, 426(6968):837-841. PMid:14685238. http://dx.doi.org/10.1038/nature02205

Siqueira MF *et al.*, 2009. Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species. *Journal for Nature Conservation*, 17:25-32. http://dx.doi.org/10.1016/j.jnc.2008.11.001

Stehmann JR *et al.*, 2009. The genus *Petunia*. In Gerats T & Strommer J (eds.). *Petunia*: Evolutionary, developmental and physiological genetics. New York: Springer Science, Business Media. p. 1-28.