Available online at www.sciencedirect.com

ScienceDirect

Journal for
Nature Conservation

Something from nothing: Using landscape similarity and ecological niche modeling to find rare plant species

Marinez Ferreira de Siqueira[a,*], Giselda Durigan[b], Paulo de Marco Júnior[c], A. Townsend Peterson[d]

[a]*Centro de Referência em Informação Ambiental–CRIA, Av Romeu Tórtima 388, CEP: 13084-791, Campinas, SP, Brazil*
[b]*Floresta Estadual de Assis-Instituto Florestal de São Paulo, CP: 104 CEP: 19800-000, Assis, SP, Brazil*
[c]*Laboratório de Ecologia Teórica e Síntese, DBG-Universidade Federal de Goiás Rodovia Goiânia-Nerópolis km 5, Campus II, Setor Itatiaia, CP 131-CEP 74001-970, GO, Brazil*
[d]*Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, KS 66045, USA*

Summary
We present a worked example of how geographic and computational tools can aid in discovery and documenting unknown or poorly known populations and distributions of rare plant species. Focusing on *Byrsonima subterranea*, a rare plant of the cerrado biome in Brazil, considered probably extinct in the state of São Paulo, we used a combination of a simple environmental matching approach to locate extant populations in the state, and then a more complex ecological niche modeling approach to predict distribution of the species over a broader area. These methodologies have the potential to assist in documenting distributions of many rare plant species.
© 2008 Elsevier GmbH. All rights reserved.

Introduction

The cerrado biome of South America is well-known for its high levels of plant endemism. Endemism reaches ~30% among trees ([Ratter et al. 2000](#)), and may be as high as 44% overall among plant species ([Mittermeier et al. 1999](#)). As such, the cerrado represents one of the great storehouses of Neotropical biodiversity.

*Corresponding author. Tel.: +55 19 3288 0466; fax: +55 19 3249 0960.
E-mail addresses: marinez@cria.org.br (M.F. de Siqueira), giselda@femanet.com.br (G. Durigan), pdemarco@icb.ufg.br (P. de Marco Júnior), town@ku.edu (A.T. Peterson).

Plant species may be represented by few locality records for at least three reasons (Kunin and Gaston 1997). First, species may indeed have small overall geographic distributions, constituting microendemic species with genuinely restricted geographic distributions. For example *Leiothrix itacambirensis* Silv. (Eriocaulaceae) is restricted to a single mountain in the Serra do Espinhaço, Brazil (Giulietti and Pirani 1988). A second possibility is that they have low abundances—indeed, some species may have populations as sparse as <1 individual/ha (Kageyama and Gandara 2000)—that cause presences to be detected only rarely in surveys (Dewdney 1998). Finally, species may be represented sparsely in collections because of insufficient collecting—a species that is otherwise not particularly uncommon and that is more or less broadly distributed may simply not have been collected enough to represent its distribution adequately.

Placing a particular "rare" species into one or more of these categories is of fundamental importance to effective conservation planning. In the case of insufficient collecting, additional field surveys may change the view of a species' distribution and abundance to a more favourable status. In the case of restricted distributions, sparse populations, or low local abundances, however, more complex actions may be necessary, such as setting aside particular areas or sets of areas for protection. To distinguish among these possibilities, field surveys are clearly necessary—does the species occur under similar conditions in other areas, or are its few known occurrences really unique?

The purpose of this contribution is to document our experience with combining ecological niche models and new field surveys towards discovery of additional populations of, and partial characterisation of the geographic distribution of, a poorly known plant species: *Byrsonima subterranea* Brade & Markgr. (Malpighiaceae). This species is a sub-shrub with yellow-orange inflorescences found in open cerrado physiognomies. It is known from very few sites in the central and southern parts of the cerrado biome. Indeed, in the Brazilian state of São Paulo, only a single record existed (Durigan et al. 1999), and that population has been declared probably extinct in the state prior to initiation of this study (D.O.E 1998). We used a combination of simple analyses of similarity of multi-temporal 'greenness' indices drawn from satellite imagery, field surveys, and ecological niche modeling. The result was, in the end, literally something from (almost) nothing—from the single known São Paulo occurrence (Figure 1), seven occurrence sites are now documented, and a working hypothesis has been developed for its geographic distribution in the state.
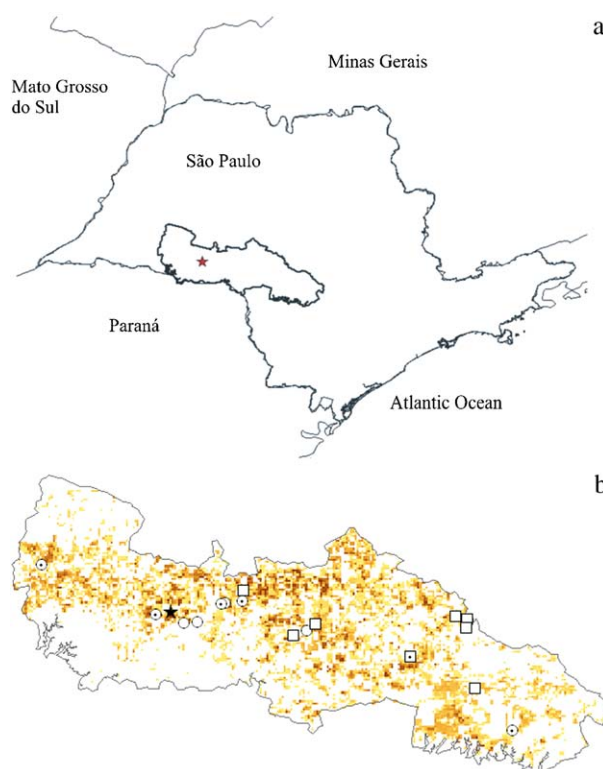

Figure 1. (a) Location of the study area (the Middle Paranapanema Basin in São Paulo state, Brazil). (b) Summary of environmental distances across the Middle Paranapanema Basin to the known occurrence point (star); also shown are localities of high similarity to the known point (circles) and of low similarity (squares)—points at which the species was found have a dot inside the symbol.

Methods

Our initial information was a single point of known occurrence for the species in the entire southern portion of its distributional area, which was determined by means of a GPS with a precision of ~100 m by Durigan. Because initially we knew of but a single occurrence site, typical ecological niche modeling approaches (Soberón and Peterson 2004) were inapplicable, as they require at least minimal sample sizes to function appropriately (Stockwell and Peterson 2002; Wisz et al. 2008). As a consequence, we used a very simple measure of similarity in environmental space to prioritise initial field efforts.

Study area

The Middle Paranapanema Basin is located in the southwestern part of São Paulo state (Figure 1a), covering ~19,622 km^2 in the area bounded by 22°1′ 59″–23° 23′ 43″ S and 50° 11′ 49″–53° 1″ W, with

altitudes ranging 334–1007 m. This region presents significant seasonality of temperatures, with coldest months averaging 9.9 °C, and warmest months 31.5 °C, with an annual mean temperature of 20.7 °C. Winters are generally dry, and summers rainy, and frost is frequent, particularly in the southeastern portion of the area (FUNDAG 1998). The basin holds two principal vegetation types: seasonal semi-deciduous forest and cerrrado, as well as broad areas of transition between the two (Figure 4) (Kronka et al. 2003), with 38 soil types, of which several are appropriate for agriculture (FUNDAG 1998).

Environmental distance calculations

We calculated similarity measures in a 16-dimensional environmental space. Environmental data sets included 12 monthly Normalized Difference Vegetation Index (NDVI) measurements derived from the Advanced Very High-Resolution Radiometer (AVHRR) sensor that are well-known as sources of detailed, fine-scale information regarding photosynthetic mass, or 'greenness' of the landscape (Tucker 1979). Because climate stations are sparse in the study region, climate data would be entirely interpolated and as such are inappropriate, but the directly-measured remote-sensing data used in our analyses provide reasonably precise, on-ground measurements. The multi-temporal nature of these data sets effectively summarises a combination of land use/land cover and vegetation phenology (UMD 2001). It is commonly used as an estimator of photosynthetic mass, or "greenness" (Chapman et al. 2005); more recently, NDVI data have been used as inputs into ecological niche models, providing detailed and robust hypotheses regarding geographic distributions of species (Egbert et al. 2000; Parra et al. 2004; Roura-Pascual et al. 2005). The remaining environmental data sets were four data layers derived from the Hydro-1 K global digital elevation model (http://edcdaac.usgs.gov/gtopo30/hydro/): elevation; slope; aspect; and, compound topographic index.

In this 16-dimensional environmental space, we summarised environmental characteristics across the study region (Middle Paranapanema Basin; Figure 1a) in a table by merging values from each data layer into a single GIS raster coverage using the COMBINE option in ArcView GIS (version 3.2). We exported the attributes table associated with this coverage in ASCII format. We converted each variable to a standard normal variable (mean 0, standard deviation 1) by subtracting the mean and

dividing by the standard deviation from each observation; this manipulation avoids differential influences of environmental dimensions with different variances, although we note that factor (principal components) analysis may provide an alternative 'view' of environmental variation across the study area. We then measured environmental similarity of each pixel in the landscape as a simple Euclidean distance (ED) to the characteristics of the known occurrence point, as

$$d = \sqrt{(\Delta x_i)^2},$$

where x_i represents the ith environmental variable, and Δ indicates difference from the reference locality at which the species is known. These distances were then reimported into ArcView, and joined with the original combined table to produce a raster GIS coverage summarising distance in environmental space to the known occurrence point. We refined this map of environmental similarity further by reducing it to just those zones documented as holding original cerrado vegetation in a recent, detailed survey (Kronka et al., 2005).

Field surveys

Based on the initial environmental distance analyses, we designed a sampling scheme to encounter additional *Byrsonima subterranea* populations. Sites were chosen based on the refined environmental distance map described above and GPS units. We located and surveyed nine cerrado patches showing high environmental similarity to the known occurrence point; only a few (8) low environmental similarity sites could be surveyed due to lack of time and resources. Each survey followed standard methodologies (four experienced observers walking throughout the small fragment, recording species encountered every 15 min; when accumulation curves reach asymptotes, survey effort ceases, generally within 1 h) that have been seen to produce consistent and repeatable inventories (Durigan et al. 2003), and herbarium specimens were prepared, submitted to experts for identification, and deposited in the herbaria of the Estação Ecológica de Assis and Instituto de Botânica de São Paulo. We used a one-tailed Fisher's exact test to assess the statistical significance of predictions in anticipating the success of on-ground searches in finding the species.

Ecological niche modeling

Once we had assembled a somewhat larger set of occurrences for the species in question, we

explored its broader-scale potential distribution using ecological niche modeling. The ecological niche of a population or species can be defined as the suite of environmental conditions and resources within which that population or species can maintain populations without immigrational subsidy (Grinnell 1917). The algorithm used for generating ecological niche models (ENMs) was the Genetic Algorithm for Rule-Set Prediction (GARP, version 1.1.3) (Stockwell and Noble 1992; Stockwell and Peters 1999) GARP is an evolutionary-computing method that builds ENMs based on non-random associations between known occurrence points for species and sets of digital GIS data layers describing the ecological landscape (i.e., in this case, the same 16 raster GIS data layers as were used in the similarity analysis described above).

Within the processing of the GARP program, initial rules are created by applying a method chosen randomly from a set of inferential tools (e.g., simple rule-generating algorithms including logistic regression, BIOCLIM, "atomic" rules, and negated range rules; Stockwell and Noble 1992; Stockwell and Peters 1999). The genetic algorithm consists of specially defined operators (e.g., means of perturbing initial rules to explore model solution space, such as crossover, mutation) that modify the initial rules randomly, and thus the result are models that have "evolved"—after each modification, the quality of the rule is tested (to maximise both significance and predictive accuracy), and a size-limited set of best rules is retained. Because rules are tested based on independent data (intrinsic test data), performance values reflect the expected performance of the rule, an independent verification that gives a more reliable estimate of true rule performance. The final result is a set of rules that can be projected onto a map to identify areas that fit the ecological 'profile' of the species, effectively a hypothesis of a potential geographic distribution for the species.

Following recent best-practice recommendations (Anderson et al. 2003), we developed 100 replicate GARP models, of which we discarded 90% based on consideration of error statistics, as follows. The 'best subsets' methodology (Anderson et al. 2003) consists of an initial filter removing models that omit (omission error = predicting absence in areas of known presence) heavily in predicting independent subsets of the input data, and a second filter based on an index of commission error (= predicting presence in areas of known absence), in which models predicting very large and very small areas are removed from consideration. Specifically, in GARP, we used intrinsic testing data (because of very small sample sizes), a soft omission threshold of 20%, and 50% retention based on commission considerations. The result was 10 'best subsets' models (binary raster data layers) that were summed to produce a best estimate of a potential geographic distribution for the species across the Middle Paranapanema Basin.

Results

The original known occurrence point for *Byrsonima subterranea* in São Paulo state, Brazil (Figure 1a), was used as a starting point for our investigations. From it, we generated maps of São Paulo state, and in particular of the Middle Paranapanema Basin (Figure 1b), summarising distance in a 16-dimensional standardised environmental space to the known occurrence point. Distances in this environmental space ranged 0.0001–3.9725 (units represent distances in a 16-dimensional space of the standard normal variables), and thus ranged across several orders of magnitude within the Middle Paranapanema Basin.

We conducted field surveys at nine sites with high similarity to the known point of occurrence; of these surveys, five yielded additional populations. These sites fell into five counties: Assis; Echaporã; Martinópolis; Águas de Santa Bárbara; and Campos Novos Paulista (Figure 1b). Surveys at eight sites of low similarity to the known point of occurrence yielded only one presence record; the interaction between level of similarity (high versus low) and success of on-ground searches was only suggestive of a relationship (Fisher's exact test, $P = 0.088$). The six additional presence sites were closely similar in environmental space to the known occurrence, with five of the six presenting distances of <0.07 units, and the sixth point more distant, at 0.159 units. The most different pixel in the Basin was 3.9725 units distant from the known occurrence point in environmental space—random sextets (to match the six occurrences we have available for Byrsonima) were all considerably more distant in environmental space than the six new occurrence sites (Figure 2).

Given (now) seven known occurrences of B. subterranea in the Middle Paranapanema Basin, development of more traditional ENM predictions was feasible. Although seven points is minimal for processing in GARP, the method can produce predictions with such small input data sets, and those predictions can provide a further guide for additional field surveys. As such, we used the 7 known occurrences in GARP to identify potential distributional areas across the Middle Paranapanema Basin (Figure 3).
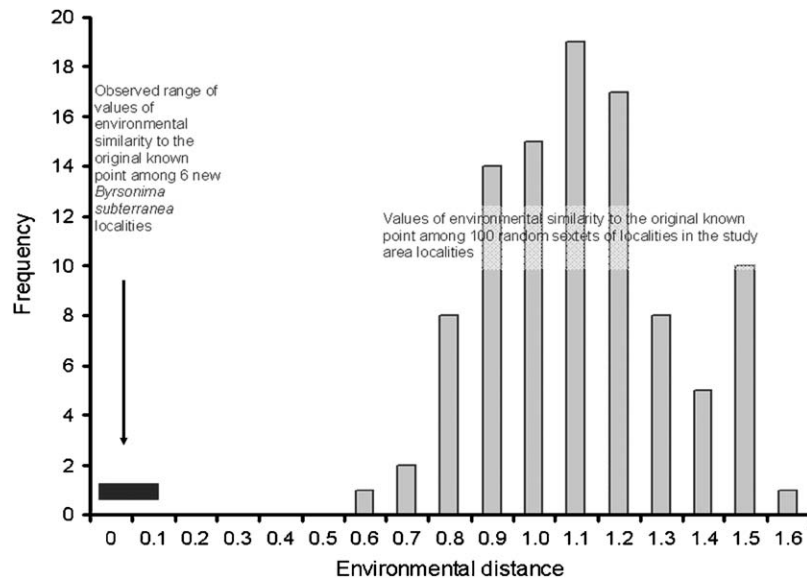
Figure 2. Comparison of observed values of environmental distance of the six new *Byrsonima subterranea* to the single previously known occurrence in the study area (0.034–0.159) with distance values derived from random sextets of pixels across the study area.
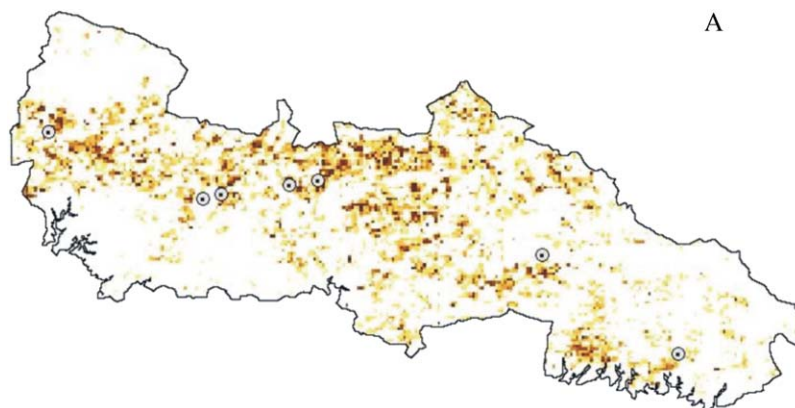

Figure 3. GARP model based on seven known occurrences of *Byrsonima subterranea* across the Middle Paranapanema Valley, São Paulo state, Brazil.

Discussion

This study has several limitations that must be acknowledged up front. First and most importantly, due to lack of time and resources, we were unable to prioritise surveys at numerous sites that were relatively distant in environmental space from the known point. As such, we do not present a clear test of the predictive nature of the simple environmental similarity approach. Also, with a greater number of sites sampled, it would be possible to determine similarity levels that are associated with presence more consistently, thus permitting a more complete prediction of the distribution of the species. Other limitations include the numerous access constraints that the complex cerrado landscape presents, thus limiting the suite of sites that we could visit and sample.

Nonetheless, given the relative ease with which we identified additional occurrence sites for the species, we consider the apparent rarity of *Byrsonima subterranea* to constitute a case of insufficient collecting effort. We suspect that difficulties of field identification and recognition for this species have limited the success of collectors in locating populations. An additional problem is the limited season in which it is apparent (i.e., flowering). This species is probably in reality fairly widely distributed in the cerrado zone, and not particularly rare, although it is known almost exclusively only in open cerrado physiognomies (i.e., *campo sujo* and *campo cerrado*),

which is quite uncommon in the state of São Paulo, presently covering only 0.05% of the state (Kronka et al., 2005). We suspect that, as knowledge of its range evolves (Lobo et al. 2007), the species will prove not to be in particular danger of extinction or regional extirpation.

Our simple, distance-based modeling effort turned out to be quite useful in directing field survey efforts and maximising their efficiency and efficacy. Although more information did exist from elsewhere on the species range, the initiation of our study with a single point (the only record in existence from São Paulo state) mimics situations in which very little is known about the geographic distribution of a species. This approach allowed us to locate six new (additional) populations of *B. subterranea* in São Paulo much more quickly and easily than we could have without the distance maps as guides.

Once the additional occurrence data were available, the ENM predictions produced using GARP provided a view of the species' potential distributional area across the Middle Paranapanema Basin. Comparing this map with that based on environmental distances (Figures 1b and 3), subtle differences in the two views become evident—the GARP prediction identifies smaller areas in the central portion of the Basin, which perhaps reflects better the rarity of this vegetation formation in the state. As such, we suspect that the GARP model will prove superior to the distance-based predictions—certainly, it is based on more occurrence information. However, as the goal of this effort was to generate *some* knowledge of the species' distribution, the distance-based predictions served their purpose adequately, and certainly started the process off in a way that GARP could not have. It is worth re-emphasising that *B. subterranea* was considered likely extinct in São Paulo state *prior to* this study, but as a result of this study, it has been removed from the list of species of conservation concern in the state completely (Mamede et al., 2007).

Perhaps most importantly, the resources and time needed to develop this new distributional understanding for a rare plant species were reduced considerably compared with exhaustive surveys, a critical saving in the face of the rapid habitat destruction occurring in this region. Given that the Middle Paranapanema Basin holds soils and terrains excellent for agricultural development, the region is seeing rapid conversion to agriculture and grazing. At present, only 6.4% of the area of the Basin remains covered with original vegetation (Kronka et al., 2005), and these remnants are extremely fragmented across the basin (Figure 4). As such, biodiversity conservation efforts in the region are under extreme pressure to identify important fragments and prioritise key fragments for conservation action.

Clearly, the simple environmental similarity-based methodology explored herein will bias discoveries of additional populations of such little-known species towards characterising a narrower niche space than these species may actually inhabit. The similarity-based approach is usable in conditions under which GARP (and other niche modeling approaches) is inapplicable (Stockwell and Peterson 2002; Wisz et al. 2008). Only through successive passes of prediction and discovery can the view of a species' niche be broadened sufficiently, but some discovery and extrapolation is possible. More to the point, something *is* better than nothing, and in many cases of this sort, next to nothing is all that is known—as such, this study echoes earlier attempts in this direction (Raxworthy et al. 2003; Bourg et al. 2005; Jarvis et al. 2005; Guisan et al. 2006), but goes to new
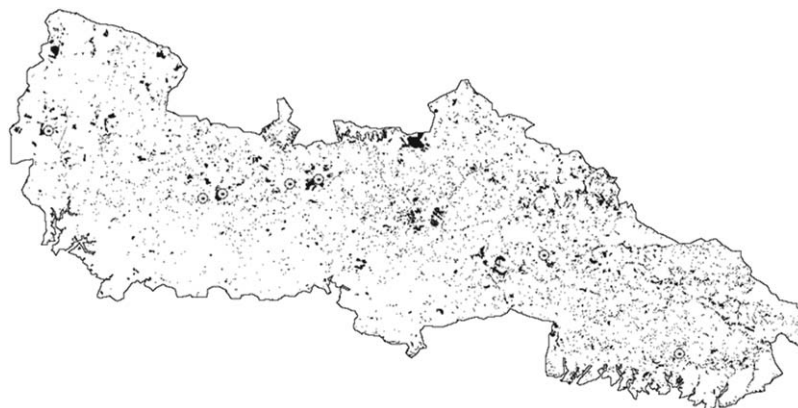

Figure 4. Present-day distribution of original natural vegetation types across the Middle Paranapanema Basin, São Paulo state, Brazil.

extremes in dealing with small sample sizes. This methodology may be broadly applicable to species that are little-known and poorly understood.

Acknowledgments

References

Anderson, R. P., Lew, D., & Peterson, A. T. (2003). Evaluating predictive models of species' distributions: Criteria for selecting optimal models. *Ecological Modelling, 162*, 211–232.

Bourg, N. A., McShea, W. J., & Gill, D. E. (2005). Putting a CART before the search: Successful habitat prediction for a rare forest herb. *Ecology, 86*, 2793–2804.

Chapman, A. D., Munõz, M. E. S., & Kock, I. (2005). Environmental information: Placing biodiversity phenomena in an ecological and environmental context. *Biodiversity Informatics, 2*, 24–41.

Dewdney, A. K. (1998). A general theory of the sampling processes with applications to the "veil line". *Theoretical Population Biology, 54*, 294–302.

DOE. (1998). Espécies da flora ameaçadas de extinção no Estado de São Paulo. In: Diário Oficial do Estado de São Paulo (pp. 1–23).

Durigan, G., Bacic, M. C., Franco, G. A. D. C., & Siqueira, M. F. (1999). Inventário florístico do cerrado na Estação Ecológica de Assis, SP. *Hoehnea, 26*, 149–172.

Durigan, G., Siqueira, M. F., Franco, G. A. D. C., Bridgewater, S., & Ratter, J. A. (2003). The vegetation of priority areas for cerrado conservation in São Paulo State, Brazil. *Edinburgh Journal of Botany, 60*, 217–241.

Egbert, S. L., Ortega-Huerta, M. A., Martínez-Meyer, E., Price, K. P., Peterson, A. T. (2000). Time-series analysis of high-temporal resolution AVHRR NDVI imagery of Mexico. In: *Proceedings IEEE 2000 International Geoscience and Remote Sensing Symposium (IGARSS).* (pp. 1978–1980).

FUNDAG. (1998). Diagnóstico do meio físico da região do Médio Paranapanema. Relatório Final Fundação de Apoio a Pesquisa Agrícola, Assis.

Giulietti, A. M., & Pirani, J. R. (1988). Patterns of geographic distribution of some plant species from the Espinhaço range, Minas Gerais and Bahia, Brasil. In H. W. Vanzolini (Ed.), *Proceedings of a workshop of neotropical distribution patterns* (pp. 39–69). Rio de JaneiroL: Academia Brasileira de Ciências.

Grinnell, J. (1917). Field tests of theories concerning distributional control. *American Naturalist, 51*, 115–128.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., et al. (2006). Using niche-based models to improve the sampling of rare species. *Conservation Biology, 20*, 501–511.

Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P., & Mottram, G. (2005). Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genetic Resources and Crop Evolution, 52*, 671–682.

Kageyama, P., & Gandara, F. B. (2000). Recuperação de áreas ciliares. In H. F. Leitão Filho (Ed.), *Matas ciliares: Conservação e recuperação*. São Paulo: Editora da Universidade de São Paulo/FAESP.

Kronka, F. J. d. N., Nalon, M. A., Baitello, J. B., Matsukuma, C. K., Pavão, M., Ywane, M. S. S.-I., et al. (2003). Levantamento da vegetação natural e caracterização de uso do solo no estado de são paulo. In *XI SBSR—Simpósio Brasileiro de Sensoriamento Remoto* (pp. 2779–2785). Belo Horizonte: INPE.

Kronka, F. J. N., Nalon, M. A., Matsukuma, C. K., Kanashiro, M. M., Ywane, M. S. S.-I., Lima, L. M. P. R., et al. (2005). Monitoramento da vegetação natural e do reflorestamento no Estado de São Paulo. In *XII simpósio brasileiro de sensoriamento remoto* (pp. 1569–1576). Goiania, GO: INPE.

Kunin, W. E., & Gaston, K. J. (1997). *The biology of rarity*. London: Chapman & Hall.

Lobo, J. M., Baselga, A., Hortal, J., Jiménez-Valverde, A., & Gómez, J. F. (2007). How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? *Diversity and Distributions, 13*, 772–780.

Mamede, M. C. H., Souza, V. C., Prado, J., Barros, F., Wanderley, M. G. L., Rando, J. G. (2007). Livro Vermelho das Espécies Vegetais Ameaçadas do Estado de São Paulo. Instituto de Botanica, Secretaria de Estado do Meio Ambiente, São Paulo, Brasil, São Paulo.

Mittermeier, R. A., Myers, N., Mittermeier, C. G., & Gil, P. R. (1999). *Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions*. Mexico: CEMEX Conservation International.

Parra, J. L., Graham, C. C., & Freile, J. F. (2004). Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography, 27*, 350–360.

Ratter, J. A., Bridgewater, S., Ribeiro, J. F., Dias, T. A. B., & Silva, M. R. (2000). Distribuição das espécies lenhosas da fitofisionomia cerrado sentido restrito nos estados compreendidos no bioma cerrado. *Boletim do Herbário Ezechias Paulo Heringer, 5*, 5–43.

Raxworthy, C. J., Martinez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A., et al. (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature, 426*, 837–841.

Roura-Pascual, N., Suarez, A., Gómez, C., Pons, P., Touyama, Y., Wild, A. L., et al. (2005). Geographic potential of Argentine ants (Linepithema humile Mayr) in the face of global climate change. *Proceedings of the Royal Society of London, B271*, 2527–2535.

Soberón, J., & Peterson, A. T. (2004). Biodiversity informatics: Managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London, B359*, 689–698.

Stockwell, D. R. B., & Noble, I. R. (1992). Induction of sets of rules from animal distribution data: A robust and informative method of analysis. *Mathematics and Computers in Simulation, 33*, 385–390.

Stockwell, D. R. B., & Peters, D. P. (1999). The GARP modelling system: Problems and solutions to auto-mated spatial prediction. *International Journal of Geographic Information Systems, 13*, 143–158.

Stockwell, D. R. B., & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling, 148*, 1–13.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*, 127–150.

UMD. (2001). AVHRR NDVI Data Set. In: University of Maryland, ⟨http://glcf.umiacs.umd.edu/index.shtml⟩, College Park, Maryland.

Wisz, M. S., Hijmans, R., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions, 14*, 763–773.