

Essay

Assumption- versus data-based approaches to summarizing species' ranges

A. Townsend Peterson*

Biodiversity Institute, University of Kansas, Lawrence, Kansas 66045, U.S.A.

Adolfo G. Navarro-Sigüenza and Alejandro Gordillo

Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico, D.F., Mexico

*email town@ku.edu

Running head: Summarizing species' ranges

Abstract

Species' geographic distributions are mapped using various approaches for use in conservation decision-making. Some such mapping efforts have relied on modifications of coarse-resolution extent-of-occurrence maps to downscale them to fine resolutions for conservation planning. This contribution examines (1) the quality of the extent-of-occurrence maps as range summaries, and (2) the utility of refining those maps into fine-resolution distributional hypotheses. In both cases, we found significant problems: the extent-of-occurrence maps are overly simple, omit many known and

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/cobi.12801](#).

This article is protected by copyright. All rights reserved.

well-documented populations, and likely frequently include many areas not holding populations. Refinement steps involve typological assumptions about habitat preferences and elevational ranges of species, which can introduce significant error in anticipating species' true distributional areas; however, as no model evaluation steps are taken to assess predictive ability of models, "bad" models are not noticed. Whereas range summaries derived by these methods may be useful in coarse-grained, global-extent studies, their continued use in on-the-ground conservation challenges at fine resolutions is not advisable. On the other hand, data-driven techniques that integrate primary biodiversity occurrence data with remotely sensed data summarizing environmental dimensions, termed ecological niche modeling or species distribution modeling, with rigorous and quantitative testing of model predictions prior to any use. These data-driven approaches constitute a well-founded, widely accepted alternative with a minimum of assumptions.

Introduction

A series of recent papers (Harris & Pimm 2008; Harris & Pimm 2004; Jenkins et al. 2011; Li & Pimm 2015; Ocampo-Peñuela & Pimm 2014; Schnell et al. 2013) has used an assumption-based methodology for summarizing species' geographic distributions in conservation applications, particularly for vertebrate species for which natural history and general distributional patterns are relatively well known. Many more papers (e.g., Ficetola et al. 2015; Rondinini et al. 2011) and initiatives (see MapOfLife; <http://mol.org/>) have used such methodologies in applications outside of conservation. The approach is based initially on polygon-format range summaries that have been produced by the IUCN, BirdLife International, and NatureServe (BirdLife International & NatureServe 2015; see summary of data available at <http://www.iucnredlist.org/technical->

documents/spatial-data). These rather simple extent-of-occurrence outlines are then refined by reducing to areas presenting elevations used by the species, and to areas presenting land-cover types used by the species (Harris & Pimm 2008). The result of this two-fold reduction is a picture of the distribution of the species that appears quite detailed, and is considerably more attractive and realistic-looking than the initial polygon; the question, however, is the degree to which the methodology has merit, particularly in fine-scale applications to conservation.

In this contribution, we evaluate the degree to which this assumption-based methodology is appropriate as a means of summarizing species' geographic distributions for conservation efforts. The method is certainly attractive, in that it offers a simple, straightforward, and data-inexpensive approach to summarizing species' distributions. Our concern is that in seeking simplicity, significant costs take the form of loss of predictive ability: these simple models (1) are based on input data (range outlines) and assumptions (e.g., stationary elevational ranges across the distribution of the species) that are frequently inaccurate, (2) make serious mistakes in combining spatial information at diverse extents and resolutions (the modifiable areal unit problem, see Openshaw 1984), and (3) lack the ability to take into account the diversity of distributional ecology across species' ranges. We present a worked example comparing the assumption-based methodology with a much more data-driven alternative based in ecological niche modeling; we then go on to evaluate how general are the problems that we have noted, and make suggestions about ways forward.

A Worked Example: The Black-throated Jay (*Cyanolyca pumilo*)

The Black-throated Jay is non-migratory, restricted to humid montane forests, and is endemic to Mesoamerica, in southern Mexico and northern Central America. Harris and Pimm (2008), to illustrate their methodology, identified the elevational range of this species as 1600-3000 m, and its

habitat association as ‘wet forest.’ In their figure legend, they appeared to cite Olson et al. (2001) as a source for habitat association data, although upon checking the data set associated with that source, only ‘montane forest’ ecoregions were identified, with no mention of ‘wet’ forest, so we used montane forest systems within the general range area of the species in an attempt to replicate their analyses. Figure 1 of their paper shows the triple intersection of the IUCN/BirdLife range outline for this species, elevations between 1600 and 3000 m, and montane forest. Indeed, in agreement with Harris and Pimm (2008), this species’ range *is* quite fragmented, and probably is vulnerable to land use change in the region.

Our concern, however, is about what is lost using such a simple approach that is based only on (1) expert knowledge about extent of occurrence, albeit at a coarse spatial resolution (Jetz et al. 2008), which is informed to some degree by reference to known occurrences of the species (Ocampo-Peñuela & Pimm 2014); and (2) assumptions based on the frequently-incomplete information about species’ elevational and land-cover associations. Vast primary biodiversity data (i.e., records that link a species to a place and a time of occurrence) are now available as Digital Accessible Knowledge (DAK; Sousa-Baena et al. 2013): data records that are digital, openly available to all via the Internet, and integrated into the broader universe of such data (= “knowledge,” as opposed to individual, isolated data points). Such data now total more than 6.4×10^8 records (see, e.g., <http://www.gbif.org>, <http://www.vertnet.org>, and others), including many millions of records for rare bird species, thanks to legions of citizen scientists (e.g., aVerAves for Mexico, within eBird, <http://ebird.org>). These data, of course, are not without problems: sampling is focused in accessible areas, precision of geographic coordinates is not always specified in data records, and identification errors cannot be detected without voucher information. However, citizen-scientist-generated data represent a fast-growing data resource for documenting present-day distributions of species (Boakes et al. 2010; Peterson et al.

2015), and can be analyzed and interpreted to create useful, detailed information about species' occurrences.

For the Black-throated Jay example, we downloaded GMTED 7.5" digital elevation layers to provide topographic information across the region, as well as the Olson et al. (2001) ecoregions data layer mentioned above; specimen and observation data for the species were downloaded from VertNet and GBIF. These data streams were interpreted following Harris and Pimm (2008): elevation between 1600 and 3000 m, and montane forest only (the satellite-derived existing habitat dataset mentioned in their Methods was not available for use, nor was the final distributional estimate; Figure 1). Immediately clear is that the elevational profile and the range outlines coincide broadly, although the latter has considerably less detail. The ecoregions data layer provides finer resolution than the range outline, and appears to distinguish well between interior slopes (too dry for this species) and those that receive sufficient moisture input to support cloud forest, but it remains coarsely resolved (as is expected from a global data set).

However, the actual occurrence data did not always coincide well with the assumption-based range summary (Figure 1). That is, most known occurrences fell within the extent-of-occurrence polygon (a notable exception being a specimen collected by A. L. Gardner in 1961, WFVZ 5043, locality given as "Tonalá, 9 mi SE, 10 mi NE"; although the directions from the locality are clearly problematic, a record from anywhere near Tonalá would extend the species' known range farther west, at least historically, than the extent-of-occurrence polygon would suggest). More significantly for conservation, numerous recent records place the species farther west in the Sierra Madre of Chiapas than the montane forest designations extend (see first inset in Figure 1). The species also ranges farther south in the Sierra Norte of Chiapas than Harris and Pimm (2008) anticipated (see second inset in Figure 1, in comparison to their Figure 1). The latter set of records comprises 41

observational records by multiple observers, ranging from 1992 to 2015, such that the presence of populations of the species at those sites is not in doubt.

To illustrate the data-driven approach, from the occurrence data mentioned above, we derived 173 spatially unique occurrence points across the species' range. We used ecological niche modeling approaches (Maxent algorithm, version 3.3.3k; Phillips et al. 2006; 10 bootstrap replicate analyses to permit assessment of model confidence, otherwise default parameters; model calibrated across the area within 110 km of the BirdLife range polygon) to relate these occurrences to environmental data in the form of 11 months (February-December 1995; January omitted owing to spatial artifacts in the data) of normalized difference vegetation indices (NDVI) from the Advanced Very High Resolution Radiometer (AVHRR) satellite-based sensor (30" resolution, or ~1 km; AVHRR data used so that the imagery would match the time of origin of the bulk of the occurrence data). The multitemporal nature of these data provides rich information about climate, land cover, and vegetation phenology (Cord & Rodder 2011; Moody & Johnson 2001; Ortega-Huerta et al. 2000); finer spatial resolution could be achieved via reduction to land-cover types in finer-resolution imagery via assumptions about habitat preferences.

The details of such data-driven analyses have been detailed in numerous publications on niche modeling, summarized in a recent book (Peterson et al. 2011). Further detail on use of remote-sensing imagery in such analyses is provided in Bodbyl-Roels et al. (2011). We emphasize that this particular example was developed with rather "off the shelf" data and tools: were we to be developing this example for conservation application, considerably more time would be devoted to all steps in the process.

The result of even this simple data-driven analysis was a striking improvement in detail (Figure 1)—we emphasize that any such models developed for real-world application (rather than for illustration only) would involve detailed quality-control of the input data, model evaluation, and other

steps (Peterson et al. 2011). The data-driven distributional hypothesis was smaller in extent and more detailed than the assumption-based approaches. Such analyses can include elements of both lower suitability and uncertainty in model predictions, depicted as lighter shades of red in Figure 1 (model uncertainty can be illustrated explicitly as well, see, e.g., Peterson et al. 2013), and applications for real-world use would include considerable experimentation with parameter settings (Warren & Seifert 2011), beyond default settings. Such data-driven approaches have seen massive-scale application to many questions in conservation biology, biogeography, and macroecology (e.g., Costa et al. 2010; de Pous et al. 2011; Moreno et al. 2011).

How General is this Problem?

Range polygons have been developed for a large number of species of birds, mammals, anurans, and several aquatic groups, and they have proven quite useful for global-scale, coarse-resolution applications (e.g., Rodrigues et al. 2004). The Black-throated Jay analyses presented above illustrate problems that these data sets may hold for finer-resolution applications. However, the question remains as to how common such problems are, across the broader suite of data; answering this question, however, depends on the availability of comprehensive, quality-controlled data documenting occurrences of species. To get some idea of the frequency of such problems, we used the *Atlas of Mexican Bird Distributions* dataset (Navarro Sigüenza & Peterson 2007; Navarro-Sigüenza 2002; Navarro-Sigüenza et al. 2002; Navarro-Sigüenza et al. 2003), a large-scale centralized database for Mexican birds (see Appendix for list of institutions contributing data) that has seen careful taxonomic scrutiny (Navarro-Sigüenza & Peterson 2004; Peterson & Navarro-Sigüenza 2009) and data quality control (Peterson et al. 2004), and that has been the cornerstone of many analyses of Mexican biodiversity (Peterson et al. 2010; Peterson & Navarro-Sigüenza 1999; Peterson et al. 2003; Peterson et al. 2015; Peterson et al. 2006; Rojas-Soto et al. 2003).

We filtered the BirdLife International and NatureServe (2015) range polygons to retain only species with ranges entirely within a rectangle slightly larger than Mexico (14-33 °N, 86-117.5 °W), and from those endemic or quasi-endemic species chose 25 at random. As discussed above, at least three sorts of error may be manifested in the assumption-based models: (1) the extent-of-occurrence polygons leave out distributional areas, (2) the reduced distributional hypotheses leave out still more distributional areas, or (3) the reduced distributional hypotheses are often overly broad and include unsuitable areas. Because the latter two error types depend on specific assumptions about the species' use of elevational bands and habitats, we focus simply on the first: to what degree to the initial range polygons omit known occurrence localities of species?

Of the 25 species inspected, omission rates were zero only for three (*Cyanocorax yucatanicus*, *Icterus auratus*, *Aimophila notosticta*). For the remaining 22 species, omission rates ranged from 5.3% of occurrence records (*Aulacorhynchus wagleri*) to 76.9% (*Amazilia wagneri*); note, that these numbers do not correspond to percentages of range area, but rather to percentages of raw occurrence records (i.e., including duplicate records from localities). We noted no clear relationship between omission rates and numbers of records available for species (Figure 2). Figure 2 also provides three mapped examples that illustrate various phenomena such as peripheral populations being left out (e.g., northeastern Mexican populations of *Atlapetes pileatus*), and overly simply polygons leaving out details of distributions (see, e.g., long, straight range limit drawn for *Buarremon virenticeps*, and populations falling north of that limit). In sum, the problems noted above for Black-throated Jays are systemic throughout the dataset, at least for Mexican birds.

Beresford et al. (2011) presented parallel analyses of the BirdLife International and NatureServe (2015) range summaries for African birds, and using observed occurrences in Important Bird Areas (IBAs). They found substantial error rates in the range summaries: of 3577 bird species occurrences in IBAs indicated by the BirdLife International and NatureServe (2015) polygons, only

847 were corroborated by observational records. These error rates did not change appreciably when the BirdLife International and NatureServe (2015) polygons were reduced to more specific hypotheses based on methods parallel to those of Harris and Pimm (2008), with only 630 of 2517 presence predictions corroborated. Hence, problems with use of the BirdLife International and NatureServe (2015) range polygons at fine resolutions appear to be very general, and not dependent on a particular evaluation method or region or group of birds (see, e.g., Ramirez-Bastida et al. 2008).

Conclusions and Recommendations

The assumption-based approach used by Harris and Pimm (2008) is certainly attractive, in that it is simple and easy to implement. However, as should be apparent from the example presented above, the approach is also limited in its ability to characterize the distribution of a species effectively (Mota-Vargas & Rojas-Soto 2012). The failing of the approach lies in its reliance on overly simple, coarse-resolution initial hypotheses of range extents, and on crude and data-poor assumptions about the details of species' distributions with respect to elevation and habitat. More fundamentally, these approaches suffer from the effects of multi-level conflicts among scales and resolutions: the extent-of-occurrence polygons are particularly coarse spatially, whereas the land cover and elevation data are very fine in resolution. This set of problems of integrating data at different spatial resolutions is called the modifiable areal unit problem (Openshaw 1984): imposing artificial units of spatial reporting or crossing among spatial reference system generates artificial spatial patterns (Heywood et al. 1998). Careless integration of massively different resolutions—as would occur in meshing the very simple and coarse range outline maps with very fine-resolution maps of elevation and land cover—is not justified in spatial terms, and can only lead to artifacts and inaccuracies in data products that result.

One might be tempted to dismiss these comments and critiques as quibbles among “modelers,” and of little real-world importance. Indeed, many of the uses of the extent-of-occurrence range polygon data have been in global-extent analyses in macroecology (e.g., Somveille et al. 2013), where biases and omissions are perhaps of least concern. The assumption-based maps, however, are being used in real-world, on-the-ground conservation efforts (<http://www.savingspecies.org/>), yet without any model testing to assure predictivity or concern about the complexities of such data integration. We note that our explorations of Black-throated Jay distributions identified several real, existing populations of this species that were left out of the modified extent-of-occurrence maps.

Indeed, very real conservation decisions are being based explicitly on the results of these analyses. The case in point comprises the laudable initiatives of the Saving Species program, which involve support for purchasing specific parcels of land to increase connectivity, reduce fragmentation, and improve effective size of habitable areas for key species. The design and objectives of the initiative are much to be admired, but our concern is in regard to their specific methodology for summarizing species' distributions. Choice of sites for these efforts is apparently centered on the assumption-based methodology (S. L. Pimm, pers. comm. with ATP, October 2015). Such local-landscape-scale applications are those that would be of greatest concern in terms of the biases and omissions that we have documented herein. That is, our complaint is not with the idea of using distributional summaries for species of conservation concern to guide conservation efforts, nor with the extent-of-occurrence maps *per se* (they have been very useful at coarse resolutions and global extents), but rather with their being downscaled to fine spatial resolutions for fine-resolution conservation planning.

We suggest that data-driven approaches that take advantage of the massive biodiversity data resources that now exist as DAK are likely to provide a much more solid foundation for such decisions. The suite of tools under the rubric of ecological niche modeling (sometimes termed species

distribution modeling) offers rich methodologies for summarizing species' geographic ranges in considerable detail (Bodbyl-Roels et al. 2011), if used well and applied to data that have been quality controlled carefully (e.g., Kumar & Stohlgren 2009; Menon et al. 2012; Urbina-Cardona & Loyola 2008). Modern implementations of these methodologies involve large-scale input data (fine-resolution occurrence data, remotely sensed environmental summaries), careful consideration of calibration areas (Barve et al. 2011) and biogeographic scenarios (Saupe et al. 2012), experimentation with many different algorithms (Qiao et al. 2015) and parameter settings (Anderson & Gonzalez 2011; Boria et al. 2014) to obtain optimal models, and detailed model testing (Peterson et al. 2008), among other methodological considerations. Although these data-driven models are limited when information available is minimal, the assumption-based approaches will be limited similarly by lack of information; several contributions have now explored data-driven applications even to the least-well-known species (Almeida et al. 2009; Engler et al. 2004; Menon et al. 2010; Menon et al. 2012). Such approaches offer quantitative, testable distributional hypotheses on which conservation decisions can be based.

Acknowledgments

We thank Stuart Pimm for his well-illustrated lecture in Lima, Peru, that led to this analysis being developed. We thank Jorge Soberón and two anonymous reviewers for insightful comments and ideas. We thank the many persons and institutions who have shared their primary occurrence data, which make studies of this sort possible.

Literature Cited

This article is protected by copyright. All rights reserved.

- Almeida, C. E., E. Folly-Ramos, A. T. Peterson, V. Lima-Neiva, M. Gumiel, R. Duarte, M. Locks, M. Beltrão, and J. Costa. 2009. Could the bug *Triatoma sherlocki* be vectoring Chagas disease in small mining communities in Bahia, Brazil? *Medical and Veterinary Entomology* **23**:410-417.
- Anderson, R. P., and I. Gonzalez. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling* **222**:2796-2811.
- Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* **222**:1810-1819.
- Beresford, A. E., G. M. Buchanan, P. F. Donald, S. H. M. Butchart, L. D. C. Fishpool, and C. Rondinini. 2011. Poor overlap between the distribution of protected areas and globally threatened birds in Africa. *Animal Conservation* **14**:99-107.
- BirdLife International, and NatureServe 2015. *Range Maps of Bird Species of the World*. BirdLife International, Cambridge.
- Boakes, E. H., P. J. K. McGowan, R. A. Fuller, D. Chang-qing, N. E. Clark, K. O'Connor, and G. M. Mace. 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology* **8**:e1000385.
- Bodbyl-Roels, S., A. T. Peterson, and X. Xiao. 2011. Comparative analysis of remotely-sensed data products via ecological niche modeling of avian influenza case occurrences in Middle Eastern poultry. *International Journal of Health Geographics* **10**:21.

- Boria, R. A., L. E. Olson, S. M. Goodman, and R. P. Anderson. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* **275**:73-77.
- Cord, A., and D. Rodder. 2011. Inclusion of habitat availability in species distribution models through multi-temporal remote-sensing data? *Ecological Applications* **21**:3285-3298.
- Costa, G. C., C. Nogueira, R. B. Machado, and G. R. Colli. 2010. Sampling bias and the use of ecological niche modeling in conservation planning: A field evaluation in a biodiversity hotspot. *Biodiversity and Conservation* **19**:883-899.
- de Pous, P., W. Beukema, M. Weterings, I. Dümmer, and P. Geniez. 2011. Area prioritization and performance evaluation of the conservation area network for the Moroccan herpetofauna: A preliminary assessment. *Biodiversity and Conservation* **20**:89-118.
- Engler, R., A. Guisan, and L. Rechsteiner. 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.
- Ficetola, G. F., C. Rondinini, A. Bonardi, D. Baisero, and E. Padoa-Schioppa. 2015. Habitat availability for amphibians and extinction threat: A global analysis. *Diversity and Distributions* **21**:302-311.
- Harris, G., and S. L. Pimm. 2008. Range size and extinction risk in forest birds. *Conservation Biology* **22**:163-171.
- Harris, G. M., and S. L. Pimm. 2004. Bird species' tolerance of secondary forest habitats and its effects on extinction. *Conservation Biology* **18**:1607-1616.

- Heywood, D. I., S. Cornelius, and S. Carver 1998. An Introduction to Geographical Information Systems. Addison Wesley Longman, New York.
- Jenkins, C. N., S. L. Pimm, and M. d. S. Alves. 2011. How conservation GIS leads to Rio de Janeiro, Brazil. *Natureza & Conservação* **9**:152-159.
- Jetz, W., C. H. Sekercioglu, and J. E. Watson. 2008. Ecological correlates and conservation implications of overestimating species geographic ranges. *Conservation Biology* **22**:110-119.
- Kumar, S., and T. J. Stohlgren. 2009. Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. *Journal of Ecology and natural Environment* **1**:94-98.
- Li, B. V., and S. L. Pimm. 2015. China's endemic vertebrates sheltering under the protective umbrella of the giant panda. *Conservation Biology*.
- Menon, S., B. I. Choudhury, M. L. Khan, and A. T. Peterson. 2010. Ecological niche modeling predicts new populations of *Gymnocladus assamicus*, a critically endangered tree species. *Endangered Species Research* **11**:175-181.
- Menon, S., M. L. Khan, A. Paul, and A. T. Peterson. 2012. *Rhododendron* species in the Indian Eastern Himalayas: New approaches to understanding rare plant species' distributions. *Journal of the American Rhododendron Society* **Spring 2012**:78-84.
- Moody, A., and D. S. Johnson. 2001. Land surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sensing of Environment* **75**:305-323.
- Moreno, R., R. Zamora, J. R. Molina, A. Vasquez, and M. Á. Herrera. 2011. Predictive modeling of microhabitats for endemic birds in South Chilean temperate forests using Maximum entropy (Maxent). *Ecological Informatics* **6**:364-370.

- Mota-Vargas, C., and O. R. Rojas-Soto. 2012. The importance of defining the geographic distribution of species for conservation: The case of the Bearded Wood-Partridge. *Journal for Nature Conservation* **20**:10-17.
- Navarro Sigüenza, A. G., and A. T. Peterson 2007. Mapas de las Aves de México Basados en WWW; final report SNIB-CONABIO project CE015. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico, D.F.
- Navarro-Sigüenza, A. G. 2002. Atlas de las Aves de México: Fase II; database from SNIB-CONABIO projects E018 y A002. Facultad de Ciencias, Universidad Nacional Autónoma de México, México, D.F.
- Navarro-Sigüenza, A. G., and A. T. Peterson. 2004. An alternative species taxonomy of Mexican birds. *Biota Neotropica* **4**:
<http://www.biotaneotropica.org.br/v4n2/pt/abstract?article+BN02304022004>.
- Navarro-Sigüenza, A. G., A. T. Peterson, and A. Gordillo-Martínez. 2002. A Mexican case study on a centralised database from world natural history museums. *CODATA Journal* **1**:45-53.
- Navarro-Sigüenza, A. G., A. T. Peterson, and A. Gordillo-Martínez. 2003. Museums working together: The atlas of the birds of Mexico. *Bulletin of the British Ornithologists' Club* **123A**:207-225.
- Ocampo-Peñuela, N., and S. L. Pimm. 2014. Setting practical conservation priorities for birds in the western Andes of Colombia. *Conservation Biology* **28**:1260-1270.
- Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. N. Powell, E. C. Underwood, J. A. Amico, I. Itoua, H. E. Strand, J. C. Morrison, C. J. Loucks, T. F. Allnutt, T.

- H. Ricketts, Y. Kura, J. F. Lamoreux, W. W. Wettengel, P. Hedao, and K. R. Kassem. 2001. Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience* **51**:933-938.
- Openshaw, S. 1984. *The Modifiable Areal Unit Problem*. Geo Books, Norwich.
- Ortega-Huerta, M. A., E. Martínez-Meyer, S. L. Egbert, K. P. Price, and A. T. Peterson. 2000. Mapping the land cover of Mexico using AVHRR time-series data sets. *GeoCarto International* **15**:5-17.
- Peterson, A. T., X. Li, and A. G. Navarro-Sigüenza. 2010. Joint effects of marine intrusion and climate change on the Mexican avifauna. *Annals of the Association of American Geographers* **100**:908-916.
- Peterson, A. T., and A. G. Navarro-Sigüenza. 1999. Alternate species concepts as bases for determining priority conservation areas. *Conservation Biology* **13**:427-431.
- Peterson, A. T., and A. G. Navarro-Sigüenza. 2009. Constructing check-lists and avifauna-wide reviews: Mexican bird taxonomy revisited. *Auk* **126**:915-921.
- Peterson, A. T., A. G. Navarro-Sigüenza, B. E. Hernández-Baños, G. Escalona-Segura, F. Rebón-Gallardo, E. Rodríguez-Ayala, E. M. Figueroa-Esquivel, and L. Cabrera-García. 2003. The Chimalapas region, Oaxaca, Mexico: A high-priority region for bird conservation in Mesoamerica. *Bird Conservation International* **13**:227-254.
- Peterson, A. T., A. G. Navarro-Sigüenza, E. Martínez-Meyer, A. P. Cuervo-Robayo, H. Berlanga, and J. Soberón. 2015. Twentieth century turnover of Mexican endemic avifaunas: Landscape change versus climate drivers. *Science Advances* **1**:e1400071.
- Peterson, A. T., A. G. Navarro-Sigüenza, and R. S. Pereira. 2004. Detecting errors in biodiversity data based on collectors' itineraries. *Bulletin of the British Ornithologists' Club* **124**:143-151.

- Peterson, A. T., M. Papeş, and J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modelling. *Ecological Modelling* **213**:63-72.
- Peterson, A. T., T. Radocy, E. Hall, J. P. Kerbis, and G. G. Celesia. 2013. The potential distribution of the African lion in the face of changing global climate. *Oryx* **48**:555-564.
- Peterson, A. T., V. Sánchez-Cordero, E. Martínez-Meyer, and A. G. Navarro-Sigüenza. 2006. Tracking population extirpations via melding ecological niche modeling with land-cover information. *Ecological Modelling* **195**:229-236.
- Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**:231-259.
- Qiao, H., J. Soberón, and A. T. Peterson. 2015. No silver bullets in correlative ecological niche modeling: Insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution* **6**:1126-1136.
- Ramirez-Bastida, P., A. G. Navarro-Sigüenza, and A. T. Peterson. 2008. Aquatic bird distributions in Mexico: Designing conservation approaches quantitatively. *Biodiversity and Conservation* **17**:2525-2558.
- Rodrigues, A. S. L., H. R. Akçakaya, S. J. Andelman, M. I. Bakarr, L. Boitani, T. M. Brooks, J. S. Chanson, L. D. C. Fishpool, G. A. B. Da Fonseca, K. J. Gaston, M. Hoffmann, P. A. Marquet, J. D. Pilgrim, R. L. Pressey, J. Schipper, W. Sechrest, S. N. Stuart, L. G. Underhill, R. W.

- Waller, M. E. J. Watts, and X. Yan. 2004. Global gap analysis: Priority regions for expanding the global protected-area network. *BioScience* **54**:1092-1100.
- Rojas-Soto, O. R., O. Alcantara-Ayala, and A. G. Navarro. 2003. Regionalization of the avifauna of the Baja California Peninsula, Mexico: A parsimony analysis of endemism and distributional modelling approach. *Journal of Biogeography* **30**:449-461.
- Rondinini, C., M. Di Marco, F. Chiozza, G. Santulli, D. Baisero, P. Visconti, M. Hoffmann, J. Schipper, S. N. Stuart, M. F. Tognelli, G. Amori, A. Falcucci, L. Maiorano, and L. Boitani. 2011. Global habitat suitability models of terrestrial mammals. *Philosophical Transactions of the Royal Society of London B* **366**:2633-2641.
- Saupe, E. E., V. Barve, C. E. Myers, J. Soberon, N. Barve, C. M. Hensz, A. T. Peterson, H. L. Owens, and A. Lira-Noriega. 2012. Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. *Ecological Modelling* **237**:11-22.
- Schnell, J. K., G. M. Harris, S. L. Pimm, and G. J. Russell. 2013. Quantitative analysis of forest fragmentation in the Atlantic Forest reveals more threatened bird species than the current Red List. *PLoS ONE* **8**:e65357.
- Somveille, M., A. Manica, S. H. Butchart, and A. S. Rodrigues. 2013. Mapping global diversity patterns for migratory birds. *PLoS ONE* **8**:e70907.
- Sousa-Baena, M. S., L. C. Garcia, and A. T. Peterson. 2013. Completeness of Digital Accessible Knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions* **20**:369-381.

Urbina-Cardona, J. N., and R. D. Loyola. 2008. Applying niche-based models to predict endangered-hylid potential distributions: Are neotropical protected areas effective enough. *Tropical Conservation Science* **1**:417-445.

Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications* **21**:335-342.

Figure Legends

Figure 1. Summary of the distribution of the Black-throated Jay (*Cyanolyca pumilo*), showing the three elements in the assumption-based approach (IUCN range polygons, elevations, land cover type), as well as actual occurrence data in the form of data associated with specimens and observations and two “zooms” to provide additional detail. This set of analyses attempts to follow Harris and Pimm (2008) in estimating the species’ extent of occurrence as those areas falling within the polygon (dashed outline), elevational profile (red shading), and currently existing montane forest (“ecoregion,” hashed area). The bottom panel provides the results of a preliminary data-driven analysis for this species, in which known occurrences were related to vegetation index data from 11 months derived from the AVHRR sensor (see Methods).

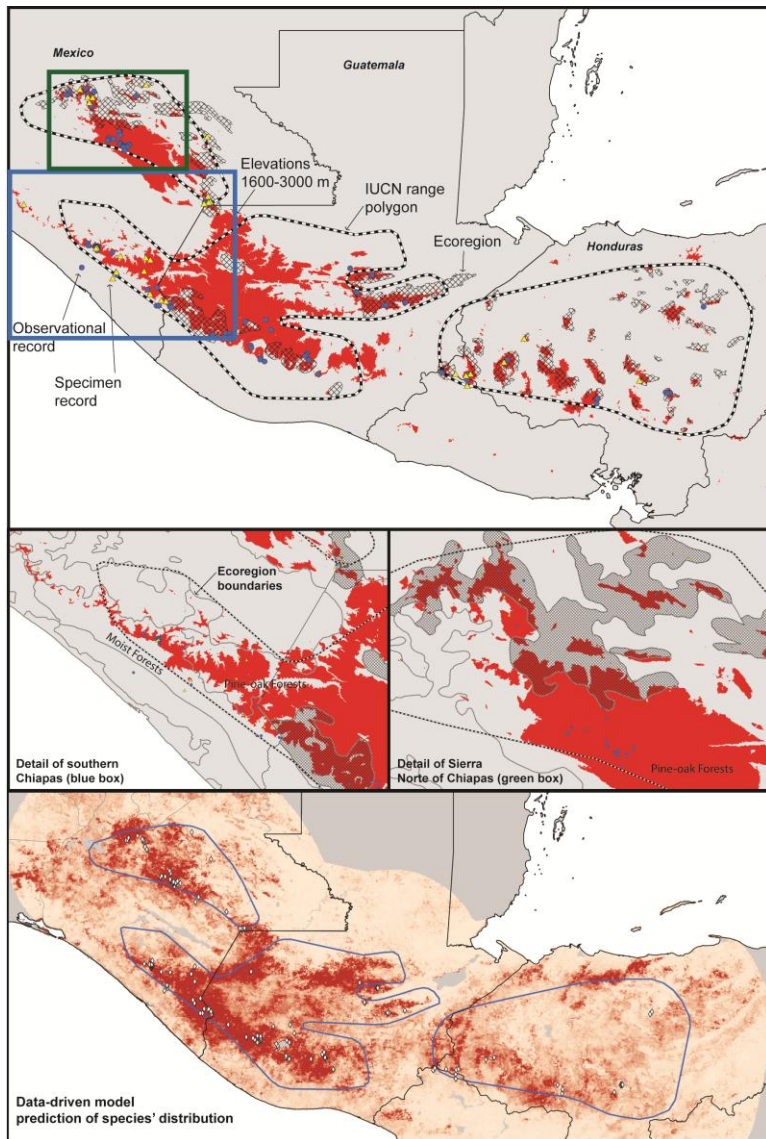


Figure 2. (a) Percentage of omission of good-quality distributional data related to numbers of occurrence points available across a sample of 25 endemic Mexican bird species. (b-d) Comparison of available distributional data with IUCN distributional maps polygons for three Mexican endemic species.

