Conservation Biology

Comment

Unfulfilled promise of data-driven approaches: response to Peterson et al. 2016

Stuart L. Pimm,[1,2]* Grant Harris,[3] Clinton N. Jenkins,[4,2] Natalia Ocampo-Peñuela,[1,5] and Binbin V. Li[1]

[1]Nicholas School of the Environment, Box 90328, Duke University, Durham, NC, 27708, U.S.A.
[2]SavingSpecies, 5625 Sunset Lake Road, #12, Holly Springs, NC 27540, U.S.A.
[3]U.S. Fish and Wildlife Service, P.O. Box 1306, Albuquerque, NM 87103-1306, U.S.A.
[4]Instituto de Pesquisas Ecológicas, Nazaré Paulista, São Paulo 12960-000, Brazil
[5]Current address: Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland

Understanding where species occur is vital to their conservation. Distribution models of threatened and endangered species, however, are necessarily based on few data and consequently are controversial. In this regard, the data-driven approaches Peterson et al. (2016) attempt promise much and in time may deliver. Unfortunately, they demonstrate how a statistical model can make poor predictions. We used their case study to illustrate some of the advantages of expert-driven approaches in data-poor situations. We identify the limitations of what they term "data-driven approaches," respond to their criticisms, and suggest improvements.

The IUCN Red List (IUCN 2016) and BirdLife International (2016) attempt global assessments of species' risk of extinction. Key criteria are the size of a species' geographical range and anthropogenic threats. The IUCN has assessed 76,000 species, and the process aspires to triple that number by 2020. When it succeeds, it will have assessed approximately 10% of described species and perhaps 1% of all species (Scheffers et al. 2012). Of the assessed terrestrial species, only birds, mammals, and amphibians have been mapped comprehensively.

In previous papers, we developed approaches to assess which species may be at risk of extinction that the IUCN Red List overlooks and identified where such species concentrate. Using these updated lists of species, we provide explicit recommendations for practical conservation actions. In several cases, land purchases for conservation followed our recommendations (Jenkins et al. 2011; Ocampo-Penuela & Pimm 2014). In particular, Harris and Pimm (2008), Schnell et al. (2013*a*), and Li et al. (2016) show that the existing IUCN range maps contain substantial commission errors—that is, they suggest many species are present outside their known elevational limits, or where habitat no longer exists. The problem is a global one that leads to substantial underestimation of the number of species likely to be at risk of extinction (Ocampo-Peñuela et al. 2016). Further, the IUCN range maps do not report the extent of habitat fragmentation (Schnell et al. 2013*a*, 2013*b*).

Peterson et al.'s maps make even greater errors of commission than do IUCN's and predict a continuous range where habitat is severely fragmented. Both errors could make a species seem of least concern when, in fact, it is at risk.

Peterson et al. suggest abandoning the IUCN maps, recommending instead "data-driven approaches" that depend strictly on models derived from species' locality data. Such data are few and sometimes unavailable for many threatened species and regions. Their approach would force planners to consider even fewer species than at present. (For any taxonomically comprehensive sets of species for which IUCN has not produced maps but for which abundant locality data are available, their approaches may be warranted.) In contrast, our methods include all species for which there are range maps. We modify maps with data on elevations (90-m scale), forest cover (30-m scale), and protected areas (Alves et al. 2008; Ocampo-Penuela & Pimm 2014; Li & Pimm 2015; Li et al. 2016). These data are freely available and global in extent, and our methods are simple applications of GIS. Whatever the uncertainties in elevational ranges and habitat cover, these are small relative to the

*email stuartpimm@me.com
Paper submitted August 1, 2016; revised manuscript accepted December 1, 2016.

Conservation Biology, Volume 31, No. 4, 944–947

DOI: 10.1111/cobi.12928

very large fractions of mapped ranges that are outside known elevational limits and lack habitat. Key aspects and annotations of these methods are in Supporting Information.

Our approach mirrors the National Gap Analysis Program (GAP), active for over 2 decades in the United States (Scott et al. 1993), and similar conservation exercises (e.g., Jennings 2000; Cowling et al. 2003). These and our methods are applicable to many species of conservation concern and are globally practical. Those of Peterson et al.'s may work well only with better-sampled species.

With 171 observations (number in table supplied to us by Peterson et al.), the Black-throated Jay (*Cyanolyca pumilo*) is a richly documented species relative to many we considered. Closer inspection of these locations on high-resolution imagery in Google Earth confirms that 119 (69%) are in forest (Fig. 1). Remaining locations include 1 in a lake, 3 on mountain summits, and 10 in cities. Much more troubling are those 38 observations (21% of the 171) in areas converted to human use, principally agricultural fields. There is no indication the jay uses such areas. Likely the jays were in forests some unknown distance away or the area was deforested in the past. Considering these points valid records of occurrence greatly expands the modeled range of jay habitats.

Worse yet, Peterson et al.'s points are not spatially independent: 30% of forest points are within 1 km of another point; two-thirds are within 5 km; and 26% are close to El Triunfo, Chiapas. In contrast, few points in agricultural landscapes are close to each other. Clustered records are problematic because they may represent multiple observations from one area frequented by birdwatchers because it affords the best chance of seeing the species. Were one to discount half the forest records as replicates, one would conclude that approximately 60 observations come from forests and 38 come from areas converted to human use. Peterson et al. consider the species to be restricted to "humid montane forests" (as we do), but their model assigns it to many other vegetation types.

All studies of species habitat should take great care to select explanatory variables that represent, as closely as possible, the biological drivers that determine a species' habitat. Modelers have a propensity to use explanatory variables for the sake of convenience rather than for biological realism. Peterson et al. chose AVHRR images and the NDVI index. The AVHRR was not designed to assess vegetation, and it predicts vegetation classes (and hence a bird's range) in the wrong locations more than other imagery types (Harris et al. 2005). Moreover, AVHRR is spatially crude (1-km pixel size) relative to the 30-m resolution of Landsat images. In contrast, Harris and Pimm (2008) used a supervised classification of imagery based on field sampling of habitats. At the 1-km scale, even if all the jay locations were accurate, the pixel in which many of them were found would contain substantial fractions of agricultural landscapes. This, too, would predict the jay to be present across a much larger range of vegetation types than those in which it occurs. Overall, Peterson et al.'s map predicts a relatively high probability of the jay's presence over extensive areas, whereas higher-resolution satellite imagery and our field work suggest this is unlikely (e.g., near Tapachula on the Guatemala–Mexico border) (Fig. 1, area A) (G.H., personal observation).

Most researchers using Maxent ignore the sensitive assumptions that it demands—a key one being systematic sampling (Yackulic et al. 2013). Maxent's appeal stems from its not requiring absence data, which are not always available. Whatever modeling approach one chooses, absence data can provide insights into the predicted range's veracity. Compelling absence data are available from eBird (2016). Peterson et al. predict this species in El Salvador near Volcán de Santa Ana (Fig. 1, area B). There are no records of the species there, yet eBird's checklists from this location include 2 with >130 species and many others with >50 species (eBird 2016). Simply, people survey this area vigorously but do not find the jay.

When Peterson et al. claim IUCN maps are not data driven, they mean IUCN and BirdLife International produce maps from "a variety of published and unpublished data sources" rather than a formal algorithm (BirdLife International 2016). What matters is how well the maps perform. For example, in Fig. 1 nearly all the point observations lie within the polygon that BirdLife International supplies (all within 7 km of it). It represents the boundary within which one expects to see a given species in its habitat. These boundaries are data driven, and for birds BirdLife International uses an open, democratic assessment process that readily solicits and incorporates new data and updates species maps and decisions accordingly—as we do when necessary (Ocampo-Penuela & Pimm 2014). Harris and Pimm (2008) digitized the range maps from Howell and Webb (1995) (no other spatial data were available at the time). We used the latest maps from BirdLife International and IUCN in subsequent publications.

Readily available data on elevations and high-resolution data on forest cover provide the best clues to where habitat remains. Indeed, the higher the resolution of satellite data, the more easily one can assess whether an area contains habitat. Coarse data provide an inadequate guide to where a species might be, as Peterson et al.'s example demonstrates.

Regardless, for the jay and other species, very little habitat remains, and what is left is severely fragmented (Schnell et al. 2013*a*, 2013*b*), something else Peterson et al.'s models miss. The proximity of our 2008 predictions to these newly available observations is striking

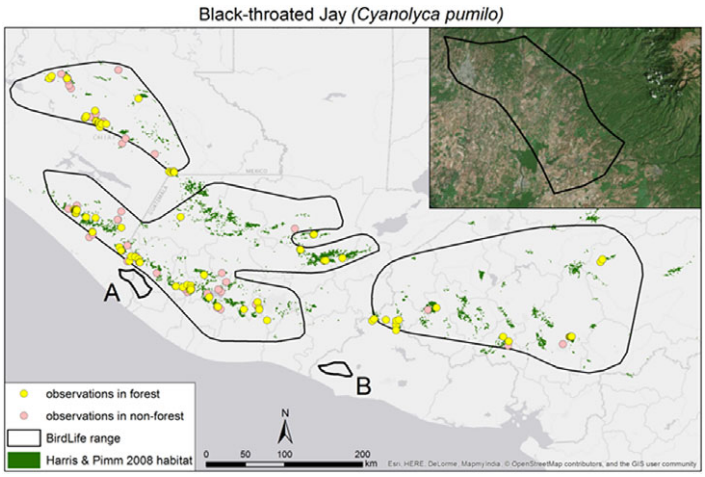

Figure 1. The range for the Black-throated jay as supplied by BirdLife International, Peterson et al.'s (2016) compilation of all the known observations of the species (circles) and those observations within forest (yellow). Areas A and B are examples of where Peterson et al. predict the jay occurs (A, approximately 10 × 20 km, shown in the inset satellite image, has been largely converted to human use; B, montane forest well outside the species' known range, well visited by birdwatchers none of whom have sighted the species). Imagery source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, AeroGRID, IGN, IGP, swisstopo, and the GIS user community.

(Fig. 1). In contrast, taking the specific locations of the observations literally as grist for models caused Peterson et al. to predict ranges poorly. Peterson et al.'s analyses fail to support their critique of our work. Instead, it illustrates why our approach is useful in a range of practical, data-poor situations in which automated techniques may be more susceptible to sample bias and poorly resolved data.

Again, what matters is how well the maps predict. How do we move forward? For the few species with many observational data, Peterson et al. could screen and clean these data appropriately, correct for sample bias (Elith & Leathwick 2009), and use appropriate and finer-resolution satellite imagery. One must assess the likelihood that an observation was in a species' habitat or at some unspecified distance from it, an approach that unavoidably and usefully includes assumptions about the suitability of habitat and the reliability of each record. One might consider exploring outside a species known range when there is a clear absence of any bird records in the suspected areas—as we did for birds in Brazil (Alves et al. 2008) and Colombia (Ocampo-Penuela & Pimm 2014).

At SavingSpecies, our aim is to connect isolated habitat fragments in areas with a high probability of containing the threatened target species we seek to conserve. We do so in close coordination with local conservation groups. We would not invest in areas that Peterson et al.'s methods speculate might contain species of interest but are well outside their known ranges and in habitats natural history suggests are unsuitable.

Supporting Information

A list of papers in which our methods were used and annotated methods from Harris and Pimm (2008) (Appendix S1) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

Alves MAS, Pimm SL, Storni A, Raposo MA, Brooke MdL, Harris G, Foster A, Jenkins CN. 2008. Mapping and exploring the distribution of the vulnerable grey-winged cotinga *Tijuca condita*. Oryx **42:**562–566.

BirdLife International. 2016. Datazone. BirdLife International, Cambridge, United Kingdom. Available from http://datazone.birdLife.org/home (accessed December 2016).

Cowling R, Pressey R, Rouget M, Lombard A. 2003. A conservation plan for a global biodiversity hotspot—the Cape Floristic Region, South Africa. Biological Conservation **112:**191–216.

eBird. 2016. Available from http://ebird.org (accessed December 2016).

Elith J, Leathwick JR 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics **40:** 677–697.

Harris G, Pimm SL. 2008. Range size and extinction risk in forest birds. Conservation Biology **22:**163–171.

Harris GM, Jenkins CN, Pimm SL. 2005. Refining biodiversity conservation priorities. Conservation Biology **19:**1957–1968.

Howell SN, Webb S. 1995. A guide to the birds of Mexico and northern Central America. Oxford University Press, New York.

IUCN (Interntional Union for Conservation of Species). 2016. IUCN red list of threatened species. IUCN, Gland, Switzerland. Available from http://www.iucnredlist.org/ (accessed December 2016).

Jenkins CN, Pimm SL, Alves MdS. 2011. How conservation GIS leads to Rio de Janeiro, Brazil. Natureza & Conservação **9:**152–159.

Jennings MD. 2000. Gap analysis: concepts, methods, and recent results. Landscape Ecology **15:**5–20.

Li BV, Hughes AC, Jenkins CN, Ocampo-Penuela N, Pimm SL. 2016. Remotely sensed data informs Red List evaluations and conservation priorities in Southeast Asia. PLOS ONE **11** (e0160566) DOI: http://dx.doi.org/10.1371/journal.pone.0160566.

Li BV, Pimm SL. 2015. China's endemic vertebrates sheltering under the protective umbrella of the giant panda. Conservation Biology **30:**329–339.

Ocampo-Penuela N, Pimm SL. 2014. Setting practical conservation priorities for Birds in the Western Andes of Colombia. Conservation Biology **28:**1260–1270.

Ocampo-Peñuela N, Jenkins CN, Vijay V, Li BV, Pimm SL. 2016. Incorporating explicit data geospatial shows more species at risk of extinction than the current Red List. Science Advances **2** (e1601367) DOI: 10.1126/sciadv.1601367.

Peterson AT, Navarro-Sigüenza AG, Gordillo A. 2016. Assumption-versus data-based approaches to summarizing species' ranges. Conservation Biology DOI: 10.1111/cobi.12801.

Scheffers BR, Joppa LN, Pimm SL, Laurance WF. 2012. What we know and don't know about Earth's missing biodiversity. Trends in Ecology & Evolution **27:**501–510.

Schnell JK, Harris GM, Pimm SL, Russell GJ. 2013*a*. Quantitative analysis of forest fragmentation in the Atlantic Forest reveals more threatened bird species than the current Red List. PLOS ONE **8** (e65357) DOI: http://dx.doi.org/10.1371/journal.pone.0065357.

Schnell JK, Harris GM, Pimm SL, Russell GJ. 2013*b*. Estimating Extinction risk with metapopulation models of large-scale fragmentation. Conservation Biology **27:**520–530.

Scott JM, Davis F, Csuti B, Noss R, Butterfield B, Groves C, Anderson H, Caicco S, D'Erchia F, Edwards TC Jr. 1993. Gap analysis: a geographic approach to protection of biological diversity. Wildlife Monographs:3–41.

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Campbell Grant EH, Veran S. 2013. Presence-only modelling using MAXENT: When can we trust the inferences? Methods in Ecology and Evolution **4:**236–243.