RESEARCH ARTICLE

Frontiers of Biogeography
the scientific journal of
the International Biogeography Society

Operationalizing expert knowledge in species' range estimates using diverse data types

Cory Merow[1], Peter J. Galante[2], Jamie M. Kass[3], Matthew E. Aiello-Lammens[4], Cecina Babich Morrow[2], Beth E. Gerstner[5], Valentina Grisales-Betancur,[6] Alex C. Moore[2], Elkin A. Noguera-Urbano[7], Gonzalo E. Pinilla-Buitrago[8,9], Jorge Velásquez-Tibatá[10], Robert P. Anderson[8,9,11] and Mary E. Blair[2]

[1]Eversource Energy Center and Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs CT, USA.
[2]Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY, USA.
[3]Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Kunigami-gun, Okinawa, Japan.
[4]Department of Environmental Studies and Science, Pace University, Pleasantville, NY, USA.
[5]Department of Fisheries & Wildlife and Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA.
[6]Universidad EAFIT, Medellín, Antioquia, Colombia.
[7]Instituto de Investigación de Recursos Biológicos, Alexander von Humboldt, Bogotá, D.C., Colombia.
[8]Department of Biology, City College of New York, City University of New York, New York, NY, USA.
[9]Ph.D. Program in Biology, Graduate Center, City University of New York, New York, NY, USA.
[10]International Alliances Program, National Audubon Society, Bogotá, Colombia.
[11]Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, New York, NY, USA.
Correspondence: Cory Merow, cory.merow@gmail.com

Abstract

Estimates of species' ranges can inform many aspects of biodiversity research and conservation-management decisions. Many practical applications need high-precision range estimates that are sufficiently reliable to use as input data in downstream applications. One solution has involved expert-generated maps that reflect on-the-ground field information and implicitly capture various processes that may limit a species' geographic distribution. However, expert maps are often subjective and rarely reproducible. In contrast, species distribution models (SDMs) typically have finer resolution and are reproducible because of explicit links to data. Yet, SDMs can have higher uncertainty when data are sparse, which is an issue for most species. Also, SDMs often capture only a subset of the factors that determine species distributions (e.g., climate) and hence can require significant post-processing to better estimate species' current realized distributions. Here, we demonstrate how expert knowledge, diverse data types, and SDMs can be used together in a transparent and reproducible modeling workflow. Specifically, we show how expert knowledge regarding species' habitat use, elevation, biotic interactions, and environmental tolerances can be used to make and refine range estimates using SDMs and various data sources, including high-resolution remotely sensed products. This range-refinement approach is primed to use various data sources, including many with continuously improving spatial or temporal resolution. To facilitate such analyses, we compile a comprehensive suite of tools in a new R package, maskRangeR, and provide worked examples. These tools can facilitate a wide variety of basic and applied research that requires high-resolution maps of species' current ranges, including quantifications of biodiversity and its change over time.

Highlights

- Species distribution models (SDMs) are widely used to estimate species' distributions; however, there are drawbacks when data are sparse, with implications for rare and threatened species.
- Expert knowledge—which we define broadly to include understanding of species' habitat use, dispersal barriers, biotic interactions, and other factors that limit their distributions—is a mostly untapped resource for post-processing statistically modeled range predictions to increase spatial and temporal accuracy.
- We formalize a framework to incorporate expert knowledge to reproducibly generate species' realized distribution estimates. Additionally, we operationalize a workflow with a new R package, maskRangeR.
- maskRangeR facilitates improved transparency and reproducibility of integrating diverse expert knowledge into refined estimates of species' ranges at high spatial and temporal resolutions to better describe geographic distributions for poorly sampled species, estimate biodiversity, and inform conservation decisions.

Keywords: biotic interactions, dispersal barriers, ecological niche models, expert map, land cover, land use, realized distribution, reproducibility, species distribution models

e-ISSN: 1948-6596 https://escholarship.org/uc/fb doi:10.21425/F5FBG53589

Introduction

Many important biological and biogeographical applications require reliable information on species' current geographic distributions (Whittaker et al. 2005). Ideally, data regarding species ranges would cover the entire geographic extent of the realized distribution at a fine spatial resolution and include a temporal resolution useful for detecting changes. However, the majority of species are known from very few localities (e.g., (Soberón et al. 2000)), a common problem termed the 'Wallacean shortfall' that can make constructing data-intensive species distribution models (SDMs) prohibitive for many species (Di Marco et al. 2017). This highlights the need for tools that can improve spatial and temporal characterizations of species' distributions. Such advances will enable more comprehensive and precise biodiversity assessments, particularly for conservation planning. For example, as low-data species are concentrated in some of the most biodiverse places in the world, the ability to improve taxonomic coverage of distribution knowledge and increase its spatial and temporal resolution is critical to understanding, predicting, and conserving biodiversity globally (Heberling et al. 2021, Jetz et al. 2019, Jung et al. 2020).

Expert knowledge may be essential to estimate distributions for species with only limited occurrence and ecological data available. We define expert knowledge broadly to include information regarding species' habitat use, dispersal barriers/accessible areas, biotic interactions, and other factors limiting their ranges. Expert knowledge could be reflected in something as simple as a hand-drawn map of accessible areas to reflect known dispersal barriers. Or it could be reflected in remotely sensed land use/land cover data that identify obligate habitat types (e.g., forests), or in conjunction with a threshold that indicates minimum/maximum suitable values of such a layer (e.g., minimum percent of forest cover). Alternatively, expert knowledge might constitute a known physiological threshold (e.g., thermal or freeze tolerance) based on lab experiments that can be used in conjunction with environmental layers to identify suitable locations. Such information is a largely untapped resource for post-processing statistically modeled range predictions to increase spatial and temporal accuracy.

Expert knowledge can include information on the focal species' environmental constraints: elevational limits, associated habitat types and land use/land cover, necessary levels of habitat quality, tolerance for anthropogenic influence, required distance from resources, key competitors or other biotic interactors, dispersal limitations, and historical contingencies (Anderson 2012, Araújo and Peterson 2012, Ocampo-Peñuela and Pimm 2014, Barve et al. 2011, Anderson and Raza 2010). Expert knowledge for a single species is often based on informal observations of field biologists that have not been entered into any database. Consequently, such information is a useful complement to the large online data aggregators that provide spatially precise observations from natural history museums and citizen scientist networks (Anderson et al. 2020). For data-limited species, we address the use of expert knowledge for determining: a) which abiotic, biotic, and/or spatial variables are important for estimating species' distributions either in the absence of—or ideally in conjunction with—a statistical model; and b), which values of those variables indicate that a given area is suitable and/or occupied. Here, we employ such information in a variety of approaches to refine SDMs.

Whether species distributions are initially estimated using expert-drawn maps, delimiting polygons (e.g., convex hulls), SDMs (Ocampo-Peñuela et al. 2016, Brooks et al. 2019, Peterson et al. 2018), or some combination of these (Merow et al. 2016, Merow et al. 2017), all benefit from some form of post-processing to improve the estimation of currently occupied locations. Indeed, recent work has shown that overestimation of range size from SDMs remains a problem in conservation applications (Velazco et al. 2020). The process of refining species range maps with expert-derived filters can be interpreted as moving from potential distributions towards realized distributions. Information derived from expert knowledge is often useful for determining unsuitable and unoccupied locations, and thus can help refine estimates (Velásquez-Tibatá et al. 2019, Calixto-Pérez et al. 2018, Skroblin et al. 2021). Generally, the initial maps (especially SDM predictions) are best interpreted as potential distributions, since other factors restrict species' distributions besides abiotic conditions. Such factors include dispersal limitations, historical contingencies, competitive interactions, social dynamics, and human modification of habitat. Realized distributions constitute those locations that are actually occupied by the species and are usually smaller than potential distributions. Importantly, relevant expert knowledge is typically related to non-climatic factors, making it complementary to SDMs commonly fit with climatic variables (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Elith and Leathwick 2009). Beginning with an expert-drawn map or SDM, we can use a series of filters to refine estimates of realized distributions based on expert knowledge (Anderson and Martínez-Meyer 2004), enabling the calculation of range sizes and other useful metrics of biodiversity and its change.

An example of documented, reproducible expert knowledge in action is a recent innovation by the Colombia Biodiversity Observation Network (BON) through the tool *BioModelos* (Velásquez-Tibatá et al. 2019). The *BioModelos* user community of scientists employs expert opinion to clean occurrence data and then refine resulting SDM predictions by identifying regions where the model either over- or underpredicts the realized distribution. This template, developed on a national scale, is primed to be applied more broadly using the principles and tools we develop below. We build upon these efforts in partnership with the Colombia BON by operationalizing tools to refine range maps in various transparent and reproducible ways. These tools can form the basis for supporting many applied conservation and management uses, including biodiversity assessments, reintroduction plans, land-use planning, and assessing progress toward global targets and biodiversity offset manuals (Velásquez-Tibatá et al. 2019, Araújo et al. 2019, Khoury et al. 2019).

Here, we formalize a conceptual framework and workflow to integrate expert knowledge into estimates of species' current distributions and develop code to make this process transparent and reproducible according to modern open science standards (Wolkovich et al. 2012). We focus on methods that begin with range predictions from either expert-drawn maps or SDMs, and then employ expert knowledge to refine these predictions using a series of filters (or geographic 'masks') to better estimate current realized distributions. Our approach aims to define methods for operationalizing expert knowledge, often by harnessing available *in situ* and *ex situ* biological and environmental data regarding factors identified as important by experts. We provide an R package, maskRangeR, that automates processes within this workflow, thus improving the transparency and reproducibility of integrating such knowledge into refined estimates of species' ranges at high spatial and temporal resolutions.

Operationalizing expert knowledge

Our approach incorporates expert knowledge as either a complement or alternative to statistical models (Fig. 1). We advocate that statistical models

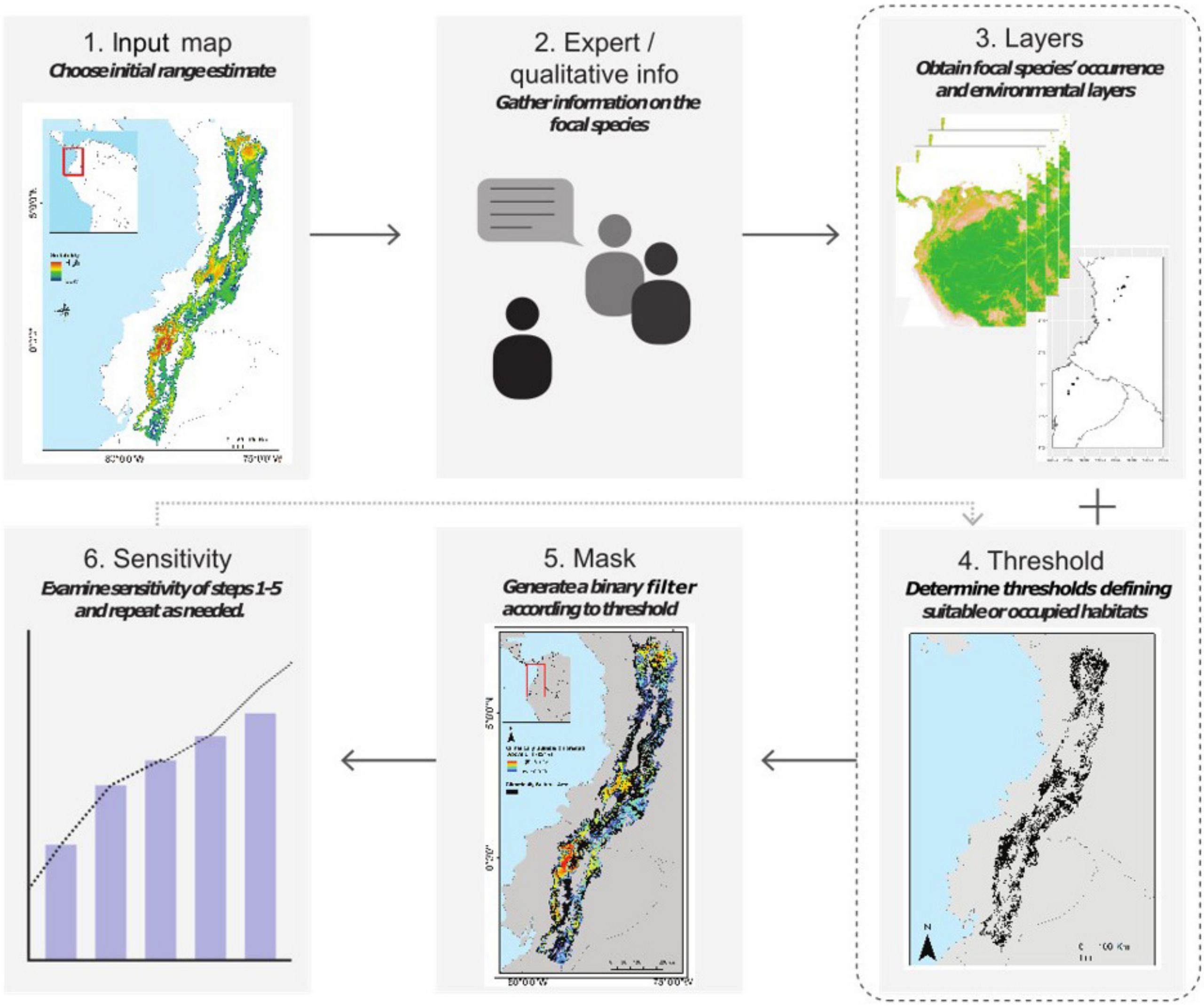

Figure 1. An overview of the steps in the maskRangeR workflow. 1) Input map: For example, choose a map from a previously built species distribution model (SDM) prediction as an initial range estimate. 2) Expert information: Gather information (such as habitat associations) on the species' distribution to inform Steps 3 and 4. 3) Filtering layers: Obtain/generate environmental layers of abiotic or biotic factors that influence the species' distribution but were not considered in building the SDM, as well as (optionally) occurrence records for associating with the environmental layers. 4) Generate a mask: Determine thresholds for those layers (not the SDM) defining suitable or occupied habitat using either expert opinion, environmental values at known localities, or both. 5) Mask: Generate a binary filter by indicating values of layers above/below (as appropriate) the threshold in Step 3. 6) Sensitivity Analysis: Sensitivity of Steps 2-5 should be examined with respect to each expert decision to either synthesize across competing plausible scenarios or to determine the most plausible scenario and assess its dependence on key assumptions. After the sensitivity analysis, any previous step can be repeated with additional layers, thresholds, and masks as needed. The output from one procedure (Steps 1-5) can be used as the input for another. Images shown are for the first use case described in the text (the olinguito; *Bassaricyon neblina*).

are preferable when sufficient occurrence data are available, as they should reconstruct expert-derived relationships if the data support them. For example, forest cover may be a limiting factor for a species' distribution; therefore, a measure of it should ideally be incorporated as an SDM predictor variable along with aspects of climate when the desired final result is a map of the species' current range. However, when sufficient occurrence data are not available to appropriately characterize a limiting factor, expert knowledge can be crucial for identifying the bounds of the species' tolerances for such variables. Below, we describe common scenarios in which sufficient occurrence data are lacking or statistical models would be inappropriate or excessively complex, and in each case outline how expert knowledge can instead be combined with existing data to refine range estimates.

We partition the process into a series of 'expert filters' applied to an initial range estimate, which we term an 'input map' (see Methods, below). These filters define various attributes of suitable environmental conditions, habitat requirements, and other limiting factors (e.g., dispersal barriers); when combined together, they result in what we term an 'expert-refined map.' For example, one filter may define a reasonable geographic domain (thereby filtering out the rest of the globe from consideration), a second may define suitable elevational limits, and a third may define minimum forest-cover requirements. Any locations that pass through all three filters are considered amenable for the species and are collectively denoted as the expert-refined map. In this way, using a series of filters can be considered analogous to the strictest implementation of ensemble modeling (Thuiller et al. 2009), where complete agreement is required between all component models (i.e., the filters) to denote a location as suitable/occupied.

This range-refinement approach is primed to use various data sources, including many with continuously improving spatial or temporal resolution. Products derived from remote sensing are increasingly available at high spatial resolutions. They can provide environmental information key to improving species' distribution knowledge and quantifying how species' distributions respond to environmental change over space and time. For example, new environmental data are coming online (e.g., soil moisture from NASA's SMAP project[1]), while others are increasingly resolved at finer resolution (e.g., forest structure based on NASA's GEDI mission[2]). Habitat characterization (e.g., classification, metrics, and continuous fields/fraction cover) is particularly critical because it tends to vary more rapidly over a fine spatial resolution than the climatic data often used to characterize distributions. Habitat information may help improve rare species' predictions, as their distributions are more likely to be limited by habitat type than by climate (Brooks et al. 2019). Expert knowledge of how these data sources can inform species' distributions is critical for improving range estimates, particularly for poorly sampled species for which statistical models are not viable.

Below, we present three use cases that represent common scenarios for such a workflow, and for each of them demonstrate how expert knowledge can be operationalized to refine or define range predictions. Although we provide a specific worked example for each of the use cases using an SDM or an expert-drawn map, either could be used as the input map. The first use case refines an SDM prediction based on environmental tolerances characterized through remote-sensing data (e.g., forest cover). The second refines expert-drawn maps using inferred environmental tolerances via a sequential filtering approach. The third removes biotically unsuitable areas (i.e., within the range of a competitor) from SDM predictions of congeneric, parapatric species. While the masking procedures for all three use cases have been used in various specific circumstances to generate or refine maps (Peterson et al. 2018, Brooks et al. 2019, Anderson and Martínez-Meyer 2004, Kass et al. 2021), we offer a generalized operational framework to conform these processes to modern open science standards.

Materials and Methods

Distribution maps can be generated or refined by a variety of mechanisms depending on the level of detail of the expert's knowledge and data available. Each of the following examples refines an initial range estimate using expert knowledge (without statistical inference) to determine which biotic or abiotic variables limit a species' distribution and/or what ranges of these variables constitute suitable conditions. In the literature, the term 'expert map' generally refers to a map composed of polygons that have been drawn by hand or with GIS software; for clarity, we refer to these with the term '**expert-drawn map**' (e.g., one option of the input into the filtering steps). As mentioned above, we generalize the concept of harnessing expert information by terming the output an '**expert-refined map**': a map that incorporates any type of expert knowledge, as in post-processing via the three use cases described below. This generalization anticipates the many types of expert knowledge that are challenging or impossible to incorporate into a statistical model or to readily draw on a map.

Step 1: An input map

The first step consists of selecting or building a preliminary input map to refine in subsequent steps. We focus on two cases where input maps are most readily available: expert-drawn maps and SDMs. However, this approach can generally be applied to range maps derived from any type of previous model which is (ideally) independent of the information used for expert filtering (including convex hulls (Busby 1991), phenological models (Chapman et al. 2014), or

1 https://smap.jpl.nasa.gov/

2 https://gedi.umd.edu/

demographic models (Merow et al. 2014, Merow et al. 2017), among others).

For many rare and endangered species, there may be limited observations available to use as inputs for an SDM. Rather, IUCN openly serves expert-drawn polygon maps for most vertebrate species and other taxa as they become available[3]. In these cases, expert maps may be the only estimates of a species' range. However, these are often not reproducible, are difficult to update as new data become available, and have a precision that is too coarse to be meaningful for many conservation activities (Ocampo-Peñuela et al. 2016, Hurlbert and Jetz 2007, Rotenberry and Balasubramaniam 2020, Di Marco et al. 2017). Nevertheless, expert-drawn maps are still valuable, as they often reflect a range of processes, including habitat selection, dispersal limitation, and biotic interactions that must be considered to estimate a species' realized distribution. Expert maps are often successful at delimiting areas beyond polygon boundaries as outside the species' distribution, but because they are typically drawn at a coarse spatial resolution, they tend to suffer from inferred false presences within polygon boundaries (e.g., (Mainali et al. 2020)). Hence, they are prime candidates for further refinement based on high-resolution data filters guided by expert knowledge.

In the absence of an expert-drawn map (particularly for a very poorly sampled or rare species), one might coarsely define a relevant geographic domain where the focal species is hypothesized to occur. This domain might constitute a biome (within a conscripted geographic region) or an ecoregion to impose rough boundaries based on the types of communities in which the species is expected to occur, a political unit where interest lies in refining range knowledge, or a buffered area around known occurrences. In such cases, our expert-based workflow describes a formalized protocol for generating a map based strictly on expert knowledge and harnessing relevant available data (i.e., without statistical inference).

SDMs constitute the third common type of input map. They can inform a wide range of applications for biodiversity and conservation, including prioritizing target species and high-diversity areas as well as guiding surveys, reintroductions, and monitoring efforts (e.g., Raxworthy et al. 2003). However, many challenges remain to maximize SDM utility for conservation and other applications (reviewed in (Urbina-Cardona et al. 2019, Villero et al. 2017)), including issues of small sample sizes, sampling bias, and interpolation/extrapolation to unsampled locations. Nevertheless, SDMs may be available as part of published studies from large data aggregators such as the Botanical Information and Ecology Network (BIEN; > 300,000 plant species[4]), or from users building their own SDMs with the intention of further refining them with expert knowledge. SDMs excel at predicting species' potential distributions and hence are ripe for refinement via expert knowledge to make better estimates of realized distributions.

Generally, SDMs are preferable to expert-drawn maps as an input in our workflow. They directly use observations in a reproducible and largely objective way and often reflect the experience of many people (data collectors), in contrast to expert-drawn maps that often involve the subjective discretion of a single or few individuals. However, occurrence data may exhibit sampling bias, which can have considerable effects on inferred ranges (Phillips et al. 2009, Yackulic et al. 2013). In contrast, the implicit assumption associated with expert maps is that bias related to variation in knowledge across the region is low. Further, sufficient sample sizes for fitting SDMs may not be available, leaving expert maps as a better option. Nevertheless, if SDMs are feasible, they can provide additional information on variation in suitability (whereas expert maps are typically binary), make predictions to unsampled areas, and be used to quantify uncertainty in range estimates.

Step 2: Expert knowledge and estimating thresholds

Expert knowledge can often be represented as a conceptual model for binary classification, usually relating to a single variable (e.g., forest cover, occurrence of competing species, dispersal) that indicates species' presence or absence. For example, a montane species may be known to occur only above 3000 m elevation, an invasive species may occur only in locations with anthropogenic disturbance, a plant may require serpentine soil, or a forest specialist may occur only in locations above a critical threshold of forest cover. Hence, generating a binary filter consists of identifying relevant filtering layers in Step 3 below and selecting a suitable threshold that can be used to omit portions of the study region that are either unsuitable or unoccupied. In GIS terminology, the operation of applying such binary filters to a map is referred to as 'masking'—thus, implementing these expert mapping tools corresponds to generating a collection of masking layers.

We consider two approaches to generating thresholds that delimit suitable/occupied locations from unsuitable/unoccupied ones. We refer to the first approach as '**expert-driven thresholding**', in which an expert has *a priori* knowledge of requirements for a given filter to distinguish suitable conditions from unsuitable ones. For example, the elevation limits often found in field guides for montane species can be combined with a digital elevation model to refine which locations are above or below the expert's threshold (Ocampo-Peñuela et al. 2016). Similarly, the expert may estimate that a grid cell must contain at least 50% forest to be suitable for a forest-obligate species.

A second approach, '**data-driven thresholding**', helps an expert determine an appropriate threshold by overlaying observed presence data on filter layers (Gavrutenko et al. 2021). This approach involves first extracting values of the filter layer at presence

3 https://www.iucnredlist.org/resources/spatial-data-download
4 https://biendata.org

locations, then either choosing a threshold (based on expert opinion) or suitable quantile from these values such that the level of omission is acceptable for the given application. An expert may have knowledge of habitat requirements but only possess a rough estimate of the quality of habitat required to distinguish suitable from unsuitable areas, e.g., the critical threshold of the percentage of forest cover in a particular area (grid cell) necessary to qualify as suitable. In this case, known presence locations of the species can be used to guide the expert's estimate of this threshold by using forest cover data available in those pixels (time-matched to correspond to the species' observation). Reliability of the data-driven thresholding will depend in part on how well occurrences match environmental data, in terms of resolution (georeferencing error vs. spatial resolution) and timing.

Step 3: Filtering layers

In this step, we focus on obtaining relevant layers (Table 1) for filtering out unsuitable or unoccupied portions of the initial map from Step 1. Filtering layers can be either binary or continuous, and this choice determines how they can be used with the input map, which we discuss further in Step 5: Masking Methods.

Common data sources on environmental conditions for post-processing an input map include land use/land cover (Hurtt et al. 2011), elevation (Robinson et al. 2014) and its heterogeneity (Amatulli et al. 2018) (https://www.earthenv.org/), habitat type (Tuanmu and Jetz 2014) and its heterogeneity (Tuanmu and Jetz 2015), soil type (soilgrids.org), and ecoregion (Dinerstein et al. 2017). Such information is often available on a continuous scale of measurement (e.g., describing the proportion of a grid cell classified in a given category, such as forest cover, which can be interpreted to represent habitat quality). Alternatively, at low spatial resolution such layers

Table 1. Examples of useful large-extent data sources for masking.

Data type	Summary	Spatial Resolution	Temporal Resolution	Citation
Ecoregion	846 terrestrial ecoregions updated from (Olson et al. 2001)	Various	Present	(Dinerstein et al. 2017)
Elevation	Based on a variety of fused data sets to provide near global coverage	90 m	Present	(Robinson et al. 2014)
Land Use/ Landcover	Fractional land use change used in the IPCC assessment	0.5 degrees	Annual	(Chini et al. 2014)
Forest Cover	Forest extent, loss, and gain	30 m	Annual	(Hansen et al. 2013)
Habitat Heterogeneity	14 metrics quantifying spatial heterogeneity of global habitat based on the textural features of the Enhanced Vegetation Index	1 km, 5 km, 25 km	Present	(Tuanmu and Jetz 2015)
Cloud Cover	Mean cloud frequency with near global coverage	1 km	Present	(Wilson and Jetz 2016)
MODIS Active Fire/Burned Area	Timing and spatial distribution of fires and their characteristics	1 km/500 m	8 days	(Giglio et al. 2016, Giglio et al. 2018)
SoilGrids	Physical and chemical soil properties estimated for each horizon	250 m	Present	(Hengl et al. 2017)
Human Footprint	Globally standardized measure of the cumulative human footprint based on infrastructure, land cover and human access into natural areas	1 km	Annual	(Venter et al. 2016)
Human modification	13 anthropogenic stressors and their estimated impacts	1 km	Present	(Kennedy et al. 2019)
Freshwater metrics	Near-global, spatially continuous, and freshwater-specific environmental variables	1 km	Present	(Domisch et al. 2015)
Surface water	History of annual water detection (1984 - 2019)	30 m	Present	(Pekel et al. 2016)

may simply indicate a majority classification for a cell, which can be interpreted as a binary map indicating whether or not a given class (e.g. soil type or land use) is present (Tuanmu and Jetz 2014). Aggregated anthropogenic factors, such as measures of human footprint (Venter et al. 2016) or human modification (Kennedy et al. 2019), can also be useful for omitting regions that would otherwise be inferred as suitable based on environmental conditions.

Biotic interactions often further restrict species distributions markedly within the environmental conditions that are suitable (Wisz et al. 2013). Therefore, knowledge of key interactions between species is relevant, and the distributions of interacting species can help to further refine range estimates of the focal species (Gutiérrez et al. 2014, Anderson 2017, Freeman and Mason 2015). For example, if a species has obligate pollinators, dispersers, or other mutualisms, one can use range estimates of these mutualistic species as a filter (retaining only the areas where the needed interactor exists). As another example, an expert may have knowledge of parapatric boundaries between closely related species (e.g., competitors) based on an observed lack of co-occurrence, different habitat requirements, or other factors leading to mutual exclusion. Based on this expert expectation of parapatry, one can partition a region with overlapping range estimates for two or more species into spatially distinct ranges for each species (Anderson et al. 2002). This can be particularly helpful near range boundaries, where the climatic conditions might be mutually suitable, yet the species do not co-occur. Rules for partitioning geographic space among species in such circumstances are described below with a recently developed spatial approach based on support vector machines (Kass et al. 2021).

Step 4: Generate a mask from a filter layer and threshold

A mask is generated by converting a continuous layer identified in Step 3 to a binary layer by identifying values above/below the threshold identified in Step 2. Alternatively, in the absence of knowledge of the specific factor driving occurrence in particular locations (e.g., Steps 2 and 3 are not possible), experts often refine the input map by drawing polygons to include/exclude regions from a species' range estimate based on their knowledge (e.g., for areas with high hunting pressure). This will often be done at a comparatively coarse spatial resolution. For example, *BioModelos* applies this approach to advance distribution knowledge in Colombia by providing graphical tools to experts, who then make modifications to existing maps (Velásquez-Tibatá et al. 2019).

Step 5: Masking methods

We consider three use cases for different combinations of input maps and ways to apply the thresholds for filter layers:

1. **Binary input map and binary filter**. In the simplest case, using a binary filter to mask a binary input map results in a binary expert-refined range estimate. For example, an expert-drawn map could be masked by elevational limits obtained from a field guide. Alternatively, a binary (thresholded) SDM could be masked by currently forested areas.
2. **Continuous input map and binary filter**. Using a binary filter to mask a continuous input map (e.g., describing probability of presence or relative abundance) results in a continuous, expert-refined range estimate. For example, a continuous input map, such as an SDM prediction, could be masked with a binary filter generated from the expert-defined lower limit of the species' tolerance for a human footprint layer.
3. **Binary input map and continuous filter**. Binary range estimates (e.g., expert-drawn maps or thresholded SDM predictions) can be refined based on continuous filters that describe either the proportion of a cell that is suitable/occupied or the corresponding habitat quality. For example, if a species is an obligate forest-dweller, a continuous filter describing the proportion of forest in a cell can be interpreted as the proportion of the cell that is suitable. Operationally, it is simplest to use the input binary map as a mask on the continuous filter to compute the expert-refined map.

It is usually not advisable to combine both a continuous input map and a continuous filter in their raw form (rather, one or the other should be converted to binary as in Cases 2 and 3 above). Typically, input maps or filter layers with continuous values represent probabilities or proportions, but it is often unclear how these values interact within a cell (i.e., the probabilities are not independent and their joint distribution is unknown). Careful applications of fundamental rules of probability must be considered in order to use these continuous values effectively; however, the information needed to know how to apply these rules is often lacking. For example, if the probability of presence in a cell is 0.6 and the proportion of forest is 0.5, we cannot simply assume that the probability of presence should be reduced to 0.6 * 0.5 = 0.3 using the logic that only half the cell is actually viable. If the species is an obligate forest-dweller (i.e., the occurrence probability is not independent of forest cover), this would imply that the occurrence probability of 0.6 applies only to the forested portion of the cell. Yet another possibility is that the occupied proportion may be between these values, due to the species' exclusive preference for forest. Mathematically, the assumption here is that the probability of presence and the proportion of forest are independent, which is biologically unrealistic. While special cases surely exist that clearly define the relationships between continuous layers, we do not consider them further here.

A few special cases of masking techniques can extend the utility of our proposed framework. Thus far we have considered a single snapshot of a species' range. However, time-series of filters may be available in some cases, which can allow one to estimate temporal trends in suitable areas. For example, land use/land cover products such as forest cover (Hansen et al. 2013) can be available at an annual resolution. In the examples below, we demonstrate a case where a binary input map was estimated from a SDM and annual forest cover layers were then used to estimate changes in range size over the last two decades.

Step 6: Sensitivity analysis

While many of the approaches presented above have appeared as applications in various studies, sensitivity of expert refinements and characterizing uncertainty in predictions has largely been ignored. In our expert-refined mapping framework, sensitivity analyses to modeling decisions involve up to four types of comparisons, following the steps outlined above. First, one can compare outputs using different combinations of biotic or abiotic filter layers. Second, one can compare different masking approaches: e.g., an expert-determined threshold compared with a data-driven one. Third, one can compare different plausible values of thresholds used to create a binary mask (Fig. 2). Fourth, one could compare maps developed by different experts and either average them or use them as a platform to achieve consensus, as in BioModelos (Velásquez-Tibatá et al. 2019). These analyses can be used to make decisions that are relatively insensitive to perturbations for creating a single model, or to consider an ensemble of plausible models that characterize uncertainty. Furthermore, reporting this uncertainty (e.g., via maps or in estimates of range size) helps to convey the level of confidence in the final expert-refined map.

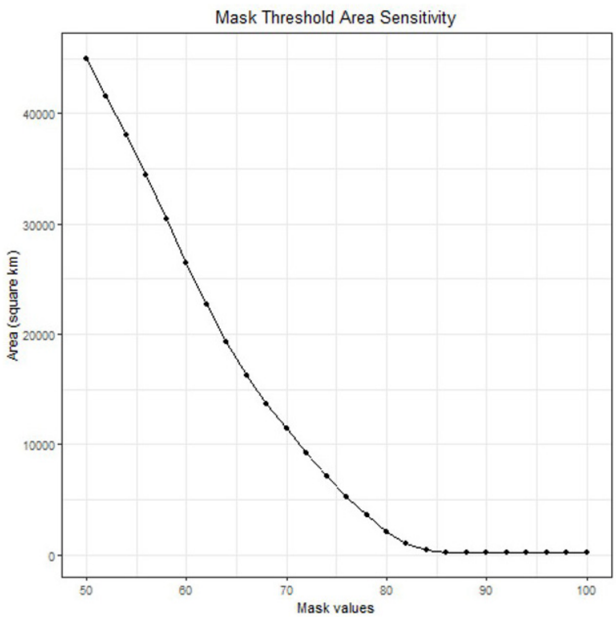

Figure 2. Areal range estimates based on masking the species distribution model for the olinguito (*Bassaricyon neblina*) at regular threshold intervals between 50-100% forest cover. While expert knowledge can be used to determine the appropriate threshold, it is important to note the considerable variation in range size that is associated with different thresholds. For more, see 'Example 1: Masking by forest cover' under 'Use Case examples'.

It may be particularly informative to compare maps from different expert- or data-driven threshold values, as expert knowledge or estimation of the precise quantitative threshold is often approximate. As a first step, one can visually inspect maps for ecological realism (Guevara et al. 2018). Additionally, it can be useful to plot summary statistics of the range, such as measures of range size or range geometry, as a function of different threshold values (Fig. 2). Ideally, one can select a threshold value from within the plausible range where changes in threshold do not result in large changes to areal predictions, indicating low sensitivity to that decision (defined as the first derivative of a plot of the range metric vs. the threshold). Whether one selects a threshold from the upper, middle, or lower value of such a low-sensitivity region will depend on the relative importance of omission versus commission in a given study, as well as on the biological interpretation of the filter involved.

Step 7: Repeat

Steps 2-6 can be repeated with as many filters as are available, and the resulting expert-refined maps will only include areas that have passed through all the filters.

Use Case examples

Each of the following is a summary of a worked example, with further details and associated R code available in the vignette of the newly developed maskRangeR package (Appendix S1). Each example uses key functions from maskRangeR to complete the analyses, which are detailed in the vignette, accessible at https://cmerow.github.io/maskRangeR/maskRangeR_Tutorial.html. Note that package plans, updates, and training materials are available on the package website (https://cmerow.github.io/maskRangeR/).

Example 1: Masking by forest cover (Fig. 3a-c**).** Determining thresholds for masking SDM predictions using recent records represents a simple methodology that can be used for the many species with limited recent records. Time-series of filters may be available in some cases, which can allow one to estimate temporal trends in suitable areas. For example, land use/land cover products such as forest cover (Hansen et al. 2013) can be available at an annual resolution. The olinguito (*Bassaricyon neblina*) is a recently described carnivoran discovered from previously misidentified museum specimens (Helgen et al., 2013). It lives in Northern Andean cloud forests in Colombia and Ecuador; according to experts, it likely has strict tolerances for forest cover (Helgen et al. 2013). The species' range was updated and estimated via an SDM (Gerstner et al. 2018) but without consideration of recent deforestation. Remotely sensed percent forest cover data now can be used to perform simple data-driven masking based on information corresponding to recent records (Gavrutenko et al. 2021), with percent forest cover

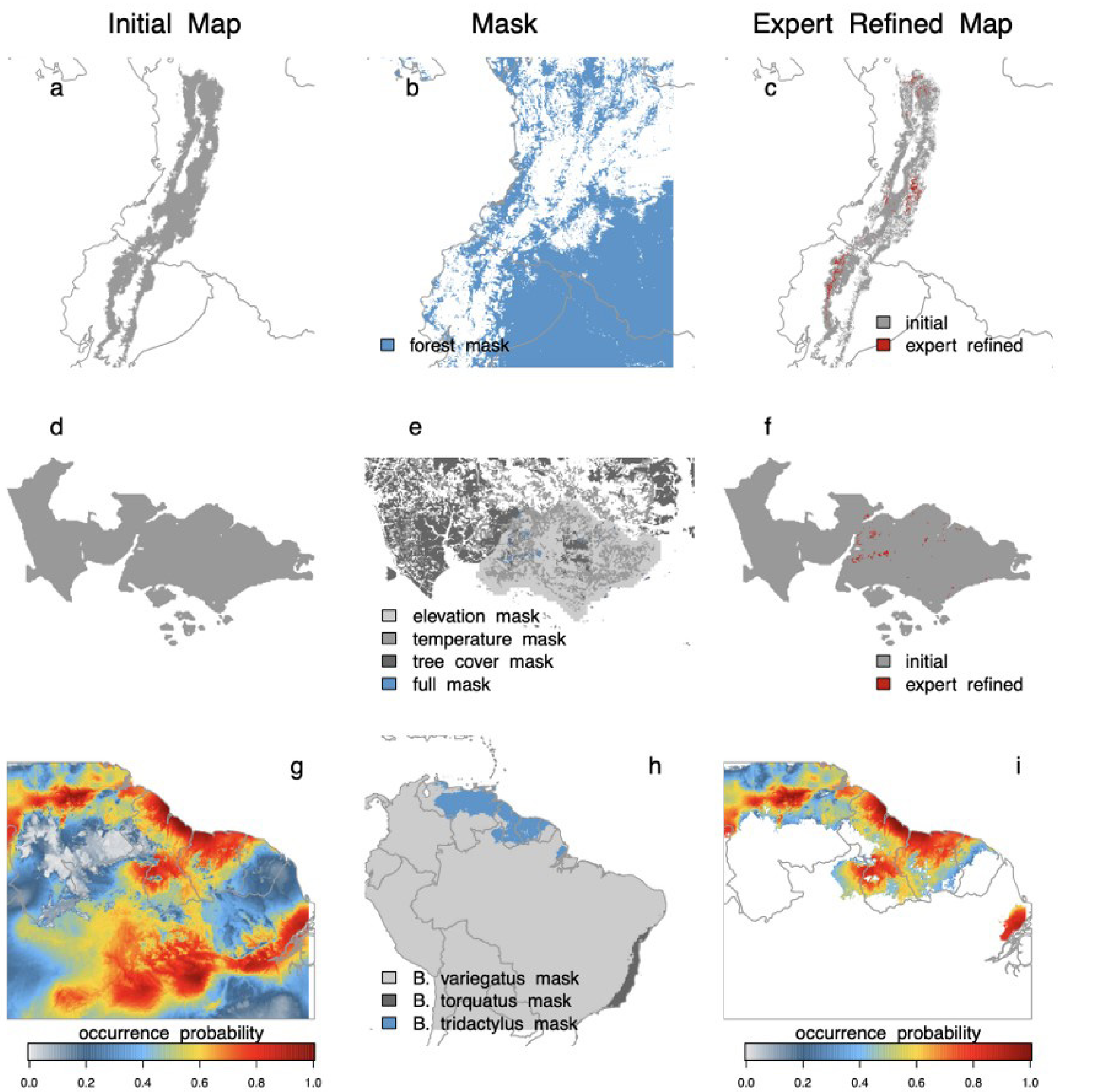

Figure 3: Use-case examples. (a-c) SDM predictions can be refined based on available forest habitat. Panel a shows the binary SDM prediction in Colombia and Ecuador for the olinguito (gray), and panel b shows suitable forest cover levels (blue). Panel c shows the same olinguito SDM prediction within areas with suitable forest cover (red, the expert-refined map) after areas with insufficient forest cover are removed through masking (gray). (d-f) The use of multiple masks can considerably refine coarsely estimated ranges. The initial map for the swamp forest crab in Singapore (d) can be processed by intersecting three masks (e) to compose the 'full' mask (blue, amenable based on all three), which is used to generate the expert-refined map (f). (g-i) Support Vector Machine (SVM) masks can be used to resolve overlapping SDM predictions of parapatric species. (g) Initial SDM prediction made with bioclimatic predictor variables and occurrence localities for *Bradypus tridactylus,* one of three parapatric sloth species in South America. The SVM (h) classifier layer was created using occurrence localities and SDMs (spatial-environmental SVM) for each of the three species. (i) Expert-refined SDM of *B. tridactylus* masked by the SVM classifier to remove areas more likely to be part of the ranges of its parapatric relatives.

extracted (e.g. MODIS Vegetation Continuous Fields (Daac 2018); (250 m) from the years matching the most recent occurrence records: 2006-2016). There are only limited occurrence records available for this species (N = 30) (and this is common for species of conservation concern (Di Marco et al. 2017)). Half of these are from museum specimens, with 80% of those from years where high-resolution forest cover estimates are not available. Only N = 18 recent (post-2000) occurrences can be matched to the corresponding year of MODIS forest cover data to calculate a forest-cover threshold appropriate for the species (here determined as 77% forest cover in a cell). This threshold can then be used to mask the SDM output for the most recent year of MODIS data for an upper estimate of the current species' range. Such a methodology should also prove useful for processing expert range maps or other pre-existing range estimates that do not take into account human modifications of the environment.

Example 2: Masking by multiple expert inferred tolerances (Fig. 3d-f**).** Some species may not have sufficient occurrence records with which to define environmental associations using an SDM. In these cases, for a given geographic area, multiple environmental layers can be thresholded to the estimated tolerances for the species based on expert knowledge, providing a series of masks. The swamp forest crab (*Parathelphusa reticulata*), a freshwater crab endemic to Singapore's last remaining patch of freshwater swamp forest, is currently listed as Critically Endangered by the IUCN and is documented by only a single occurrence record. Here, we perform an expert-driven species range estimate using three masks. We identified areas of overlap between canopy cover, mean annual temperature, and elevation that are considered by experts (Chua et al. 2015) as within the bounds of tolerance for this species, and then used these areas to constrain a relevant geographic region (also described by expert opinion) to an area of suitable environmental conditions.

Example 3: Masking by biotic factors (Fig. 3g-i**).** In addition to abiotic factors, species' distributions are also affected by biotic interactions, especially when species distributions are mutually exclusive (Wisz et al. 2012, Anderson 2017). For example, closely related species commonly show parapatric distributions (abutting but non-overlapping ranges); yet, including biotic interactors as predictors should be avoided in standard SDMs if both species affect each other; Anderson 2017). While it is straightforward to mask a focal species' distribution by the known range of a competing species that fully excludes it, the solution is less clear when the distributions of both species—and their effects on each other—are imperfectly known (which is likely the case for poorly sampled species) (Anderson et al. 2002, Kass et al. 2021). Some studies have addressed this by identifying the species with the highest suitability value predicted by an SDM fit with abiotic variables (Anderson and Martìnez-Meyer 2004, Gutiérrez et al. 2014). However, high abiotic suitability predictions for one species can be made far from the contact zone between both species. Such areas may be abiotically suitable but nonetheless unoccupied due to biotic and/or dispersal limitations. Therefore, considering the spatial positions of occurrence records, in addition to the SDM predictions, can help remedy this issue. Kass et al. (2021) used support vector machines (SVMs) (Drake et al. 2006) to classify grid cells by the species most likely to be present based on either the spatial patterns of the records alone (spatial), or those in addition to SDM predictions (which they termed spatial-environmental). These approaches can be used to delimit ranges for two or more parapatric species by using the classification output as filters on the focal species' respective input maps.

The three-toed sloths (*Bradypus* spp.) comprise four species distributed across Central and South America. With the exception of the microendemic *B. pygmaeus*, the other three species range across mainland Central and South America: *B. variegatus* is widely distributed from eastern Honduras to northern Argentina, *B. tridactylus* is found in the Guianan shield region, and *B. torquatus* occurs only in the Brazilian Atlantic Forest (Anderson and Handley 2001). *Bradypus variegatus* and *B. tridactylus* exhibit parapatry (with a few sites of inferred contact), which is likely caused by competition for resources and/or suitable habitat (de Moraes-Barros et al. 2010). On the other hand, *B. variegatus* and *B. torquatus* show overlapping ranges in some portions of the Brazilian Atlantic Forest, where they exhibit more localized geographic separations (Hirsch and Chiarello 2012). Using support vector machines (Drake et al. 2006), we delimited the distributions of these three species by masking out regions classified as more likely to be within the range of one or more congeners based on spatial and (optionally) also non-spatial predictors. This methodology fits SVMs with two kinds of data: spatial, based on the occurrence coordinates of each species, or spatial-environmental, based on a combination of the coordinates and the predicted suitabilities from SDMs (Kass et al. 2021). The SVM classifications are then used to mask the SDM predictions of each species to those regions not predicted to be occupied by the other congeners.

The maskRangeR package

To facilitate the use of our proposed framework for expert refinement of range maps, we developed the R package maskRangeR (website: https://cmerow.github.io/maskRangeR/; download from CRAN at https://cran.r-project.org/web/packages/maskRangeR/index.html or the development version at https://github.com/cmerow/maskRangeR). This package fills key gaps in existing tools needed for the workflow presented above, and additionally provides a convenient and unified approach to the variety of masking strategies discussed. The package also includes tools to visualize the consequences of various expert decisions as part of sensitivity analyses. Notably, in developing maskRangeR we leveraged a number of existing, high-quality tools for components of our workflow and focused on (1) adding tools missing from existing R packages and (2) providing all of these tools in a unified framework, accompanied by an extensive vignette with worked examples aimed at practitioners (Appendix S1).

A workflow with maskRangeR follows the steps described in Methods and Materials using all the core maskRangeR functions as follows:

- **Step 1**: The user provides an input map in raster format.
- **Step 2**: Users can define their own thresholds *a priori* or use maskRangeR's annotate() function to extract environmental values from timestamped occurrence locations and a time series of environmental layers. For static (single) layers, the raster package's extract() function is sufficient (Hijmans 2021). The user then chooses a threshold based on the distribution of these extracted values.
- **Step 3**: The user can provide filtering layers in raster format. Users can take advantage of data-download functionality from other R packages such as dismo (Hijmans et al. 2010) for elevation or WorldClim climate layers (Hijmans et al. 2005), forestChange for forest change indices (Lara and Gutierrez-Velez 2019), or MODIStsp for forest cover (Busetto and Ranghetti 2016). A wide variety of layers are openly available if users can download and process them in R (e.g., alternative bioclimatic layers from CHELSA (Karger et al. 2016). The function focalCompare() makes it easy to aggregate these layers with moving-window choices (using raster::focal()), so that users can compare the results of different decisions. This can be useful when aligning layers originally at different resolutions.
- **Step 4 and 5**: maskRangeR provides a number of functions to perform different types of masking. The core function, maskRanger(), generates binary masks by thresholding filter layers, and then applies them to the input map. The function lotsOfMasks() readily implements multiple masks at once, while continuousMask() uses a continuous filter to mask a binary input map. Finally, rangeSVM() runs an iterative cross-validation procedure to fit a support vector machine with optimal complexity, and rangeSVM_predict() uses this model to output masks for delimiting the ranges of multiple species that biotically constrain each other's ranges.
- **Step 6**: Sensitivity analysis can be performed with thresholdSensitivity(), which generates plots showing how range size depends on threshold choice.

Discussion

By refining range maps based on expert knowledge, we illustrate how to improve spatial and/or temporal precision by incorporating different types of information that would often be omitted from species distribution models. This results in a transparent and reproducible workflow and illustrates the complementarity of expert knowledge and SDMs. With these improvements to range estimates, we can better describe geographic distributions for poorly sampled species, estimate biodiversity, and inform conservation decisions. While formal frameworks exist for elicitation of expert knowledge (McBride et al. 2012), as well as formal methods to incorporate this knowledge into SDMs (Choy et al. 2009, Merow et al. 2017), our focus here is to offer a highly practical workflow to elicit and utilize expert knowledge for species' range estimates for a wide range of situations and applications.

Although practicality has driven the development of our workflow, it is useful to note that our approach is also consistent with theoretical constructs of the ecological niche that help to define a species' geographic distribution (Soberón 2007). In particular, Drake (2014) argued that the niche is better represented by hulls which delimit niche boundaries rather than the probabilities associated with species distribution (niche) models. Drake (2014) emphasized that "the niche is the range of environments in which the species can persist" and that this range is the "interval between two extremes". As such it is well represented by thresholds and manifests as a binary map in geographic space. The binary filters we generate here can be interpreted as convex hulls in environmental space, or the intervals between extremes (thresholds), and a collection of such filters represents a niche hypervolume, at least along the dimensions considered by the filters. Hence, our filtering approach is fully consistent with at least one interpretation of the niche concept.

Expert-refined range maps have the potential to advance the spatial and temporal resolution of range estimates in multiple ways. In the spatial domain, our masking workflow may allow larger sample sizes when fitting coarse-grain SDMs to be used as input maps, because spatially imprecise data can be included for a coarse-grain SDM (which might otherwise be discarded for a fine grain SDM (Moudrý and Šímová 2012, Mitchell et al. 2017, Naimi et al. 2014)). Improved spatial resolution could be achieved with a fine-resolution filter applied to such a coarse-grain input map, allowing for the use of all available data. In the temporal domain, it is uncommon for range estimates to be updated frequently (e.g., every year), but this is feasible when filtering layers are available as a time series. Such filter layers might include annual climate layers, as are now available from CHELSA or ERA5, or remote sensing layers describing land use/land cover. Range estimates with higher temporal resolution are valuable for examining the timing of various threats to species persistence (Trisos et al. 2020).

Our presentation has thus far assumed that the expert-driven filters are a better representation of reality than the input map, though this may not always be the case. For example, an experts' expectation of absence due to dispersal limitation may be informed by biased sampling; in this case, an unmasked SDM that adequately accounts for sampling bias (Phillips et al. 2009, Warton et al. 2013) may better reflect the species' true distribution. As another example, when using remotely sensed layers for masking (e.g., forest cover), extremely fine-scale variation may be undetectable from available satellite resolution, leading to a mask that inappropriately filters out suitable yet small locations. In practice, researchers are unlikely to know *a priori* whether one data source is more reliable than the other; hence, such disagreements should be characterized as uncertainty in range estimates. Perhaps the most conservative option is for downstream analyses (such as range size estimates) to consider all range estimates supported by at least one data source. Alternatively, one could use an ensemble approach, where each candidate range estimate contributes one 'vote' toward determining whether a cell is likely occupied. The number of votes a cell receives can be interpreted as the degree of evidence supporting the range estimate at that location. The *BioModelos* network takes this ensemble approach one step further and allows multiple experts to weigh in, each casting votes that contribute to an ensemble range estimate (Velásquez-Tibatá et al. 2019).

A relevant question is: Why would one choose to use data-driven masking rather than including the presence records in a statistical model with the filter layer as a covariate? We consider a variety of scenarios, although other applications will undoubtedly emerge:

1. **Additional map updates**: If maps originate from a previous study, either expert-drawn or based on an SDM, one could refine them based on additional information beyond what was originally used. Relatedly, it may not be necessary to refit statistical models if an expert already knows that the species is fundamentally restricted to particular features of the study region (e.g. land use/land cover).
2. **Finer resolution filter:** If the filtering layer represents variation at a finer resolution than the precision of the occurrence observations, one could use a filter to obtain a higher-resolution range estimate. For example, presence data could be accurate to within a few hundred meters while habitat type is available at 30 m resolution.
3. **Change over time:** Although there may be sufficient data to build a single, static SDM, sparse sampling over time may limit the ability to infer change in a distribution. In Example 1, this is demonstrated with annually varying forest cover data used to filter a static SDM. Relatedly, temporal mismatch between covariates may preclude using them in a single model. For example, commonly used WorldClim data represents the mean conditions between 1970-2000 (Fick and Hijmans 2017) whereas data derived from MODIS satellites, such as fire products (Table 1), are available only after 2000.
4. **Low number of records with filter data:** Filter layer values may be available only at a subset of the presences used to make the input model. For example, in Example 1 we considered a case where forest-cover filter layers were available only beginning in the year 2001, after which 9 observations of the species were recorded.
5. **Failure to meet assumptions of SDM:** If an expert has good evidence that two species distributions are mutually exclusive (e.g., due to competition), using the distribution of one as a predictor variable for the other does not meet SDM assumptions (it is not a scenopoetic variable, *sensu* (Soberón 2007, Anderson 2017)). Instead, masking one estimated potential distribution by the other may be an effective option as in Example 3 (Kass et al. 2021).

Ongoing uptake of tools from maskRangeR by Colombia's Biodiversity Observation network *BioModelos* (http://biomodelos.humboldt.org.co; see Introduction), exemplifies how this workflow can inform conservation (Velásquez-Tibatá et al. 2019). Importantly, a series of formal end-user consultation workshops with Colombian conservation practitioners and other biodiversity experts informed maskRangeR's development, which was expanded through use cases (see the Vignette, Appendix S1) to be broadly applicable. maskRangeR now provides the capability for real-time model updates by the *BioModelos* community. A next step in the collaboration will be integrating maskRangeR as new modules in the *Wallace* ecological modeling application (Kass et al. 2018), a modular, R-based software for reproducible modeling of species niches and distributions. Adding maskRangeR functionalities to *Wallace* should facilitate its use by conservation practitioners and researchers.

Acknowledgments

All authors acknowledge funding from NASA Grant No. 80NSSC18K0406. CM, JMK, GEPB, MAL, and RPA acknowledge funding from NSF grant DBI-1661510. CM acknowledges funding from NSF Grant DBI-1913673. JMK acknowledges funding from the Japan Society for the Promotion of Science Postdoctoral Fellowships for Foreign Researchers program. CBM acknowledges funding from the Helen Gurley Brown Revocable Trust. BEG was supported by NASA FINESST Award 80NSSC19K133. VGB was supported by the National Association of Business of Colombia (ANDI) scholarship. We thank N. Horning, A. Paz, B.A. Johnson, N. Gazit, and all participants of our software end-user consultation workshops in Bogotá, Colombia for their input on the development of this package, and especially O. Rojas-Soto and L.A. Peña Peniche for their assistance with beta-testing. We also acknowledge workshop support from the Alexander von Humboldt Biological Resources Research Institute and the Center for Biodiversity and Conservation at the American Museum of Natural History.

Data Accessibility Statement

All code and vignette data are available at https://github.com/cmerow/maskRangeR and stable updates to the maskRangeR package will be maintained on CRAN (https://CRAN.R-project.org/package=maskRangeR)

Supplementary Material

The following materials are available as part of the online article at https://escholarship.org/uc/fb

Appendix S1. Vignette of maskRangeR package

References

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J. & Jetz, W. (2018) A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Scientific Data, 5, 180040. https://doi.org/10.1038/sdata.2018.40

Anderson, R.P. (2012) Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. Annals of the New York Academy of Sciences, 1260, 66–80. https://doi.org/10.1111/j.1749-6632.2011.06440.x

Anderson, R.P. (2017) When and how should biotic interations be considered in models of species niches and distributions? Journal of Biogeography, 44, 8–17. https://doi.org/10.1111/jbi.12825

Anderson, R.P., Araújo, M.B., Guisan, A., Lobo, J.M., Martínez-Meyer, E., Peterson, A.T. & Soberón, J.M. (2020) Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. Frontiers of Biogeography, 12. https://doi.org/10.21425/F5FBG47839

Anderson, R.P. & Handley, C.O. (2001) A new species of three-toed sloth (Mammalia: Xenarthra) from Panama, with a review of the genus *Bradypus*. Proceedings-Biological Society of Washington, 114, 1–33.

Anderson, R.P. & Martínez-Meyer, E. (2004) Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. Biological Conservation, 116, 167–179.

Anderson, R.P. & Raza, A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. Journal of Biogeography, 37, 1378–1393. https://doi.org/10.1111/j.1365-2699.2010.02290.x

Anderson, R.P., Peterson, A. T. & Gómez-Laverde, M. (2002) Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos, 98, 3–16.

Araújo, M.B., Anderson, R.P., Barbosa, A. M., et al. (2019) Standards for distribution models in biodiversity assessments. Science Advances, 5, eaat4858. https://doi.org/10.1126/sciadv.aat4858

Araújo, M.B. & Peterson, A.T. (2012) Uses and misuses of bioclimatic envelope modelling. Ecology, 93, 1527–1539. https://doi.org/10.1890/11-1930.1

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J. & Villalobos, F. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modelling, 222, 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Brooks, T.M., Pimm, S.L., Akçakaya, H.R., et al. (2019) Measuring Terrestrial Area of Habitat (AOH) and Its Utility for the IUCN Red List. Trends in Ecology and Evolution, 34, 977–986. https://doi.org/10.1016/j.tree.2019.06.009

Busby, J.R. (1991) BIOCLIM - a bioclimatic analysis and prediction system. In: Nature conservation: cost effective biological surveys and data analysis (ed by C.R. Margules and M.P. Austin), pp. 64-68. CSIRO, Canberra.

Busetto, L. & Ranghetti, L. (2016) MODIStsp: an R package for preprocessing of MODIS Land Products time series. Computers and Geosciences, 97, 40–48.

Calixto-Pérez, E., Alarcón-Guerrero, J., Ramos-Fernández, G., et al. (2018) Integrating expert knowledge and ecological niche models to estimate Mexican primates' distribution. Primates, 59, 451–467.

Chapman, D.S., Haynes, T., Beal, S., Essl, F. & Bullock, J.M. (2014) Phenology predicts the native and invasive range limits of common ragweed. Global Change Biology, 20, 192–202. https://doi.org/10.1111/gcb.12380

Chini, L.P., Hurtt, G.C. & Frolking, S. (2014) Harmonized global land use for years 1500-2100, V1. ORNL DAAC. Available at: https://daac.ornl.gov/VEGETATION/guides/Land_Use_Harmonization_V1.html

Choy, S.L., O'Leary, R. & Mengersen, K. (2009) Elicitation by design in ecology: using expert opinion to inform priors for bayesian statistical models. Ecology, 90, 265–277. https://doi.org/10.1890/07-1886.1

Chua, K.W.J., Ng, D.J.J., Zeng, Y. & Yeo, D.C.J. (2015) Habitat characteristics of tropical rainforest freshwater crabs (Decapoda:

Brachyura: Potamidae, Gecarcinucidae) in Singapore. Journal of Crustacean Biology, 35, 533–539. https://doi.org/10.1163/1937240X-00002346

Daac, O. (2018) MODIS and VIIRS land products global subsetting and visualization tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Available at: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1379

Di Marco, M., Watson, J.E.M., Possingham, H.P. & Venter, O. (2017) Limitations and trade-offs in the use of species distribution maps for protected area planning. Journal of Applied Ecology, 54, 402–411. https://doi.org/10.1111/1365-2664.12771

Dinerstein, E., Olson, D., Joshi, A., et al. (2017) An ecoregion-based approach to protecting half the terrestrial realm. BioScience, 67, 534–545. https://doi.org/10.1093/biosci/bix014

Domisch, S., Amatulli, G. & Jetz, W. (2015) Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Scientific Data, 2, 150073. https://doi.org/10.1038/sdata.2015.73

Drake, John M. (2015) Range bagging: a new method for ecological niche modelling from presence-only data. Journal of the Royal Society Interface, 12, 20150086–89. https://doi.org/10.1098/rsif.2015.0086

Drake, J.M., Randin, C. & Guisan, A. (2006) Modelling ecological niches with support vector machines. Journal of Applied Ecology, 43, 424–432. https://doi.org/10.1111/j.1365-2664.2006.01141.x

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37, 4302–4315. https://doi.org/10.1002/joc.5086

Freeman, B.G. & Mason, N.A. (2015) The geographic distribution of a tropical montane bird is limited by a tree: acorn woodpeckers (*Melanerpes formicivorus*) and colombian oaks (*Quercus humboldtii*) in the Northern Andes. PLoS ONE, 10, e0128675. https://doi.org/10.1371/journal.pone.0128675

Gavrutenko, M., Gerstner, B.E., Kass, J.M., Goodman, S.M. & Anderson, R.P. (2021) Temporal matching of occurrence localities and forest cover data helps improve range estimates and predict climate change vulnerabilities. Global Ecology and Conservation, 27, e01569. https://doi.org/10.1016/j.gecco.2021.e01569

Gerstner, B.E., Kass, J.M., Kays, R., Helgen, K.M. & Anderson, R.P. (2018) Revised distributional estimates for the recently discovered olinguito (*Bassaricyon neblina*), with comments on natural and taxonomic history. Journal of Mammalogy, 99, 321–332. https://doi.org/10.1093/jmammal/gyy012

Giglio, L., Boschetti, L., Roy, D.P., Humber, M.L. & Justice, C.O. (2018) The Collection 6 MODIS burned area mapping algorithm and product. Remote Sensing of the Environment, 217, 72–85. https://doi.org/10.1016/j.rse.2018.08.005

Giglio, L., Schroeder, W. & Justice, C.O. (2016) The collection 6 MODIS active fire detection algorithm and fire products. Remote Sensing of the Environment, 178, 31–41. https://doi.org/10.1016/j.rse.2016.02.054

Guevara, L., Gerstner, B.E., Kass, J.M. & Anderson, R.P. (2018) Toward ecologically realistic predictions of species distributions: a cross-time example from tropical montane cloud forests. Global Change Biology, 24, 1511–1522. https://doi.org/10.1111/gcb.13992

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8, 993–1009. https://doi.org/ 10.1111/J.1461-0248.2005.00792.X

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. Ecological Modelling, 135, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Gutiérrez, E.E., Boria, R.A. & Anderson, R.P. (2014) Can biotic interactions cause allopatry? Niche models, competition, and distributions of South American mouse opossums. Ecography, 37, 741–753. https://doi.org/10.1111/ecog.00620

Hansen, M.C., Potapov, P.V., Moore, R., et al. (2013) High-resolution global maps of 21st-century forest cover change. Science, 342, 850–853. https://doi.org/10.1126/science.1244693

Heberling, J.M., Miller, J.T., Noesgaard, D., Weingart, S.B. & Schigel, D. (2021) Data integration enables global biodiversity synthesis. Proceedings of the National Academy of Sciences USA, 118. https://doi.org/10.1073/pnas.2018093118

Helgen, K.M., Pinto, C.M., Kays, R., Helgen, L.E., Tsuchiya, M.T.N., Quinn, A., Wilson, D.E. & Maldonado, J.E. (2013) Taxonomic revision of the olingos (*Bassaricyon*), with description of a new species, the Olinguito. ZooKeys, 1–83. https://doi.org/10.3897/zookeys.324.5827

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., et al. (2017) SoilGrids250m: global gridded soil information based on machine learning. PLoS ONE, 12, e0169748. https://doi.org/10.1371/journal.pone.0169748

Hijmans, R.J. (2021) raster: Geographic data analysis and modeling.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965–1978. https://doi.org/10.1002/joc.1276

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2010) Dismo: species distribution modeling; 2013. R package version 0. 9-1, 1–87.

Hirsch, A. & Chiarello, A.G. (2012) The endangered maned sloth *Bradypus torquatus* of the Brazilian Atlantic forest: a review and update of geographical distribution and habitat preferences. Mammal Review, 42, 35–54. https://doi.org/10.1111/j.1365-2907.2011.00188.x

Hurlbert, A.H. & Jetz, W. (2007) Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proceedings of the National Academy of Sciences USA, 104, 13384–13389. https://doi.org/10.1073/pnas.0704469104

Hurtt, G.C., Chini, L.P., Frolking, S., et al. (2011) Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Climatic Change, 109, 117–161. https://doi.org/10.1007/s10584-011-0153-2

Jetz, W., McGeoch, M.A., Guralnick, R., et al. (2019) Essential biodiversity variables for mapping and monitoring species populations. Nature Ecology and Evolution. 3, 539-551. https://doi.org/10.1038/s41559-019-0826-1

Jung, M., Arnell, A., De Lamo, X. & Garcia-Rangel, S. (2020) Areas of global importance for terrestrial biodiversity, carbon, and water. Nature Ecology and Evolution. https://doi.org/10.1038/s41559-021-01528-7

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P. & Kessler, M. (2016) CHELSA climatologies at high resolution for the earth's land surface areas (Version 1.1). World Data Center for climate at DKRZ. https://doi.org/10.1594/WDCC/CHELSA_v1_1

Kass, J.M., Meenan, S.I., Tinoco, N., Burneo, S.F. & Anderson, R.P. (2021) Improving area of occupancy estimates for parapatric species using distribution models and support vector machines. Ecological Applications, 31, e02228. https://doi.org/10.1002/eap.2228

Kennedy, C.M., Oakleaf, J.R., Theobald, D.M., Baruch-Mordo, S. & Kiesecker, J. (2019) Managing the middle: a shift in conservation priorities based on the global human modification gradient. Global Change Biology, 25, 811–826. https://doi.org/10.1111/gcb.14549

Khoury, C.K., Amariles, D., Soto, J.S., et al. (2019) Comprehensiveness of conservation of useful wild plants: an operational indicator for biodiversity and sustainable development targets. Ecological Indicators, 98, 420–429. https://doi.org/10.1016/j.ecolind.2018.11.016

Lara, W. & Gutierrez-Velez, V. (2019) forestChange: computing essential biodiversity variables from global forest change data. In: AGU Fall Meeting Abstracts, pp. B23F-2606. Available at: https://ui.adsabs.harvard.edu/abs/2019AGUFM.B23F2606L

Mainali, K., Hefley, T., Ries, L. & Fagan, W.F. (2020) Matching expert range maps with species distribution model predictions. Conservation Biology, 34, 1292–1304. https://doi.org/10.1111/cobi.13492

McBride, M.F., Garnett, S.T., Szabo, J.K., et al. (2012) Structured elicitation of expert judgments for threatened species assessment: a case study on a continental scale using email. Methods in

Ecology and Evolution, 3, 906–920. https://doi.org/10.1111/j.2041-210X.2012.00221.x

Merow, C., Bois, S.T., Allen, J.M., Xie, Y. & Silander, J.A. (2017) Climate change both facilitates and inhibits invasive plant ranges in New England. Proceedings of the National Academy of Sciences USA, 114, E3276–E3284. https://doi.org/10.1073/pnas.1609633114

Merow, C., Allen, J.M., Aiello-Lammens, M. & Silander, J.A., Jr (2016) Improving niche and range estimates with Maxent and point process models by integrating spatially explicit information: Minxent. Global Ecology and Biogeography, 25, 1022–1036. https://doi.org/10.1111/geb.12453

Merow, C., Wilson, A.M. & Jetz, W. (2017) Integrating occurrence data and expert maps for improved species range predictions. Global Ecology and Biogeography, 26, 243–258. https://doi.org/10.1111/geb.12539

Merow, C., Latimer, A.M., Wilson, A.M., McMahon, S.M., Rebelo, A.G. & Silander, J.A. (2014) On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data. Ecography, 37, 1167–1183. https://doi.org/10.1111/ecog.00839

Mitchell, P.J., Monk, J. & Laurenson, L. (2017) Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. Methods in Ecology and Evolution, 8, 12–21. https://doi.org/10.1111/2041-210X.12645

de Moraes-Barros, N., Giorgi, A.P., Silva, S. & Morgante, J.S. (2010) Reevaluation of the Geographical Distribution of *Bradypus tridactylus* Linnaeus, 1758 and *B. variegatus* Schinz, 1825. Edentata, 11, 53–61. https://doi.org/10.1896/020.011.0110

Moudrý, V. & Šímová, P. (2012) Influence of positional accuracy, sample size and scale on modelling species distributions: a review. International Journal of Geographical Information Science, 26, 2083–2095. https://doi.org/10.1080/13658816.2012.721553

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014) Where is positional uncertainty a problem for species distribution modelling? Ecography, 37, 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

Ocampo-Peñuela, N., Jenkins, C.N., Vijay, V., Li, B.V. & Pimm, S.L. (2016) Incorporating explicit geospatial data shows more species at risk of extinction than the current Red List. Science Advances, 2, e1601367. https://doi.org/10.1126/sciadv.1601367

Ocampo-Peñuela, N. & Pimm, S.L. (2014) Setting practical conservation priorities for birds in the Western Andes of Colombia: bird conservation priorities in Colombia. Conservation Biology, 28, 1260–1270. https://doi.org/10.1111/cobi.12312

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., et al. (2001) Terrestrial ecoregions of the world: a new map of life on earth. BioScience, 51, 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A.S. (2016) High-resolution mapping of global surface water and its long-term changes. Nature, 540, 418–422. https://doi.org/10.1038/nature20584

Peterson, A.T., A., Navarro-Sigüenza, A.G. & Gordillo, A. (2018) Assumption-versus data-based approaches to summarizing species' ranges. Conservation Biology, 32, 568–575. https://doi.org/10.1111/cobi.12801

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications, 19, 181–197. https://doi.org/10.1890/07-2153.1

Robinson, N., Regetz, J. & Guralnick, R.P. (2014) EarthEnv-DEM90: a nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 57–67. https://doi.org/10.1016/j.isprsjprs.2013.11.002

Rotenberry, J.T. & Balasubramaniam, P. (2020) Connecting species' geographical distributions to environmental variables: range maps versus observed points of occurrence. Ecography, 43, 897–913. https://doi.org/10.1111/ecog.04871

Skroblin, A., Carboon, T., Bidu, G., Chapman, N., Miller, M., Taylor, K., Taylor, W., Game, E.T. & Wintle, B.A. (2021) Including indigenous knowledge in species distribution modeling for increased ecological insights. Conservation Biology, 35, 587–597. https://doi.org/10.1111/cobi.13373

Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters, 10, 1115–1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Soberón, J.M., Llorente, J.B. & Oñate, L. (2000) The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. Biodiversity and Conservation, 9, 1441–1466. https://doi.org/0.1023/A:1008987010383

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. Ecography, 32, 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Trisos, C.H., Merow, C. & Pigot, A.L. (2020) The projected timing of abrupt ecological disruption from climate change. Nature, 580, 496–501. https://doi.org/10.1038/s41586-020-2189-9

Tuanmu, M.-N. & Jetz, W. (2014) A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. Global Ecology and Biogeography, 23, 1031–1045. https://doi.org/10.1111/geb.12182

Tuanmu, M.-N. & Jetz, W. (2015) A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. Global Ecology and Biogeography, 24, 1329–1339. https://doi.org/10.1111/geb.12365

Urbina-Cardona, N., Blair, M.E., Londoño, M.C., Loyola, R., Velásquez-Tibatá, J. & Morales-Devia, H. (2019) Species Distribution Modeling in Latin America: a 25-Year retrospective review. Tropical Conservation Science, 12, 1940082919854058. https://doi.org/10.1177/1940082919854058

Velásquez-Tibatá, J., Olaya-Rodríguez, M.H., López-Lozano, D., Gutiérrez, C., González, I. & Londoño-Murcia, M.C. (2019) BioModelos: a collaborative online system to map species distributions. PloS ONE, 14, e0214522. https://doi.org/10.1371/journal.pone.0214522

Velazco, S.J.E., Ribeiro, B.R., Laureto, L.M.O. & De Marco Júnior, P. (2020) Overprediction of species distribution models in conservation planning: a still neglected issue with strong effects. Biological Conservation, 252, 108822. https://doi.org/10.1016/j.biocon.2020.108822

Venter, O., Sanderson, E.W., Magrach, A., et al. (2016) Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. Nature Communications, 7. https://doi.org/10.1038/ncomms12558

Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J. & Brotons, L. (2017) Integrating species distribution modelling into decision-making to inform conservation actions. Biodiversity and Conservation, 26, 251–271. https://doi.org/10.1007/s10531-016-1243-2

Warton, D.I., Renner, I.W. & Ramp, D. (2013) Model-based control of observer bias for the analysis of presence-only data in ecology. PloS ONE, 8, e79168–9. https://doi.org/10.1371/journal.pone.0079168

Whittaker, R.J., Araújo, M.B. & Jepson, P. (2005) Conservation biogeography: assessment and prospect. Diversity and Distributions, 11, 3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Wilson, A.M. & Jetz, W. (2016) Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. PLoS Biology, 14, e1002415. https://doi.org/10.1371/journal.pbio.1002415

Wisz, M.S., Pottier, J., Kissling, W.D., et al. (2012) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biological Reviews of the Cambridge Philosophical Society, 88, 15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Wolkovich, E.M., Regetz, J. & O'Connor, M.I. (2012) Advances in global change research require open science by individual researchers. Global Change Biology, 18, 2102–2110. https://doi.org/10.1111/j.1365-2486.2012.02693.x

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. & Veran, S. (2013) Presence-only modelling using MAXENT: when can we trust the inferences? Methods in Ecology and Evolution, 4, 236–243. https://doi.org/10.1111/2041-210x.12004

Submitted: 22 June 2021.
First decision: 20 July 2021.
Accepted: 8 December 2021.

Edited by Janet Franklin.

