SAPIENZA
UNIVERSITÀ DI ROMA

Dottorato in Biologia Ambientale ed Evoluzionistica, XXXIV ciclo

Doctorate in Environmental and Evolutionary Biology, 34th cycle

PhD thesis

Production and validation of updated Area of Habitat maps for terrestrial birds and mammals

Candidate

Prabhat Raj Dahal

Supervisors

Prof. Dr. Carlo Rondinini (Sapienza University of Rome)

Dr. Paul Donald (BirdLife International)

Dr. Stuart Butchart (BirdLife International)

Submission date: 31st January, 2022

Acknowledgment

This PhD was funded by the Inspire4Nature Innovative Training Network, funded by the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement no. 766417. I would like to thank the funding agency.

I would also like to expresses a deep gratitude to my supervisors Prof. Dr. Carlo Rondinini, Dr. Paul Donald and Dr. Stuart Butchart for guiding me throughout my PhD journey. I have learned a lot from them about habitat modeling, IUCN Red List, scientific writing among many other things. I would also like to thank BirdLife International for hosting me in 2019.

I am also thankful to my father Prof. Dr. Mahesh Raj Dahal, my mother Mrs. Shanta Dahal and my sister Pragati Dahal who have always supported my academic endeavors.

Table of Contents

Abstract....1
List of Figures....3
List of Tables....4
Chapter 1: Introduction....5
1.1 References....10
Chapter 2: A global map of terrestrial habitat types....15
2.1 Abstract....15
2.2 Background & Summary....15
2.3 Methods....17
2.4 Data records....20
2.5 Technical validation....22
2.5.1 Approach....22
2.5.2 Results....23
2.6 Usage notes....25
2.6.1 Validation interpretation....25
2.6.2 Known limitations....27
2.6.3 Suggestions to improve the IUCN Habitat Classification Scheme....28
2.7 Code availability....28
2.8 Author contributions....28
2.9 References....29
Chapter 3: A validation standard for Area of Habitat maps for terrestrial birds and mammals....34
3.1 Abstract....34
3.2 Introduction....35
3.3 Methods....36
3.3.1 A modeling approach to identify outliers....39
3.3.2 Point validation of AOH maps of terrestrial birds and mammals....41
3.4 Results....41
3.4.1 Point validation....45
3.4.2 Representativeness of validation sample....47
3.5 Discussion....48
3.6 Author contributions....50
3.7 Code availability....50
3.8 References....50

Chapter 4: A comparison and validation assessment of different Area of Habitat mapping approaches for terrestrial birds and mammals....54
4.1 Abstract....54
4.2 Introduction....55
4.3 Methods....57
4.3.1 Production of AOH maps using global map of terrestrial habitat types....57
4.3.2 Comparison of AOH maps....58
4.4 Results....60
4.5 Discussion....65
4.6 Author contributions....67
4.7 References....67
Chapter 5: Discussion....70
5.1 References....75
Appendices and supporting information....78
6.1 Appendices and supporting information for chapter 2....78
Appendix 2.A1: Summary table with summary rules to map the different level-2 habitat types. 78
6.2 Appendices and supporting information for chapter 3....79
Appendix 3.A1: Metadata table with summary of validation analyses....79
Figure 3.S1: AOH map for species *Zimmerius chicomendesi*....79
Figure 3.S2: AOH map for the species *Icterus graduacauda*....*80*
Figure 3.S3: AOH map for the species *Semnopithecus entellus*....81
Figure 3.S4: Point validation of the AOH maps using model and point prevalence....82
Figure 3.S5: Histogram of mean percentage of suitable AOH pixels inside the 300 m buffer for mammals and birds species used in point validation....82
Figure 3.S6: Comparison of species with and without validation points for mammals....83
Figure 3.S7: Comparison of species with and without validation points for birds....83
6.3 Appendices and supporting information for chapter 4....84
Appendix 4.A1: Metadata table with validation metrics for AOH_T and AOH_L....84
Figure 4.S1: GIS workflow to produce the Area of Habitat maps using global map of terrestrial habitat types....84

Abstract

An accurate representation of the geographical distribution of species is central to ecological research and conservation science and practice. Species' distributions can be represented using a variety of approaches: geographical ranges, which represent the geographical limits of distributions; point locality data, which represent species' known occurrences; or inductive or deductive models, which usually represent species' habitat within geographic ranges. Representations of distributions may contain false presences (commission errors) and/or false absences (omission errors). Recently, Area of Habitat (AOH) maps, a type of deductive model, have gained traction as a tool to represent global distribution of species, reducing the often high rate of commission errors in range maps. AOH models map the distribution of suitable habitat for a species inside its distributional limits. One of the key challenges in producing AOH maps is to translate knowledge of a species' habitat (a complex and species-specific concept) into specific land-cover classes in existing land use/cover layers. Three different methods (expert-based crosswalk, translation table and global maps of terrestrial habitat types) have been developed to date to produce the AOH maps. ('Crosswalk' is a table translating habitat types in a Habitat Classification Scheme to land-cover classes in a land-cover layer.) However, the performance of these methods has not yet been tested. One of the key parts of modeling is validation of the model outputs. This is done by comparing the model output with real world observations, to quantify omission and commission errors in the models. The aim of this thesis is to produce and compare AOH models for terrestrial mammals and birds using different habitat mapping methods, and to assess their relative utility for applications in ecology and conservation. In the second chapter, I developed a map of global terrestrial habitat types based on the IUCN Red List Habitat Classification Scheme, and a novel method to estimate the omission and commission error of the global map of terrestrial habitat types using presence-only data of habitat specialist species downloaded from open repositories like GBIF (Global Biodiversity Information Facility), eBird (www.ebird.com), PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) and the IBA (Important Bird and Biodiversity Areas) dataset. To date, AOH maps have been validated using presence-only data for small subsets of species for different taxonomic groups, but no standard validation method exists for cases where absence data are not available. In Chapter 3, I developed a novel two-step validation protocol for AOH maps which includes first a model-based evaluation of model prevalence (i.e, the proportion of a species' range that contains suitable habitat), and second a validation using species point localities (point prevalence) using presence-only data. I used 48,336,141 point localities for 4,889 bird species and

107,061 point localities for 420 mammal species. Where point prevalence exceeded model prevalence, the AOH was taken to be a better reflection of species' distribution than random. In Chapter 4, I used the global map of terrestrial habitat types to produce AOH maps for 10,651 terrestrial birds and 4,581 terrestrial mammals. I then applied the validation protocol developed in Chapter 3 to AOH maps of terrestrial birds and mammals produced using translation table and global maps of terrestrial habitat types. I found that the average model prevalence for AOH maps produced using the global map of terrestrial habitat type was lower (0.55±0.28 for birds and 0.51±0.29 for mammals) than those produced using the translation table (0.64±27 for birds and 0.65±0.28 for mammals). This led to higher omission errors in the AOH maps produced using the global map of terrestrial habitat types. Also, the number of AOH maps which were better than random was higher in the AOH mapset produced using the translation table. I also found a high congruence between these two sets of maps (53.44% mapped as suitable and 23.22% mapped as unsuitable in both datasets for birds and 58% mapped as suitable and 19% mapped as unsuitable in both datasets for mammals). Each AOH map produced using the global map of terrestrial habitat types was effectively a subset of the equivalent AOH map produced using the translation table, because the former was based on a single map for each habitat type, whereas the latter was based on one-to-many relationships between habitat types and land-cover classes. I conclude that, overall, AOH maps based on the translation table are likely to be of greater utility than AOH maps based on the global map of terrestrial habitat types. However, for species occurring primarily in human-modified habitats, the AOH maps based on the global map of terrestrial habitat types are more accurate. Furthermore, for some purposes, using data on the congruence between the two types of AOH map for each species may be most appropriate. The AOH modeling and validation methods developed in this thesis can help update the AOH maps in the future with latest data on land-cover, habitat and elevation. Furthermore, the validation metrics can be used as a guideline by the users to select the most appropriate AOH map for their use.

List of Figures

Figure 2.1: Sequential order in which habitat types were identified using our decision tree approach..……...18

Figure 2.2: Distribution of IUCN habitat types globally……21

Figure 2.3: Validation results for the habitat map……..25

Figure 3.1: Percentage contribution of elevation range, habitat and both in clipping the IUCN range to produce AOH maps for birds……..43

Figure 3.2: Percentage contribution of elevation range, habitat and both in clipping the IUCN range to AOH for mammals…….44

Figure 3.3: Point prevalence vs model prevalence for terrestrial birds……..46

Figure 3.4: Point prevalence vs model prevalence for terrestrial mammals……..47

Figure 3.5: Taxonomic representativeness of validation sample for birds and mammals……..48

Figure 4.1: Histogram of model prevalence for terrestrial mammals and birds for AOH_T and AOH_L ……60

Figure 4.2: Model prevalence vs point prevalence for birds of AOH_L……...61

Figure 4.3: Model prevalence vs point prevalence for mammals of AOH_L. …….62

Figure 4.4: Difference in mean model and point prevalence for each level-1 habitat class for habitat specialist species with standard errors……63

Figure 4.5: Summary of intersection between AOH_T and AOH_L. …….64

List of Tables

Table 3.1: Mean model and point prevalence for AOH maps with standard deviation of 4,889 bird species across 3 different thresholds……42

Table 3.2: Mean model and point prevalence for AOH maps with standard deviation of 420 mammal species across 3 different thresholds……..42

Table 4.1: Summary of model prevalence and point prevalence for AOH_T and AOH_L for mammal and bird species having point localities…….62

Chapter 1: Introduction

An accurate estimate of the geographical distribution of species is central to ecological research and conservation science and practice (Gregory et al., 2006, Lawler et al., 2011). Such distributions are often represented by vector or raster maps of the species. There is a history of use of distribution maps for conservation of biodiversity (Ferrier, 2002). These include identification of protected areas for conservation planning (Watts et al., 2017), gap analysis (Scott et al., 1993) and assessment of threat to species (Mace et al., 2008) among several other uses in biodiversity conservation.

There are three different classes of information on the distribution of species (Rondinini et al., 2006). These are: 1) point localities (latitude and longitude) where individuals of the species have been recorded; 2) geographical ranges, which represent the geographic limits of the distribution and are typically derived by drawing boundaries around known point localities, and incorporating other sources of information including expert knowledge; and 3) species distribution models, which use environmental and other relevant variables associated with the species to identify suitable habitats where it is likely to occur. Representations of distributions may contain false presences (commission errors) and/or false absences (omission errors), in variable amounts (Rondinini & Boitani, 2012).

Mapping the global distribution of species using point localities (from specimens, surveys, and citizen science), expert knowledge, distribution atlases, and other sources to produce vector polygons showing distributional limits is standard practice. For example, Schipper et al. (2008), BirdLife International and HBW (2020) and Stuart et al. (2004) produced global range maps for mammals, birds and amphibians respectively.

The International Union for Conservation of Nature (IUCN) Red List (www.iucnredlist.org) is the most comprehensive and current database which provides range maps and information on population, habitat, ecology and threats for species from different taxonomic groups. Over 142,000 species have been assessed, but the only large groups with all species assessed and range maps provided are mammals, birds and amphibians (www.iucnredlist.com). The range maps inform assessments of EOO (Extent of Occurrence), which is one of the key parameters used by the IUCN

Red List criteria to assign species to categories of extinction risk (IUCN, 2017). These categories are widely used to set priorities for conservation action.

Despite the extensive utility of the IUCN range maps, they come with some limitations. One of their major limitations is they are drawn to minimize errors of omission, leading to overestimation of presence of species and hence high commission errors (Ficetola et al., 2014; Di Marco et al., 2017). This leads to the inclusion of unsuitable as well as suitable habitat inside the geographic range of the species.

The advent of remote sensing products (Bartholomé & Belward 2005; ESA 2008; ESA 2017; Buchhorn et al., 2020; Karra et al., 2021) and large datasets of point localities for species (eBird (www.ebird.org) and the Global Biodiversity Information Facility (GBIF) (www.gbif.org)) provides the opportunity to greatly improve global distribution maps of species. One of the main tools to harness the power of these big datasets is species distribution models, which is used to estimate the spatial occurrence of species. Species distribution models are of two types (Stoms et al., 1992). The first are deductive models, which use information on species' habitat use to model the suitable areas for the species. The second type are inductive models, in which the environmental conditions at point localities where the species were recorded are interpolated over wider areas. Inductive models require accurate point localities of the species to estimate its distribution based on conditions at these localities. Deductive models on the other hand do not require point localities to estimate the distribution.

One class of deductive model is the Area of Habitat (AOH; also known as Extent of Suitable Habitat, ESH) which maps the distribution of suitable habitat for a species inside its geographical range (Brooks et al., 2019). It aims to reduce commission errors present in the range map without introducing omission errors. Several sets of AOH maps for different taxonomic groups at continental and global scales have already been produced (Rondinini et al., 2005; Rondinini et al., 2006; Catullo et al., 2008; Jenkins and Giri, 2008; Rondinini et al., 2011; Beresford et al. 2011; Jetz et al., 2012; Ficetola et al., 2015; Tracewski et al., 2016).

The AOH maps have several applications, like identification of target areas for sampling rare species (Brooks et al., 2019), identification of Key Biodiversity Areas (IUCN, 2016), gap analysis of protected areas and environmental assessment among others. The AOH maps can also guide IUCN Red List assessment of species as they can be used to estimate the upper limits of Area of

Occupancy (AOO) which is which is one of the metrics applied to the IUCN Red List criteria (Brooks et al., 2019). Furthermore, time series of AOH maps can provide information on habitat change and fragmentation at species level. Moreover, AOH maps can also be used to study the impacts of climate change by projecting the suitable habitat of species under different climate change scenarios.

One of the key challenges in producing AOH maps is to map the habitats of species, which are often a combination of different factors like climate and land use/cover (Lumbierres et al., 2021). Most of the global land-cover maps (For example: ESA 2017; Buchhorn et al., 2020; Karra et al., 2021) have land-cover classes that do not directly represent a particular habitat type, and hence need to be modeled. Three different approaches have been developed to model the different habitat types as per the IUCN Habitat Classification Scheme to produce AOH maps:

1) Using an expert-based 'crosswalk' i.e. matching of habitat types from the IUCN Red List Habitats Classification Scheme to land-cover classes from a land-cover map using expert opinion (for example: Rondinini et al., 2010; Ficetola et al., 2015 and Tracewski et al., 2016). For some habitat types, this is straightforward; for example habitat type "Forest" is also a land-cover type in most land-cover products. However, for other habitat types, there is no direct match to land-cover classes. For example, in the Copernicus land-cover map (CGLS-LC100) (Buchhorn et al., 2020), the habitat type "Savanna" could be associated with land-cover classes like open forest, grasslands, shrubland or herbaceous vegetation. In such cases, expert-based matching of land-cover class and habitat type is subjective.

2) In order to reduce the subjectivity of expert-based crosswalks and to standardize the habitat modeling process to make it data-driven, empirically derived habitat - land-cover translation table was developed by Lumbierres et al. (2021). The translation table associates habitat types from the IUCN habitats classification scheme with different land-cover classes in the Copernicus land-cover map (Buchhorn et al., 2020) and ESA (2017) using a logistic regression model at three different levels of association (called thresholds 1, 2 and 3 hereafter). The logistic model was calibrated using point localities for 6,986 terrestrial vertebrates that also have habitats coded on the IUCN Red List. I used the translation table for the Copernicus land-cover map (Buchhorn et al., 2020) since it has finer resolution of 100 m as compared to 300 m of ESA (2017). The translation table provides habitat – land-cover association for 11 habitat types (those in level-1 of the IUCN Habitats Classification Scheme). The degree of association between the habitat types and land-cover classes

is exclusive at threshold 3 and most inclusive at threshold 1. Therefore, a habitat is usually mapped to more land-cover classes at threshold 1 than at thresholds 2 and 3. Identifying the optimal threshold to use among the three thresholds is also one of the key challenges of AOH mapping using the translation table.

3) A third approach was developed by Jung et al. (2020), who mapped 47 different IUCN habitat types (those at level-2 in the IUCN habitat classification scheme) as a global raster layer using data on climate (Beck et al., 2020) and land-cover (Copernicus land-cover map CGLS-LC100, Buchhorn et al., 2020) along with other ancillary data. This layer can be used directly to extract the suitable habitats of species inside their geographic range to produce AOH maps.

One of the key steps of modeling is validation of the model outputs. This is usually done by comparing the model output with real world observations to quantify the omission and commission errors of the models. Omission errors can be quantified by comparing the model predictions with observation data for presence. Commission errors can be quantified by comparing the model predictions with observation data for absence. By combining the omission and commission errors several validation metrics have been developed such as True Skill Statistics (TSS) (Allouche et al., 2006), the Boyce Index (Boyce et al., 2002) and Area Under the Curve (AUC) (Jiménez-Valverde, 2012) to validate models with presence and absence observation data. But in the case of global data related to habitats and species occurrence, well-sampled presence-absence data sets are scarce – proving that a species is present requires only one positive record, whereas proving that a species is absent is far more difficult. However, there is a large amount of presence-only data (latitude and longitude of the point of occurrence of species) in global repositories like GBIF and eBird which can be freely accessed. With a proper data cleaning protocol, these global data sets can be used to validate AOH maps.

In the second chapter of this thesis, I developed a novel method to estimate the omission and commission errors of the global habitat layer of Jung et al. (2020) using presence-only data of habitat specialist species downloaded from open repositories like GBIF, eBird, PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) (Hudson et al., 2017) and the IBA (Important Bird and Biodiversity Areas) database (Donald et al., 2019). I first identified habitat specialists (i.e. species coded to only one level-1 habitat in the IUCN Red List habitat classification scheme) among terrestrial vertebrates (mammals, birds, amphibians and reptiles) and then acquired point localities for these species from GBIF, eBird and PREDICTS, and

polygons of IBAs supporting the species from the IBA database. The point localities were cleaned using a filtering protocol. Since habitat specialist species occur only in one habitat, the point and polygon localities of habitat specialist species can be translated as point or polygon localities of the habitat itself. Then using these presence-only data of habitats I first computed the omission errors for 31 level-2 habitat types. Then I used the presence data of one habitat type as absence for other habitat types to create an absence only dataset which I used to estimate the commission errors of different habitat types. Using the commission and omission error I computed the balanced accuracy (Kuhn et al., 2020) which is a validation metric ranging from 0 – 1 (a score of 1 means a perfect model) for 31 level-2 habitat types to validate the global map of terrestrial habitat types.

For AOH maps produced for species when absence data are not available, no standard validation method exists. Rondinini et al. (2011) and Ficetola et al. (2015) used presence-only point localities from GBIF to validate AOH maps for mammals and amphibians respectively. AOH maps for South Asian mammals (Catullo et al., 2008) and African vertebrates (Rondinini et al., 2005) were also validated using presence-only point localities. Brooks et al. (2019) recommended using point localities for validation and inclusion of AOH maps for IUCN Red List assessment. However, presence-only point localities are often not available for many species and are biased towards certain taxonomic groups and well-studied areas.

In the third chapter of this thesis, I developed a novel two-step validation protocol for AOH which includes: a) a model-based evaluation of model prevalence (i.e., the proportion of a species' range that comprises AOH), and b) a validation using species point localities (presence-only) data (point prevalence). I demonstrated the use of this approach by validating AOH maps produced by 1) using the translation table and 2) by using the global map of terrestrial habitat types. I also used model prevalence and point prevalence to identify the most appropriate threshold among the three thresholds for AOH maps produced by using the translation table. The validation method developed here is an iterative process whereby systematic errors in the production of AOH (e.g. in the matching of habitat types to land-cover maps) were identified using logistic regression models, then corrected where possible and a new set of AOH maps produced. Then I employed a point validation analysis for the subset of species for which presence-only point localities were available to assess the performance of the AOH maps. Finally, I assessed the extent to which the subset of species for which presence-only point locality data were available were representative of those for which no point data were available.

To date, there has been no comparison of the performance of AOH maps based on three habitat mapping methods discussed above (expert-based crosswalk, translation table and global map of terrestrial habitat types). Lumbierres et al. (2021) compared the accuracy of an expert-based crosswalk and the translation table, and found that both performed equally well. In the fourth chapter of this thesis I compared the performance of AOH maps produced by using the translation table and the global map of terrestrial habitat types using six different validation metrics.

1.1 References

Allouche O, Tsoar A, and Kadmon R (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS), *Journal of Applied Ecology* , 43, 6, 1223 -1232, DOI: 10.1111/j.1365-2664.2006.01214.x

Bartholomé E and Belward A. S (2005). GLC2000: a new approach to global land-cover mapping from Earth observation data, *International Journal of Remote Sensing*, 26, 9, 1959-1977, DOI: 10.1080/01431160412331291297

Beck H. E, Zimmermann N. E, McVicar T. R, et al. (2020). Publisher Correction: Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data* , 7.1, p. 274. ISSN: 2052-4463. DOI:10.1038/s41597-020-00616-w. PMID: 32807783.

BirdLife International and Handbook of the Birds of the World.: Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world. Version 5, url:http://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBWBirdLife_Checklist_v5_Dec20.zip, 2020

Bivand R. (2019). rgrass7: Interface Between GRASS 7 Geographical Information System and R. R package version 0.2-1. https://CRAN.R-project.org/package=rgrass7

Boyce M. S, Vernier P. R, Nielsen S. E, et al. (2002). Evaluating resource selection functions, *Ecological Modeling*, 157, 281-300, DOI: 10.1016/S0304-395 3800(02)00200-4

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 & Evolution*, **34**, (11), 977–986.

Buchhorn M, Smets B, Bertels L, et al. (2020). *Copernicus Global Land Service: land-cover 100m: Version 3 Globe 2015–2019: Product User Manual*. Geneve: Zenodo.

Catullo G, Masi M, Falcucci A, et al. (2008). A gap analysis of Southeast Asian mammals based on habitat suitability models, *Conservation Biology*, 141, 11, 2730-2744, DOI: 10.1016/j.biocon.2008.08.019

Di Marco M, Watson J. E. M, Possingham H. P, et al. (2017). Limitations and trade-offs in the use of species distribution maps for protected area planning. *Journal of Applied Ecology*, **54**, 402–411.

Donald P, Fishpool L, Ajagbe A, et al. (2019). Important Bird and Biodiversity Areas (IBAs): The development and characteristics of a global inventory of key sites for biodiversity. *Bird Conservation International, 29*(2), 177-198. doi:10.1017/S0959270918000102

ESA (2008) GlobCover land-cover v2 2008 database, European Space Agency GlobCover Project, led by MEDIAS-France

ESA (European Space Agency). (2017). *land-cover CCI product user guide version 2. Technical report*. Paris: ESA.

Ferrier S (2002). Mapping Spatial Pattern in Biodiversity for Regional Conservation Planning: Where to from Here?, *Systematic Biology*, 51, 2, 1,331–363.

Ficetola G. F, Rondinini C, Bonardi A, et al. (2014). An evaluation of the robustness of global amphibian range maps. *Journal of Biogeography*, 41, 211–221.

Ficetola G. F, Rondinini C, Bonardi A, et al. (2015). Habitat availability for amphibians and extinction threat: A global analysis. *Diversity and Distributions*, 21, 302–311.

Hudson L. N, Newbold T, Contu S, et al. (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecology and evolution*, 7, 1, 145-188. ISSN: 2045-7758. DOI: 10.1002/ece3.2579. PMID: 28070282.

IUCN. (2012). *Habitats classification scheme (version 3.1)*. *IUCN*, Gland, Switzerland.

IUCN. (2016). *A Global Standard for the Identification of Key Biodiversity Areas*, *Version 1.0. IUCN*, Gland, Switzerland.

Jarvis A, Reuter H. I, Nelson A, et al. (2008). Hole-filled SRTM for the globe version 4. *CGIAR-CSI SRTM 90 M Database*, srtm.csi.cgiar.org.

Jenkins C. N and Giri C (2008). Protection of mammal diversity in Central America, *Conservation Biology*, 22, 4, 1037-44, DOI: 10.1111/j.1523-1739.2008.00974.x

Jiménez-Valverde A (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21, 498–507

Jetz W, McPherson J. M and Guralnick R. P (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology and Evolution*, 27, 151-159. DOI:10.1016/j.tree.2011.09.007

Jung M, Dahal P. R, Butchart S. H. M, et al. (2020). A global map of terrestrial habitat types. *Scientific Data*, 7, 1–8.

Karra K, Kontgis C, Statman-Weil Z, et al. (2021). *Global land use/land-cover with Sentinel-2 and deep learning.* IGARSS 2021-2021 IEEE International Geoscience and Remote Sensing Symposium.

Kuhn, M (2020). Caret: Classification and regression training. (R Project, 2020).

Lawler J. J, Wiersma Y. F and Huettmann F (2011). Using Species Distribution Models for Conservation Planning and Ecological Forecasting. In: *Drew C., Wiersma Y., Huettmann F. (eds)*

Predictive Species and Habitat Modeling in Landscape Ecology. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-7390-0_14

Lumbierres M, Dahal P. R, Di Marco M, Butchart S. H. M, Donald P. F, & Rondinini C (2021). Translating habitat class to land-cover to map area of habitat of terrestrial vertebrates. *Conservation Biology*, 1–11. https://doi.org/10.1111/cobi.13851

Mace G. M, Collar N. J, Gaston K. J, et al. (2008). Quantification of Extinction Risk: IUCN's System for Classifying Threatened Species. *Conservation Biology*, 22, 6, pp 1124-1142

McDermid G. J, Franklin S. E, and LeDrew E. F (2005). Remote sensing for large-area habitat mapping. *Progress in Physical Geography*, 29,4.

Rondinini C, Stuart S, and Boitani, L (2005). Habitat suitability models and the shortfall in conservation planning for African vertebrates. *Conservation Biology*, 19, 1488–1497.

Rondinini C, Wilson K. A, Boitani L, et al. (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology letters*, 9, 10, pp. 1136-45.

Rondinini C, Di Marco M, Chiozza F, et al. (2011). Global habitat suitability models of terrestrial mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 2633–2641.

Rondinini C and Boitani L (2012). Mind the map: trips and pitfalls in making and reading maps of carnivore distributon. *Boitani, L. and Powell, R.A. (eds.) Carnivore ecology and conservation: a handbook of techniques*. Oxford University Press.

Schipper J, Chanson J. S, Chiozza F, et al. (2008). The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science*, 322, 5899, 225-30.

Scott J. M, Davis F, Csuti B, et al. (1993). Gap Analysis: A Geographic Approach to Protection of Biological Diversity. *Wildlife Monographs*, 123, pp 3-38.

Stoms D. M, Davis F. W, and Cogan C. B (1992). Sensitivity of wildlife habitat models to uncertainties in GIS data. *Photogrammetric Engineering and Remote Sensing*, 58, 843- 850.

Stuart S. N, Chanson J. S, Cox N. A, et al. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, 306, 5702, 1783-6.

Tracewski Ł, Butchart S. H, Di Marco M, et al. (2016). Toward quantification of the impact of 21st century deforestation on the extinction risk of terrestrial vertebrates. *Conservation Biology*, *30*(5), pp.1070-1079.

Watts M. E, Stewart R. R, Martin T. G, et al. (2017). Systematic Conservation Planning with Marxan. *Gergel S, Turner M(eds) Learning Landscape Ecology*. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-6374-4

www.ebird.com

www.gbif.com

www.iucnredlist.com

Chapter 2: A global map of terrestrial habitat types

This chapter is based on the publication: Jung M, **Dahal P. R**, Butchart S. H. M. *et al.* A global map of terrestrial habitat types. *Sci Data* **7,** 256 (2020). https://doi.org/10.1038/s41597-020-00599-8

2.1 Abstract

We provide a global, spatially explicit characterization of 47 terrestrial habitat types, as defined in the International Union for Conservation of Nature (IUCN) habitat classification scheme, which is widely used in ecological analyses, including for quantifying species' Area of Habitat. We produced this novel habitat map for the year 2015 by creating a global decision tree that intersects the best currently available global data on land-cover, climate and land use. We independently validated the map using occurrence data for 828 species of vertebrates (35152 point plus 8181 polygonal occurrences) and 6026 sampling sites. Across datasets and mapped classes we found on average a balanced accuracy of 0.77 (∓ 0.14 SD) at level-1 and 0.71 (∓ 0.15 SD) at level-2, while noting potential issues of using occurrence records for validation. The maps broaden our understanding of habitats globally, assist in constructing area of habitat refinements and are relevant for broad-scale ecological studies and future IUCN Red List assessments. Periodic updates are planned as better or more recent data becomes available.

Keywords Species' habitats, IUCN Red List, land-cover, conservation planning, area of habitat

2.2 Background & Summary

Habitat loss is one of the primary causes of biodiversity decline[1–4]. There are many definitions of 'habitat', but they can broadly be described as the entirety of the physical conditions - including land-cover and climate - that enable a species' population to persist in space and time[5]. There is a strong positive relationship between the extent and intactness of a species' habitat and its population persistence[6–8], which may help species extinction risk assessments when information about other symptoms of risk is limited. Knowledge about species' habitats is critical to design landscape management plans[9], conservation planning[10,11] and analysis of past trends and future scenarios of species' extinction risk[12–14].

There are many ways to delimit species' habitats types[15–17], which can be represented as either continuous variables[17,18] or discrete classes[19]. The International Union for Conservation of Nature (IUCN) Red List of Threatened Species uses a global standard typology (https://www.iucnredlist.org/resources/habitat-classification-scheme) that aims to categorize all species-relevant habitats into a system of pre-defined habitat types[16]. In this scheme 16 different broad habitat types are listed at level-1 (e.g. forest, wetlands), with 119 more specific classes listed at level-2 (e.g. Forest – Subtropical/tropical moist lowland). Although detailed descriptions of the habitat types in this classification scheme are unfinished - with the latest available documentation draft dating to December 2012 - it is used by IUCN Red List assessors to describe species' habitats preferences[20].

IUCN Red List assessments also involve compiling distribution maps showing the range boundaries for each species, typically based on point locality data, presence/absence data from atlases, published maps in field guides and monographs, remote sensing data on habitat extent, and expert inference (e.g.[20–22]). Such maps are typically used to estimate Extent of Occurrence (the area of a minimum convex polygon that contains all occurrence records) in Red List assessments, and are also used in aggregate to quantify spatial biodiversity patterns at regional and global scales[23]. However, maps showing distributional boundaries often considerably overestimate the occurrence of a species at finer scales[11,24], a type of error commonly known as commission error. To obviate these types of errors, one approach is to use the habitat preferences and elevational range documented in IUCN Red List assessments to exclude all land-cover classes and altitudes that are not considered suitable for a species in order to map its 'Area of Habitat' (AOH,[20]). This requires a 'crosswalk' that establishes the relationships between each habitat and land-cover class in a particular land-cover product[13,25,26]. However establishing such relationships between different thematic legends can be problematic.

Differences in thematic resolution and definitions can lead to large variations in area-based land-cover estimates[27], and errors have been shown to increase uncertainty and decrease accuracy of any subsequent analysis[28]. These problems are likely to affect AOH estimates as described above, for instance by treating climatically distinct habitats - such as savannah-dominated and subtropical-moist shrub-covered land - as equivalent. Although the potential distribution of a species can be estimated statistically[29,30], it is challenging to do so in a robust, consistent and reproducible

manner[31,32] and in most cases the primary biodiversity data necessary to do so are not available [33]. There is therefore a need to explore alternative approaches to mapping AOH.

Here we describe a method to map the IUCN habitats classification scheme directly for most terrestrial and inland water habitats. We do so by overlaying the best available data on land-cover, climate and other ancillary data sources using simple map algebra. The derived map describes the global distribution of habitats at levels 1 and 2 as outlined by the IUCN classification scheme in the year 2015[16]. We validated the classes from this global map using independent spatially-explicit estimates. To our knowledge this is the first attempt to map IUCN habitat types at a global scale.

2.3 Methods

We delineated terrestrial habitat types following the IUCN classification scheme by intersecting data on land-cover, climate and land use. This intersection was done following a decision tree approach (Figure 2.1), i.e. if the conditions for class 1.9 (Forest – Subtropical/tropical moist montane) were not true for a grid cell then class 1.7 (Forest – Subtropical/tropical mangrove vegetation) was tested. Thus each grid cell of the habitat map is allocated to a single IUCN habitat class. For global land-cover, we used the Copernicus land-cover product[34], which has 23 thematic classes at a ~100 m resolution and an overall average accuracy of ~80%. We used the discrete land-cover classification as well as the Copernicus fractional forest cover estimates available for the year 2015. For climate, we used data on the world's climatic zones based on the global Köppen-Geiger climate classification system[35] at approximately 1 km resolution for the present climate (climatology 1980-2016). We also used the distribution of some terrestrial 'biomes'[36,37] for additional fine adjustment of climatic zones and to create a global mask of the subtropics & tropics.

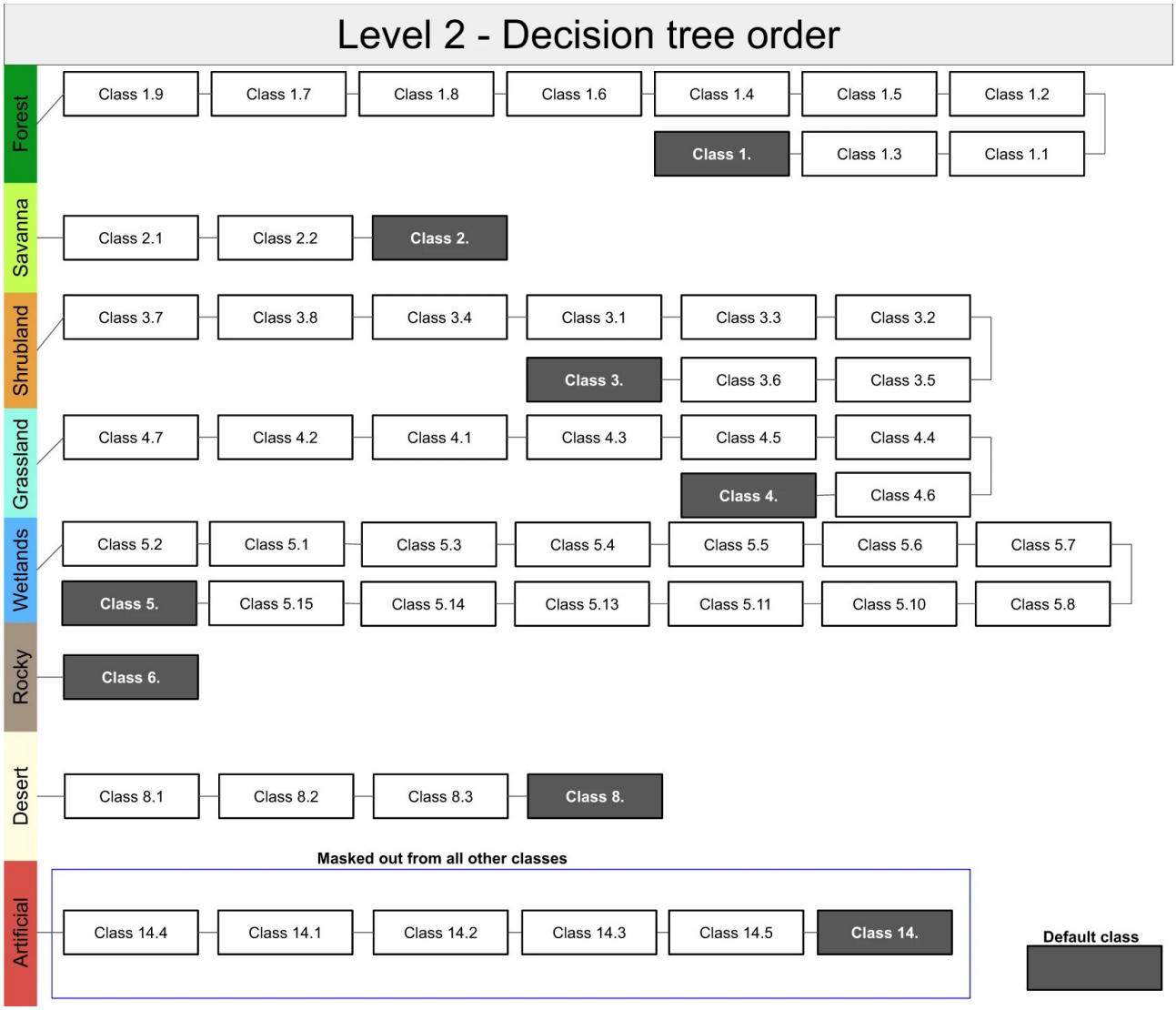

Figure 2.1: Sequential order in which habitat types were identified using our decision tree approach. For instance, if the conditions do not match for IUCN habitat class 1.9, then the conditions for class 1.7 are tested afterwards. Black boxes indicate default classes (level-1 code) in case no conditions could be met at level-2. Artificial habitat types (blue border) are masked out from all other habitat types. Codes and rulesets for each habitat class are further explained in Appendix 2.A1.

In addition, we also considered a number of ancillary data layers for predominantly natural and anthropogenically defined habitat types (Appendix 2.A1). To separate lowland and mountainous habitat types, we used the 'K1' global mountain mask[38], as well as elevation data from the Shuttle Radar Topography Mission (SRTM) mission at ~90 m resolution[39]. For IUCN wetland habitat types - which follow the Ramsar Wetland type classification system[16] - we used the Global Lakes and

Wetlands Database (GLWD) at ~1 km resolution[40], which we expanded with a ~5 km modal filter to account for small-scale differences in water cover (compared to Copernicus). For seasonal and intertidal wetlands and lakes we also considered information from HydroLAKES and other remotely-sensed water surface data[41–43]. To represent tropical and subtropical swamp and mangrove forests we used expert-based estimates for the subtropics and tropics[44].

For terrestrial anthropogenically modified habitat types not already mapped by Copernicus, we relied on existing and novel human pressure datasets. Rural gardens were identified by (1) creating a boundary area of 500 m around urban land-cover classes in the Copernicus data and (2) intersecting arable land-cover within that boundary area with the "very small field size" category according to data on global field size distribution[45]. The 14.3 Plantations class is based on a novel global forest-management layer for the year 2015[46] (available here[47]). To create this layer, separate labelling campaigns were run on sampled forested grid cells (according to Copernicus[34] and Hansen forest cover change dataset[48]) for the tropical, temperate, boreal climate region using the GEO-WIKI platform[49]. Labellers were asked to classify the forest grid cells into several human-dominated forest classes. Finally the global forest management layer was created using a random-forest classifier applied on full PROBA-V time series for the year 2015[46]. We considered all replanted forests (rotation period longer than 20 years), short-rotation woody plantations, agroforestry and fruit plantations as plantation forests.

For pastureland we investigated several existing global pasture datasets for their suitability to serve as a pasture mask[50–52], however we found them either too coarse or outdated, failing to highlight for instance the expansion of pastoral land in Brazil or unable to distinguish between different livestock management systems, for instance grazing in natural grassland versus man-made pastures. For the release (version 002) of the global habitat map we defined 'Pastureland' as grid cells with non-tree covered vegetation with at least 1 head per km² of a grazing livestock-unit (LSU) on land climatically suitable for forest cover, that is, where trees would grow in the absence of grazing. To define the pasture mask we used the latest estimates of all grazing and browsing livestock (buffalo, cattle, goats, horses, sheep) from the gridded livestock density of the world dataset[53] and converted them to LSU using region-specific conversion factors[54]. Originally forest-covered land was defined as those grid cells that are not in a grass, tundra, steppe or meadow defined ecoregion[37] and which are in predominantly tree-covered climatic zones (Tropical, Temperate, Continental) according to the Köppen-geiger climate classification system[35].

All aforementioned datasets were intersected to construct the global habitat map (Figure 2.1) using a decision tree approach (see Appendix 2.A1 for coded rules)**.** This was done in a hierarchical way, by first identifying the IUCN habitat class at level-1 i.e. Forest, Savanna, Shrubland, Grassland, Wetlands (inland), Rocky Areas, Deserts & Artificial habitats (but see 2.A1), followed by the level-2 classifications nested within the respective level-1 class through a decision tree (Figure 2.1). The sequential order is important, with anthropogenically modified habitat types always being mapped first and therefore masking all other 'natural' habitat types. All calculations were implemented in Google Earth Engine (GEE), a cloud-based platform for remote sensing data processing[55]. Whenever the input layers differed in spatial resolution with Copernicus, we resampled those layers by taking the nearest-neighbor. The particular benefits of using GEE are computational speed (taking less than 4h to create and export a new version), clear reproducibility and the ability to update the map easily as new or improved input layers become available. We provide a publicly accessible interface that lets users navigate the map and make all GEE code necessary to reproduce the map available (see code availability).

2.4 Data records

The global habitat map for the year 2015 (version 0032, Figure 2.2) is made interactively available through Google Earth Engine (https://uploads.users.earthengine.app/view/habitat-types-map). As part of this manuscript, the map for level-1 and level-2 habitat types has been made available on a public Zenodo repository at both the Copernicus ~100m resolution and at fractional aggregated 1km resolution[56]. The GEE code to recreate the map is available at (https://github.com/Martin-Jung/Habitatmapping). Asset data used in GEE are publicly readable and directly available from the original sources (see methods). The extent of global planted trees needed to reproduce the map has been made available here[47]. Users are advised to check the data repository for newer versions of both code and map, as we consider this product a "living map" that can be improved in the future pending better data availability. Soon, annual updates to Copernicus up to 2019 will be available[34] and we also plan to create variants relying on the potential distribution of land-cover and biomes[57].

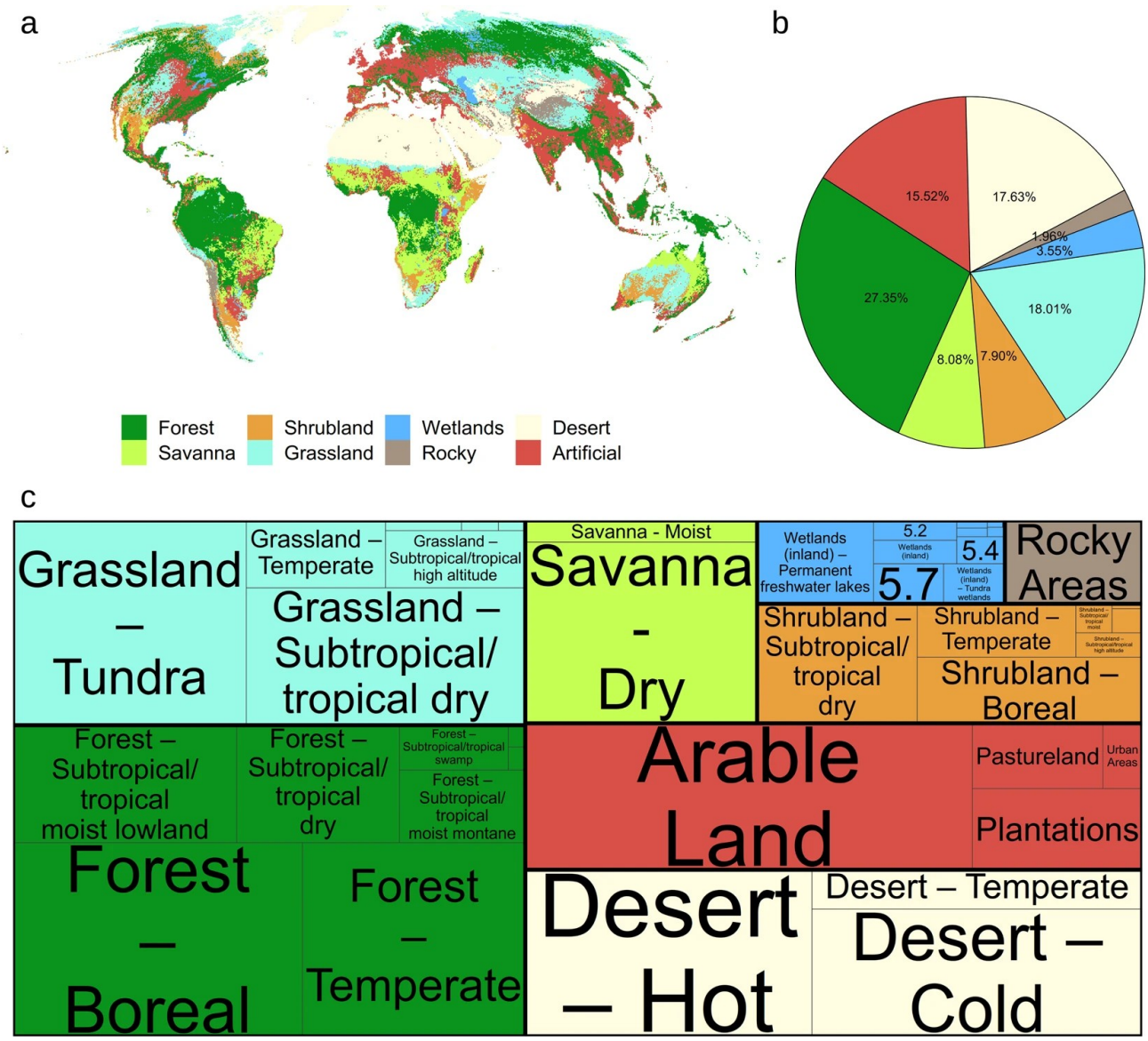

Figure 2.2: Distribution of IUCN habitat types globally (a) Showing the level-1 classification (coarsened to ~5km for this visualization). (b) Proportion of global land area occupied by each level-1 IUCN habitat class. (c) Tree map showing the most dominant IUCN habitat class at level-2 16 nested within the level-1 classes. Colors as in (a) with classes scaled proportional to the land area. level-2 classes with very long names were converted to their id number 16, while small proportions might not be mapped.

2.5 Technical validation

2.5.1 Approach

Since the global habitat map was thematically created to match the IUCN habitats classification system, we mainly relied on existing, independently derived habitat information data to assess its accuracy. We relied on four different data sources for the validation, recognizing that none of them are without spatial bias[58] and that it was not possible to find suitable validation data for all mapped habitat types.

As a first source, we obtained occurrence records of all terrestrial 'habitat specialist' species (those considered to occur only in a single level-2 habitat class according to IUCN Red List assessors) observed during 2005-2019 from the Global Biodiversity Information Facility (GBIF) and eBird (https://ebird.org/). We excluded observations outside the geographical range of a species (as mapped for IUCN Red List assessments), which result largely from misidentifications, vagrants or taxonomic mismatches. Only unique observations with a coordinate uncertainty smaller than 300 m (GBIF) or 30 m (eBird) were retained and we furthermore applied a conservative buffer of 300 m to all observations to account for positional errors. A total of 35152 points were used in this analysis associated with 828 habitat specialist species, 50% of which are birds, 22% reptile, 20% mammals and 8% amphibian species.

Second, we used data from Important Bird and Biodiversity Areas (IBAs,[22,59]) in which habitat specialist birds were known to occur. Specifically, available species checklists were used to identify those IBAs where a given habitat specialist bird species was known to occur, and we checked for the occurrence of that habitat within the IBA. In total, 2142 IBA polygons were used (mean area of 2584 km² with 54% being smaller than 500 km²); however IBA polygons were tested multiple times for different habitats as IBAs can contain more than one habitat. Altogether, a total of 8181 IBA polygons (representing 758 habitat specialist bird species) were tested for the presence of the preferred habitat of species recorded there.

Third, we used species coordinates from the Projecting Responses of Ecological Diversity In Changing Terrestrial Systems (PREDICTS) database[60,61], specifically for artificial habitat types (14) that are usually not found as habitat specialism. Here we selected only those sites that were sampled

after the year 2000, and furthermore we buffered each point by the sampling extent (measured in m). For artificial habitat types in total, we used 1506 validation sites for ‘Arable Land’ (14.1), 1130 for ‘Pastureland’ (14.2), 732 for ‘Plantations’ (14.3) and 429 for ‘Urban Areas’ (14.5).

Fourth, we used the LACO-Wiki platform to visually assess the mapped habitat types at level-2 using publicly available high-resolution satellite imagery[62]. Half the points were placed at random and half were stratified by habitat class, thus ensuring an even spatial and thematic spread globally. People familiar with the IUCN habitat classification system were then asked to label the respective point with a provided level-2 class. NDVI time series from Landsat and the PROBA-IV satellites as well as Flickr™ images taken in the vicinity were provided as guidance. An initial comparison of label agreement between experts reached a 81.5% agreement at level-1 and a 62.5% agreement at level-2. Given that many climatically similar classes at level-2 are very hard or impossible to distinguish visually from satellite imagery, we decided to use this data source only for habitats mapped at level-1 of the IUCN habitat type legend, plus for level-2 deserts, rocky and artificial habitats, which could be most robustly visually identified. In total, 2,229 points were collected as part of this exercise.

We then calculated the match between all observed habitat types (from the three data sources) and the predicted habitat class from the habitat map at ~100 m resolution (the resolution of the Copernicus land-cover data) and at level-1 and level-2. We considered only habitat types for validation for which at least 10 suitable independent validation records were available. For both levels and each dataset we calculated the overall accuracy and the balanced accuracy (to account for an imbalanced number of testing observations) per class and overall using the ‘caret’ package [63].

In addition to the technical validation, we also presented the map to a number of regional experts to ask for their feedback on mapped classes, which helped to fine-tune the ruleset for creating the habitat map.

2.5.2 Results

Across all considered datasets we found an overall accuracy of 0.62 for level-1 and 0.55 for level-2 of the mapped IUCN habitat types. However there was a large disparity among validation datasets and number of classes. For the point records from GBIF/eBird/PREDICTS the overall accuracy at

level-1 was 0.55 (level-2: 0.49), for the IBA data 0.91 (level-2: 0.82), for the artificial habitats from the PREDICTS database 0.79 (level-2: 0.45) and for the visual labeled sites at level-1 0.60 (level-2: 0.65). The average balanced accuracy across validation datasets was 0.76 (∓ 0.12 SD) at level-1 and 0.72 (∓ 0.15 SD) at level-2. We found the greatest balanced accuracy at level-1 for ‘1. Forests’ with 0.88 and the lowest for ‘5. Wetlands’ with 0.65, while the difference in balanced accuracy between datasets was greatest between ‘6. Rocky areas’ and ‘8. Deserts’ (Figure 2.3).

We were able to validate 29 of 48 habitat types mapped at level-2 of the IUCN habitat classification system (Figure 2.3). Across datasets, the largest number of independent validation records was available for ‘1.6. Forest – Subtropical/tropical moist lowland’ (N = 8574) with the lowest being for ‘5.3. Wetlands (inland) – Shrub dominated wetlands’ and ‘4.3 Grassland – Subantarctic’ (both 12). For those habitat types that could be validated at level-2 (Figure 2.3), we found the highest balanced accuracy for ‘4.3 Grassland – Subantarctic’ (0.96), ‘8.3. Desert – Cold’ (0.92), 1.9. Forest – Subtropical/tropical moist montane’ (0.918) and ‘1.4. Forest – Temperate’ (0.88), and the lowest for, ‘5.3. Wetlands (inland) – Shrub dominated wetlands’ and ‘5.4. Wetlands (inland) – Bogs, marshes, swamps, fens, peatlands’ (all 0.5). The balanced accuracy for artificial habitat types was found to be highest for ‘14.5 Urban Areas’ (0.81) and lowest for ‘14.4 Rural Gardens’ (0.54).

Overall, we stress that all of the validation data sources have characteristics that limit their utility for validating a habitat map, and the presented validation results should be interpreted with caution (see Usage notes).

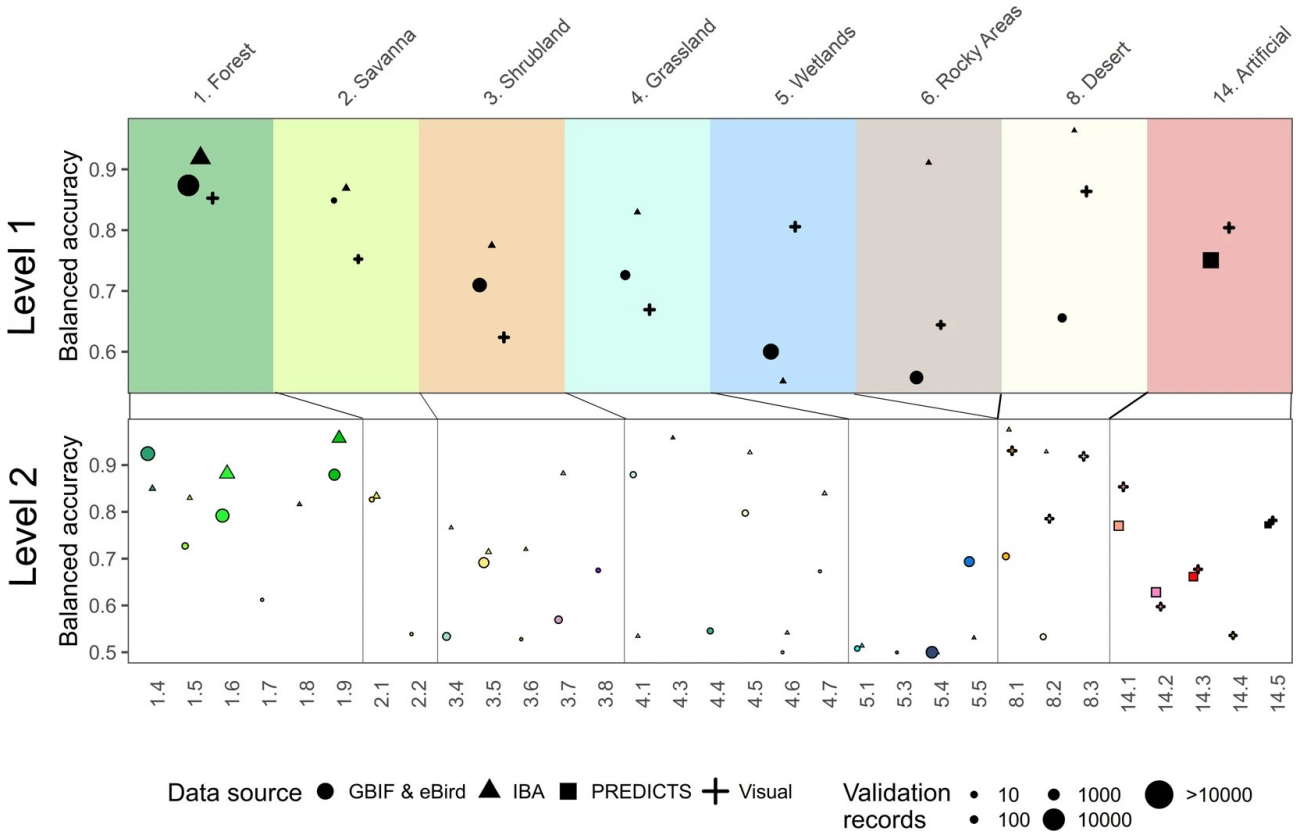

Figure 2.3: Validation results for the habitat map. Estimates of the balanced accuracy are shown for all habitats for which suitable validation data existed. Symbols indicate the validation data source, while point size shows the number of contributing records. The id corresponding to the specific IUCN habitat types is indicated at the bottom and top of the figure. Colors match those of the online interactive interface (https://uploads.users.earthengine.app/view/habitat-types-map).

2.6 Usage notes

2.6.1 Validation interpretation

Independently validating a global habitat map is challenging. In this manuscript we mainly relied on biodiversity observations and sampling sites for validation, recognizing that doing so can be problematic for several reasons: (*a*) These observations can be spatially and taxonomically imprecise. For instance most vertebrates, particularly birds, are highly mobile and non-systematically collected observations (e.g. citizen-science initiatives like eBird) can occur in atypical habitats, for instance if a species is wrongly identified or a migrating bird recorded during

passage. In addition, species occurrences obtained by direct, opportunistic observation tend to be biased towards accessible areas, therefore species tend to be observed at the margins of natural habitats rather than the core, which can result in attributing a record to the wrong habitat type. The fact that we had generally better accuracy for static sampling sites with observations performed by experts (IBAs and PREDICTS sites, Figure 2.3) with larger sampling extent may confirm this assumption; (*b*) For the validation, we used records for those species which had only a single habitat listed as their preference, however it is quite likely that is an incomplete characterization of a species habitat preference. For instance, *Montifringilla nivalis* is said to exclusively occur in '6. Rocky Areas', however within its range the species regularly occurs also in '4.4. Grassland – Temperate' and '14.2 Pastureland'; (*c*) There can be errors in the assigned habitat preferences themselves. For instance, the endemic Japanese macaque (*Macaca fuscata*) is listed to occur exclusively in '1.6. Forest - Subtropical/Tropical Moist Lowland'[64], although most of Japan (where the species is endemic, albeit widespread) is of temperate climate[35,37]. The fact that we were able to programmatically and quickly identify several incorrect habitat preferences in the IUCN Red List database suggests that mapping the IUCN habitat types would help Red List assessors to code species' habitat preferences more accurately swiftly, because it allows them to immediately visualize their mental model of a species' habitats, and correct wrong or missing habitat preferences as well as validating their own assumptions about species ecology; (*d*) All biodiversity observations have obvious geographic and sampling biases, occurring predominantly in temperate regions and more accessible habitats and locations[58]. This is exemplified by the fact that we were not able to validate all mapped IUCN habitat types directly, with boreal habitats missing entirely, while other habitat types such as mangrove forests had very few records (Figure 2.3).

In addition to the biodiversity observations and sampling sites, we also relied on a visual assessment of the habitat types based on satellite imagery, which however also has limitations as a validation data source. Visual labeling of habitats is prone to human errors, depends on - often patchy distributed and outdated - high resolution satellite imagery coverage[65] and is often not easily done for climatically similar classes. Indeed, particularly at level-2 some classes are very hard or impossible to distinguish visually even for experts, such as for instance '1.6. Subtropical/tropical moist lowland forest' from '1.8 Subtropical/tropical swamp forest'.

The habitat map presented is an intersection of multiple existing datasets, each with its own uncertainty in the mapped classes. This uncertainty in the mapped input layers has only been explicitly mapped for land-cover and climate data (Supplementary Figure 2.1), making it challenging to evaluate the influence of input data uncertainty on the mapped habitats[28]. We visually interpreted many of the mismatching species observations used for validation and often found fine-scale differences in land-cover (e.g. '4.4. Grassland – Temperate' to ' 3.4. Shrubland – Temperate') to be the origin.

2.6.2 Known limitations

The documentation of the IUCN Habitat Classification Scheme is unfinished, with ~20% of all class descriptions lacking further elaboration[16]. In this study we aimed to follow the habitat classification system outlined by IUCN[16] to facilitate links with other IUCN data, realizing that other - often more detailed - habitat classification systems exist at national scale[66,67], using land-cover and climate data of higher spatial and thematic resolution[19]. For instance, in an expert-based visual assessment of the habitat map we found that the most common error source were mistakes in the underlying global land-cover data. Based on a precautionary principle and known limitations (see text file on the data repository), we recommend to use the habitat map at a coarsened resolution and supply fractional aggregated maps of each individual class at 1km resolution with every release[56].

Furthermore not all habitat types can be adequately mapped spatially, with some being only seasonally present[41], having intra-annual sequences[68] or being of 'mixed' nature, such as lightly-grazed savanna habitats which can be considered grassland, shrubland or forest depending on the vegetation cover. Other IUCN habitat types are very hard to map spatially, such as '16. Introduced vegetation'. Better spatial information on other anthropogenic classes, such as sown pasture / rangelands, are also necessary to better represent this class in the global habitat map. In addition, four terrestrial IUCN habitat types (four level-2 habitat types) are not represented in the current version of the global habitat map, i.e. all marine habitats (habitat types 9 to 13) as well as artificial aquatic habitats (15). We stress that the habitat map will be updated in the future as new or improved ancillary data become available, which will likely also help to improve many mapped classes.

2.6.3 Suggestions to improve the IUCN Habitat Classification Scheme

In the process of producing the first map of IUCN habitat types, the potential for several improvements to the IUCN habitat classification system became apparent. Firstly, we suggest that additional classes could be added to represent managed forests other than plantations: specifically natural and semi-natural forests that are regularly logged, and recently cleared forests outside the tropics (category 14.6 is limited to heavily degraded or former forest within the subtropics and tropics); and mixed classes of forest / shrubland / grassland, for instance for ‘Temperate open woodland’. For anthropogenic IUCN classes, we suggest that, besides the existing ‘14.2 Pastureland’ class, another class ‘14.7 Rangeland’ could be established, that explicitly relates to anthropogenically grazed natural grasslands in arid regions, like the Kalahari or Western Australia Shrublands[37] and rangelands in the Chaparral. The definition of ‘14.2 Pastureland’ is limited to intensively managed ‘fertilized or re-seeded permanent grasslands, sometimes treated with selective herbicides, with very impoverished flora and fauna’[16] which is an extremely small fraction of all areas that are grazed by livestock. In addition, many existing habitat types without defined descriptions require additional documentation to make it feasible to map them spatially.

2.7 Code availability

All programming code necessary to reproduce the map in Google Earth Engine is supplied together with the data (see Data records) and on https://github.com/Martin-Jung/Habitatmapping.

2.8 Author contributions

Martin Jung conceived the idea of the study, the creation of the habitat map, analysis and writing of the manuscript with support from Piero Visconti. Prabhat Raj Dahal led the technical validation of the habitat map. Paul Donald, Stuart Butchart, Xavier De Lamo, Valarie Kapos & Carlo Rondinini assisted in the technical validation and helped with the writing of the manuscript. Myroslava Lesiv provided unpublished data on forest management and helped with the writing of the manuscript. All authors contributed in interpreting the results and writing of the manuscript.

2.9 References

1. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015).
2. Joppa, L. N. *et al.* Filling in biodiversity threat gaps. *Science* **352**, 416–418 (2016).
3. Maxwell, S. L., Fuller, R. A., Brooks, T. M. & Watson, J. E. M. Biodiversity: The ravages of guns, nets and bulldozers. *Nature* **536**, 143–145 (2016).
4. Díaz, S. *et al.* Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* **366**, eaax3100 (2019).
5. Kearney, M. Habitat, environment and niche: what are we modelling? *Oikos* **115**, 186–191 (2006).
6. Hanski, I. & Ovaskainen, O. The metapopulation capacity of a fragmented landscape. *Nature* **404**, 755–758 (2000).
7. Owens, I. P. F. & Bennett, P. M. Ecological basis of extinction risk in birds: Habitat loss versus human persecution and introduced predators. *Proc. Natl. Acad. Sci.* **97**, 12144–12148 (2000).
8. Brooks, T. M. *et al.* Habitat loss and extinction in the hotspots of biodiversity. *Conserv. Biol.* **16**, 909–923 (2002).
9. Lindenmayer, D. *et al.* A checklist for ecological management of landscapes for conservation. *Ecol. Lett.* **11**, 78–91 (2008).
10. Rodrigues, A. S. L. Improving coarse species distribution data for conservation planning in biodiversity-rich, data-poor, regions: no easy shortcuts. *Anim. Conserv.* **14**, 108–110 (2011).
11. Di Marco, M., Watson, J. E. M., Possingham, H. P. & Venter, O. Limitations and trade-offs in the use of species distribution maps for protected area planning. *J. Appl. Ecol.* **54**, 402–411 (2017).
12. Visconti, P. *et al.* Projecting Global Biodiversity Indicators under Future Development Scenarios. *Conserv. Lett.* **9**, 5–13 (2016).
13. Santini, L. *et al.* Applying habitat and population density models to land cover time series to inform IUCN Red List assessments. *Conserv. Biol.* **00**, cobi.13279 (2019).
14. Powers, R. P. & Jetz, W. Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. *Nat. Clim. Change* (2019) doi:10.1038/s41558-019-0406-z.
15. Fischer, J., Lindenmayer, D. B. & Fazey, I. Appreciating Ecological Complexity: Habitat Contours as a Conceptual Landscape Model. *Conserv. Biol.* **18**, 1245–1253 (2004).
16. IUCN. Habitats Classification Scheme, Version 3.1. 1–14 (2012).

17. Tuanmu, M.-N. & Jetz, W. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Glob. Ecol. Biogeogr.* **24**, 1329–1339 (2015).
18. Radeloff, V. C. *et al.* The Dynamic Habitat Indices (DHIs) from MODIS and global biodiversity. *Remote Sens. Environ.* **222**, 204–214 (2019).
19. Weiss, M. & Banko, G. *Ecosystem Type Map v3.1 – Terrestrial and marine ecosystems*. 1–79 https://www.eionet.europa.eu/etcs/etc-bd/products/etc-bd-reports/ecosystem_mapping_v3_1/@@download/file/Ecosystem_mapping_v3_1.pdf (2018).
20. Brooks, T. M. *et al.* Measuring Terrestrial Area of Habitat (AOH) and Its Utility for the IUCN Red List. *Trends Ecol. Evol.* **34**, 977–986 (2019).
21. IUCN. *IUCN 2016. The IUCN Red List of Threatened Species. Version 2016.1*. (2016).
22. BirdLife International and Handbook of the Birds of the World. Bird species distribution maps of the world. *Bird species distribution maps of the world* http://datazone.birdlife.org/species/requestdis (2019).
23. Pimm, S. L. *et al.* The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344**, 1246752–1246752 (2014).
24. Hurlbert, A. H. & Jetz, W. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proc. Natl. Acad. Sci.* **104**, 13384–13389 (2007).
25. Rondinini, C. *et al.* Global habitat suitability models of terrestrial mammals. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2633–2641 (2011).
26. Ficetola, G. F., Rondinini, C., Bonardi, A., Baisero, D. & Padoa-Schioppa, E. Habitat availability for amphibians and extinction threat: a global analysis. *Divers. Distrib.* **21**, 302–311 (2015).
27. Sexton, J. O. *et al.* Conservation policy and the measurement of forests. *Nat. Clim. Change* **6**, 192–196 (2016).
28. Estes, L. *et al.* A large-area, spatially continuous assessment of land-cover map error and its impact on downstream analyses. *Glob. Change Biol.* **24**, 322–337 (2018).
29. Guisan, A. & Thuiller, W. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009 (2005).
30. Pineda, E. & Lobo, J. M. The performance of range maps and species distribution models representing the geographic variation of species richness at different resolutions. *Glob. Ecol. Biogeogr.* **21**, 935–944 (2012).
31. Araújo, M. B. & Guisan, A. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* **33**, 1677–1688 (2006).

32. Golding, N. *et al.* The zoon r package for reproducible and shareable species distribution modelling. *Methods Ecol. Evol.* **9**, 260–268 (2018).
33. Boitani, L. *et al.* What spatial data do we need to develop global mammal conservation strategies? *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2623–2632 (2011).
34. Buchhorn, M. *et al.* Copernicus Global land-cover Layers—Collection 2. *Remote Sens.* **12**, 1044 (2020).
35. Beck, H. E. *et al.* Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* **5**, 180214 (2018).
36. Olson, D. M. *et al.* Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* **51**, 933 (2001).
37. Dinerstein, E. *et al.* An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience* **67**, 534–545 (2017).
38. Sayre, R. *et al.* A New High-Resolution Map of World Mountains and an Online Tool for Visualizing and Comparing Characterizations of Global Mountain Distributions. *Mt. Res. Dev.* **38**, 240–249 (2018).
39. Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. Hole-filled SRTM for the globe version 4. *Available CGIAR-CSI SRTM 90 M Database* srtm.csi.cgiar.org (2008).
40. Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296**, 1–22 (2004).
41. Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016).
42. Messager, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nat. Commun.* **7**, 13603 (2016).
43. Murray, N. J. *et al.* The global distribution and trajectory of tidal flats. *Nature* **565**, 222–225 (2019).
44. Gumbricht, T. *et al.* An expert system model for mapping tropical wetlands and peatlands reveals South America as the largest contributor. *Glob. Change Biol.* **23**, 3581–3599 (2017).
45. Lesiv, M. *et al.* Estimating the global distribution of field size using crowdsourcing. *Glob. Change Biol.* **25**, 174–186 (2019).
46. Lesiv, M. *et al. Methodology for generating a global forest management layer*. (Zenodo, 2020). doi:10.5281/zenodo.3933966.
47. Lesiv, M. *et al. Global planted trees extent 2015*. (Zenodo, 2020). doi:10.5281/zenodo.3933966.
48. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–3 (2013).

49. Fritz, S. *et al.* Geo-Wiki: An online platform for improving global land-cover. *Environ. Model. Softw.* **31**, 110–123 (2012).
50. Ramankutty, N., Evan, A. T., Monfreda, C. & Foley, J. A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Glob. Biogeochem. Cycles* **22**, (2008).
51. Hoskins, A. J. *et al.* Downscaling land-use data to provide global 30″ estimates of five land-use classes. *Ecol. Evol.* **6**, 3040–3055 (2016).
52. Klein Goldewijk, K., Beusen, A., Doelman, J. & Stehfest, E. New anthropogenic land use estimates for the Holocene; HYDE 3.2. *Earth Syst. Sci. Data Discuss.* 1–40 (2016) doi:10.5194/essd-2016-58.
53. Gilbert, M. *et al.* Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci. Data* **5**, 180227 (2018).
54. Chilonda, P. & Otte, J. Indicators to monitor trends in livestock production at national, regional and international levels. *Livest. Res. Rural Dev.* **18**, (2006).
55. Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
56. Jung, M. *et al.* A global map of terrestrial habitat types. *Zenodo* (2020) doi:10.5281/zenodo.3666245.
57. Hengl, T., Jung, M. & Visconti, P. Potential distribution of land-cover classes (Potential Natural Vegetation) at 250 m spatial resolution (Version v0.1). *Zenodo* (2020) doi:http://doi.org/10.5281/zenodo.3631254.
58. Sorte, F. A. L. & Somveille, M. Survey completeness of a global citizen science database of bird occurrence. *Ecography* **43**, 34–43 (2020).
59. Donald, P. F. *et al.* Important Bird and Biodiversity Areas (IBAs): the development and characteristics of a global inventory of key sites for biodiversity. *Bird Conserv. Int.* **29**, 177–198 (2019).
60. Hudson, L. N. *et al.* The PREDICTS database: a global database of how local terrestrial biodiversity responds to human impacts. *Ecol. Evol.* **4**, 4701–4735 (2014).
61. Hudson, L. N. *et al.* The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecol. Evol.* **7**, 145–188 (2017).
62. See, L. *et al.* LACO-Wiki: A New Online land-cover Validation Tool Demonstrated Using GlobeLand30 for Kenya. *Remote Sens.* (2017) doi:10.3390/rs9070754.
63. Kuhn, M. *et al. Caret: Classification and regression training.* (R Project, 2020).
64. Watanabe, K. & Tokita, K. Macaca fuscata . The IUCN Red List of Threatened Species 2008: e.T12552A3355997. *Macaca fuscata* https://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T12552A3355997.en (2008).

65. Lesiv, M. *et al.* Characterizing the Spatial and Temporal Availability of Very High Resolution Satellite Imagery for Monitoring Applications. *Earth Syst. Sci. Data Discuss.* 1–24 (2018) doi:10.5194/essd-2018-13.
66. Bunce, R. G. H. *et al.* A standardized procedure for surveillance and monitoring European habitats and provision of spatial data. *Landsc. Ecol.* **23**, 11–25 (2008).
67. Janssen, J. *et al.* European Red list of Habitats. Part 2. Terrestrial and freshwater habitats. (2016) doi:10.2779/091372.
68. Kleyer, M. *et al.* Mosaic cycles in agricultural landscapes of Northwest Europe. *Basic Appl. Ecol.* **8**, 295–309 (2007).

Chapter 3: A validation standard for Area of Habitat maps for terrestrial birds and mammals

This chapter is based on the publication: **Dahal, P. R.**, Lumbierres, M., Butchart, S. H. M., Donald, P. F., and Rondinini, C.: A validation standard for Area of Habitat maps for terrestrial birds and mammals, Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2021-245, in review, 2021.

3.1 Abstract

Area of Habitat (AOH) is a deductive model which maps the distribution of suitable habitat at suitable altitudes for a species inside its broad geographical range. AOH maps have been validated using presence-only data for small subsets of species for different taxonomic groups, but no standard validation method exists when absence data are not available. We develop a novel two-step validation protocol for AOH which includes first a model-based evaluation of model prevalence (i.e, the proportion of suitable habitat within a species' range), and second a validation using species point localities (presence-only) data. We applied the protocol to AOH maps of terrestrial birds and mammals. In the first step we built logistic regression models to predict expected model prevalence (the proportion of the range retained as AOH) as a function of each species' elevation range, mid-point of elevation range, number of habitats, realm and, for birds, seasonality. AOH maps with large difference between observed and predicted model prevalence were identified as outliers and used to identify a number of sources of systematic error which were then corrected when possible. For the corrected AOH, only 1.7% of AOH maps for birds and 2.3% of AOH maps for mammals were flagged as outliers in terms of the difference between their observed and predicted model prevalence. In the second step we calculated point prevalence, the proportion of point localities of a species falling in pixels coded as suitable in the AOH map. We used 48,336,141 point localities for 4,889 bird species and 107,061 point localities for 420 mammals. Where point prevalence exceeded model prevalence, the AOH was a better reflection of species' distribution than random. We also found that 4,689 out of 4,889 (95.9%) AOH maps for birds, and 399 out of 420 (95.0%) AOH maps for mammals were better than random. Possible reasons for the poor performance of a small proportion of AOH maps are discussed.

3.2 Introduction

An accurate estimate of the distribution of species is central to ecological and conservation research and action. There are three different classes of information on the distribution of species (Rondinini and Boitani, 2006). These are 1) point localities (latitude and longitude) of individuals; 2) geographic ranges, which are derived by mapping the extent of known point localities along with expert knowledge; and 3) species distribution models, which use environmental and other relevant variables associated with the species to refine geographical ranges. Species distribution models are of two types (Stoms et al., 1992). The first are deductive models, which use expert-based information on species' habitat use to model the suitable areas for the species. The second type are inductive models, in which the environmental conditions at point localities where the species were recorded are interpolated over wider areas.

Area of Habitat (AOH; also known as Extent of Suitable Habitat, ESH) is a deductive model which maps the distribution of suitable habitat for a species inside its broad geographical range (Brooks et al., 2019). It aims to reduce commission errors present in the range map while minimizing omission errors. Several sets of AOH maps for different taxonomic groups at continental and global scales have already been produced (Rondinini et al., 2005; Rondinini et al., 2006; Catullo et al., 2008; Jenkins and Giri, 2008; Rondinini et al., 2011; Ficetola et al., 2015; Tracewski et al., 2016; Lumbierres et al., 2021b).

Habitat models are prone to two major types of errors: omission errors occur when suitable habitat areas for the species are wrongly mapped as being unsuitable, commission errors occur when areas unsuitable for the species are wrongly mapped as being suitable. Quantification of these errors is one of the key parts of the habitat modeling process and is done by validation. The omission and commission errors could both be quantified only when independent presence and absence data on the species are available. In such cases standard validation metrics such as True Skill Statistics (TSS) (Allouche et al., 2006) and the Boyce Index (Boyce et al., 2002) are used. In case of AOH maps produced for large taxonomic groups when true absence data are not available, no standard validation method exists.

Rondinini et al. (2011) and Ficetola et al. (2015) used point localities from GBIF (Global Biodiversity Information Facility) (www.gbif.org) to validate AOH maps for mammals and

amphibians respectively. AOH maps for South Asian mammals (Catullo et al., 2008) and African vertebrates (Rondinini et al., 2005) were also validated using point localities. Brooks et al. (2019) recommend using point localities for validation and inclusion of AOH maps for IUCN (International Union for Conservation of Nature) Red List assessment. However, point localities are often not available for many species and are biased towards certain taxonomic group and well-studied areas.

In this paper, we developed a novel two-step validation protocol for AOH which includes: a) a model-based evaluation of model prevalence (i.e., the proportion of a species' range that comprises AOH), and b) a validation using species point localities (presence-only) data. We demonstrate the use of this approach by validating a new set of AOH maps produced by Lumbierres et al. (2021b) for all terrestrial birds and mammals. The validation method developed here is an iterative process whereby systematic errors in the production of AOH (e.g. in the matching of habitat types to land-cover maps) were identified using logistic regression models, then corrected where possible and a new set of AOH maps produced. Then we employed a point validation analysis for the subset of species for which point localities were available to assess the performance of the AOH maps. Finally, we assessed the extent to which the subset of species for which point locality data were available were representative of those for which no point data were available.

3.3 Methods

The new set of AOH maps (Lumbierres et al., 2021b) was produced at a resolution of 100 m using a novel habitat-land-cover model (Lumbierres et al., 2021a) which associated the different land-cover classes in the Copernicus land-cover map (Buchhorn et al., 2019) with the level-1 habitat types of the IUCN Habitat Classification Scheme (IUCN, 2012). The IUCN Habitat Classification Scheme is a hierarchy of habitat types, and each species assessed in the IUCN Red List is assigned to one or more of these habitat types, based on available information in the literature, unpublished reports and expert knowledge. The habitat-land-cover model (Lumbierres et al., 2021a) has the provision of associating IUCN habitat types to land-cover classes using three different thresholds (1, 2 and 3). Lower thresholds permit weaker associations between land-cover and habitat types. Therefore, with threshold 1 each land-cover class is associated with more habitat types than with threshold 3. Lumbierres et al. (2021b) produced a set of AOH maps for each of the three different thresholds by

clipping out of each species' range any cells of land-cover that were not linked by the model to the habitat class(es) to which the species was coded, then further clipping out parts of the range falling outside the elevation range of the species.

In order to identify the best threshold among the three thresholds and to validate the set of AOH maps with the best threshold at species level, we quantified two measures: 'model prevalence' and 'point prevalence'. Model prevalence is defined as the proportion of pixels inside the range that were retained in the AOH. For example, if 25% of the pixels present in the original range map are clipped out because they contain unsuitable habitat, fall outside the species' elevation range or both, the model prevalence is 0.75. Point prevalence is defined as the proportion of point localities (or their buffers) out of all points inside the range of a species falling inside the suitable pixels. For example the Red-tailed Comet (*Sappho sparganurus*) had a total of 71 point localities within its range, of which 62 fell in pixels coded as suitable in the species' AOH map, giving a point prevalence of 62/71 = 0.88.

Because the number of habitats associated with each land-cover class decreases with increasing thresholds, model prevalence is highest for threshold 1 models and lowest for threshold 3 models. With increasing threshold, commission errors are expected to decrease (which is the main purpose of AOH) but omission errors might increase. Our validation protocol therefore aimed to control for omission errors. We did this by calculating point prevalence and model prevalence across the three thresholds and identified the set of AOH maps for which the mean model prevalence was lowest without compromising the mean point prevalence.

The point localities for bird species were downloaded from eBird (www.ebird.org), the largest global repository for data on point localities of birds. eBird provides a metadata file called "eBird basic data set" (Cornell Lab of Ornithology, 2020) which is a compilation of all the validated point localities at species level and is updated monthly. These point localities are submitted by citizen scientists as well as experts worldwide and are checked by local experts to remove obvious misidentifications before they are made available for download (Sullivan et al., 2009). We first downloaded the metadata file from eBird updated in January 2020 which was then queried in R (R Core Team, 2018) using the *auk* package (Strimas-Mackey et al., 2018), as recommended by eBird, to extract the point localities at species level. The taxonomy of Birdlife International (BirdLife International and Handbook of the Birds of the World, 2020), which is that followed by the IUCN, was matched with eBird's taxonomy and point localities of only those species common to both were

queried and extracted from the metadata. Of the 10,813 species listed in Birdlife International's list for which AOH maps were produced, 9,628 species matched by name. Of these 9,628 species, 8,998 species shared the same taxonomic concept and for 730 species the scientific names matched but the taxonomic concept did not.

To ensure that only high-accuracy points were used for the validation, we selected the stationary points from eBird's metadata. The stationary points are those that have coordinate uncertainty of less than 30 m. We then applied a temporal filter of 2019-2020 because the point localities from 2005-2018 were used to calibrate the habitat-land-cover model in Lumbierres et al. (2021a). This ensured there was no overlap between the calibration and validation data. The points were further filtered by the range polygon of the species provided by the IUCN Red List website (IUCN, 2020) to remove the small number of points falling outside the range (many of them likely to be misidentifications). Since the AOH maps in question only include a certain combination of presence, origin and seasonality of the range, we used the same combination to filter the point localities. This ensured that we only included points which fell inside the boundaries of the selected range maps. We also made sure that only one point locality was allowed per pixel of the AOH map to avoid clustering of points. Finally, we excluded species which had fewer than 10 point localities after all the filters were applied. A total of 4,889 bird species had 4,836,141 point localities after filtering. For mammals, point localities were downloaded from GBIF (Cold Spring Harbor Laboratory, 2021) following the taxonomy of Global Mammal Assessment (which is followed by IUCN) with same temporal and spatial filters as with birds except the filter of coordinate uncertainty which was set to 300 m for mammals. This was done because far too many mammal species would be excluded in the validation if we only considered point localities with coordinate uncertainty of less than 30 m. The *rgbif* package (Chamberlain et al., 2021) in R was used to download the points for mammals. A total of 107,061 point localities for 420 species were available for mammals after applying all the filters.

A buffer of 300 m was applied around all the point localities to account for the positional uncertainty of the points and for the fact that the location usually records that of the observer at the time of observation and not the focal animal, following Jung et al. (2020). The buffers of point localities were then overlaid on top of the AOH maps across all three thresholds at species level and if at least one pixel coded to suitable habitat was found inside the buffer, the pixel was considered to be validated at that point locality. The count of validated pixels was used to calculate point prevalence at species level across all three thresholds.

We identified the threshold that produced a set of AOH maps for which the mean model prevalence was lowest without detriment to the mean point prevalence. We then employed a two-step approach to validate the set of AOH maps with the optimal threshold. In the first step, we identified potential systematic errors in the AOH maps using a modeling approach that aimed to identify species whose model prevalence was larger or smaller than expected, given the characteristics of the species concerned. In the second step, we validated the AOH maps using point localities following Rondinini et al. (2011).

3.3.1 A modeling approach to identify outliers

We used logistic generalized linear models to predict model prevalence of the set of AOH maps produced using the optimal threshold as a function of a number of independent variables, and identified outliers whose observed model prevalence was significantly higher or lower than predicted by the model. Outliers were then examined to identify systematic errors in, for example, the way habitats were coded to land-cover classes in the production of the AOH maps, and to identify species that might be coded to the wrong habitats or elevation limits. For example, if a species' range includes a high proportion of a particular land-cover type not associated with the suitable habitats of the species in the land-cover-habitat association table (Lumbierres et al., 2021b), or if errors in coding species to elevation limits mean that most of the range is outside the species' stated limits, the model prevalence would be lower than predicted by the model.

The predictors fitted to the logistic models included: elevation range of the species (upper elevation limit minus lower elevation limit), mid-point of the elevation range, number of habitats to which the species is coded against in the IUCN Red List, seasonality of species (breeding and non-breeding ranges in case of migratory birds) and the geographical realm of the species. In case of migratory birds, Lumbierres et al. (2021b) has three different classes (resident, breeding and non- breeding seasonalities) of AOH maps based on seasonality of the species. We merged resident seasonality to breeding and non breeding seasonalities to have AOH maps with only two seasonalities (breeding and non-breeding). The dependent variable was the model prevalence of the AOH maps. Data from a total of 10,475 AOH maps for 9,163 bird species (including for some species with separate breeding and non-breeding ranges) and 2,758 AOH maps for 2,758 mammal species were used to build logistic regression models for birds and mammals separately using the *lme4* (Bates et al., 2015) package in R . Data on elevation were lacking for many mammal and bird species which is

the reason why not all species could be included in the logistic model. After testing taxonomic genus, family and order as random effects in the model to control the non-independence of closely related taxa, family was selected for fitting as the residual variance was lowest for the models with family as the random effect for both birds and mammals. The predictive power of the model was assessed by calculating marginal R^2 and conditional R^2 using the *insight* (Lüdecke et al., 2019) package in R. The marginal R^2 expresses how much of the variation in data is explained by the fixed effects and conditional R^2 tells how much of the variation in data is explained by both fixed and random effects.

The Tukey fences outlier detection test (Wilcox, 2017) was used to identify outliers based on the difference between the estimated and observed values of model prevalence. This test uses the interquartile ranges to estimate the outliers in a data-set. The outlier test identified mild lower and upper threshold values for the difference between estimated and observed values.

*Mild upper threshold = (interquartile range * 1.5) + upper quartile*

*Mild lower threshold = lower quartile - (interquartile range * 1.5)*

The AOH maps identified as mild upper outliers have an observed model prevalence much larger than their predicted model prevalence, whereas maps identified as mild lower outliers have an observed model prevalence much smaller than their predicted model prevalence.

In order to investigate the sources of errors in the outliers, we produced two more sets of AOH maps for the outliers. One set included AOH maps which were produced by clipping the range of the species by the altitudinal range only ($AOH_{Elevation\ only}$). Similarly, the other set included AOH maps which were derived by clipping the range with only suitable habitat of the species ($AOH_{Habitat\ only}$). If the model prevalence of an outlier was equal or nearly equal to the model prevalence of its $AOH_{Elevation\ only}$, then we concluded that the under-representation of model prevalence could be attributed to errors in elevation range of the species. If the model prevalence of an outlier was equal or nearly equal to the model prevalence of $AOH_{Habitat\ only}$, then the source of error could be attributed to the mapping of the habitats inside the range using the habitat-land-cover crosswalk (Lumbierres et al., 2021a) or to errors in the species' habitat coding. Furthermore, in some of the outliers the under-representation could result from inclusion of large proportion of habitats which were unsuitable for the species but were inside the range map of the species. Outliers do not necessarily

represent errors in AOH, as species might legitimately have very high or low model prevalence, but by identifying suites of outliers sharing common characteristics we were able to identify and correct a number of systematic errors in AOH production. The models also allowed us to identify species whose AOH maps might be unreliable and whose habitat and elevation coding needs to be checked.

3.3.2 Point validation of AOH maps of terrestrial birds and mammals

We validated 4,889 bird and 420 mammal species' AOH maps using the filtered point localities. The point validation was done by comparing the model and point prevalence at species level. If the point prevalence exceeded model prevalence at species level, the AOH maps performed better than random, otherwise they were no better than random. We also calculated the percentage of suitable habitat pixels inside the buffers to ensure that the validation success wasn't due to a one off pixel falling inside the 300 m buffer.

One of the major issues with citizen science data is that there is often a non-representative spread of data across species. It is therefore possible that the species included in the point validation analysis are not representative, in terms of the ratio between point prevalence and model prevalence, of the species not included. We assessed how representative the validation sample size was by comparing the representation of variables such as family, order, genus, realm, elevation range, mid-point of the elevation range, range size and extinction risk categories for birds and mammals between species with and without point data. The point validation was done in R and GRASS (GRASS Development Team, 2017).

3.4 Results

After comparing point and model prevalence of 4,889 birds and 420 mammal species across all the three thresholds, we selected the set of AOH maps derived by using threshold 3 in the habitat-land-cover model. At threshold 3, the mean model prevalence decreased as compared to thresholds 1 and 2 with much lower change in the mean point prevalence (Table 3.1 and 3.2) for both birds and mammals.

	Threshold 1	Threshold 2	Threshold 3
Mean model prevalence	0.81 ± 0.21 SD	0.77 ± 0.23 SD	0.65 ± 0.25 SD
Mean point prevalence	0.95 ± 0.14 SD	0.94 ± 0.14 SD	0.90 ± 0.17 SD

Table 3.1: Mean model and point prevalence for AOH maps with standard deviation of 4,889 bird species across 3 different thresholds.

	Threshold 1	Threshold 2	Threshold 3
Mean model prevalence	0.87 ± 0.21 SD	0.83 ± 0.22 SD	0.73 ± 0.24 SD
Mean point prevalence	0.95 ± 0.14 SD	0.95 ± 0.15 SD	0.93 ± 0.17 SD

Table 3.2: Mean model and point prevalence for AOH maps with standard deviation of 420 mammal species across 3 different thresholds.

We also assessed the relative contribution of elevation range, habitat, and both in reducing the range to AOH. For both birds and mammals, most of the pixels removed from the range were because either the habitat or the elevation were unsuitable, with a relatively small proportion being removed because both were unsuitable (Figures 3.1, 3.2). The proportion of the range that was clipped out on the basis of having unsuitable habitat at suitable elevations increased as model prevalence decreased, whereas there was little change across the same axis in the proportion of the range that was excluded on the basis of having suitable habitat at unsuitable elevations (Figures 3.1, 3.2). The number of both bird and mammal species peaked at model prevalence of 95-100% and gradually decreased as the model prevalence decreased.

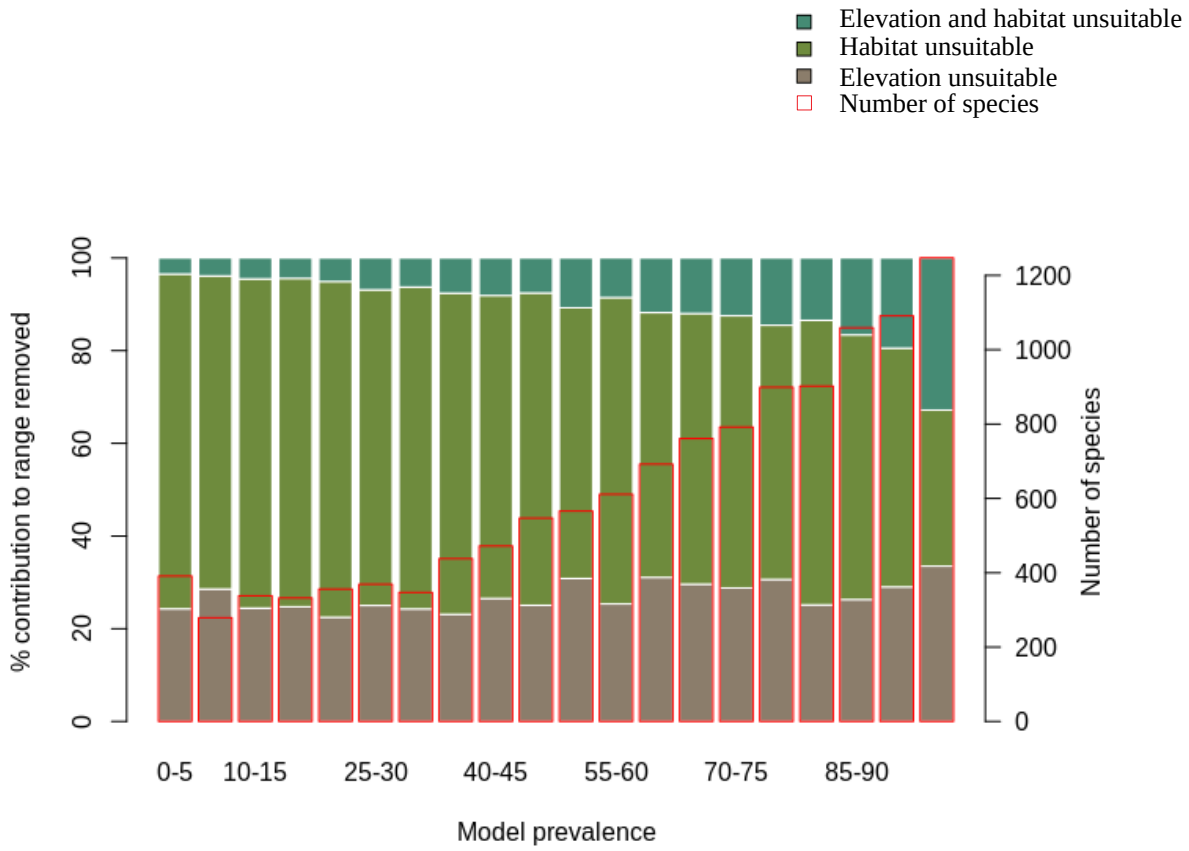

Figure 3.1: Percentage contribution of elevation range, habitat and both in clipping the IUCN range to produce AOH maps for birds. Each bar represents a 5% bin of model prevalence, divided to show how much of the range was clipped out due to unsuitable habitat at suitable elevations ("Habitat unsuitable"), by suitable habitat at unsuitable elevations ("Elevation unsuitable") and by unsuitable habitat at unsuitable elevations ("Elevation and habitat unsuitable"). The red blocks correspond to the second *y*-axis and show the number of species falling into each 5 % bin of model prevalence.

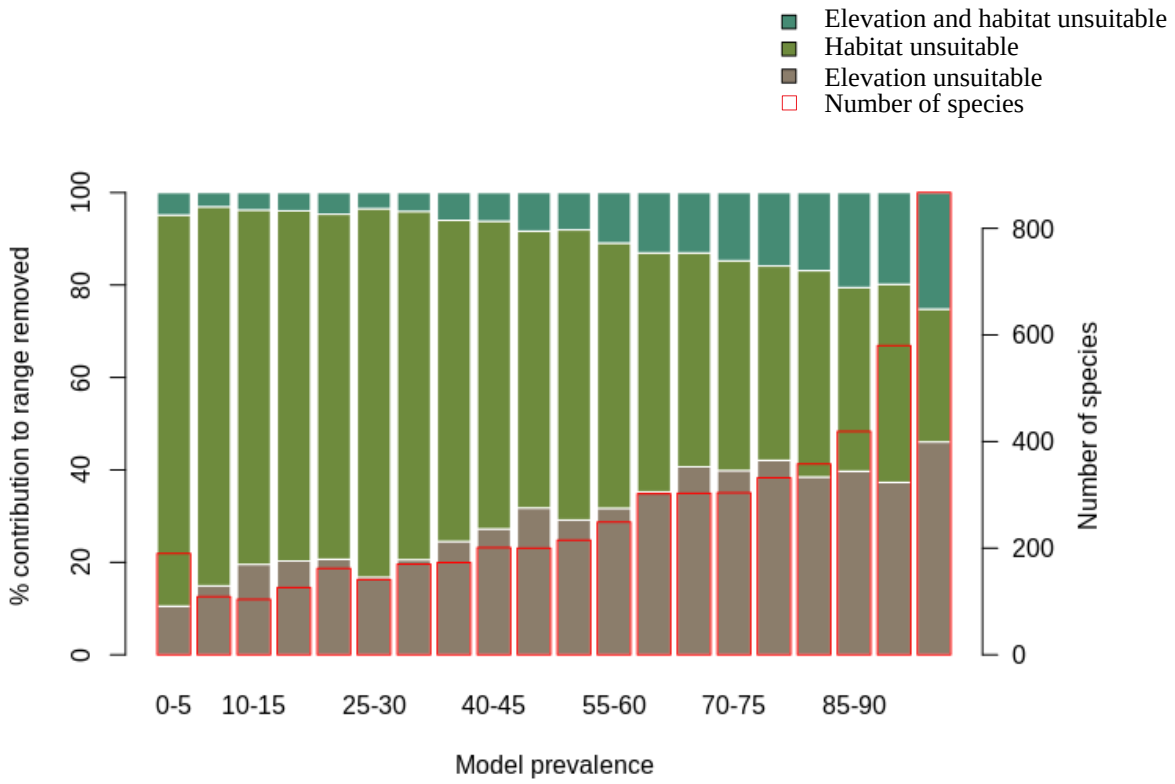

Figure 3.2: Percentage contribution of elevation range, habitat and both in clipping the IUCN range to AOH for mammals. See caption to Figure 3.1 for interpretation.

For birds, the logistic model identified 178 AOH maps (1.7%) as lower outliers and 118 AOH maps (1.1%) as upper outliers out of 10,475 AOH maps for 9,163 terrestrial bird species. Similarly for mammals, the logistic model was applied to the AOH maps of 2,758 species and identified 64 (2.3%) as lower outliers and 21 (0.8%) as upper outliers.

The mean of mid-point of elevation of the bird and mammal species identified as upper outliers was 2,725 m and 3,193 m respectively while the mid-point of elevation for species which were not identified as upper outliers was 1,261 m for birds and 1,289 m for mammals. This suggests that species identified as upper outliers were those found in higher elevation. These species were identified as upper outliers because the logistic models predicted low model prevalence at higher elevations. Also, the range maps for high-altitude species are drawn using contour maps, therefore most of the range is within the correct attitudinal band leading to high model prevalence for these species.

The lower outliers indicate where model prevalence was possibly underestimated due to potential

errors in habitat mapping/coding and elevation range of the species. We found that the habitats “Shrubland” and “Savannah” in the habitat-land-cover crosswalk were not associated with the land-cover class “Herbaceous cover”, leading to under-representation of these habitat types and hence lower model prevalence than estimated by the logistic model (Figure 3.S1). We also found mismatch in the elevation range and geographical range for the lower outliers (Figure 3.S2). There were few cases where the range included large proportion of a particular land-cover type which was not associated with the suitable habitat of the species (Figure 3.S3). Moreover, we found that there was no land-cover information in the Copernicus land-cover map for very small range polygons located on oceanic islands which caused the AOH maps for these species to be empty. Furthermore, the land-cover class “open forest unknown” was discarded in the habitat land-cover model. This led to low model prevalence of AOH maps for some species whose ranges included this land-cover. This was corrected and a new set of AOH maps produced.

3.4.1 Point validation

Out of 4,889 bird species (45% of all bird species) for which point data were available, 4,689 (95.9%) had higher point prevalence than model prevalence and 200 species had lower point prevalence than model prevalence (Figure 3.3). The mean percentage of pixels coded as suitable inside the 300 m buffers of point localities of 4,889 species of birds was 62% (Figure 3.S5).

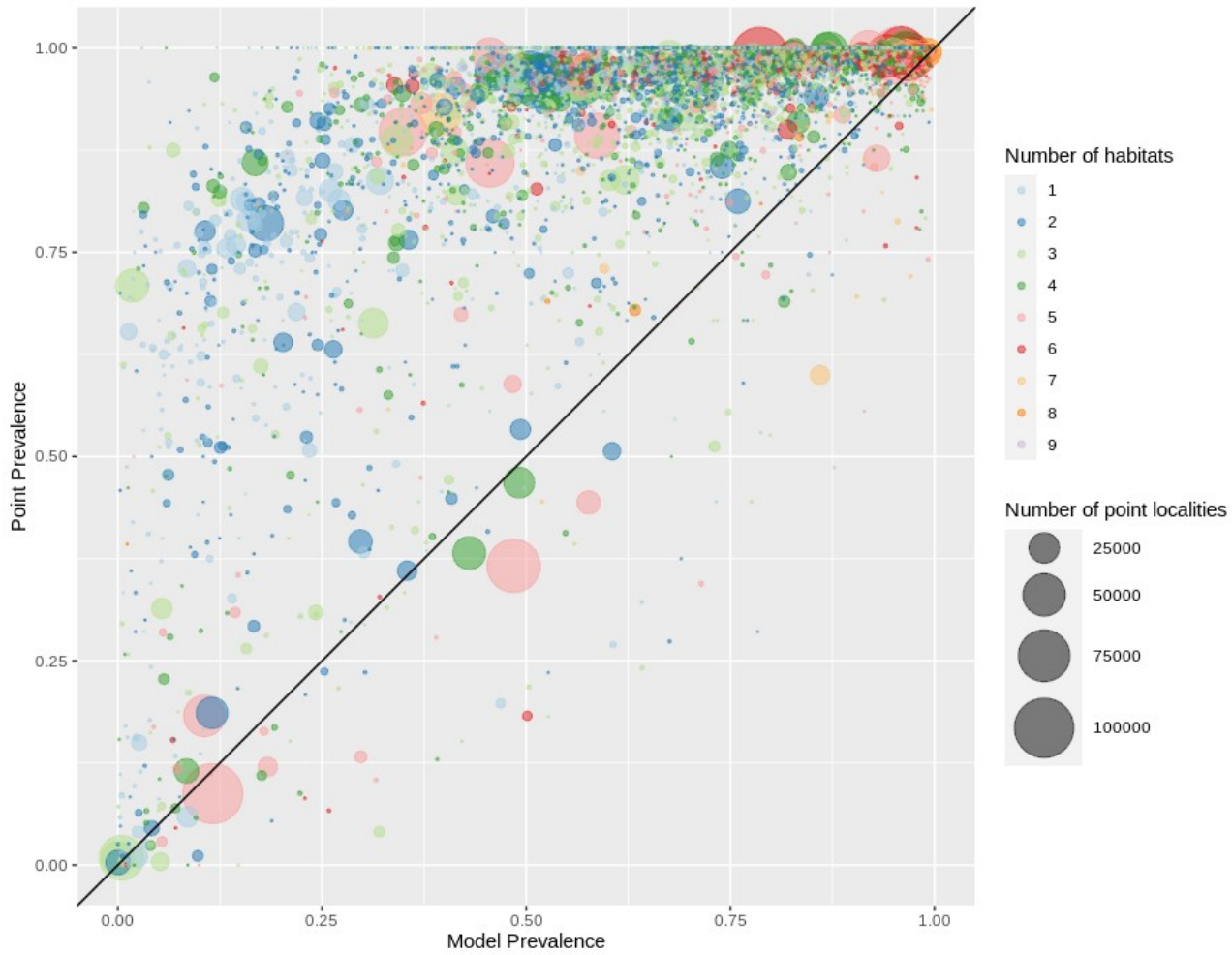

Figure 3.3: Point prevalence vs model prevalence for terrestrial birds. Colors indicate the number of habitats each species is coded to, size of circles indicates the number of point localities.

Out of 420 mammal species (8% of all mammal species) for which point data were available, 399 (95.0%) had point prevalence higher than model prevalence (Figure 3.4). The mean percentage of pixels coded as suitable inside the 300 m buffers of point localities of 420 species of mammals was 78% (Figure 3.S5).

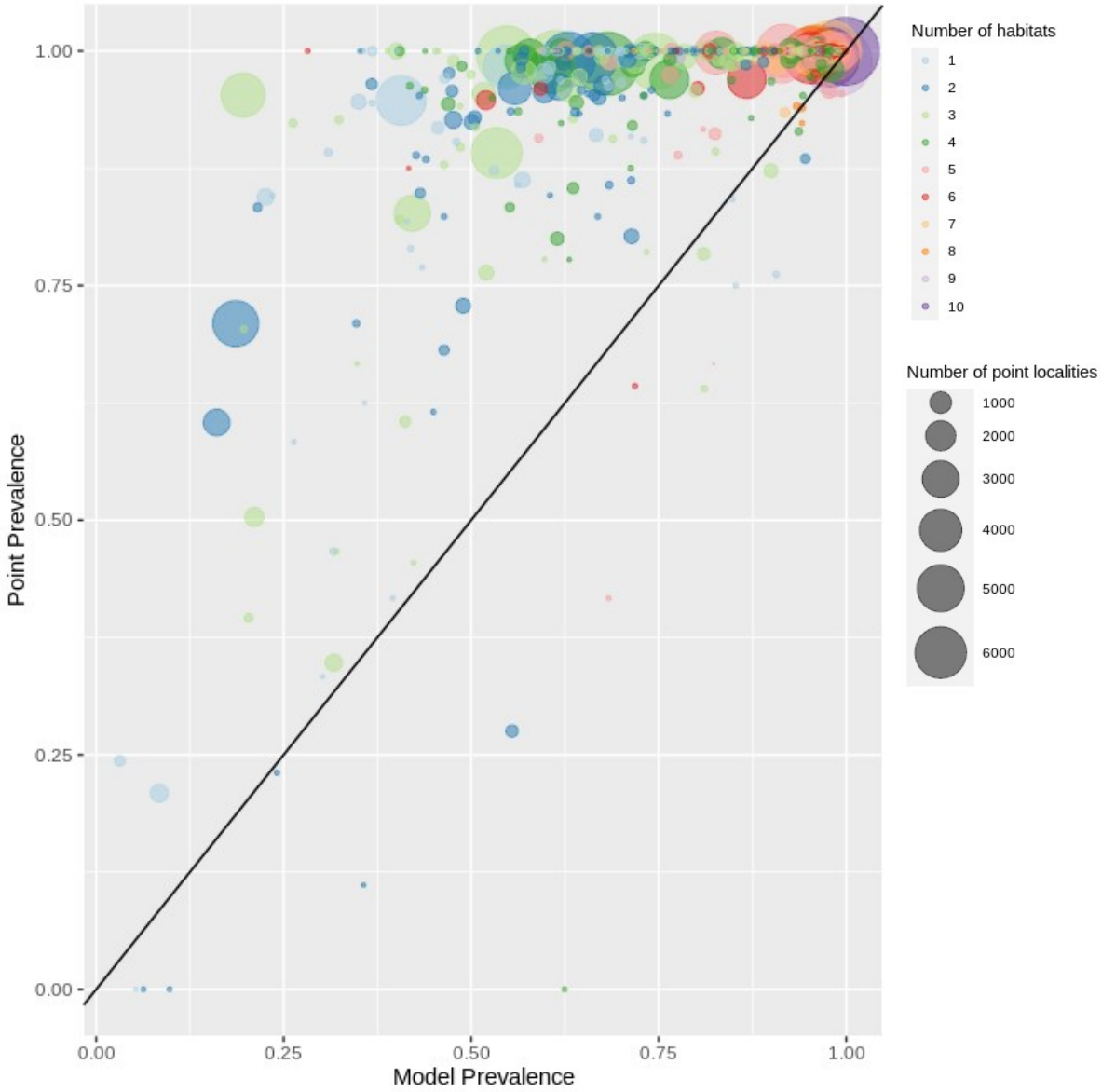

Figure 3.4: Point prevalence vs model prevalence for terrestrial mammals. Interpretation as in Figure 3.3.

3.4.2 Representativeness of validation sample

We found that for birds over 60% all families, genera and orders were represented in the sample included in the point validation and species from all biomes were represented but representation for mammals was lower, as expected due to the much lower proportion of mammal species for which point locality data were available (Figure 3.5).

The validation points were spread across all of the variables and majority of their sub-classes (Figure 3.S6, Figure 3.S7). Species with validation points tended to have larger range sizes, wider elevation ranges and to be coded to more habitat types than those without. Furthermore, validation points were not available for any critically endangered or endangered mammals as these species are rare in the wild.

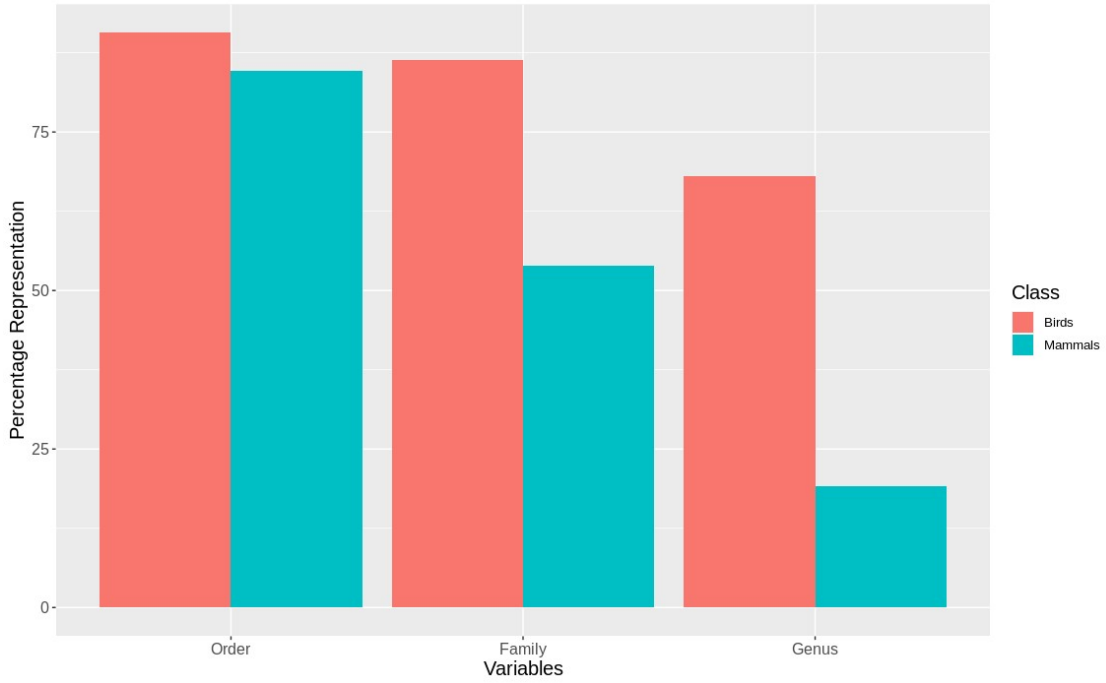

Figure 3.5: Taxonomic representativeness of validation sample for birds and mammals.

3.5 Discussion

On comparing our point validation results with previous validation analysis of AOH maps, we found that validation results are similar to or better than previous exercises. For mammals, Rondinini et al. (2011) evaluated AOH maps for 263 species at 300 m resolution, of which 241 (91.6%) were better than random as compared to 95.0% in our analysis. However, it should be noted that the mean model prevalence for AOH maps of Rondinini et al. (2011) was 54.8 ± 21.5 SD as compared to 65.16 ± 25.42 for our AOH maps. The ratio of mean point prevalence to mean model prevalence for Rondinini et al. (2011) was 1.4 compared to 1.38 in our case. Ficetola et al. (2015) found that AOH for 94% of 115 amphibian species used in the validation analysis were better than random with the mean model prevalence for species with validation points being 0.79 ±

0.21 SD. The ratio of mean point prevalence to mean model prevalence was 1.18 in this case.

Moreover, Catullo et al. (2008) found that 140 AOH maps out of 190 (73.7%) South Asian mammal species gave positive validation results while Rondinini et al. (2005) found the mean proportion of suitable habitats correctly mapped inside the range for 181 species of African vertebrates was 0.55 ± 0.01 SE using presence-absence data sets. The high validation success in our analyses could be attributed to the use of novel habitat-land-cover model (Lumbierres et al., 2021a), the use of logistic regression models to identify systematic errors and the larger validation sample as compared with previous exercises. Furthermore, the underlying land-cover map used in Lumbierres et al. (2021b), has the highest resolution among the global land-cover maps providing with more detailed land-cover classification.

The point validation identified a small proportion of AOH maps which were no better than random. Some of these had high model prevalence. In such cases, point prevalence must be exceptionally high for the models to be better than random since even if a majority of point localities fall within the AOH these maps may perform no better than random. For the AOH maps which were no better than random and had low point prevalence, this was usually due to an apparent error in the coding of elevation range of the species, the areas inside the range of the species where the point localities fell being clipped out by what was assumed to be an erroneous elevation range. A list of species with probably erroneous elevation coding will be forwarded to IUCN Red List team for future corrections.

AOH maps aim to minimize the commission errors known to be present in species ranges without increasing omission errors (Rondinini and Boitani, 2006). One of the limitations of this validation analysis is the inability to quantify the commission errors of the AOH maps as we don’t have the true absence data of the species. Therefore, some uncertainty remains in AOH maps regarding the commission errors.

Also, there are some intrinsic errors in the models as identified by the logistic regression analysis. The species which are coded only to habitats like “Shrubland” might have under-represented model prevalence as discussed above. However, the number of AOH maps identified as lower outliers by the application of the logistic model was low for birds (178/10,475) and for mammals (64/2,758), indicating that for the majority of AOH maps the observed model prevalence was fairly close to that predicted by the model.

3.6 Author contributions

Prabhat Raj Dahal, Paul Donald and Carlo Rondinini conceptualized the idea. Prabhat Raj Dahal did the formal data analysis. Prabhat Raj Dahal led the manuscript writing with contributions from all the authors. Paul Donald, Carlo Rondinini and Stuart Butchart supervised the whole process.

3.7 Code availability

The point localities used in the validation analyses along with the metadata tables summarizing the validation analyses can be found at http://doi.org/10.5281/zenodo.5109073. The same DOI can be used to access the code used for validation and to also access some sample AOH maps which were validated.

3.8 References

Allouche, O., A. Tsoar, and R. Kadmon.: Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS), J APPL ECOL., 43, 6, 1223 – 1232, DOI: 10.1111/j.1365-2664.2006.01214.x, 2006

Bates, D., M. Mächler, B. Bolker, and S. Walker.: Fitting Linear Mixed-Effects Models Using lme4, J STAT SOFTW., 1406 ,1, DOI: 10.18637/jss.v067.i01, 2015

BirdLife International and Handbook of the Birds of the World.: Bird species distribution maps of the world., http://datazone.birdlife.org/species/requestdis, 2019

BirdLife International and Handbook of the Birds of the World.: Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world. Version 5, url:http://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBWBirdLife_Checklist_v5_Dec20.zip, 2020

Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. Schmiegelow.: Evaluating resource selection functions, Ecological Modelling., 157, 281-300, DOI: 10.1016/S0304-3800(02)00200-4

Brooks, T. M., S. L. Pimm, H. R. Akçakaya, G. M. Buchanan, S. H. Butchart, W. Foden, C. Hilton-Taylor, M. Hoffmann, C. N. Jenkins, L. Joppa, B. V. Li, V. Menon, N. Ocampo-Peñuela, and C. Rondinini.: Measuring Terrestrial Area of Habitat (AOH) and Its Utility for the IUCN Red List, TRENDS ECOL EVOL., 34, 11, 977-986, DOI: https://doi.org/10.1016/j.tree.2019.06.009, 2019.

Buchhorn, M., B. Smets, L. Bertels, M. Lesiv, N. Tsendbazar, M. Herold and S. Fritz.: Copernicus Global Land Service: land-cover 100m: epoch 2015: Globe, Dataset of the global component of the Copernicus Land Monitoring Service, doi:10.5281/zenodo.3243509, 2019

Catullo, G., M. Masi, A. Falcucci, L. Maiorano, C. Rondinini, and L. Boitani.: A gap analysis of Southeast Asian mammals based on habitat suitability models, BIOL CONSERV., 141, 11, 2730-2744, DOI: 10.1016/j.biocon.2008.08.019, 2008

Chamberlain, S., V. Barve, D. Mcglinn, D. Oldoni, P. Desmet , L. Geffert , K. Ram .: rgbif:Interface to the Global Biodiversity Information Facility API. R package version 3.5.2, https://CRAN.R-project.org/package=rgbif, 2021

Cold Spring Harbor Laboratory.: Data used in Dahal PR, Lumbierres M, Butchart SHM, Donald PF and Rondinini C (2021) A validation standard for Area of Habitat maps for terrestrial birds and mammals, Available at: https://doi.org/10.1101/2021.07.02.450824, 2021

Cornell Lab of Ornithology.: eBird Basic Dataset. Version: EBD_Jan 2020, Ithaca, New York, 2020

Dahal, P. R., M. Lumbierries, S. H. M. Butchart, P. F. Donald, & C. Rondinini.: Data used, summary and codes: A validation standard for Area of Habitat maps for terrestrial birds and mammals [Data set]. Zenodo. http://doi.org/10.5281/zenodo.5109073, 2021

Ficetola, G. F., C. Rondinini, A. Bonardi, D. Baisero, and E. Padoa Schioppa.: Habitat availability for amphibians and extinction threat: A global analysis, DIVERS DISTRIB., 21, 3, DOI: 10.1111/ddi.12296, 2015

GRASS Development Team.: Geographic Resources Analysis Support System (GRASS) Software, Version 7.2. Open Source Geospatial Foundation. Electronic document:. http://grass.osgeo.org, 2017

Habitats Classification Scheme (Version 3.1).: IUCN, 2012

https://ebird.org, last access: 1st January 2020

https://www.gbif.org, last access: 1st January 2020

Jenkins, C. N. and C. Giri.: Protection of mammal diversity in Central America, CONSERV BIOL.,

22, 4, 1037-44, DOI: 10.1111/j.1523-1739.2008.00974.x, 2008

Jung, M., P. R. Dahal, S. H. M. Butchart, P. F. Donald, X. De Lamo, M. Lesiv, V. Kapos, C. Rondinini, and P. Visconti.: A global map of terrestrial habitat types, Scientific data., 7, 1, 256, DOI: 10.1038/s41597-020-00599-8, 2020

Lüdecke, D., P. D. Waggoner, and D. Makowski.: insight: A Unified Interface to Access Information from Model Objects in R, Journal of Open Source Software, 4, 38, 2019

Lumbierres, M., P. R. Dahal, M. Di Marco, S. H. Butchart, P. F. Donald, and C. Rondinini.: Area of Habitat maps for the world's terrestrial birds and mammals, in preparation., 2021b

Lumbierres, M., P. R. Dahal, M. Di Marco, S. H. Butchart, P. F. Donald, and C. Rondinini.: A habitat class to land-cover translation model for mapping Area of Habitat of terrestrial vertebrates, bioRxiv [pre-print], doi: https://doi.org/10.1101/2021.06.08.447053, 2021a

R Core Team.: R: A language and environment for statistical computing, R Foundation for Statistical Computing, https://www.R-project.org/, 2018

Rondinini, C., S. Stuart, and L. Boitani.: Habitat suitability models and the shortfall in conservation planning for African vertebrates, CONSERV BIOL., 19, 5, 1488 – 1497, DOI: 10.1111/j.1523-1739.2005.00204.x, 2005

Rondinini C.& Boitani L.: Differences in the umbrella effects of African amphibians and mammals based on two estimators of the area of occupancy, CONSERV BIOL., 20, 170-179, DOI: 10.1111/j.1523-1739.2005.00299.x, 2006

Rondinini, C., M. D. Marco, F. Chiozza, G. Santulli, D. Baisero, P. Visconti, M. Hoffmann, J. Schipper, S. N. Stuart, M. F.Tognelli, G. Amori, A. Falcucci, L. Maiorano, and L. Boitani.: Global habitat suitability models of terrestrial mammals, PHILOS T R SOC B., 366, 1578, 2633-41, DOI: 10.1098/rstb.2011.0113, 2011

Stoms, D. M., F. W. Davis, and C. B. Cogan. : Sensitivity of wildlife habitat models to uncertainties in GIS data, PHOTOGRAMM ENG REM S., 58, 843- 850, 1992.

Strimas-Mackey, M., E. Miller and W. Hochachka .: auk: eBird Data Extraction and Processing with AWK. R package version 0.3.0, https://cornelllabofornithology.github.io/auk/, 2018

Sullivan, L., B., C. L.Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling.: eBird: A citizen-based bird observation network in the biological sciences, BIOL CONSERV., 142, 10, 2009

The IUCN Red List of Threatened Species. Version 2020-2.: IUCN, 2020

Tracewski, L., S. H. Butchart, M. Di Marco, G. F. Ficetola, C. Rondinini, A. Symes, H. Wheatley, A. E. Beresford, and G. M.Buchanan (2016).: Toward quantification of the impact of 21 st -century deforestation on the extinction risk of terrestrial vertebrates, CONSERV BIOL., 30, 5, 2016

Wilcox, R. R., Introduction to robust estimation and hypothesis testing.: 4th edition, Elsevier,713 Waltham, Massachusetts, USA, 2017

Chapter 4: A comparison and validation assessment of different Area of Habitat mapping approaches for terrestrial birds and mammals

4.1 Abstract

Area of Habitat (AOH) maps are deductive maps which use information on species' habitat preferences and elevational range to model the distribution of species within their geographical ranges. Three different approaches have been developed to model the distribution of suitable habitats to produce the AOH maps: 1) Using an expert-based land-cover to habitat 'crosswalk', 2) Using an empirically derived habitat – land-cover translation table and 3) Using a spatially explicit global raster layer of terrestrial habitat types. However, there has been no comparison of the performance of AOH maps based on these three methods. We produced AOH maps for terrestrial birds and mammals using a spatially explicit global raster layer of terrestrial habitat types (AOH_L) and compare them with AOH maps produced from the translation table (AOH_T) using six different validation metrics based on model prevalence (the proportion of pixels coded as suitable inside the range of the species) and point prevalence (the proportion of point localities of each species falling in the pixels coded as suitable out of all the point localities of the species). The mean model prevalence for AOH_T was 0.64 for birds and 0.65 for mammals, compared to 0.55 and 0.51 respectively for AOH_L. Of the AOH_T models for birds, 96% performed better than random, compared to 89% for AOH_L; for mammals the equivalent values were 95% and 91% respectively. In terms of the ratio between the reduction in commission error and the introduction of omission errors, 77% of validated AOH_T maps for birds and 73% for mammals performed better than AOH_L. There was high congruence between AOH_T and AOH_L after intersection, with a mean of 77% of pixels within the range being consistently mapped as suitable or unsuitable for both birds and mammals. As the aim of AOH is to reduce commission errors within the range without introducing omission errors, we conclude that the AOH_T mapset is more appropriate for most uses.

4.2 Introduction

An accurate estimate of the geographical distribution of species and their habitats is a first step in species conservation and planning (Gregory et al., 2006, Lawler et al., 2011). Range maps are available for many species, but these are generally drawn to minimize omission errors (i.e. they aim to capture all known locations of a species), which inevitably leads to the introduction of commission errors (many places within a species' range will be unsuitable for it). Area of Habitat (AOH) maps are deductive maps which use information on species' habitat preferences and elevational range to model the distribution of a species within its broad geographical range and thereby reduce commission errors (Brooks et al., 2019). They have a wide range of applications in ecology and conservation; for example, AOH maps can be used to identify Key Biodiversity Areas (Eken et al., 2004).

One of the key challenges in producing AOH maps is to model the habitats of the species, which are often a combination of different factors like climate and land use/cover (Lumbierres et al., 2021). Three different approaches have been developed to produce the AOH maps. The first is to use an expert-based 'crosswalk' to match habitat types from the IUCN (International Union for Conservation of Nature) Red List Habitat Classification Scheme (https://www.iucnredlist.org/resources/habitat-classification-scheme) to land-cover classes from a global land-cover layer (e.g. Rondinini et al., 2010, Ficetola et al., 2015 and Tracewski et al., 2016). For example the IUCN habitat type "Forest" is easily matched to forest classes in the legend of a land-cover layer. However, there are cases when such a direct match between habitat type and land-cover class is not possible; for example the IUCN habitat type "Savanna" could be matched to any of several land-cover classes in the legend of the Copernicus land-cover map (CGLS-LC100; Buchhorn et al., 2020), such as open forest, grasslands, shrublands or herbaceous vegetation habitat class. In such cases, expert-based matching of habitat types to land-cover classes becomes more subjective. This subjectivity can be overcome in a second method, which is to use translation table (Lumbierres et al., 2021), which matches IUCN habitat types to one or more land-cover classes using a logistic regression model of known locations of each species. The third way to produce AOH maps is to use the spatially explicit global raster layer of terrestrial habitat types developed by Jung et al. (2020) using data on climate, land-cover, and land use.

Lumbierres et al. (2021) compared the performance of AOH maps produced using an expert-based crosswalk with those from the translation table and found that both performed equally well. However, the translation table offers considerable advantages over expert-based methods as it is less subjective, it can be applied to different land-cover layers and it offers users a range of thresholds for different purposes. To date, there has been no comparison of the performance of AOH maps produced by translation table with that of maps produced using the new global layer of terrestrial habitat types, so it is unclear which set of maps offers the most promise for a range of conservation applications.

Different validation metrics are available to assess the performance of AOH maps in predicting at a higher spatial resolution where within their broad ranges each species actually occurs. Validation can use presence or presence-absence data from sampled surveys or opportunistic sightings. The majority of data available for species globally are presence-only data, such as those generated by citizen science recording schemes. Such data allow the quantification of omission errors (i.e. cases where the species is present despite being mapped as absent), while absence data allow the quantification of commission errors (i.e. cases where the species is absent despite being mapped as present). Validation metrics like TSS, kappa, prevalence (Allouche et al., 2006) and the Boyce Index (Boyce et al., 2002) are commonly used to quantify commission and omission errors and assess the performance of distribution maps. When absence data are not available, it is still possible to determine if a distribution map is better than random (Rondinini et al., 2010). If no presence or absence data are available, logistic regression models can be used to evaluate AOH maps (Dahal et al., 2021).

In this paper, we produced AOH maps for 10,651 terrestrial birds and 5,481 mammals using information on habitat preferences and elevational range from the IUCN Red List, and the global map of terrestrial habitat types of Jung et al. (2020) (referred to as AOH_L hereafter). We then compared their performance in reducing commission errors without unduly increasing omission error rates with that of AOH maps based on translation table (referred to as AOH_T hereafter) between habitat types and land-cover classes (Lumbierres et al., 2022) using the validation protocol developed by Dahal et al. (2021).

4.3 Methods

4.3.1 Production of AOH maps using global map of terrestrial habitat types

We used the global map of terrestrial IUCN habitat types (Jung et al., 2020) to map apparently suitable habitats inside the elevational and geographic range of each species to produce AOH maps for 5,481 terrestrial mammals and 10,651 terrestrial birds. The data set of species and their habitat associations and elevation data from the IUCN Red List were identical to those used by Lumbierres et al. (2022) in order to allow meaningful comparisons between the two sets of AOH maps. The global map of terrestrial habitat types is a global raster map at c. 100 m resolution of 47 different habitat types as defined in the IUCN Habitat Classification Scheme (IUCN, 2021). The classification scheme categorizes broad habitat types at level-1 (e.g. Forest) while level-2 refers to finer level habitat types (e.g. Subtropical/tropical dry forest). The global map of terrestrial habitat types from Jung et al. (2020) maps habitat types at both levels.

The geographical range polygons for mammals from IUCN (2020) and for birds from BirdLife International and Handbook of the Birds of the World (2020) are coded for different combinations of species' presence, origin and seasonal occurrence. We selected polygons with presence coded as "extant" and "probably extant", origin coded as "native" or "re-introduced" or "assisted colonization". For migratory birds, we combined polygons coded as "resident", "breeding" or "uncertain" and treated them as the breeding range of the species, and we combined polygons coded as "resident", "non-breeding" and "uncertain" and treated them as the non-breeding range of the species.

To produce the AOH maps, we first selected the areas inside the geographic range of each species falling within its elevation range using Shuttle Radar Topographic Mission Digital Elevation Model (SRTM DEM) (Jarvis et al., 2008), re-sampled at 100-m resolution. Then within these areas we extracted the relevant level-1 habitat types from the global map of terrestrial habitat types. We used only level-1 habitat types to allow comparison with Lumbierres et al. (2022).

The only difference between AOH_T and AOH_L is the habitat mapping method. AOH_T used a habitat-land-cover translation table developed in Lumbierres et al. (2021) which associates each habitat

class with one or many land-cover classes, while AOH_L uses habitat maps directly extracted from Jung et al. (2020).

We assessed the accuracy of the AOH_L maps using the approach developed by Dahal et al. (2021). This protocol includes a two-step validation using logistic models to first estimate the proportion of suitable habitat pixels inside the geographical range of the species (i.e. 'model prevalence') using several independent variables (elevation range, mid-point of elevation range, number of habitats associated with the species, geographical realm of the species and seasonality in case of birds). These estimates were compared with the observations to identify outliers that were then carefully examined to identify systematic errors. Logistic models were built for 10,475 AOH maps for 9,163 bird species and 2,758 AOH maps for 2,758 mammal species for which we had minimum and maximum elevations. The second step involved validation using the point localities downloaded from eBird (eBird Basic Dataset, 2020) for birds and GBIF (Global Biodiversity Information Facility) (www.gbif.org) for mammals. The point localities for each species were overlapped with its AOH map(s) to calculate point prevalence (the proportion of points falling within the AOH map). The point data set was identical to that used in Dahal et al. (2021) to validate the AOH maps in Lumbierres et al. (2021), and included point localities for 420 mammal (8% of all terrestrial mammals) and 4,889 bird species (45% of all terrestrial birds), with each species having at least 10 occurrence points from 2019-2020. The data acquisition, filtering process and logistic modeling are described in detail in Dahal et al. (2021).

4.3.2 Comparison of AOH maps

To compare the two sets of AOH maps, we assessed 1) the total number of logistic outliers identified by the logistic regression models, i.e. those whose model prevalence appeared to be too high or too low for the characteristics of the species; 2) the total number of AOH maps that performed no better than random; 3) model prevalence; 4) point prevalence; 5) the mean omission rates relative to the reduction in model prevalence (from the mean difference between model prevalence of AOH_T and AOH_L and the mean difference in point prevalence of AOH_T and AOH_L); and 6) a comparison metric based on the ratios of model and point prevalence in such a way that models with higher omission errors were penalized. This comparison metric was used to select the best AOH maps among AOH_T and AOH_L for the same species. The comparison metric was calculated for each set of maps as:

((1-point prevalence)/(1-model prevalence))*(1-point prevalence)

For any given species, the lower of the two values was considered to reflect better performance in terms of reducing commission errors while penalizing omission errors. When values were the same, the mode with the lower model preference was considered to have better performance (i.e. it reduced commission errors more without adding to omission errors).

In order to identify particular habitats with higher omission rates relative to the reduction in model prevalence, we assessed the mean model prevalence and mean point prevalence for habitat specialist species (defined here as species associated with only one habitat class). There were 683 forest specialist species (combining both mammals and birds) with 181,989 point localities, 4 savanna specialists (227 points), 52 shrubland specialists (11,778 points), 29 grassland specialists (7,781 points) and 164 wetland specialists (205,295 points). Other habitats had zero or one specialist species, and hence were not considered.

Finally, we also intersected AOH_T and AOH_L for each species to identify the pixels which were 1) mapped as suitable in both AOH maps, 2) mapped as unsuitable in both AOH maps, 3) mapped as suitable in AOH_T but unsuitable in AOH_L and 4) mapped as suitable in AOH_L but unsuitable in AOH_T.

4.4 Results

The average model prevalence for AOH_L was 0.55±0.28 SD for birds and 0.51±0.29 SD for mammals compared to 0.64±0.27 SD for birds and 0.65±0.28 SD for mammals in AOH_T (Figure 4.1).

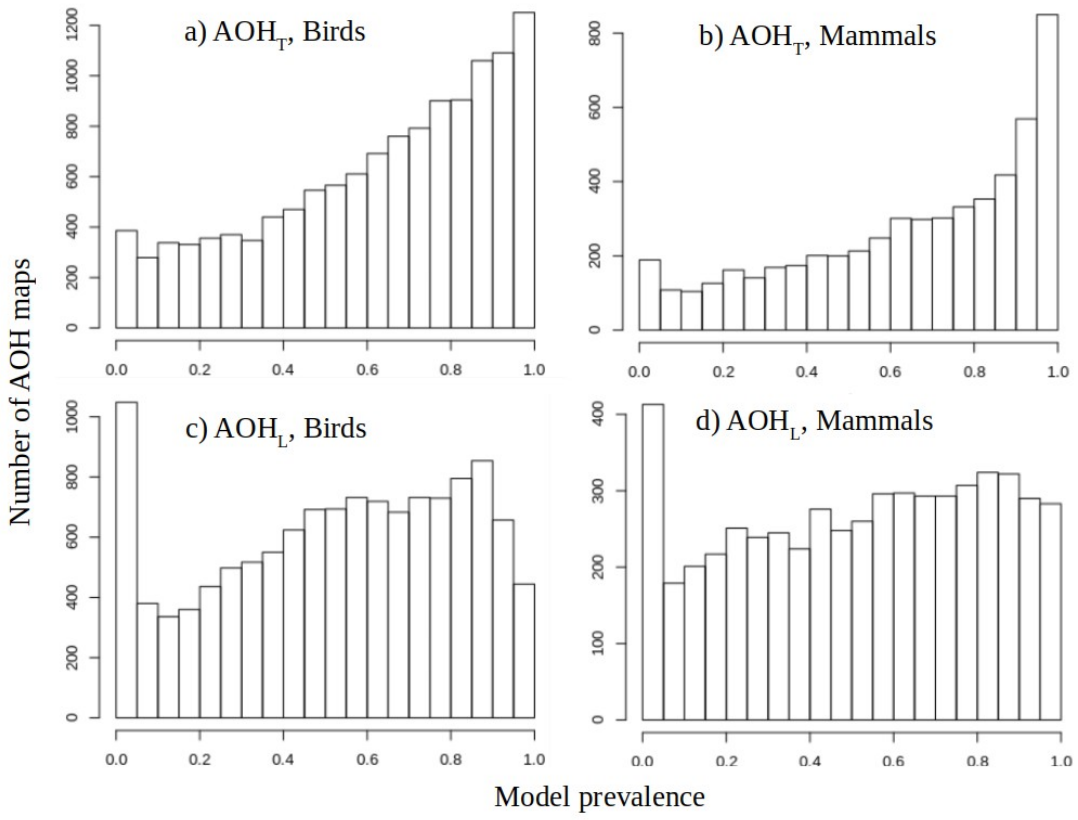

Figure 4.1: Histogram of model prevalence for terrestrial mammals and birds for AOH_T and AOH_L.

After applying the logistic regression models, 24 (< 1%) upper and 64 (< 1%) lower outliers out of 10,475 AOH_L maps for birds were identified compared with 18 (1.1%) upper and 178 (1.7%) lower outliers in AOH_T. For mammals, there were 18 (0.65%) upper and 23 (< 0.83%) lower outliers out of 2,758 AOH_L maps as compared with 21 (0.8%) upper and 64 (2.3%) lower outliers in AOH_T.

For AOH maps of birds in AOH_L, we found that for 4,350 (90.0%) species the point prevalence was higher than the model prevalence , hence AOH maps for these bird species were better than random

(Figure 4.2). For AOH_T, the total number species of birds whose AOH maps were better than random was 4,689 (95.9%) species.

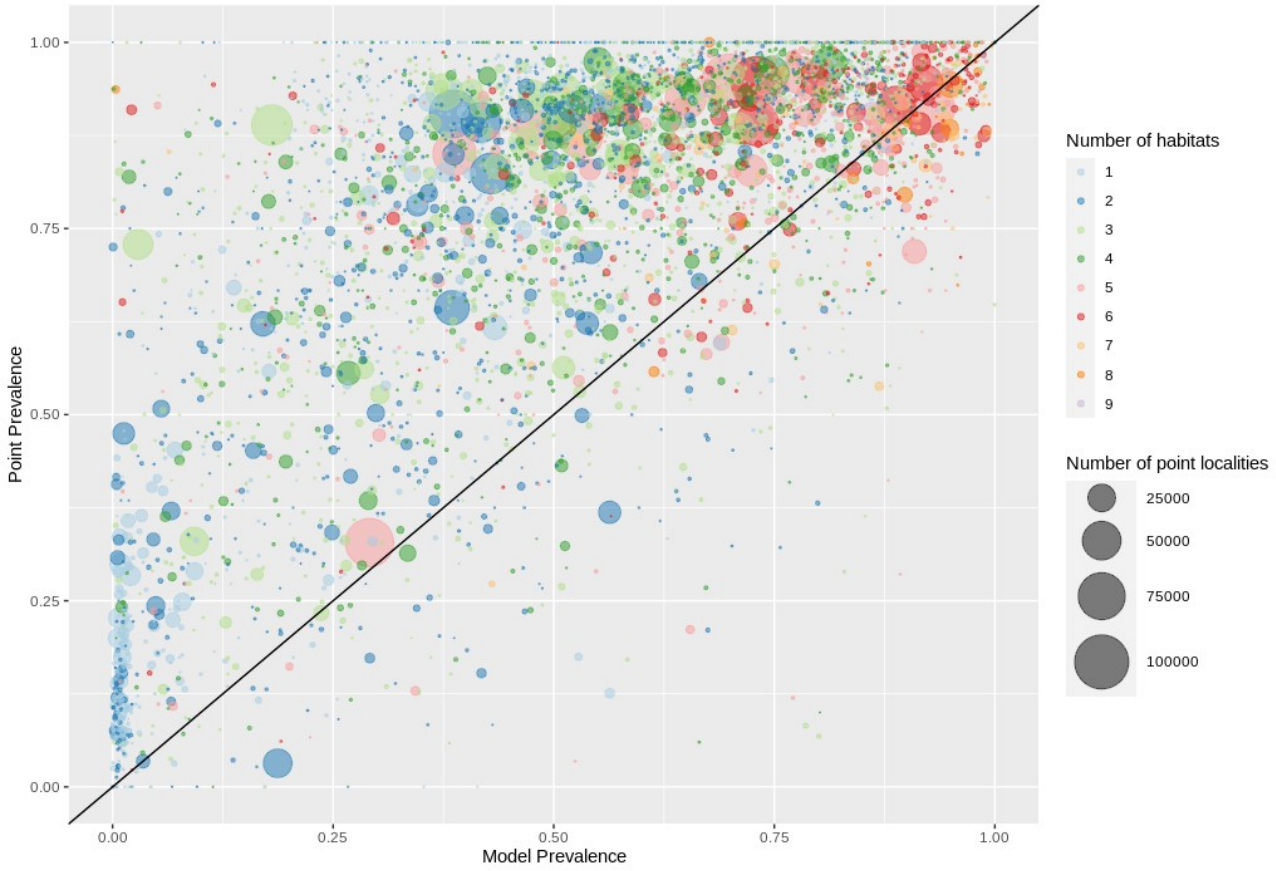

Figure 4.2: Model prevalence vs point prevalence for birds of AOH_L. Species with better than random AOH maps are represented by circles above the diagonal line.

Similarly for mammals in $AOH_{L,}$ we found that for 381 (90.7%) species the model prevalence was less than the point prevalence hence the AOH maps are better than random for these species (Figure 4.3). For AOH_T, the total number AOH maps of mammals that were better than random was 399 (95.0%). Similar scatter plots can be found in Dahal et al. (2021) for AOH_T for birds and mammals.

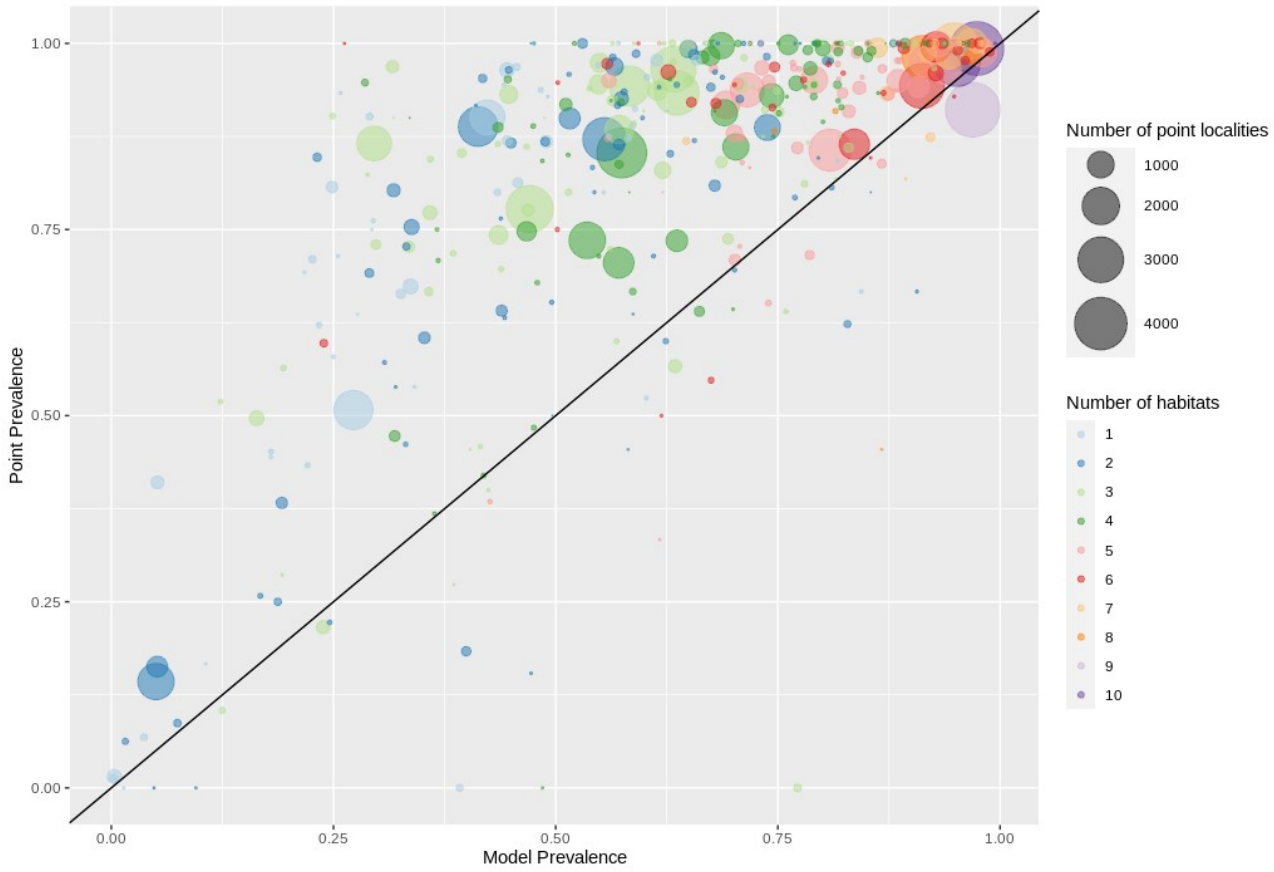

Figure 4.3: Model prevalence vs points prevalence for mammals of AOH_L. Species with better than random AOH maps are represented by circles above the diagonal line.

	Model prevalence				**Point prevalence**			
	1st Q	Median	Mean	3rd Q	1st Q	Median	Mean	3rd Q
AOH_L (Birds)	0.35	0.56	0.53	0.75	0.68	0.86	0.77	0.95
AOH_T (Birds)	0.49	0.70	0.65	0.85	0.90	0.97	0.90	1
AOH_L(Mammals)	0.47	0.66	0.63	0.81	0.77	0.92	0.83	0.98
AOH_T (Mammals)	0.61	0.77	0.74	0.94	0.95	1	0.93	1

Table 4.1: Summary of model and point prevalence of AOH_T and AOH_L for mammal and bird species having point localities.

For mammals, the mean model prevalence was lower by 0.11 for AOH_L than AOH_T and the corresponding mean point prevalence was 0.10 lower. For birds, the mean model prevalence was low by 0.11 for AOH_L than AOH_T and the corresponding mean point prevalence was 0.13 lower (Table 4.1).

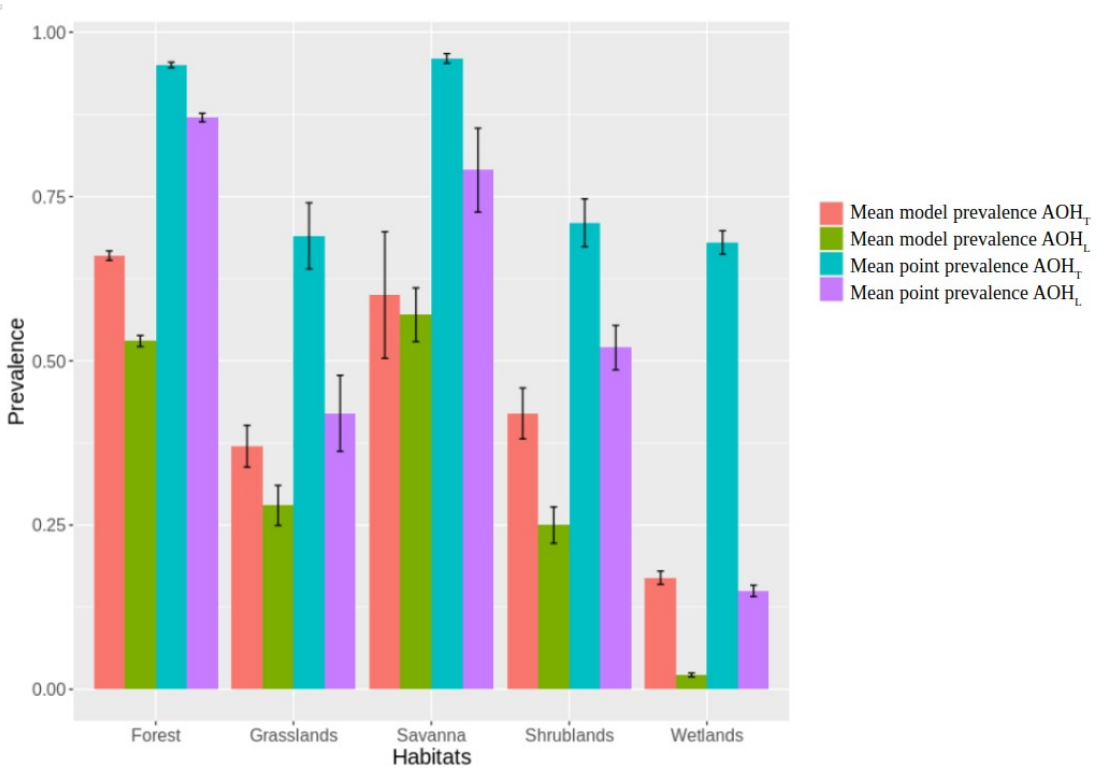

Figure 4.4: Difference in mean model and point prevalence for each level-1 habitat class for habitat specialist species with standard errors.

The difference in the mean point prevalence between AOH_T and AOH_L for wetland specialist species is the highest (0.53) among the five habitat types shown in Figure 4.4 indicating high omission errors for species associated with the habitat class "Wetlands" in AOH_L. For grassland and savanna specialist species the difference in the mean point prevalence between AOH_T and AOH_L is 0.27 and 0.17 respectively. For forest and shrubland specialist species the difference in the mean point prevalence between AOH_T and AOH_L is similar to difference in the mean model prevalence.

When we intersected AOH_T and AOH_L we found that a mean of 53.4% of pixels inside the range of the species were mapped as suitable in both sets of AOH maps for birds and 58% for mammals, while 23.22% of the pixels were mapped as unsuitable in both sets of AOH maps for birds and 19%

for mammals. A total of 15% of pixels were identified as suitable in AOH_T but unsuitable in AOH_L for birds and 17% for mammals, while 7.4% of pixels were identified as unsuitable in AOH_T but suitable in AOH_L for birds and 3% for mammals (Figure 4.5).

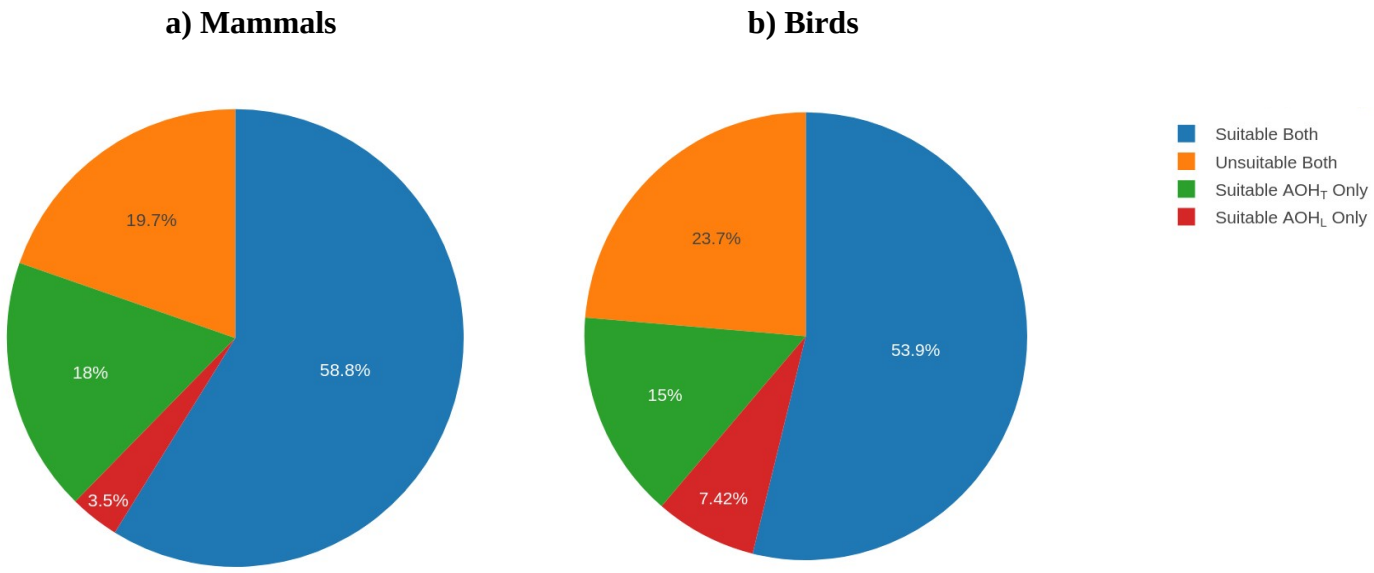

Figure 4.5: Summary of intersection between AOH_T and AOH_L for a) mammals and b) birds.

Using the comparison metric described in the methods (metric number 6 in methods) the two sets of AOH maps were compared. For birds, out of 4,889 AOH maps with point localities, 3,813 (77%) AOH maps in AOH_T performed better than AOH maps in AOH_L while 1059 (22%) AOH maps in AOH_L performed better than AOH maps in AOH_T and for 17 AOH maps (0.34%) performance was identical. For mammals, out of 420 AOH maps with point localities, 305 (73%) AOH maps in AOH_T performed better than AOH maps in AOH_L while 112 (26%) AOH maps in AOH_L performed better than AOH maps in AOH_T and for 3 AOH maps (0.7%) performance was identical.

4.5 Discussion

Using the validation metrics described above, we found that AOH_T generally performed better than AOH_L in terms of number of AOH maps which were better than random and in terms of number of AOH maps with fewer omission errors. However, we found a high congruence between AOH_T and AOH_L after intersection (53.9% mapped as suitable and 23.7% mapped as unsuitable in both the mapsets for birds and 58.8% mapped as suitable and 19.7% mapped as unsuitable in both the mapsets for mammals).

The average model prevalence was lower for the AOH_L compared to AOH_T. This can be attributed to the difference in number and types of land-cover classes associated with each habitat. For example, Lumbierres et al. (2021) included 100% fractional tree cover while mapping the habitat class “Forest” and “Savanna” whereas Jung et al. (2021) included more than 50% fractional tree cover causing mean model prevalence and corresponding point prevalence to be lower for species associated with habitat class “Forest” and “Savanna” in AOH_L as compared to AOH_T. Also, “Savanna” habitat class in Jung et al 2020 is associated with land-cover class “Herbaceous vegetation” while Lumbierres et al. (2021) did not include this land-cover. The habitat class “Shrubland” is associated with the land-cover class “Shrubland” but other habitat types like “Desert” and “Savanna” were also associated with this land-cover in Jung et al. (2020) whereas in Lumbierres et al. (2021) all of land-cover class “Shrubland” gets mapped. The habitat class “Grassland” and “Wetlands” are associated with land-cover class “Cropland” in Lumbierres et al. (2021) however this is not the case in Jung et al. (2020). This has caused large omissions for the species coded to habitat types “Grassland” and “Wetlands” in AOH_L.

The AOH_L maps clearly perform better for species associated with artificial habitat types, as “Artificial terrestrial plantations and heavily degraded forest”, “Pastureland” and “Rural gardens” are mapped in Jung et al 2020, but omitted in Lumbierres et al. (2021), leading to omission errors for AOH_T.

For birds, there were 64 lower outliers in AOH_L out of which 14 outliers were common with the lower outliers in AOH_T. These were cases where the elevational range was narrow and excludes a lot of areas inside the geographical range of the species. For other 50 lower outliers, errors related to habitat mapping was identified (very low suitable habitats mapped inside the range). For

mammals, there were 23 lower outliers in AOH_L out of which 16 were common with the lower outliers in AOH_T. These are the cases where the elevational range excludes a lot of areas inside the geographical range of the species and the remaining outliers are associated with errors in habitat mapping. We did not consider the upper outliers as the logistic model estimates very low model prevalence at higher elevations as described in Dahal et al. (2021).

The number of AOH maps for which the model prevalence is less than the point prevalence (no better than random models) is higher for the AOH_L (10%) as compared to AOH_T (5%). This is due the omissions discussed earlier while mapping different habitat types in Jung et al. (2021).

The intersection results show that only 3% of pixels on average inside the range is mapped as suitable in AOH_L but unsuitable in AOH_T for mammals and 7.5% for birds. If AOH_T is used, all the pixels in the AOH_L will also be mapped except the ones which are mapped as suitable in AOH_L but unsuitable in AOH_T. This implies that AOH_T includes 97% of the pixels included in AOH_L for mammals and 93.5% for birds, indicating that AOH_L is largely a subset of AOH_T.

After comparing AOH_T and AOH_L, we conclude that AOH_L has higher level of omissions for natural habitat types whereas AOH_T has higher levels of omission for artificial habitat types. Therefore, we provide the intersected maps which can be used by the users as per their need in combination with the validation statistics provided in the metadata. The pixels which are mapped as suitable or unsuitable in both the mapsets are of importance as they show congruence in habitat mapping methods.

One of the major limitations of this analysis is the inability to quantify the commission errors using presence-only data sets. The global data repositories have presence-only data for most of the species with potential geographic and taxonomic biases. Furthermore, all species could not be validated using the presence data due to unavailability of data in global repositories like GBIF and eBird. Furthermore, due to lack of elevation ranges all the AOH maps could not be validated using the logistic model. For the production of AOH maps there are also some limitations like incorrect IUCN habitat associations of the species in the IUCN Red List as pointed out in Jung et al. (2020). Moreover, the underlying remote sensing products like the Copernicus land-cover map can have limitations such as inability to distinguish between cropland and herbaceous vegetation. Furthermore, AOH maps without elevation range data will include all areas including suitable habitats only inside the range leading to commission error. We conclude that for most purposes, the

use of AOH_T is likely to be preferable to AOH_L however for the species associated artificial habitat types AOH_L is likely to perform better.

4.6 Author contributions

Prabhat Raj Dahal, Paul Donald and Carlo Rondinini conceptualized the idea. Prabhat Raj Dahal led the data analyses and manuscript writing. All other authors contributed in writing the manuscript.

4.7 References

Allouche O, Tsoar A, and Kadmon R (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS) . *Journal of Applied Ecology*, 43, 6.

BirdLife International and Handbook of the Birds of the World (2020) Bird species distribution maps of the world. Version 2020.1. Available at http://datazone.birdlife.org/species/requestdis

Boyce M. S, Vernier P. R, Nielsen S. E, et al. (2002). Evaluating resource selection functions. *Journal of Applied Ecology*, 43, 6.

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 & Evolution*, 34, (11), 977. –986.

Buchhorn M, Smets B, Bertels L, et al. (2020). Copernicus Global Land Service: land-cover 100m: Version 3 Globe 2015–2019: Product User Manual. Geneve: Zenodo.

Dahal P R, Lumbierres M, Butchart S. H. M, et al. (2021). A validation standard for Area of Habitat maps for terrestrial birds and mammals, *Geosci. Model Dev. Discuss.* [preprint], https://doi.org/10.5194/gmd-2021-245, in review.

eBird. (2019). *Basic dataset*. Version January 2019. Ithaca, NY: Cornell Lab of Ornithology.

Eken G, Bennun L, Brooks T. M, et al. (2004). Key Biodiversity Areas as Site Conservation Targets. *BioScience*, 54,12.

Ficetola G. F, Rondinini C, Bonardi A, et al. (2015). Habitat availability for amphibians and extinction threat: A global analysis. *Diversity and Distributions*, 21, 3.

IUCN. (2012). *Habitats classification scheme (version 3.1)*. Gland: IUCN.

IUCN. (2013). *Documentation standards and consistency checks for IUCN Red List assessments and species accounts. Version 2*.Gland: IUCN.

IUCN. (2020). *The IUCN Red List of Threatened Species. Version 2020-2*. https://www.iucnredlist.org. Downloaded on 09 May (2020).

Jarvis A, Reuter H. I, Nelson A, et al. (2008). Hole-filled SRTM for the globe version 4. *CGIAR-CSI SRTM 90 M Database*, srtm.csi.cgiar.org.

Jung M, Dahal P. R, Butchart S. H. M, et al. (2020). A global map of terrestrial habitat types. *Scientific Data*, **7**, 1–8.

Lawler J. J, Wiersma Y. F and Huettmann F (2011). Using Species Distribution Models for Conservation Planning and Ecological Forecasting. *In: Drew C., Wiersma Y., Huettmann F. (eds) Predictive Species and Habitat Modeling in Landscape Ecology*. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-7390-0_14

Lumbierres M, Dahal P. R, Di Marco M, et al. (2021). Translating habitat class to land-cover to map area of habitat of terrestrial vertebrates. *Conservation biology* , ISSN: 1523-1739. DOI: 10.1111/cobi.13851. PMID: 34668609

Lumbierres, M, Dahal P. R, Di Marco M, et al. (2022). Area of Habitat maps for the world's terrestrial birds and mammals. *in preparation*

McDermid G. J, Franklin S. E, and LeDrew E. F (2005). Remote sensing for large-area habitat mapping. *Progress in Physical Geography*, 29, 4.

Rondinini C, Di Marco M,Chiozza F, et al. (2011). Global habitat suitability models of terrestrial mammals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 366, 1578, pp. 2633-41. ISSN: 1471-2970. DOI: 10.1098/rstb.2011.0113. PMID: 21844042.

Tracewski L, Butchart S. H. M, Di Marco M, et al. (2016). Toward quantification of the impact of 21st-century deforestation on the extinction risk of terrestrial vertebrates. *Conservation biology*, 30, 5, pp.1070-9. ISSN: 1523-1739. DOI: 10.1111/cobi.12715. PMID: 26991445.

www.ebird.org

www.gbif.org, last access: 1st January 2020

Chapter 5: Discussion

In this thesis, AOH maps of 10,651 terrestrial birds and 4,581 terrestrial mammals were produced using two new habitat modeling methods (Jung et al., 2020, Lumbierres et al., 2021). The AOH mapsets produced using the translation table (Lumbierres et al., 2021) and the global map of terrestrial habitat types (Jung et al., 2020) will be referred hereafter as AOH_T and AOH_L respectively. The maps are based on the latest geographic ranges, habitat information and elevation data available from the IUCN Red List. These maps were validated using a novel standard validation protocol that I developed. Moreover, the best threshold among the three thresholds for the translation table (Lumbierres et al., 2021) was identified by comparing model and point prevalence across all three thresholds. A validation method that was able to quantify both commission and omission errors using presence-only point localities of the habitat specialist species was also developed. This method was used to validate the global map of terrestrial habitat types (Jung et al., 2020).

A habitat model with high omission or commission error can have serious implications (Rondinini et al., 2012). For example: habitat models with more omission errors can lead to exclusion of areas of high importance in terms of conservation. This can reduce the effectiveness of any conservation plans based on habitat models. Similarly, habitat models with commission errors can include areas that might not be important thereby increasing the cost of conservation. Due to the uncertainties in data associated with species, perfect models with zero omission and commission errors are not feasible. Therefore, it is important to quantify these errors and take them into consideration when using the models for conservation applications (Rondinini et al., 2006).

The availability of validation data determines if omission or commission errors can be quantified. As discussed in the introduction section, presence - absence data for a large number of species at global scale are scarce. In the second chapter, using the presence-only data set of habitat specialist species, both the commission and omission error for each of the habitat types were quantified, and balanced accuracy (Kuhn et al., 2020) was calculated. Balanced accuracy considers both commission and omission errors and provides a holistic assessment of the performance of the model. We found the highest balanced accuracy for the habitat type “Forests” and the lowest for “Wetlands”. The habitat type “Forests” is one of the most prevalent habitat types and majority of validation points also belonged to this habitat. The habitat type “Wetlands” are often hard to map

using remote sensing as they are often covered by canopies and too small to be captured by the resolution of the land-cover maps. Also, the habitat type “Wetlands” had the lowest validation sample. The mapped habitats with high balanced accuracy are robust in terms of both commission and omission errors and can be used with high confidence.

As majority of species are found in more than one habitat, the approach used in the second chapter could not be applied to validate the AOH maps for most birds and mammals in the third chapter. Therefore, it was only possible to estimate the omission errors by calculating point prevalence for the AOH maps. The objective of the AOH map is to reduce commission errors of the geographic range map without introducing omission errors (Brooks et al., 2019). By comparing the model and point prevalence it was possible to quantify the omission errors in the AOH maps relative to the reduction in commission errors of the geographic range maps. Ideally AOH maps with lower model prevalence with a point prevalence of 1 is desired. Based on the omission errors introduced relative to the reduction in commission error of the geographic range map, AOH_T performed better than AOH_L. However, some of the artificial habitat types like “Plantations”, “Pastureland” and “Rural gardens” could not be mapped in AOH_T but are mapped in AOH_L.

According to Bunce et al. (2012), land-cover/land use classes are less suitable for assessing biodiversity distribution compared with habitat types, which are linked to species, communities and biotopes. However, using point locality and habitat data of species, habitats can be modeled as a combination of different land-cover classes (Lumbeirres et al., 2021). In doing so the uncertainties associated with the data used in modeling propagates into the AOH models. Such uncertainties arise from the land-cover maps, the point locality data and the habitat association of the species used for modeling. For example: remote sensing sensors distinguish poorly between land-cover classes whose spectral signatures are similar (for example: grasslands and cropland), which adds to the uncertainty of the land-cover maps. Also, the presence-only data sets used to model the habitat types are opportunistic sighting data and can have taxonomic and spatial biases. Furthermore, the habitat associations of the species in the IUCN Red List can sometimes be erroneous, as pointed out in Jung et al. (2020).

The AOH maps inherit the resolution of the land-cover maps used to model the habitats. However, whether this is the optimal resolution to represent the habitat of the species has not been examined yet. With the resolution of the land-cover maps regularly increasing, the cost and benefits of mapping at finer resolution must be assessed (Jetz et al., 2012). During the course of this thesis, the

resolution of available global land-cover maps increased 10-fold: the latest global land-cover map released by ESRI (Karra et al., 2021) has a resolution of 10 m as compared with 100 m for the Copernicus land-cover map (Buchhorn et al., 2020). The benefit of using high resolution land-cover maps will be the greater level of detail provided by them. However, the large computational power to process and store such data need to be accounted for. For species with large range sizes living in a relatively specialized habitat, fine resolutions might not be optimal for representing distribution. On the other hand, the mapping of species with smaller ranges distributed in multiple habitats will benefit from the greater level of detail in fine resolution maps. Also, there seems to be some temporal discrepancy among the habitat association and land-cover maps which are used to model the habitats. The land-cover map used to model the habitats is for the year 2015 while the habitat associations in the IUCN Red List are updated in 2020 for birds and mammals (although the coding may not have been revised since previous assessments of each species, which in some cases may be more than a decade ago). However, the AOH production methods developed allow for the data to be easily updated using updated land-cover maps.

The AOH models should be revised as data associated with each species are updated, including range boundaries, habitats and elevation ranges. The AOH maps with the most recent data on geographical range, land-cover, habitat preferences and point localities will be able to represent the current distribution of species most accurately. The availability of open source tools like “rredlist” (Chamberlain, 2020), “rgbif” (Chamberlain et al., 2022), “rgrass7” (Bivand, 2019) and “Google Earth Engine” (Gorelick et al., 2017) provide an opportunity to construct an automated online system to produce updated AOH maps. However, to update large number of AOH maps regularly, there is a large computational cost associated.

The updated AOH maps will be now available for mammals and birds species listed in the IUCN Red List. However, updated AOH maps for other taxonomic groups like amphibians and reptiles are missing. The AOH production and evaluation methods developed in this thesis can be used to produce the AOH maps for other taxonomic groups as well.

The updated AOH maps can be of high value for the IUCN Red List assessments of the species. As the distribution of the species and the threats associated with the species change with time, the IUCN Red List is also updated regularly to capture the current status of the species. As these updates are done by the experts based on several collated data sources, the process slow and costly (Rondinini et al., 2014). There is a need to make the IUCN Red List assessment more systematic

and comprehensive (Santini et al., 2019). The AOH maps can be updated regularly and combined with other data relevant to the species offers a more automated and systematic approach to the IUCN Red List assessments. The AOH maps can be used to estimate key parameters (extent of occurrence, maximum area of occupancy, population size and trend, and degree of fragmentation) which can be used to assess the threat status of the species under different IUCN Red List criteria (Brooks et al., 2019). Santini et al. (2019) used the AOH maps of ca. 15,000 bird and mammal species to assess the threat status of the species under criteria A2, B1, B2, C1, D and D2 by first estimating the upper limits of AOO using the AOH maps and then by estimating population of the species inside the AOH maps using population density models. Furthermore, others (Buchanan et al., 2008; Rondinini et al., 2011; Bird et al., 2012; Visconti et al., 2016; Tracewski et al., 2016) have also used AOH maps to assess the threat status of the species under different IUCN Red List criteria.

Of the five different criteria, criterion E of the IUCN Red List is the most under utilized criteria to assess the species. To assess the threat status of a species using this criteria requires the estimation of probability of extinction of a species via population viability analysis (IUCN, 2017). Population viability analysis requires data on population of the species over a period of time to estimate the probability of extinction of species (Brook et al., 2000). Time series of AOH maps combined with population density models can provide the opportunity to estimate the population of species over a long period of time which can be used to quantify the probability of extinction of species under criterion E. This can help to compare IUCN Red List assessments using different criteria for the same species to check for consistency among the criteria.

The recent “post 2020 global biodiversity framework” (Convention on Biological Diversity, 2022) is one of the important global framework related to biodiversity conservation. It includes 21 action-oriented targets for urgent action over this decade. Some of these targets envision integrated biodiversity-inclusive spatial planning, well-connected systems of protected areas and other effective area-based conservation measures, reducing human-wildlife conflict and minimizing the impact of climate change on biodiversity. In order to implement and monitor these targets accurate information on distribution of species is required. The AOH maps for several taxonomic groups can be used to measure and track the achievement towards these targets. Furthermore, AOH maps can be projected under different land use and climate change scenarios and can be used as tools to measure the progress towards the sustainable development goals (UN General Assembly, 2015) like “Life on land” and “Climate action”.

Despite the development of different habitat modeling methods, there is still room for improvement of the habitat models. One of the major issues that remains is the inability to quantify both omission and commission errors for many of the habitat models due to unavailability of absence data of the species on a global scale. Therefore, there is scope to model or compile absence data sets of species that can be used to assess and improve the performance of the habitat models. Global repositories like eBird also have checklists for sampling locations which can be used to infer the absences of the species. Also, for the majority of mammal species, the data on elevation range are missing; these could be complied using published literature.

The inductive species distribution models like MAXENT (Maximum Entropy Model) (Phillips et al., 2006) and biomod2 (Biodiversity Modeling 2) (Thuiller et al., 2009) have gained popularity in the recent years. These models use regression and machine learning models along with the presence only point localities of the species with several bio-climatic variables to model the potential habitat distribution of the species. There has been little work on integrating inductive and deductive models, which could further improve the representation of distribution of species. Niamir et al. (2012) created an expert based deductive model using expert based information on habitat of species. Random points for presence of the species were generated from the expert based map which was then used as input to MAXENT species distribution model along with other predictor variables related to climate and land use. By doing so expert based information as well as information on climate and land use were captured in the final model with good validation scores. However, this approach has been tested to one species with accurate sampled validation data. Expanding this approach using the updated AOH maps at a global scale will be an interesting avenue to explore.

The updated AOH maps produced in this thesis will be freely accessible after publication. The users will have the option to choose among AOH_T, AOH_L and the intersected maps. The different validation metrics can be used by the end users to choose the best AOH map for their species of interest using Appendix 3.A1 and Appendix 4.A1.

5.1 References

Bird J. P, Buchanan G. M, Lees A. C, et al. (2012). Integrating spatially explicit habitat projections into extinction risk assessments: a reassessment of Amazonian avifauna incorporating projected deforestation. *Diversity and Distributions* 18:273–281.

Bivand R (2019). rgrass7: Interface Between GRASS 7 Geographical Information System and R. R package version 0.2-1. https://CRAN.R-project.org/package=rgrass7

Brook B, O'Grady J, Chapman A, et al. (2000). Predictive accuracy of population viability analysis in conservation biology. *Nature* 404, 385–387. https://doi.org/10.1038/35006050

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 & Evolution*, 34, (11), 977–986.

Buchanan G. M, Butchart S. H. M and Dutson G (2008). Using remote sensing to inform conservation status assessment: Estimates of recent deforestation rates on New Britain and the impacts upon endemic birds. *Biological Conservation* 141:56–66.

Buchhorn M, Smets B, Bertels L, et al. (2020). *Copernicus Global Land Service: land-cover 100m: Version 3 Globe 2015–2019: Product User Manual*. Geneve: Zenodo.

Bunce R. G. H, Bogers M. M. B, Evans D, et al. (2012). The significance of habitats as indicators of biodiversity and their links to species. *Ecological Indicators*, 33, 19–25

Convention on Biological Diversity (2022). *REPORT OF THE OPEN-ENDED WORKING GROUP ON THE POST-2020 GLOBAL BIODIVERSITY FRAMEWORK ON ITS SECOND MEETING*. Rome.

Chamberlain S (2020). rredlist: 'IUCN' Red List Client. R package version 0.6.0. https://CRAN.R-project.org/package=rredlist

Chamberlain S, Barve V, Mcglinn D, et al. (2022). rgbif: Interface to the Global Biodiversity Information Facility API_. R package version 2.1.0, <URL: https://CRAN.R-project.org/package=rgbif>.

Gorelick N, Hancher M, Dixon M, et al. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone.

IUCN (International Union for Conservation of Nature). (2017). Guidelines for using the IUCN Red List categories and criteria. Version 13. IUCN, Gland, Switzerland. Available from http://www.iucnredlist.org/documents/RedListGuidelines.pdf.

Jung M, Dahal P. R, Butchart S. H. M, et al. (2020). A global map of terrestrial habitat types. *Scientific Data*, 7, 1–8.

Jetz W, McPherson J. M, and Guralnick R. P (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology and Evolution*, 27, 151-159. DOI:10.1016/j.tree.2011.09.007

Karra K, Kontgis C, Statman-Weil Z, et al. (2021). *Global land use/land-cover with Sentinel-2 and deep learning.* IGARSS 2021-2021 IEEE International Geoscience and Remote Sensing Symposium.

Kuhn, M (2020). Caret: Classification and regression training. (R Project, 2020).

Lumbierres M, Dahal P. R, Marco M. Di, et al. (2021). Translating habitat class to land-cover to map area of habitat of terrestrial vertebrates. *Conservation Biology*, 1–11. https://doi.org/10.1111/cobi.13851

Niamir A, Skidmore A. K, Toxopeus A. G, et al. (2011). Finessing atlas data for species distribution models. *Diversity and Distributions*, 17: 1173-1185. https://doi.org/10.1111/j.1472-4642.2011.00793.x

Phillips S. J, Anderson R. P, and Schapire R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.

Rondinini C, Di Marco M, Chiozza F, et al. (2011). Global habitat suitability models of terrestrial mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 2633–2641.

Rondinini C and Boitani L (2012). Mind the map: trips and pitfalls in making and reading maps of carnivore distributon. *In: Boitani, L. and Powell, R.A. (eds.) Carnivore ecology and conservation: a handbook of techniques.* Oxford University Press.

Rondinini C, Di Marco M, Visconti P, et al. (2014). Update or outdate: long-term viability of the IUCN Red List. *Conservation Letters* 7:126–130.

Santini L, Butchart S. H. M, Rondinini C, et al. (2019). Applying habitat and population-density models to land-cover time series to inform IUCN Red List assessments. *Conservation Biology*, 33: 1084-1093. https://doi.org/10.1111/cobi.13279

Tracewski Ł., Butchart S. H, Di Marco M, et al. (2016). Toward quantification of the impact of 21st century deforestation on the extinction risk of terrestrial vertebrates. *Conservation Biology*, *30*(5), pp.1070-1079.

Thuiller W, Lafourcade B, Engler R, et al. (2009). BIOMOD a platform for ensemble forecasting of species distributions, *Ecography* 32: 369373, doi: 10.1111/j.1600-0587.2008.05742.x

UN General Assembly. (2015). *Transforming our world : the 2030 Agenda for Sustainable Development*, A/RES/70/1, available at: https://www.refworld.org/docid/57b6e3e44.html [accessed 24 January 2022]

Visconti P, Bakkenes M, Baisero D, et al. (2016). Projecting Global Biodiversity Indicators under Future Development Scenarios, *Conservation Letters*, 9: 5-13. https://doi.org/10.1111/conl.12159

Appendices and supporting information

6.1 Appendices and supporting information for chapter 2

Appendix 2.A1: Summary table with summary rules to map the different level-2 habitat types. (https://doi.org/10.1038/s41597-020-00599-8.)

6.2 Appendices and supporting information for chapter 3

Appendix 3.A1: Metadata table with summary of validation analyses.
(http://doi.org/10.5281/zenodo.5109073)

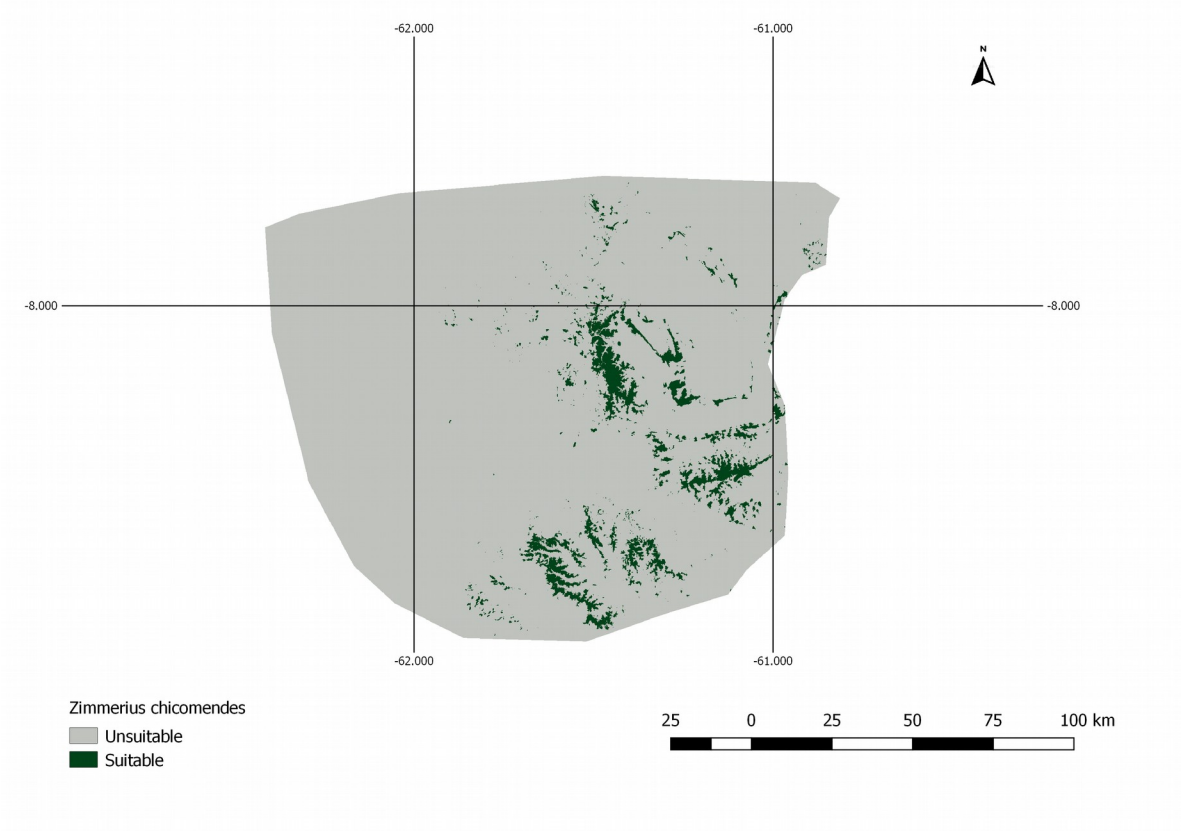

Figure 3.S1: AOH map for species *Zimmerius chicomendesi*.

The species is coded against “Forest” and “Shrubland” habitats and the elevation range falls inside the IUCN range. However, the land-cover inside this range map includes a high proportion of “Herbaceous cover” land-cover type which is not associated with “Shrubland” habitat in the habitat – land-cover association table. Therefore, the model prevalence of this AOH is much lower than expected.

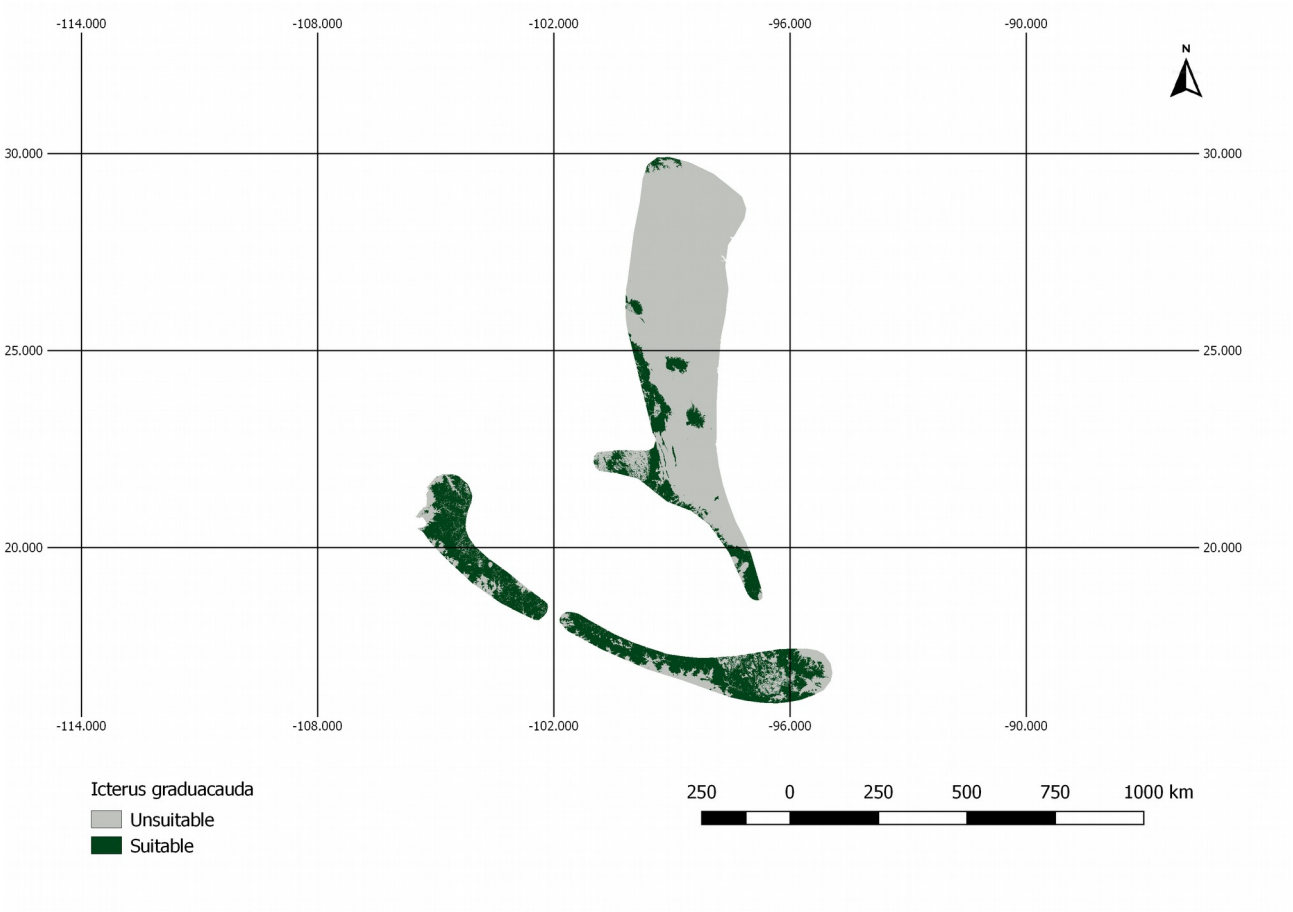

Figure 3.S2: AOH map for the species *Icterus graduacauda.*

The IUCN range of the species doesn't cover much of the elevation range. Therefore, the model prevalence of this species is lower than estimated.

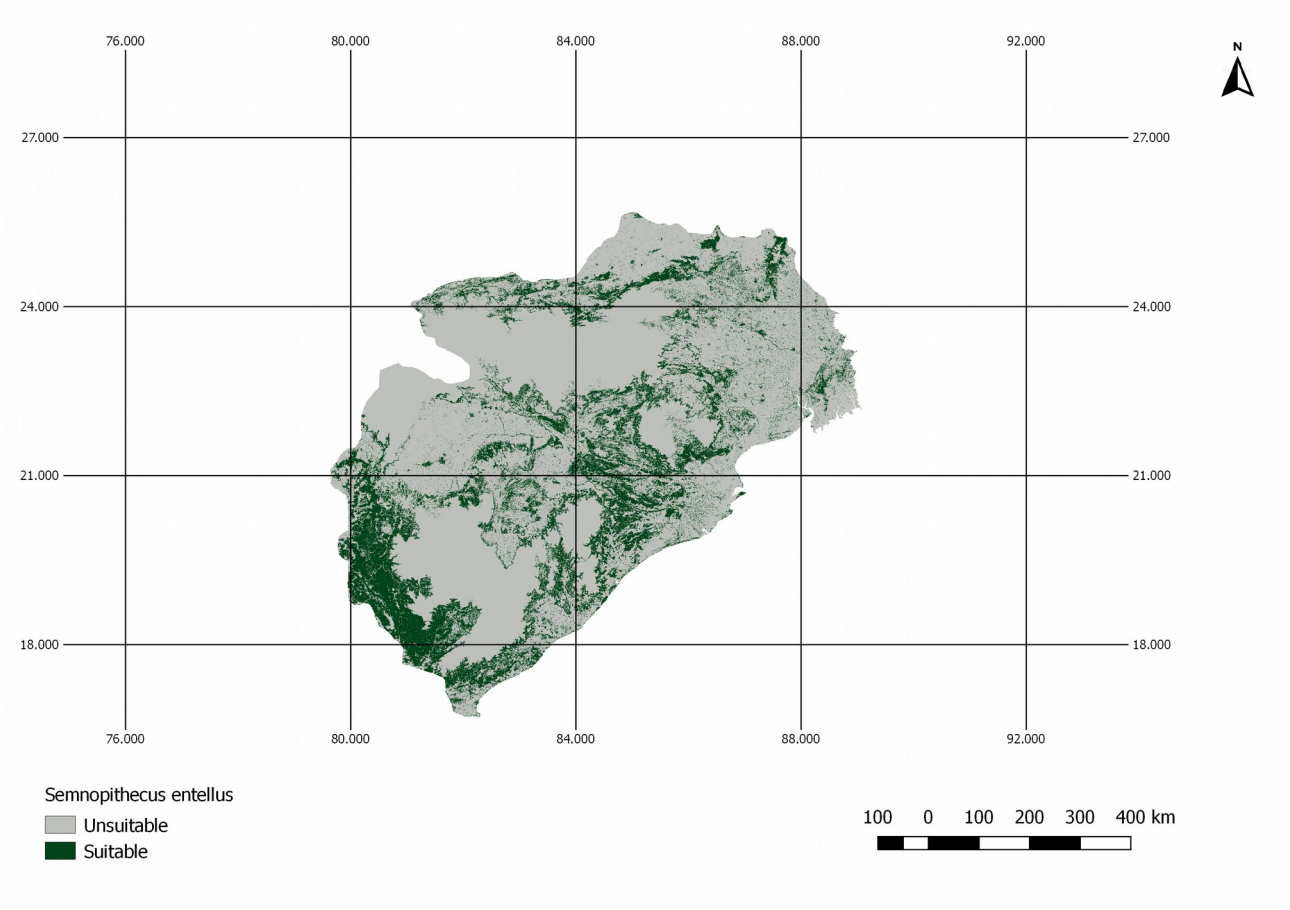

Figure 3.S3: AOH map for the species *Semnopithecus entellus*.

There is a large proportion of land-cover class “Cropland” inside the range map of this species. However, this species is not coded to habitats that are associated with the land-cover “Cropland”. Therefore, the model prevalence is lower than estimated.

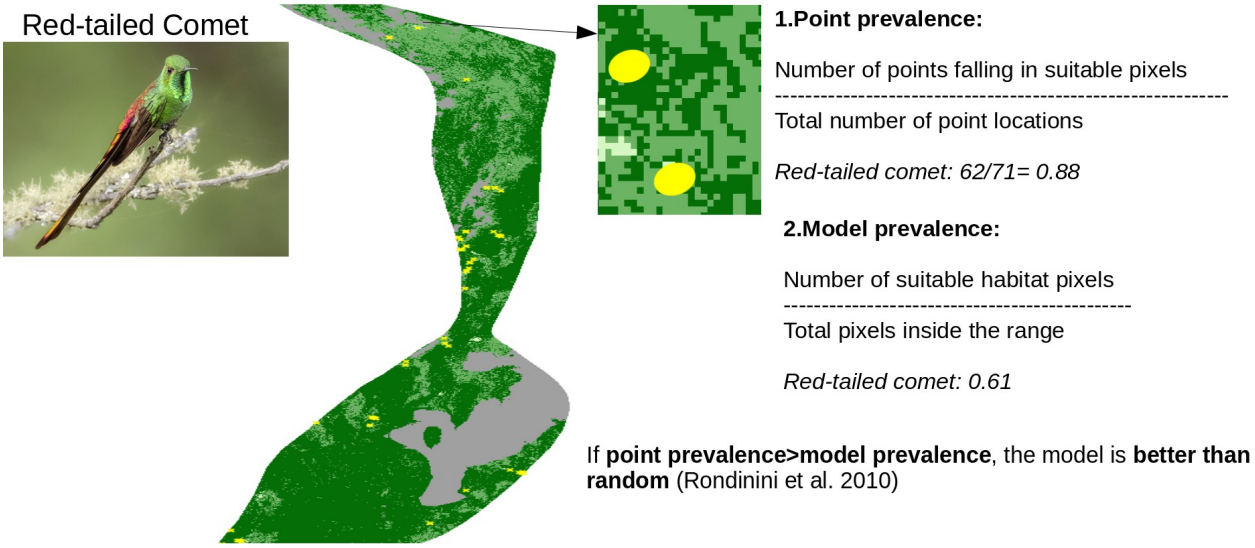

Figure 3.S4: Point validation of the AOH maps using model and point prevalence.

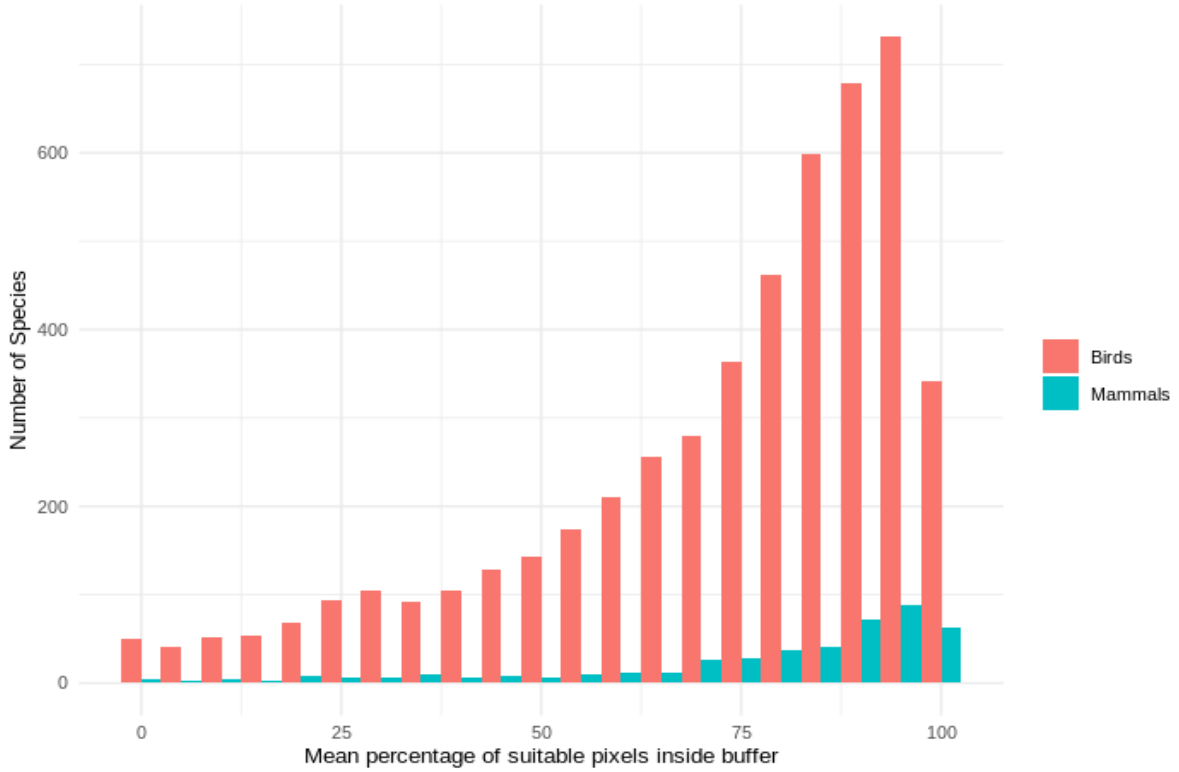

Figure 3.S5: Histogram of mean percentage of suitable AOH pixels inside the 300 m buffer for mammals and birds species used in point validation.

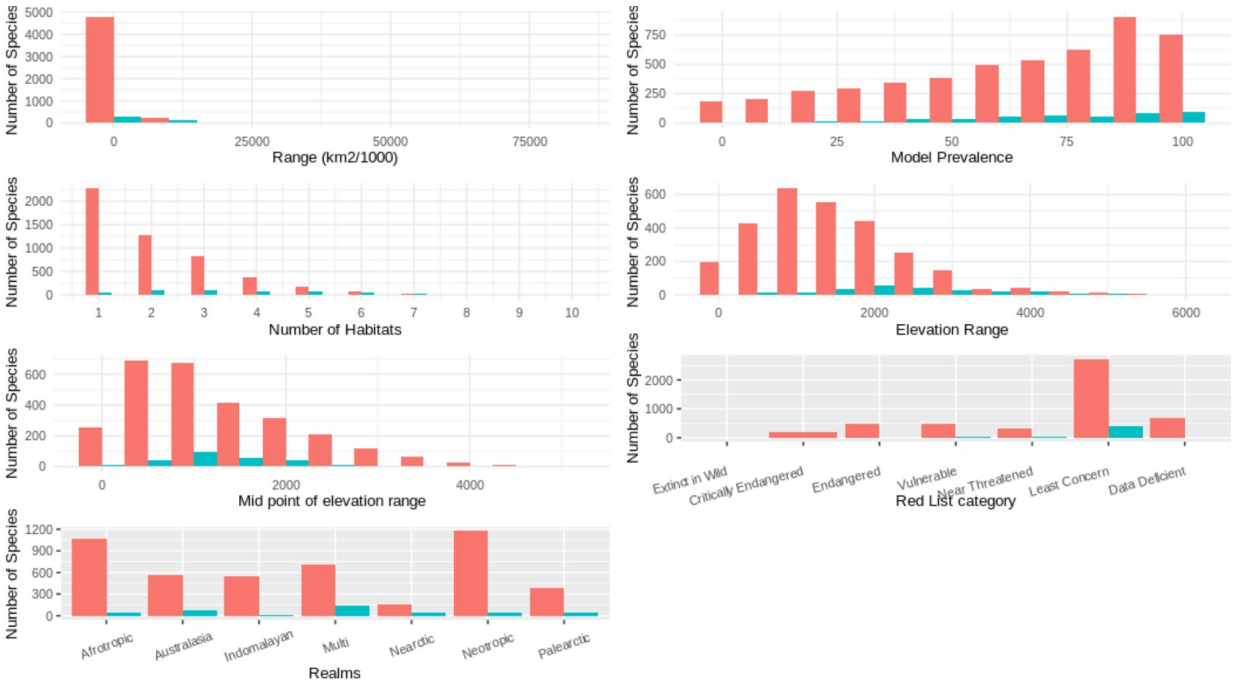

Figure 3.S6: Comparison of species with and without validation points for mammals.

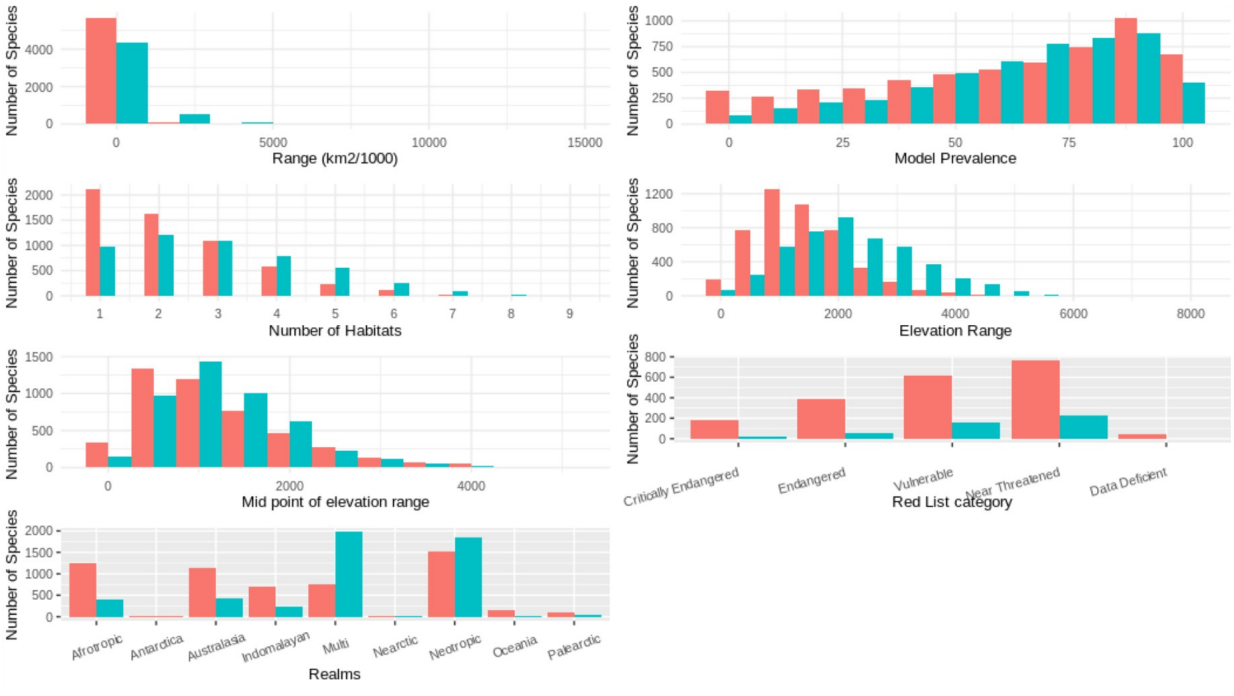

Figure 3.S7: Comparison of species with and without validation points for birds.

6.3 Appendices and supporting information for chapter 4

Appendix 4.A1: Metadata table with validation metrics for AOH_T and AOH_L. (https://doi.org/10.5281/zenodo.5918224)

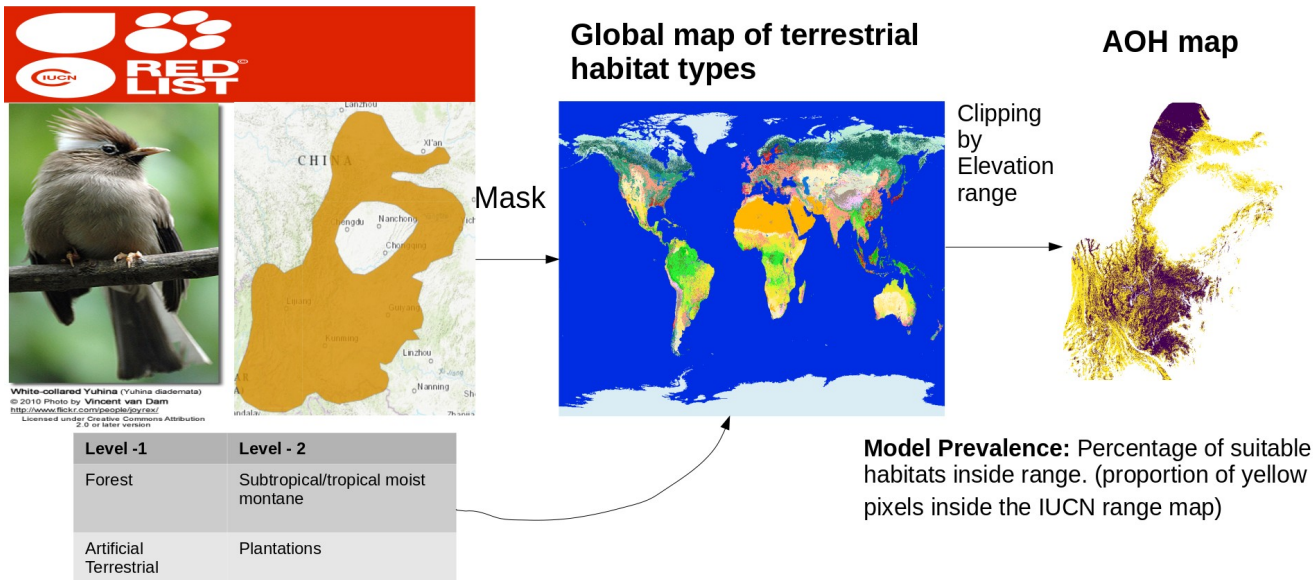

Figure 4.S1: GIS workflow to produce the Area of Habitat maps using global map of terrestrial habitat types.