

Translating habitat class to land cover to map area of habitat of terrestrial vertebrates

Maria Lumbierres^{1,2}  | Prabhat Raj Dahal^{1,2}  | Moreno Di Marco³ |
Stuart H. M. Butchart^{3,4}  | Paul F. Donald^{3,4}  | Carlo Rondinini¹ 

¹ Global Mammal Assessment Program, Department of Biology and Biotechnologies, Sapienza University of Rome, Rome, Italy

² BirdLife International, Cambridge, UK

³ Department of Biology and Biotechnologies, Sapienza University of Rome, Rome, Italy

⁴ Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, UK

Correspondence

Maria Lumbierres, Global Mammal Assessment Program, Department of Biology and Biotechnologies, Sapienza University of Rome, Viale dell'Università 32, 00185 Rome, Italy.
Email: maria.lumbierrescivit@uniroma1.it

Article Impact Statement: Point-locality data can be used to translate IUCN habitat classes to land cover to produce area of habitat maps.

Abstract

Area of habitat (AOH) is defined as the “habitat available to a species, that is, habitat within its range” and is calculated by subtracting areas of unsuitable land cover and elevation from the range. The International Union for the Conservation of Nature (IUCN) Habitats Classification Scheme provides information on species habitat associations, and typically unvalidated expert opinion is used to match habitat to land-cover classes, which generates a source of uncertainty in AOH maps. We developed a data-driven method to translate IUCN habitat classes to land cover based on point locality data for 6986 species of terrestrial mammals, birds, amphibians, and reptiles. We extracted the land-cover class at each point locality and matched it to the IUCN habitat class or classes assigned to each species occurring there. Then, we modeled each land-cover class as a function of IUCN habitat with (SSG, using) logistic regression models. The resulting odds ratios were used to assess the strength of the association between each habitat and land-cover class. We then compared the performance of our data-driven model with those from a published translation table based on expert knowledge. We calculated the association between habitat classes and land-cover classes as a continuous variable, but to map AOH as binary presence or absence, it was necessary to apply a threshold of association. This threshold can be chosen by the user according to the required balance between omission and commission errors. Some habitats (e.g., forest and desert) were assigned to land-cover classes with more confidence than others (e.g., wetlands and artificial). The data-driven translation model and expert knowledge performed equally well, but the model provided greater standardization, objectivity, and repeatability. Furthermore, our approach allowed greater flexibility in the use of the results and uncertainty to be quantified. Our model can be modified for regional examinations and different taxonomic groups.

KEYWORDS

commission and omission errors, Copernicus Global Land Service Land Cover (CGLS-LC100), ESA Climate Change Initiative (ESA-CCI), IUCN Habitat Classification Scheme, IUCN Red List, habitat suitability models

Conversión de la Categoría de Hábitat a Cobertura de Terreno para Mapear el Área de Hábitat de los Vertebrados Terrestres

Resumen: El área del hábitat (AOH) está definida como “el hábitat disponible para una especie, es decir, el hábitat dentro del área de distribución de la especie” y se calcula mediante la sustracción de las áreas de terreno inadecuado y la elevación del área de distribución. El Esquema de Clasificación de Hábitats de la Unión Internacional para la Conservación de la Naturaleza proporciona información sobre las asociaciones entre los hábitats de las especies y con frecuencia se utilizan las opiniones no validadas de expertos para cotejar el hábitat con los tipos de cobertura de terreno, lo que genera una fuente de incertidumbre

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by-nc/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Conservation Biology* published by Wiley Periodicals LLC on behalf of Society for Conservation Biology.

en los mapas de AOH. Desarrollamos un método orientado por datos para convertir las categorías de hábitat que maneja la UICN en cobertura de terreno basado en los datos de localidad puntual de 6,986 especies de mamíferos terrestres, aves, anfibios y reptiles. Extrajimos la categoría de cobertura de terreno en cada localidad puntual y la cotejamos con la categoría o categorías de hábitat de UICN asignada a cada especie incidente en la localidad. Después modelamos cada categoría de cobertura de terreno como función del hábitat según la UICN usando modelos de regresión logística. Las proporciones de probabilidad resultantes fueron usadas para evaluar la solidez de la asociación entre cada categoría de hábitat y de cobertura de terreno. Después comparamos el desempeño de nuestro modelo orientado por datos con el desempeño de una tabla de conversión publicada basada en el conocimiento de expertos. Calculamos la asociación entre las categorías de hábitat y las de cobertura de terreno como una variable continua, pero para mapear el AOH como una presencia o ausencia binaria, fue necesario aplicar un umbral de asociación. Este umbral puede ser elegido por el usuario de acuerdo con el balance requerido entre los errores de omisión y comisión. Algunos hábitats (p. ej.: bosques y desiertos) fueron asignados a las categorías de cobertura de terreno con más confianza que otros (p. ej.: humedales y artificiales). El modelo de conversión orientado por los datos y el conocimiento de los expertos tuvieron un desempeño igual de eficiente, pero el modelo proporcionó una mayor estandarización, objetividad y repetitividad. Además, nuestra estrategia permitió una mayor flexibilidad en el uso de los resultados y de la incertidumbre para ser cuantificados. Nuestro modelo puede modificarse para análisis regionales y para diferentes grupos taxonómicos.

PALABRAS CLAVE

errores de comisión y omisión, Copernicus Global Land Service Land Cover (CGLS-LC100), Esquema de Clasificación de Hábitats de la UICN, Iniciativa de Cambio Climático ESA (ESA-CCI), Lista Roja de la UICN, modelos de idoneidad de hábitat

【摘要】

Brooks 等人 (2019) 将栖息地范围定义为“一个物种可利用的栖息地, 即其分布范围内的栖息地”, 其计算方法是从物种分布范围中减去土地覆盖和海拔不适宜的区域。《世界自然保护联盟 (IUCN) 栖息地分类方案》提供了物种栖息地关联的信息, 但它大多使用未经验证的专家意见来匹配栖息地与土地覆盖类型, 这也是栖息地地图不确定性的来源之一。本研究开发了一种数据驱动的方法, 基于 6986 种陆生哺乳动物、鸟类、两栖动物和爬行动物的点位数据, 将 IUCN 栖息地类型转换为土地覆盖。我们提取了每个点位的土地覆盖类型, 并将其与 IUCN 栖息地类型或该点位出现物种的栖息地类型相匹配。接下来, 我们用逻辑回归模型建立了土地覆盖类型与 IUCN 栖息地类型的函数, 并用得到的比值比来评估每个栖息地类型与土地覆盖类型之间的关联程度。我们还比较了这个数据驱动的模式与已发表的基于专家知识的转换表的表现情况。我们将栖息地类型和土地覆盖类型之间的关联作为连续变量来计算, 但在绘制栖息地地图时为了将其作为存在或不存在的二元变量来处理, 需要采用一个关联程度的阈值。这个阈值可以由使用者根据所需的漏分与错分误差平衡来确定。一些栖息地 (如森林和沙漠) 相比于其它栖息地 (如湿地和人工栖息地) 有更大的把握被分配到对应的土地覆盖类型。我们还发现, 该数据驱动的转换模型与基于专家知识的转换表的表现相当, 但模型具有更好的标准化、客观性和可重复性。此外, 我们的方法在结果应用和不确定性量化上有更为灵活; 该模型还可以针对区域性分析或不同生物类群的研究进行修改。【翻译: 胡怡思; 审校: 聂永刚】

关键词: 错分与漏分误差, 哥白尼全球土地服务土地覆盖 (CGLS-LC100), 欧洲航天局气候变化倡议 (ESA-CCI), 《IUCN 栖息地分类方案》, 《IUCN 红色名录》, 栖息地适宜性模型

INTRODUCTION

Because habitat loss is the most important driver of biodiversity decline (Díaz et al., 2019), there is an urgent need to

determine where habitat is located within each species' distribution (Brooks et al., 2019; Pimm et al., 2014). Several approaches have been developed to map global species' distributions, but accurate spatial data are only available for a limited

number of species (Rondinini et al., 2005; Rondinini & Boitani, 2012).

The most complete data set of maps of species' ranges is that available in the International Union for Conservation of Nature (IUCN) Red List (www.iucnredlist.org). The IUCN Red List has assessed comprehensively more than 134,400 species and species groups, including mammals, amphibians, and birds. The IUCN range maps are generally drawn to minimize errors of omission (i.e., false absence), with the result that they often contain substantial areas that are not occupied by the species and so contain errors of commission (i.e., false presence) (Ficetola et al., 2014; Di Marco et al., 2017).

Area of habitat (AOH) (previously known as extent of suitable habitat, or ESH) is the "habitat available to a species, that is, habitat within its range" (Brooks et al., 2019). Maps of AOH are produced by subtracting unsuitable areas from range maps based on data on each species' associations with land cover and elevation (Beresford et al., 2011; Rondinini et al., 2011; Ficetola et al., 2015), the aim of which is to reduce commission errors in range maps. Therefore, the production of AOH maps requires an understanding of a species' habitat and where such areas are within its range.

Information on habitat preferences is documented for each species assessed on the IUCN Red List (IUCN, 2013) following the IUCN Habitats Classification Scheme (IUCN habitat) (IUCN, 2012), a classification and coding system of habitats that ensures global consistency. The habitats are defined independently of taxonomy or geography. However, IUCN habitat classes are not spatially explicit, although attempts have been made to delimit them (Jung et al., 2020). Land-cover classes derived from remote sensing have been used widely as a surrogate of habitat (e.g., Buchanan et al., 2008; Beresford et al., 2011; Rondinini et al., 2011; Tomaselli et al., 2013; Montesino Pouzols et al., 2014; Corbane et al., 2015; Santini et al., 2019), although habitat is a complex multidimensional concept that is difficult to simplify into land-cover classes.

A table that translates habitat into land-cover classes is typically used to represent IUCN habitat classes spatially and to produce AOH maps. Such tables have been based solely on expert knowledge, raising concerns about the accuracy and objectivity of the resulting associations because the assumptions generated in the translation process are rarely considered in detail and the errors are difficult or impossible to quantify (Bradley et al., 2012). Furthermore, there is a lack of standardization and repeatability in the procedure (Seoane et al., 2005), which is subject to variability in expert opinion (Johnson & Gillingham, 2004).

Repositories of point locality data (i.e., locational records in which particular species have been recorded [Rondinini et al., 2006]) primarily from citizen science have been used successfully in habitat suitability models (e.g., Gueta & Carmel, 2016; Bradter et al., 2018; Crawford et al., 2020). The potential, therefore, exists to use such data to develop an objective data-driven table that can be used to translate habitat into land-cover classes by extracting information on land cover from point localities of species with different habitat associations.

We sought to devise an objective, transparent, repeatable, and data-driven method to produce a table that can be used to translate IUCN habitat classes into land cover based on two widely used global land-cover maps, the Copernicus Global Land Service Land Cover (CGLS-LC100) (Buchhorn et al., 2020; Buchhorn et al., 2019) and the European Space Agency Climate Change Initiative land cover 2015 (ESA-CCI) (ESA, 2017) and point-locality data for mammals, birds, amphibians, and reptiles (the best documented groups of species). The aim of this analysis was to develop a translation table that quantifies the power of association between land cover and habitat classes. In doing so, we aimed to illustrate a method that improves on expert opinion by quantifying errors in associations between habitat and land-cover classes and being flexible to the needs of the user in terms of the required trade-off between reducing commission errors and increasing omission errors and that can be developed at different spatial scales, for different taxa, based on any set of habitat or land-cover classes.

METHODS

Data cleaning and preparation

We downloaded point-locality data for mammals (GBIF, 2019; GBIF, 2020), amphibians (GBIF, 2020), and reptiles (GBIF, 2020) from the Global Biodiversity Information Facility (GBIF) and for birds from GBIF (GBIF, 2019; GBIF, 2020) and eBird (eBird Basic Dataset, 2019). The data were restricted to point localities dated from January 2005 to December 2018 for the model building (70% training and 30% test) and from January 2019 to December 2020 for the evaluation of the model. For eBird data, we selected only stationary point localities with a coordinate uncertainty of <30 m. To minimize errors and uncertainties inherent to repositories of point locality data, we included only the most precisely georeferenced points (Rondinini et al., 2006; Meyer, 2012) and applied a set of filters following the guidelines of Boitani et al. (2011). The main attributes considered were currentness, spatial accuracy, and spatial coverage (Figure 1). To make it clear where we are referring to explicit classes, we present land-cover class names in quotation marks and IUCN habitat class names in italics.

The habitat class or classes association for each species was extracted from IUCN (2020). The IUCN habitat classes are standardized terms describing the major habitat types in which taxa occur globally. They follow a hierarchical classification of habitat with three levels. The definitions consider land cover, biogeography, latitudinal zonation, and in marine systems, depth. We used level-1 habitat classes for all habitats except for artificial terrestrial, for which we used a modification of level 2 (Appendix S1). We subdivided artificial terrestrial into three subclasses because in terms of land cover these are distinct habitat classes that could aggregate different species (Ducatez et al., 2018).

Because the land-cover classes from the two remote sensing products are exclusively terrestrial, we limited the analysis to species coded only to terrestrial habitat classes, thus

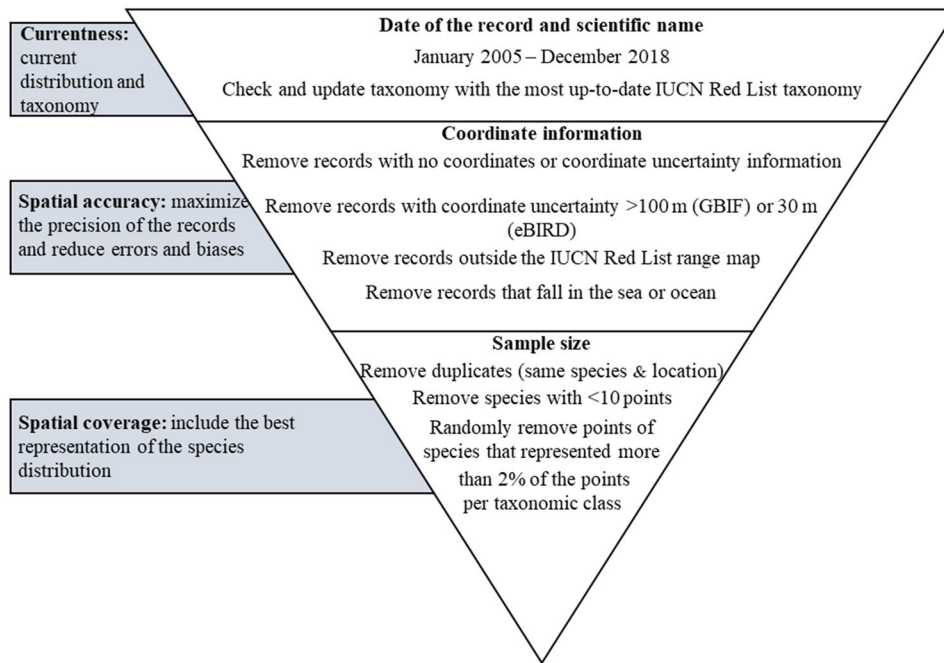


FIGURE 1 Description of the repository point-locality cleaning process, following Boitani et al. (2011). The factors considered are currentness, spatial accuracy, and spatial coverage and are applied from top to bottom (GBIF, 2019; GBIF, 2020; eBird Basic Dataset, 2019)

excluding species coded to one or more IUCN marine habitats. We also excluded species coded to more than five level-1 habitat classes because habitat generalists are likely to add little information to the habitat–land cover relationship. In contrast, specialist species coded to only one habitat class provide more insight into the relationship between habitat and land cover. For that reason, for each taxonomic class, we randomly subsampled point records from species coded to more than one habitat class to match the number of points of species coded to one habitat and thereby gave equal weight to habitat specialists even when they had fewer points.

The two land-cover products used in the analysis have different characteristics. The CGLS-LC100 has a 100-m spatial resolution and a global classification accuracy of 80.2% (Tsendbazar et al., 2020). The ESA-CCI has a 300-m spatial resolution and a global classification accuracy of 71.1% (ESA, 2017). It is part of a time series from 1992 to 2015, of which we used the 2015 map. Both products use the United Nations Food and Agriculture Organization Land Cover Classification System (UN FAO-LCCS; Di Gregorio & Jansen, 2000), although they have different legends. The CGLS-LC100 has 12 land-cover classes at level 1 and 23 classes at level 3 (level 2 is not used by CLGS-LC100); we used level 3. The ESA-CCI has 22 land-cover classes at level 1 and 38 classes at level 2. We used only level 1 because level 2 is only available for some regions of the globe.

To prepare the data for the model, we extracted the land-cover class at the coordinates of each point locality. Some land-cover classes did not have enough point localities within them to be modeled, although in all cases these were land-cover classes with very low global coverage. For CGLS-LC100, the underrepresented land-cover classes were “open forest deciduous needle

leaf” (10 points, 0.03% of global land surface), “snow and ice” (108 points, 3.1% of global land surface), “moss and lichen” (124 points, 2.3% of global land surface), and “closed forest deciduous needle leaf” (383 points, 3.0% of global land surface). For ESA-CCI, the only class represented too infrequently for analysis was “lichens and mosses” (713 points, 2.2% of global land surface).

Modeling of habitat–land cover associations

To quantify the relationship between IUCN habitat classes and land-cover classes, we modeled the presence or absence of each land-cover class as a function of the IUCN habitat class or classes of the species whose point localities fell within it (Figure 2). An important consideration for modeling was that the number of habitat classes per species varied from 1 to 5. Therefore, it was impossible to model land-cover class as a one-to-one relationship with habitat class because each point location was associated with one or multiple habitat classes. This consideration restricted the number of models we could use for our analysis. We required a flexible model that allowed a many-to-many match between habitat classes and land-cover classes to model this matrix of habitat versus land-cover class relations. In multinomial logistic regression models, the data and the computational power requirements increase exponentially with the number of response categories. In our case, with more than 20 land-cover categories, this option was not feasible. Therefore, we modeled each land-cover class separately, transforming each class into a binary variable of 1 (land cover present) or 0 (land cover not present). Then, we used logistic regressions to model

IUCN habitat class \ Land-cover class	Number point	Forest	Savanna	Shrubland	Grassland	Wetlands	Rocky areas	Desert	Artificial arable and pasture lands	Artificial degraded forest and plantation	Artificial urban areas and rural gardens	Artificial aquatic	AUC
Shrubs	2,64,166	0.309	1.870	2.683	1.188	-	1.383	4.225	-	0.711	-	-	0.882
Herbaceous vegetation	5,31,007	0.372	1.180	1.622	2.369	1.204	1.506	1.517	1.296	0.880	1.262	-	0.793
Cultivated and managed vegetation agriculture	4,70,123	0.588	1.570	1.395	1.748	1.875	0.687	-	1.743	-	1.330	1.523	0.807
Urban / built up	4,12,978	-	-	1.362	-	1.351	-	-	1.488	1.293	3.183	1.696	0.763
Bare / sparse vegetation	30,746	0.139	-	2.026	1.524	-	3.380	8.192	-	0.489	-	-	0.924
Permanent water bodies	1,12,799	0.603	-	-	-	2.189	-	0.698	1.236	-	1.447	1.712	0.745
Herbaceous wetland	8,70,084	0.732	1.240	-	1.248	3.185	0.631	0.367	1.215	1.220	1.396	2.360	0.827
Closed forest, evergreen needle leaf	4,15,369	3.824	0.384	-	0.706	-	-	0.229	0.592	0.530	-	0.674	0.885
Closed forest, deciduous needle leaf	6,30,019	13.720	0.455	0.382	0.475	-	0.483	0.052	0.735	1.516	0.748	-	0.940
Closed forest, deciduous broad leaf	3,11,541	2.520	1.704	-	-	1.560	-	0.174	-	-	1.435	-	0.867
Closed forest, mixed	60,555	5.461	0.548	0.671	-	1.523	-	-	-	-	-	-	0.906
Closed forest, unknown	1,28,975	1.971	-	-	-	1.312	-	0.389	-	1.264	-	1.345	0.736
Open forest, evergreen needle leaf	1,08,686	2.171	0.476	-	-	-	1.790	0.578	-	0.454	-	-	0.856
Open forest, evergreen broad leaf	49,487	2.442	0.801	-	-	-	0.562	0.160	-	1.566	-	-	0.894
Open forest, deciduous broad leaf	1,03,555	1.391	2.172	-	-	1.658	-	0.251	-	-	1.398	-	0.854
Open forest, mixed	2,758	3.038	-	-	-	-	-	-	-	-	-	-	0.899
Open forest, unknown	7,73,407	1.138	1.368	1.284	-	1.302	-	0.673	1.221	1.138	1.167	1.341	0.644

Positive association (odds ratio = 1.138–1.351)

Positive association (odds ratio = 1.362–1.712)

Positive association (odds ratio = 1.743–13.720)

Negative association

Nonsignificant association

FIGURE 2 Odds ratio values describing the association between Copernicus Global Land Service Land Cover (CGLS-LC100) classes and International Union for the Conservation of Nature (IUCN) Habitat Classification Scheme (AUC, area under the curve from a receiver operating characteristic [ROC] curve, a measure of accuracy of a classification mode). Odds ratio values <1 indicate a negative association, and values >1 indicate a positive association. The positive associations are divided into tertiles (green), indicating three possible options for setting a threshold to convert continuous variables into a binary association-nonassociation variable for creating area of habitat maps

the binary land-cover class variable as a function of the different habitat classes:

$$\log \frac{p_{lc}}{1-p_{lc}} = \beta_0 + \beta_1 H_{forest} + \beta_2 H_{savanna} + \beta_3 H_{shrubland} + \beta_4 H_{grassland} + \beta_5 H_{wetlands} + \beta_6 H_{rocky\ areas} + \beta_7 H_{desert} + \beta_8 H_{artificial1} + \beta_9 H_{artificial2} + \beta_{10} H_{artificial3} + \beta_{11} H_{artificial4} + \epsilon, \tag{1}$$

where $(p_{lc}/[1-p_{lc}])$ is the land-cover odds ratio, β_x is the model parameter for the IUCN habitats H_x , and ϵ is the error.

The transformation of the land-cover class into a binary form for each of the models generated a highly unbalanced variable, with many more zeroes than ones. In a logistic regression model, unbalanced data underestimate the probability of an event, so it is recommended that the number of 1s and 0s be adjusted (King & Zeng, 2001; Pozzolo et al., 2015). We, therefore, randomly subsampled the 0s in the training set before running the model. The assumption behind this is that in the majority class, there are many redundant observations and randomly removing some of them does not change the estimation of the within-class distribution (Pozzolo et al., 2015).

To reduce the intrinsic spatial and taxonomic bias of point-locality data (Boitani et al., 2011; Meyer et al., 2016) and to

account for multiple but varying numbers of point localities per species, we added taxonomic and spatial variables as random effects in the model (Bird et al., 2014). As taxonomic variables, we added species nested within taxonomic class (Amphibia, Reptilia, Aves, and Mammalia). Adding intermediate taxonomic groupings (e.g., family or genus) in the nesting would result in many factor levels with single or very few replicates. To test whether there was any bias among taxonomic classes, we produced separate models for each class. This test showed that the association between land-cover and habitat classes from the different translation tables was very similar; therefore, we decided to model all classes together. As a spatial variable, we added the country of the point record as a random effect.

We used the coefficients of the models to evaluate the association between land-cover class and habitat classes. The intercept did not provide any information on the relationship between land-cover class and habitat class because it represented the odds of a point locality falling within a particular land-cover class after the subsampling of the data set, independently of the habitat (Ranganathan et al., 2017). The coefficients represented the odds ratio, in other words, the odds of a point locality falling within a particular land-cover class (when the species to

which the point locality relates is coded for a particular habitat class) divided by the odds of the species occurring in that land-cover class when it is not coded for that habitat class. The ratio, therefore, indicates the extent to which being coded to a particular habitat class increases or decreases the odds of a species being found in a particular land-cover class. The units of the logit function are $\log(\text{odds ratio})$, but for easier interpretation, we changed them to the exponential and present the results as odds ratios.

Odds ratio values below 1 indicate a negative association between land cover and habitat classes, whereas those above 1 indicate a positive association. Because the odds ratio is a continuous variable, it is necessary to set a threshold to transform the results into a binary translation table that can be used to assign, or not, a particular habitat class to a particular land-cover class. The threshold can be modified according to the needs of the user based on the required balance between minimizing commission errors (land-cover classes incorrectly associated with a habitat class) and increasing omission errors (land-cover classes incorrectly omitted from a habitat class). Coefficients that had $p > 0.05$ were considered to indicate a lack of association between land cover and habitat classes. To adjust the significance threshold of the p values for multivariable analysis, we used Bonferroni corrections.

To validate the models, we set aside 30% of the point occurrence data for testing, leaving 70% to train the model. As a validation test, we used the area under the curve (AUC) from a receiver operating characteristic curve (Jiménez-Valverde, 2012). The AUC is a model accuracy measure that provides information on how well a model can distinguish among classes. In our case, we used it to test how well the models predicted the presence or absence of a point locally in a given land-cover class. The AUC values range from 0 to 1; a value of 0.5 meant the model did not perform better than random, whereas a value of 1 indicated the model perfectly separated the two groups.

The results of the models can also be mapped spatially with one of the three thresholds of associations between habitat and land-cover classes. In such maps, habitats are overlaid because the same land-cover class may represent more than one habitat class or because both habitats occur in the same geographical areas. The overlap among habitats increases as the threshold of association is reduced.

We then compared the performance of the data-driven translation table with that of an expert-knowledge translation table (Santini et al., 2019) based on the same ESA-CCI land-cover classification used here. We did not find any published translation table that used CGLS-LC100. Santini et al. (2019) compared the ESA CCI land-cover classes with level-2 IUCN habitat classes, so we aggregated the habitat classes to level-1 IUCN habitat classes to make the two translation tables comparable. We limited the comparison to birds and mammals because they were the taxonomic groups considered by Santini et al. (2019) For each species, we mapped habitat based on both tables. We assessed the proportion of points located in the suitable areas (point prevalence) and compared it with the proportion of habitat inside the species' range (model prevalence) to determine whether the results were better than a randomly

assigned set of points (Rondinini et al., 2011). We used 211,304 point localities for 489 species of mammal and 461,277 point localities for 2112 species of bird.

RESULTS

The number of point localities and species available for analysis was 200,683 and 455, respectively, for mammals, 4,083,510 and 5154 for birds, 92,327 and 479 for amphibians, and 131,077 and 898 for reptiles. For the CGLS-LC100 land-cover product, 71 coefficients showed a significantly positive association (odds ratio > 1) and 38 coefficients showed a significantly negative association (odds ratio < 1) between land-cover classes and habitat classes (Figure 2). For the ESA-CCI land-cover product, 101 coefficients showed a significantly positive association, and 40 coefficients showed a significantly negative association (Figure 3).

Higher odds ratios (> 1) indicated stronger positive associations between land-cover and habitat classes, and lower odds ratios (nearer to zero) indicated stronger negative associations. We divided the significantly positive values into tertiles to identify three potential thresholds for creating a table of binary association and nonassociation variables for producing AOH maps: 1.138–1.351, 1.362–1.712, and 1.743–13.720 for CGLS-LC100, and 1.121–1.393, 1.396–1.704, and 1.708–19.148 for ESA-CCI.

Forest and *desert* had the strongest positive associations between land-cover and habitat classes. The *forest* habitat class was associated with almost all the forest and tree cover land-cover classes (CGLS-LC100 average positive odds ratio = 3.8; ESA-CCI average positive odds ratio = 4.0) and with no other land-cover classes. The *desert* habitat class was also strongly associated with particular land-cover classes: “shrubs,” “herbaceous vegetation,” and “bare/sparse vegetation” in CGLS-LC100 (average positive odds ratio = 4.6) and “shrubland,” “grassland,” “sparse vegetation (tree, shrub, herbaceous cover $< 15\%$),” and “bare areas” in ESA-CCI (average positive odds ratio = 3.0). *rocky areas* were associated with almost the same land-cover classes as *desert* but had lower odds ratios.

Savanna, *shrubland*, and *grassland* habitat classes were associated with “shrubs,” “herbaceous vegetation,” and “cultivated and managed vegetation agriculture” in CGLS-LC100 land cover and “cropland,” “herbaceous cover,” “shrubland,” “grassland,” “sparse vegetation,” “mosaic cropland,” and “mosaic herbaceous cover” in ESA-CCI. However, the power of association varied between these different combinations. The *savanna* habitat class was also associated with some forest classes, whereas *shrubland* and *grassland* habitats were also associated with bare areas.

We divided artificial terrestrial habitats into three different classes: *artificial arable and pasture lands*, *artificial degraded forest and plantations*, and *artificial urban and rural gardens*. These classes had the least certain relationships because the odds ratio values were the closest to 1 (CGLS-LC100 average positive odds ratio = 1.367, 1.333, and 1.577, respectively; ESA-CCI average positive odds ratio = 1.468, 1.370, and 1.579, respectively). Some

IUCN habitat class \ Land-cover class	Number point	Forest	Savanna	Shrubland	Grassland	Wetlands	Rocky areas	Desert	Artificial arable and pasture lands	Artificial degraded forest and plantation	Artificial urban areas and rural gardens	Artificial aquatic	AUC
Cropland, rainfed	2,47,928	0.662	1.609	1.417	1.415	1.660	-	-	1.617	1.309	1.393	1.613	0.824
Cropland, rainfed: herbaceous cover	2,99,354	0.732	1.434		1.674	1.581	-	-	1.922	-	1.453	1.678	0.821
Cropland, rainfed: tree or shrub cover	13,384	-	-	-	-	-	-	0.354	-	1.613	-	-	0.960
Cropland irrigated or post-flooding	40,494	-	-	-	1.704	1.996	-	-	1.763	1.396	1.357	1.708	0.954
Mosaic cropland (>50%) /natural vegetation (tree, shrub, herbaceous cover)(<50%)	1,22,990	-	1.183	1.234	1.253	1.363	-	0.750	1.475	1.283	1.232	-	0.773
Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%/cropland (<50%))	1,81,508	1.281	-	1.121	1.163	1.326	-	0.609	1.249	1.222	1.202	-	0.709
Tree cover, broadleaved, evergreen, closed to open (>15%)	7,44,491	12.230	0.495	0.354	0.505	0.734	0.391	0.036	0.686	1.354	-	-	0.949
Tree cover, broadleaved, deciduous, closed to open (>15%)	2,94,126	2.016	1.477	1.564	-	1.496	-	0.294	-	0.813	1.451	-	0.834
Tree cover, needleleaved, evergreen, closed to open (>15%)	4,44,643	3.695	0.250	-	-	-	2.011	0.643	0.704	0.431	-	-	0.886
Tree cover, needleleaved, deciduous, closed to open (>15%)	20,580	-	-	-	-	1.960	-	-	-	-	-	-	0.868
Tree cover, mixed leaf type (broadleaved and needleleaved)	88,506	5.990	0.520	0.705	-	1.510	-	0.463	-	-	-	-	0.898
Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	1,70,385	1.217	-	1.294	-	1.318	-	0.429	1.152	-	1.285	-	0.724
Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	20,667	0.788	1.543	1.232	1.675	-	-	-	-	-	-	-	0.874
Shrubland	4,71,187	0.535	1.758	2.133	1.309	-	1.678	1.603	-	0.702	-	0.626	0.873
Grassland	3,23,351	0.468	1.469	1.612	2.132	1.348	1.928	1.280	1.281	0.700	1.240	1.453	0.806
Lichens and mosses	778	-	-	-	19.148	-	-	-	-	-	-	-	0.972
Sparse vegetation (tree, shrub, herbaceous cover)(<15%)	80,470	0.111	1.547	2.073	2.205	-	1.770	3.919	-	0.449	-	-	0.937
Tree cover, flooded, fresh or brakish water	33,641	1.824	-	0.622	0.660	2.793	-	-	-	-	-	2.109	0.886
Tree cover, flooded, saline water	28,755	-	-	-	-	2.021	-	0.292	-	1.374	-	1.733	0.888
Shrub or herbaceous cover, flooded, fresh/ saline/brakish water	75,650	0.649	1.430	1.191	-	2.585	0.636	-	1.242	-	-	1.990	0.802
Urban area	5,75,679	-	-	1.352	-	1.611	-	0.676	1.755	1.409	3.582	1.718	0.768
Bare areas	33,857	0.174	-	2.147	1.457	-	2.717	5.375	-	0.558	-	-	0.877
Water bodies	2,17,197	0.651	-	-	-	2.167	-	0.487	1.221	-	1.598	1.939	0.749

Positive association (odds ratio = 1.121–1.393)	Positive association (odds ratio = 1.396–1.704)	Positive association (odds ratio = 1.708–19.148)
Negative association	Nonsignificant association	

FIGURE 3 . Odds ratio values describing the association between Copernicus Global Land Service Land Cover (CGLS-LC100) classes and International Union for the Conservation of Nature (IUCN) Habitat Classification Scheme. Odds ratio values < 1 indicate a negative association, and values > 1 indicate a positive association (AUC, area under the curve from a receiver operating characteristic [ROC] curve, a measure of accuracy of a classification mode). The positive associations are divided into tertiles (green), indicating three possible options for setting a threshold to convert continuous variables into a binary association-nonassociation variable for creating area of habitat maps

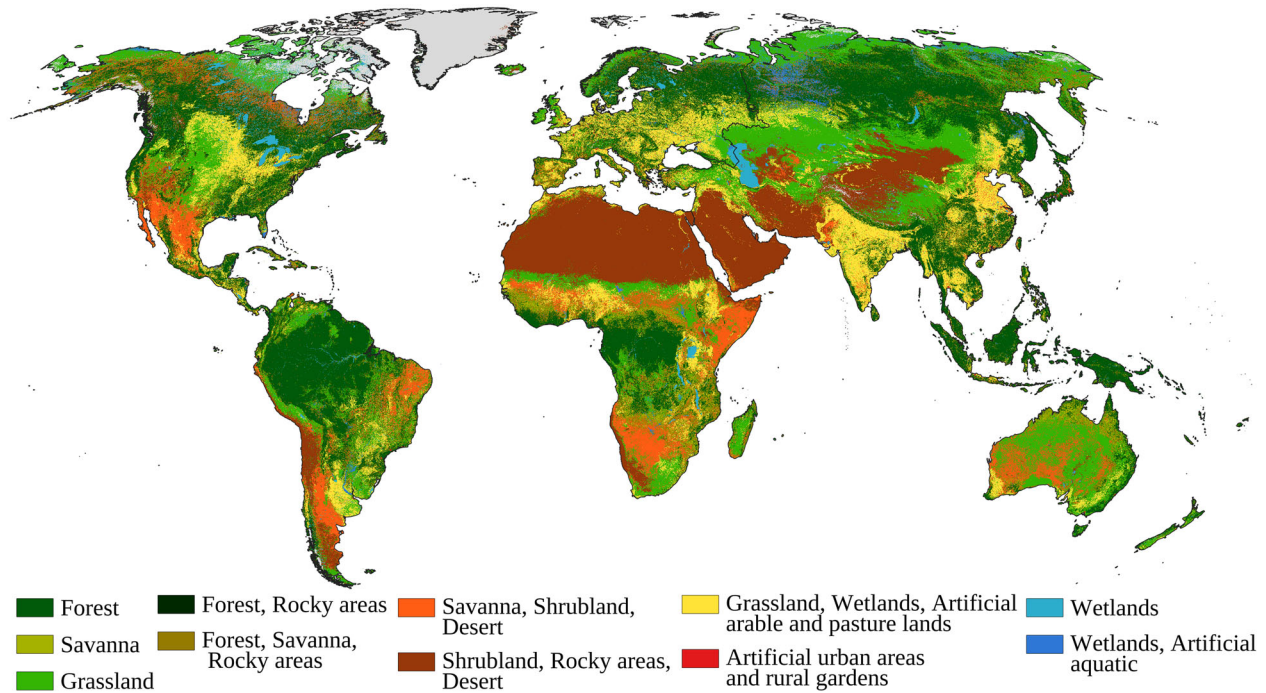


FIGURE 4 Map of habitat classes (level 1) from the International Union for the Conservation of Nature Habitat Classification Scheme based on the highest threshold for Copernicus Global Land Service Land Cover (CGLS-LC100) data-derived translation (Figure 2) (Geotiff version Appendix S2)

unexpected land-cover classes were associated with these habitat classes; for example, *Arable and pasture lands* and *degraded forest and plantations* were associated with “urban areas.” However, these unexpected associations disappeared when the threshold increased.

Wetland and *artificial aquatic* habitats had intermediate odds ratio values (CGLS-LC100 average positive odds ratio = 1.7; ESA-CCI average positive odds ratio = 1.8). In terms of land-cover associations, they were associated (in some cases strongly) with land-cover classes related to water, but also to some land-cover classes that have no relation with wetlands or aquatic environments (e.g., some type of forest or cultivated areas).

The AUC of models for CGLS-LC100 ranged from 0.644 to 0.940. The land-cover classes with the lowest AUC were the “open and closed unknown forest” (AUC = 0.644 and 0.736) classes, followed by “water bodies” (AUC = 0.745) and “urban areas” (AUC = 0.763). Those with the highest AUC values were the other forest classes (AUC range 0.854–0.940) and “bare and sparse vegetation” (AUC = 0.924). For ESA-CCI, the AUC ranged from 0.709 to 0.972. The land-cover classes with the lowest AUC were mosaic land-cover classes (AUC range 0.709–0.874), followed by “water bodies” (AUC = 0.750) and “urban areas” (AUC = 0.768). The land covers with the highest AUC values were “lichens and mosses” (AUC = 0.972), “cropland irrigated or post-flooding” (AUC = 0.954), “sparse vegetation” (AUC = 0.937), and tree cover land classes (AUC range 0.834–0.949). The spatial representation of the models showed the geographical distribution of the habitat classes (Figure 4 & Appendices S2 and S3). Habitat classes *savanna*, *shrubland*, *desert*, and *rocky areas* had the same geographical extent. In contrast, forest

had its own geographical distribution. *Grassland* had its own distribution and appeared in combination with *artificial arable and pasture and wetlands*.

Point prevalence in Santini et al. (2019) was similar to the point prevalence we found from our model when using the middle and high odds-ratio thresholds (Table 1). The ratio between point prevalence and model prevalence (proportion of the range remaining after apparently unsuitable land-cover classes are excluded) between the two methods was also very similar, and higher than 1, indicating that the habitat associations were better than random for both approaches.

DISCUSSION

By modeling the relationship between IUCN habitat classes and the CGLS-LC100 and ESA-CCI land-cover classes, we generated two translation tables, quantifying the strength of association between habitat and land-cover classes. Among habitat classes, *forest*, *desert*, and *rocky areas* had the strongest associations with land-cover classes, perhaps owing to the higher accuracy of the relevant land-cover classes. For both CGLS-LC100 and ESA-CCI, the highest classification accuracy classes were “forest,” “tree cover areas,” and “bare soil.” Using a different approach based on a decision tree, Jung et al. (2020) found that *Forest* has the highest validation accuracy, although they obtained lower validation accuracy for *rocky areas* and *desert* habitat classes.

In contrast, *wetlands* and *artificial* habitats were more difficult to represent with land-cover maps. Wetland-related land-cover

TABLE 1 Mean point prevalence^a and model prevalence^b for birds and mammals using the three tertile thresholds for ESA CCI land cover derived from data-driven assessment (see Figure 4) and the expert-knowledge-based assessment of Santini et al. (2019)

Model parameters	Lower tertile threshold	Middle tertile threshold	Upper tertile threshold	Santini et al. (2019)
Birds				
point prevalence	0.94	0.81	0.66	0.74
model prevalence	0.91	0.76	0.59	0.68
Mammals				
point prevalence	0.93	0.82	0.67	0.73
model prevalence	0.90	0.80	0.62	0.70

^aProportion of points located in the habitat.

^bProportion of habitat inside the species' range.

classes have the lowest classification accuracy in both land-cover maps. From a remote sensing perspective, wetlands are difficult to map because they are highly dynamic; rapid phenological changes occur throughout the year (Gallant, 2015; Lumbrerres et al., 2017). Remote sensing products at a global scale cannot distinguish small ponds or temporary water bodies (Pekel et al., 2016; Klein et al., 2017). Therefore, wetland land-cover classes had more omission errors, and this had a direct impact on the results of our model.

Artificial land-cover classes are also difficult to map because they tend to be more heterogeneous (Álvarez-Martínez et al., 2018), producing misclassifications among land-cover classes. Moreover, it is difficult to separate artificial land-cover classes from natural ecosystems (e.g., plantation from forest, grassland from cropland, and lake from reservoir) with land-cover maps (Álvarez-Martínez et al., 2018). Overall, species richness and average abundance are often lower in artificial environments than in their natural equivalent, even if there is variation across different biogeographical contexts (Barlow et al., 2007; Newbold et al., 2015), and this introduces commission errors. Moreover, we found that artificial land covers are associated with some natural habitat classes. This is likely a consequence of citizen science sampling bias produced by the greater accessibility of these habitats (Meyer et al., 2015). Because a high proportion of citizen science point location data are recorded in artificial land-cover classes, there is an increased probability that species primarily associated with natural habitats are reported there, so a data-driven method may associate some natural habitats with artificial land-cover classes. Addressing the biases in citizen science data is complex. For small data sets, accessibility maps are a useful tool for estimating sample bias (Monsarrat et al., 2019). However, at the global scale, accessibility is driven by an interplay of geographic and socioeconomic factors that require complex modeling approaches in addition to more effective and structured data sampling techniques.

Land-cover maps have an associated error that varies among different land-cover classes (Grekousis et al., 2015) and continents (Tsendbazar et al., 2020). Moreover, land-cover classes that do not occur in extensive blocks have edge effects (Smith et al., 2002), which, combined with the mobility of animals, introduces errors in the association of the point data with the land cover. There are several differences between the two land-cover

layers used to produce the translation tables that could determine the use of the table. The CGLS-LC100 has a resolution of 100 m, whereas ESA-CCI has a coarser resolution of 300 m, also CGLS-LC100 has an overall classification accuracy of 80.2% compared with 71.1% for ESA-CCI. Moreover, CGLS-LC100 avoids mosaic classes and in general; mapping areas with homogenous land cover is easier than mapping areas with heterogeneous land cover (Corbane et al., 2015; Álvarez-Martínez et al., 2018). The mosaic land-cover classes in the ESA-CCI table had very low odds ratio values and AUC. However, ESA-CCI has the advantage of being available for a longer time series (1992–2020 for ESA-CCI vs. 2015–2019 for CGLS-LC100), which may be important for some applications. For both land-cover maps, we excluded some land-cover classes because of the lack of point localities. We recommend adding these land-cover classes manually when using the translation tables, according to the user's needs.

The coding of habitats to each species on the IUCN Red List could introduce some noise to the modeling process. Coding is based on the qualitative assessment by experts and is, therefore, susceptible to individual biases (Brooks et al., 2019; Santini et al., 2019). The current version of the IUCN Habitat Classification Scheme on IUCN's website is described as a draft version. We, therefore, recommend that IUCN update and improve this document and anticipate this would influence our odds ratio estimates. Our analysis also illustrates the complexity of linking habitat and land cover (Tomaselli et al., 2013). Based on IUCN usage, *habitat* is a description of the environments of organisms (Kearney, 2006), whereas *land cover* is used to describe the biophysical material over the Earth's surface (Grekousis et al., 2015). Different habitat or land-cover schemes, stemming from the particular needs for each product, translate into different definitions of classes. This problem is exacerbated in transitional zones, where landscape heterogeneity is higher (Grekousis et al., 2015). Although the FAO-LCCS scheme, a scheme that defines the classes based on both land-cover maps, can better cope with the complexity of habitat description compared with other land-cover classification schemes (Grekousis et al., 2015), it is important to understand that these classes are not optimized for biodiversity conservation studies (Joppa et al., 2016), so they do not directly relate to the habitat of species.

Both the data-driven table and the expert-knowledge translation table represented land-cover distribution inside

the range better than random. However, our data-driven approach presents several advantages compared with an expert-knowledge approach. First, it defines the relationship between IUCN habitat and land-cover classes as a continuous variable, allowing greater flexibility in its application. For example, for producing AOH maps, the user is able to decide a threshold of association to transform the results into a binary table according to the required balance between omission and commission errors. Second, a data-driven approach allows quantification of the uncertainty associated with the habitat to land-cover association. Third, it represents a more objective approach: several expert-knowledge translation tables exist, but there is no clear basis for choosing among them.

These data-driven translation tables have a direct applicability in the production of AOH maps because they provide a more objective way of removing unsuitable areas from the range map based on the information from the IUCN Habitat Classification Scheme and enable evaluation of uncertainties in the AOH maps. Our approach can be adapted to develop a translation table between any set of habitat codes for any set of species and any set of land-cover classes at a global or regional scale. As better data (including land-cover maps, species point localities, elevations, and habitat associations) become available, the translation table can be improved, ensuring objectivity and repeatability.

ACKNOWLEDGMENTS


This research is part of the Inspire4Nature Innovative Training Network, funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement 766417. M.D.M. acknowledges support from the MIUR Rita Levi MontalcinLevii program.

Open Access Funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

ORCID

Maria Lumbierres  <https://orcid.org/0000-0001-8442-1264>

Prabhat Raj Dabal  <https://orcid.org/0000-0003-3490-4078>

Stuart H. M. Butchart  <https://orcid.org/0000-0002-1140-4049>

Paul F. Donald  <https://orcid.org/0000-0003-0023-6200>

Carlo Rondinini  <https://orcid.org/0000-0002-6617-018X>

LITERATURE CITED

- Álvarez-Martínez, J. M., Jiménez-Alfaro, B., Barquín, J., Ondiviela, B., Recio, M., Silió-Calzada, A., & Juanes, J. A. (2018). Modelling the area of occupancy of habitat types with remote sensing. *Methods in Ecology and Evolution*, *9*, 580–593.
- Barlow, J., Gardner, T. A., Araujo, I. S., Avila-Pires, T. C., Bonaldo, A. B., Costa, J. E., Esposito, M. C., Ferreira, L. V., Hawes, J., Hernandez, M. I. M., Hoogmoed, M. S., Leite, R. N., Lo-Man-Hung, N. F., Malcolm, J. R., Martins, M. B., Mestre, L. A. M., Miranda-Santos, R., Nunes-Gutjahr, A. L., Overal, W. L., ..., & Peres, C. A. (2007). Quantifying the biodiversity value of tropical primary, secondary, and plantation forests. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 18555–18560.
- Beresford, A. E., Buchanan, G. M., Donald, P. F., Butchart, S. H. M., Fishpool, L. D. C., & Rondinini, C. (2011). Poor overlap between the distribution of protected areas and globally threatened birds in Africa. *Animal Conservation*, *14*, 99–107.
- Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., & Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, *173*, 144–154.
- Boitani, L., Maiorano, L., Baisero, D., Falcucci, A., Visconti, P., & Rondinini, C. (2011). What spatial data do we need to develop global mammal conservation strategies? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 2623–2632.
- Bradley, B. A., Olsson, A. D., Wang, O., Dickson, B. G., Pelech, L., Sesnie, S. E., & Zachmann, L. J. (2012). Species detection vs. habitat suitability: Are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling*, *244*, 57–64.
- Bradter, U., Mair, L., Jönsson, M., Knappe, J., Singer, A., & Snäll, T. (2018). Can opportunistically collected citizen science data fill a data gap for habitat suitability models of less common species? *Methods in Ecology and Evolution*, *9*, 1667–1678.
- Brooks T. M., Pimm S. L., Akçakaya H. R., Buchanan G. M., Butchart S. H. M., Foden W., Hilton-Taylor C., Hoffmann M., Jenkins C. N., Joppa L., Li B. V., Menon V., Ocampo-Peñuela N., Rondinini C. (2019). Measuring Terrestrial Area of Habitat (AOH) and Its Utility for the IUCN Red List. *Trends in Ecology & Evolution*, *34*, (11), 977–986. <http://doi.org/10.1016/j.tree.2019.06.009>
- Buchanan, G. M., Butchart, S. H. M., Dutson, G., Pilgrim, J. D., Steining, M. K., Bishop, K. D., & Mayaux, P. (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., Lesiv, M., Tsendbazar, N. - E., Herold, M., & Fritz, S. A. (2019). Copernicus Global Land Service: Land Cover 100m: Epoch 2015: Globe. Dataset of the global component of the Copernicus Land Monitoring Service le.
- Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N. E., Linlin, L., & Tarko, A. (2020). *Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015–2019: Product User Manual*. Geneva: Zenodo.
- Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V. E., Strasser, T., Vanden Borre, J., Toon, S., & Michael, F. (2015). Remote sensing for mapping natural habitats and their conservation status — New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation*, *37*, 7–16.
- Crawford, B. A., Olds, M. J., Maerz, J. C., & Moore, C. T. (2020). Estimating population persistence for at-risk species using citizen science data. *Biological Conservation*, *243*, 108489.
- Di Gregorio, A., & Jansen, L. J. M. (2000). *Land cover classification system (LCCS): Classification concepts and user manual*. Rome: FAO.
- Di Marco, M., Watson, J. E. M., Possingham, H. P., & Venter, O. (2017). Limitations and trade-offs in the use of species distribution maps for protected area planning. *Journal of Applied Ecology*, *54*, 402–411.
- Díaz S., Settele J., Brondízio E. S., Ngo H. T., Agard J., Arneth A., Balvanera P., Brauman K. A., Butchart S. H. M., Chan K. M. A., Garibaldi L. A., Ichii K., Liu J., Subramanian S. M., Midgley G. F., Miloslavich P., Molnár Z., Obura D., Pfaff A., Polasky S., Purvis A., Razzaque J., Reyers B., Chowdhury R. R., Shin Y.-J., Vissers-Hamakers I., Willis K. J., Zayas C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*, (6471), <http://doi.org/10.1126/science.aax3100>
- Ducatez, S., Sayol, F., Sol, D., & Lefebvre, L. (2018). Are urban vertebrates city specialists, artificial habitat exploiters, or environmental generalists? *Integrative and Comparative Biology*, *58*, 929–938.
- eBird. (2019). *Basic dataset*. Version January 2019. Ithaca, NY: Cornell Lab of Ornithology.
- ESA (European Space Agency). (2017). *Land cover CCI product user guide version 2. Technical report*. Paris: ESA.
- Ficetola, G. F., Rondinini, C., Bonardi, A., Baisero, D., & Padoa-Schioppa, E. (2015). Habitat availability for amphibians and extinction threat: A global analysis. *Diversity and Distributions*, *21*, 302–311.
- Ficetola, G. F., Rondinini, C., Bonardi, A., Katariya, V., Padoa-Schioppa, E., & Angulo, A. (2014). An evaluation of the robustness of global amphibian range maps. *Journal of Biogeography*, *41*, 211–221.

- Gallant, A. (2015). The challenges of remote monitoring of wetlands. *Remote Sensing*, *7*, 10938–10950.
- GBIF.org. (2020). GBIF occurrence (download 4 March). Available from <https://doi.org/10.15468/dl.tvtiqq>.
- GBIF.org. (2019). GBIF occurrence (download 14 January). Available from <https://doi.org/10.15468/dl.tk87g2>.
- GBIF.org. (2020). GBIF occurrence (download 23 December). Available from <https://doi.org/10.15468/dl.swey54>.
- GBIF.org. (2020). GBIF occurrence (download 24 February). Available from <https://doi.org/10.15468/dl.5vqa7s>.
- GBIF.org. (2019). GBIF occurrence (download 26 January). Available from <https://doi.org/10.15468/dl.8bff5p>.
- Grekoussis, G., Mountrakis, G., & Kavouras, M. (2015). An overview of 21 global and 43 regional land-cover mapping products. *International Journal of Remote Sensing*, *36*, 5309–5335.
- Gueta, T., & Carmel, Y. (2016). Quantifying the value of user-level data cleaning for big data: A case study using mammal distribution models. *Ecological Informatics*, *34*, 139–145.
- 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*. Gland: IUCN.
- 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.
- Johnson, C. J., & Gillingham, M. P. (2004). Mapping uncertainty: Sensitivity of wildlife habitat ratings to expert opinion. *Journal of Applied Ecology*, *41*, 1032–1041.
- Joppa, L. N. et al. (2016). Filling in biodiversity threat gaps. *Science Translational Medicine*, *352*, 199.
- Jung, M., Dahal, P. R., Butchart, S. H. M., Donald, P. F., De Lamo, X., Lesiv, M., Kapos, V., Rondinini, C., & Visconti, P. (2020). A global map of terrestrial habitat types. *Scientific Data*, *7*, 1–8.
- Kearney, M. (2006). Habitat, environment and niche: What are we modelling? *Oikos*, *115*, 186–191.
- King, G., & Zeng, L. (2001). Logistic regression in rare. *Political Analysis*, *9*, 137–163.
- Klein, I., Gessner, U., Dietz, A. J., & Kuenzer, C. (2017). Global WaterPack – A 250 m resolution dataset revealing the daily dynamics of global inland water bodies. *Remote Sensing of Environment*, *198*, 345–362.
- Lumbierres, M., Méndez, P., Bustamante, J., Soriguer, R., & Santamaría, L. (2017). Modeling biomass production in seasonal wetlands using MODIS NDVI land surface phenology. *Remote Sensing*, *9*, 1–18.
- Meyer, C. (2012). Limitations in global information on species occurrences. *Frontiers of Biogeography*, *4*, 217–220.
- Meyer, C., Jetz, W., Guralnick, R. P., Fritz, S. A., Kreft, H., & Gillespie, T. (2016). Range geometry and socio-economics dominate species-level biases in occurrence information. *Global Ecology and Biogeography*, *25*(10), 1181–1193.
- Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*, 1–8.
- Monsarrat, S., Boshoff, A. F., & Kerley, G. I. H. (2019). Accessibility maps as a tool to predict sampling bias in historical biodiversity occurrence records. *Ecography*, *42*, 125–136.
- Montesino Pouzols, F., Toivonen, T., Di Minin, E., Kukkala, A. S., Kullberg, P., Kuusterä, J., Lehtomäki, J., Tenkanen, H., Verburg, P. H., & Moilanen, A. (2014). Global protected area expansion is compromised by projected land-use and parochialism. *Nature*, *516*, 383–386.
- Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Lysenko, I., Senior, R. A., Börger, L., Bennett, D. J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverría-Londoño, S., Edgar, M. J., Feldman, A., Garon, M., Harrison, M. L. K., Alhusseini, T., ..., & Purvis, A. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, *520*, 45–50.
- Pekel, J. - F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*, 418–422.
- Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*, 1246752.
- Pozzolo, A. D., Caelen, O., Johnson, R. A., Bontempi, G., Dal Pozzolo, A., Caelen, O., Bontempi, G., & Johnson, R. A. (2015). Calibrating probability with undersampling for unbalanced classification. 2015 IEEE Symposium Series on Computational Intelligence, 159–166.
- Ranganathan, P., Pramesh, C., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Logistic regression. *Perspectives in Clinical Research*, *8*, 148–151.
- Rondinini, C., Di Marco, M., Chiozza, F., Santulli, G., Baisero, D., Visconti, P., Hoffmann, M., Schipper, J., Stuart, S. N., Tognelli, M. F., Amori, G., Falcucci, A., Maiorano, L., & Boitani, L. (2011). Global habitat suitability models of terrestrial mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 2633–2641.
- Rondinini, C., & Boitani, L. (2012). Mind the map: Trips and pitfalls in making and reading maps of carnivore distribution. Boitani, L. & Powell, R. A., *Carnivore ecology and conservation. A handbook of techniques* (pp. 31–46). Oxford University Press.
- Rondinini, C., Stuart, S., & 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., Grantham, H., & Possingham, H. P. (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, *9*, 1136–1145.
- Santini, L., Butchart, S. H. M., Rondinini, C., Benítez-López, A., Hilbers, J. P., Schipper, A. M., Cengic, M., Tobias, J. A., & Huijbregts, M. A. J. (2019). Applying habitat and population-density models to land-cover time series to inform IUCN Red List assessments. *Conservation Biology*, *33*, 1084–1093.
- Seoane, J., Bustamante, J., & Diaz-Delgado, R. (2005). Effect of expert opinion on the predictive ability of environmental models of bird distribution. *Conservation Biology*, *19*, 512–522.
- Smith, J. H., Wickham, J. D., Stehman, S. V., & Yang, L. (2002). Impacts of patch size and land-cover heterogeneity on thematic image classification accuracy. *Photogrammetric Engineering and Remote Sensing*, *68*, 65–70.
- Tomaselli, V., Dimopoulos, P., Marangi, C., Kallimanis, A. S., Adamo, M., Tarantino, C., Panitsa, M., Terzi, M., Veronico, G., Lovregine, F., Nagendra, H., Lucas, R., Maiorano, P., Múcher, C. A., & Blonda, P. (2013). Translating land cover/land use classifications to habitat taxonomies for landscape monitoring: A Mediterranean assessment. *Landscape Ecology*, *28*, 905–930.
- Tsendsbazar, N. E., Tarko, A., Linlin, L., Herold, M., Lesiv, M., Fritz, S., & Maus, V. (2020). *Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015–2019: Validation Report*. Geneva: Zenodo.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Lumbierres, M., Dahal, P. R., Marco, M Di, Butchart, S. H. M., Donald, P. F., & Rondinini, C. (2022). Translating habitat class to land cover to map area of habitat of terrestrial vertebrates. *Conservation Biology*, *36*, e13851.

<https://doi.org/10.1111/cobi.13851>