ECOGRAPHY

Research

Habitats as predictors in species distribution models: Shall we use continuous or binary data?

Lukáš Gábor, Petra Šímová, Petr Keil, Alejandra Zarzo-Arias, Charles J. Marsh, Duccio Rocchini, Marco Malavasi, Vojtěch Barták and Vítězslav Moudrý

L. Gábor (https://orcid.org/0000-0001-6137-0994) ✉ (gabor@fzp.czu.cz), P. Šímová (https://orcid.org/0000-0003-2480-1171), P. Keil (https://orcid.org/0000-0003-3017-1858), A. Zarzo-Arias (https://orcid.org/ 0000-0001-5496-0144), D. Rocchini (https://orcid.org/0000-0003-0087-0594), M. Malavasi (https://orcid.org/0000-0002-9639-1784), V. Barták (https://orcid.org/0000-0001-9887-1290) and V. Moudrý (https://orcid.org/0000-0002-3194-451X), Dept of Spatial Sciences, Faculty of Environmental Sciences, Czech Univ. of Life Sciences Prague, Praha – Suchdol, Czech Republic. – LG and C. J. Marsh, Dept of Ecology and Evolutionary Biology, Yale Univ., New Haven, CT, USA and Center for Biodiversity and Global Change, Yale Univ., New Haven, CT, USA. DR also at: BIOME Lab, Dept of Biological, Geological and Environmental Sciences, Alma Mater Studiorum Univ. of Bologna, Bologna, Italy.

The representation of a land cover type (i.e. habitat) within an area is often used as an explanatory variable in species distribution models. However, it is possible that a simple binary presence/absence of the suitable habitat might be the most important determinant of the presence/absence of some species and, thus, be a better predictor of species occurrence than the continuous parameter (area). We hypothesize that the binary predictor is more suitable for relatively rare habitats (e.g. wetlands) while for common habitats (e.g. forests) the amount of the focal habitat is a better predictor. We used the Third Atlas of Breeding Birds in the Czech Republic as the source of species distribution data and CORINE Land Cover inventory as the source of the landcover information. To test our hypothesis, we fitted generalized linear models of 32 water and 32 forest bird species. Our results show that for water bird species, models using binary predictors (presence/absence of the habitat) performed better than models with continuous predictors (i.e. the amount of the habitat); for forest species, however, we observed the opposite. Thus, future studies using habitats as predictors of species occurrences should consider the prevalence of the habitat in the landscape, and the biological role of the habitat type in the particular species' life history. In addition, performing a preliminary comparison of the performance of the binary and continuous versions of habitat predictors (e.g. using information criteria) prior to modelling, during variable selection, can be beneficial. These are simple steps that will improve explanatory and predictive performance of models of species distributions in biogeography, community ecology, macroecology and ecological conservation.

Keywords: binary data, continuous data, land cover, niche models, variable selection

Introduction

Species distribution models (SDMs) are an important tool in macroecology, biogeography and wildlife management. The goal of SDMs is to map species distributions

or to estimate species niches, and there is an ongoing effort to improve their reliability (Araújo et al. 2019, Zurell et al. 2020, Merow et al. 2022). Selecting appropriate environmental predictors is a major methodological challenge of species distribution modeling (Dormann et al. 2007, Austin and Van Niel 2011, Williams et al. 2012, Mod et al. 2016, Misiuk et al. 2018, Moudrý et al. 2019, Smith and Santos 2020, Zurell et al. 2020). These environmental predictors, such as landcover or habitat type, are most often included in SDMs as the area or percentage of a particular land cover type within the individual sites (e.g. grid cells or atlas mapping squares; Milanesi et al. 2017, Halstead et al. 2019, Lecours et al. 2020, Tessarolo et al. 2021).

But what if, for some species, the total area of habitat is less relevant than the simple fact that a particular habitat is present or absent? To our knowledge, this possibility has been considered neither theoretically, nor empirically. In conservation biology, this is somehow related to the concept of critical habitat area (Fahrig 2001, Melo et al. 2018), i.e. to the idea that there is a certain habitat amount (threshold) below which a species cannot survive, leading to a step-like, rather than continuous, response of species probability of occurrence to habitat area. To our knowledge this has not been explored in the context of SDMs. Further, a guideline on whether habitat predictors should be included in SDMs as continuous, or binary variables would be directly applicable in many subfields of biogeography and community ecology.

In this study, we evaluate the effect of using forest and water habitats as binary or continuous predictors in species distribution modelling of 64 forest and water specialist bird species. Specifically, we propose two alternative hypotheses linking the probability of occurrence (P) of a species to either (a) the amount or (b) the presence/absence of a particular habitat within a spatial unit (e.g. grid cell).

The first hypothesis (H1) assumes that P is driven by continuous areas (Fig. 1a), i.e. that P increases continuously with the increase in the habitat area within a spatial unit. Reasoning supporting H1 is as follows: 1) Larger habitat areas support larger populations due to their carrying capacity and food and shelter availability, so that populations are less susceptible to stochastic extinctions, competition, predation and inbreeding depression (Hanski 1999, Lande et al. 2003), and 2) larger habitat areas are bigger targets for colonizing individuals from the surrounding habitat matrix (Buckley and Knedlhans 1986), increasing the probability of rescue effects after extinction events (Brown and Kodric-Brown 1977). We propose that these mechanisms will operate in the most common habitat types. H1 will also apply to species specializing in these common habitats. In Central Europe, forests can be considered an example of such habitats, with forest specialist species such as the long-tailed tit *Aegithalos caudatus*, the goldcrest *Regulus regulus* or the crested tit *Parus cristatus*.

An alternative hypothesis (H2) is that P is driven by binary presence or absence of a habitat (Fig. 1b, 2b). In other words, the amount of habitat within a spatial unit is irrelevant, and what matters for the species is that the habitat is simply there. However, we first need to know how small, or large such habitat needs to be to be able to support a viable population of the species. Therefore, H2 assumes that there is a threshold of habitat amount (e.g. 20% as in Fig. 1b, 2b), below which
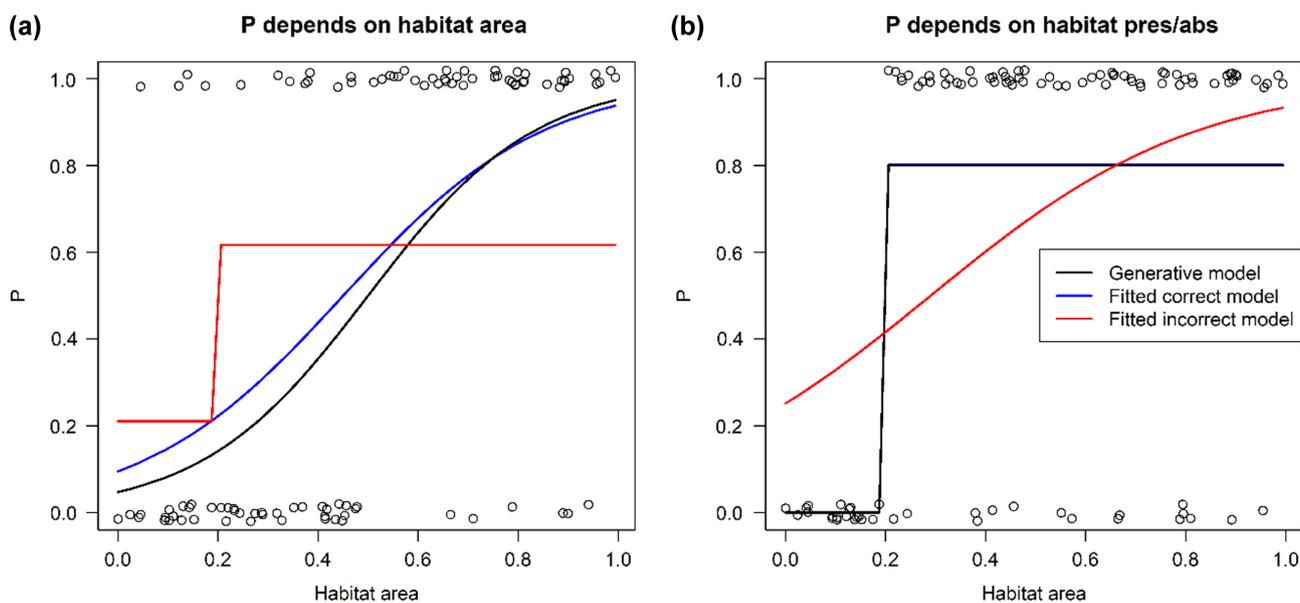
Figure 1. Two alternative hypotheses for the effect of the habitat area on the species' probability of occurrence (P), illustrating the theoretical possibility that habitat area be an accurate or inaccurate predictor of species distributions depending on if it is fitted as a continuous or binary variable and the process that generated the data. In the left panel we modelled P as a sigmoidal curve (generative model, black line) to generate 100 presences/absences of a species, drawn from a Bernoulli distribution with parameter P (jittered points). In right panel we used a binary habitat classification and a step function to generate the data. We then fitted binomial GLM with either continuous area, or binary area, as predictor (red and blue lines).

the species is unlikely to occur, and above the threshold the species will persist. We assume that an increase of the habitat area beyond the threshold size will not increase P (note that the threshold of habitat amount is affected by resolution of the habitat data). The presence of such threshold has been predicted both theoretically (Andrén 1994, Fahrig 2001) and documented empirically for birds (Melo et al. 2018).

We propose that H2 applies particularly to species specializing in rare (i.e. less prevalent) habitats, and species with good dispersal abilities and ability to readily identify the habitats in the landscape. Consequently, if a fragment of suitable habitat (irrespectively of its area) appears in the landscape, it will quickly attract a population of the species, thus causing high P. In Central Europe, water bodies can be considered an example of such habitat for water specialist species such as the common teal *Anas crecca*, the great-crested grebe *Podiceps cristatus* or the black tern *Chlidonias niger*.

Material and methods

Study area and bird distribution data

The study area was the territory of the Czech Republic, a central European country covering almost 79 000 km² (Fig. 2a). Data on bird species were obtained from the Third Atlas of Breeding Bird Distribution in the Czech Republic (Šťastný et al. 2006). The study area is divided into 628 grid squares of approx. 134 km² (10′ east longitude × 6′ north latitude; hereafter referred to as mapping squares) to which bird occurrences and environmental predictors are referred. The fieldwork for the atlas was conducted by volunteers between 2001 and 2003 where the breeding status of all species was recorded in each mapping square. Field observations of the bird species occurring in each mapping square were recorded using 17 numerical breeding codes (Hagemeijer and Blair 1997). Breeding occurrence of each bird species within a given mapping square was classified into one of the following categories: 0 – non-breeding (where no observations of the species were made, or where the species was observed but no breeding evidence was found), A – possible breeding, B – probable breeding or C – confirmed breeding. For the purpose of our study, all breeding categories (A, B and C) were used as presences whereas category 0 was used as absences. We prepared data for 85 bird species, 36 of them nesting in wetlands and surrounding habitats (e.g. standing water, littoral zones of ponds, swamps), and 49 species nesting in forests, following classification of Reif et al. (2006). Nevertheless, we had to remove 21 species with relatively small (less than 30 presence cells out of 628 cells), and relatively high occupancy (more than 598 presence cells out of 628 cells). Therefore, 32 water and 32 forest bird species (Supporting information) were included in the study.

Habitat variables

We derived four habitat predictors from the CORINE Land Cover database at 100 m resolution (Feranec et al. 2010). Specifically, within mapping squares, we derived the area of

agricultural areas (CORINE class 2), *artificial surfaces* divided into four classes (0, 0–20, 20–40, > 40 km²; CORINE class 1), continuous *area of water bodies* (CORINE class 5.1.2) and *area of forest* (CORINE class 3.1). In addition, binary factors representing presence or absence of water bodies and forests, respectively, were calculated. In order to generate binary habitat maps, it is necessary to determine an area threshold that defines the presence-absence of the habitat. An appropriate threshold should consider the prevalence of the habitat across the region of interest, the grain size at which the variable is being considered (i.e. the size of the grid cells at which the species are recorded) and the original grain size that the habitat variable is being aggregated from (i.e. the size of the grid cells of the original land-cover data, which is then aggregated to the larger modelling grain size). Due to the uncommonness of water habitats as well as due to the coarse resolution of CORINE Land Cover, we considered any amount of the water habitat in a cell as presence (i.e. the proportion of the cell occupied by one hectare set to > 0%). Forest pixels are, on the other hand, present in all mapping squares across the study region and, for this reason, we tested several thresholds (10, 20, 30, 40 and 50%) to derive the binary predictor.

Other environmental variables

Although the habitat predictors were our main focus, other predictors, such as climate, may also be important in determining the distributions of species. As climatic predictors, we used current climatic data from WorldClim (Hijmans et al. 2005). Following previous studies, we used two predictors: mean temperature and mean precipitation during the breeding season, i.e. in April–June (Moudrý and Šímová 2013, Venne and Currie 2021). We downloaded these at a resolution of 30 arc seconds (~ 1 km²) and averaged them inside each mapping square to match the grid resolution of the species distribution data (~ 100 km²). We also considered usage of elevation predictors such as maximum, minimum and range of elevation derived from Shuttle radar topography mission (SRTM, Farr et al. 2007, Moudrý et al. 2018) as they might be ecologically important to birds (Kosicki 2017). However, as these variables were highly correlated with the mean temperature in April–June, we eventually decided not to include them. The data were processed in arcgis ver. 10.7.1 (ESRI, CA, USA) and R (<www.r-project.org>) software.

Species distribution models

We fitted SDMs for each species using the climate variables and the water (or forest cover) variable for water (or forest) species; modelling was always performed separately for the continuous as well as binary water (or forest) variable. We did not use the information about forest areas for wetland species and vice versa (see Supporting information for used formulas). In addition, we also considered using forest and wetland area transformed with arcsine, log or square root, all of which are often adopted in ecological studies for areal predictors (for more details, Roberts 1986, Palmer 1993). However, these

(A) Study area

Forests

Forests and water bodies within a grid cell

Mapping squares

Water bodies

(B) Threshold of habitat amount

Threshold of water bodies > 0%
Habitat prevalence = 0.46

Threshold of forests > 20%
Habitat prevalence = 0.75

Threshold of forests > 40%
Habitat prevalence = 0.29
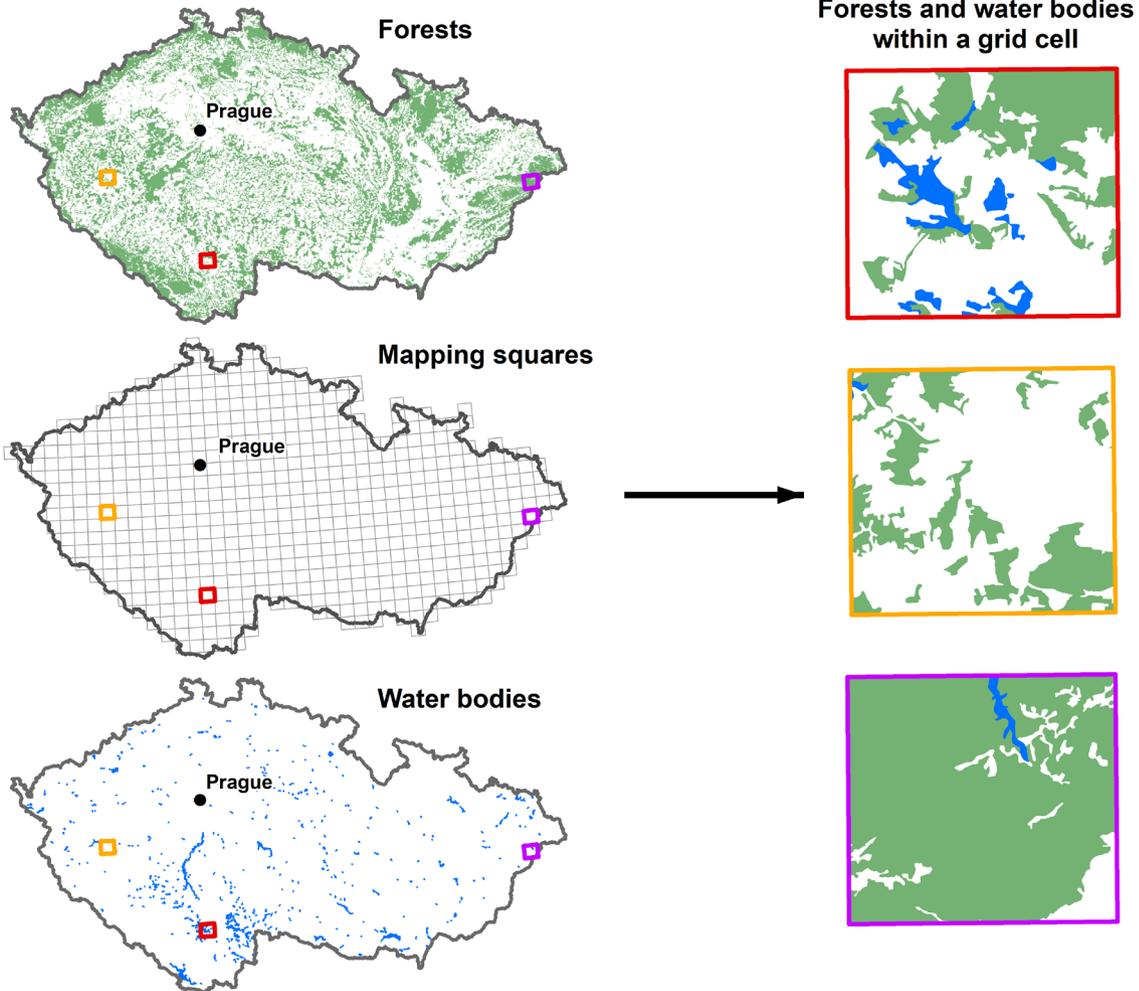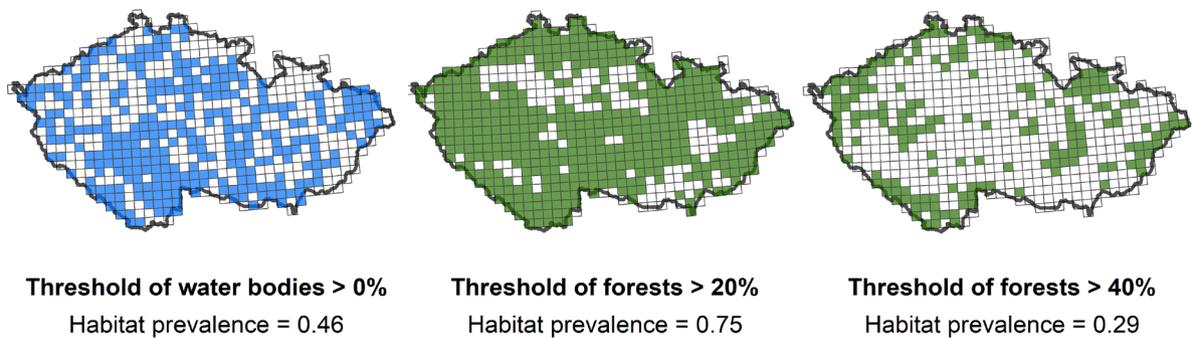
Figure 2. (A) The study region covers the territory of the Czech Republic, Europe. The grid consists of cells of 10′ east longitude × 6′ north latitude (approximately 12 × 11.2 km, n = 628), as used in the breeding birds atlases of the Czech Republic (Šťastný et al. 2006). Water bodies shown on the left side have a 500 m buffer for better visibility. (B) Representation of binary variables based on different threshold of habitat amount. We considered any amount of the water habitat in a cell as presence (i.e. the proportion of the cell occupied by one hectare set to > 0%). In addition, we tested several thresholds (e.g. 20% and 40%) to derive the binary predictor for forest variables.

transformations did not improve the models and were not further considered in this study. We used generalized linear models (GLMs; McCullagh and Nelder 1989), with binomial error distribution and a logit link function implemented in the R function glm. Environmental predictors were used as monotonic sigmoidal functions on the probability scale of the response (i.e. linear in logit space).

Model calibration and evaluation

We assessed the performance of the models using calibration and discrimination metrics where calibration refers to the accuracy of description of the environmental relationships, and discrimination refers to the ability to separate presences from absences (Lawson et al. 2014). We used five-fold cross-validation where the data were randomly divided into fifths to evaluate the models. Four fifths of the data were used to train the model and the remaining one fifth was used to assess the performance. To assess model calibration, we used likelihood-based McFadden's pseudo R^2 (Smith and McKenna 2013), which indicates the proportion of the deviance in the dependent variable that is explained by the model (Agresti 2003). To assess the model discrimination ability, we used the area under the curve of the receiver operating characteristic plot (AUC). The AUC is a threshold independent measure of model performance that ranges from 0 to 1, where a score of 1 indicates perfect discrimination, and a score of 0.5 indicates random performance (Fielding and Bell 1997).

Results

The occurrence of water birds was better modelled using the binary variable (prevalence of water habitat = 0.4; Fig. 3 4, Supporting information), suggesting that their distribution is driven simply by presence rather than the area of water habitat. In contrast, the models with continuous environmental variables outperformed those using binary predictors in modelling forest birds (Fig. 3, 4). This result was observed independently of the forest amount threshold for most of the species (Supporting information); however, to maintain clarity, we present results of models fitted with a 40 % threshold (prevalence of forest habitat = 0.29; see Supporting information for results using remaining tested thresholds).

For forest species, models fitted using the area of forest (i.e. a continuous habitat variable) achieved poor to excellent model calibration (R^2: min = 3.34%, max = 42%, mean = 16.2%) and discrimination performances (AUC: min = 0.60, max = 0.91, mean = 0.74; see Supporting information for the performance of individual models). Both model calibration (R^2: min = 2.69%, max = 38.42%, mean = 14.3%) and discrimination (AUC: min = 0.6, max = 0.89, mean = 0.72) were lower in models using the forest presence (i.e. a binary habitat variable) in 28 out of 32 species, although differences in model performances were relatively small. The differences in model calibration between models fitted using the area of forest and forest presence were negligible, except for six species where R^2 increased by up to 7% (Fig. 4). Similarly,
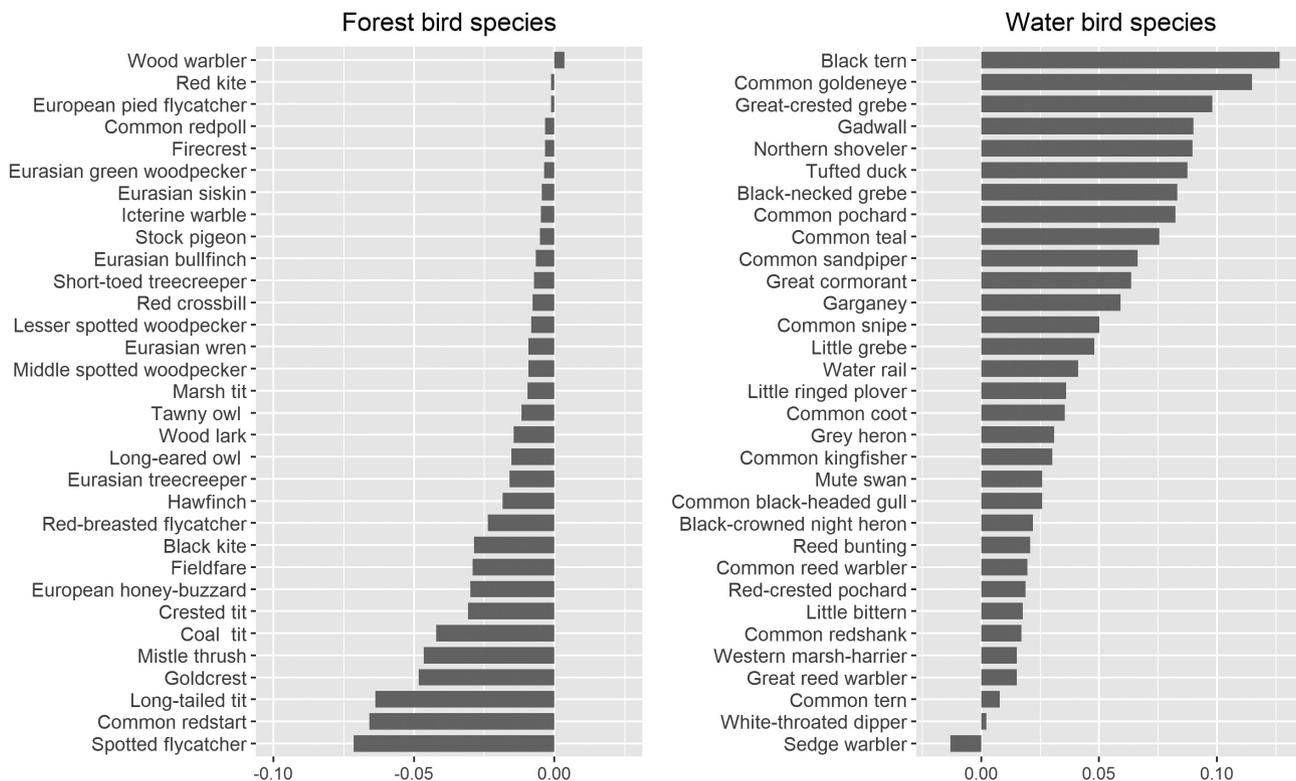
Figure 3. Differences in AUC between models fitted with binary habitat presence/absence versus continuous area as predictors. Positive values indicate that models with binary habitat predictors performed better than those with continuous predictors and vice versa.
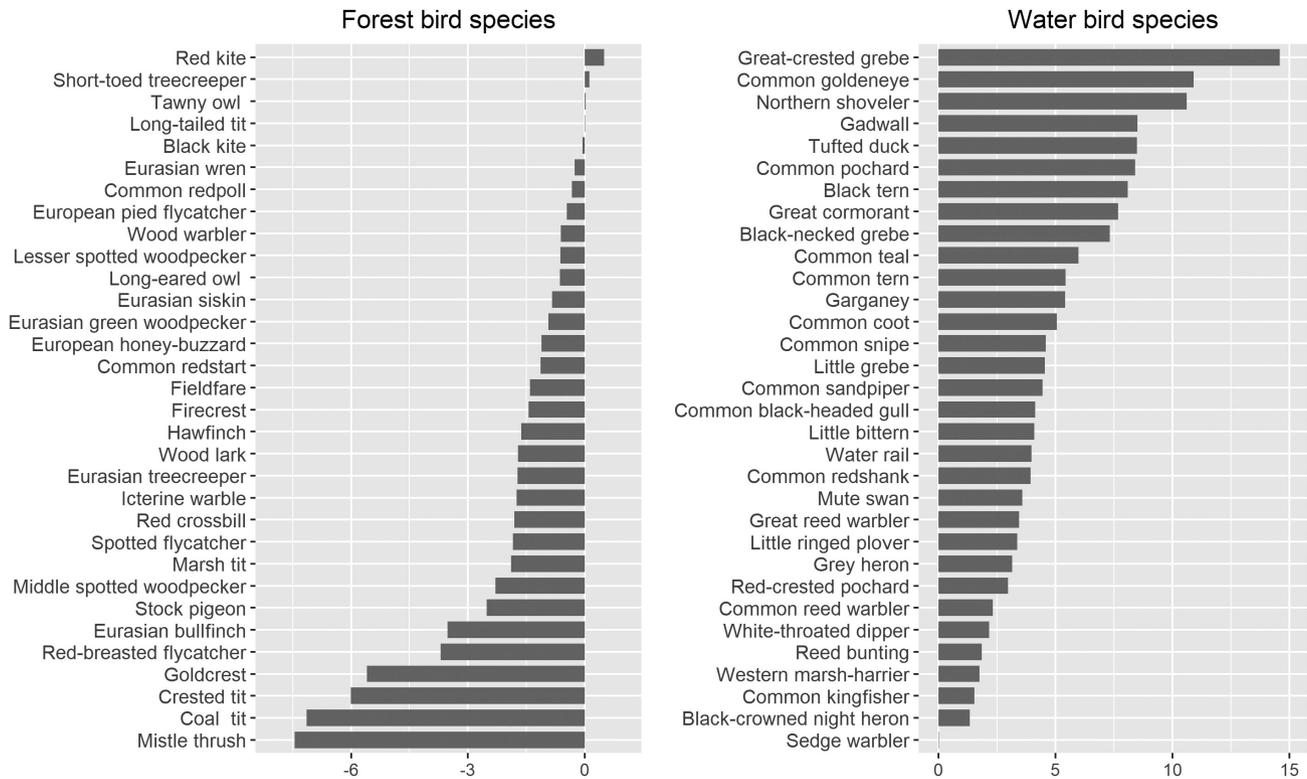
Figure 4. Differences in R² (%) between models fitted with binary habitat presence/absence vs continuous area as predictors. Positive values indicate that models with binary habitat predictors performed better than models with continuous predictors and vice versa.

the difference in AUC was < 0.02 for 22 out of 32 species. The highest AUC differences (~0.07) were recorded for long-tailed tit *Aegithalos caudatus*, black redstart *Phoenicurus ochruros*, and spotted flycatcher *Muscicapa striata*.

Models fitted for water species using the water presence performed better in both calibration (R²: min = 9.62%, max = 38.62%, mean = 22.1%) and discrimination (AUC: min = 0.69, max = 0.91, mean = 0.79; see Supporting information for the performance of individual models) compared to those using area of water (R²: min = 5.05%, max = 37.29%, mean = 16.96%; AUC: min = 0.63, max = 0.89, mean = 0.75) in nearly all cases. R² was on average ~5 %, and up to almost 15%, higher when using the water presence (Fig. 3). For 13 out of 32 water bird species, model discriminations (AUC) were increased by > 0.05 when considering the water presence compared to the area of water. In three cases (black tern *Chlidonias niger*; common goldeneye *Bucephala clangula*; and great-crested grebe *Podiceps cristatus*) the improvement in AUC was close to or even greater than 0.1. The model using the area of water was superior to that using the water presence only for a single species (sedge warbler, *Acrocephalus schoenobaenus*).

Discussion

Our results are in line with the hypotheses presented in the introduction. As expected, for species for which a widespread habitat (e.g. forest) is sufficient, models discriminated presences from absences better and explained more variability when a continuous, not binary, measure of the forest (habitat) area was used. On the other hand, the opposite was found for species specializing in a relatively rare habitat–water. In this case, models using water as a binary predictor outperformed those with water as a continuous area. As we have suggested, there are biological reasons for this: the relationship between species biology and specific habitat (and its rarity) determines how a binary habitat predictor stands out against continuous one. For instance, in waterbirds, the presence or absence of wetland or water habitats, which worked well in binary models, is directly related to food and shelter availability (Wiens 1992, Weller 1999, Gatto et al. 2008). Moreover, highly mobile waterbird species such as common redshank *Tringa totanus*, common tern *Sterna hirundo*, or gadwall *Anas strepera* are able to spot such habitat and colonize it, even if the habitat is rare and isolated in an otherwise dry landscape matrix. Thus, a patch of isolated wetland within a grid cell is almost guaranteed to host the species, despite the habitat being rare. We argue that such biological reasoning should precede any decisions about the specific form (binary or continuous) of predictors in SDMs. However, future studies are needed to show if this explanation based on habitat rarity applies to other environments, habitats and taxa.

In addition, we propose that the relative merit of continuous versus binary predictors depends on the interplay between spatial resolution of the habitat data (Domisch et al.

2015, Friedrichs-Manthey et al. 2020), spatial grain at which habitats are aggregated for modelling (response grain; Seoane et al. 2004, Venier et al. 2004, Convertino et al. 2011, Moudrý and Šímová 2012, Tuanmu and Jetz 2014, Šímová et al. 2019), as well as the home range size of the species, and its degree of specialisation to the habitat (Jedlikowski et al. 2016, Mertes et al. 2020). For highly specialised species, the ratio between the home range size and the grain size of the response variable may be particularly important (Jedlikowski et al. 2016), as it determines whether the species can gather resources from multiple grid cells, or whether it is confined to a single cell. For example, if the area of a single cell classified as water is larger than the home range of a highly specialised species, the binary predictor (water presence) should be used. However, if that area is smaller than the species home range, considering the habitat area is preferable, as the higher is the representation of water within a grid cell, the higher is the probability that the cell contains the habitat area necessary for species persistence. In addition, our habitat predictors were derived from the CORINE database (Feranec et al. 2010), with a minimum mapping unit of 25 ha, and it may be that forest species require habitats larger than that. Indeed, SDM studies using common land use categories as predictors, such as the proportion of forests, reported low improvement using finer resolution data (Seoane et al. 2004, Venier et al. 2004). In contrast, atlas squares with binary presence of water almost always contain a substantial area of water bodies, possibly enough to support a persistent breeding population of a waterbird species, leading to the good performance of the binary water predictor. In line with this, Šímová et al. (2019) showed that the area of water bodies derived from high-resolution (30 m) datasets explain distributions of waterbirds better than predictors derived from coarser water datasets (including CORINE Land Cover). This may be a reason why Tuanmu and Jetz (2014) found the Global Consensus Land cover (1 km resolution) performed worse for water species than for species that from other environments. In addition, Seoane et al. (2004) found considerable improvement of models for riparian species when finer-resolution data were used. However, this should be further validated, especially using finer resolutions than ours. Moreover, our results suggest that a hectare of wetland may be enough to be used by many water bird species and thus in future studies water habitats with equal or larger area then one hectare can be used as presence of habitat.

Other reason for the better performance of binary wetland predictor is that the threshold for absence was 0%, whereas for forest predictors 40% of habitat cover. In the threshold chosen in the results, the binary predictor is higher or lower than 40% cover of forest, which could be expected to have lower discrimination capacity than absolute absence of forest versus presence of forest. Note, however, that with the resolution of response variable approx. 12 × 11 km (used in our study), the absolute absence of forest in most of the Europe is unlikely.

It is fair to point out that only few species show considerable difference between models fitted using binary and continuous variables (i.e. the differences are relatively small for most of the species). Thus, models adopting traditional continuous variables will likely produce useful predictions. Nevertheless, in terms of practical recommendations, we advocate for testing both types of such variables during variable selection; this could be done using model selection criteria (e.g. AIC, BIC, DIC), cross-validation or measures of model fit (R^2, AUC). In addition, some of the biological mechanisms outlined in the introduction can also help with the decision on the preferable form of the predictor, and the importance of selecting biologically meaningful habitat predictors prior to modelling cannot be overstated. To summarize, categorical predictors would be preferred for 1) highly mobile species where even a small fragment of habitat is sufficient, 2) species for which one can expect a threshold response to the environment (e.g. at least a 20% coverage of the habitat within a mapping square is needed for the occurrence); and 3) highly mobile species specializing in less prevalent habitats, which could be quickly identified in the landscape and colonized.

The fact that habitat variables can, depending on the commonness of the habitat, perform best either as a binary or a continuous predictor, has relevance beyond simple species distribution models. After all, the information on the probability of species presence is sought after in many fields, from epidemiology to metacommunity ecology. Particularly in the latter, estimation of species responses to environmental conditions (including habitats) is at the core of the assessment of the relative role of niche versus spatial processes structuring ecological communities (Cottenie 2005, Leibold and Chase 2017). Our results suggest that if an inappropriate response of species to habitat amount (threshold versus continuous) is used, it can lead to underestimation of the importance of the niche processes. Furthermore, our results are relevant for conservation. Specifically, it is encouraging that, for many species, the presence of a (rare) habitat above a certain threshold (Radford et al. 2005, Melo et al. 2018) is important irrespectively of its area. If the critical threshold is low, even the protection of small areas is meaningful.

Author contributions

Lukáš Gábor: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Funding acquisition (lead); Methodology (lead); Supervision (equal); Validation (lead);

Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **Petr Keil**: Conceptualization (lead); Formal analysis (equal); Methodology (lead); Supervision (equal); Writing – original draft (lead); Writing – review and editing (lead). **Alejandra Zarzo-Arias**: Conceptualization (supporting); Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (equal). **Charles J. Marsh**: Methodology (equal); Validation (equal); Writing – original draft (supporting); Writing – review and editing (equal). **Petra Šímová**: Conceptualization (supporting); Data curation (equal); Funding acquisition (supporting); Methodology (equal); Writing – review and editing (equal). **Duccio Rocchini**: Conceptualization (supporting); Methodology (supporting); Writing – review and editing (equal). **Marco Malavasi**: Formal analysis (supporting); Writing – review and editing (equal). **Vojtěch Barták**: Validation (equal); Writing – review and editing (supporting). **Vítězslav Moudrý**: Conceptualization (lead); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (lead); Supervision (lead); Validation (equal); Visualization (equal); Writing – original draft (lead); Writing – review and editing (equal).

Data availability statement

The R script with step-by-step guideline as well as species and environmental data we used in the study are available from the Dryad Digital Rrepository, <https://doi.org/10.5061/dryad.4j0zpc8dk> (Gábor et al. 2022).

Supporting information

The supporting information associated with this article is available from the online version.

References

Agresti, A. 2003. Categorical data analysis, vol. 482. – Wiley.

Andrén, H. 1994. Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. – Oikos 76: 355–366.

Araújo, M. B. et al. 2019. Standards for distribution models in biodiversity assessments. – Sci. Adv. 5: eaat4858.

Austin, M. P. and Van Niel, K. P. 2011. Improving species distribution models for climate change studies: variable selection and scale. – J. Biogeogr. 38: 1–8.

Brown, J. H. and Kodric-Brown, A. 1977. Turnover rates in insular biogeography: effect of immigration on extinction. – Ecology 58: 445–449.

Buckley, R. C. and Knedlhans, S. B. 1986. Beachcomber biogeography: interception of dispersing propagules by islands. – J. Biogeogr. 13: 69–70.

Convertino, M. et al. 2011. Scale- and resolution-invariance of suitable geographic range for shorebird metapopulations. – Ecol. Complex. 8: 364–376.

Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. – Ecol. Lett. 8: 1175–1182.

Domisch, S. et al. 2015. Application of species distribution models in stream ecosystems: the challenges of spatial and temporal scale, environmental predictors and species occurrence data. – Fundam. Appl. Limnol. 186: 45–61.

Dormann, F. C. et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. – Ecography 30: 609–628.

Fahrig, L. 2001. How much habitat is enough? – Biol. Conserv. 100: 65–74.

Farr, T. G. et al. 2007. The shuttle radar topography mission. – Rev. Geophys. 45: RG2004.

Feranec, J. et al. 2010. Determining changes and flows in European landscapes 1990–2000 using CORINE land cover data. – Appl. Geogr. 30: 19–35.

Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – Environ. Conserv. 24: 38–49.

Friedrichs-Manthey, M. et al. 2020. From topography to hydrology – the modifiable area unit problem impacts freshwater species distribution models. – Ecol. Evol. 10: 2956–2968.

Gábor, L. et al. 2022. Data from: Habitats as predictors in species distribution models: Shall we use continuous or binary data? – Dryad Digital Repository, <https://doi.org/10.5061/dryad.4j0zpc8dk>.

Gatto, A. et al. 2008. Feeding behavior and habitat use in a waterbird assemblage at a marine wetland in coastal Patagonia, Argentina. – Waterbirds 31: 463–471.

Hagemeijer, W. J. and Blair, M. J. 1997. The EBCC atlas of European breeding birds. – Poyser, p. 479.

Halstead, K. E. et al. 2019. Using a species-centered approach to predict bird community responses to habitat fragmentation. – Landsc. Ecol. 34: 1919–1935.

Hanski, I. 1999. Metapopulation ecology. – Oxford Univ. Press.

Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Climatol. 25: 1965–1978.

Jedlikowski, J. et al. 2016. Multi-scale habitat selection in highly territorial bird species: exploring the contribution of nest, territory and landscape levels to site choice in breeding rallids (Aves: Rallidae). – Acta Oecol. 73: 10–20.

Kosicki, J. Z. 2017. Should topographic metrics be considered when predicting species density of birds on a large geographical scale? A case of Random Forest approach. – Ecol. Model. 349: 76–85.

Lande, R. et al. 2003. Stochastic population dynamics in ecology and conservation. – Oxford Univ. Press on Demand.

Lawson, C. R. et al. 2014. Prevalence, thresholds and the performance of presence–absence models. – Methods Ecol. Evol. 5: 54–64.

Lecours, V. et al. 2020. Fine-scale habitat characterization of the Gully, the Flemish Cap and the Orphan Knoll, Northwest Atlantic, with a focus on cold-water corals. Eds Peter Harris, Elaine Baker. – In: Harris, P. and Baker, E. (eds), Seafloor geomorphology as benthic habitat. Elsevier, pp. 735–751.

Leibold, M. A. and Chase, J. M. 2017. Metacommunity ecology, Vol. 59. – Princeton Univ. Press.

McCullagh, P. and Nelder, J. A. 1989. Generalized linear models, 2nd edn. – Chapman and Hall.

Melo, I. et al. 2018. A review of threshold responses of birds to landscape changes across the world. – J. Field Ornithol. 89: 303–314.

Merow, C. et al. 2022. Operationalizing expert knowledge in species' range estimates using diverse data types. – Front. Biogeogr., doi: 10.21425/F5FBG53589.

Mertes, K. et al. 2020. Hierarchical multi-grain models improve descriptions of species' environmental associations, distribution and abundance. – Ecol. Appl. 30: e02117.

Milanesi, P. et al. 2017. Towards continental bird distribution models: environmental variables for the second European breeding bird atlas and identification of priorities for further surveys. – Vogelwelt 137: 53–60.

Misiuk, B. et al. 2018. A multiscale approach to mapping seabed sediments. – PLoS One 13: e0193647.

Mod, H. K. et al. 2016. What we use is not what we know: environmental predictors in plant distribution models. – J. Veg. Sci. 27: 1308–1322.

Moudrý, V. and Šímová, P. 2012. Influence of positional accuracy, sample size and scale on modelling species distributions: a review. – Int. J. Geogr. Inform. Sci. 26: 2083–2095.

Moudrý, V. and Šímová, P. 2013. Relative importance of climate, topography and habitats for breeding wetland birds with different latitudinal distributions in the Czech Republic. – Appl. Geogr. 44: 165–171.

Moudrý, V. et al. 2018. On the use of global DEMs in ecological modelling and the accuracy of new bare-earth DEMs. – Ecol. Model. 383: 3–9.

Moudrý, V. et al. 2019. Potential pitfalls in rescaling digital terrain model-derived attributes for ecological studies. – Ecol. Inform. 54: 100987.

Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. – Ecology 74: 2215–2230.

Radford, J. Q. et al. 2005. Landscape-level thresholds of habitat cover for woodland-dependent birds. – Biol. Conserv. 124: 317–337.

Reif, J. et al. 2006. Population trends of birds in the Czech Republic during 1982–2005. – Sylvia 42: 22–37.

Roberts, D. W. 1986. Ordination on the basis of fuzzy set theory. – Vegetatio 66: 123–131.

Seoane, J. et al. 2004. Are existing vegetation maps adequate to predict bird distributions? – Ecol. Model. 175: 137–149.

Šímová, P. et al. 2019. Fine scale waterbody data improve prediction of waterbird occurrence despite coarse species data. – Ecography 42: 511–520.

Smith, A. B. and Santos, M. J. 2020. Testing the ability of species distribution models to infer variable importance. – Ecography 43: 1801–1813.

Smith, T. J. and McKenna, C. M. 2013. A comparison of logistic regression pseudo R2 indices. – Multiple Linear Regress. Viewp. 39: 17–26.

Šťastný, K. et al. 2006. Atlas of breeding bird distribution in the Czech Republic 2001–2003. – Aventinum.

Tessarolo, G. et al. 2021. High uncertainty in the effects of data characteristics on the performance of species distribution models. – Ecol. Indic. 121: 107147.

Tuanmu, M. N. Jetz, W. 2014. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. – Global Ecol. Biogeogr. 23: 1031–1045.

Venier, L. A. et al. 2004. Climate and satellite-derived land cover for predicting breeding bird distribution in the Great Lakes Basin. – J. Biogeogr. 31: 315–331.

Venne, S. and Currie, D. J. 2021. Can habitat suitability estimated from MaxEnt predict colonizations and extinctions? – Divers. Distrib. 27: 873–886.

Weller, M. W. 1999. Wetland birds: habitat resources and conservation implications. – Cambridge Univ. Press.

Wiens, J. A. 1992. The ecology of bird communities, Vol. 1. Foundations and patterns. – Cambridge Univ. Press.

Williams, K. J. et al. 2012. Which environmental variables should I use in my biodiversity model?. – Int. J. Geogr. Inform. Sci. 26: 2009–2047.

Zurell, D. et al. 2020. A standard protocol for reporting species distribution models. – Ecography 43: 1261–1277.