2024 ERDKUNDE Vol. 78 · No. 3 · 195–224

EVALUATING URBAN BIODIVERSITY: EFFECTIVENESS OF CITIZEN SCIENCE DRIVEN SPECIES DISTRIBUTION MODELS IN URBAN ECOSYSTEMS

A case study in the Ruhr Metropolis, Germany

Malte Bührs, Harald Zepp and Thomas Schmitt

With 7 figures, 3 tables and appendix
Received 02 April 2024 · Accepted 16 September 2024

Summary: Citizen science (CS) and remote sensing (RS) approaches have become more reliable, thus providing higher resolution and generating a large amount of environmental data. When considering urban environments, where fragmented and highly diverse landscapes are predominant, the combination of citizen science data and remote sensing techniques with species distribution models (SDM) can play a vital role in comprehensively investigating and evaluating urban biodiversity. However, citizen science derived species distribution models for multiple avian species in dense and fragmented urban areas are rarely used. The study aims (I) to elaborate, whether CS driven SDMs can be effectively used in spatially complex urban environments; (II) to identify biodiversity hotspots and prioritize areas for nature conservation and (III) to examine, if existing protective areas correspond to species' hotspots. Therefore, Citizen science-based datasets of 26 breeding bird species over three years were obtained for this analysis in Germany's Ruhr Metropolis. Quality assurance, data thinning, and pseudo-absence simulations were performed. Spatial data from the ecosystem LiDAR project GEDI (Global Ecosystem Dynamics Investigation), climate data from the German Weather Service, and land use information from Copernicus were used as environmental predictors. Eleven different species distri-bution models (SDMs) were trained on species subselection by using Biomod2 for preliminary analysis. Overall model performance was evaluated via several metrics, including TSS (true skill statistics) and ROC (receiver operating characteristic). Finally, four species distribution models were used for ensemble modelling. Subsequently, a species richness analysis was performed with the aim of identifying spots with high avian biodiversity. Overall, the CS-derived SDMs performed well, with high predictive power for all of the investigated species. Within the Ruhr Metropolis, approximately 6% (250 km²) of the terrain was identified as being highly suitable for avian diversity, inhabiting at least 17 out of 26 species. Predominantly within the core urban areas, high species richness was predicted on preserved brownfields and revitalized mine sites, as well as in the remnants of formerly demarcated regional greenbelts. Additionally, regions outside of the core area, which are part of the overarching biotope network framework, proved to have high species richness capabilities for avian biodiversity. These findings aid in optimizing urban development concepts and (sub)urban green space management with respect to urban biodiversity conservation. Following the implications of the recently established Regional Biodiversity Strategy in the Ruhr Metropolis, this analysis demonstrates the importance of networked green spaces, their preservation and the need to close existing network gaps within the Ruhr Metropolis.

Keywords: Avian biodiversity, species distribution models, citizen science, urban ecology, Ruhr Area, animal geography

1 Introduction

The forthcoming decades are anticipated to experience the most significant decline in biodiversity since the last mass extinction 65 million years ago (IPBES 2019). The swift expansion of urban human populations and their utilization of land, water, timber, and energy, which commonly occur in biodiversity-rich areas, have been identified as being primary catalysts for these declines (Elmqvist et al. 2016). These urban areas are known as agglomerates of complex spatio-temporal mosaics of different habitat types and are characterized by various climate conditions, habitat compositions, and fragmented landscapes (Pickett et al. 2008, Lepczyk et al. 2023). Consequently, conventional regional and urban planning typically overlooks biodiversity and nature within cities. However, at a frequent rate, conservation planning even disregards urban environments, with the presumption that they have minimal-to-no biodiversity (Guerry et al. 2021).

However, in recent years, the perception of urban settlements as biodiversity threats and population sinks has shifted (Soanes & Lentini 2019, McDonald et al. 2020, Lepczyk et al. 2023). Cities are even increasingly acknowledged as being hotspots due to their heterogeneous mosaics of open green spaces, forest relics, open water bodies, and

https://doi.org/10.3112/erdkunde.2024.03.03 ISSN 2702-5985 (Online)

spontaneous vegetation. This recognition is accepted not only for endangered species (Soanes & Lentini 2019, McDonald et al. 2020) but also for biodiversity in general (Knapp et al. 2021, Spotswood et al. 2021). A growing body of literature highlights the importance of various types of green infrastructure for biodiversity conservation, including their functions as suitable habitats (Dearborn & Kark 2010), landscape stepping stones (Wu et al. 2022), and refuges (Lepczyk et al. 2023). Nonetheless, open green spaces in cities and their surroundings are under pressure due to densification (Haaland & van den Bosch 2015), urban development (Spangenberg 2007, Güneralp & Seto 2013), and intensification of recreational usage (Venter et al. 2020).

To support cities in their conservational efforts to preserve biodiversity in urban areas and to identify beneficial landscapes and areas to prioritize for nature conservation, an understanding of the distribution of species and the identification of biodiversity hotspots are fundamental.

Given the need for a preservation strategy for urban biodiversity, which was also recently demonstrated by the results of the newly established Regional Biodiversity Strategy for the Ruhr Metropolis (Keil & Hering 2022) and the binding targets by the newly established Nature Restoration Law (European Commission 2022), more comprehensive approaches that consider landscape context and networks can contribute to the evaluation and delimitation of existing and future green spaces and conservational management (Zepp 2018).

1.1 Bird species as biological indicators in urban areas

Ideally, the quality of urban green spaces that supplement and maintain biodiversity and their necessary habitats should be evaluated in a quantitative manner. Therefore, taxa functioning as indicators for all levels of biodiversity, which are interconnected with the biotic community and which are relevant in terms of biodiversity, are favourable candidates for assessment (Vallecillo et al. 2016). The composition of avian communities reflects variations in habitat structure, vegetation characteristics, and human activities (Fig. 1). Thus, urban environments present various challenges (but also opportunities) for avian species, leading to complex patterns of habitat selection. The array of bird species inhabiting urban green spaces ranges from opportunistic urban adapters to highly specialized urban residents with specific habitat requirements. Their presence or absence, breeding success, and foraging behaviours serve as reliable indicators of habitat quality and ecological functionality within urban environments. Therefore, birds can be regarded as being an ideal research taxon (Mekonen 2017).

Birds occupy high trophic levels, with specific habitat and food requirements. They are easy to detect, observe, survey, and monitor (Vallecillo et al. 2016); moreover, they function as flagships for nature conservation (Smith & Sutton 2008, Veríssimo et al. 2014). Additionally, the distribution of bird species reflects that of other taxa with similar requirements in the same habitat (Eglington et al. 2012). As birds are highly mobile, they quickly react to changing environmental conditions by disappearing (Mekonen 2017). Due to the fact that they are directly and indirectly influenced by the urban environment, their distribution can be modelled by selecting drivers of habitat selection in urban areas (Isaksson 2018). Several biotic and abiotic factors drive the habitat selection process, including parameters such as habitat structural richness (Donnelly & Marzluff 2006, Huang et al. 2015), urban climatic anomalies (Cai et al. 2023), predation risk, food availability and anthropogenic disturbances from noise, light pollution, and general human presence (Ciach & Fröhlich 2017).

Different bird species often exhibit different preferences for habitat features, such as vegetation density, patch size, and proximity to water sources, which can provide suitable nesting sites and forag-

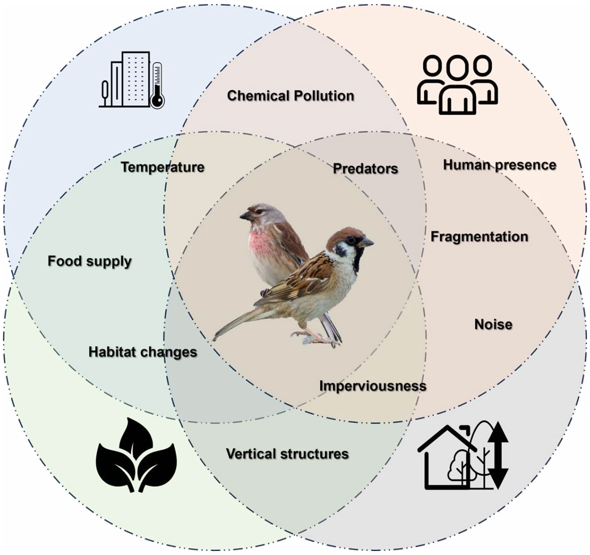

Fig. 1: Drivers of avian habitat selection in urban areas (based on Isaksson 2018)

ing opportunities (Hildén 1965). Additionally, the availability of resources, such as insects, seeds, and fruits, influences habitat selection, with some species showing adaptability to anthropogenic food sources such as bird feeders or food waste (Ottoni et al. 2009). Moreover, the level of predation risk, which is influenced by factors such as the presence of predators and human activities, shapes habitat selection patterns; however, bird community composition (*urban filter*), such as the presence of birds, balances the trade-off between foraging opportunities and escape distance and is influenced by environmental, biotic, and anthropogenic filters determined by urban environments (Aronson et al. 2016).

1.2 The concept of species distribution models

Species distribution models (SDMs) represent tools in ecological research that aid scientists in understanding and predicting the spatial distribution of various organisms across diverse landscapes (Hernandez et al. 2008). This concept involves the use of environmental variables to model potential habitats or geographical ranges that a given species may inhabit (Guisan et al. 2017). The initial definitions of the ecological niche were introduced by Elton (1927) and Grinnell (1928). Grinnell (1928) defined the niche as the specific place a species occupies in the environment. In contrast, Elton (1927) emphasized the functional role a species plays and its impact on the community. Both Grinnell's and Elton's definitions linked niches to environments. However, Hutchinson (1957, 1978) attributed the niche to the species itself, a definition that has become the most widely used and a fundamental concept in SDM theory. Hutchinson (1957) defined the fundamental niche as the sum of all physical and biological variables required for a species to persist. He accounts for competition and other negative interactions between individual species and within the environment. Thus, ecologists generally attribute the difference between the size of the fundamental and actually realized niches to negative interactions such as competition. With the years also positive interspecific interactions (Bruno et al. 2003), dispersal abilities (Pulliam 1988) and movement capacities of species (Hanski 1998) were incorporated in the understanding how the realized niche of species within a certain area are determined.

In recent years, the subsequent integration of advanced computational techniques, ecological theories and geoinformatics has significantly enhanced the precision and applicability of SDMs (He et al. 2015, Valavi et al. 2022). In general, SDMs are based on the principle that the distribution of a species is intricately linked to environmental factors which determine the potential of the species to occupy certain habitat types (Guisan & Thuiller 2005). SDMs became indispensable tools in ecology and conservation biology for predicting the geographic distribution of species for large areas (Miller 2010, Randin et al. 2020). Currently, RS data have revolutionized the field of landscape and animal ecology by providing spatially explicit environmental information at various scales (Roughgarden et al. 1991, Cavender-Bares et al. 2022). By that, remote sensing (RS) data already contribute to improve and accelerate the modelling and prediction of species' distribution in various ways (Randin et al. 2020). Additionally, the emergence of citizen science initiatives has enabled the collection of species presence-only data on a larger scale, thus complementing and even replacing traditional (but scarce) survey data (Steen et al. 2019). Despite their individual strengths, the combination of remote sensing and citizen science data might be a potent approach to enhance the applicability and predictive capacity of SDMs when used in urban environments (Randin et al. 2020).

1.3 Challenges of avian biodiversity research in the Ruhr Metropolis

Currently, fundamental research on avian biodiversity within the urban boundaries of the Ruhr Metropolis is scarce and has primarily focused on certain species (Kowallik & Rautenberg 2014) or on locations and land use aspects (Hamann 1991, Keil 2019). This focus hinders the identification of contributions and the importance of green space networks for species, as well as the analysis of regional biodiversity hotspots (Engler et al. 2017).

A broader research scope, which is necessary to prioritize nature conservation areas and their inherent biodiversity-relevant landscape structures, is lacking due to several constraints. These include financial and staff support limitations, time restrictions, and a lack of acknowledgement for urban biodiversity in policy, as recognized by the Regional Biodiversity Strategy Ruhr Metropolis (Keil et al. 2021). As such constraints impede methods that rely on exhaustive field surveys and mapping, modelling species distributions using widely reported presence-only species records and freely available RS data and their derivates has the potential to identify important areas for biodiversity in the Ruhr Metropolis.

Therefore, the utility of SDMs is mainly limited by the existence of reliable species occurrence data and spatially explicit environmental data. Typically, extensive survey data on taxa in urban areas are seldom available. However, the substantial amount of unstructured data reported by citizens, so called Citizen Science (CS) data, can provide a remedy in this scenario. Preconditions in urban areas are advantageous due to the high concentration of active reporters (Dachverband Deutscher Avifaunisten 2018).

The Ruhr Metropolis has developed a unique and specific biodiversity which stems from a combination of its historical development, specific location, and unique transformations (Keil et al. 2013). By proactively preserving and encompassing industrial brownfields into green infrastructure (Knapp 1998, Angelo 2019), the study area serves as a prime example for analysing its significance for and contribution to urban biodiversity. These distinct inner-city industrial wastelands with characteristic biocoenoses ('industrial nature'), large areas of succession, hydromorphic subsidence areas and the location at the interface between low mountain ranges and lowlands give the biodiversity of the Ruhr Metropolis its very own characteristics (Keil 2019). The variety of habitats in a small area leads to a high diversity of species that is far more pronounced and led to a development of biotic communities that do not exist in near-natural habitats. At the same time, the range of pressures and threats to biodiversity in the Ruhr region is particularly wide: land-use changes and structural change with high land requirements are increasingly leading to the disappearance of characteristic biotic communities. Contaminations in soils and water bodies are usual remnants and the density of traffic routes and sealed surfaces exceptionally high compared to other regions in Europe (Keil & Hering 2022). Global phenomena such as climate change, insect decline, and structural and financial deficits in nature conservation are also relevant to the Ruhr region and are even more pronounced than elsewhere (Keil et al. 2021).

By considering the current challenges and existing research gap within the Ruhr Metropolis, we conducted a study encompassing the entire Ruhr Metropolis and including multiple bird species to address the limitations of site- and species-specific considerations. Therefore, we used CS-derived SDMs in combination with environmental data to predict the occurrence of 26 endangered avian species (Tab. 1) in the Ruhr Metropolis based on data from breeding periods over three consecutive years between 2019 and 2021. The objectives of this study were as follows:

- To elaborate, whether the combination of CS and RS with SDMs can be effectively used to predict species distributions in spatially complex urban environments; and
- to identify biodiversity hotspots and prioritize areas for nature conservation within the urban environment.
- To examine, to what extent the existing network of protected areas overlays with species' hotspots

2 Materials and methods

2.1 Study area

The current study was conducted within the Ruhr Metropolis, which is located in North Rhine-Westphalia, Germany. The Ruhr Metropolis is situated in the western part of Germany in Central Europe and encompasses an area of approximately 4500 km^2 (Fig. 2). With its approximately five million inhabitants, it stands as the fifth largest conurbation in Europe. Extending from the Lower Rhine Basin in the west to the Westphalian Plain in the north and the Rhenish Massif in the south, the Ruhr Metropolis boasts a polycentric and administratively fragmented structure. Formerly known for coal mining and heavy industry, an ongoing transformational change towards a knowledge-based society is currently leading to severe land use changes. Divided into a densified core area (2,280 inhabitants per km^2) and a more scattered surrounding area dominated by settlement, agriculture and forest patches, the Ruhr Metropolis embodies a diverse and dynamic urban area (Regionalverband Ruhr 2021).

Hence, the core of the Ruhr Metropolis is well endowed with blue and green infrastructure, such as parks, meadows, ponds, small woods and reservoir lakes (Zepp et al. 2020). The entire area is interspersed with a green space network that has been prioritized for nature conservation. Predominantly, the green space network is outlined by the higher-level rivers Rhine, Ruhr and Lippe in the periphery of the core area and within the core area by the formerly defined old greenbelts and the Emscher River.

Tab. 1: Number of occurrences (records) derived from CS. Datasets for the 2019-2021 breeding season (Apr. to Jul.) for the investigated bird species (source: ornitho.de) before and after spatial thinning. Additionally, mean values for TSS, AUC, bias, and the Boyce index were calculated for all of the algorithms performed per species. Additionally, the mean specificity and mean sensitivity over the TSS and ROC curve are depicted.

Scientific Name	**Common Name**	**Abbr.**	**Records pre-thinning**	**Records post thinning**	**Mean**					
					TSS	**ROC**	**Bias**	**Boyce**	**Specificity**	**Sensitivity**
Accipiter gentilis	Northern Goshawk	AccGen	561	253	0.70	0.91	0.99	0.70	0.83	0.87
Accipiter nisus	Eurasian Sparrowhawk	AccNis	951	386	0.64	0.87	0.99	0.64	0.80	0.84
Acrocephalus scirpaceus	Eurasian Reed Warbler	AcrSci	985	189	0.79	0.95	0.99	0.79	0.88	0.92
Alauda arvensis	Skylark	AlaArv	2086	325	0.73	0.92	0.99	0.73	0.83	0.90
Anthus pratensis	Meadow Pipit	AntPra	1111	253	0.73	0.92	0.99	0.73	0.85	0.89
Anthus trivialis	Tree Pipit	AntTri	924	230	0.77	0.94	0.99	0.77	0.86	0.92
Ardea cinerea	Grey Heron	ArdCin	4131	730	0.61	0.86	0.99	0.61	0.77	0.84
Buteo buteo	Common Buzzard	ButBut	4403	1047	0.54	0.82	0.99	0.54	0.77	0.77
Carduelis cannabina	Eurasian Linnet	CarCan	1791	395	0.67	0.89	0.99	0.67	0.82	0.84
Charadrius dubius	Little Ringed Plover	ChaDub	2261	218	0.83	0.96	0.99	0.83	0.90	0.93
Corvus frugilegus	Rook	CarFru	1131	192	0.82	0.96	0.99	0.82	0.88	0.95
Cuculus canorus	Common Cuckoo	CucCan	1467	396	0.74	0.93	0.99	0.74	0.85	0.89
Delichon urbicum	Common House Martin	DelUrb	2034	508	0.63	0.87	0.99	0.63	0.80	0.83
Dendrocopos medius	Middle Spotted Woodpecker	DenMed	307	177	0.80	0.95	0.99	0.80	0.88	0.93
Dryobates minor	Lesser Spotted Woodpecker	DryMin	197	106	0.89	0.98	0.99	0.89	0.93	0.96
Falco subbuteo	Eurasian Hobby	FalSub	561	187	0.77	0.94	0.99	0.77	0.87	0.90
Falco tinnunculus	Common Kestrel	FalTin	3359	865	0.56	0.83	0.99	0.56	0.77	0.79
Hirundo rustica	Barn Swallow	HirRus	3628	780	0.60	0.85	0.99	0.60	0.78	0.82
Lullula arborea	Wood Lark	LulArb	361	89	0.96	0.99	1.00	0.96	0.97	0.99
Luscinia megarhynchos	Common Nightingale	LusMeg	1756	392	0.76	0.93	0.99	0.76	0.84	0.91
Milvus milvus	Red Kite	MilMil	1837	740	0.59	0.85	0.99	0.59	0.77	0.82
Passer montanus	Eurasian Tree Sparrow	PasMon	445	90	0.92	0.99	1.00	0.92	0.95	0.98
Phoenicurus phoenicurus	Common Redstart	PhoPho	1132	367	0.67	0.89	0.99	0.67	0.83	0.84
Streptopelia turtur	European Turtle Dove	StrTur	134	39	0.97	1.00	0.99	0.97	0.97	1.00
Sturnus vulgaris	Common Starling	StuVul	3181	646	0.60	0.86	0.98	0.60	0.78	0.82
Vanellus vanellus	Northern Lapwing	VanVan	2833	287	0.76	0.94	0.99	0.76	0.87	0.89

2.2 Species distribution modelling

In this study, we employed a comprehensive workflow (Fig. 3) to integrate CS- and RS-derived data sources into SDMs.

The workflow encompasses the steps of data preselection and processing, variable selection, model training, and validation, all of which are tailored specifically for presence-only occurrence data of rare species. Modelling species distribution was finally done within the biomod2 package (THUILLER et al. 2023).

2.2.1 Citizen science species occurrence datasets

CS datasets from *ornitho.de* (DDA e.V., www.ornitho.de) were utilized to generate presence-only records for modelling. The database *ornitho.de* is a nationwide platform in Germany, which enables citizens to report bird observations throughout the country. It is regarded as the most comprehensive and extensive database for bird observations in Germany. Citizens have the option to either report observations as incidental (unstructured) records or perform a semi-structured approach by submitting

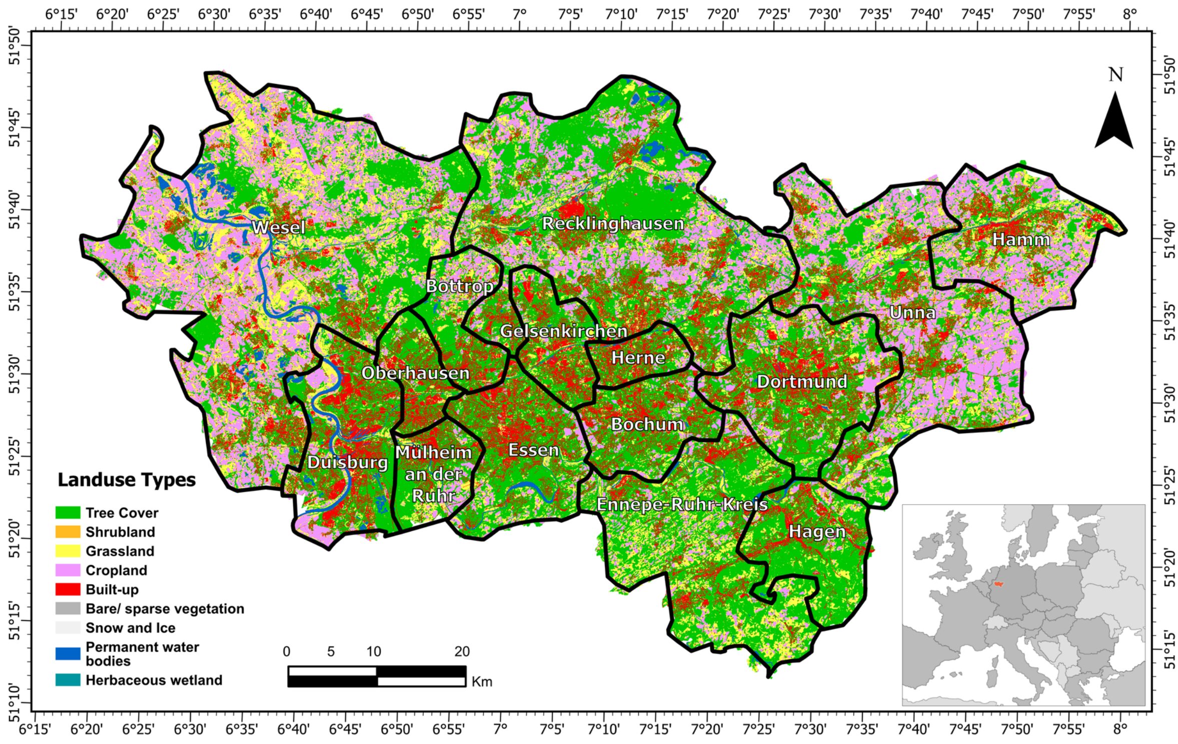

Fig. 2: Characterization of land use within the Ruhr Metropolis based on WorldCover land use classification (Zanaga et al. 2022)

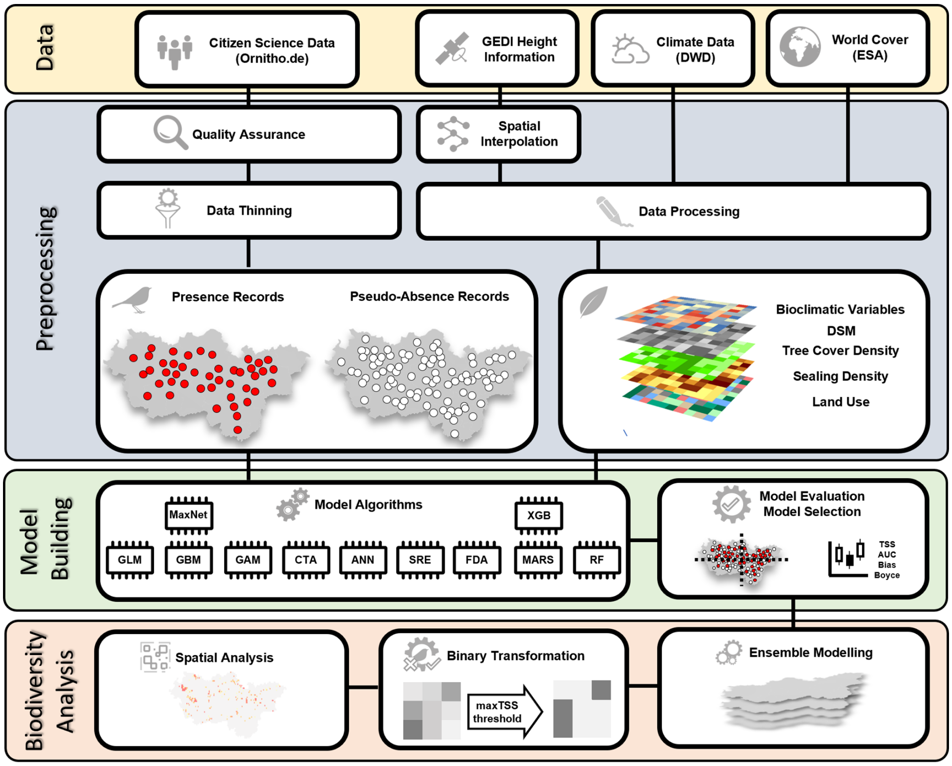

Fig. 3: Workflow for species distribution modelling with biomod2

a complete checklist of observations for a specific location. Both types of data are assigned to precise coordinates or are allocated to the centre of a grid cell with an approximately 1 km x 1 km raster size. From all of the recorded species, including semi- and unstructured data, we selected 26 species (Tab. 1), which are collectively referred to as *species relevant for planning* (Kiel 2005) because they are known to be vulnerable to urbanisation in the form of land sealing, densification and conversion of natural or semi-natural areas. Additionally, 19 out of these 26 species are listed in the Red List of North Rhine-Westphalia (Sudmann et al. 2023), 16 are in categories 1-3 (threatened by extinction, severely threatened, or threatened). We included several orders of birds, including (but not limited to) waders, passerines, woodpeckers, and different orders of birds of prey. Species of these orders differ in habitat preferences based on landscape composition requirements, as well as climate preconditions and habitat structure needs (Hildén 1965).

2.2.2 Data preprocessing and pseudo-absence data

All of the data were processed by using basic R, version 4.2.2 (R Core Team 2022), unless otherwise specified. To enhance reliability, we conducted data preprocessing. We included only species data records from three consecutive years (2019 to 2021) to align precisely with available local climate and RS datasets. To avoid distortions in habitat selection analysis due to migratory patterns, we confined our data to records from April to July, which represents the main breeding season for all bird species under study in North Rhine-Westphalia (Südbeck et al. 2005). Additionally, we only included records with a reported breeding code in that specific timespan (Keller et al. 2020). To avoid underestimating the occurrence of rare and difficult to detect species we did not discard category A (possible breeding). By incorporating both, breeding codes and restriction on breeding time we minimize possible misattributions throughout the years and better align the occurrences to the climate predictors of the breeding season.

Subsequently, we used the R package *spThin (*version 0.2.0) (Aiello-Lammens et al. 2015) to thin the species presence-only datasets and to reduce spatial autocorrelation where necessary (Tobler 1970, Legendre & Fortin 1989). However, thinning datasets of rare species with few records may lead to an incomplete representation of the species' ranges, decreased model performance and suboptimal sample sizes (Guisan et al. 2017, Coudun & Gégout 2006, Jiménez-Valverde et al. 2009, Steen et al. 2021). To address species-dependent spatial autocorrelation and unbalanced thinning, we computed the range of the semivariogram for each species dataset by using a spherical model (Fletcher & Fortin 2019, Bebber 1999). The range of the semivariogram computed with the R package *gstat* (Pebesma 2004) was then used as the range input in *spThin* to thin each corresponding species dataset (Fig. 4 and Tab. 1).

In accordance with the current literature recommendations regarding pseudo-absences for different modelling algorithms (Barbet-Massin et al. 2012) and following guidance from biomod2 developers (Thuiller et al. 2023), we standardized the number of pseudo-absences for each algorithm based on three times the number of presences post-thinning (Tab. 1). Each raster cell, which is equipped with a presence record, is excluded from pseudoabsence generation. To mitigate potential misallocation issues, multiple independent pseudo-absence datasets should be utilized for each species. As SDMs inherently over- or underpredict species' ranges according to prerequisites and assumptions made in generating pseudo-absence data. Also, density and arrangement of occurrence and pseudo-absence data affect model results and transferability as many studies highlighted lately (Barbet-Massin et al. 2012, Grimmett et al. 2020). The generation and thinning methods are highly dependent on the research question. Here, we aim to examine the fundamental niche of each species in the study area. To determine the overall habitat each corresponding species may inhabit an overprediction bias is preferably over an underestimation of the occurrence range. Thus, we chose ten independent randomly distributed pseudo-absence dataset iterations within the whole study area rather than using a disk or buffer approach which restricts the generation of pseudo-absences to a certain area around the presences or favourable environmental conditions (Márcia Barbosa et al. 2013, Thuiller et al. 2023, Whitford et al. 2024).

2.2.3 Environmental predictors

Climate raster data from the German Weather Service (DWD) were used to calculate bioclimatic variables (Tab. 2). The climate datasets are a synthesis of weather station data and RS supported interpolations to create complete raster datasets. We incorporated data for the referenced breeding period (April to July) spanning from 2019 to 2021. A

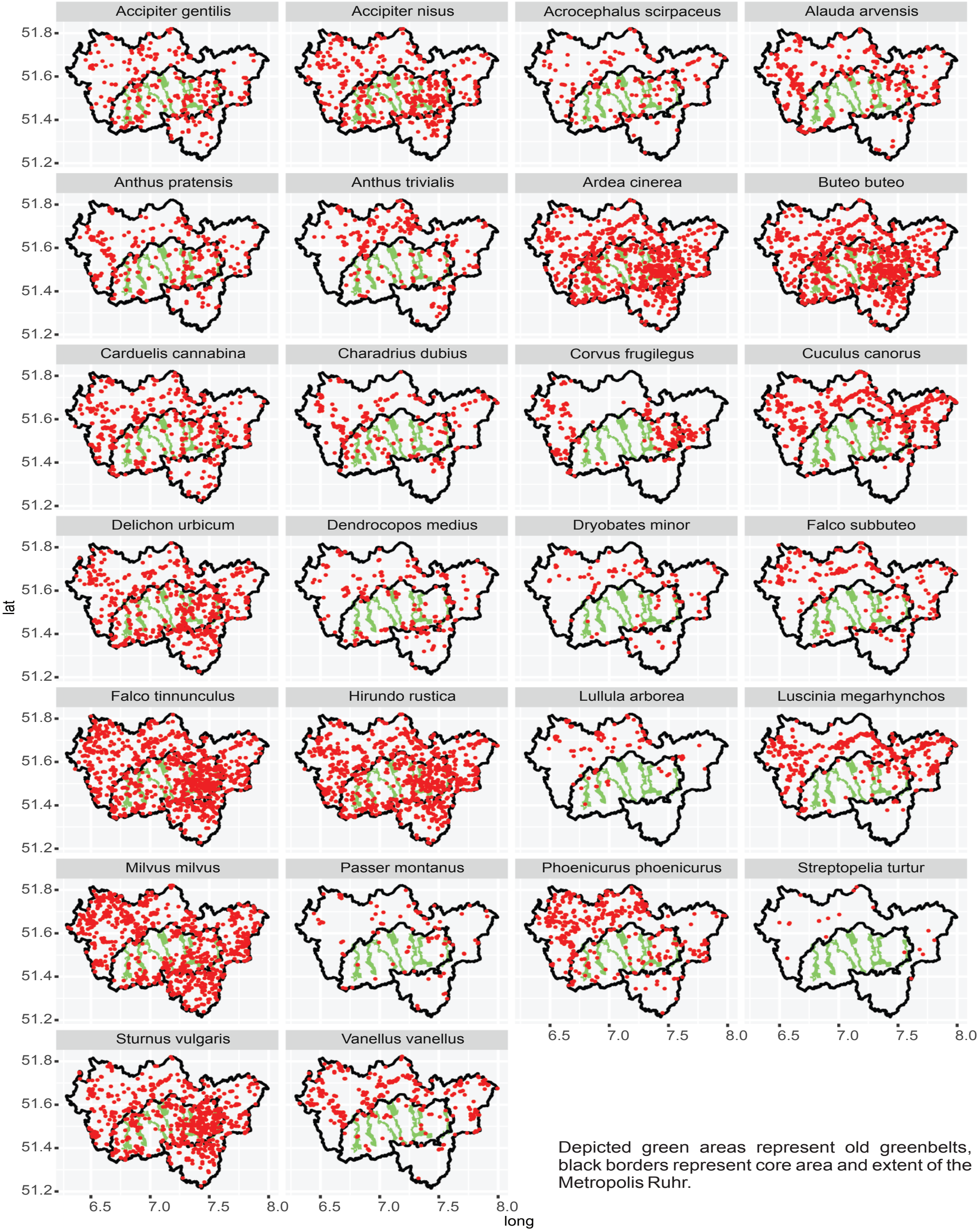

Fig. 4: Records after thinning of the investigated 26 avian species

subset of these variables served as climatic explanatory variables reflecting known influencing factors on the habitat selection of birds. Maximum and minimum temperature (bio5 and bio6), as well as maximum and minimum precipitation (bio14 and bio16) determines physiological limits, breeding success, range, and foraging behaviour (Root 1988, Thomas et al. 2004, Jiguet et al. 2006). In addition,

Tab. 2: Covariate variables for SDM perfomance

Code	Covariate Definition	Time Resolution	Scale	Original Resolution	Resampling Method	Source
bio5	*Maximum temperature in breeding periods*	monthly mean	10^2 °C	1x1 km	cubic	DWD Climate Data Center n.d. a
bio6	Minimum temperature in breeding periods	monthly mean	10^2 °C	1x1 km	cubic	DWD Climate Data Center n.d. b
bio14	Minimum precipitation in breeding periods	monthly sum	millimetres	1x1 km	cubic	DWD Climate Data Center n.d. c
bio16	Maximum precipitation in breeding periods	monthly sum	millimetres	1x1 km	cubic	DWD Climate Data Centern.d. c
bio20	*Potential evapotranspiration in breeding periods*	monthly mean	10^2 millimetres	1x1 km	cubic	DWD Climate Data Center n.d. e
bio21	*Real evapotranspiration in breeding periods*	monthly mean	10^2 millimetres	1x1 km	cubic	DWD Climate Data Center n.d. d
SEAL	Imperviousness	static	0-100%	100x100 m	cubic	European Environment Agency 2020a
TCD	Tree cover density	static	0-100%	100x100 m	cubic	European Environment Agency 2020b
GEDI	Global Ecosystem Dynamics Investigation height information	multiple dates across years	meters	Point data	cubic	Dubayah et al. 2021b
p_forest	*Proportional contribution of forest class to each 1x1 km grid cell*	static	0-1	10x10 m	AGG	Zanaga et al. 2022
p_shrub	Proportional contribution of forest class to each 1x1 km grid cell	static	0-1	10x10 m	AGG	Zanaga et al. 2022
p_grass	Proportional contribution of grassland class to each 1x1 km grid cell	static	0-1	10x10 m	AGG	Zanaga et al. 2022
p_crop	Proportional contribution of cropland class to each 1x1 km grid cell	static	0-1	10x10 m	AGG	Zanaga et al. 2022
p_built	*Proportional contribution of built-up area class to each 1x1 km grid cell*	static	0-1	10x10 m	AGG	Zanaga et al. 2022
p_soil	Proportional contribution of bare soil class to each 1x1 km grid cell	static	0-1	10x10 m	AGG	Zanaga et al. 2022
p_water	Proportional contribution of permanent water surface class to each 1x1 km grid cell	static	0-1	10x10 m	AGG	Zanaga et al. 2022
p_wetland	Proportional contribution of wetland class to each 1x1 km grid cell	static	0-1	10x10 m	AGG	Zanaga et al. 2022

Note: Variables highlighted in italics were excluded for SDM after pairwise Pearson correlation and stepwise VIF. AGG = aggregation.

we considered data on potential and real evapotranspiration (bio20 and bio21). Evapotranspiration is directly linked to water availability and capacity in ecosystems, reflecting vegetation and habitat quality status, resource availability and microclimate conditions (Zepp et al. 2023, Symonds & Johnson 2008). We further integrated land surface characteristics (Tab. 2) and height information to be known for their direct impact on habitat selection and structure of species in urban areas (Ferenc et al. 2014, Blair 1996). Height information was retrieved from GEDI LiDAR datasets (Dubayah et al. 2021a) which were computed in Google Earth Engine (Gorelick et al. 2017) and subjected to Kriging interpolation to generate a digital surface model (DSM). Data on imperviousness (SEAL) (European Environment Agency 2020a) and tree cover density (TCD) (European Environment Agency 2020b) were obtained from Copernicus Land Cover Monitoring. Finally, land use information was sourced from the ESA WorldCover based on 2021 (Zanaga et al. 2022).

Where necessary, data were resampled via cubic convolution for continuous data to a 1x1 km grid cell size (Tab. 2) by using the *Whitebox* r package (Lindsay 2016). We chose cubic convolution over other methods, such as bilinear interpolation, for continuous datasets due to its higher predictive accuracy for point and raster image resampling (Keys 1981, Smith et al. 2004). We disaggregated the land use categories that were applicable to the study area into eight distinct layers. After aggregating 10,000 10x10 m raster cells, each 1x1 km cell represents the proportional aerial contribution of a specific land use class. The applied minimum and maximum values per grid cell reflect the absolute measured climatic parameters over the span of three consecutive years. To account for multicollinearity, we performed a pairwise Pearson correlation and computed the variance inflation factor (VIF) of each pair of variables (Fig. A1). Due to the fact that variables with a high correlation may lead to unstable estimates in SDMs, we only retained variables with a Pearson correlation ≤ 0.5 and a stepwise variance inflation factor (VIF) ≤ 3 to exclude moderately to highly correlated covariates for further analysis (Kock & Lynn 2012, Naimi & Araújo 2016). Thus, from the initially selected climatic variables, the maximum temperature (bio5) and both evapotranspiration datasets (bio20, bio21) were removed. Moreover, p_forest and p_built were also omitted because they represent land use-derived variables. Ultimately, three bioclimatic variables and nine land use-derived variables were selected for inclusion in the SDM analyses (Tab. 2).

2.2.4 SDM construction, evaluation and selection

We used a multiple modelling approach that is available within the biomod2 package to individually build presence-only SDMs for all 26 species (Tab. 1), thus associating species occurrences with the prior selected environmental predictors (Guisan et al. 2017).

Eleven different modelling algorithms were used to capture the broad variability in statistical behaviour and algorithm categories (Fig. 3) (Meller et al. 2014). Algorithms can be categorized as regression-based, machine learning, enveloping, and classification-based algorithms, whereas some algorithms combine the categories (Tab. 3). Regression-based algorithms generate linear and nonlinear equations between presence-absence data and environmental predictors. Machine learning algorithms directly generate environmental spaces using all of the provided input datasets. Classification algorithms successively divide data into homogeneous partitions, whereas enveloping methods investigate provided environmental conditions at the location of presence–absence data and search for similar grid cells (Merow et al. 2014).

To determine the most effective models, we performed a preliminary analysis by using the methodology of Carroll et al. (2023) on all modelling algorithms that were implemented in the biomod2 package. We selected four representative species covering different habitat preferences, sample sizes and differences in commonness and behaviour (Fig. A2). We retained four out of eleven models due to their

Tab. 3: Investigated SDM algorithms in the biomod2 package with method references

Code	Algorithm	Category	Model reference
GLM	*Generalized linear model*	Regression-based	McCullagh 1984, Venables & Ripley 2011
GBM	Generalized boosted model	Machine-learning	Ridgeway 1999
GAM	Generalized additive model	Regression-based	Hastie & Tibshirani 1986, Yee & Mitchell 1991
CTA	*Classification tree analysis*	Classification/ Machine-learning	Breiman et al. 1984
ANN	*Artificial neural network*	Machine-learning	Ripley 2008, Hilbert 2001
SRE	*Surface range envelope*	Enveloping	Busby 1991
FDA	*Flexible discriminate analysis*	Classification	Hastie et al. 1994
MARS	*Multivariate adaptive regression splines*	Regression-based/ Machine-learning	Elith & Leathwick 2009, Friedman 1991
RF	Random Forest	Machine-learning	Breiman 2001
MAXNET	*Maximum Entropy*	Machine-learning	Phillips et al. 2006
XGBOOST	Extreme Gradient Boosting	Machine-learning	Chen & Guestrin 2016

Note: Dismissed models for modelling are highlighted in italics.

better predictive performance for a majority of the pretested species (Tab. 3). These models include the generalized boosted model (GBM), generalized additive model (GAM), random forest (RF) and extreme gradient boosting (XGBOOST).

For spatial cross-validation and to achieve a reduction in spatial sample bias (Valavi et al. 2019), we applied the *block* method adapted from the blockCV package recently implemented in the biomod2 package. This method partitions data into four separate bins. The bins are selected by splitting the dataset based on the lines of latitude and longitude, which equally divides the occurrences (Muscarella et al. 2014). The *block* method reduces data dependencies and spatial autocorrelation, thus offering advantages over random cross-validation methods (Roberts et al. 2017). In alignment with prior research recommendations, multiple test statistics were used to provide a more robust assessment of model performance and to validate model responses (Guisan et al. 2017, 2021, Baker et al. 2024).

Thus, model performance was evaluated by using the area under the receiver operating characteristic curve (ROC), true skill statistic (TSS), frequency bias index (bias), and continuous Boyce index (Hirzel et al. 2006). The AUC (area under the curve) value of ROC method is known to be an effective measure for evaluating the performance of ordinal score models and additionally serves as a threshold-independent measure of accuracy (Guisan & Thuiller 2005). The ROC index (ranging from 0 to 1 = highly accurate prediction) defines the probability that an SDM will rank a presence locality higher than a pseudo-absence. Nevertheless, the ROC value is highly ineffective when it comes to presence-pseudoabsence models as a stand-alone metric, being strongly influenced by the amount of pseudoabsence points generated (Lobo et al. 2008, Whitford et al. 2024). TSS is a threshold-dependent measure of accuracy. The TSS ranges between -1 and 1. Moreover, it tests the agreement between the expected and observed distributions and the probability of the outcome being predicted by chance alone (Allouche et al. 2006). TSS values of +1 are considered to indicate perfect agreement between the observed and expected distributions. Values below 0 indicate that the predictive performance of the models is no better than random (Fielding & Bell 1997). We selected the TSS over Cohen's kappa statistic due to its robustness against prevalence (Allouche et al. 2006). Bias is the ratio of the frequency of valid forecasts to the frequency of valid observations. It indicates whether the model predictions have a tendency to under- (bias <1) or overpredict (bias >1) (Wunderlich et al. 2019). The Boyce index is a presence-only evaluation method that measures a model's ability to predict distributions other than random distributions of observed presences across prediction gradients (Boyce et al. 2002). On a scale from -1 to +1, positive values indicate consistent models. Values close to zero indicate random predictions, whereas negative values indicate counter predictions (Hirzel et al. 2006). By that, the Boyce index measures the model's habitat suitability after it has projected into geographic space, making it an appropriate method for evaluating the model's transferability outside of its training extent (Petitpierre et al. 2017).

2.2.5 Ensemble model arrangement and binary transformation

For ensemble modelling (EM), only models exceeding the thresholds of TSS > 0.5 and ROC > 0.7 were considered. After threshold filtering, all of the SDM algorithms were combined in an EM to assess model congruence and improve model accuracy (Fig. 3). For each species, one individual EM was computed. For the EM output, we used a committee averaging (CA) algorithm. CA is both a predictive distribution model and a measure of uncertainty. The CA approach uses thresholded binary predictions that are individually calculated for each species and across all applied models to predict species presence (1) or absence (0). Raster cells with values between 0 and 1 show a ratio of uncertainty in the EM definition of species presence-absence (Caruana et al. 2004). EMs were evaluated by using specificity and sensitivity metrics. Specificity refers to the proportion of true positives that are correctly identified, whereas sensitivity accounts for the proportion of true-negatives that are correctly identified (negative refers to pseudoabsences instead of real absences in the case of presence-only data). Both test variables are reported as percentages and represent the accuracy of the prediction of EMs. Higher EM values correspond to a better ability of EM to predict the real occurrence or absence of each species (Allouche et al. 2006, Liu et al. 2011).

For species richness analysis, binary transformations of all of the EMs were performed. By applying a species-specific value maximizing TSS along different thresholds (maxTSS) for the raster values obtained from the CA, each EM output was transformed into a binary presence-absence map (Allouche et al. 2006). All of the grid cells with values above the species-specific threshold were treated as suitable, and all of the grid cells with values below the threshold were treated as not suitable. After transformation, all of the species grids were overlaid to perform species richness analysis (Scott et al. 1987, Gotelli & Colwell 2001). The calculation of quantiles was performed by using the *terra* package (Hijmans 2022).

3 Results

3.1 Main drivers of bird distribution in the Ruhr Metropolis

For all 26 considered species, all models performed well, with a mean ROC ≥0.82 (Fig. 5a) and a TSS ≥0.54 (Fig. 5b).

An exception is represented by *Streptopelia turtur*. The small number of available species occurrence points indicated that the GAM algorithm could not produce successful calculations, as the model splits surpassed the number of available presence records (Fig. A3a, A3b). Despite this exception, the other models achieved good predictive performance. Species exhibited differences in model accuracy across all of the models and between TSS and ROC (Tab. 1 and Fig. A3a, A3b). Models for birds of prey, such as *Buteo buteo*, *Milvus milvus* and *Falco tinnunculus,* showed lower performance than did models for larks or woodpeckers and exhibited a greater degree of variability across different model iterations.

The sensitivity and specificity for species' EMs were ≥0.77, thus suggesting that the ensemble models correctly classified the presence and absence of multiple species. All of the SDMs yielded bias scores near 1, thus indicating an unbiased prediction in which the predicted occurrence is forecasted almost exactly as often as it has been observed. The mean continuous Boyce index ranged from 0.54 to 0.97, thus showing an overall consistency of the present predictions with the distribution of presences in the evaluation dataset (Tab. 1) and hints that the models perform well beyond training extent.

Variable importance varied greatly across species and variables, but in general precipitation, tem-

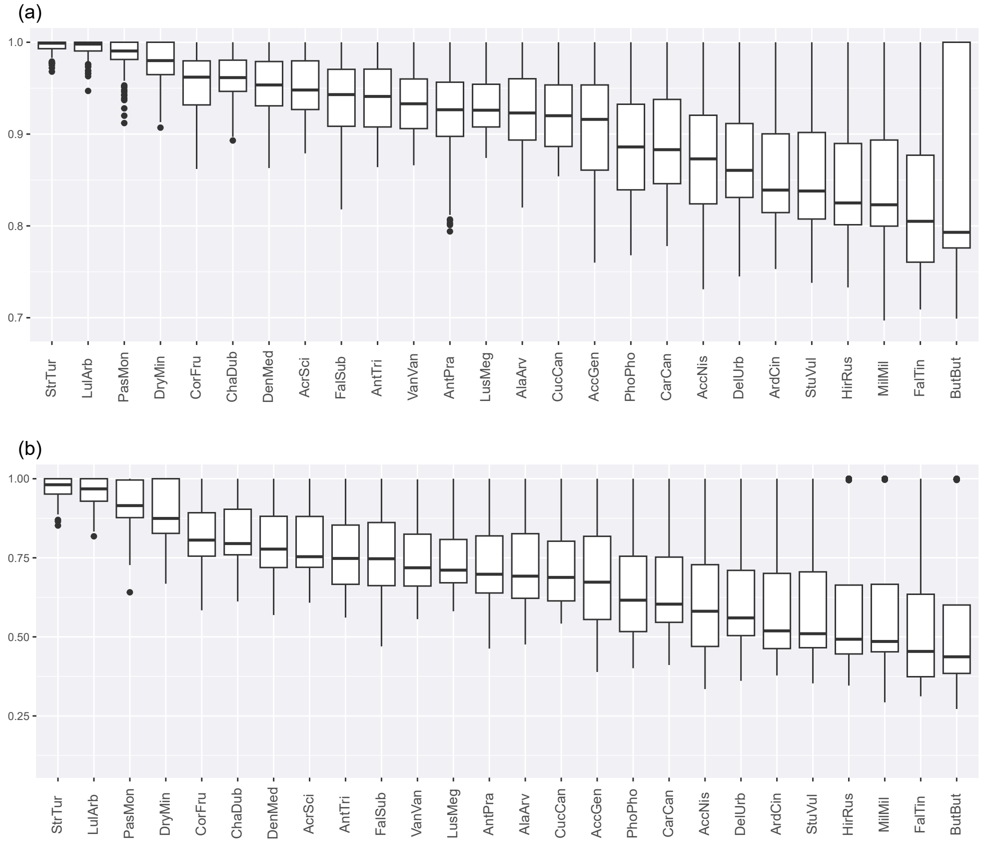

Fig. 5: Model performance metrics (a) ROC and (b) TSS individually per species over each of the model algorithms and runs

perature and the land use proportions for grassland, water and wetland had higher importance than others across all four applied algorithms (Fig. A4). Species individual predictions (Fig. 6) showed a greater dependence of water bound birds (Tab. A1) on p_water and p_wetland. Ground-close breeding grassland bird species were relying predominately on p_grass and p_soil. Height information from GEDI and the proportion of shrubland showed to have less importance for predictions for the majority of all species. Sealing intensity showed only importance for forest breeding birds. Especially model predictions on species associated to urban surroundings showed a greater dependence on temperature (bio6).

3.2 Distribution of species richness and identification of biodiversity hotspots

The combined map of all of the species presence/absence results from binary transformation and details on the frequency distribution of grid cells by total species richness (TSR) with the contribution of each TSR quantile are presented in Fig. A5. The results of the analysis of the top richness quantiles are shown in Fig. 7, overlaid with protected areas and the old greenbelts in the core area of the Ruhr Metropolis. Notably, hotspots based on 5% TSR (Q95) encompass six percent (250 km^2) of the Ruhr Metropolis, whereas Q90 covers 10.5% (441 km^2), and Q85 includes 17% (696 km^2). In total, 116 km^2

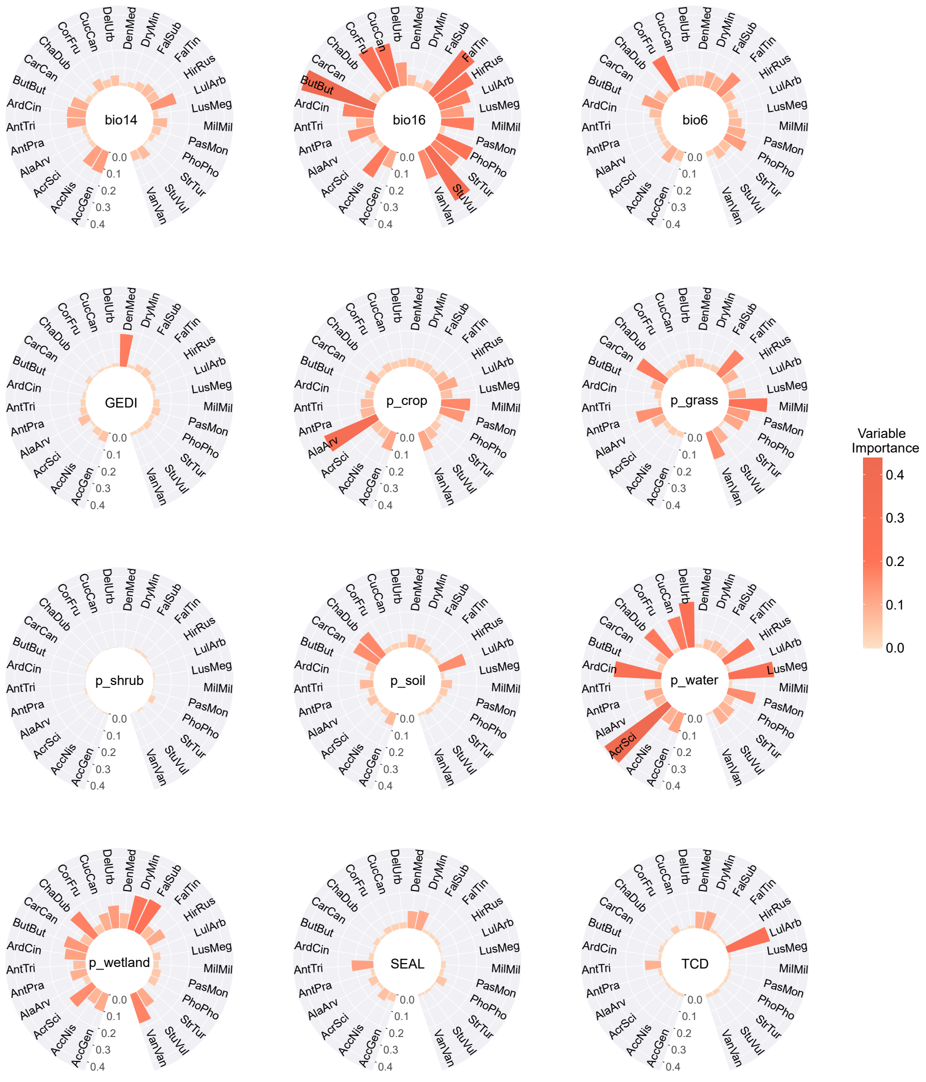

Fig. 6: Median variable importance scores per species across all algorithms

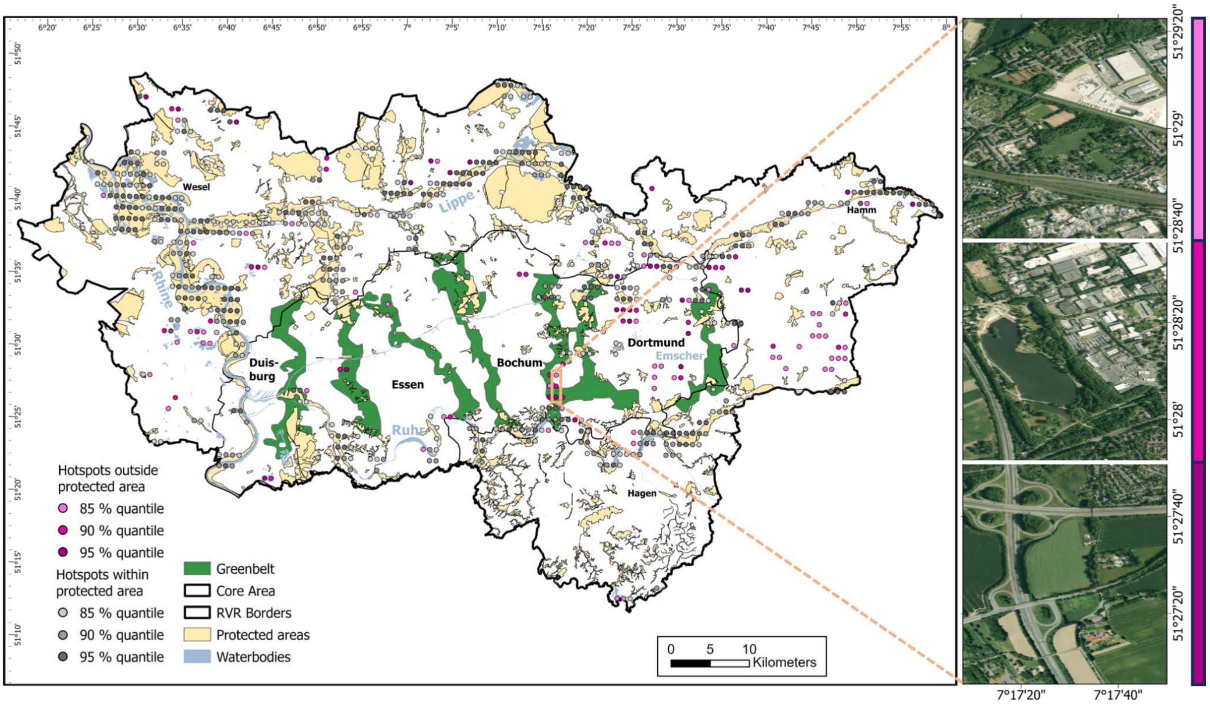

Fig. 7: Species richness analysis of the Ruhr Metropolis Hotspot analysis by species richness quantiles (Q_{SR}) of the 0.95 (Q_{95}), 0.90 (Q_{90}), and 0.85 (Q_{85}) quantiles. Grey shadings depict hotspots inside legally protected areas, red shadings hotspots outside. Water bodies of the Ruhr Metropolis are depicted in blue. Green space network composed of old greenbelts (green) and biotope network areas (beige) of special interest to conservation (protected areas). The biotope network consists of nature and landscape protection areas, as well as other legally protected biotopes important for biodiversity and preservation of connectivity. The core area depicts the densely populated area of the Ruhr Metropolis (data modified from Zepp 2018 and OpenGeodata.NRW). The map strip shows one part of the remnants of the greenbelts with high species richness evaluation.

of identified species richness areas are not covered by protected areas, including 28 km² of Q95. An example of unprotected area with potentially high species richness is shown in the map strip of Figure 7. The area is one of the remnants of the greenbelts in the south of the city of Bochum, consisting of a diverse composition of different land uses and habitat structures, including water bodies, agricultural fields and forest relicts. Although a high number of raster cells with high species richness are located outside of the core area of the Ruhr Metropolis, it is evident that even densely urbanized sectors of the Ruhr Metropolis have the potential to support a majority of the species that were examined, even though these areas are currently not under legal protection. Outside of the core area high species richness is especially common along the river axes of the Rhine, Ruhr and Lippe Rivers, with their diverse and extensive landscapes, comprising extensively used agricultural fields and grassland but also renatured wetland and forested areas. Within the central core area, species richness manifests in small patches dispersed throughout the entire metropolis region. In particular, the eastern section benefits from the old regional greenbelts that act as valuable habitat corridors for a substantial proportion of avian biodiversity.

4 Discussion

The combination of CS datasets and RS data in SDMs has not been widely used to analyse dense urban surroundings in terms of habitat suitability for bird species. This study demonstrated that citizen science-powered SDMs combined with datasets derived from remote sensing offer a sophisticated and accurate method for predicting avian species diversity within urban landscapes. Our findings confirm that SDMs have considerable predictive capabilities for biodiversity assessment at a fine scale in urban environments, despite relying solely on CS data without structured survey datasets.

Previous studies have shown that the combination of RS and SDMs is favourable for use in urban settings to differentiate the heterogeneity of urban landscapes and compensate for the lack of researchers' access to certain areas, such as private gardens and inner courtyards (Rocchini et al. 2010, Rocchini et al. 2016).

Although spatial thinning, spatial cross-validation and model evaluations were cautiously performed to minimize unwanted dependencies and inaccuracies, the reliance on CS datasets introduces potential biases derived from uneven spatial coverage and varying reporting intensities (Feldman et al. 2021). These biases may affect the overall representativeness of the avian species distribution in the study area. In general, when investigating single-species models, interspecies dependencies can not be detected, which may have biased species habitat suitabilities in this study (Araújo & Guisan 2006).

Nevertheless, overall satisfying test metrics of ROC, TSS and Boyce index for all species and models confirm good predictive power and rather few potential biases. Intentionally, we chose random sampling of pseudo-absences, variogram thinning and block cross validation to minimise spatial dependencies and to rather over- than underpredict rare species and their fundamental niches in the study area (Valavi et al. 2019, Whitford et al. 2024, Márcia Barbosa et al. 2013). The bias score hints that the models are far from over-prediction and rather tend to underpredict. As the bias scores tend to be near 1 for all species, biases of under- or overprediction are neglectable, though. We were able to show that the combination of both CS and RS is sophisticated in urban settings (such as the Ruhr Metropolis) on a fine spatial scale. Our approach yielded high accuracy and predictive power for all 26 planning-relevant species that were investigated. Although accuracy and performance varied among species and algorithms, overall predictions of species distribution patterns were deemed reliable for further analysis (Guisan et al. 2017). Species with distinct land use preferences and consistent nesting site requirements, such as woodpeckers or skylarks, were predicted with high precision, whereas species with a broader range of habitat preferences and larger range sizes as well as higher general abundance, such as birds of prey, were predicted less accurately (Fig. A3a, A3b). Birds of prey may use open spaces for hunting regardless of the land use type, whereas woodpeckers or skylarks are more restricted to certain land use types for foraging and breeding. Although, we restricted the occurrence points to the commonly used breeding codes, we still found to have less predictive power for certain species. Thus, occurrence data for abundant and noticeable species may be less associated with distinct environmental parameters, which leads to a lack of well-defined statistical results. The results show that increasing habitat tolerance of species unfavourably affects the accuracy of SDM predictions, as stated in many previous studies (McPherson & Jetz 2007, Mitchell et al. 2001, Segurado & Araújo 2004, Pearce et al. 2001). To account for different abundances, a more refined and restricted use of the breeding codes would be advisable for common species. Currently, we uniformly used equal breeding codes for all species to prevent the underprediction of rare and less detectable species. To better align occurrence points with the corresponding environmental predictors it might be necessary to differentiate the used breeding code classes according to the abundance of the investigated species. Nevertheless, the quality of predictions that had been demonstrated in this study can be considered as sufficient enough for assessing species richness (McPherson & Jetz 2007).

In our preliminary analysis, simple machine learning algorithms, such as RF and XGBOOST, outperformed more complex algorithms, such as ANNs or regression-based algorithms (GLMs). Overall, in our study Random Forest outperformed all other metrics, which has been found also by other studies fitting SDMs based on unstructured presence-only sample data (Lawson et al. 2014, Valavi et al. 2022). These patterns have been widely known in the literature for SDMs built with unstructured CS datasets in areas with structural richness and diverse landscape characteristics, such as urban environments (Hernandez et al. 2008, Williams et al. 2009). Therefore, simpler machine learning algorithms such as RF are better at predicting habitat suitability from incomplete and unstructured datasets due to their robustness to statistical outliers. Complete and structured data favour neural networks and classification algorithms.

The results of the species richness analysis pinpointed significant predicted occurrences along the Rhine River with its restored and preserved wetland, forest and agricultural habitats and less populated surroundings. In addition, high spatial concordance was also observed with the open green space system covering biotope networks and consisting of river and grassland systems located in the periphery of the core area of the Ruhr Metropolis (Fig. 7). Within the core area, patches and regional hotspots with high species richness could be identified. In particular, repurposed industrial brownfields and derelict collieries now function as habitat structures for preserved species richness. During the economic transformation in the Ruhr region, these areas were preserved as brownfields or became valuable green infrastructure, which can facilitate a high number of species by acting as refugee habitats and stepping stones in otherwise highly-sealed surroundings (Angold et al. 2006, Harrison & Davies 2002).

Some biodiversity hotspots that were demonstrated by this study coincide with locations that were highlighted in a recent publication examining biodiversity across former mining sites within the Ruhr Metropolis (Regionalverband Ruhr 2022).

Additionally, the structures and appearances of the five old regional greenbelts across the Ruhr region are mirrored in the species richness analysis. These green infrastructure-dominated landscapes work as a dense network of habitats spanning the cities of the Ruhr region (Zepp 2018), thus supporting biodiversity and ecosystem preservation. These belts consist of richly structured green infrastructure composed of extensive grassland, forest relics and water bodies interconnecting the large river axes in the North and South.

This study highlights the efficacy of SDM driven methodologies for identifying conservation-worthy habitats within inner-city environments. As urban areas continue to rapidly evolve, SDMs could play an integral role in pinpointing critical regions for species protection strategies aligned with conservation objectives (Le Louarn et al. 2018, Brun et al. 2020), thereby guiding initiatives aimed at sustaining and enhancing populations and habitats vital for biodiversity preservation.

4.1 Implications for urban green space planning in the Ruhr Metropolis

The species richness analysis shows the importance of preserving these green infrastructures not only in the larger connected networks in the outskirts of cities but also within the densely populated areas of inner city core areas, where they serve as ecological crosslinks (Keil et al. 2021). Despite the need for densification and urban sprawl, which intensifies the pressure on the Ruhr Metropolis for urban development and the use of open spaces for construction in the Ruhr Metropolis and elsewhere (Lin et al. 2015, Haaland & van den Bosch 2015), action concepts for prioritizing nature conservation and preservation were recently established with the framework of a Regional Strategy for Urban Biodiversity (Keil & Hering 2022, Keil et al. 2021). Within this framework, mission statements aimed at preserving and restoring urban nature within and beyond existing biotope networks, urban green spaces and even private greenery were developed. The framework suggests supportive measurements and monitoring concepts to identify early biodiversity changes on small or large scales. For large-scale assessments meta-barcoding, CS and RS techniques are key tools for assessing and monitoring biodiversity within the Ruhr Metropolis (Keil & Hering 2022).

From the perspective of the Regional Strategy for Biodiversity, our analysis helps to delineate and identify areas for prioritizing nature conservation. The presence of high species richness outside of protected areas suggests that conservation efforts should extend beyond legally designated zones. Urban and semi-urban green and current protected areas should be extended to safeguard the inner core green spaces. Hence, the green space network, including both old greenbelts and biotope network areas, plays a critical role in maintaining biodiversity and should be a focus for future conservation strategies. The indicator status of birds for other taxa even exceeds the considerations for conservational management beyond the examined avian biodiversity.

When considering the SLOSS (*Single Large or Several Small*) debate (Fletcher et al. 2018, Fahrig 2020, Tjørve 2010, Fahrig et al. 2022) concerning whether rather large single green spaces or several small patches better support biodiversity richness, the study initially emphasized the enhancement of biodiversity through large connected ecological corridors. In the urban core area, the implementation of extensively linked corridors may enrich biodiversity, as observed within the remnants of the crossing greenbelts. However, small patches within the core also sustain a diversity of avian species due to conservation efforts at historical industrial sites and other substantial green areas, such as cemeteries (Kowarik et al. 2016, Villaseñor & Escobar 2019) or urban and landscape parks (Nielsen et al. 2014, Yang et al. 2020). These structurally diverse sites are essential for biodiversity conservation. Thus, a balanced approach combining the maintenance of smaller rich patches with expansive high-quality green corridors throughout the Ruhr Metropolis could ensure a resilient network of green spaces. With this synergy, larger green spaces may function as source habitats supporting smaller patches and their urban surroundings during events such as local population loss due to human-induced disturbances or climate chance-related hazards (Snep et al. 2006, Savard et al. 2000).

5 Conclusions

In summary, our study successfully integrated Citizen Science (CS) data and remote sensing (RS) data into species distribution models (SDMs) to predict avian species presence in the urban landscape of the Ruhr Metropolis, Germany. The SDMs exhibited robust predictive performance, with consistently high ROC and TSS values across 26 selected species that were relevant to urban planning. Accuracy metrics, including sensitivity, specificity, and bias scores, confirmed the models' reliability in accurately classifying species occurrences. The Boyce index evidenced a high transferability beyond training space and time. We found precipitation, temperature, and the proportion of water, wetland and grassland to be the strongest drivers of bird distribution within the urban environment. The species richness analysis revealed distinct distribution patterns, thus emphasizing the pivotal role of green infrastructure across and between the green spaces along the river axis. High species richness along the Rhine River and within the urban core area highlighted the importance of preserving diverse habitats with respect to urban densification. The preservation of small patches with high biodiversity linked by large high-quality green space corridors may be favourable for biodiversity conservation in the urban area of the Ruhr Metropolis. These findings have practical implications for urban planning and conservation strategies, such as the recently implemented Regional Strategy for Biodiversity within the Ruhr Metropolis, thus guiding policy-makers to secure biodiversity hotspots and prioritizing conservation efforts. The integration of CS and RS datasets has proven to be a sophisticated approach, thus providing fine-scale predictions and enabling broad-scale identification of biodiversity patterns. In essence, our approach offers a valuable outcome for comprehending and conserving biodiversity in the dynamic urban environment of the Ruhr Metropolis and beyond.

Acknowledgments

We are grateful to the Association of German Avifaunists (DDA) and the steering committee providing us with citizen science observational data from ornitho.de. In particular, we would like to thank the volunteer birdwatchers on whose efforts many of these observation data are based. As part of the IMECOGIP project this research was partly funded by the Federal Ministry of Education and Research (BMBF) within the SURE funding priority of FONA (01LE1805A1).

References

Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38: 541–545. https://doi.org/10.1111/ecog.01132

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

Angelo H (2019) The greening imaginary: Urbanized nature in Germany's Ruhr region. *Theory and Society* 48: 645–669. https://doi.org/10.1007/s11186-019-09361-5

Angold PG, Sadler JP, Hill MO, Pullin A, Rushton S, Austin K, Small E, Wood B, Wadsworth R, Sanderson R, Thompson K (2006) Biodiversity in urban habitat patches. *Science of The Total Environment* 360: 196–204. https://doi.org/10.1016/j.scitotenv.2005.08.035

Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33: 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x

Aronson MFJ, Nilon CH, Lepczyk CA, Parker TS, Warren PS, Cilliers SS, Goddard MA, Hahs AK, Herzog C, Katti M, La Sorte FA, Williams NSG, Zipperer W (2016) Hierarchical filters determine community assembly of urban species pools. *Ecology* 97: 2952–2963. https://doi.org/10.1002/ecy.1535

Baker DJ, Maclean IMD, Gaston KJ (2024) Effective strategies for correcting spatial sampling bias in species distribution models without independent test data. *Diversity and Distributions* 30. https://doi.org/10.1111/ddi.13802

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3: 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Bauer H-G, Bezzel E, Fiedler W (2005) Das Kompendium der Vögel Mitteleuropas. Alles über Biologie, Gefährdung und Schutz. Wiebelsheim.

Bebber D (1999) Spatial autocorrelations. *Trends in Ecology & Evolution* 14: 196. https://doi.org/10.1016/S0169-5347(99)01607-9

Blair RB (1996) Land Use and Avian Species Diversity Along an Urban Gradient. *Ecological Applications* 6: 506–519. https://doi.org/10.2307/2269387

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FK (2002) Evaluating resource selection functions 157: 281–300. https://doi.org/10.1016/S0304-3800(02)00200-4

Breiman L, Friedmann JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Boca Raton.

Breiman L (2001) Random forests. *Machine Learning* 45: 5–32. https://doi.org/10.1023/A:1010933404324

Brun P, Thuiller W, Chauvier Y, Pellissier L, Wüest RO, Wang Z, Zimmermann NE (2020) Model complexity affects species distribution projections under climate change. *Journal of Biogeography* 47: 130–142. https://doi.org/10.1111/jbi.13734

Bruno JF, Stachowicz JJ, Bertness MD (2003) Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution* 18: 119–125. https://doi.org/10.1016/S0169-5347(02)00045-9

Busby JR (1991) Bioclim, a bioclimatic analysis and prediction system. Margules CR, Austin MP (eds) *Nature conservation: Cost effective biological surveys and data analysis:* 64–68. Canberra.

Cai Z, La Sorte FA, Chen Y, Wu J (2023) The surface urban heat island effect decreases bird diversity in Chinese cities. *The Science of the total environment* 902: 166200. https://doi.org/10.1016/j.scitotenv.2023.166200

Carroll KA, Pidgeon AM, Elsen PR, Farwell LS, Gudex-Cross D, Zuckerberg B, Radeloff VC (2023) Mapping multiscale breeding bird species distributions across the United States and evaluating their conservation applications. *Ecological applications: a publication of the Ecological Society of America:* e2934. https://doi.org/10.1002/eap.2934

Caruana R, Niculescu-Mizil A, Crew G, Ksikes A (2004) Ensemble selection from libraries of models. Brodley C (ed) *Design Automation Conference:* 18. Proceedings 2004. New York, N.Y., Piscataway, N.J. https://doi.org/10.1145/1015330.1015432

Cavender-Bares J, Schneider FD, Santos MJ, Armstrong A, Carnaval A, Dahlin KM, Fatoyinbo L, Hurtt GC, Schimel D, Townsend PA, Ustin SL, Wang Z, Wilson AM (2022) Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nature Ecology & Evolution* 6: 506–519. https://doi.org/10.1038/s41559-022-01702-5

Chen T, Guestrin C (2016) XGBoost. Krishnapuram B, Shah M, Smola A, Aggarwal C, Shen D, Rastogi R (eds) *KDD '16:* 785–794. New York. https://doi.org/10.1145/2939672.2939785

Ciach M, Fröhlich A (2017) Habitat type, food resources, noise and light pollution explain the species composition, abundance and stability of a winter bird assemblage in an urban environment. *Urban Ecosystems* 20: 547–559. https://doi.org/10.1007/s11252-016-0613-6

Coudun C, Gégout J-C (2006) The derivation of species response curves with Gaussian logistic regression is sensitive to sampling intensity and curve characteristics. *Ecological Modelling* 199: 164–175. https://doi.org/10.1016/j.ecolmodel.2006.05.024

Dachverband Deutscher Avifaunisten (2018) Verbreitung der Beobachter:innen. https://www.ornitho.de/index.php?m_id=1119&item=14 (date: 04.03.2024).

Dearborn DC, Kark S (2010) Motivations for conserving urban biodiversity. *Conservation Biology* 24: 432–440. https://doi.org/10.1111/j.1523-1739.2009.01328.x

Donnelly R, Marzluff JM (2006) Relative importance of habitat quantity, structure, and spatial pattern to birds in urbanizing environments. *Urban Ecosystems* 9: 99–117. https://doi.org/10.1007/s11252-006-7904-2

Dubayah R, Hofton M, Blair J, Armston J, Tang H, Luthcke S (2021a) GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002.

Dubayah R, Tang H, Armston J, Luthcke S, Hofton M, Blair J (2021b) GEDI L2B Canopy Cover and Vertical Profile Metrics Data Global Footprint Level V002.

DWD Climate Data Center (CDC) (n.d. a) Grids of monthly averaged daily maximum air temperature (2m) over Germany, version v1.0.

DWD Climate Data Center (CDC) (n.d. b) Grids of monthly averaged daily minimum air temperature (2m) over Germany, version v1.0.

DWD Climate Data Center (CDC) (n.d. c) Grids of monthly total precipitation over Germany, version v1.0.

DWD Climate Data Center (CDC) (n.d. d) Monthly grids of the accumulated actual evapotranspiration over grass, version 0.x (date: 19.08.2022).

DWD Climate Data Center (CDC) (n.d. e) Monthly grids of the accumulated potential evapotranspiration over grass, version 0.x (date: 19.08.2022).

Eglington SM, Noble DG, Fuller RJ (2012) A meta-analysis of spatial relationships in species richness across taxa: Birds as indicators of wider biodiversity in temperate regions. *Journal for Nature Conservation* 20: 301–309. https://doi.org/10.1016/j.jnc.2012.07.002

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

Elmqvist T, Fragkias M, Goodness J, Güneralp B, Marcotullio PJ, McDonald RI, Parnell S, Schewenius M, Sendstad M, Seto KS, Wilkinson C (eds) (2016) Urbanization, biodiversity and ecosystem services. Challenges and opportunities. https://doi.org/10.1007/978-94-007-7088-1

Elton CS (1927) Animal ecology. Chicago.

Engler JO, Stiels D, Schidelko K, Strubbe D, Quillfeldt P, Brambilla M (2017) Avian SDMs: current state, challenges, and opportunities. *Journal of Avian Biology* 48: 1483–1504. https://doi.org/10.1111/jav.01248

European Commission (2022) Nature restoration law – For people, climate, and planet. Brussels.

European Environment Agency (2020a) Imperviousness density 2018 (raster 100 m), Europe, 3-yearly, Aug. 2020.

European Environment Agency (2020b) Tree cover density 2018 (raster 100 m), Europe, 3-yearly, Sep. 2020.

Fahrig L (2020) Why do several small patches hold more species than few large patches? *Global Ecology and Biogeography* 29: 615–628. https://doi.org/10.1111/geb.13059

Fahrig L, Watling JI, Arnillas CA, Arroyo-Rodríguez V, Jörger-Hickfang T, Müller J, Pereira HM, Riva F, Rösch V, Seibold S, Tscharntke T, May F (2022) Resolving the SLOSS dilemma for biodiversity conservation: a research agenda. *Biological Reviews* 97: 99–114. https://doi.org/10.1111/brv.12792

Feldman MJ, Imbeau L, Marchand P, Mazerolle MJ, Darveau M, Fenton NJ (2021) Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLOS ONE* 16: e0234587. https://doi.org/10.1371/journal.pone.0234587

Ferenc M, Sedláček O, Fuchs R (2014) How to improve urban greenspace for woodland birds: site and local-scale determinants of bird species richness. *Urban Ecosystems* 17: 625–640. https://doi.org/10.1007/s11252-013-0328-x

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38–49. https://doi.org/10.1017/S0376892997000088

Fletcher R, Fortin M-J (2019) Spatial dependence and autocorrelation. Fletcher R, Fortin M-J (eds) *Spatial ecology and conservation modeling:* 133–168. Cham, Switzerland. https://doi.org/10.1007/978-3-030-01989-1_5

Fletcher RJ, Didham RK, Banks-Leite C, Barlow J, Ewers RM, Rosindell J, Holt RD, Gonzalez A, Pardini R, Damschen EI, Melo FP, Ries L, Prevedello JA, Tscharntke T, Laurance WF, Lovejoy T, Haddad NM (2018) Is habitat fragmentation good for biodiversity? *Biological Conservation* 226: 9–15. https://doi.org/10.1016/j.biocon.2018.07.022

Friedman JH (1991) Multivariate adaptive regression splines. *The Annals of Statistics* 19: 1–67. https://doi.org/10.1214/aos/1176347963

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202: 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology letters* 4: 379–391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Grimmett L, Whitsed R, Horta A (2020) Presence-only species distribution models are sensitive to sample prevalence: Evaluating models using spatial prediction stability and accuracy metrics. *Ecological Modelling* 431: 109194. https://doi.org/10.1016/j.ecolmodel.2020.109194

Grinnell J (1928) Presence and absence of animals. Berkeley.

Guerry AD, Smith JR, Lonsdorf E, Daily GC, Wang X, Chun Y (2021) Urban nature and biodiversity for cities. Washington D.C. https://doi.org/10.1596/36325

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

Guisan A, Thuiller W, Zimmermann NE (2017) Habitat suitability and distribution models. With applications in R. Cambridge. https://doi.org/10.1017/9781139028271

Güneralp B, Seto KC (2013) Futures of global urban expansion: uncertainties and implications for biodiversity conservation. *Environmental Research Letters* 8: 14025. https://doi.org/10.1088/1748-9326/8/1/014025

Haaland C, van den Bosch CK (2015) Challenges and strategies for urban green-space planning in cities undergoing densification: A review. *Urban Forestry & Urban Greening* 14: 760–771. https://doi.org/10.1016/j.ufug.2015.07.009

Hamann M (1991) Die Bedeutung von Industriebrachflächen für die Avifauna am Beispiel von Gelsenkirchen (mittleres Ruhrgebiet). *Charadrius:* 49–62.

Hanski I (1998) Metapopulation dynamics. *Nature* 396: 41–49. https://doi.org/10.1038/23876

Harrison C, Davies G (2002) Conserving biodiversity that matters: practitioners' perspectives on brownfield development and urban nature conservation in London. *Journal of Environmental Management* 65: 95–108. https://doi.org/10.1006/jema.2002.0539

Hastie T, Tibshirani R (1986) Generalized additive models. *Statistical Science* 1. https://doi.org/10.1214/ss/1177013604

Hastie T, Tibshirani R, Buja A (1994) Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association* 89: 1255–1270. https://doi.org/10.1080/01621459.1994.10476866

He KS, Bradley BA, Cord AF, Rocchini D, Tuanmu M-N, Schmidtlein S, Turner W, Wegmann M, Pettorelli N (2015) Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation* 1: 4–18. https://doi.org/10.1002/rse2.7

Hernandez PA, Franke I, Herzog SK, Pacheco V, Paniagua L, Quintana HL, Soto A, Swenson JJ, Tovar C, Valqui TH, Vargas J, Young BE (2008) Predicting species distributions in poorly-studied landscapes. *Biodiversity and Conservation* 17: 1353–1366. https://doi.org/10.1007/s10531-007-9314-z

Hijmans R (2022) terra: Spatial data analysis. R package version 1.7-81. https://rspatial.github.io/terra/, https://rspatial.org/.

Hilbert D (2001) The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. *Ecological Modelling* 46: 311–327. https://doi.org/10.1016/S0304-3800(01)00323-4

Hildén O (1965) Habitat selection in birds: A review. *Annales Zoologici Fennici* 2: 53–75. https://www.jstor.org/stable/23730835.

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences *Ecological Modelling* 199: 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Huang Y, Zhao Y, Li S, Gadow K von (2015) The effects of habitat area, vegetation structure and insect richness on breeding bird populations in Beijing urban parks. *Urban Forestry & Urban Greening* 14: 1027–1039. https://doi.org/10.1016/j.ufug.2015.09.010

Hutchinson GE (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415–427. https://doi.org/10.1101/SQB.1957.022.01.039

Hutchinson GE (1978) An introduction to population ecology. New Haven.

IPBES (2019) Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn. https://doi.org/10.5281/zenodo.3831673

Isaksson C (2018) Impact of Urbanization on Birds. Tietze DT (ed) *Bird Species:* 235–257. Cham. https://doi.org/10.1007/978-3-319-91689-7_13

Jiguet F, Julliard R, Thomas CD, Dehorter O, Newson SE, Couvet D (2006) Thermal range predicts bird population resilience to extreme high temperatures. *Ecology letters* 9: 1321–1330. https://doi.org/10.1111/j.1461-0248.2006.00986.x

Jiménez-Valverde A, Lobo J, Hortal J (2009) The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology* 10: 196–205. https://doi.org/10.1556/ComEc.10.2009.2.9

Keil P (2019) Industrial nature and species diversity in the Landscape Park Duisburg-Nord. Oberhausen. https://www.bswr.de/downloads/bswr_ep39_2019_keil_industrial_nature_and_spec.pdf

Keil P, Brosch B, Buch C (2013) Open urban-industrial brownfields of high conservation interest: A methodological approach to site selection in the Ruhr metropolis. *Natur & Landschaft* 88: 213–219. https://doi.org/10.17433/5.2013.50153218.213-219

Keil P, Hering D (eds) (2022) Regionale Biodiversitätsstrategie Ruhrgebiet. Offensive Grüne Infrastruktur 2030. Essen.

Keil P, Hering D, Schmitt T, Zepp H (eds) (2021) Positionen zu einer regionalen Biodiversitäts-Strategie Ruhrgebiet. Essen.

Keller V, Milanesi P, Kipson M, Martí D, Herrando Vila S, Klvaňová A, Voříšek P, Anton M, Rodríguez-Franch M, Kalâkin MV, Foppen RPB, Bauer H-G (2020) European breeding bird atlas 2. Distribution, abundance and change. Barcelona.

Keys R (1981) Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29: 1153–1160. https://doi.org/10.1109/TASSP.1981.1163711

Kiel E-F (2005) Artenschutz in Fachplanungen. *LÖBF-Mitteilungen:* 12–17.

Kiel E-F (2016) Geschützte Arten in Nordrhein-Westfalen. Vorkommen, Erhaltungszustand, Gefährdungen, Maßnahmen. Düsseldorf.

Knapp S, Aronson MFJ, Carpenter E, Herrera-Montes A, Jung K, Kotze DJ, La Sorte FA, Lepczyk CA, Macgregor-Fors I, MacIvor JS, Moretti M, Nilon CH, Piana MR, Rega-Brodsky CC, Salisbury A, Threlfall CG, Trisos C, Williams NSG, Hahs AK (2021) A research agenda for urban biodiversity in the global extinction crisis. *BioScience* 71: 268–279. https://doi.org/10.1093/biosci/biaa141

Knapp W (1998) The Rhine-Ruhr area in transformation: Towards a European metropolitan region? *European Planning Studies* 6: 379–393. https://doi.org/10.1080/09654319808720469

Kock N, Lynn G (2012) Lateral Collinearity and Misleading Results in Variance-Based SEM: An Illustration and Recommendations. *Journal of the Association for Information Systems* 13: 546–580. https://doi.org/10.17705/1jais.00302

Kowallik C, Rautenberg T (2014) Kiebitze Vanellus vanellus als Brutvögel im Westlichen Ruhrgebiet – ein Vergleich der Situation 2014 mit dem NRW-Brutvogelatlas (2005-2009). *Charadrius* 50: 50–55.

Kowarik I, Buchholz S, Lippe M von der, Seitz B (2016) Biodiversity functions of urban cemeteries: Evidence from one of the largest Jewish cemeteries in Europe. *Urban Forestry & Urban Greening* 19: 68–78. https://doi.org/10.1016/j.ufug.2016.06.023

Lawson CR, Hodgson JA, Wilson RJ, Richards SA (2014) Prevalence, thresholds and the performance of presence–absence models. *Methods in Ecology and Evolution* 5: 54–64. https://doi.org/10.1111/2041-210X.12123

Le Louarn M, Clergeau P, Strubbe D, Deschamps-Cottin M (2018) Dynamic species distribution models reveal spatiotemporal habitat shifts in native range-expanding versus non-native invasive birds in an urban area. *Journal of Avian Biology* 49: jav-01527. https://doi.org/10.1111/jav.01527

Legendre P, Fortin MJ (1989) Spatial pattern and ecological analysis. *Vegetatio* 80: 107–138. https://doi.org/10.1007/BF00048036

Lepczyk CA, Aronson MFJ, La Sorte FA (2023) Cities as sanctuaries. *Frontiers in Ecology and the Environment* 21: 251–259. https://doi.org/10.1002/fee.2637

Lin B, Meyers J, Barnett G (2015) Understanding the potential loss and inequities of green space distribution with urban densification. *Urban Forestry & Urban Greening* 14: 952–958. https://doi.org/10.1016/j.ufug.2015.09.003

Lindsay JB (2016) Whitebox GAT: A case study in geomorphometric analysis. *Computers & Geosciences* 95: 75–84. https://doi.org/10.1016/j.cageo.2016.07.003

Liu C, White M, Newell G (2011) Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography* 34: 232–243. https://doi.org/10.1111/j.1600-0587.2010.06354.x

Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

Márcia Barbosa A, Real R, Muñoz A-R, Brown JA (2013) New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Diversity and Distributions* 19: 1333–1338. https://doi.org/10.1111/ddi.12100

McCullagh P (1984) Generalized linear models. *European Journal of Operational Research* 16: 285–292. https://doi.org/10.1016/0377-2217(84)90282-0

McDonald RI, Mansur AV, Ascensão F, Colbert M, Crossman K, Elmqvist T, Gonzalez A, Güneralp B, Haase D, Hamann M, Hillel O, Huang K, Kahnt B, Maddox D, Pacheco A, Pereira HM, Seto KC, Simkin R, Walsh B, Werner AS, Ziter C (2020) Research gaps in knowledge of the impact of urban growth on biodiversity. *Nature Sustainability* 3: 16–24. https://doi.org/10.1038/s41893-019-0436-6

McPherson JM, Jetz W (2007) Effects of species' ecology on the accuracy of distribution models. *Ecography* 30: 135–151. https://doi.org/10.1111/j.2006.0906-7590.04823.x

Mekonen S (2017) Birds as biodiversity and environmental indicator. *Journal of Natural Science Research* 7(21): 28–34. https://www.iiste.org/Journals/index.php/JNSR/article/view/39931/41049

Meller L, Cabeza M, Pironon S, Barbet-Massin M, Maiorano L, Georges D, Thuiller W (2014) Ensemble distribution models in conservation prioritization: From consensus predictions to consensus reserve networks. *Diversity and Distributions* 20: 309–321. https://doi.org/10.1111/ddi.12162

Merow C, Dahlgren JP, Metcalf CJE, Childs DZ, Evans ME, Jongejans E, Record S, Rees M, Salguero-Gómez R, McMahon SM (2014) Advancing population ecology with integral projection models: a practical guide. *Methods in Ecology and Evolution* 5: 99–110. https://doi.org/10.1111/2041-210X.12146

Miller J (2010) Species distribution modeling. *Geography Compass* 4: 490–509. https://doi.org/10.1111/j.1749-8198.2010.00351.x

Mitchell MS, Lancia RA, Gerwin JA (2001) Using landscape-level data to predict the distribution of birds on a managed forest: Effects of scale. *Ecological application: A publication of the Ecological Society of America* 11: 1692–1708. https://doi.org/10.1890/1051-0761(2001)011[1692:ULLDTP]2.0.CO;2

Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RP (2014) ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution* 5: 1198–1205. https://doi.org/10.1111/2041-210X.12261

Naimi B, Araújo MB (2016) sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* 39: 368–375. https://doi.org/10.1111/ecog.01881

Nielsen AB, van den Bosch M, Maruthaveeran S, van den Bosch CK (2014) Species richness in urban parks and its drivers: A review of empirical evidence. *Urban Ecosystems* 17: 305–327. https://doi.org/10.1007/s11252-013-0316-1

Ottoni I, Oliveira FF de, Young RJ (2009) Estimating the diet of urban birds: The problems of anthropogenic food and food digestibility. *Applied Animal Behaviour Science* 117: 42–46. https://doi.org/10.1016/j.applanim.2008.11.002

Pearce J, Ferrier S, Scotts D (2001) An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. *Journal of Environmental Management* 62: 171–184. https://doi.org/10.1006/jema.2001.0425

Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* 30: 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A (2017) Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Global Ecology and Biogeography* 26: 275–287. https://doi.org/10.1111/geb.12530

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pickett S, Cadenasso ML, Grove JM, Nilon CH, Pouyat RV, Zipperer WC, Costanza R (op. 2008) Urban ecological systems: Linking terrestrial ecological, physical, and socioeconomic components of metropolitan areas *Urban ecology:* 99–122. New York. https://doi.org/10.1007/978-0-387-73412-5_7

Pulliam HR (1988) Sources, Sinks, and Population Regulation. *The American naturalist* 132: 652–661. https://doi.org/10.1086/284880

R Core Team (2022) R: A language and environment for statistical. Vienna, Austria.

Randin CF, Ashcroft MB, Bolliger J, Cavender-Bares J, Coops NC, Dullinger S, Dirnböck T, Eckert S, Ellis E, Fernández N, Giuliani G, Guisan A, Jetz W, Joost S, Karger D, Lembrechts J, Lenoir J, Luoto M, Morin X, Price B, Rocchini D, Schaepman M, Schmid B, Verburg P, Wilson A, Woodcock P, Yoccoz N, Payne D (2020) Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote Sensing of Environment* 239: 111626. https://doi.org/10.1016/j.rse.2019.111626

Regionalverband Ruhr (2021) Metropole Ruhr - Die Region in Zahlen. Essen.

Regionalverband Ruhr (2022) Biodiversität in der Metropole Ruhr. Raumanalyse der Industrienatur. https://www.rvr.ruhr/themen/oekologie-umwelt/gruene-infrastruktur/biodiversitaetsstrategie/raumanalyse-der-industrienatur/.

Ridgeway G (1999) The state of boosting. Berk K, Pourahmadi M (eds) *Computing Science and Statistics* 31: 172–181. Fairfax.

Ripley BD (2008) Pattern Recognition and Neural Networks. Cambridge.

Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40: 913–929. https://doi.org/10.1111/ecog.02881

Rocchini D, Balkenhol N, Carter GA, Foody GM, Gillespie TW, He KS, Kark S, Levin N, Lucas K, Luoto M, Nagendra H, Oldeland J, Ricotta C, Southworth J, Neteler M (2010) Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecological Informatics* 5: 318–329. https://doi.org/10.1016/j.ecoinf.2010.06.001

Rocchini D, Boyd DS, Féret J-B, Foody GM, He KS, Lausch A, Nagendra H, Wegmann M, Pettorelli N (2016) Satellite remote sensing to monitor species diversity: Potential and pitfalls. *Remote Sensing in Ecology and Conservation* 2: 25–36. https://doi.org/10.1002/rse2.9

Root T (1988) Environmental factors associated with avian distributional boundaries. *Journal of Biogeography* 15: 489. https://doi.org/10.2307/2845278

Roughgarden J, Running SW, Matson PA (1991) What does remote sensing do for ecology? *Ecology* 72: 1918–1922. https://doi.org/10.2307/1941546

Savard J-PL, Clergeau P, Mennechez G (2000) Biodiversity concepts and urban ecosystems. *Landscape and Urban Planning* 48: 131–142. https://doi.org/10.1016/S0169-2046(00)00037-2

Scott JM, Csuti B, Jacobi JD, Estes JE (1987) Species richness. *BioScience* 37: 782–788. https://doi.org/10.2307/1310544

Segurado P, Araújo MB (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31: 1555–1568. https://doi.org/10.1111/j.1365-2699.2004.01076.x

Smith AM, Sutton SG (2008) The role of a flagship species in the formation of conservation intentions. *Human Dimensions of Wildlife* 13: 127–140. https://doi.org/10.1080/10871200701883408

Smith SL, Holland DA, Longley PA (2004) The importance of understanding error in lidar digital elevation models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives:* 996–1001. https://www.isprs.org/proceedings/XXXV/congress/comm4/papers/488.pdf

Snep R, Opdam P, Baveco JM, WallisDeVries MF, Timmermans W, Kwak R, Kuypers V (2006) How peri-urban areas can strengthen animal populations within cities: A modeling approach. *Biological Conservation* 127: 345–355. https://doi.org/10.1016/j.biocon.2005.06.034

Soanes K, Lentini PE (2019) When cities are the last chance for saving species. *Frontiers in Ecology and the Environment* 17: 225–231. https://doi.org/10.1002/fee.2032

Spangenberg JH (2007) Biodiversity pressure and the driving forces behind. *Ecological Economics* 61: 146–158. https://doi.org/10.1016/j.ecolecon.2006.02.021

Spotswood EN, Beller EE, Grossinger R, Grenier JL, Heller NE, Aronson MFJ (2021) The biological deserts fallacy: Cities in their landscapes contribute more than we think to regional biodiversity. *BioScience* 71: 148–160. https://doi.org/10.1093/biosci/biaa155

Steen VA, Elphick CS, Tingley MW (2019) An evaluation of stringent filtering to improve species distribution models from citizen science data. *Diversity and Distributions* 25: 1857–1869. https://doi.org/10.1111/ddi.12985

Steen VA, Tingley MW, Paton PWC, Elphick CS (2021) Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution* 12: 216–226. https://doi.org/10.1111/2041-210X.13525

Südbeck P, Andretzke H, Fischer S, Gedeon K (2005) Methodenstandards zur Erfassung der Brutvögel Deutschlands. Radolfzell.

Sudmann SR, Schmitz M, Grüneberg C, Herkenrath P, Jöbges MM, Mika T, Nottmeyer K, Schidelko K, Schubert W, Stiels D (2023) Rote Liste der Brutvogelarten Nordrhein-Westfalens, 7. Fassung, Stand: Dezember 2021. *Charadrius* 57: 75–130.

Symonds MRE, Johnson CN (2008) Species richness and evenness in Australian birds. *The American naturalist* 171: 480–490. https://doi.org/10.1086/528960

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, Siqueira MF de, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extinction risk from climate change. *Nature* 427: 145–148. https://doi.org/10.1038/nature02121

Thuiller W, Georges D, Gueguen M, Engler R, Breiner F, Lafourcade B, Patin R (2023) biomod2: Ensemble platform for species distribution modeling.

Tjørve E (2010) How to resolve the SLOSS debate: lessons from species-diversity models. *Journal of theoretical biology* 264: 604–612. https://doi.org/10.1016/j.jtbi.2010.02.009

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234. https://doi.org/10.2307/143141

Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G (2019) blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution* 10: 225–232. https://doi.org/10.1111/2041-210X.13107

Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J (2022) Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs* 92: e01486. https://doi.org/10.1002/ecm.1486

Vallecillo S, Maes J, Polce C, Lavalle C (2016) A habitat quality indicator for common birds in Europe based on species distribution models. *Ecological Indicators* 69: 488–499. https://doi.org/10.1016/j.ecolind.2016.05.008

Venables WN, Ripley BD (2011) Modern applied statistics with S. Statistics and computing. New York, London.

Venter ZS, Barton DN, Gundersen V, Figari H, Nowell M (2020) Urban nature in a time of crisis: recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway. *Environmental Research Letters* 15: 104075. https://doi.org/10.1088/1748-9326/abb396

Veríssimo D, Pongiluppi T, Santos MCM, Develey PF, Fraser I, Smith RJ, MacMilan DC (2014) Using a systematic approach to select flagship species for bird conservation. *Conservation Biology* 28: 269–277. https://doi.org/10.1111/cobi.12142

Villaseñor NR, Escobar MAH (2019) Cemeteries and biodiversity conservation in cities: how do landscape and patch-level attributes influence bird diversity in urban park cemeteries? *Urban Ecosystems* 22: 1037–1046. https://doi.org/10.1007/s11252-019-00877-3

Whitford AM, Shipley BR, McGuire JL (2024) The influence of the number and distribution of background points in presence-background species distribution models. *Ecological Modelling* 488: 110604. https://doi.org/10.1016/j.ecolmodel.2023.110604

Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O'Brien JM, Schwartz MW (2009) Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions* 15: 565–576. https://doi.org/10.1111/j.1472-4642.2009.00567.x

Wu J, Zhang S, Luo Y, Wang H, Zhao Y (2022) Assessment of risks to habitat connectivity through the stepping-stone theory: A case study from Shenzhen, China. *Urban Forestry & Urban Greening* 71: 127532. https://doi.org/10.1016/j.ufug.2022.127532

Wunderlich RF, Lin Y-P, Anthony J, Petway JR (2019) Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. *Nature Conservation* 35: 97–116. https://doi.org/10.3897/natureconservation.35.33918

Yang X, Tan X, Chen C, Wang Y (2020) The influence of urban park characteristics on bird diversity in Nanjing, China. *Avian Research* 11: 1–9. https://doi.org/10.1186/s40657-020-00234-5

Yee TW, Mitchell ND (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science* 2: 587–602. https://doi.org/10.2307/3236170

Zanaga D, van de Kerchove R, Daems D, Keersmaecker W de, Brockmann C, Kirches G, Wevers J, Cartus O, Santoro M, Fritz S, Lesiv M, Herold M, Tsendbazar N-E, Xu P, Ramoino F, Arino O (2022) ESA WorldCover 10 m 2021 v200.

Zepp H (2018) Regional green belts in the Ruhr region. A planning concept revisited in view of ecosystem services. *Erdkunde* 72: 1–23. https://doi.org/10.3112/erdkunde.2018.01.01

Zepp H, Gessner M, Gruenhagen L, Bührs M (2023) Modeling the cooling effect of urban green spaces: The neglected control variables of 'soil moisture' and 'biotope types'. *Urban Forestry & Urban Greening* 90: 128137. https://doi.org/10.1016/j.ufug.2023.128137

Zepp H, Gross L, Inostroza L (2020) And the winner is? Comparing urban green space provision and accessibility in eight European metropolitan areas using a spatially explicit approach. *Urban Forestry & Urban Greening* 49: 126603. https://doi.org/10.1016/j.ufug.2020.126603

Authors

Malte Bührs
ORCID: 0000-0002-5751-3267
Malte.Buehrs@rub.de
Prof. Dr. Harald Zepp
ORCID: 0000-0001-5599-9542
Harald.Zepp@rub.de
Prof. Dr. Thomas Schmitt
ORCID: 0000-0002-3171-4132
Thomas.Schmitt@rub.de
Institute of Geography
Ruhr University Bochum
Universitätsstraße 150
44801 Bochum
Germany

Appendix

Tab. A1: Habitat preferences of the investigated avian species. Land use classification based on the ESA WorldCover (Zanaga et al. 2022). The habitat preferences of the species were derived from Kiel 2016, Bauer et al. 2005. ++: highly suitable, +: suitable, -: less suitable, --: unsuitable.

Scientific Name	**Land use classification**							
	Tree cover	**Shrubland**	**Grassland**	**Cropland**	**Built-up**	**Sparse vegetation**	**Water bodies**	**Herbaceous wetland**
Accipiter gentilis	++	+	-	-	--	--	--	--
Accipiter nisus	+	++	-	-	+	-	--	--
Acrocephalus scirpaceus	--	-	+	-	-	--	+	++
Alauda arvensis	--	-	++	++	-	+	--	+
Anthus pratensis	--	++	++	+	-	+	--	+
Anthus trivialis	++	+	+	-	--	-	--	-
Ardea cinerea	+	-	+	+	-	-	++	++
Buteo buteo	++	++	+	+	-	--	--	--
Carduelis cannabina	-	++	++	+	-	--	--	+
Charadrius dubius	--	+	+	--	+	++	-	++
Corvus frugilegus	++	++	+	+	+	-	--	-
Cuculus canorus	--	-	++	-	-	--	--	++
Delichon urbicum	--	-	+	++	++	--	--	--
Dendrocopos medius	++	+	--	--	-	--	--	--
Dryobates minor	++	+	--	--	-	--	--	--
Falco subbuteo	+	++	+	-	--	-	--	-
Falco tinnunculus	-	+	++	++	++	-	--	-
Hirundo rustica	--	-	+	++	+	-	--	--
Lullula arborea	-	+	++	-	--	++	-	+
Luscinia megarhynchos	+	++	+	--	-	--	-	+
Milvus milvus	++	+	+	+	--	-	--	--
Passer montanus	+	++	-	+	-	--	--	-
Phoenicurus phoenicurus	+	++	-	--	+	-	--	-
Streptopelia turtur	+	++	+	--	-	--	--	-
Sturnus vulgaris	++	+	+	+	+	--	--	--
Vanellus vanellus	--	--	+	++	--	+	--	-

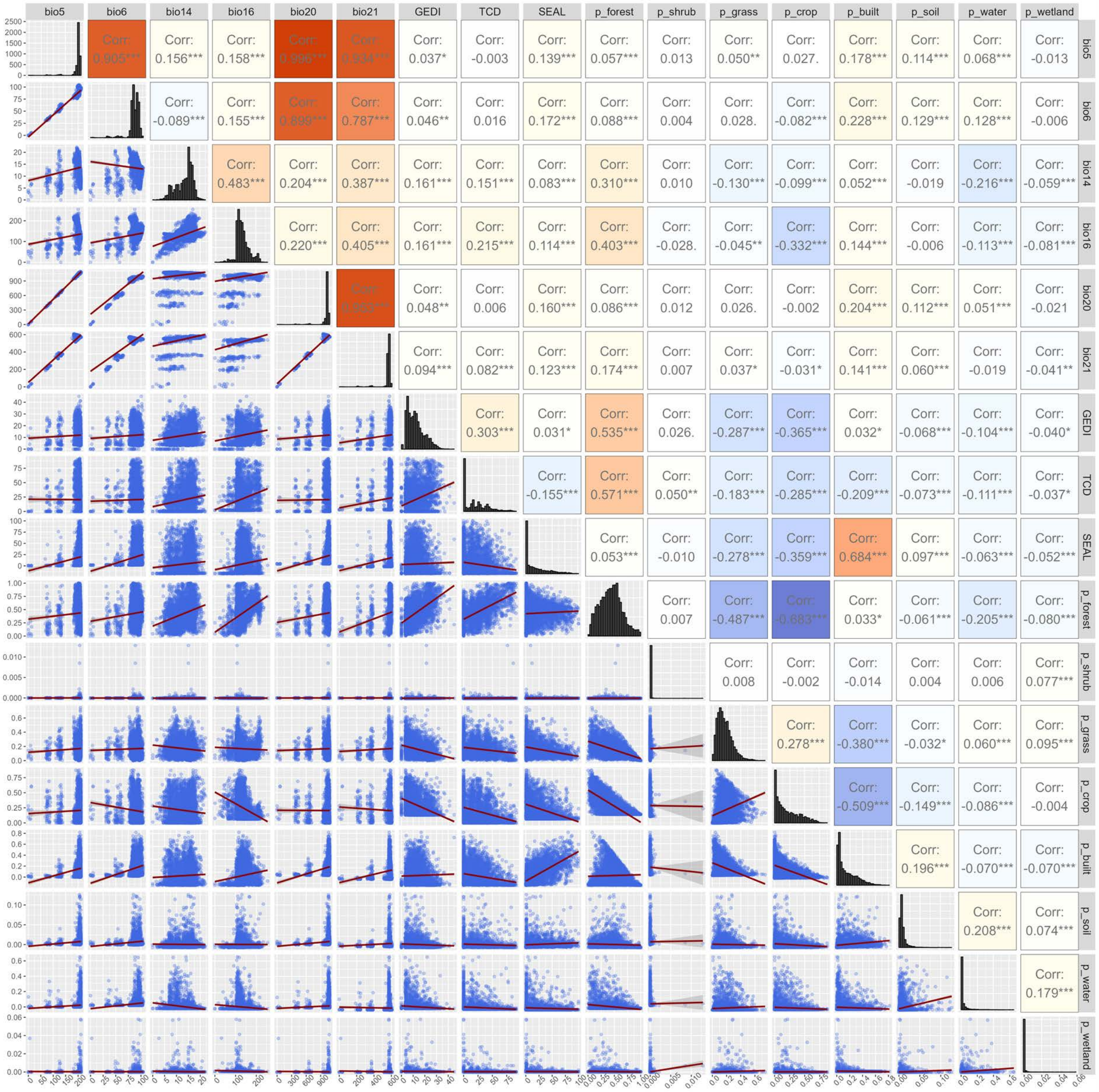

Fig. A1: Pearson correlation coefficient matrix comparing paired environmental covariates (n=7,540). Negative correlations are shaded blue; positive correlations are shaded red. The strength of the correlation is indicated by colour saturation. Stars indicate significance levels; ***: P<0.05, **: P<0.01, ***: P<0.001.

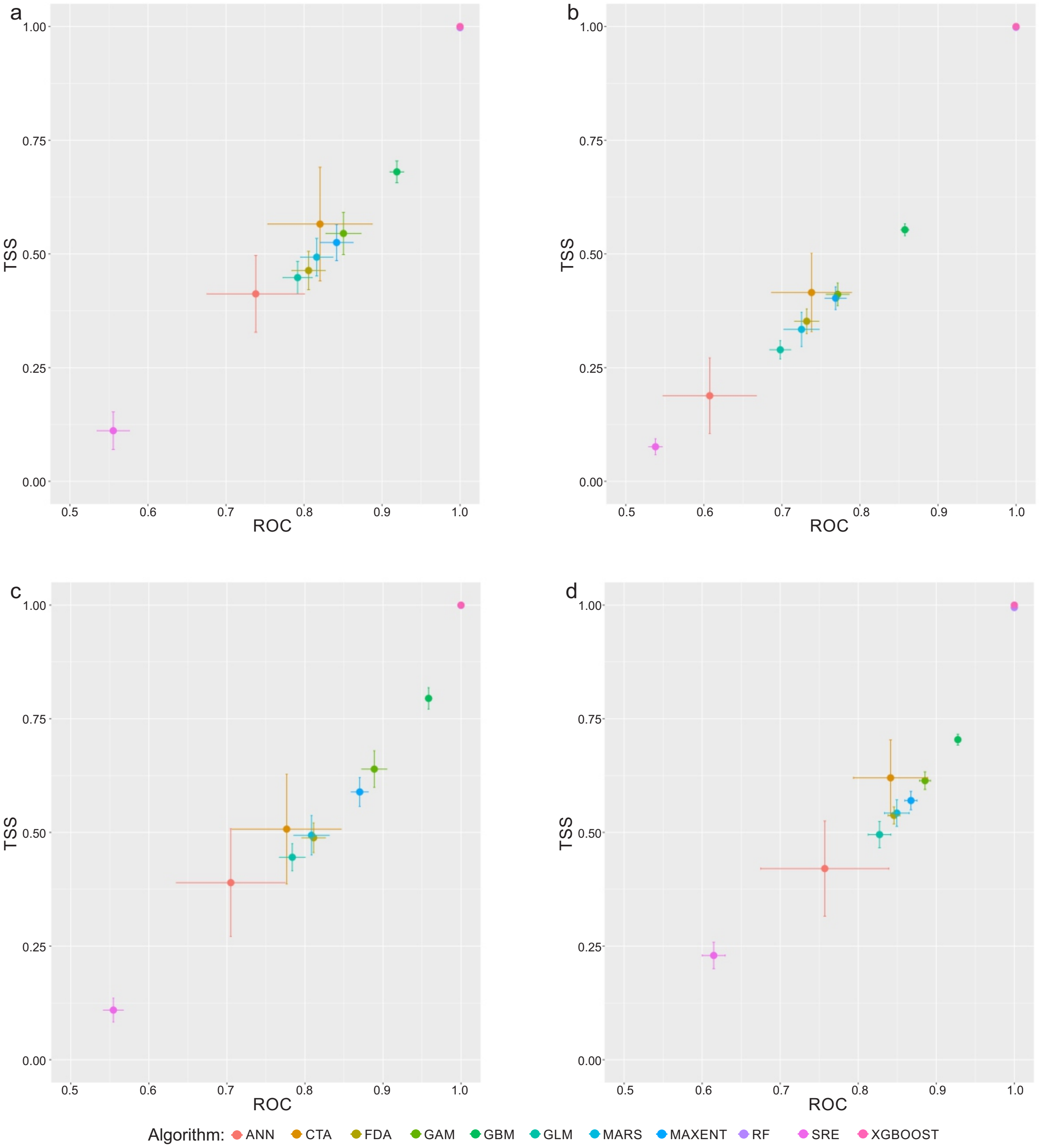

Fig. A2: Model performance of eleven algorithms for four test species: a) Alauda arvensis, b) Delichon urbicum, c) Dendrocopos medius, and d) Luscinia megarhynchos.

The bottom left indicates models that performed poorly based on both metrics, whereas points in the bottom right performed well for ROC but not TSS, points in the top left performed well for TSS but not ROC, and points in the top left performed well for both metrics. Each point includes error lines, thus indicating the possible range of actual values. We evaluated both the points and lines for model selection. We included four species covering several occurrence aspects, such as land use preferences, nesting behaviour and commonness.

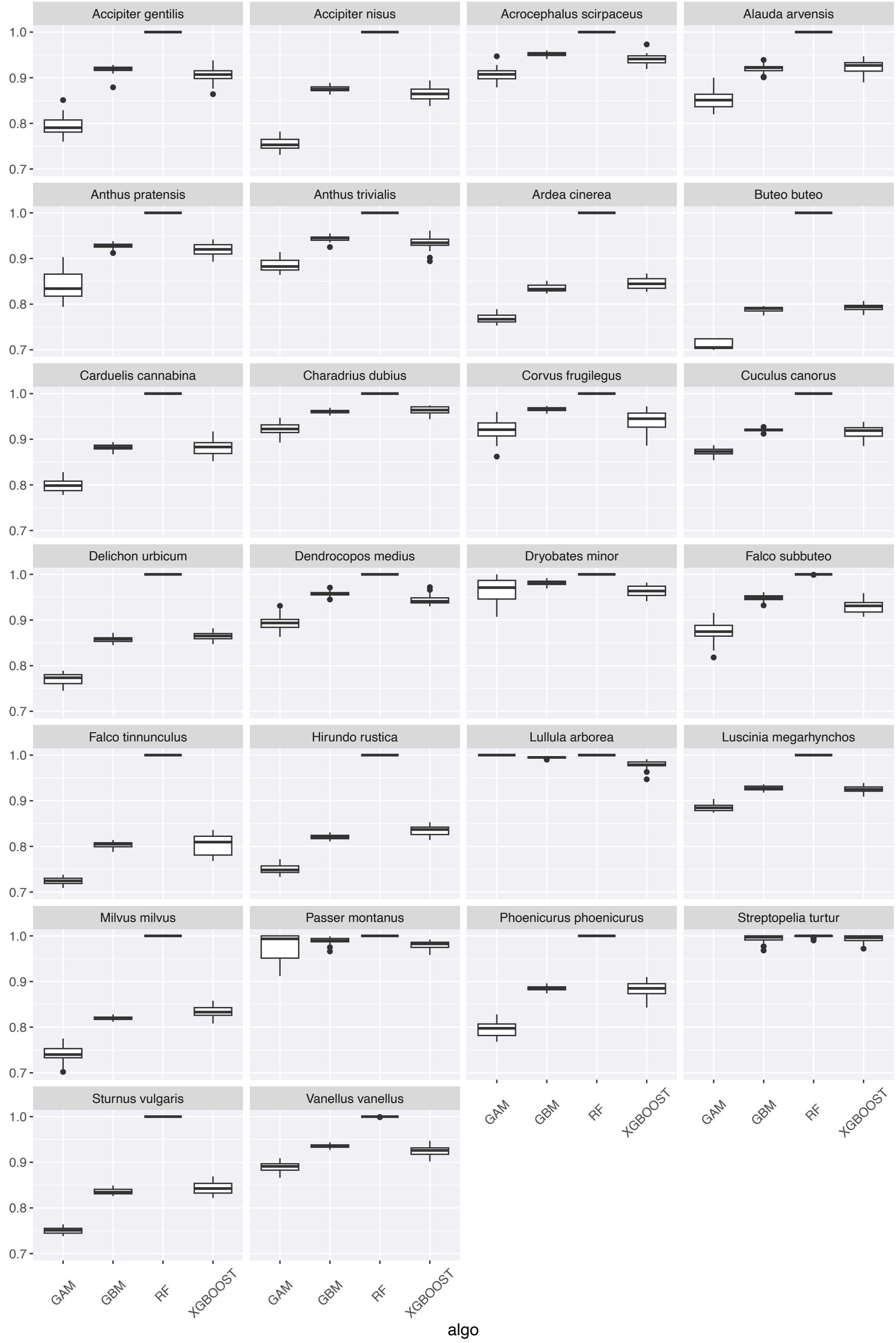

Fig. A3a: Results per species for applied algorithms evaluated by ROC

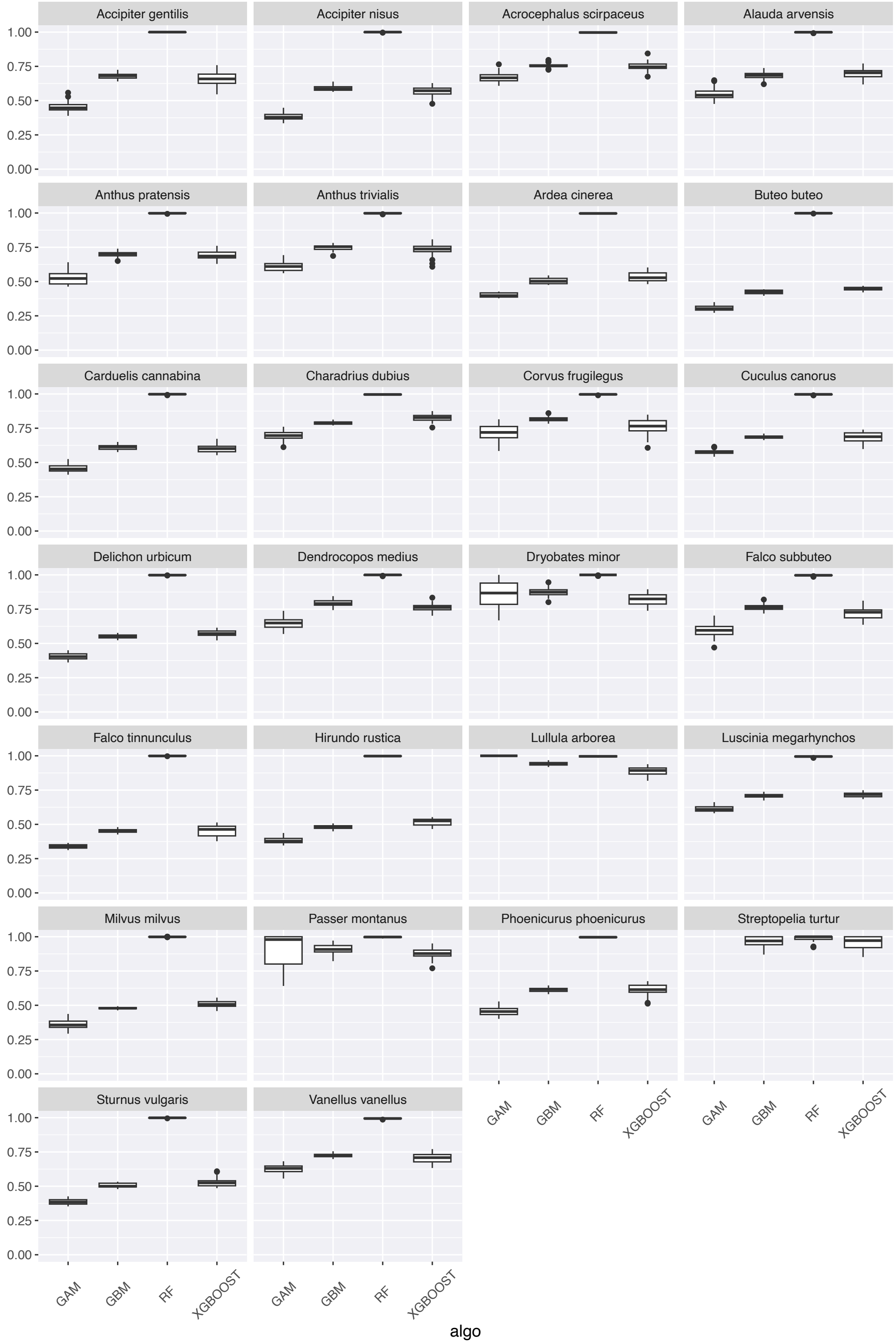

Fig. A3b: Results per species for applied algorithms evaluated by TSS

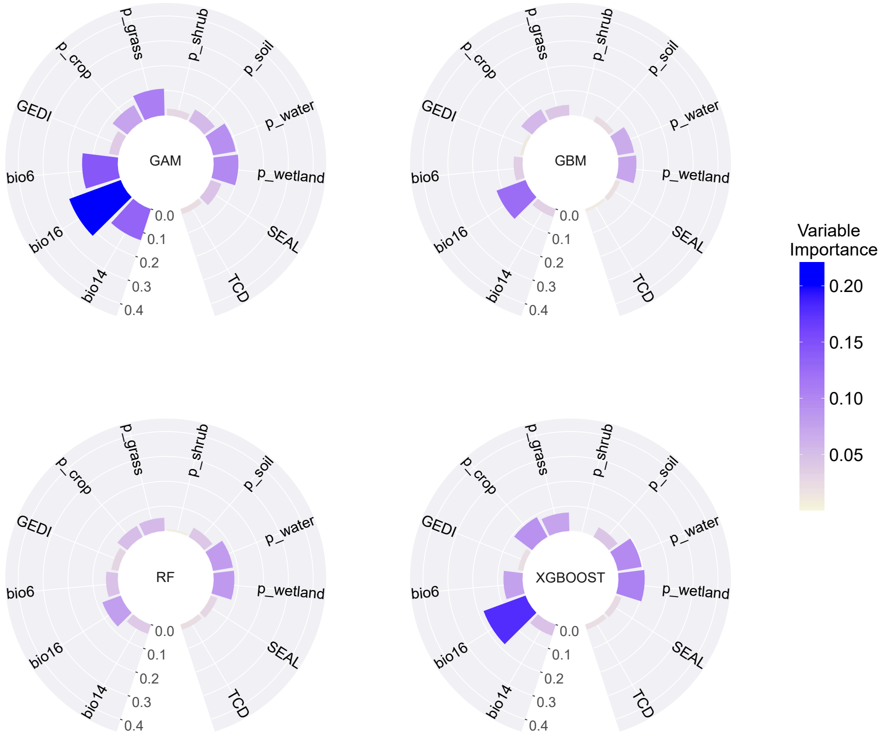

Fig. A4: Median Variable Importance per algorithm. Median contribution of all 12 predictor variables to model building regardless of the individual species. Darker colors indicate stronger importance of the predictor variable within the model.

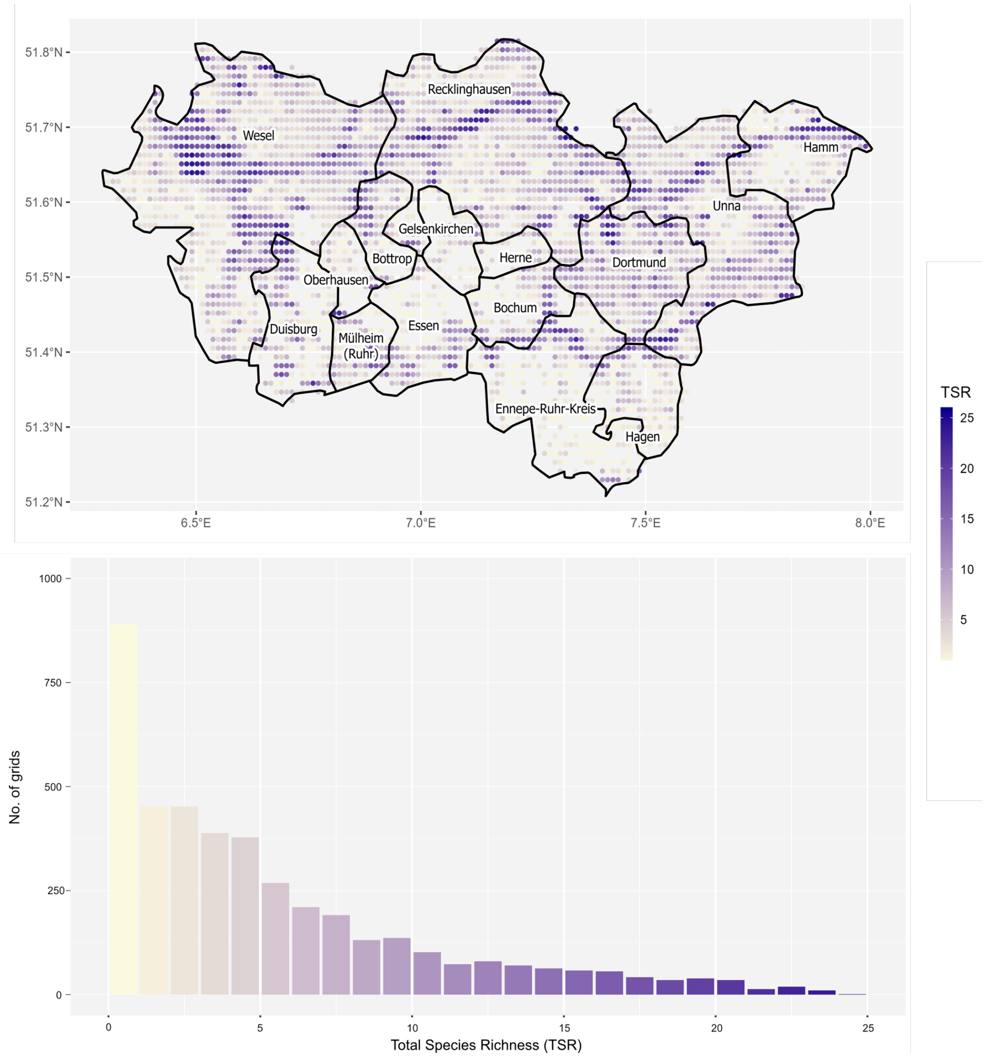

Fig. A5: Frequency distribution of total species richness (TSR). Total species richness (TSR) per grid cell. The more blue the raster cell is, the greater the number of predicted species within the raster cell. For the number of grids per TSR, Q95 indicates the 95% quantile (5%-based hotspot). Q90 and Q85 indicate the 90% and 85% quantiles, respectively.