Utilizing Remote Sensing Data for Species Distribution Modelling of Birds in Croatia

Andreja Radović[1,], Sven Kapelj[2], Louie Thomas Taylor[2]*

[1] Ruđer Bošković Institute, Department for Marine and Environmental Research, Laboratory for informatics and environmental modelling, Zagreb, Croatia, andreja.radovic@irb.hr
[2] BIOM, BirdLife Croatia, Zagreb, Croatia
* corresponding author

doi: 10.5281/zenodo.11657433

Abstract: This study focuses on modelling the distribution of bird species in Croatia using spatial variables derived from remote sensing satellite systems. The project involved data preparation, spatial analysis, and model creation to assess the entire territory of Croatia. Environmental variables were categorized into morphometric, habitat, bioclimatological, and landscape heterogeneity variables, all of which influence bird population distribution. The study utilized data from the European Space Agency (ESA) and NASA, including optical and radar sensors, to create a coverage map at a 10 m resolution. The study aslo employed the algorithm described for calculating habitat heterogeneity variables, including connectivity, diversity, and number of categories, crucial in spatial modelling of bird populations. Additionally, WorldClim bioclimatic variables were used, derived from temperature and precipitation data, reflecting the connection between bioclimatic conditions and vegetation types, and hence, bird populations. Modelling was conducted using the Maxent algorithm for habitat suitability assessment and a Random Forest algorithm (RF) for variable selection. All data preparation, manipulation, and model creation were performed in the R environment, utilizing various packages such as openxlsx, sf, terra RandomForest, and others. This interdisciplinary approach provides insights into the spatial distribution of bird populations in Croatia and demonstrates the utility of remote sensing data in species distribution modelling and habitat suitability assessments.

Keywords: bird; spatial modelling; Random Forest; Maxent; valorisation.

1 Introduction

The purpose of this research was to review the conservation objectives and measures for selected bird species (Table 1) in designated areas of the Ecological network of Croatia (Figure 1), based on species distribution modelling using all available data on birds in Croatia. Spatial valorisation of areas for bird species using species distribution modelling (SDM) is crucial for identifying critical habitats, assessing biodiversity, and planning effective conservation strategies, aligning with the objectives of the EU Birds Directive. By predicting species distributions and anticipating environmental changes, SDM informs targeted conservation efforts and sustainable land use planning, ensuring compliance with the Directive's goal of protecting all wild bird species naturally occurring in the European Union.

Table 1. Ecological groups on which habitat suitability spatial models were created

Ecological Group	Number
Woodpeckers	9
Birds of Prey	15
Rock Partridge	1
Nightjar	1
Hazel Grouse	1
Wetland Birds of Prey	3
Herons, Gulls, Spoonbills, Sacred Ibis,	15
Shoreline Nesters	2
Waterfowl, Grebes, Plovers, Coots, and	26
Songbirds	11
Waders and Cranes	16
Bitterns, Moorhens, Crakes, and Little	6
Mountain Owls	2
Western Capercaillie	1
Olive-tree Warbler	1

All analyses were based on data collected from all available sources of data on birds in Croatia available to authors and had a goal to assess, beside suitability of Croatian territory for the breeding, wintering and flyway populations of birds also the quality of data by producing spatial distribution models and evaluate it in help of expert-based system. We need the element of expert-based valorisation due to

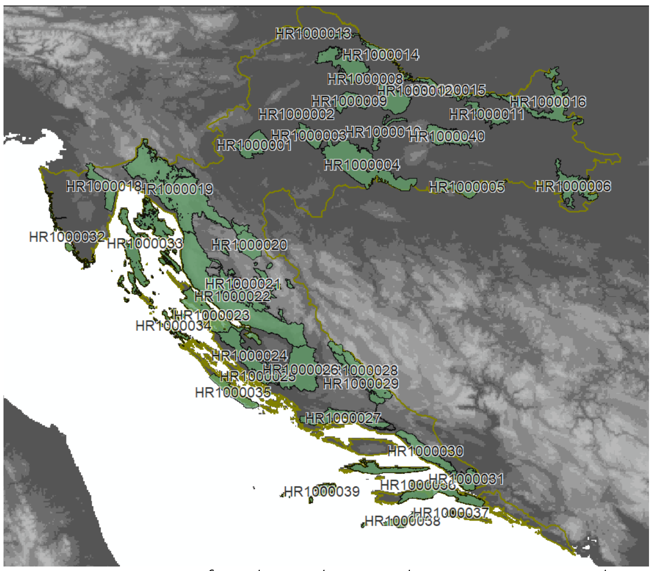

Figure 1. Position of Ecological network sites in Croatia where beside complete state territory, evaluation for each site is given, not presented in paper, available on request.

author's persuasion that for some species models cannot be used at all due to problems in the way data on birds are collected in time.

2 Materials and methods

The project combines spatial modeling with expert knowledge systems, utilizing spatial models with independent datasets on species presence within and beyond Croatia's borders. Observational data for all seasons were collected from the a) Global Biodiversity Information Facility (GBIF) (GBIF.org 09 July 2023; bird dataset 1) and b) from national sources BirdLife Croatia and Ministry of Economy and Sustainable Development and are data from numerous projects and research (bird dataset 2).

Only environmentally available variables, accessible to all for basic insights into Croatian bird populations' climatic conditions, were collected, aiding expert decision-making decisions based on modelling results. Modelling on national data are performed with breeding, wintering and flyover populations of 111 bird species. Results of Maxent model results base od GBIF datasets are available upon request and not presented in details in the paper.

Second part was the modelling bird populations from nationally collected data on birds. We prepared numerous environmental variables in order to detect, by using advanced machine learning methods, the most important set influencing detected population patterns. Data from satellite systems were obtained from EU Copernicus, NASA and JRC web services. The final goal was to prepare base for estimation population sizes as providing input for adequately defining conservation objectives and measures for insufficiently known bird species. Spatial datasets characterizing environmental conditions were methodically prepared at diverse spatial resolutions and harmoniously aggregated to meet specific requirements. All datasets and spatial models adhered to the projection system outlined by the EU Environment Agency EEA (EPSG: 3035) and only final results transformed to Croatia's official two-dimensional projection (EPSG: 3765). Prior to generating any spatial datasets, reference grids were established for the territory of Croatia. Reference raster grids of various spatial resolutions (10 m, 100 m, 500 m, 1,000 m, 10,000 m) were crafted to facilitate the creation of appropriate spatial datasets. Environmental variables selected for spatial model development can be classified into four categories: morphometric, habitat, bioclimatic, and area heterogeneity variables, all pivotal in understanding the spatial distribution of bird populations. These categories were informed by extensive scientific literature (Riitters et al. 2002, Fahrig 2003, Elith et al. 2006, Abatzoglou et al. 2018). Given the complexity of modeling numerous species and populations, a comprehensive suite of spatial datasets was prepared. Morphometric variables, including digital elevation models (EU-DEM (raster) version 1.1, Apr. 2016). and derivatives like the wetness index (WI) and slope were prepared. Habitat datasets largely relied on existing habitat maps of Croatia (Bardi et al. 2016). Two types of datasets were formulated: 1) indicating habitat presence or absence in reference grid units and 2) the surface area of each habitat type in reference grids. For model development, all nationally available data were utilized, encompassing data from projects such as the EU Natura 2000 Integration Project (NIP) (Mikulić et al. 2016) and CroFauna, alongside data collected during the SMART (Kapelj et al. 2024) project. Integrating these datasets required substantial data manipulation due to variations in data structures. Bioclimatic variables (Hijjmans et al. 2023) are utilized to generate biologically significant variables, crucial for modeling species distribution due to documented associations with vegetation types and bird populations.

Assessment of Croatia's territory for each species/population combination is obtained with both algorithms: 1) Maxent giving information of habitat suitability for the species in range 1–100; later reclassified bases on descriptive statistics onto 4 classes) and 2) classification type Random Forest (RF) for providing classification of Croatia's territory into two categories suitable (1) and unsuitable (0) with environmental variables that are the most important for obtained classification. For the purpose of the modelling procedure we defined, for all bird species and territory of Croatia seasons in the way: 1) breeding (April – August); 2) migration/flyway (September – December) and 3) wintering (January – March). This, not optimal way was the only possible for modelling such a big number of species and populations. The Maxent algorithm was implemented in two ways: using open species occurrence data from the GBIF database for all species in the project, 5000 observations were selected with a unique seed, covering all seasons (breeding, wintering, migration). Observations were divided by seasons. Background pseudo-absence points were generated in the same spatial extent, such as areas covering breeding points for each species. Environmental variables were probed at species occurrence and background points. The algorithm, through mathematical methods, attempts to recognize maximum entropy, pinpointing locations where environmental differences between presence and absence points are greatest. This model ultimately supported the expert-based component of defining species abundance concerning Croatia's position in the species' global range. Additionally, previously described variables specific to Croatia, describing habitats at the national classification level, were utilized. The result is a habitat suitability index for species/populations on a scale of 1–100, categorized into habitat quality classes. Models with an AUC statistic below 70% were not used in further analyses. The RF algorithm was used to classify Croatian territories into two classes: areas where breeding/wintering/migratory bird populations are likely to exist and areas where such populations are unlikely to be found. Confusion matrix statistics were used for model evaluation, focusing on the correct prediction of Class 1 (presence of species). The model was built on a random forest of 1,000 classification trees. As the algorithm builds classification trees by selecting a smaller subset of variables, the absolute importance of variables in describing the detected species distribution is determined. With 1000 trees, adequate replication was ensured, enabling the calculation of relevant statistics for model evaluation. The traditional 70% data training and 30% testing split, common in statistical model development, was not feasible due to very small data sets, and it was unnecessary given the large number of classification trees for error calculation, so we used only 15% of data for model evaluation.

3 Results

Results are presented for only one species, one population – breeding population of *Microcarbo pygmaeus*, as example. The species is one whose distribution is mainly defined by climatological (bioclimatological) variables with 10 most important variables identified by the RF classification algorithm model as decisive for classifying a pixel at 1 km spatial resolution as suitable habitat (1) or unsuitable habitat (0) are as follows: 1) elevation; 2) bio5; 3) area under habitat class A4 national classification - Reeds, cattails, tall sedges, and tall rushes; 4) bio9; 5) bio18; 6) bio11; 7) bio11; 8) bio3; 9) bio10 and 10) area under habitat class A1 national classification - Temporary standing water bodies (description of bioclim variables at: https://www.worldclim.org/data/bioclim.html) (Figures 2, 3). The modeling results from a) the Maxent algorithm (HIS values) and b) the RF classification algorithm (0,1). Maxent values are then reclassified into 4 classes of suitability over Croatian territory and overall statistics given for complete territory and separately for each Ecological network site for further comparison.

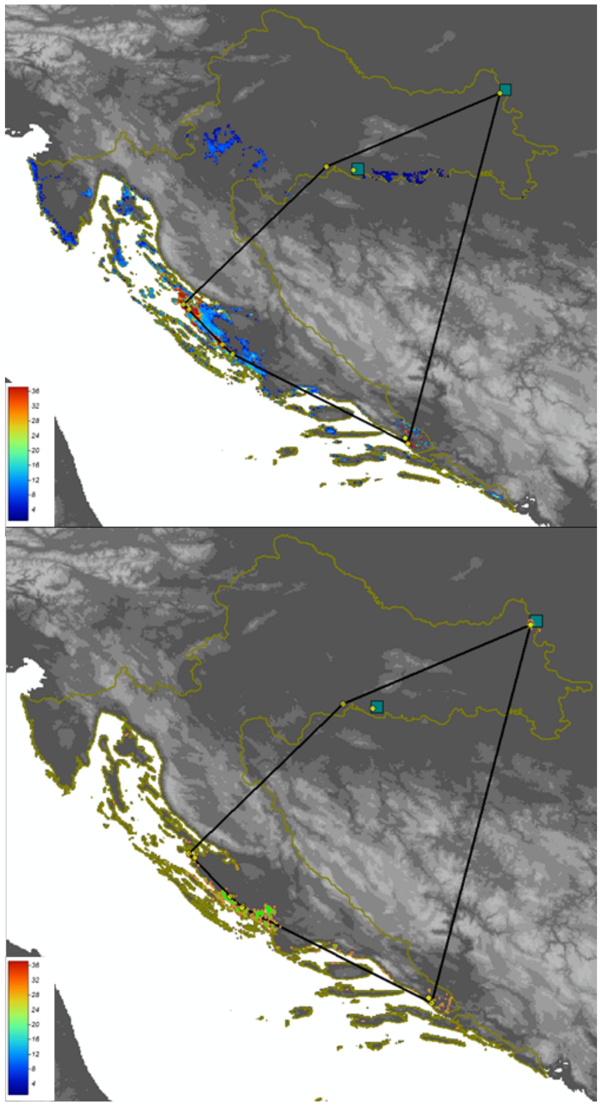

Figure 2. Breeding population habitat suitability for species Microcarbo pygmaeus base map – DEM, top) Maxent algorithm and bottom) classification RF (green), national data not older than 25 years (yellow circles) with convex hull around (black polygon); data from last 5 years (green squares); Maxent valorization of Croatian territory with legend

4 Discussion

Developing spatial models for individual populations in this project presented several challenges or factors to consider when interpreting modeling results. Bird data were not collected over time systematically (covering complete set of environmental gradients) or completely randomly, inherently carrying errors in models violating the assumptions the data for the model should meet. There is an error in data collected for territorial birds distributed across complete territory with evident lack of data on areas

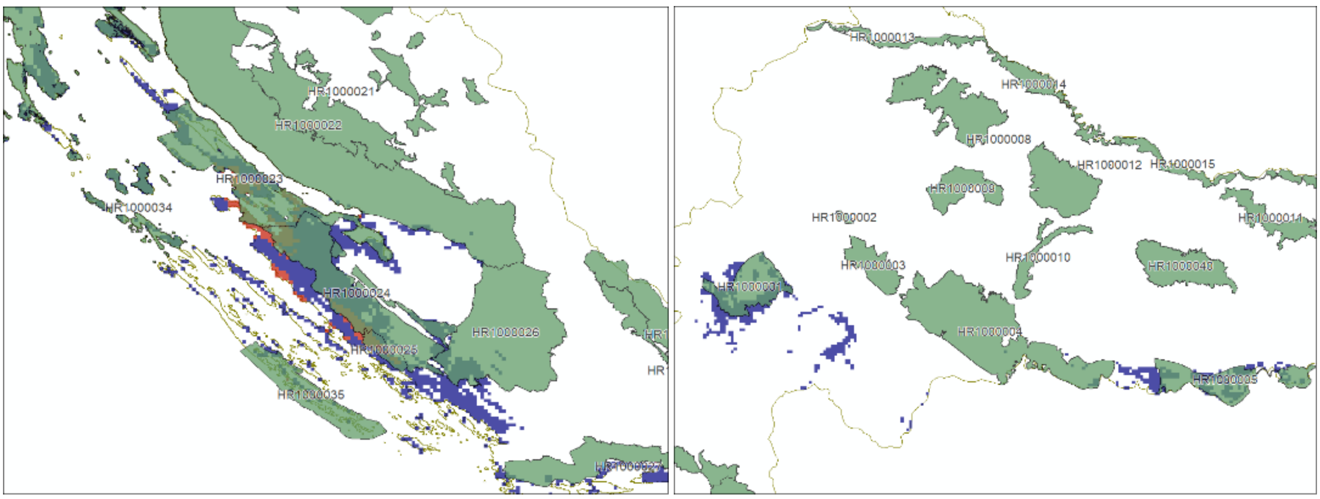

Figure 3. Final statistics of Maxent algorithm habitat suitability valorization at national scale for breeding population of Microcarbo pygmaeus zoom at 1) North Dalmatia (left) and 2) central Croatia (right).

that are less interesting to researchers. On the modeling resolution (1 km), the rarefaction curve (results not presented here) did not reach an asymptote on a five-year scale for most of the species/habitat at 5 km scale, indicating insufficient investigation of habitat types. The inconsistent collection of data on species absence requires careful scrutiny of resulting findings. Data are gathered only in environmental conditions deemed suitable for the species, lacking representation across the entire feature space. Uneven sampling efforts, correlated with environmental gradients, contribute to the heterogeneity of recorded species. Furthermore, modeling a large set of populations in the same environment presents challenges, requiring unique parameters for each species across different seasons and geographic locations. Despite data limitations, models were developed using presence-only data. Given these constraints, reliance on fauna and flora databases in Croatia is essential for interpreting analyses and results. Addressing these concerns, we advocate for improved data collection practices within the Croatian scientific community and encourage alignment with established procedures to enhance the reliability of mathematical modeling, particularly in mitigating sampling bias, crucial for algorithms like Maxent and Random Forest.

5 References

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. Scientific Data 5, 170191.

Bardi, A., Papini, P., Quaglino, E., Biondi, E., Topić, J., Milović, M., Pandža, M., Kaligarič, M., Oriolo, G., Roland, V., Batina, A., Kirin, T., 2016. Karta prirodnih i poluprirodnih ne-šumskih kopnenih i slatkovodnih staništa Republike Hrvatske. AGRISTUDIO s.r.l., TEMI S.r.l., TIMESIS S.r.l., HAOP.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2), 129-151.

Elith, J., Graham, C.H., ... 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2), 129-151.

Fahrig, L., 2003. Effects of habitat fragmentation on biodiversity. Annual Review of Ecology, Evolution, and Systematics 34 (1), 487-515.

Hijmans, R.J., Barbosa, M., Ghosh, A., Mandel, A., 2023. geodata: Download.

Kass, J.M., Muscarella, R., Galante, P.J., Bohl, C.L., Pinilla-Buitrago, G.E., Boria, R.A., Kapelj, S., Radović, A., Zec ... 2023. Završno izvješće Usluge definiranja SMART ciljeva očuvanja i osnovnih mjera očuvanja ciljnih vrsta i stanišnih tipova - Grupa 5: Definiranje ciljeva i mjera očuvanja za nedovoljno poznate vrste ptica, Udruga BIOM, Geonatura, DOPPS, Zagreb, 36.

Mikulić, K., Kapelj, S., Zec, ... 2016. Završno izvješće za skupinu Aves. In: Mrakovčić, M., Mustafić, P., Jelić, D., Mikulić, K., Mazija, M., Maguire, I., Šašić Kljajo, M., Kotarac, M., Popijač, A., Kučinić, M., Mesić, Z. (Eds.), Projekt integracije u EU Natura 2000 - Terensko istraživanje i laboratorijska analiza novoprikupljenih inventarizacijskih podataka za taksonomske skupine: Actinopterygii i Cephalaspidomorphi, Amphibia i Reptilia, Aves, Chiroptera, Decapoda, Lepidoptera, Odonata, Plecoptera, Trichoptera. OIKON-HID-HYLA-NATURA-BIOM-CKFF-GEONATURA-HPM-TRAGUS, Zagreb, pp. 1-49.

Riitters, K.H., Wickham, J.D., O'Neill, R.V., Jones, K.B., Smith, E.R., Coulston, J.W., Wade, T.G., Smith, J.H., 2002. Fragmentation of continental United States forests. Ecosystems 5, 815-822.