D4.3
Ensemble species distribution models for all study areas

Deliverable for the Horizon Europe Project BirdWatch

Version 1.0

Legal Disclaimer

Document Information

GA Number	101082634	**Type of Action**	Horizon-IA
Full Title	BirdWatch - a Copernicus-based service for the improvement of habitat suitability of farmland birds via satellite-enabled monitoring, evaluation and optimisation of CAP greening measures		
Project Acronym	BirdWatch		
Start Date	February 1st, 2023	**Duration**	36 Months
Project URL	https://birdwatch-europe.org/		
Deliverable	D4.3 Ensemble species distribution models for all study areas*		
Work Package	WP4		
Project Month of Delivery	**Contractual** M28	**Actual**	M34
Nature	Report	**Dissemination Level**	PUB
Lead Beneficiary	University of Potsdam (UP)		
Responsible Author	Levin Wiedenroth - UP		
Contributions from	Levin Wiedenroth - UP Damaris Zurell - UP Emma Underwood - UP Johanne Burmeister - UP Josh Hauer – UP Pedro J. Leitão - LUP Sascha Gey - LUP		

* Please note that this deliverable was originally suggested as "D4.3: Ensemble and joint species distribution models for all study areas". As detailed in deliverable D4.1, joint distribution models were discarded from the further workflow and are thus not considered here.

History of Changes

Version	Issue Date	Stage	Description	Comments	Contributor
1.0	25.11.2025	Draft	First version of D4.3		Levin Wiedenroth Damaris Zurell Emma Underwood Johanne Burmeister Josh Hauer Pedro J. Leitão Sascha Gey

Table of Contents

1. Introduction

Here we present the third deliverable of work package 4000, which applies the workflows discussed in deliverable 4.1 “Data, algorithms, and workflows of SDM” for three test regions, i.e. Germany, Lithuania and South-Tyrol, while results for the test region Flanders had already been presented in deliverable 4.2 “D4.2 Ensemble species distribution models estimated and validated for first study region”. We present the results of the individual model building steps: (1) data preparation and model training, (2) model validation, and (3) model predictions. The models and final predictions of habitat suitability will then feed into the BirdWatch optimization algorithm in WP5000.

2. Data and Methods

As detailed in deliverables 4.1 and 4.2, the Birdwatch work package WP4000 uses species distribution models (SDMs) to relate species occurrence data to environmental predictors and predict bird habitat suitability (Guisan et al., 2017). We utilised a nested SDM framework as described in deliverable 4.1 in which we combine European-wide, coarse-grained SDMs of climate suitability at 50 km spatial resolution with fine-grained SDMs describing suitability of land use practices and intensity at 200 m spatial resolution. At both spatial scales, occurrence data were spatially thinned prior to modelling with the thinning distance set twice as large as the spatial resolution for each respective scale i.e., 100 km for the coarse-grained European data, and 400 m for the fine-grained region-specific data. The workflow for the SDMs included a variable pre-selection to avoid problems with multicollinearity, fitting of four different SDM algorithms (generalized linear models - GLMs, generalized additive models - GAMs, random forest - RF, boosted regression trees - BRTs), model validation using a five-fold spatial block cross-validation and four different performance metrics (AUC, TSS, sensitivity, specificity), calculation of a binarization threshold based on maximising TSS, ensemble building (using mean probability of occurrence across algorithms as well as committee average), and prediction (cf. deliverables D4.1-4.2, S1). As described below, there are three deviations from the previous workflow. First, we discarded Maxent from the SDM ensemble for these study regions because all occurrence data stem from standardised survey data and thus provide presence-absence data while Maxent is typically applied to opportunistic data and large sets of background data. Second, we changed the climate data input for the European SDMs from CHELSA to ERA5. Last, we separated the effect of land cover variables in the SDMs and the effect of land use intensity, specifically mowing frequency and mowing date.

The European SDMs were already developed in deliverable 4.2 but were updated because the climatic input data changed. Specifically, deliverable 4.2 used the CHELSA climate dataset (Climatologies at high resolution for the earth’s land surface areas; Karger et al., 2017, 2021), which provides monthly climate data for the years 1980-2019. This period coincided well with the land cover data for Flanders that were compiled for the year 2018 (cf. deliverable 4.2). However, as described in deliverable 4.1, the more recent year 2022 was chosen as target year for Germany, Lithuania and South-Tyrol because in that year all bird species had comparably high numbers of occurrences to build SDMs. This change in target year also necessitated a change in the underlying climate data to match the time period under study. Consequently, we now directly use the hourly ERA5 reanalysis data from the Copernicus Climate Change Service (Hersbach et al., 2023a, 2023b), available at 0.25° spatial resolution. We downloaded three hourly daily temperature values (2 meters above ground level) as well as monthly total precipitation values for the years 2012-2022. This period covers the time span of the European bird data (EBBA2, cf. deliverable 4.1) until the most recent year of analysis. The EBBA2 observation period was 2013 to 2017, and we additionally included the year preceding the start of observations to account for lag effects of weather on a species distribution (Sofaer et al., 2018). Using the ERA5 temperature data we calculated monthly minimum, average, and maximum temperature. The average ERA5 monthly precipitation values were multiplied with the number of days for each respective month to calculate the monthly total precipitation sum. Using the averaged climatic predictors across the time period we calculated 19 bioclimatic variables commonly used in SDMs (Booth et al., 2014). To match the

EBBA2 data we reprojected the climate data to the LAEA equal area projection (EPSG: 3035) and aggregated them to a 50 x 50 km resolution matching the EBBA2 grid. European SDMs were then re-calibrated using the ERA5 data from 2012-2017 and then predicted climatic suitability for the period 2017-2022, in line with the approach described in deliverables D4.1-4.2. As a further complication, we found that nesting the European climatic suitability within the fine-scale regional models through covariate nesting introduced artefacts in the fine-scale habitat suitability based on machine-learning, and we hence had to design a new nesting approach based on meta-learner ensembles that were introduced in Oeser et al., (2024). These clearly separate the different scales, meaning that climate is no longer included as covariate in the fine-scale model but only the final predictions are ensembled. As the climate nesting is mostly relevant for cross-predicting between regions (Deliverable 4.4), we here separately present the European-scale climate model predictions (section 3.1) and the fine-scale habitat predictions (sections 3.2-3.4).

As described in deliverable D4.1, the fine-scale data of bird occurrences for Germany and Lithuania stem from standardised bird survey data including presence and absence data. For Germany, data were provided from Monitoring of Common Breeding Birds (MCB) by the German avifaunistic umbrella organisation DDA ("Dachverband Deutscher Avifaunisten"). For Lithuania, data were obtained from the Common Bird Population Abundance Monitoring project of the Lithuanian Ornithological Society. For South-Tyrol, point count data were provided by the Museum of Nature South Tyrol and regional surveys but had overall very few occurrences. After spatial thinning, limited sample sizes allowed to model only three species: Yellowhammer, Eurasian skylark, and Red-backed shrike. A summary of the final number of occurrences for the different regions is shown in Table 1. Maps of the final presence points for these regions are shown in Figures 1-3. The European occurrence data were already presented in deliverable D4.2 and are not repeated here.

As a deviation to the previous workflow described in deliverables D4.1-4.2, we now excluded the land use intensity variables mowing frequency and mowing intensity from the fine-scale SDMs and instead separately quantified the effect of these variables on fine-scale bird occurrence probability. This step was necessary as for Germany, the mowing regime was only quantified for grassland areas while previously for Flanders, mowing regime was quantified for the entire study region although being less informative for non-grassland plots. We now harmonized the procedure across all study regions by first calibrating fine-scale SDMs with climatic suitability and land cover variables as predictors and then separately analysing how mowing frequency and mowing date increased or decreased habitat suitability in grassland plots. For the latter analyses, we selected only 200 m cells with at least 50% grassland coverage and then fitted random forest models (with 1000 trees and nodesize = 10), with bird occurrence as response and SDM-derived fine-scale habitat suitability, mowing frequency and mowing date as predictors. Separating these steps allowed us to quantify bird habitat suitability across all cells in the study region and additionally predict the effect of mowing regimes in grassland cells, which is an important effect to consider in the land use optimization in WP5000.

Table 1: Occurrence numbers after thinning over all study regions. The * indicates species excluded from model construction for at least one of the regions due to a low number of occurrences.

Species	Germany presences	Germany absences	Lithuania presences	Lithuania absences	South-Tyrol presences	South-Tyrol absences
Eurasian skylark (*Alauda arvensis*)	1737	476	719	714	45	573
Meadow pipit (*Anthus pratensis*)	198	970	154	1131	0*	0
Yellowhammer (*Emberiza citrinella*)	2013	358	601	627	70	900
Red-backed shrike (*Lanius collurio*)	574	702	73	1199	109	1825
Black-tailed godwit (*Limosa limosa*)	0*	0	3*	1723	0*	0
Tree sparrow *(Passer montanus)*	878	616	87	1121	16*	190
Whinchat (*Saxicola rubetra*)	146	975	396	809	22*	280
European turtle dove (*Streptopelia turtur*)	124	1007	13*	1630	0*	0
Common starling (*Sturnus vulgaris*)	2290	224	350	775	12*	130

Northern lapwing (*Vanellus vanellus*)	174	983	155	1017	0*	0

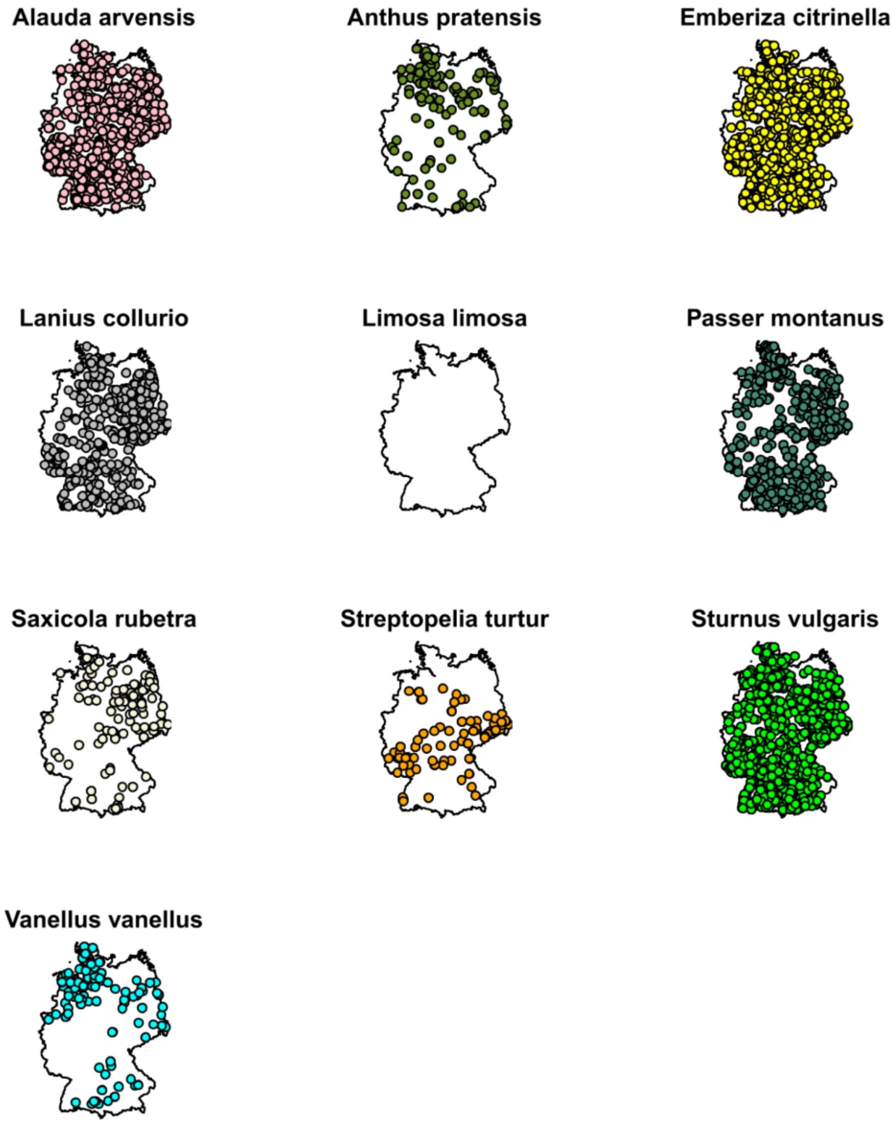


Figure 1: Final occurrence data of study species in Germany. Note that *Limosa limosa* did not have enough presence records to be included in the SDM workflow.

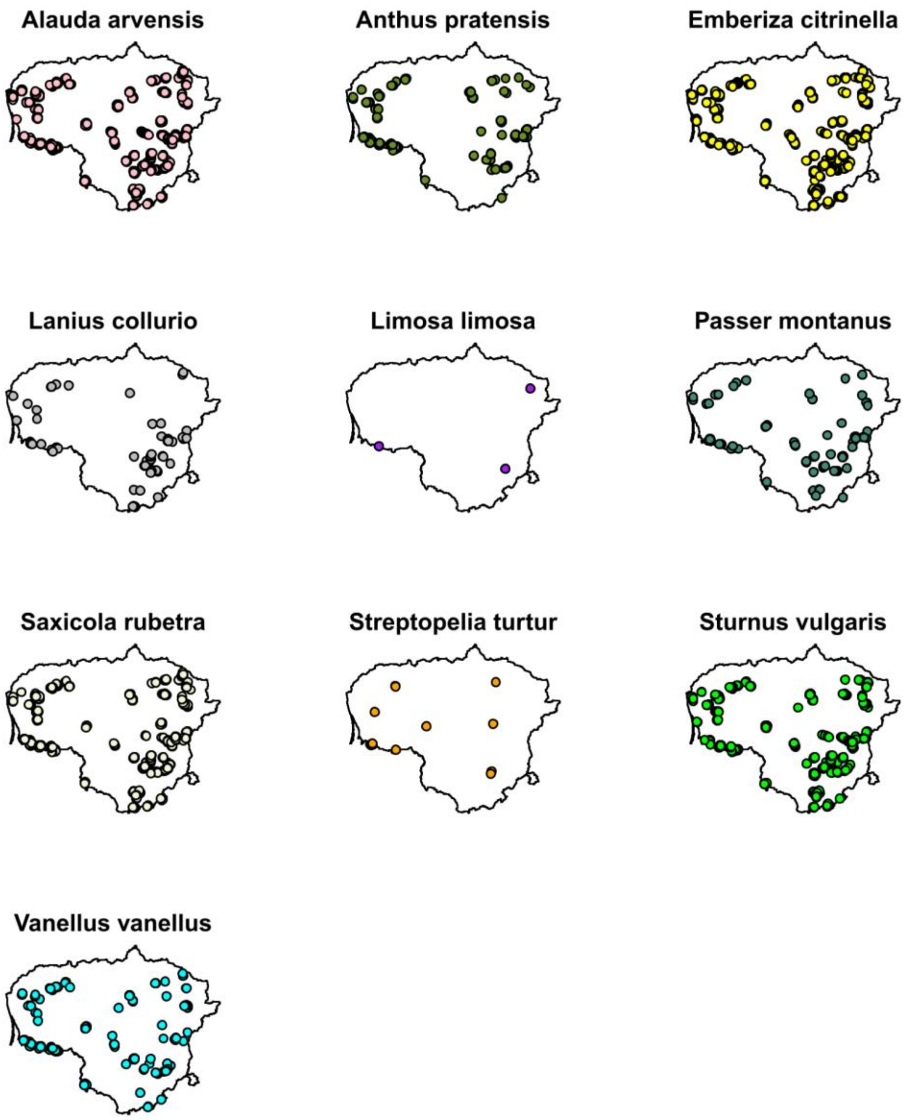


Figure 2: Final occurrence data of study species in Lithuania. Note that *Limosa limosa* and *Streptopelia turtur* did not have enough presence records to be included in the SDM workflow.

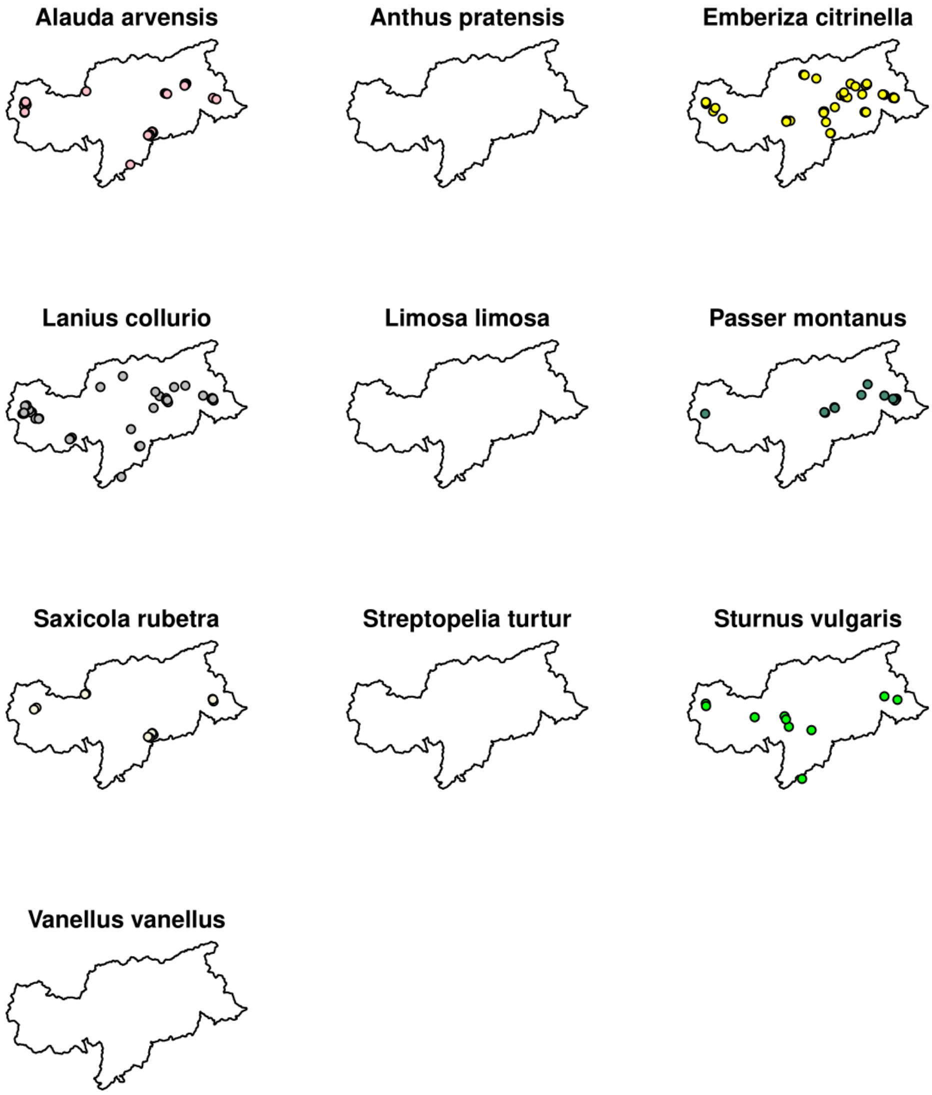


Figure 3: Final occurrence data of study species in South-Tyrol. Note that *Anthus pratensis*, *Limosa limosa, Passer montanus, Saxicola rubetra, Streptopelia turtur, Sturnus vulgaris* and *Vanellus vanellus* did not have enough presence records to be included in the SDM workflow.

3. Results

3.1 Europe

As we re-calibrated the European SDM using the ERA5 data, we here again summarize the model performance and predicted climatic suitability for completeness (cf. deliverable D4.2).

Model performance

Model performance changes only marginally between the European SDM fitted to CHELSA climate data presented in deliverable D4.2 and the SDM fitted to ERA5 climate data. All ensemble models showed very good to excellent predictive performance with AUC scores ranging 0.83 - 0.93 and TSS ranging 0.54-0.73 (Table 2). Sensitivity values ranged 0.76 - 0.86 and specificity ranged 0.66 - 0.87 (Table 3). Only the Black-tailed godwit had specificity values below 0.75 indicating slight overprediction of their European range.

Table 2: Performance metrics of the European mean probability ensemble. The ensemble includes the algorithms: GLM, GAM, RF, and BRT. The threshold was used to binarize the continuous habitat suitability predictions and convert them into predicted presences and absences. It is the threshold that maximizes TSS (or the sum of sensitivity and specificity) in the cross-validated predictions.

Species	AUC	TSS	Sensitivity	Specificity	Explained Deviance	Threshold
Eurasian skylark (*Alauda arvensis*)	0.91	0.67	0.81	0.87	0.68	0.61
Meadow pipit (*Anthus pratensis*)	0.93	0.73	0.86	0.87	0.73	0.52
Yellowhammer (*Emberiza citrinella*)	0.90	0.67	0.83	0.84	0.72	0.43
Red-backed Shrike (*Lanius collurio*)	0.90	0.62	0.77	0.85	0.71	0.69
Black-tailed godwit (*Limosa limosa*)	0.83	0.54	0.86	0.68	0.48	0.16
Tree sparrow (*Passer montanus)*	0.88	0.62	0.77	0.85	0.63	0.70
Whinchat (*Saxicola rubetra*)	0.91	0.67	0.87	0.79	0.74	0.55
European turtle dove (*Streptopelia turtur*)	0.90	0.64	0.78	0.86	0.66	0.60
Common starling (*Sturnus vulgaris*)	0.89	0.62	0.76	0.86	0.69	0.84
Northern lapwing (*Vanellus vanellus*)	0.90	0.61	0.80	0.81	0.62	0.57

Model predictions

The final maps of predicted climate suitability in Europe based on ERA5 data are congruent with the models and resulting maps based on CHELSA data in deliverable D4.2. Additionally, we now also provide the climate suitability predictions for the Red-backed shrike and the Whinchat, which were not included in deliverable D4.2 for the Flanders region as these species did not have enough occurrence records in Flanders to allow fitting SDMs. Both species have widespread climate suitability across all central, eastern, and western parts of Europe (Figure 4).

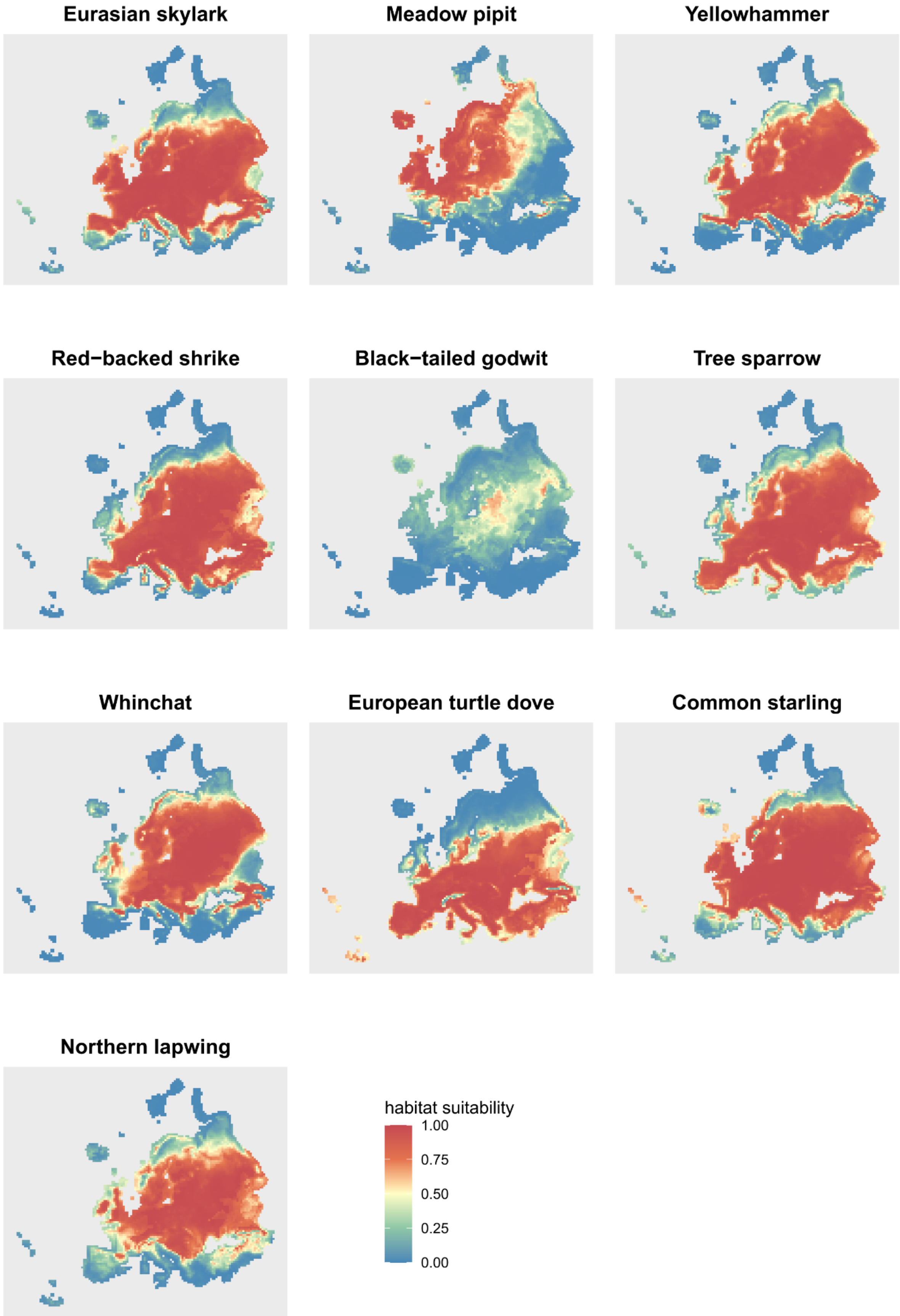

Eurasian skylark
Meadow pipit
Yellowhammer
Red−backed shrike
Black−tailed godwit
Tree sparrow
Whinchat
European turtle dove
Common starling
Northern lapwing
habitat suitability
1.00
0.75
0.50
0.25
0.00

Figure 4: Climatic suitability predictions for Europe based on the mean probability ensemble per species. The ensembles include the model algorithms GLM, GAM, RF, and BRT.

3.2 Germany

Model performance

For Germany, all fine-scale habitat ensemble models had fair to excellent predictive performance with AUC ranging 0.81 - 0.95 (Table 3). The TSS, sensitivity and specificity scores also indicated fair to excellent predictive accuracy ranging from 0.49 - 0.78, 0.75 - 0.89 and 0.64 - 0.89, respectively (Table 3). For the Red-backed shrike and the Tree sparrow, absences were predicted less accurately than for other species.

Table 3: Performance metrics of the German mean probability ensemble. The ensemble includes the algorithms: GLM, GAM, RF, and BRT. The threshold was used to binarize the continuous habitat suitability predictions and convert them into predicted presences and absences. It is the threshold that maximizes TSS in the cross-validated predictions.

Species	AUC	TSS	Sensitivity	Specificity	Explained Deviance	Threshold
Eurasian skylark (*Alauda arvensis*)	0.95	0.78	0.89	0.89	0.69	0.80
Meadow pipit (*Anthus pratensis*)	0.87	0.60	0.86	0.74	0.54	0.13
Yellowhammer (*Emberiza citrinella*)	0.90	0.66	0.83	0.82	0.54	0.85
Red-backed Shrike (*Lanius collurio*)	0.82	0.49	0.85	0.64	0.41	0.42
Tree sparrow (*Passer montanus)*	0.84	0.53	0.86	0.67	0.44	0.56
Whinchat (*Saxicola rubetra*)	0.87	0.57	0.75	0.82	0.51	0.16

European turtle dove (*Streptopelia turtur*)	0.81	0.49	0.77	0.71	0.45	0.11
Common starling (*Sturnus vulgaris*)	0.86	0.58	0.78	0.80	0.52	0.92
Northern lapwing (*Vanellus vanellus*)	0.93	0.72	0.88	0.84	0.58	0.15

Model predictions

For each species, we predicted the ensemble mean habitat suitability for Germany (Figure 5). Among the studied species, Yellowhammer, Common starling and Eurasian skylark are most common across Germany and showed high habitat suitability across all parts of Germany. Also, for Tree sparrow and Red-backed shrike suitable habitat is predicted across large parts of Germany, while for Meadow pipit, Northern lapwing and Whinchat suitable habitat is found mainly in northern Germany. For European turtle dove, habitat suitability is generally low with highest suitability in Central Germany. According to the committee averages of the binary predictions, the lowest amount of suitable area across Germany is predicted for the Meadow pipit, and the largest for Yellowhammer, Red-backed shrike and Starling (Figure 6). Standard deviations in predicted habitat suitability across model algorithms are low (Figure 7). Differences in habitat preferences are also visible in the species-specific response curves (Figure 8).

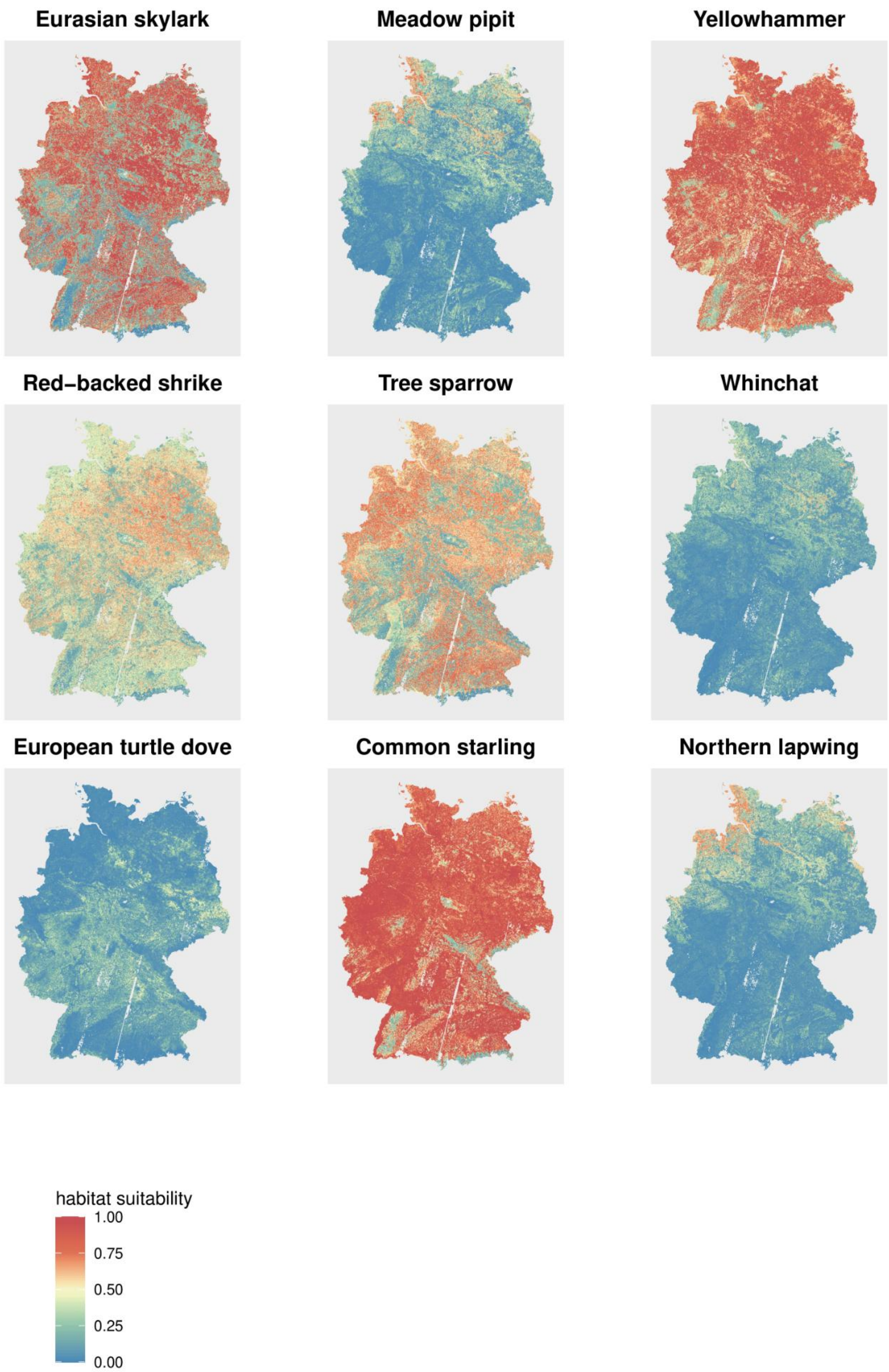


Figure 5: Habitat suitability predictions for Germany based on the mean probability ensemble per species. The ensembles include the model algorithms GLM, GAM, RF, and BRT.

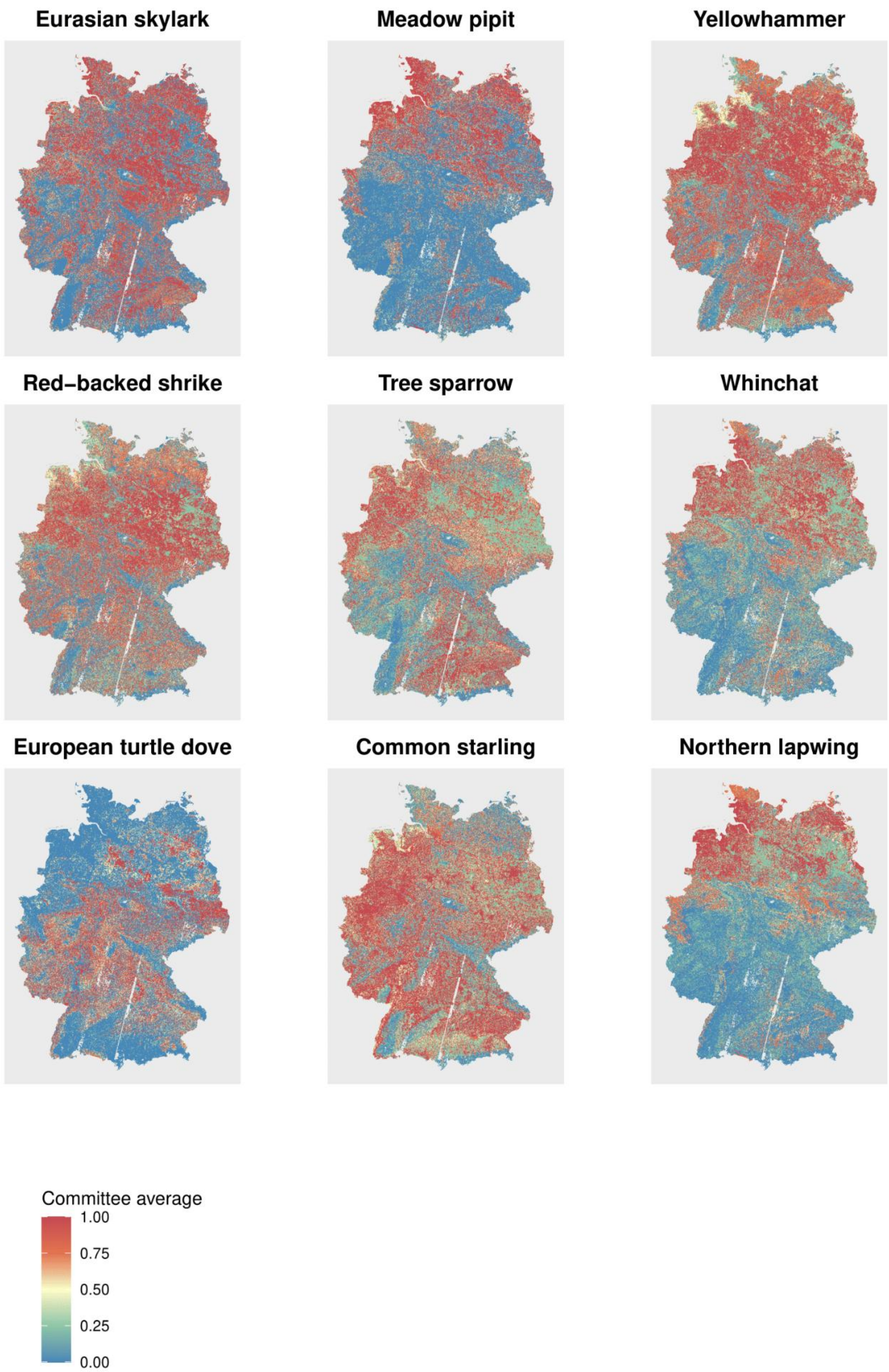


Figure 6: Committee averaged predictions for Germany per species. Committee averages indicate the proportion of algorithms agreeing on a species being present or absent. The model algorithms GLM, GAM, RF, and BRT were used in the ensembles.

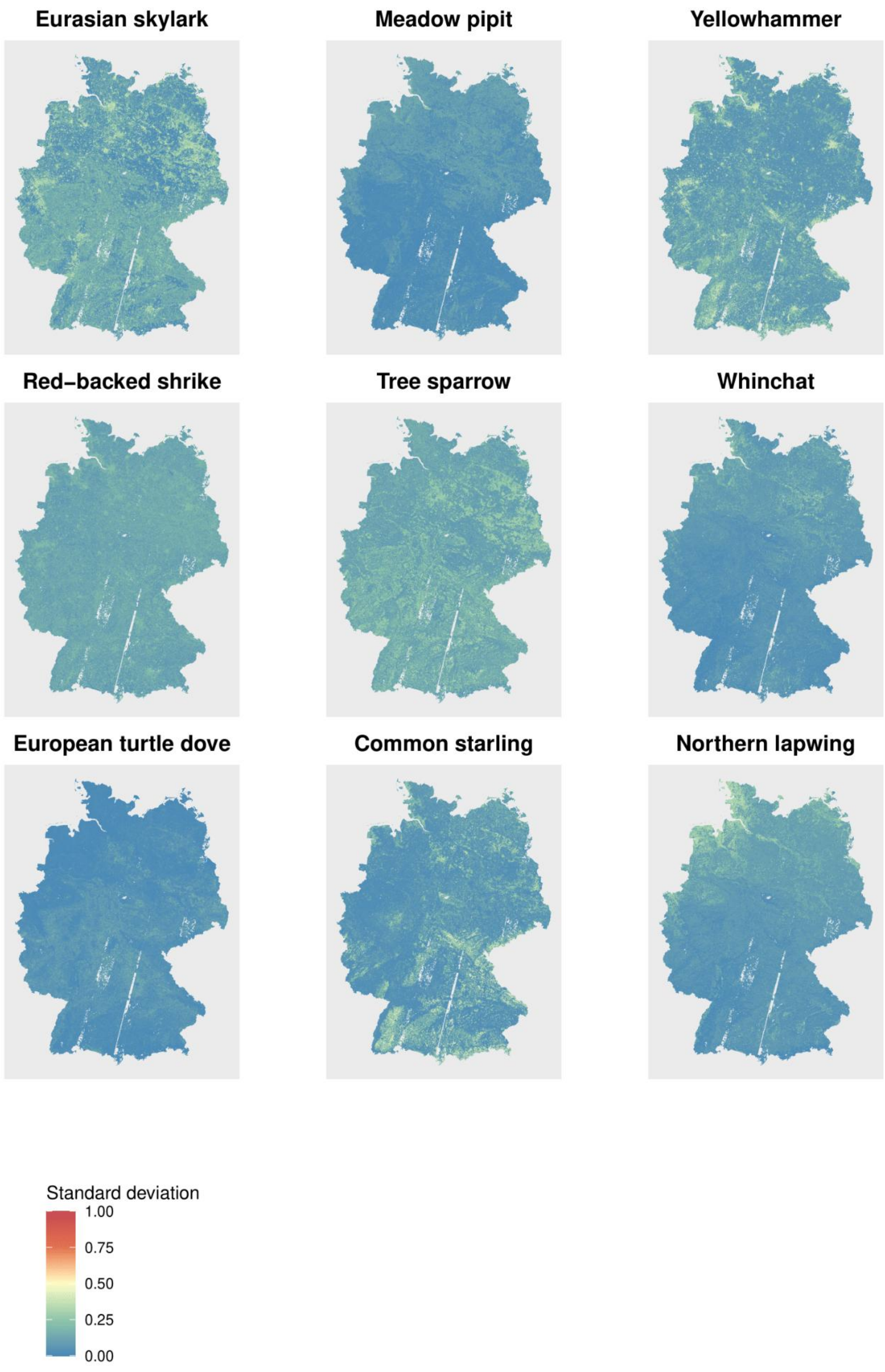


Figure 7: Standard deviation calculated from the habitat suitability predictions for Germany by the model algorithms (GLM, GAM, RF, and BRT).

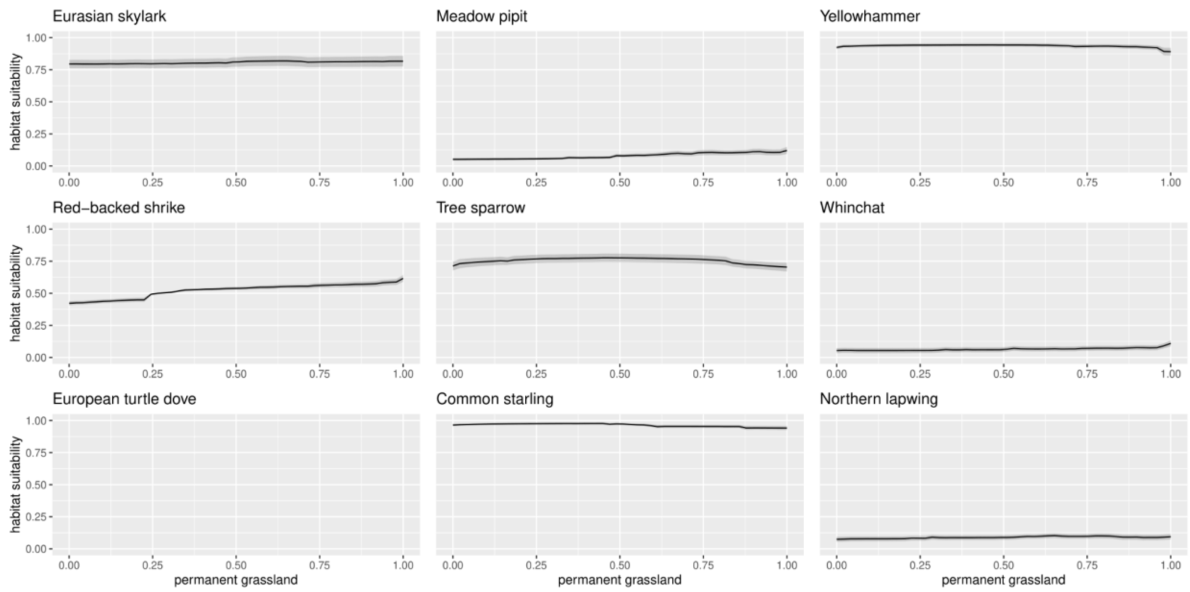


Figure 8: Partial response curves per species and Germany for the environmental parameter "permanent grassland" given in % cover. The solid line represents the mean ensemble predictions, the gray bands around the line plot represent the standard error across the SDM algorithm.

Mowing effects

For easier interpretation, we show the effects of mowing frequency and first mowing date on occurrence probability of all species for an average habitat suitability of 0.8 in grasslands (Figure 9). We see that mowing effects can be substantial for many species, such that occurrence probability of the species increases to almost 1.0 or drops as low as 0.25 in grasslands that otherwise have an average habitat suitability of 0.8. The Eurasian skylark, Meadow pipit, Red-backed shrike, and Tree sparrow benefit most from early mowing dates, although a similar but less pronounced pattern was also found for the Yellowhammer and Northern lapwing. The Eurasian skylark, Red-backed shrike, and Tree sparrow benefit from medium to low mowing frequency while the European turtle dove and Whinchat benefit from medium to high mowing frequency.

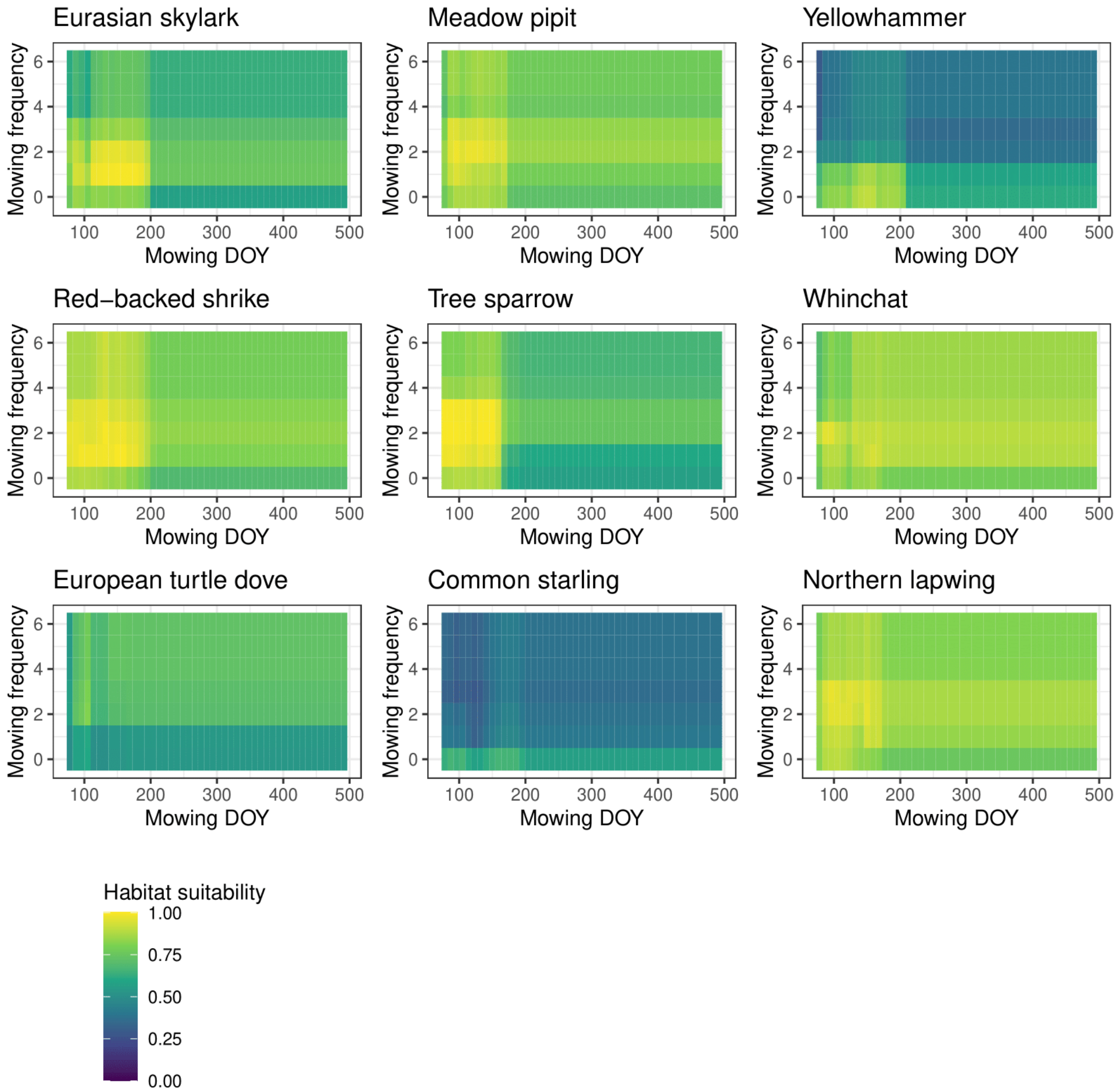


Figure 9: Mowing level plots indicating the effect of mowing frequency and first mowing date (DOY: day of year) on habitat suitability across species for Germany. The plots refer to a land-use suitability of 0.8.

3.3 Lithuania

Model performance

For Lithuania, the ensemble models generally performed poorer than for Germany. All ensemble models had fair to very good predictive performance with AUC ranging 0.73 - 0.81 (Table 4). The TSS, sensitivity and specificity scores ranged from 0.38 - 0.52, 0.59 - 0.84 and 0.66 - 0.86, respectively (Table 4). Especially for the Meadow pipit, sensitivity was low and thus presences were slightly underpredicted while specificity was high. In contrast, for the Whinchat specificity was comparably low and presences were slightly overpredicted while sensitivity was high.

Table 4: Performance metrics of the Lithuania mean probability ensemble. The ensemble includes the algorithms: GLM, GAM, RF, and BRT. The threshold was used to binarize the continuous habitat suitability predictions and convert them into predicted presences and absences. It is the threshold that maximizes TSS in the cross-validated predictions.

Species	AUC	TSS	Sensitivity	Specificity	Explained Deviance	Threshold
Eurasian skylark (*Alauda arvensis*)	0.74	0.38	0.74	0.63	0.43	0.54
Meadow pipit (*Anthus pratensis*)	0.78	0.43	0.58	0.85	0.40	0.21
Yellowhammer (*Emberiza citrinella*)	0.71	0.37	0.78	0.59	0.43	0.47
Red-backed Shrike (*Lanius collurio*)	0.69	0.36	0.81	0.56	0.34	0.05
Tree sparrow (*Passer montanus*)	0.70	0.30	0.65	0.65	0.30	0.08
Whinchat (*Saxicola rubetra*)	0.77	0.45	0.82	0.63	0.47	0.37
Common starling (*Sturnus vulgaris*)	0.68	0.27	0.69	0.58	0.41	0.34
Northern lapwing (*Vanellus vanellus*)	0.75	0.39	0.66	0.73	0.40	0.17

Model predictions

For each species, we predicted the ensemble mean habitat suitability for entire Lithuania (Figure 10). Among the studied species, Skylark and Yellowhammer showed highest habitat suitability across Lithuania, followed by Whinchat and Starling. Meadow pipit and Lapwing only have few areas with highly suitable habitats in Western Lithuania while Red-backed shrike and Tree sparrow habitat suitability is low in most parts of the country. Judging from the committee averages of the binary predictions, the lowest amount of suitable area is predicted for the Meadow pipit and Northern lapwing. Yellowhammer, Skylark, Whinchat, Starling and Red-backed shrike have largely overlapping ranges (Figure 11). Standard deviations in predicted habitat suitability between different algorithms were overall low (Figure 12). Differences in habitat preferences are also visible in the species-specific response curves (Figure 13). On average, only the meadow pipit shows clear habitat suitability increases with increasing permanent grassland coverage.

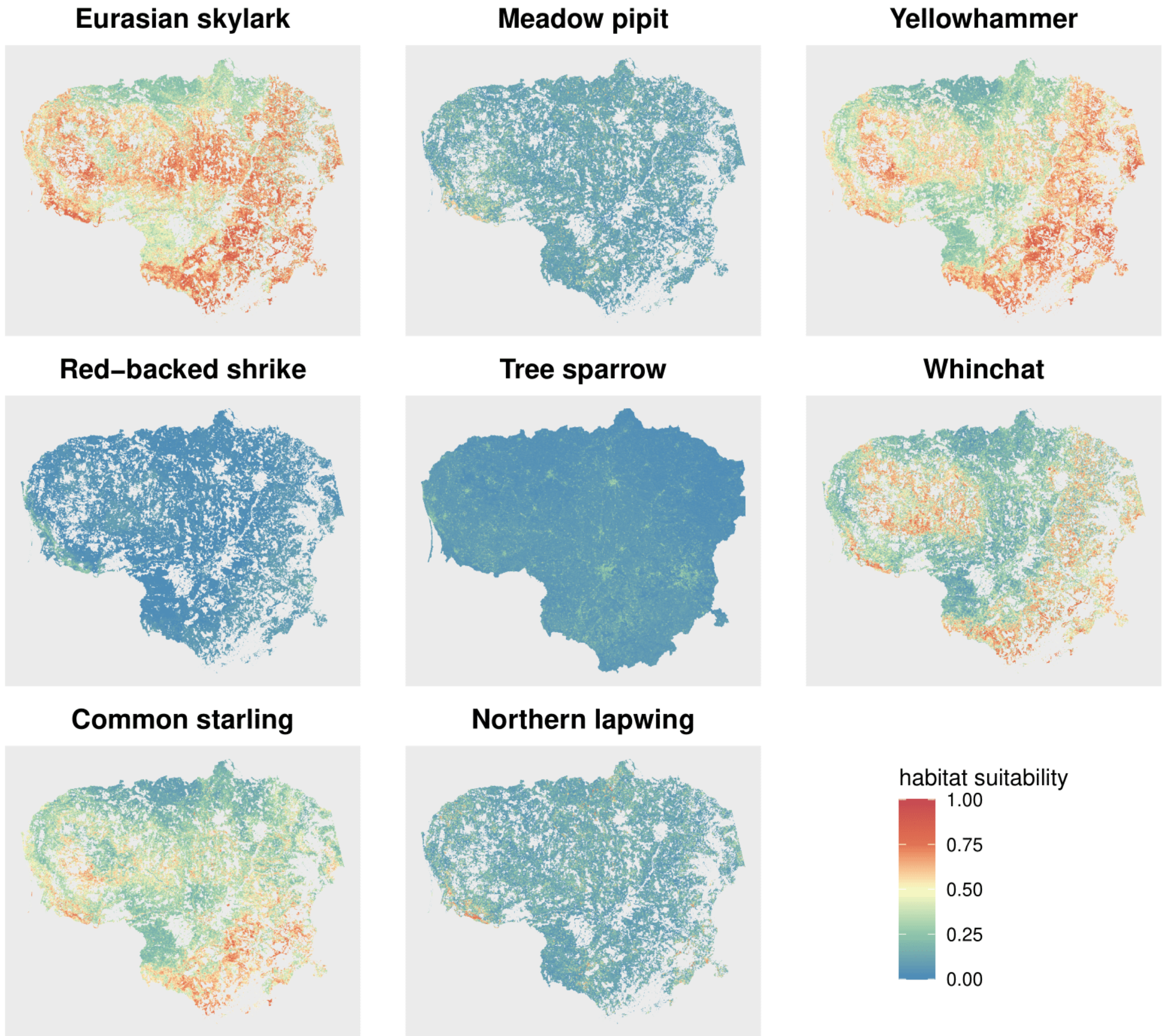


Figure 10: Habitat suitability predictions for Lithuania based on the mean probability ensemble per species. The ensembles include the model algorithms GLM, GAM, RF, and BRT.

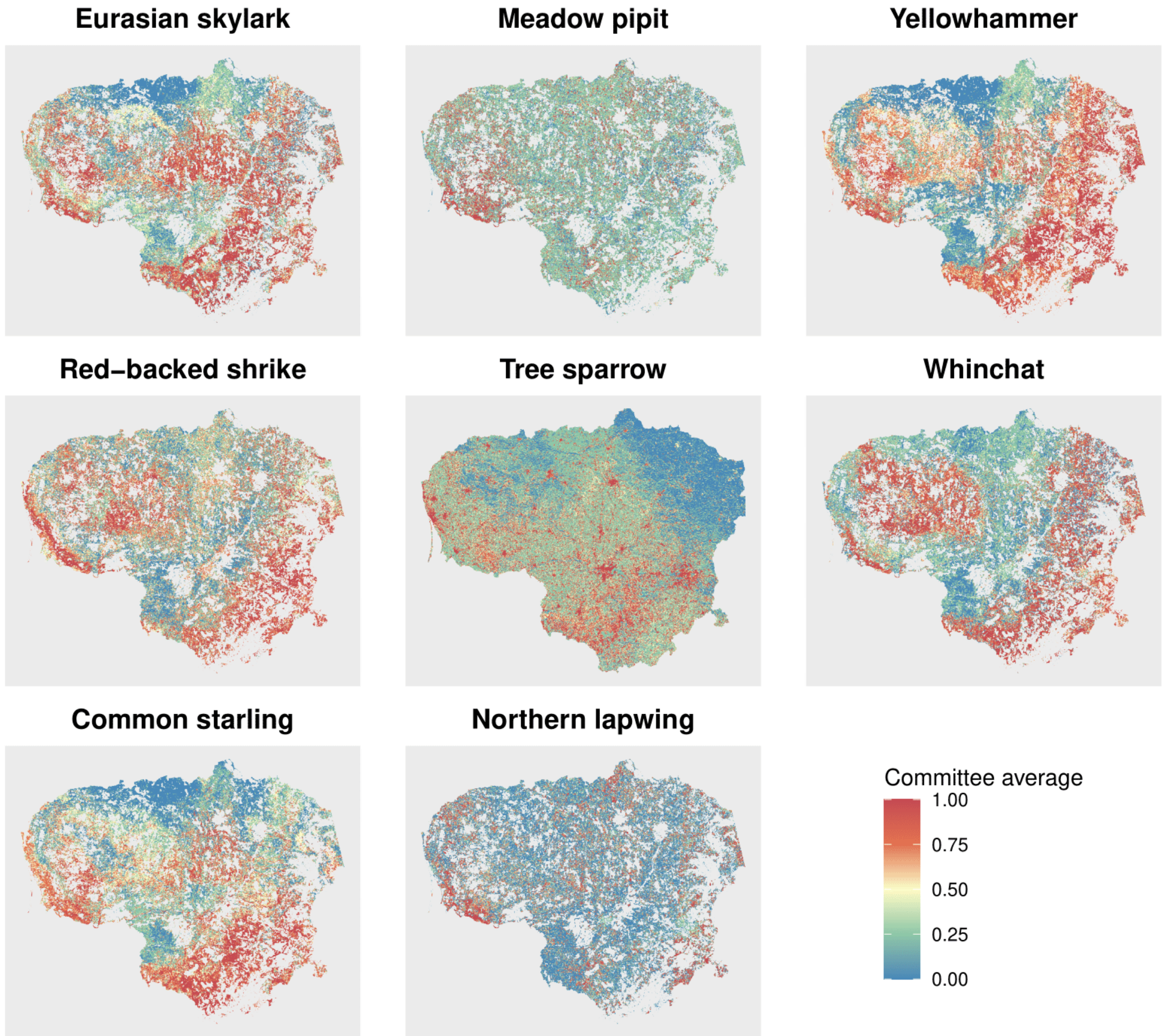


Figure 11: Committee averaged predictions for Lithuania per species. Committee averages indicate the proportion of algorithms agreeing on a species being present or absent. The model algorithms GLM, GAM, RF, and BRT were used in the ensembles.

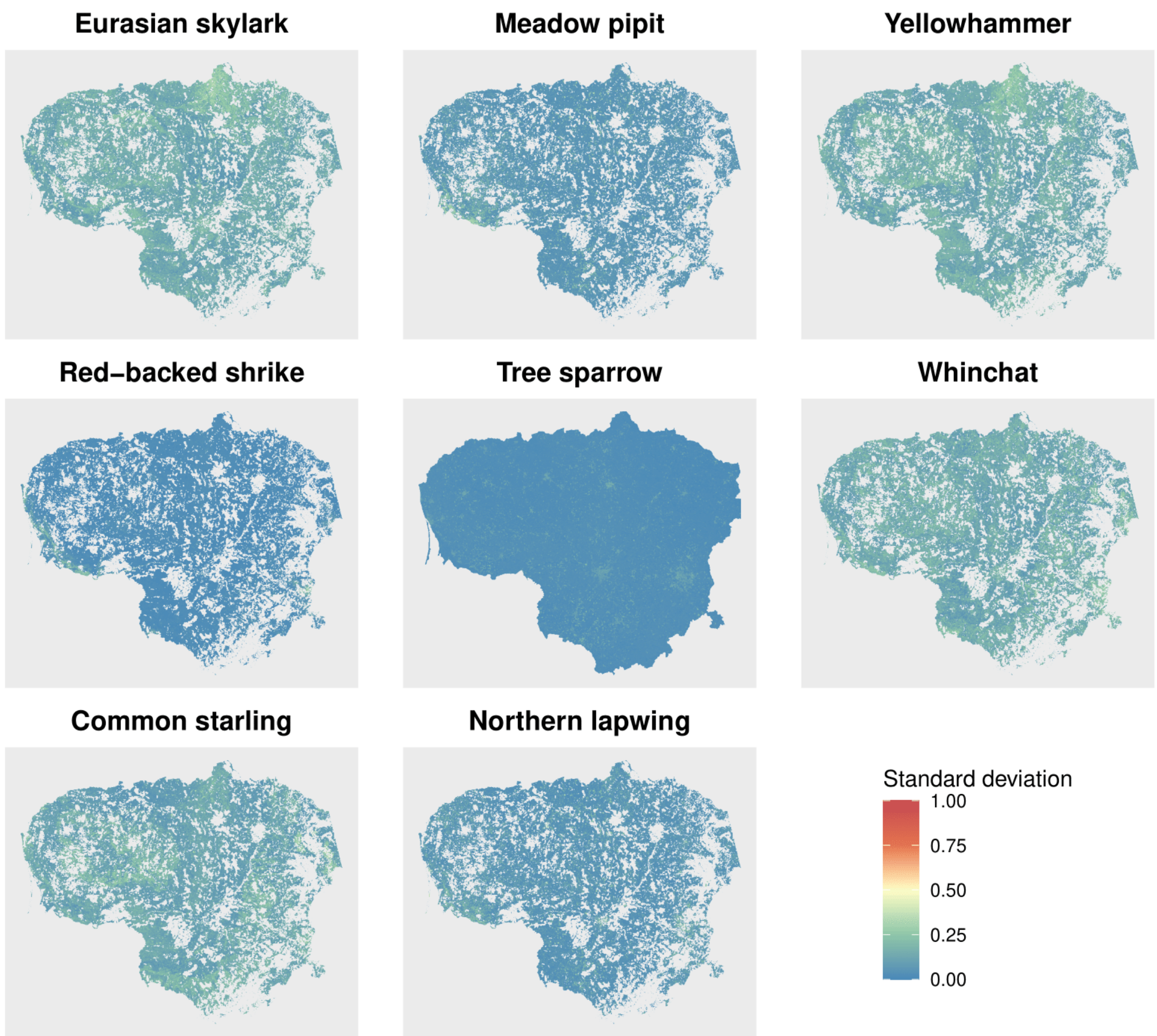


Figure 12: Standard deviation calculated from the habitat suitability predictions for Lithuania by the model algorithms (GLM, GAM, RF, and BRT).

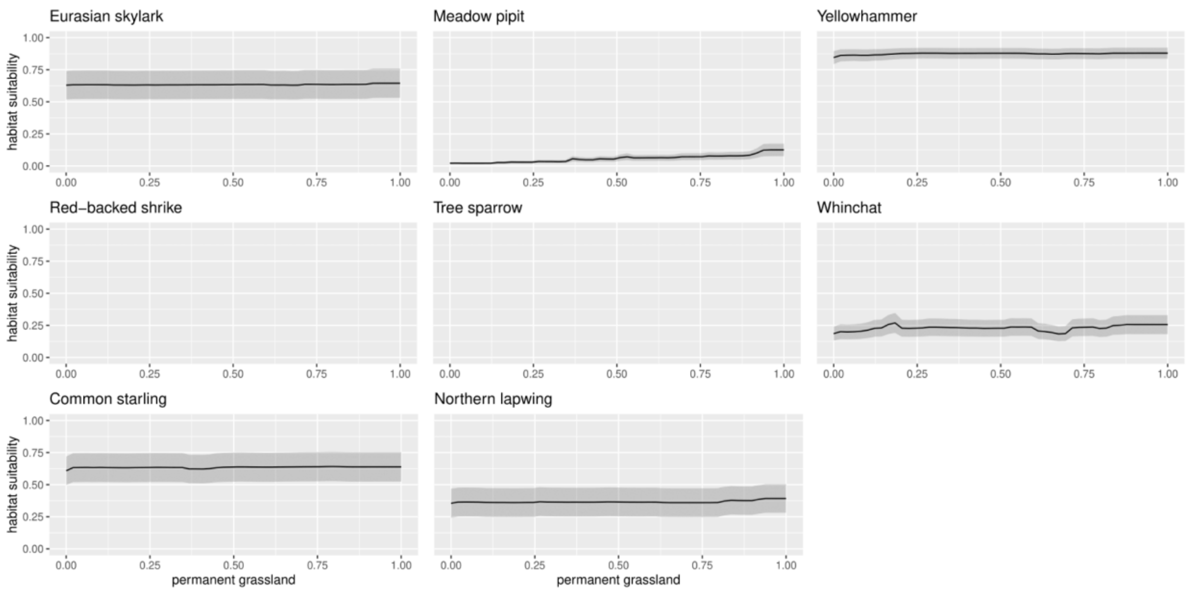


Figure 13: Partial response curves per species and Lithuania for the environmental parameter “permanent grassland” given in % cover. The solid line represents the mean ensemble predictions, the grey bands around the line plot represent the standard error across the SDM algorithm.

Mowing effects

For easier interpretation, we show the effects of mowing frequency and first mowing date on occurrence probability of all species for an average habitat suitability of 0.8 in grasslands (Figure 14). We see that mowing effects can be substantial for many species, such that occurrence probability of the species increases to almost 1.0 in grasslands that otherwise have an average habitat suitability of 0.8. The Yellowhammer, Eurasian skylark and the Common starling benefit most from lower mowing frequency, whereas the Whinchat and the Northern lapwing most from earlier mowing dates.

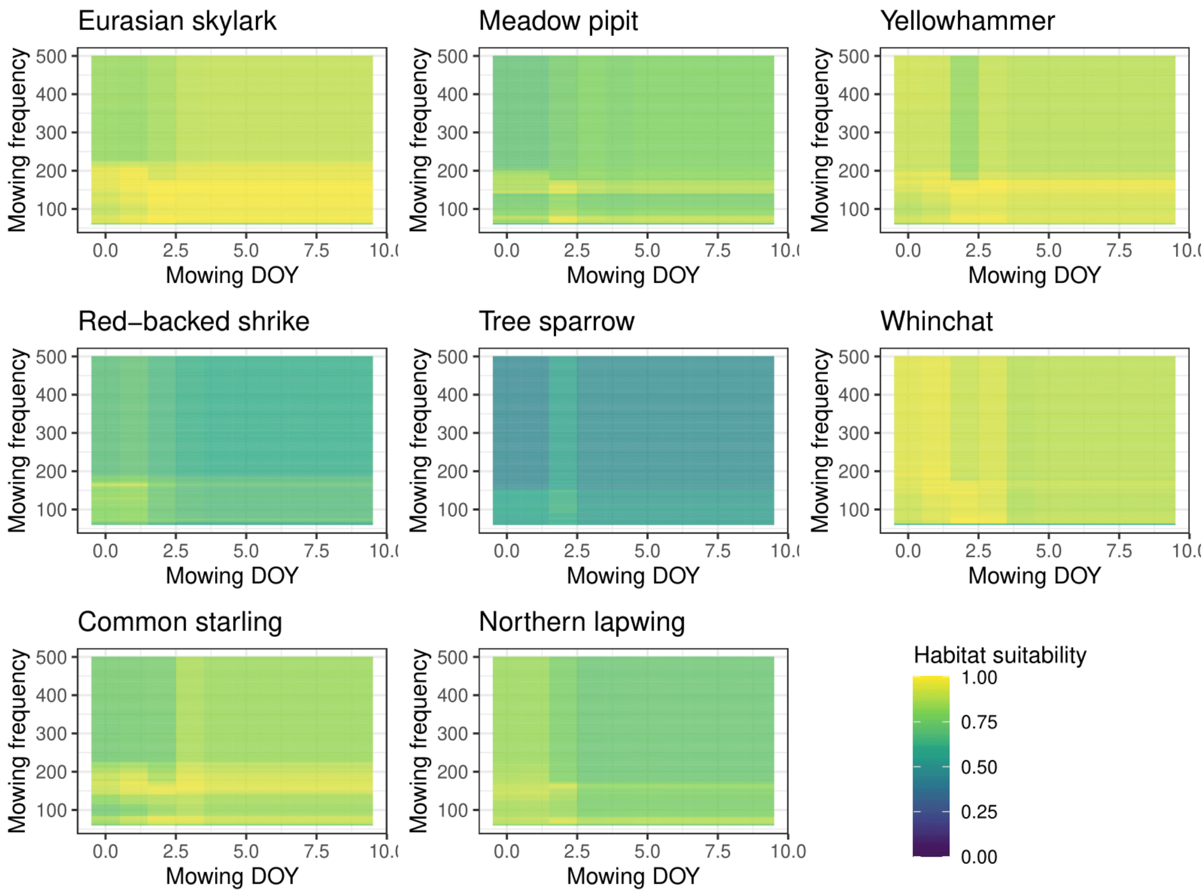


Figure 14: Mowing level plots indicating the effect of mowing frequency and first mowing date (DOY: day of year) on habitat suitability across species for Lithuania. The plots refer to a land-use suitability of 0.8.

3.4 South-Tyrol

Model performance

For South-Tyrol, the ensemble models also performed poorer than for Germany, with high variation between the species. The ensemble models ranged from poor to excellent predictive performance with AUC spanning from 0.55 - 0.88 (Table 5). The TSS, sensitivity and specificity scores ranged from 0.21 - 0.68, 0.39 - 0.89 and 0.51 - 0.85, respectively (Table 5). While both presences and absences were predicted fairly accurately for the Eurasian Skylark, sensitivity for the Yellowhammer was low indicating underprediction of presences. In contrast, for the red-backed Shrike, specificity was low and thus presences were overpredicted.

Table 5: Performance metrics of the mean probability ensemble for South-Tyrol. The ensemble includes the algorithms: GLM, GAM, RF, and BRT. The threshold was used to binarize the continuous habitat suitability predictions and convert them into predicted presences and absences. It is the threshold that maximizes TSS in the cross-validated predictions.

Species	AUC	TSS	Sensitivity	Specificity	Explained Deviance	Threshold
Eurasian skylark (*Alauda arvensis*)	0.88	0.68	0.86	0.81	0.48	0.23
Yellowhammer (*Emberiza citrinella*)	0.55	0.21	0.36	0.85	0.09	0.25
Red-backed Shrike (*Lanius collurio*)	0.75	0.40	0.89	0.51	0.42	0.17

Model predictions

For each species, we predicted the ensemble mean habitat suitability for South-Tyrol (Figure 15). For the Eurasian Skylark and the Red-backed shrike, suitable habitat is mostly predicted in the valleys while the Yellowhammer showed medium habitat suitability also in mid elevations across South-Tyrol. According to the committee average of the binary predictions, the lowest amount of suitable area is predicted for the Yellowhammer and the largest for the Red-backed Shrike (Figure 16). Standard deviations in predicted habitat suitability across model algorithms were overall low (Figure 17). Differences in habitat preferences are also visible in the species-specific response curves indicating important differences in habitat preferences (Figure 18).

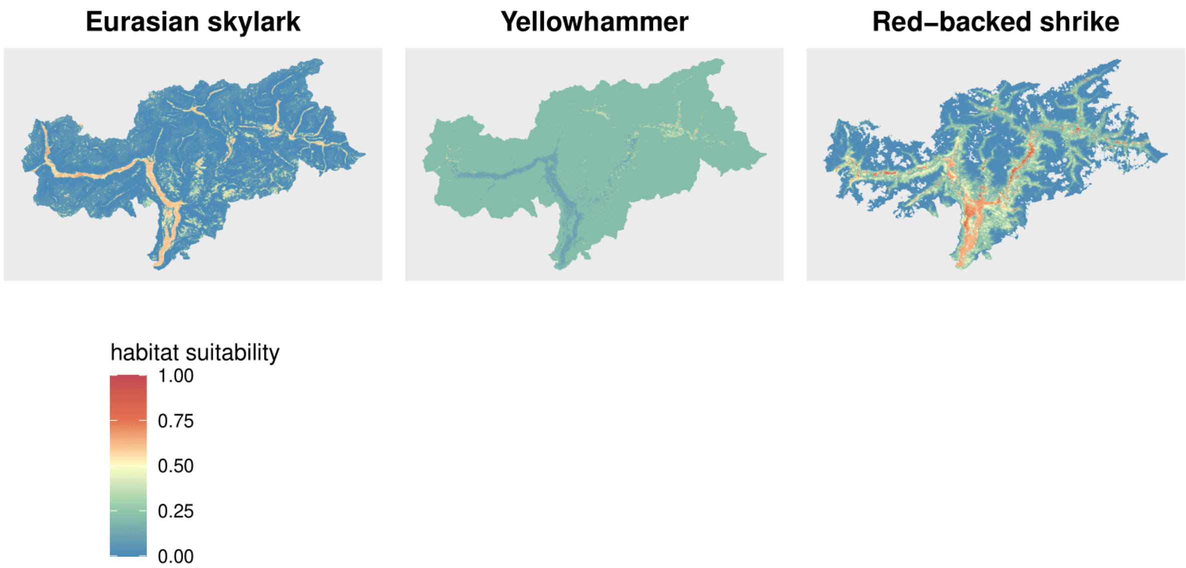


Figure 15: Habitat suitability predictions for South-Tyrol based on the mean probability ensemble per species. The ensembles include the model algorithms GLM, GAM, RF, and BRT.

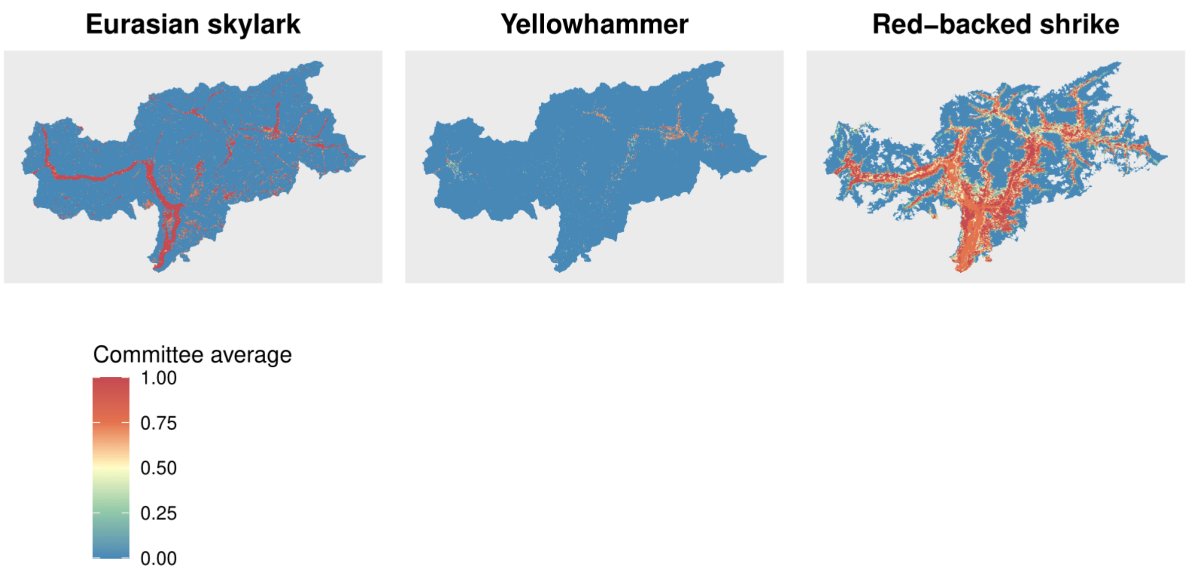


Figure 16: Committee averaged predictions for South-Tyrol per species. Committee averages indicate the proportion of algorithms agreeing on a species being present or absent. The model algorithms GLM, GAM, RF, and BRT were used in the ensembles.

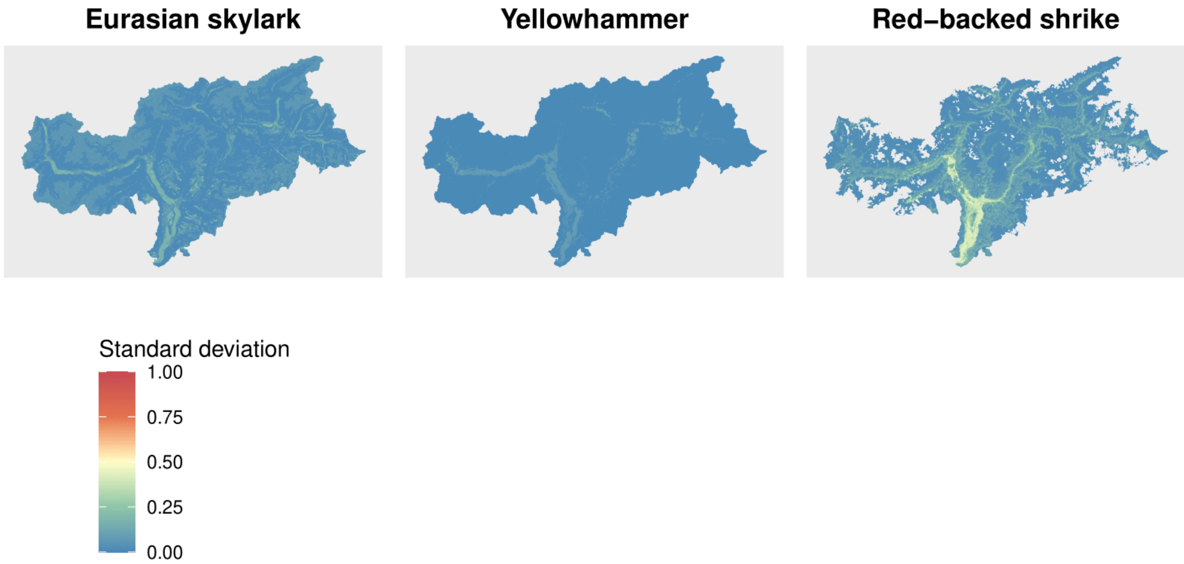
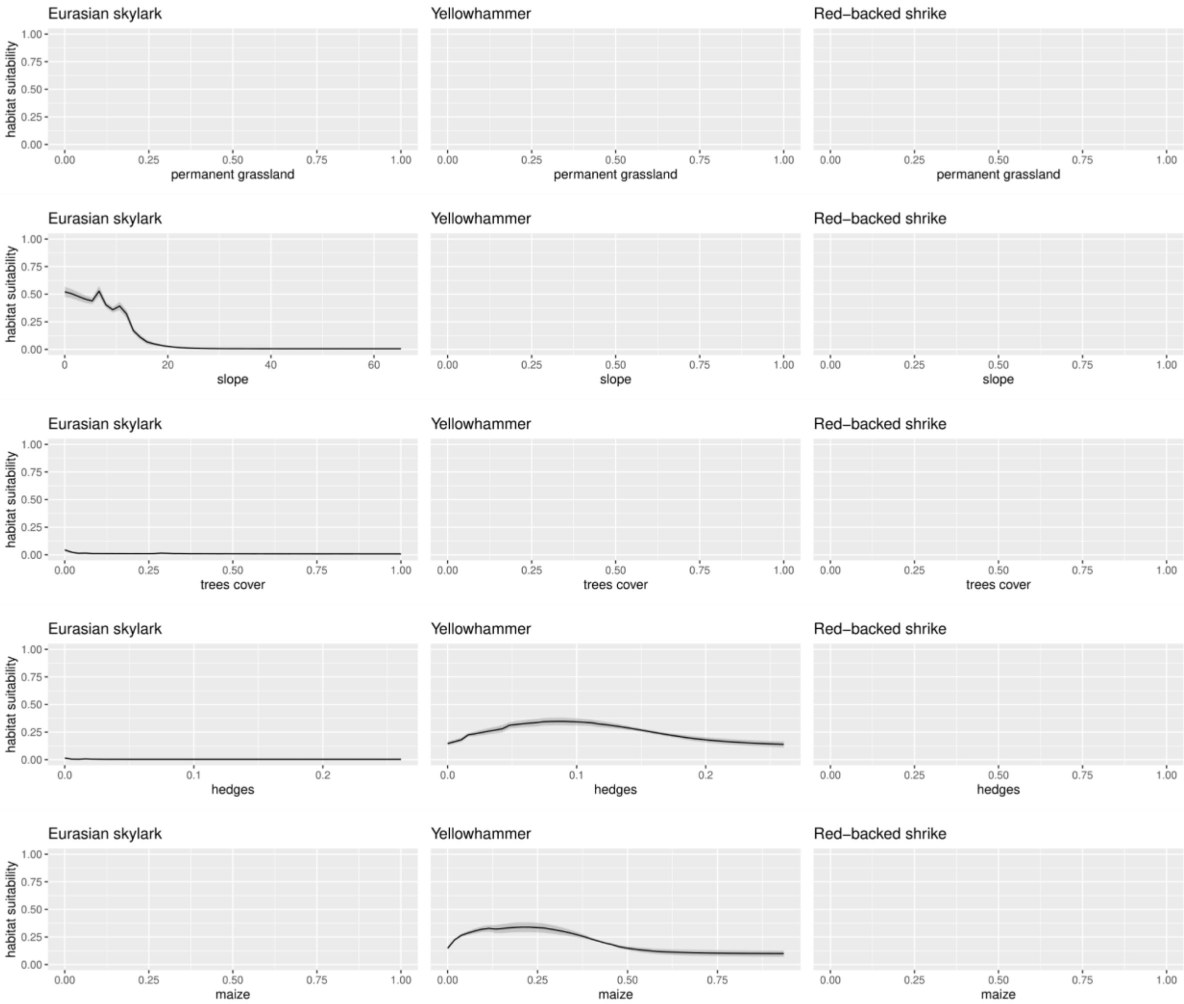


Figure 17: Standard deviation calculated from the habitat suitability predictions for South-Tyrol by the model algorithms (GLM, GAM, RF, and BRT).



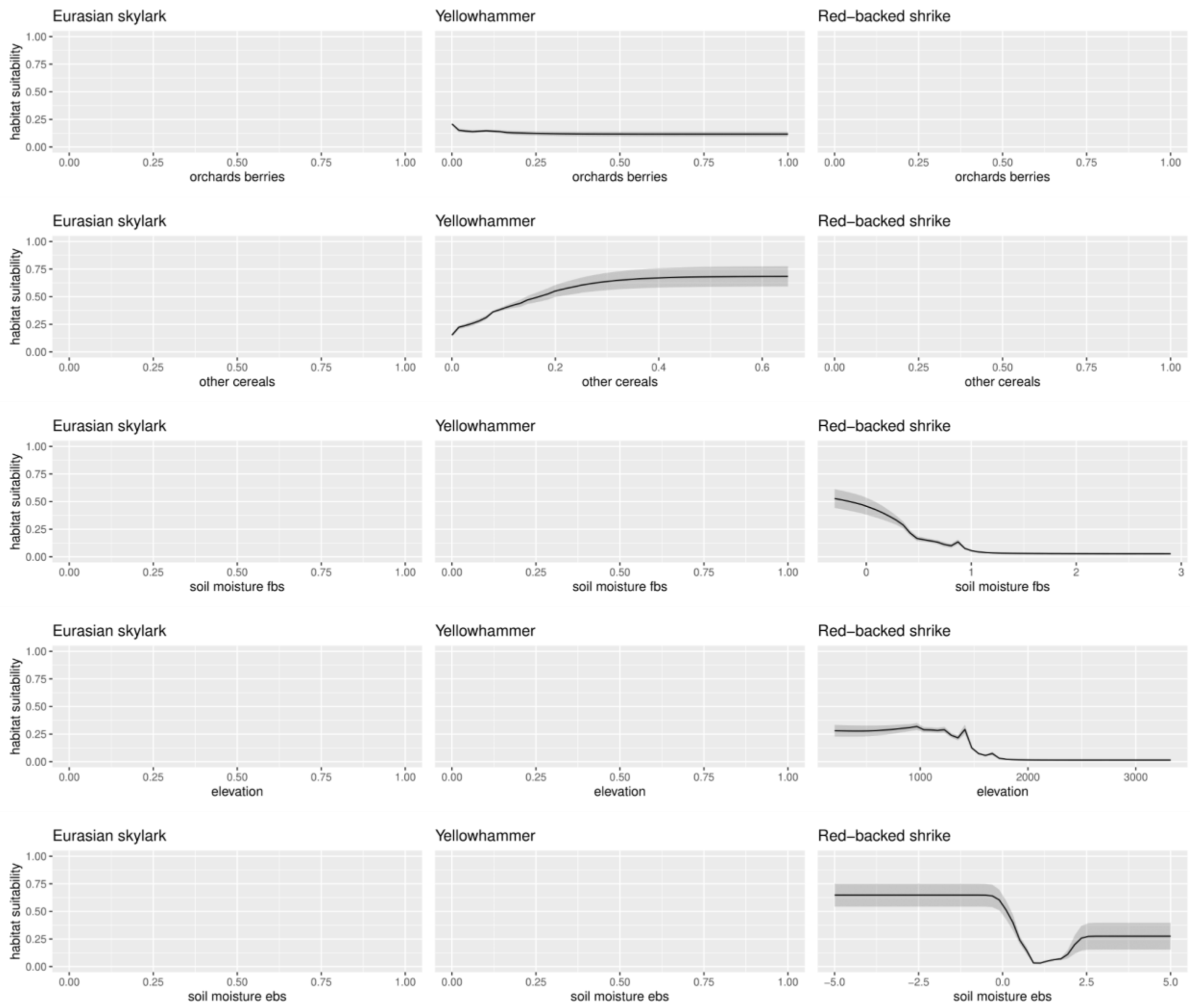


Figure 18: Partial response curves per species and South-Tyrol for all environmental parameters. The solid line represents the mean ensemble predictions, the gray bands around the line plot represent the standard error across the SDM algorithm.

Mowing effects

For easier interpretation, we show the effects of mowing frequency and first mowing date on occurrence probability of all species for an average habitat suitability of 0.8 in grasslands (Figure 19). We see that mowing effects can be substantial for many species, such that occurrence probability of the species increases to almost 1.0 or drops as low as 0.4 in grasslands that otherwise have an average occurrence probability of 0.8. The Yellowhammer benefits from earlier mowing dates. The Eurasian Skylark and the Red-backed Shrike benefit from slightly later mowing dates with low to medium mowing frequency.

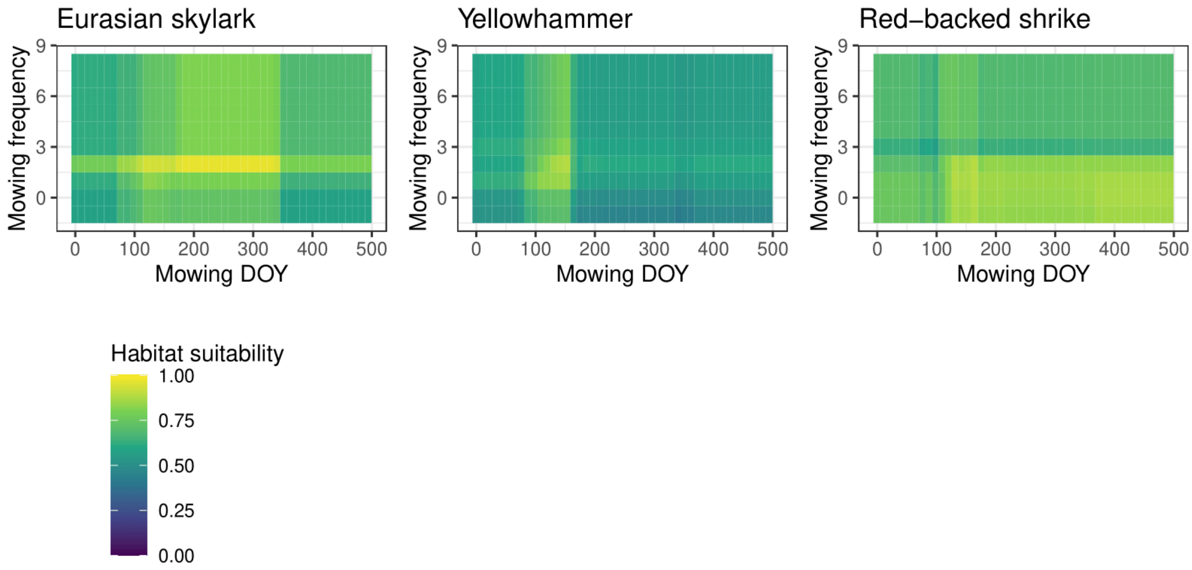


Figure 19: Mowing level plots indicating the effect of mowing frequency and first mowing date (DOY: day of year) on habitat suitability across species for South-Tyrol. The plots refer to a land-use suitability of 0.8.

Conclusion

SDMs were successfully applied to all study regions, although the number of species that could be successfully modelled differed between regions as did the model performances. The new approach of separately assessing the effect of mowing date and mowing frequency proved very insightful and provides more tangible information for the BirdWatch optimization algorithm of WP5000. Specifically, this allows the optimization algorithm to separate decisions about land use types from land use intensities.

References

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, *20*(1), 1–9. https://doi.org/10.1111/ddi.12144

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. (2023a). *ERA5 hourly data on single levels from 1940 to present* [Dataset]. Copernicus Climate Change Service (C3S). https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. (2023b). *ERA5 monthly averaged data on single levels from 1940 to present* [Dataset]. Copernicus Climate Change Service (C3S). https://doi.org/10.24381/cds.f17050d7

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth’s land surface areas. *Scientific Data*, *4*(1), 170122. https://doi.org/10.1038/sdata.2017.122

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2021). *Climatologies at high resolution for the earth´s land surface areas*. https://doi.org/10.16904/envidat.228.v2.1

Oeser, J., Zurell, D., Mayer, F., Çoraman, E., Toshkova, N., Deleva, S., Natradze, I., Benda, P., Ghazaryan, A., Irmak, S., Hasanov, N., Guliyeva, G., Gritsina, M., & Kuemmerle,

T. (2024). The Best of Two Worlds: Using Stacked Generalisation for Integrating Expert Range Maps in Species Distribution Models. *Global Ecology and Biogeography*, *33*(12), e13911. https://doi.org/10.1111/geb.13911

Sofaer, H. R., Jarnevich, C. S., & Flather, C. H. (2018). Misleading prioritizations from modelling range shifts under climate change. *Global Ecology and Biogeography*, *27*(6), 658–666. https://doi.org/10.1111/geb.12726

Supplementary material

ODMAP

Table S1: ODMAP protocol describing the main steps of model construction for all study regions.

ODMAP section	ODMAP subsection	ODMAP elements
Overview	Authorship	**Authors:** Levin Wiedenroth, Damaris Zurell, Emma Underwood, Pedro J. Leitão, Johanne Burmeister, Josh Haue **Contact email**: wiedenroth@uni-potsdam.de **Title:** Birdwatch nested SDM ensemble Germany, Lithuania, South-Tyrol
	Model objective	Mapping / interpolation, creating maps of relative probability of presence
	Taxon	European farmland birds: *Alauda arvensis, Anthus pratensis, Emberiza citrinella, Lanius collurio, Limosa limosa, Passer montanus, Saxicola rubetra, Streptopelia turtur, Sturnus vulgaris, Vanellus vanellus*
	Location	Germany; Lithuania; South-Tyrol; Europe (for quantifying broad-scale climate niche)
	Scale	**Spatial extent:** (minX, minY, maxX, maxY EPSG: 3035): Germany: 4031200, 4672600, 2684000, 3551600 Lithuania: 5007400, 5380000, 3507200, 3800200 South-Tyrol: 4350200, 4510800, 2568000, 2666800 Europe: 871891.5, 7671891, 894717.3, 6894717 **Spatial resolution:** 50 km / 200m **Temporal extent:** 2022 (Germany, Lithuania, South-Tyrol), 2012-2022 (Europe) **Temporal resolution:** 1 year (Germany, Lithuania, South-Tyrol), 6 years (Europe) **Boundary:** political

	Biodiversity data	**Observation type:** Presence-absence data
	Type of predictors	Climatic, topographic, land cover, land-use intensity
	Conceptual model	**Hypotheses about species-environment relationships:** We used a nested ensemble method that utilised European scale climate data and local land cover data including crop types, to predict habitat suitability of ten species of farmland birds across the regions of Germany, Lithuania and South-Tyrol.
	Assumptions	**Our assumptions were that:** 1. species are at equilibrium within their environment; 2. any biases in sampling and surveying are negligible; 3. the most relevant ecological drivers are included as proxies in the model variables.
	SDM algorithms	Model algorithms: Generalized linear model (GLM), Generalized Additive Model (GAM), Random Forest (RF), Boosted Regression Trees (BRT). **Justification of model complexity:** We chose to utilise an ensemble of different statistical (GLM, GAM) and supervised learning algorithms (RF, BRT) to get the best overall mean response. The algorithms were fitted with medium-high model complexity to ensure improved interpolation and mapping while keeping the fitted species-environment relationship reasonably general. **Model averaging/ensemble modelling used?**: Yes
	Model workflow	**Conceptual description of modelling steps including model fitting, assessment and prediction:** 01 - European climate data pre-preparations: reprojecting the raw climate layers before stacking and calculating month-wise means. 02 - SDM Europe - transforming the EBBA2 bird occurrence data to match the climatic data, extraction of climate data for each occurrence cell, spatial distance thinning (with distance of 100 km), 5-fold spatial blocking, collinearity testing (select07), SDM modelling, model validation, ensemble building, prediction of mean climatic suitability across Europe. 03 - Ensemble nested SDM: occurrence data preparation for Germany, Lithuania and South-Tyrol: filtering bird occurrences based on breeding season and considering

		only counts within 100 m from observer, removing duplicates (from within the same 200m cell), spatial distance thinning (with distance of 400m), 5-fold spatial blocking, collinearity testing (select07), nested SDM modelling using mean climatic suitability extracted from European SDM and local environmental and land cover variables, model validation, ensemble building, prediction of local habitat suitability.
	Software, codes and data	We used R version 4.4.1. With packages: terra, dplyr, sfheaders, corrplot, mecofun, randomForest, gbm, dismo, mgcv, sf, blockCV, ecospat **Availability of codes and data:** currently for internal circulation within project
Data	Biodiversity data	**Taxon names:** *Alauda arvensis, Anthus pratensis, Emberiza citrinella, Lanius collurio, Limosa limosa, Passer montanus, Saxicola rubetra, Streptopelia turtur, Sturnus vulgaris, Vanellus vanellus* **Biodiversity data source:** EBBA2 (Europe); Monitoring of Common Breeding Birds (MCB) collected by Dachverband Deutscher Avifaunisten (DDA) (Germany); Common Bird Population Abundance Monitoring project by Lithuanian Ornithological Society (Lithuania); point count data by Museum of Nature South Tyrol and local surveys (South-Tyrol) **Sampling design:** Europe: EBBA2 data are non-standardised monitoring data collected between 2013-2017 across entire Europe. EBBA2 provides presence-absence data at 50 km spatial resolution. Germany: 2,600, 1 x 1 km sample plots randomly stratified across Germany; simplified territory mapping in 4 surveys during breeding season along c. 3 km long transects within 1 x 1 km cells. Lithuania: point count surveys 2 times a year during breeding season, point count surveys along c. 10 km long routes consisting of 20 stops, roughly 500 m apart South-Tyrol: point count surveys 2-3 times per year during breeding season, point count sampling along transects with stops every 200 m and 10 min counts **Sample size per taxon**: After spatial distance thinning - Europe 156-633 presences (cf. Table 1 in Birdwatch deliverable D4.2); Germany 0-2290 presences (cf. Table 1 in this Birdwatch deliverable D4.3); Lithuania 3-719 presences (cf. Table 1 in this Birdwatch deliverable D4.3); South-Tyrol 0-109 presences (cf. Table 1 in this Birdwatch deliverable D4.3) **Details on scaling**: all regional occurrence data were

		scaled to the target spatial resolution of 200 m. Occurrences were spatially thinned to account for spatial autocorrelation. **Data cleaning/filtering steps:** Occurrences were filtered to only include the specific breeding period for each species and to include only observations within 100 m of the observer. For full details on this process, see D4.1 including accuracy of observation, validation by an expert, and type of behaviours observed. **Absence data:** non-observation of a species within the replicate surveys was considered an absence. **Potential errors and biases**: misidentification was accounted for by expert validation, georeferencing errors minimized based on scale of modelling, sampling bias should be reduced in standardized surveys and by our pre-processing steps including spatial distance thinning.
	Data partitioning	**Training data**: The final models were trained on all data. For several model building steps (collinearity checks, threshold optimization) and for validation, a 5-fold spatial block cross-validation was used. **Selection of validation data:** hexagonal spatial blocks were generated using the cv_spatial_block_autocor function of the blockCV package.
	Predictor variables	Climatic: 19 bioclimatic variables; Topographic: elevation, slope, solar radiation; Habitat (percentage cover): built-up cover, tree cover, water cover, soil moisture entire breeding season, soil moisture early breeding season; Crop types (percentage cover): cultivated grassland, orchards and berries, other cereals, permanent grassland, rapeseed, root crops, summer cereals, sunflower, vegetables, winter cereals, fallow, grapevine, hedges, hops, legumes, maize, and rapeseed. **Details on data sources:** Climate data: ERA5 hourly data from Copernicus Climate Change Service, https://doi.org/10.24381/cds.adbb2d47, Accessed on 01-02-2025 Topographic/Habitat/Crop type data: Birdwatch data products described in Birdwatch deliverable D3.4. **Spatial resolution of raw data**: ERA5 0.25°. For Birdwatch data products cf. Birdwatch deliverable D3.4. **Spatial extent of raw data**: ERA5 global. For Birdwatch data products cf. Birdwatch deliverable D3.4. **Map projection:** EPSG: 3035 or WGS 84 transformed to EPSG: 3035. **Temporal resolution of raw data:** ERA5 hourly **Details on data processing and on spatial, temporal and thematic scaling**: ERA5 data were upscaled to 50

		km spatial resolution to match the EBBA2 occurrence grid using bilinear interpolation; and monthly temporal resolution. Birdwatch data products were scaled to 200 m spatial resolution to match the average home range across all considered species, and 1 year temporal resolution (cf. Birdwatch deliverable D3.4).
Model	Multi-collinearity	We used the select07 method from Dormann et al. (2013) to check for and reduce multicollinearity. We calculated Spearman's rank correlation for each pair of variables and from pairs with |rho|>0.7 we removed the less important variable in terms of cross-validated univariate importance. The climate suitability predicted from European SDM was always included in the fine-scale SDMs.
	Model settings	GLM: linear and quadratic terms, AIC-based step-wise variable selection. binomial link. GAM: cubic smoothing splines with 4 degrees of freedom. binomial link. RF: nodesize= 5, number of trees = 1000. BRT: learning rate = model specific to end up between 1000 and 5000 trees, bag fraction = 0.75, tree complexity = 2. binomial link.
	Model estimates	**Assessment of coefficients:** not applicable. **Uncertainties:** not applicable. **Variable importance:** estimated for RF and BRT.
	Model selection / averaging / ensemble	**Nesting method**: we used the covariate method (Adde et al., 2023) **Ensemble method**: Ensemble predictions were derived by calculating the mean habitat suitability across the different SDM algorithms. For the fine-scale SDMs, we additionally computed the standard deviations of habitat suitability across algorithms, and the committee average (agreement of predicted presence across algorithms, using maxTSS threshold for binarizing prediction).
	Threshold selection	**Transforming continuous predictions into binary predictions:** We use the maxTSS approach to find the optimal threshold for binarizing continuous habitat suitability predictions and deriving predicted presences and absences. The threshold was optimized using cross-validated predictions.
Assessment	Performance statistics	**Performance statistics estimated on validation data:** model performance was assessed with a 5-fold spatial block cross-validation.The following model performance metrics were used: AUC (area under the receiver

		operating characteristic curve), TSS (true skill statistic), sensitivity, specificity
Prediction	Prediction output	**Prediction unit:** All models from all algorithms were combined into a single data frame, then thresholded for presence-absence predictions, and the mean of the probabilities was calculated.
	Uncertainty quantification	**Algorithmic uncertainty:** we assessed uncertainty in predictions by deriving the standard deviation across model predictions from the different SDM algorithms, and the committee average. Model ensembles with AUC < 0.7 were flagged for high uncertainty in predictions.