D4.4
Cross-region transferability assessed

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	D4.4 Cross-region transferability assessed*			
Work Package	WP4			
Project Month of Delivery	**Contractual**	M36	**Actual**	M36
Nature	Report	**Dissemination Level**		PUB
Lead Beneficiary	University of Potsdam (UP)			
Responsible Author	Levin Wiedenroth - UP			
Contributions from	Levin Wiedenroth - UP Damaris Zurell - UP Johanne Burmeister - UP			

* Please note that this deliverable was originally suggested as “D4.4 Cross-region transferability assessed and predictions of habitat suitability for entire Europe”. The EO-DateCube from WP3000 provided environmental data only for the four case study regions, thus prohibiting habitat suitability predictions for entire Europe.

History of Changes

Version	Issue Date	Stage	Description	Comments	Contributor
1.0	28.01.2026	Draft	Initial version of D4.4		Levin Wiedenroth Damaris Zurell Johanne Burmeister

Table of Contents

1. Introduction

Here we present the fourth and final deliverable of work package 4000, which uses the models presented in deliverables D4.2 and D4.3 and tests cross-region transferability of these models. This step does not feed into the BirdWatch optimization algorithm in WP5000 but rather assesses robustness of the habitat suitability predictions across regions with important implications for future upscaling of the Birdwatch tool to other regions. The data and methods have already been described in previous deliverables D4.1-4.3, and we here focus on predictive accuracy of the models when transferred to another region.

2. Data and Methods

As detailed in deliverables D4.1 - D4.3, the Birdwatch work package WP4000 uses species distribution models (SDMs) to relate species occurrence data to prevailing environmental conditions in order to predict bird habitat suitability (Guisan et al., 2017). To account for scale dependent predictor relevance, a nested SDM framework was implemented (cf. deliverable D4.1) that combines European-wide, coarse-grained SDMs of climate suitability at 50 km spatial resolution with fine-grained SDMs of habitat suitability at 200 m spatial resolution. This final 200 m spatial resolution represents the scale at which decisions can be supported regarding land use practices and intensity, while the large-scale climate suitability represents boundary conditions. For the present deliverable, we are mostly interested in how far SDMs trained on land use data in one region can reliably inform about habitat suitability in other regions. We thus focus only on the fine-scale models. Occurrence data for Germany, Lithuania, and South-Tyrol stem from standardized bird surveys (cf. deliverable D4.3). For Flanders, opportunistic observations were combined with pseudo-absences (cf. deliverable D4.2). Prior to modelling, occurrence data were spatially thinned (cf. deliverables D4.2 - 4.3). 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) and building ensembles using the mean habitat suitability across SDM algorithms. Within-region predictive performance was assessed using a five-fold spatial block cross-validation and four different performance metrics (AUC, TSS, sensitivity, specificity), whereby the threshold-dependent performance metrics (TSS, sensitivity, specificity) used the maximumTSS threshold to binarize predictions (cf. deliverables D4.1-4.3). We used the same performance metrics for quantifying cross-region transferability, quantifying for each training region how well the models predicted to the independent test regions. As not all species occurred in all regions, cross-region transferability could only be assessed for those farmland bird species that occurred in each training region. In a few cases, species occurred in a region but in such low numbers that no SDM could be fitted for that training region, e.g. the Red-back shrike in Flanders. In those cases, the occurrences of that species could still be used for testing cross-region transferability of models trained in other regions.

3. Results

The predictive performances within and across regions were very heterogeneous. While within-region predictive accuracy in terms of AUC yielded fair to excellent predictions for all species and regions (except for Red-backed shrike and Common starling in Lithuania; cf. deliverable D4.3), many models yielded poor predictions when transferred across regions (Tables 1-9). Highest transferability was achieved by models trained in Germany (AUC ranging 0.52 - 0.95, mean 0.73, median 0.73) followed by models trained in Flanders (AUC ranging 0.50 - 0.92, mean 0.71, median 0.72). The mean cross-prediction AUC achieved by models trained in Lithuania and South Tyrol was poor (Lithuania: AUC ranging 0.50 - 0.90, mean, 0.66, median 0.67; South Tyrol AUC ranging 0.52 - 0.88, mean 0.62, median 0.56). According to AUC, models trained in Germany yielded fair predictions (or better) in 43% of the transfers (species x region), models trained in Flanders in 35% of transfers, models trained in Lithuania in 32%, and models trained in South Tyrol in 11% of transfers.

When assessing model predictive performance in terms of TSS, which summarizes the model's ability to distinguish presences from absences and correct predictions from misclassifications, transferability was even lower (Tables 1-9). Again highest accuracy was achieved by models trained in Germany (TSS ranging 0.02 - 0.90, mean 0.40, median 0.41) followed by models trained in Flanders (TSS ranging -0.02 - 0.7, mean 0.35, median 0.35) while mean cross-region TSS were considerably lower for models trained in the other two regions (Lithuania: TSS ranging 0.00 - 0.90, mean 0.26, median 0.28; South Tyrol: TSS ranging 0.02 - 0.69, mean 0.21, median 0.14). According to TSS, models trained in Germany yielded fair predictions (or better) in 22% of the transfers, models trained in Flanders in 18% of transfers, models trained in Lithuania in 9%, and models trained in South Tyrol in zero transfers.

In very few cases, models showed higher cross-region predictive performance than within-region predictive performance. For example, the models trained in Germany for Red-backed shrike and Whinchat achieved higher predictive performance in Flanders than in Germany. Similarly, the models trained in Lithuania for the Tree sparrow and the Whinchat achieved higher predictive performance in Germany and Flanders, respectively.

4. Conclusion

Overall, cross-region transferability of the fine-scale models was poor for most species, potentially inhibiting a European-wide tool, based on only a few regions, for optimizing land use practices under consideration of habitat suitability for different farmland bird species. The likely reasons are region-specific farming practices and region-specific habitat preferences of the selected species. In general, models trained in regions with long environmental gradients such as Germany transferred better to other regions, although not for all species and regions. For future use of the Birdwatch optimization algorithm, we thus recommend careful selection of training regions to reduce bias through regional peculiarities, both in terms of farming practices as well as bird behaviour.

Table 1: Predictive performances of SDMs trained for Eurasian skylark (*Alauda arvensis*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

		Predicted region			
		FI	Ger	Lith	ST
Training region	FI	**0.90**	**0.89**	0.59	**0.70**
	Ger	**0.82**	**0.95**	0.53	**0.82**
	Lith	0.59	**0.71**	**0.74**	**0.83**
	ST	0.65	**0.77**	0.53	**0.88**

AUC

	Predicted region			
	FI	Ger	Lith	ST
FI	**0.64**	**0.68**	0.15	0.34
Ger	0.49	**0.78**	0.14	**0.58**
Lith	0.17	0.34	0.38	**0.55**
ST	0.26	0.46	0.09	**0.69**

TSS

		FI	Ger	Lith	ST
Training region	FI	**0.89**	**0.90**	**0.85**	0.57
	Ger	**0.77**	**0.89**	**0.88**	**0.86**
	Lith	**0.78**	**0.76**	0.74	**0.95**
	ST	0.49	**0.8**	**0.91**	**0.86**

Sensitivity

	FI	Ger	Lith	ST
FI	**0.76**	**0.78**	0.30	**0.77**
Ger	0.73	**0.89**	0.26	0.72
Lith	0.40	0.58	0.63	0.60
ST	**0.78**	0.66	0.18	**0.83**

Specificity

Table 2. Predictive performances of SDMs trained for Meadow pipit (*Anthus pratensis*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

Predicted region

Training region

	FI	Ger	Lith	ST
FI	**0.90**	**0.82**	**0.71**	-
Ger	**0.71**	**0.87**	**0.74**	-
Lith	0.66	**0.73**	**0.78**	-
ST	-	-	-	-

AUC

Predicted region

	FI	Ger	Lith	ST
FI	**0.62**	0.49	0.37	-
Ger	0.34	**0.60**	0.34	-
Lith	0.29	0.43	0.43	-
ST	-	-	-	-

TSS

Training region

	FI	Ger	Lith	ST
FI	**0.84**	0.64	0.58	-
Ger	0.67	**0.86**	0.48	-
Lith	0.74	0.66	0.58	-
ST	-	-	-	-

Sensitivity

	FI	Ger	Lith	ST
FI	0.77	**0.84**	**0.79**	-
Ger	0.67	0.74	**0.86**	-
Lith	0.54	**0.77**	**0.85**	-
ST	-	-	-	-

Specificity

Table 3. Predictive performances of SDMs trained for Yellowhammer (*Emberiza citrinella*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

Predicted region

Training region

	Fl	Ger	Lith	ST
Fl	**0.92**	**0.82**	0.59	0.50
Ger	**0.84**	**0.90**	0.53	0.52
Lith	0.56	0.55	**0.71**	0.58
ST	0.57	0.54	0.52	0.55

AUC

Predicted region

	Fl	Ger	Lith	ST
Fl	**0.70**	**0.55**	0.23	0.04
Ger	**0.54**	**0.66**	0.02	0.11
Lith	0.01	0.00	0.37	0.05
ST	0.15	0.14	0.04	0.21

TSS

Training region

	Fl	Ger	Lith	ST
Fl	**0.95**	**0.77**	0.51	0.05
Ger	**0.83**	**0.83**	**0.95**	0.29
Lith	**0.94**	0.00	**0.78**	1.00
ST	**0.76**	0.31	0.06	0.36

Sensitivity

	Fl	Ger	Lith	ST
Fl	**0.75**	**0.79**	0.72	**0.99**
Ger	0.71	**0.82**	0.07	**0.83**
Lith	0.07	**1.00**	0.59	0.05
ST	0.38	**0.82**	**0.98**	**0.85**

Specificity

Table 4. Predictive performances of SDMs trained for Red-backed shrike (*Lanius collurio*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

Predicted region

Training region

	FI	Ger	Lith	ST
FI	-	-	-	-
Ger	**0.84**	**0.82**	0.56	0.62
Lith	0.50	0.57	0.69	0.59
ST	0.52	0.58	0.55	**0.75**

AUC

Predicted region

	FI	Ger	Lith	ST
FI	-	-	-	-
Ger	**0.70**	0.49	0.10	0.23
Lith	0.00	0.00	0.36	0.16
ST	0.05	0.06	0.02	0.40

TSS

Training region

	FI	Ger	Lith	ST
FI	-	-	-	-
Ger	**1.00**	**0.85**	**0.97**	0.51
Lith	0.00	0.00	**0.81**	0.46
ST	**1.00**	**0.99**	0.05	**0.89**

Sensitivity

	FI	Ger	Lith	ST
FI	-	-	-	-
Ger	0.70	0.64	0.14	0.72
Lith	**1.00**	**1.00**	0.56	0.70
ST	0.05	0.07	**0.97**	0.51

Specificity

Table 5. Predictive performances of SDMs trained for Tree sparrow (*Passer montanus*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

Predicted region

Training region

	FI	Ger	Lith	ST
FI	**0.72**	0.55	0.56	0.51
Ger	0.55	**0.84**	0.64	**0.76**
Lith	0.54	**0.74**	**0.70**	**0.76**
ST	-	-	-	-

AUC

Predicted region

	FI	Ger	Lith	ST
FI	0.42	-0.02	0.01	0.17
Ger	0.15	**0.53**	0.26	**0.56**
Lith	0.02	0.37	0.30	0.48
ST	-	-	-	-

TSS

Training region

	FI	Ger	Lith	ST
FI	**0.86**	0.07	**0.81**	0.33
Ger	**0.95**	**0.86**	0.61	**0.92**
Lith	**0.98**	**0.83**	0.65	**0.75**
ST	-	-	-	-

Sensitivity

	FI	Ger	Lith	ST
FI	0.56	**0.91**	0.21	**0.84**
Ger	0.20	0.67	0.65	0.64
Lith	0.04	0.54	0.65	0.73
ST	-	-	-	-

Specificity

Table 6. Predictive performances of SDMs trained for Whinchat (*Saxicola rubetra*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

Predicted region

Training region

	Fl	Ger	Lith	ST
Fl	-	-	-	-
Ger	**0.90**	**0.87**	0.52	0.59
Lith	**0.90**	0.53	**0.77**	0.60
ST	-	-	-	-

AUC

Predicted region

	Fl	Ger	Lith	ST
Fl	-	-	-	-
Ger	**0.90**	**0.57**	0.06	0.13
Lith	**0.90**	0.07	0.45	0.20
ST	-	-	-	-

TSS

Training region

	Fl	Ger	Lith	ST
Fl	-	-	-	-
Ger	**1.00**	**0.75**	**0.83**	0.58
Lith	**1.00**	**0.92**	**0.82**	**0.83**
ST	-	-	-	-

Sensitivity

	Fl	Ger	Lith	ST
Fl	-	-	-	-
Ger	**0.90**	**0.82**	0.23	0.55
Lith	**0.90**	0.15	0.63	0.37
ST	-	-	-	-

Specificity

Table 7. Predictive performances of SDMs trained for European turtle dove (*Streptopelia turtur*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

Predicted region

Training region

	FI	Ger	Lith	ST
FI	**0.85**	0.53	-	-
Ger	0.64	**0.81**	-	-
Lith	-	-	-	-
ST	-	-	-	-

AUC

Predicted region

	FI	Ger	Lith	ST
FI	**0.54**	0.03	-	-
Ger	0.28	0.49	-	-
Lith	-	-	-	-
ST	-	-	-	-

TSS

Training region

	FI	Ger	Lith	ST
FI	0.68	**0.97**	-	-
Ger	0.74	**0.77**	-	-
Lith	-	-	-	-
ST	-	-	-	-

Sensitivity

	FI	Ger	Lith	ST
FI	**0.86**	0.07	-	-
Ger	0.55	0.71	-	-
Lith	-	-	-	-
ST	-	-	-	-

Specificity

Table 8. Predictive performances of SDMs trained for Common starling (*Sturnus vulgaris*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

Predicted region

Training region

	FI	Ger	Lith	ST
FI	**0.80**	0.66	0.53	0.57
Ger	0.61	**0.86**	0.57	**0.70**
Lith	0.63	0.54	0.68	0.52
ST	-	-	-	-

AUC

Predicted region

	FI	Ger	Lith	ST
FI	0.44	0.28	0.09	0.10
Ger	0.19	**0.58**	0.14	0.49
Lith	0.01	0.16	0.27	0.06
ST	-	-	-	-

TSS

Training region

	FI	Ger	Lith	ST
FI	0.73	0.70	0.56	**1.00**
Ger	0.49	**0.78**	0.70	**1.00**
Lith	**1.00**	0.34	0.69	0.60
ST	-	-	-	-

Sensitivity

	FI	Ger	Lith	ST
FI	0.71	0.58	0.53	0.10
Ger	0.70	**0.80**	0.44	0.49
Lith	0.01	**0.82**	0.58	0.46
ST	-	-	-	-

Specificity

Table 9. Predictive performances of SDMs trained for Northern lapwing (*Vanellus vanellus*). Gray shading indicates within-region cross-validation, while all other metrics refer to cross-region transferability. Bold numbers indicate model predictions that reached fair predictive accuracy or better.

Predicted region

Training region

	FI	Ger	Lith	ST
FI	**0.84**	**0.92**	**0.73**	-
Ger	0.69	**0.93**	**0.72**	-
Lith	0.66	**0.74**	**0.75**	-
ST	-	-	-	-

AUC

Predicted region

	FI	Ger	Lith	ST
FI	0.51	**0.71**	0.33	-
Ger	0.27	**0.72**	0.34	-
Lith	0.26	0.35	0.39	-
ST	-	-	-	-

TSS

Training region

	FI	Ger	Lith	ST
FI	**0.80**	**0.88**	0.48	-
Ger	0.67	**0.88**	0.61	-
Lith	**0.83**	0.50	0.66	-
ST	-	-	-	-

Sensitivity

	FI	Ger	Lith	ST
FI	0.71	**0.83**	**0.84**	-
Ger	0.60	**0.84**	0.73	-
Lith	0.43	**0.86**	0.73	-
ST	-	-	-	-

Specificity

5. References

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