

Supplemental Information for:

Species-observer link and kernel density estimation of background points allow for sampling bias correction in bird species distribution models

Table of Contents:

1		
2		
3		
4		
5		
6		
7		
8	Table of Contents:	
9	Introduction	3
10	Species data	4
11	Presence-absence dataset (PAD)	7
12	Presence-only dataset (POD)	7
13	Spatial autocorrelation in environmental data	9
14	Predictors	10
15	Background points simulation	11
16	Results (part A).....	12
17	Results (part B).....	14
18	GLM.....	14
19	AUC.....	14
20	COR.....	19
21	Jaccard.....	23
22	TSS.....	27
23	Maxnet	31
24	AUC.....	31
25	COR.....	35
26	Jaccard.....	39
27	TSS.....	43
28	Random forest	47
29	AUC.....	47
30	COR.....	52
31	Jaccard.....	56
32	TSS.....	60

33	Results (part C).....	64
34		

35 Introduction

36 The scripts and anonymised data used in this study are available for download at:

37

- 38 • Data and scripts for replicating study analyses

39 <[- 44 • Target-Group Background Generator '*tbg*' package:](https://zenodo.org/records/15497910?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImZmMjU1ZjNjLWU2M2EtNDE4NC1hYzgzLWQ1NDMwYzdiMDAxZCIsImRhdGEiOnt9LCJyYW5kb20iOiI2NGNhMzc2NzE4ZWE5MDQ4ZWQzNTM4ZDQ3NmM4ZGlwNyJ9.hQD2KLbfUSY_hHQsfewuxizWI6nniMnct3pwk0L2QBXjopQckDJRVnK6jca148uyaMk47I8p4zvE8-pUNYcJtQ></p>
</div>
<div data-bbox=)

45 <[- 50 • The '*sdm*' package \(based on version 1.2-56; Naimi and Araújo, 2016\) was modified to enable sampling with replacement and to add Jaccard and Sorensen indices:](https://zenodo.org/records/15497383?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImZmMjU1ZjNjLWU2M2EtNDE4NC1hYzgzLWQ1NDMwYzdiMDAxZCIsImRhdGEiOnt9LCJyYW5kb20iOiI1ODU0ZmZjYjI2MDZlOTQzZTU0OTAyZjE4ODQzN2FhZTgzMCI9.1z0260yK5XkRQmDrbO1NyJ06fVVpDK7jfmZe-5RcZSZaoKCCoJpEpINf-r9TtTfS601Mm2f-YBWDh1c20F5usg></p>
</div>
<div data-bbox=)

51 <[57](https://zenodo.org/records/15497164?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImZmMjU1ZjNjLWU2M2EtNDE4NC1hYzgzLWQ1NDMwYzdiMDAxZCIsImRhdGEiOnt9LCJyYW5kb20iOiI1ODE4NDMwODY4YzExOTI0YmVmOTkyYTQwNTQyM2M5MyJ9.Rhx9kkm3gmkSvAoG7igwslNdFnIzes7qSwGyegAghHW2qTy_r8IJYoiRGDnQ89IDdTLpkf7GXA26r5cVIGLYw></p>
</div>
<div data-bbox=)

58 All of the above-referenced data are originally maintained as separate GIT repositories, each
59 accompanied by a description in a *readme.md* file and, in the case of the packages, also by the
60 corresponding generated documentation (using '*roxygen2*').

61 Species data

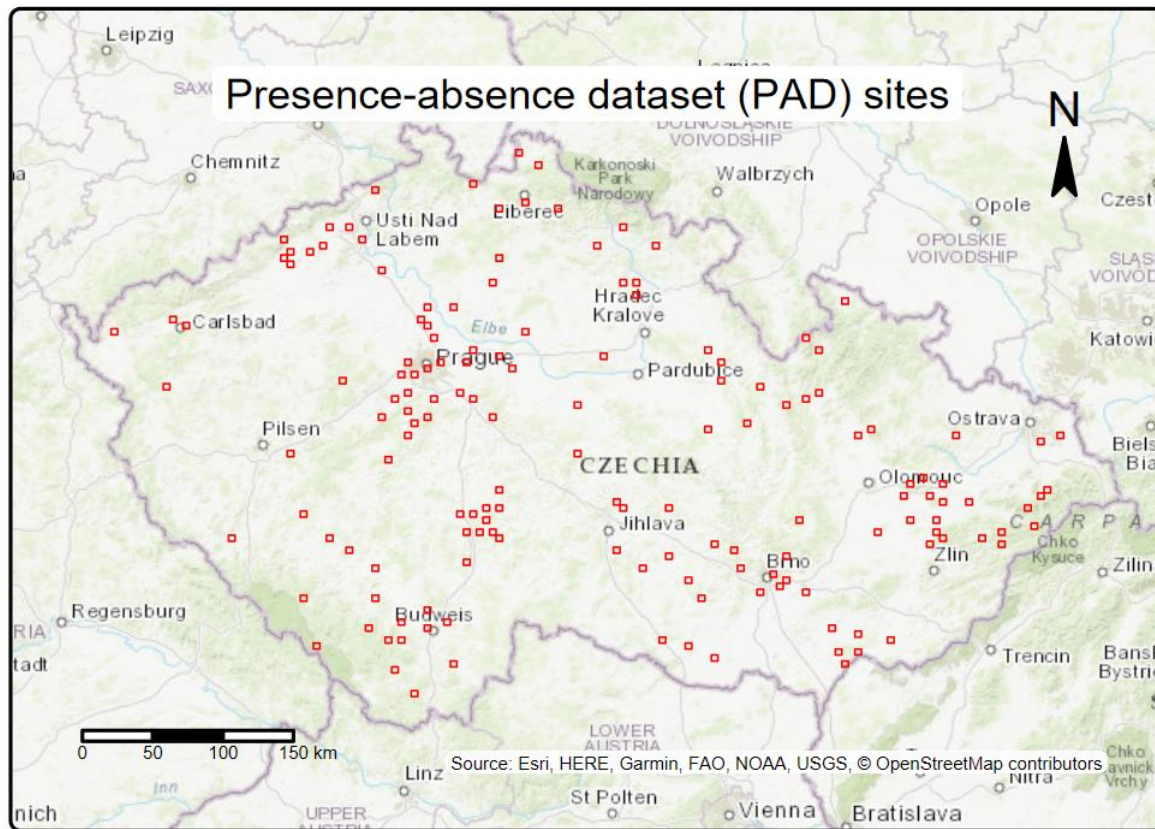
62 Table S1. Basic characteristics regarding the number of presence-only (POD) and presence-
63 absence (PAD) datasets for 109 bird species of the Czech Republic used in this study.

Species	Presences (POD)	Presences (PAD)	Absences (PAD)	Prevalence (PAD)	Observers (POD)
sp102	624	10	144	0.06	512
sp16	95	10	144	0.06	100
sp37	182	10	144	0.06	116
sp79	261	11	143	0.07	153
sp98	507	11	143	0.07	364
sp85	200	12	142	0.08	106
sp96	510	12	142	0.08	405
sp101	904	13	141	0.08	544
sp13	181	13	141	0.08	113
sp54	421	13	141	0.08	326
sp44	383	14	140	0.09	212
sp86	144	14	140	0.09	101
sp99	232	14	140	0.09	193
sp47	290	15	139	0.10	244
sp5	560	15	139	0.10	378
sp18	509	16	138	0.10	345
sp30	681	16	138	0.10	396
sp100	500	17	137	0.11	330
sp45	612	17	137	0.11	399
sp73	709	17	137	0.11	447
sp26	520	18	136	0.12	298
sp22	496	19	135	0.12	298
sp35	306	19	135	0.12	299
sp10	465	21	133	0.14	270
sp76	1344	21	133	0.14	596
sp88	349	21	133	0.14	214
sp50	761	22	132	0.14	477
sp41	649	23	131	0.15	352
sp20	433	25	129	0.16	279
sp40	347	25	129	0.16	153
sp24	456	26	128	0.17	282
sp29	348	26	128	0.17	119
sp46	471	27	127	0.18	373
sp64	805	27	127	0.18	308
sp95	576	28	126	0.18	329
sp61	433	30	124	0.19	230
sp27	215	31	123	0.20	119
sp75	707	31	123	0.20	269
sp34	530	32	122	0.21	204
sp60	473	36	118	0.23	230

sp63	835	37	117	0.24	341
sp66	930	37	117	0.24	388
sp81	963	38	116	0.25	578
sp56	533	41	113	0.27	340
sp48	320	43	111	0.28	103
sp108	939	46	108	0.30	498
sp83	718	46	108	0.30	320
sp72	485	47	107	0.31	205
sp21	227	48	106	0.31	166
sp103	689	49	105	0.32	329
sp25	362	49	105	0.32	124
sp2	1007	50	104	0.32	478
sp74	774	51	103	0.33	335
sp9	464	53	101	0.34	142
sp4	836	56	98	0.36	370
sp90	588	59	95	0.38	210
sp33	1752	63	91	0.41	682
sp82	443	63	91	0.41	166
sp7	456	71	83	0.46	198
sp71	441	75	79	0.49	248
sp8	1119	76	78	0.49	362
sp109	793	78	76	0.51	279
sp92	780	78	76	0.51	275
sp39	1129	80	74	0.52	447
sp1	430	81	73	0.53	208
sp84	585	82	72	0.53	176
sp106	615	85	69	0.55	201
sp32	733	85	69	0.55	292
sp65	1148	85	69	0.55	463
sp77	793	87	67	0.56	206
sp68	725	88	66	0.57	192
sp104	857	90	64	0.58	407
sp17	877	90	64	0.58	243
sp14	741	91	63	0.59	288
sp19	1266	91	63	0.59	523
sp59	985	92	62	0.60	303
sp69	591	93	61	0.60	232
sp28	840	94	60	0.61	275
sp78	903	96	58	0.62	263
sp23	1376	97	57	0.63	638
sp6	547	97	57	0.63	191
sp43	1201	99	55	0.64	424
sp93	900	106	48	0.69	256
sp31	1865	107	47	0.69	605
sp58	813	108	46	0.70	345
sp105	1905	109	45	0.71	671

sp57	909	110	44	0.71	391
sp80	802	110	44	0.71	321
sp11	1896	111	43	0.72	686
sp70	858	112	42	0.73	319
sp52	984	115	39	0.75	339
sp55	1621	115	39	0.75	516
sp67	819	118	36	0.77	301
sp87	1384	118	36	0.77	545
sp51	1225	120	34	0.78	432
sp107	1041	122	32	0.79	336
sp94	1466	124	30	0.81	388
sp53	1227	125	29	0.81	505
sp12	1038	126	28	0.82	315
sp42	1805	126	28	0.82	541
sp62	815	126	28	0.82	298
sp97	1168	126	28	0.82	365
sp36	2421	133	21	0.86	696
sp91	1119	134	20	0.87	422
sp15	2295	135	19	0.88	546
sp38	1254	136	18	0.88	451
sp3	1877	142	12	0.92	508
sp89	1670	144	10	0.94	505

65 Presence-absence dataset (PAD)



66 Figure S1. The red squares show the distribution of the unbiased, independent presence-
67 absence dataset (PAD) in the study area of the Czech Republic.
68

69

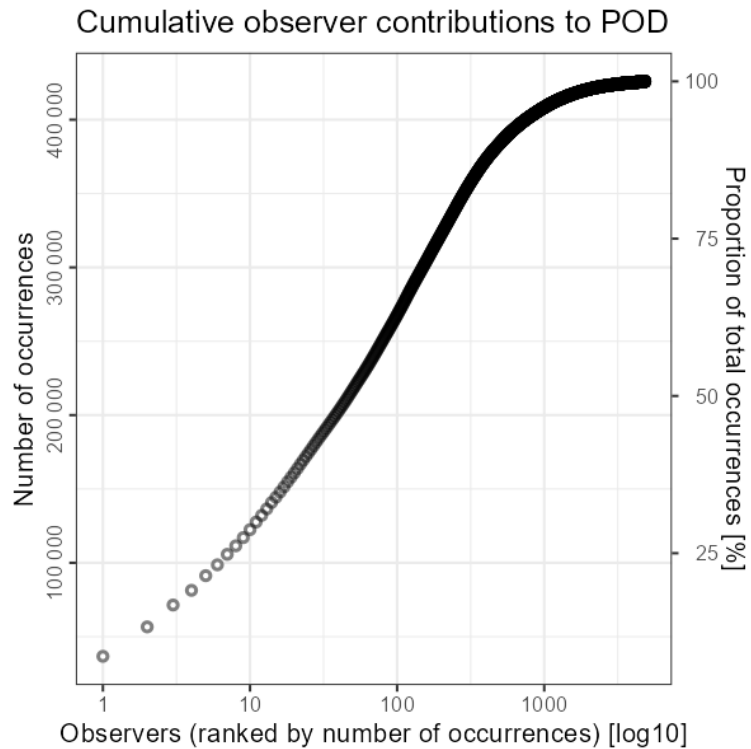
70 Presence-only dataset (POD)

71 Table S2. Pearson's correlations between human population, observers, occurrences and
72 species counts. See Figure 2 in the main text.

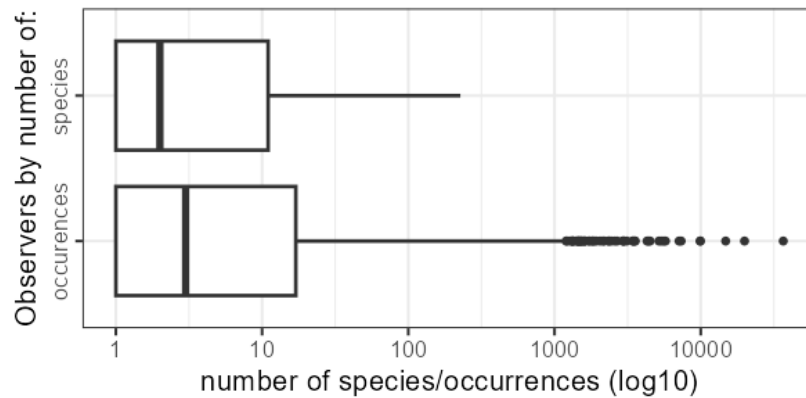
	human population	observers	occurrences	species
human population	1.00	0.48	0.14	0.14
observers	0.48	1.00	0.59	0.57
occurrences	0.14	0.59	1.00	0.74
species	0.14	0.57	0.74	1.00

73

74



75
76 Figure S2. Presence-only database (POD): Relationship between the number of observers and
77 number of observations. X-axis: observers sorted by the number of observations (left observers
78 with the most observations). Y-axis: Cumulative number of observations. The graph reveals that
79 the top 100 "most active" observers accounted for >60% of all the occurrences.
80



81
82 Figure S3. Presence-only database (POD): Distribution of observers by number of species and
83 occurrences. Half of the observers (i.e., the median) contributed only minimally to POD, with no
84 more than two species and three occurrences.
85

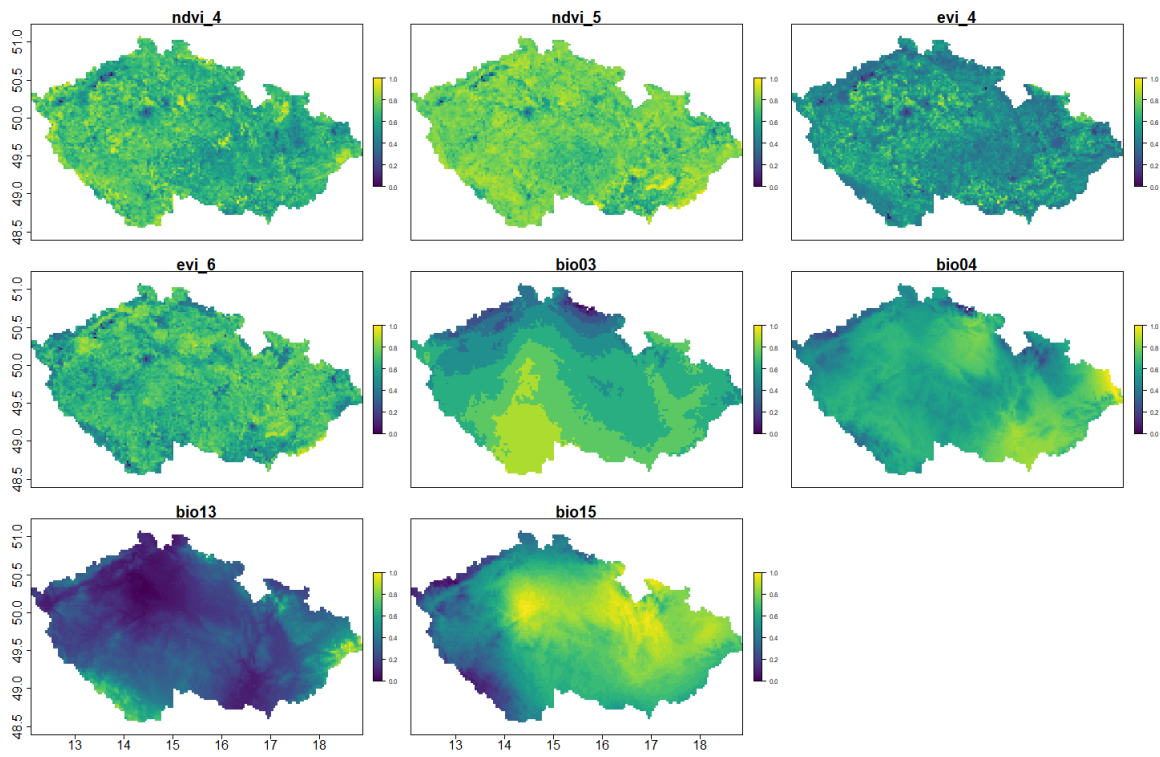
86 Spatial autocorrelation in environmental data

87 Table S3. Moran's I Autocorrelation Index (=Observed column) of PAD sites for each predictor
88 (=Layer column).

Layer	Observed	Expected	SD	P-value
ndvi_4	0.1270	-0.0065	0.0123	0.0000
ndvi_5	0.0748	-0.0065	0.0121	0.0000
evi_4	0.0936	-0.0065	0.0124	0.0000
evi_6	0.0671	-0.0065	0.0123	0.0000
bio03	0.2890	-0.0065	0.0124	0.0000
bio04	0.2124	-0.0065	0.0123	0.0000
bio13	0.3390	-0.0065	0.0123	0.0000
bio15	0.2674	-0.0065	0.0124	0.0000

89

90 Predictors



91
92 Figure S4. Overview of the final predictors used in the study.
93

94 Background points simulation

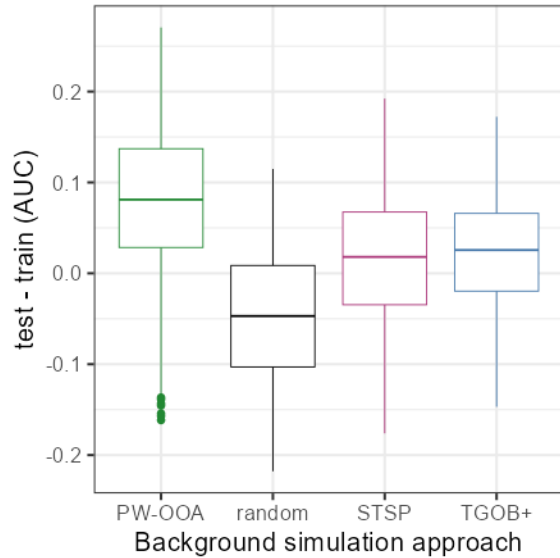
95 Table S4. Conversion table of kernel bandwidths and pixel sizes for OW derived as multiples of
96 the default pixel size (Multiply=1). Kernel density (KD), observer weight (OW).

97 * The last row (1/60, ~ 50 m) specifies the pixel size for aggregating unique observer occurrences.

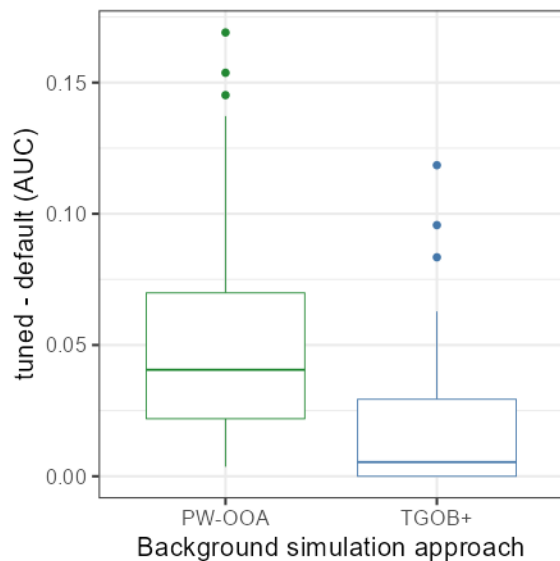
Multiply	Width (deg)	Height (deg)	Width (km)	Height (km)	Used within
0.25	0.010	0.006	0.747	0.695	KD, OW
0.50	0.021	0.013	1.494	1.390	KD, OW
1.00	0.042	0.025	2.987	2.781	KD, OW
2.00	0.083	0.050	5.975	5.561	KD, OW
3.00	0.125	0.075	8.961	8.343	KD
4.00	0.167	0.100	11.949	11.123	KD, OW
8.00	0.333	0.200	23.899	22.246	OW
1/60*	0.001	0.000	0.050	0.046	-

98

99 Results (part A)



100
101 Figure S5: Difference between test and train AUC of all Maxnet model versions including kernel
102 bandwidth change for TGOB+, PW-OOA and observers weights for PW-OOA.
103



104
105 Figure S6: Performance improvement gained by tuning the TGOB+ and PW-OOA approaches
106 compared to their baseline configurations. Baselines are: "traditional" TGOB (using a default bias
107 raster and nKD_{TGOB} of 0.25) and default PW-OOA (using a default bias raster, nKD_{PW-OOA} of 0.25,
108 and an OW pixel size of 1).
109

110 Table S5. The number of species (sum of species), ideal presence thinning distance (STSP version),
111 and kernel bandwidth (TGOB+, PW-OOA version) were selected based on the best prediction

112 performance (AUC) using Maxnet. Thinning distances and kernel bandwidths were derived as
113 multiples of the unit square (Table S4, multiply column).

Approach	Ideal presence thinning distance (STSP) or kernel bandwidth (TGOB+, PW-OOA)	Number of species for which the parameter performed best
PW-OOA	0.25	21
PW-OOA	0.50	18
PW-OOA	1.00	27
PW-OOA	2.00	10
PW-OOA	3.00	11
PW-OOA	4.00	21
STSP	1.00	56
STSP	2.00	19
STSP	3.00	33
TGOB+	0.25	35
TGOB+	0.50	5
TGOB+	1.00	19
TGOB+	2.00	11
TGOB+	3.00	16
TGOB+	4.00	22

114
115 Table S6. Counts of species (sum species) and ideal aggregate pixel size (multiples of the unit
116 square, see Table S4, column multiply) were selected by the best prediction performance (AUC)
117 using Maxnet for PW-OOA. The aggregate pixel size was used to calculate the relative number
118 of unique presences for each author per focus species. The relative number of unique presences
119 serves as a weight (multiple) for the "unit" KD estimate sampling effort of the observer, thus
120 accounting for the individual contribution of each observer to the models.

Ideal pixel size (PW-OOA)	Number of species for which the parameter performed best
0.25	8
0.50	6
1.00	18
2.00	24
4.00	24
8.00	28

121

122 Results (part B)

123 GLM

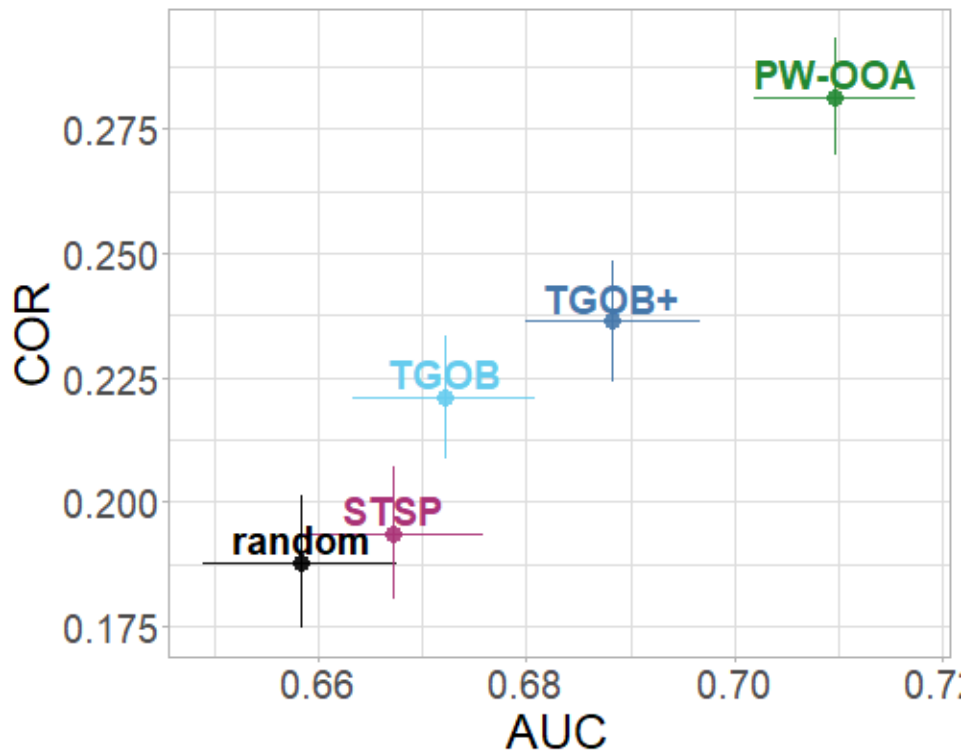
124 AUC

125

126 *Table S7. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using GLM in*
 127 *terms of prediction performance (AUC) for the 108 species of interest. The values represent the number*
 128 *of species for which the model in the row significantly outperformed the model in the column (Unpaired*
 129 *Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the mean (AUC)*
 130 *of the respective method across all species. Spatial thinning of species presences (STSP), Target Group*
 131 *Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths)*
 132 *and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	91	88	87	76	0.710
TGOB+	76	65	62	0.688	22
TGOB	64	54	0.672	0	14
STSP	54	0.667	46	33	12
random	0.658	24	37	27	12

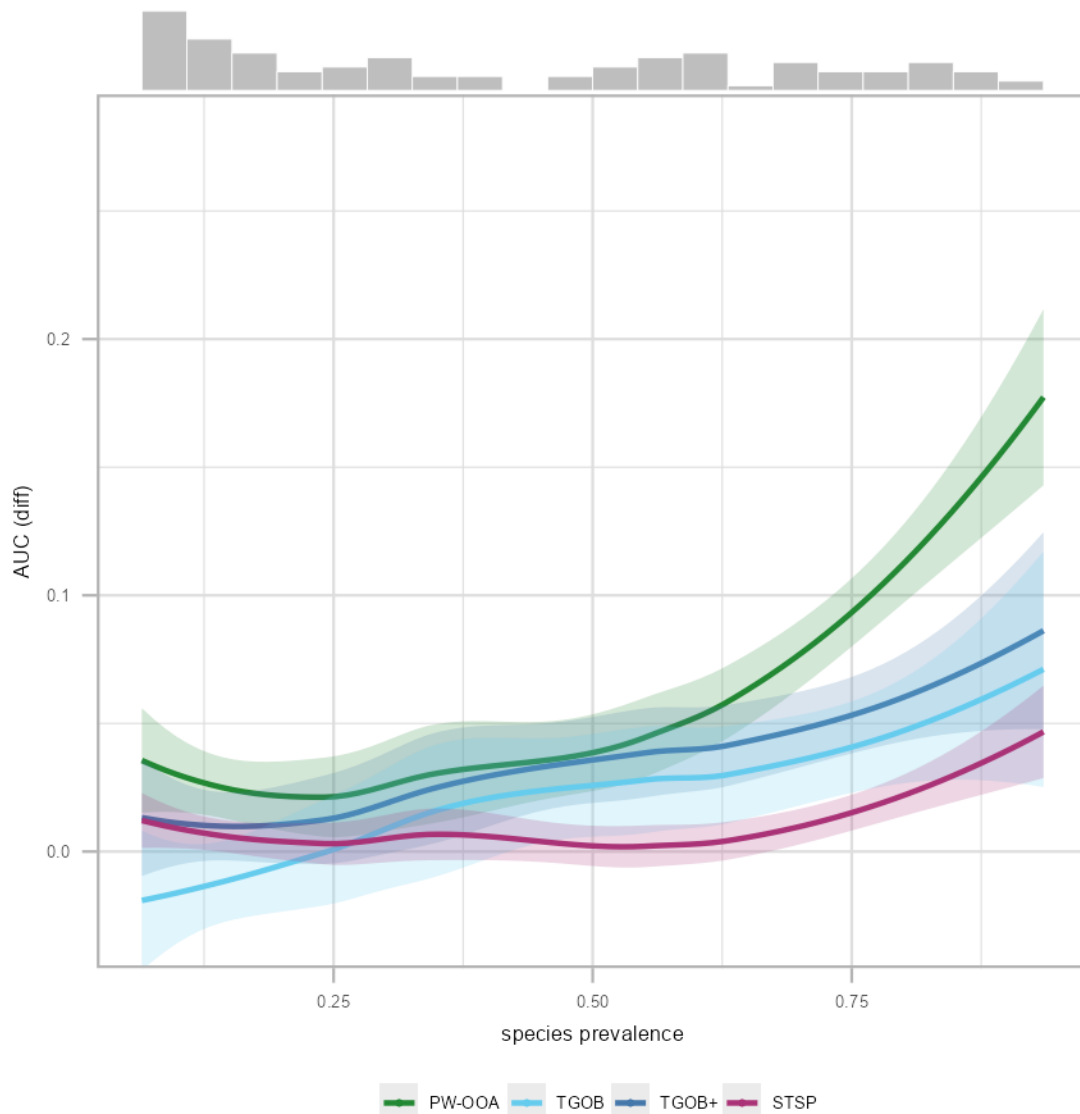
133



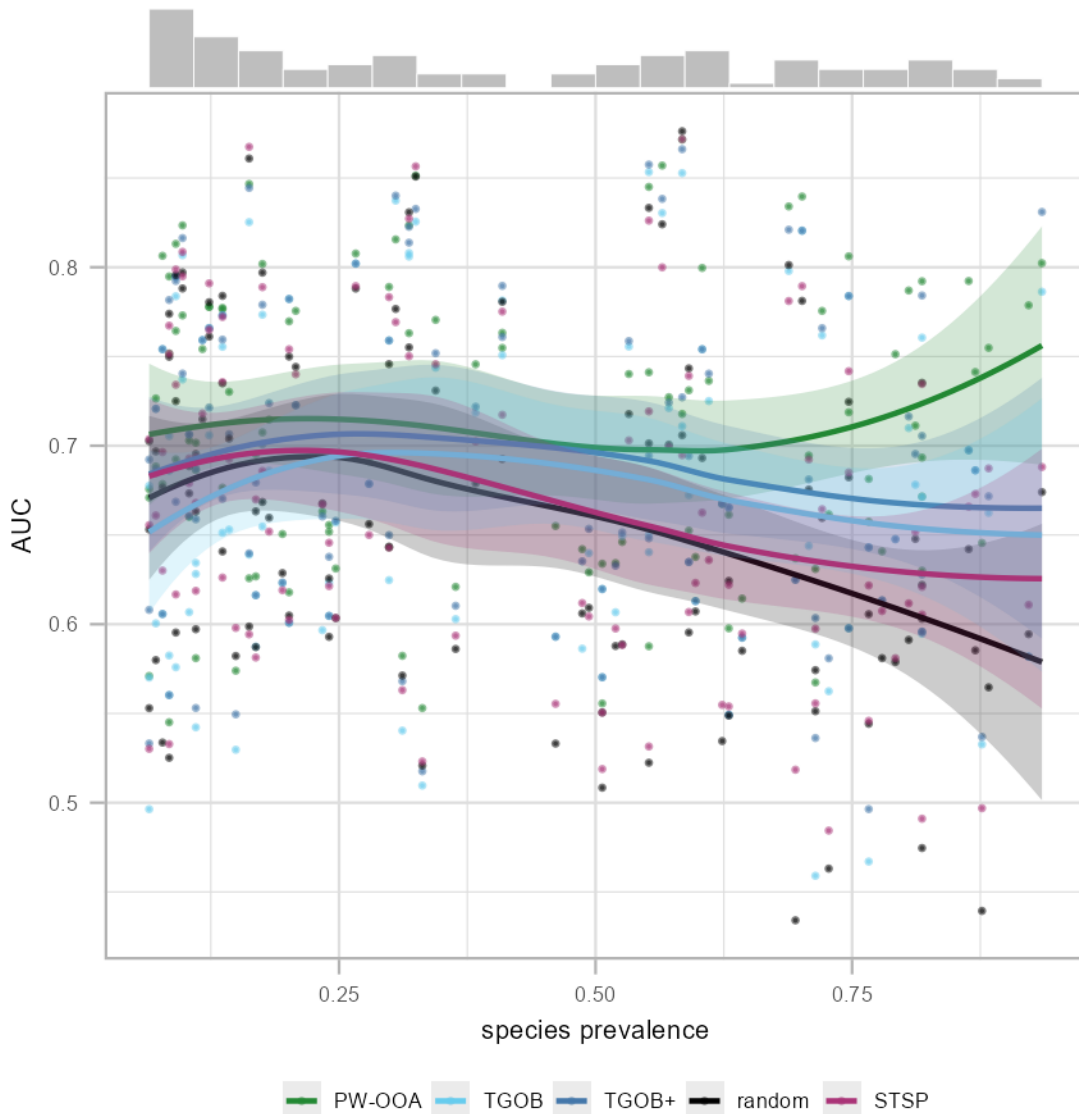
134

135 *Figure S7. The mean AUC (x-axis) and COR (y-axis) of the five presence and/or background simulation*
 136 *methods calculated from 108 bird species using GLM. The bars (around each model) represent standard*
 137 *errors that reflect the variation among all species. Spatial thinning of species presences (STSP), Target*

- 138 *Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
139 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
140

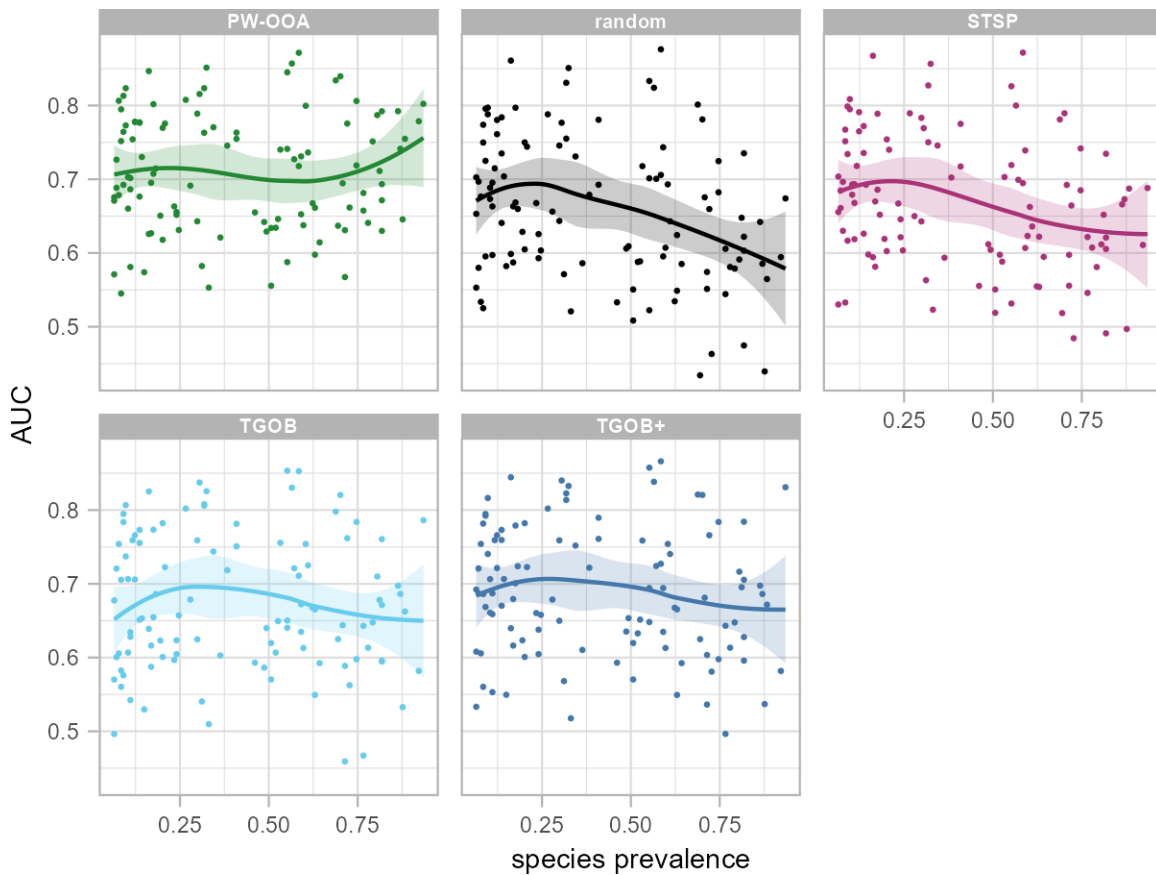


141
 142 *Figure S8. Comparison of SDM prediction performance of 108 bird species differing in species*
 143 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 144 *measured as AUC improvement over the random method (AUC (diff) = difference between the STSP,*
 145 *TGOB, TGOB+, or PW-OOA versus the random method) using GLM. Locally estimated scatterplot*
 146 *smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally*
 147 *weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 148 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 149 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 150 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 151 *(PW-OOA).*
 152



153
154
155
156
157
158
159
160
161
162

Figure S9. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as AUC using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



163
 164 *Figure S10. Comparison of SDM prediction performance of 108 bird species differing in species*
 165 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 166 *measured as AUC using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-*
 167 *parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with*
 168 *95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of species*
 169 *presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 170 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 171

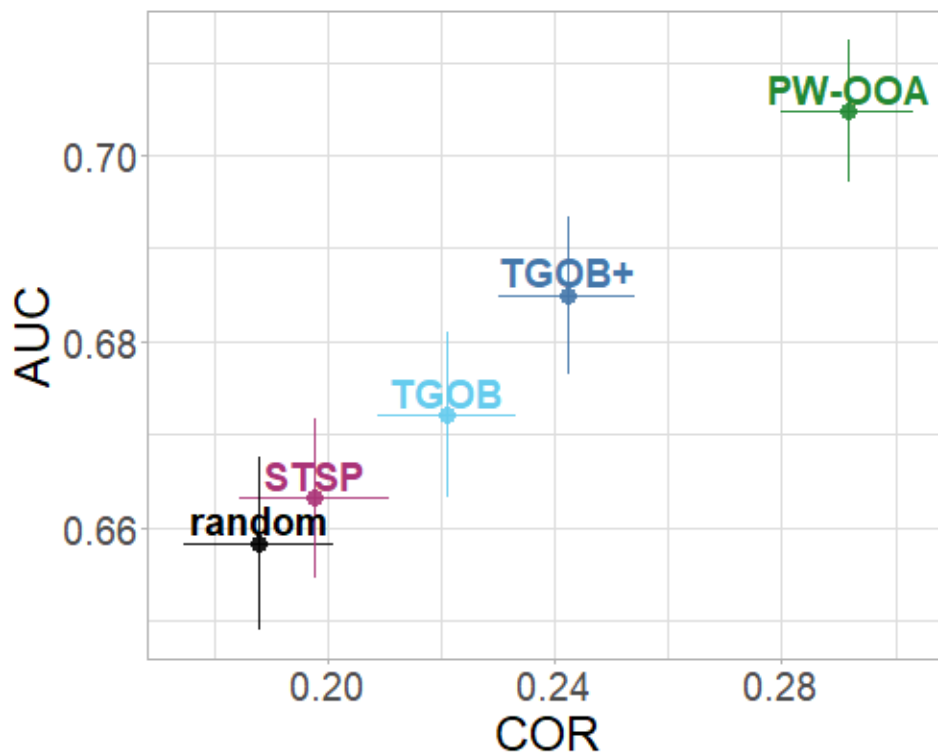
172 COR

173

174 *Table S8. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using GLM in*
 175 *terms of prediction performance (COR) for the 108 species of interest. The values represent the number*
 176 *of species for which the model in the row significantly outperformed the model in the column (Unpaired*
 177 *Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the mean (COR)*
 178 *of the respective method across all species. Spatial thinning of species presences (STSP), Target Group*
 179 *Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths)*
 180 *and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	102	96	97	89	0.292
TGOB+	76	72	61	0.242	12
TGOB	64	59	0.221	0	9
STSP	40	0.198	42	29	5
random	0.188	25	41	27	3

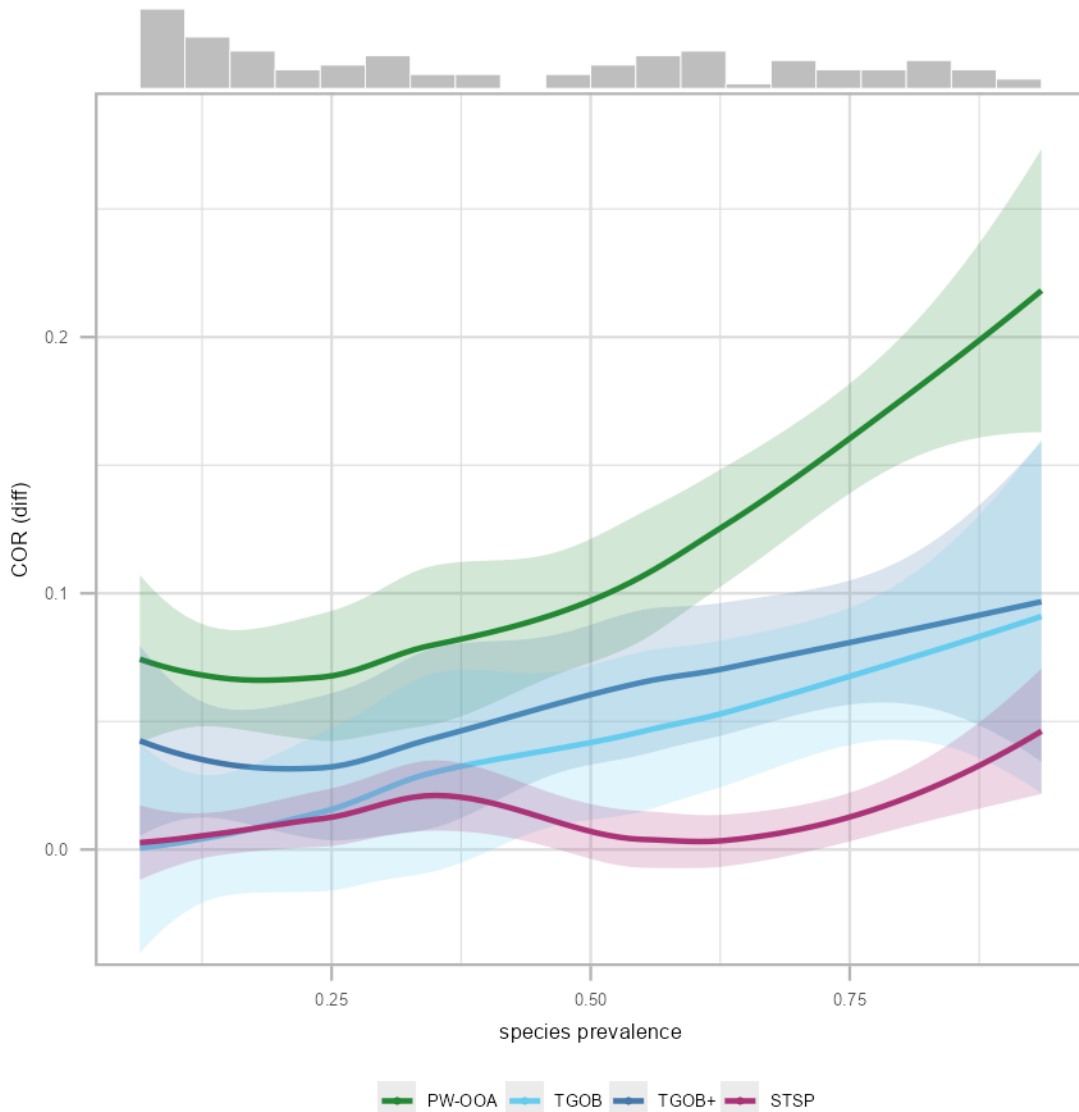
181



182

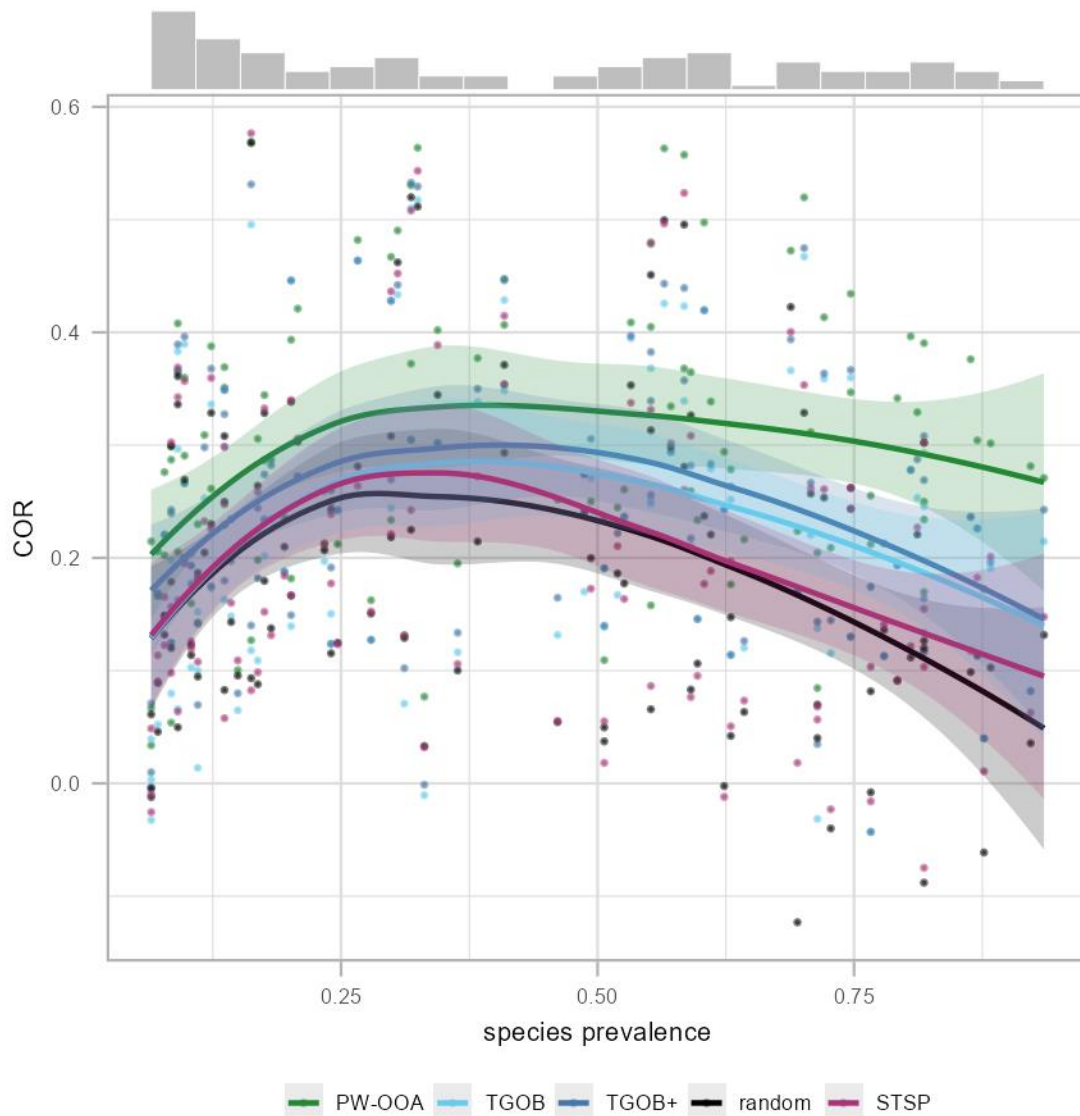
183 *Figure S11. The mean COR (x-axis) and AUC (y-axis) of the five presence and/or background simulation*
 184 *methods calculated from 108 bird species using GLM. The bars (around each model) represent standard*
 185 *errors that reflect the variation among all species. Spatial thinning of species presences (STSP), Target*
 186 *Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 187 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

188



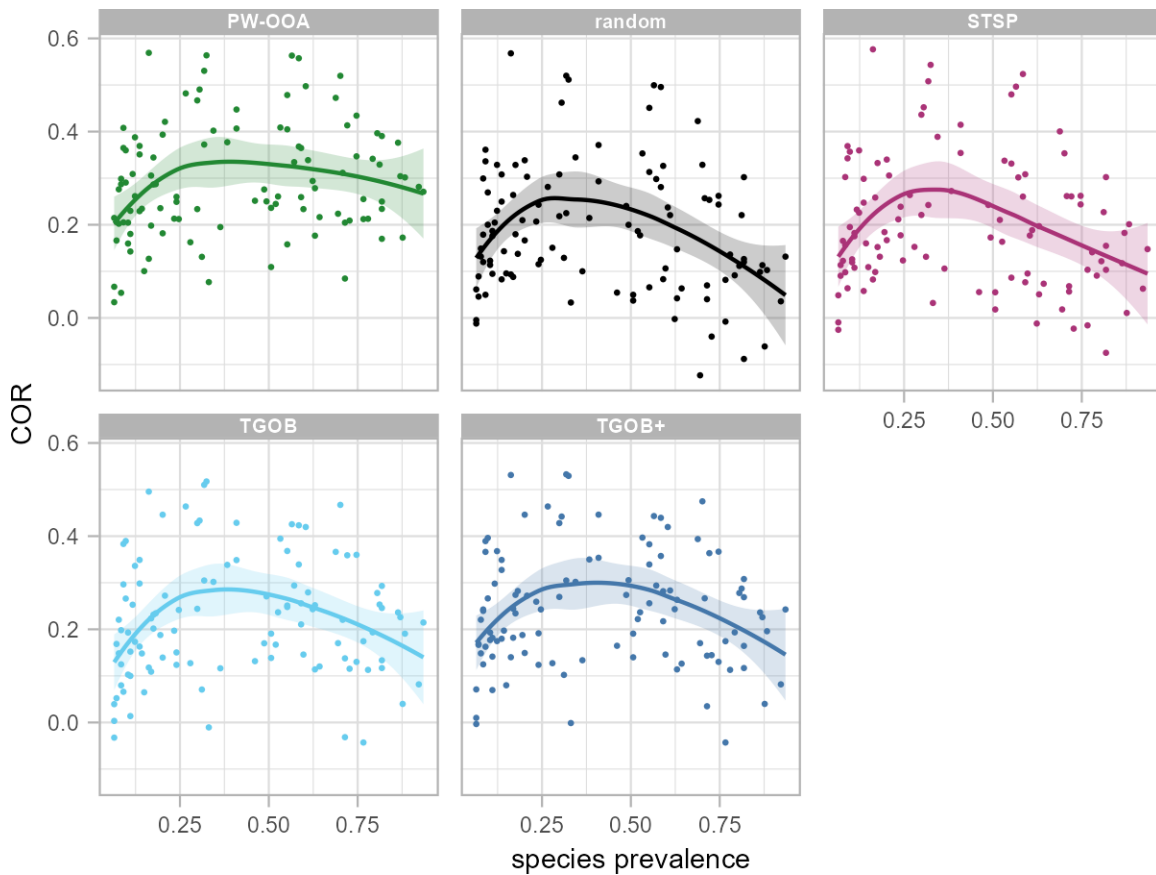
189
190
191
192
193
194
195
196
197
198
199
200

Figure S12. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as COR improvement over the random method ($COR\ (diff) = \text{difference between the STSP, TGOB, TGOB+, or PW-OOA versus the random method}$) using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



201
202
203
204
205
206
207
208
209
210
211

Figure S13. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as COR using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



211
 212 *Figure S14. Comparison of SDM prediction performance of 108 bird species differing in species*
 213 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 214 *measured as COR using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-*
 215 *parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with*
 216 *95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of species*
 217 *presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 218 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 219

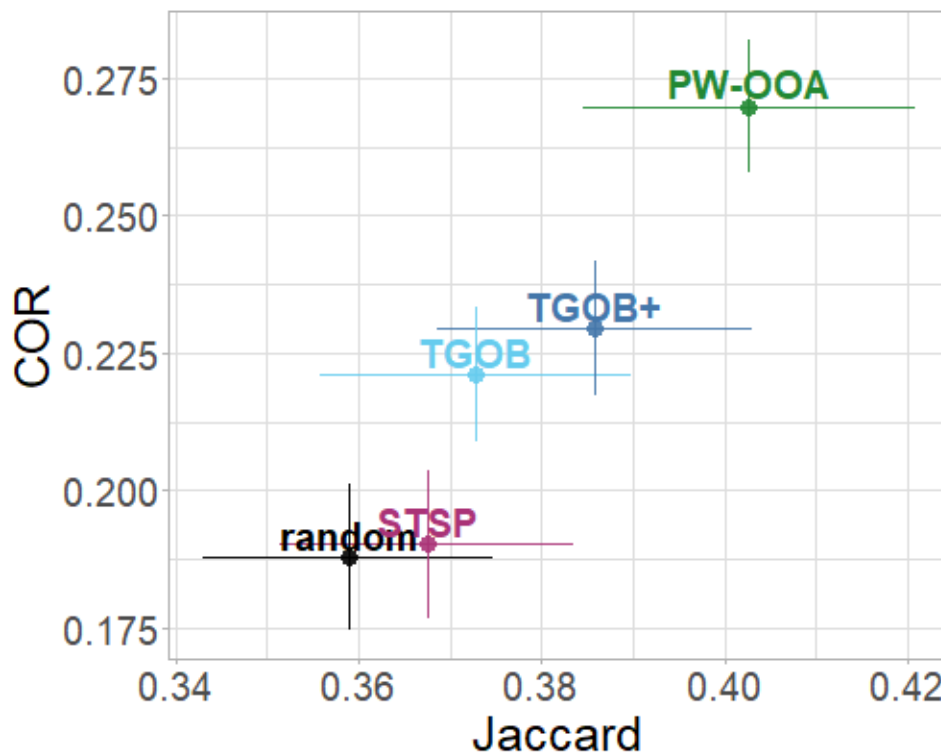
220 Jaccard

221

222 *Table S9. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using GLM in*
 223 *terms of prediction performance (Jaccard) for the 108 species of interest. The values represent the*
 224 *number of species for which the model in the row significantly outperformed the model in the column*
 225 *(Unpaired Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the*
 226 *mean (Jaccard) of the respective method across all species. Spatial thinning of species presences*
 227 *(STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel*
 228 *smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	86	83	79	65	0.403
TGOB+	72	67	54	0.386	26
TGOB	59	54	0.373	0	13
STSP	49	0.367	42	24	11
random	0.359	17	39	23	10

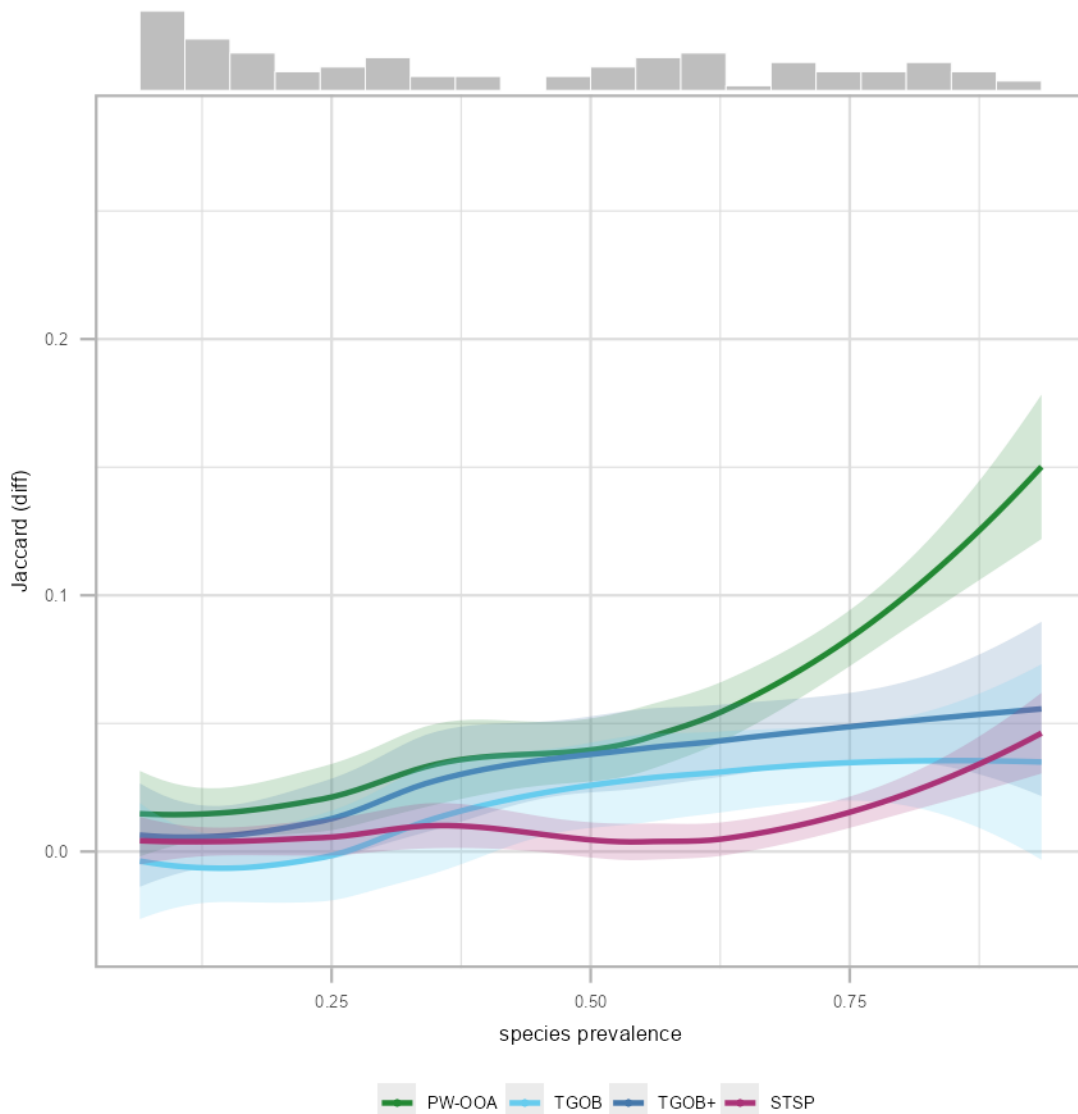
229



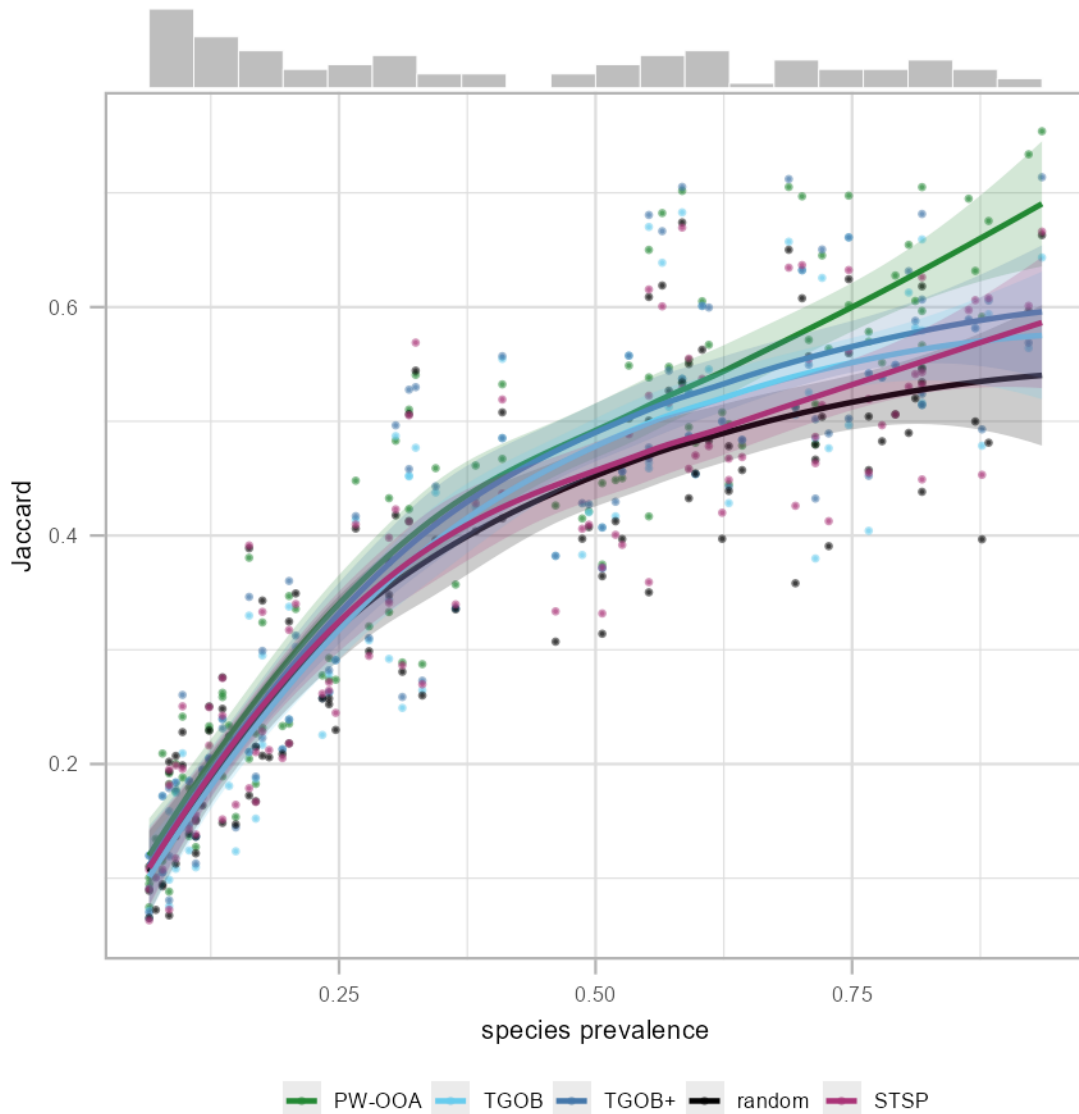
230

231 *Figure S15. The mean Jaccard (x-axis) and COR (y-axis) of the five presence and/or background*
 232 *simulation methods calculated from 108 bird species using GLM. The bars (around each model)*
 233 *represent standard errors that reflect the variation among all species. Spatial thinning of species*
 234 *presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 235 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

236

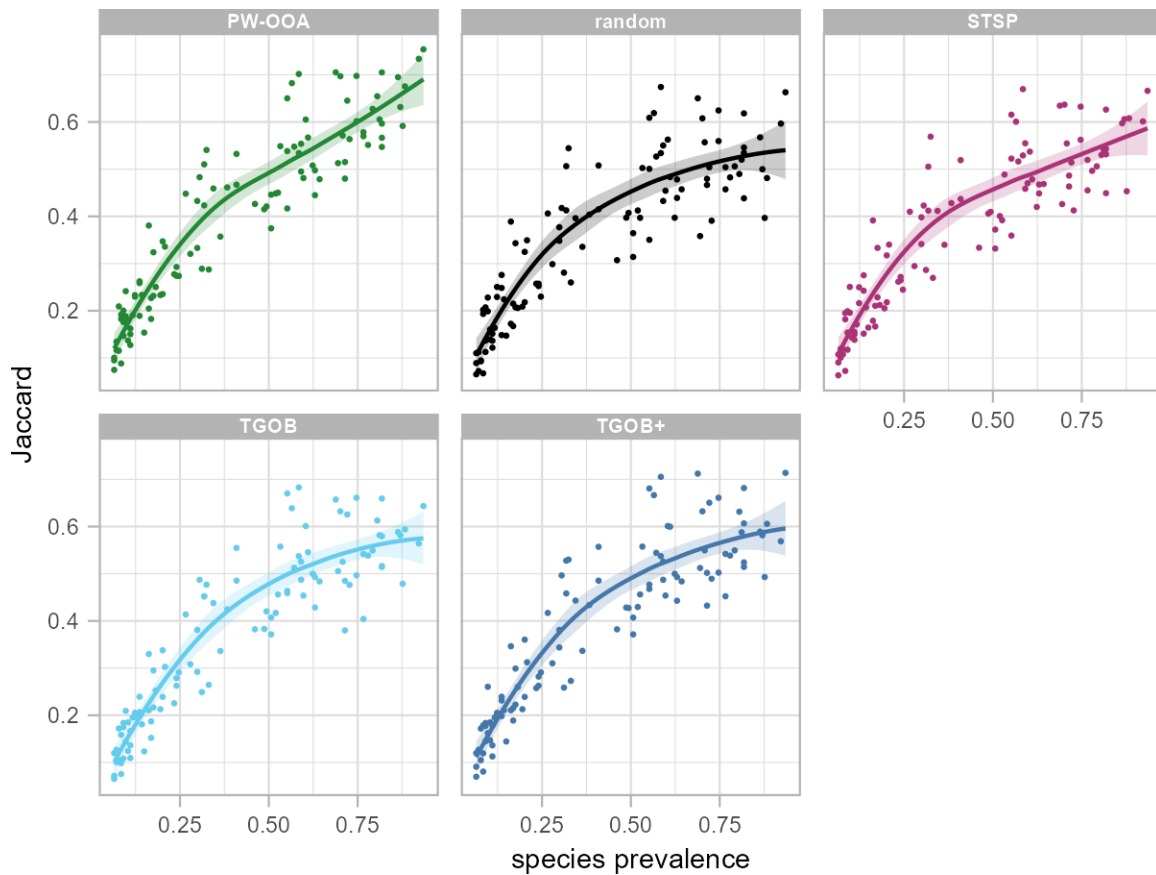


237
 238 *Figure S16. Comparison of SDM prediction performance of 108 bird species differing in species*
 239 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 240 *measured as Jaccard improvement over the random method (Jaccard (diff) = difference between the*
 241 *STSP, TGOB, TGOB+, or PW-OOA versus the random method) using GLM. Locally estimated scatterplot*
 242 *smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally*
 243 *weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 244 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 245 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 246 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 247 *(PW-OOA).*
 248



249
250
251
252
253
254
255
256
257
258

Figure S17. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as Jaccard using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



259
260
261
262
263
264
265
266
267

Figure S18. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as Jaccard using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).

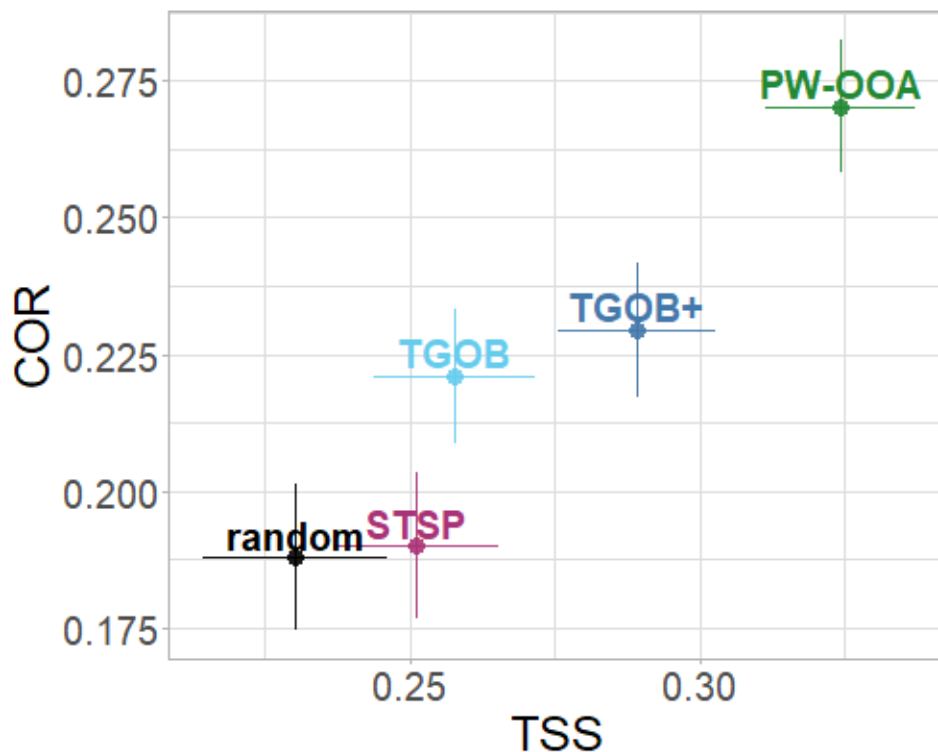
268 TSS

269

270 *Table S10. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using GLM in*
 271 *terms of prediction performance (TSS) for the 108 species of interest. The values represent the number*
 272 *of species for which the model in the row significantly outperformed the model in the column (Unpaired*
 273 *Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the mean (TSS) of*
 274 *the respective method across all species. Spatial thinning of species presences (STSP), Target Group*
 275 *Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths)*
 276 *and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	86	83	79	65	0.324
TGOB+	72	67	54	0.289	26
TGOB	59	54	0.258	0	13
STSP	49	0.251	42	24	11
random	0.230	17	39	23	10

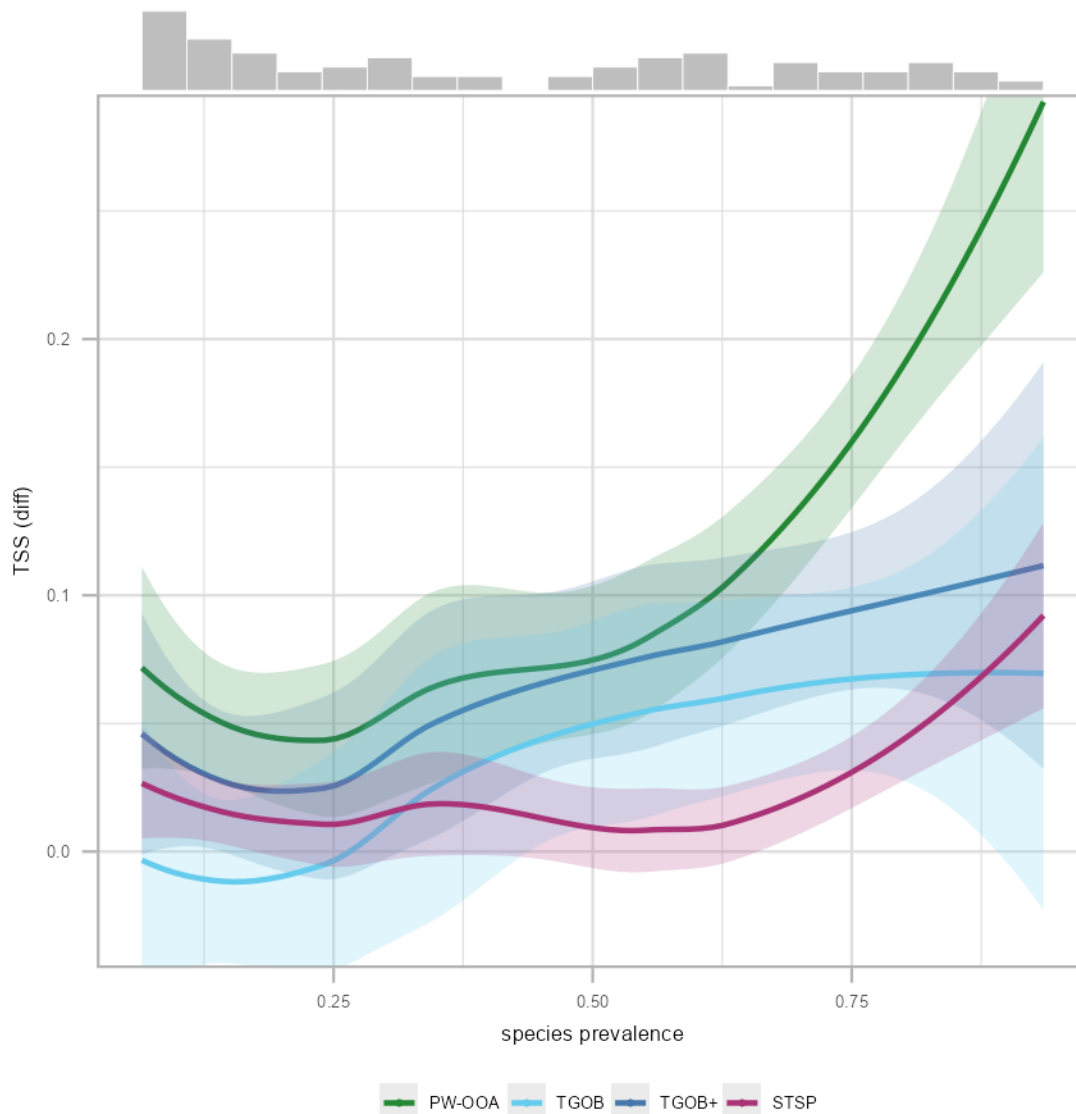
277



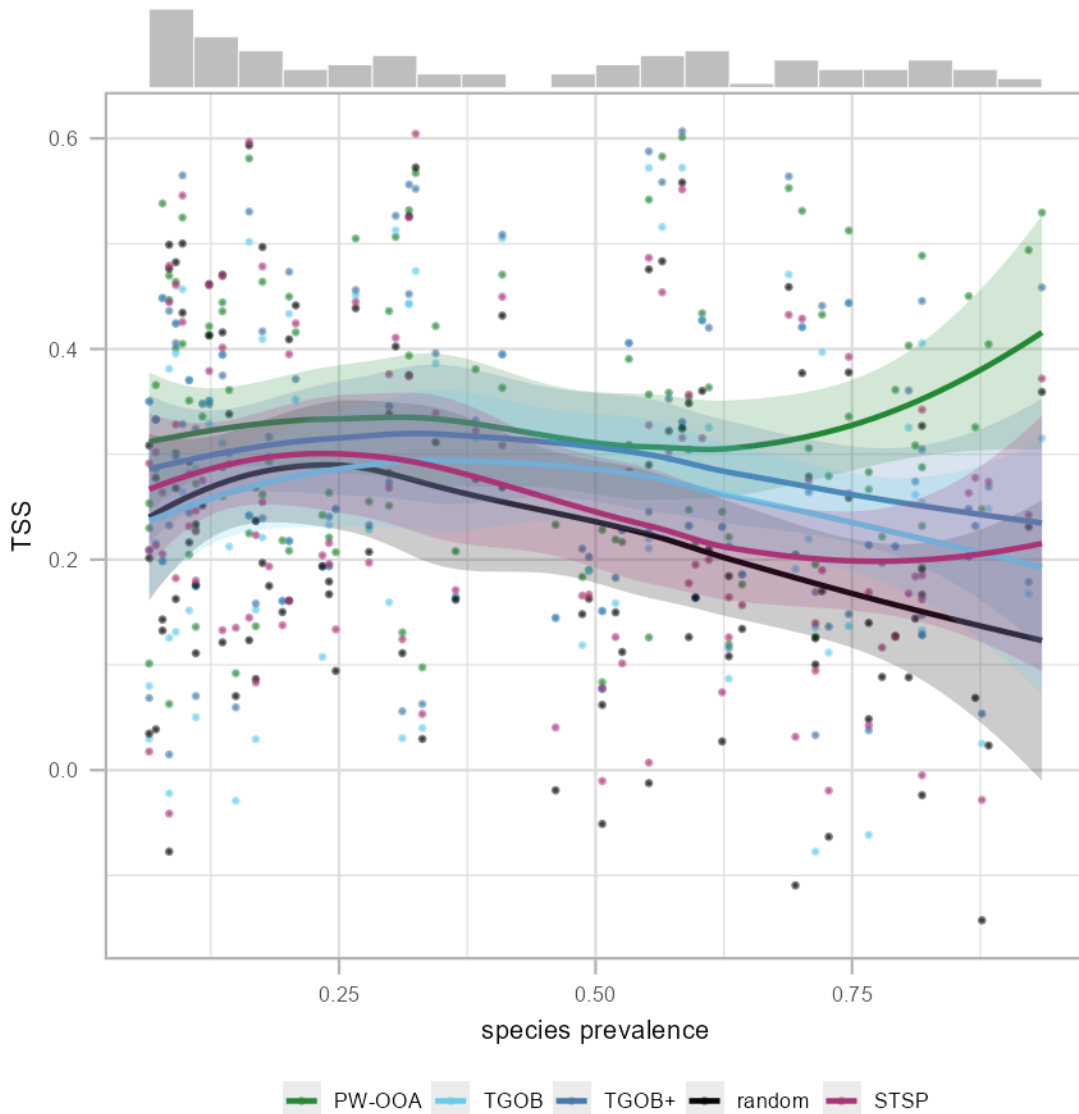
278

279 *Figure S19. The mean TSS (x-axis) and COR (y-axis) of the five presence and/or background simulation*
 280 *methods calculated from 108 bird species using GLM. The bars (around each model) represent standard*
 281 *errors that reflect the variation among all species. Spatial thinning of species presences (STSP), Target*
 282 *Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 283 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

284

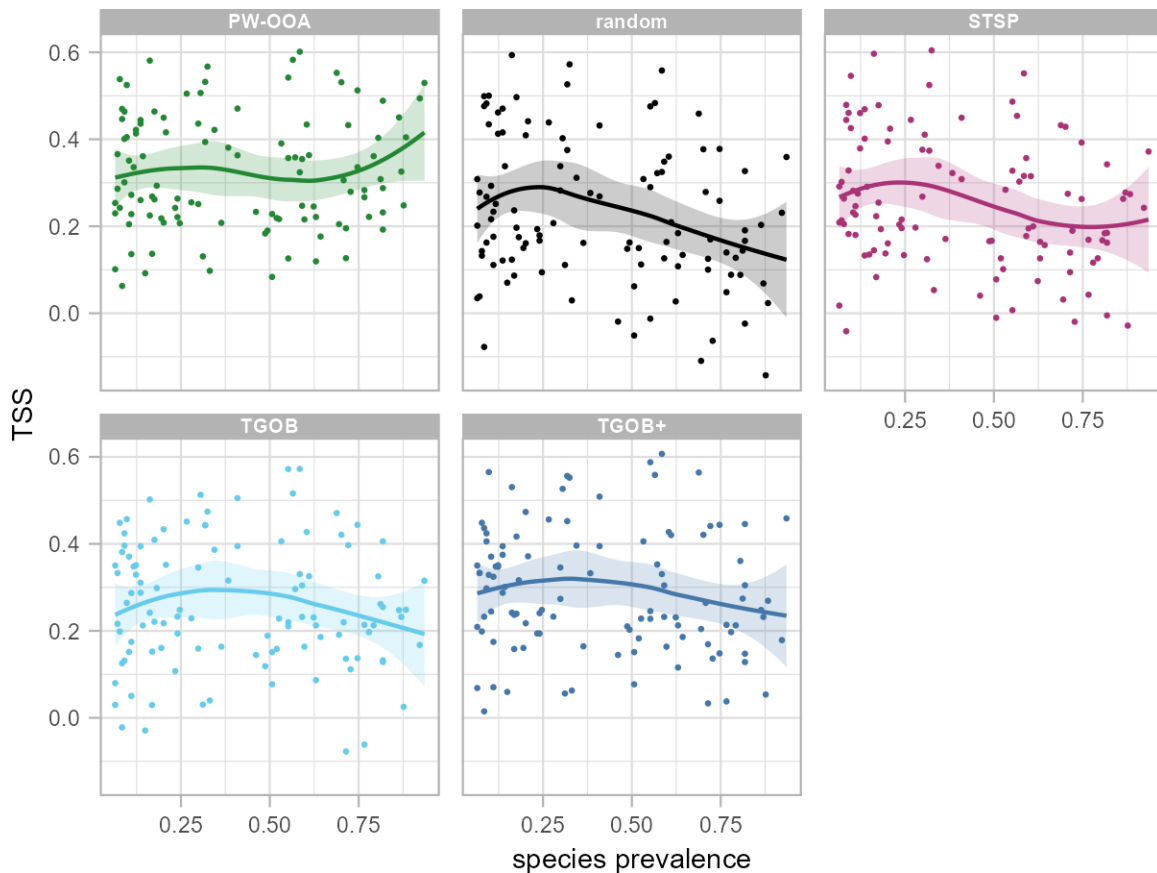


285
 286 *Figure S20. Comparison of SDM prediction performance of 108 bird species differing in species*
 287 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 288 *measured as TSS improvement over the random method ($TSS (diff) = \text{difference between the STSP,}$*
 289 *TGOB, TGOB+, or PW-OOA versus the random method) using GLM. Locally estimated scatterplot*
 290 *smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally*
 291 *weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 292 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 293 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 294 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 295 *(PW-OOA).*
 296



297
298
299
300
301
302
303
304
305
306

Figure S21. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as TSS using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



307
 308 *Figure S22. Comparison of SDM prediction performance of 108 bird species differing in species*
 309 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 310 *measured as TSS using GLM. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-*
 311 *parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with*
 312 *95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of species*
 313 *presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 314 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 315

316 Maxnet

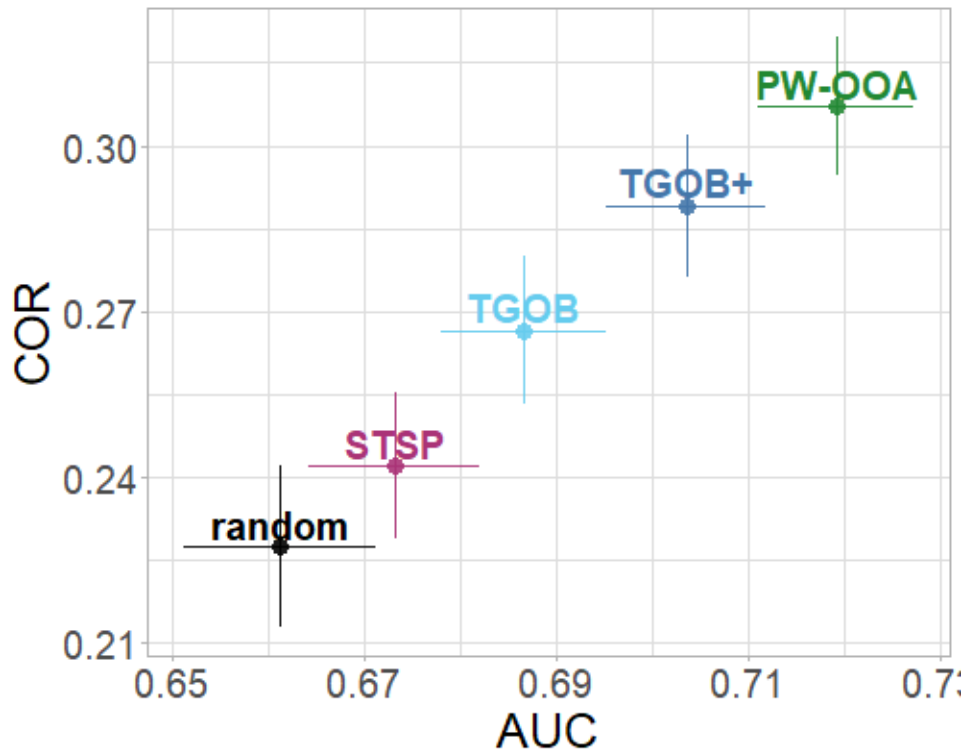
317 AUC

318

319 *Table S11. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using Maxnet in*
 320 *terms of prediction performance (AUC) for the 108 species of interest. The values represent the number*
 321 *of species for which the model in the row significantly outperformed the model in the column (Unpaired*
 322 *Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the mean (AUC)*
 323 *of the respective method across all species. Spatial thinning of species presences (STSP), Target Group*
 324 *Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths)*
 325 *and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	95	90	79	59	0.719
TGOB+	84	76	54	0.703	25
TGOB	69	57	0.686	0	15
STSP	49	0.673	36	18	13
random	0.661	31	36	17	8

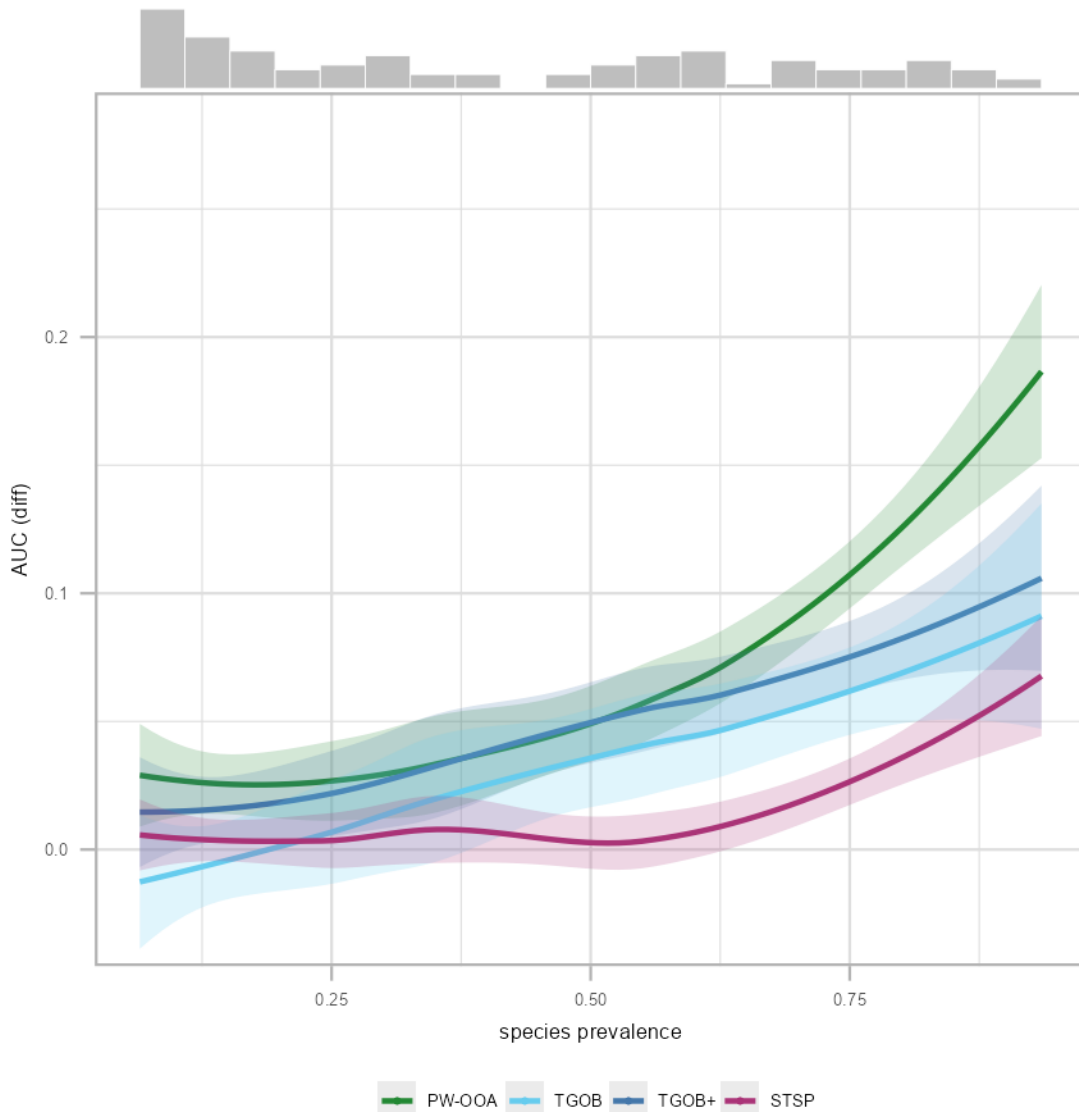
326



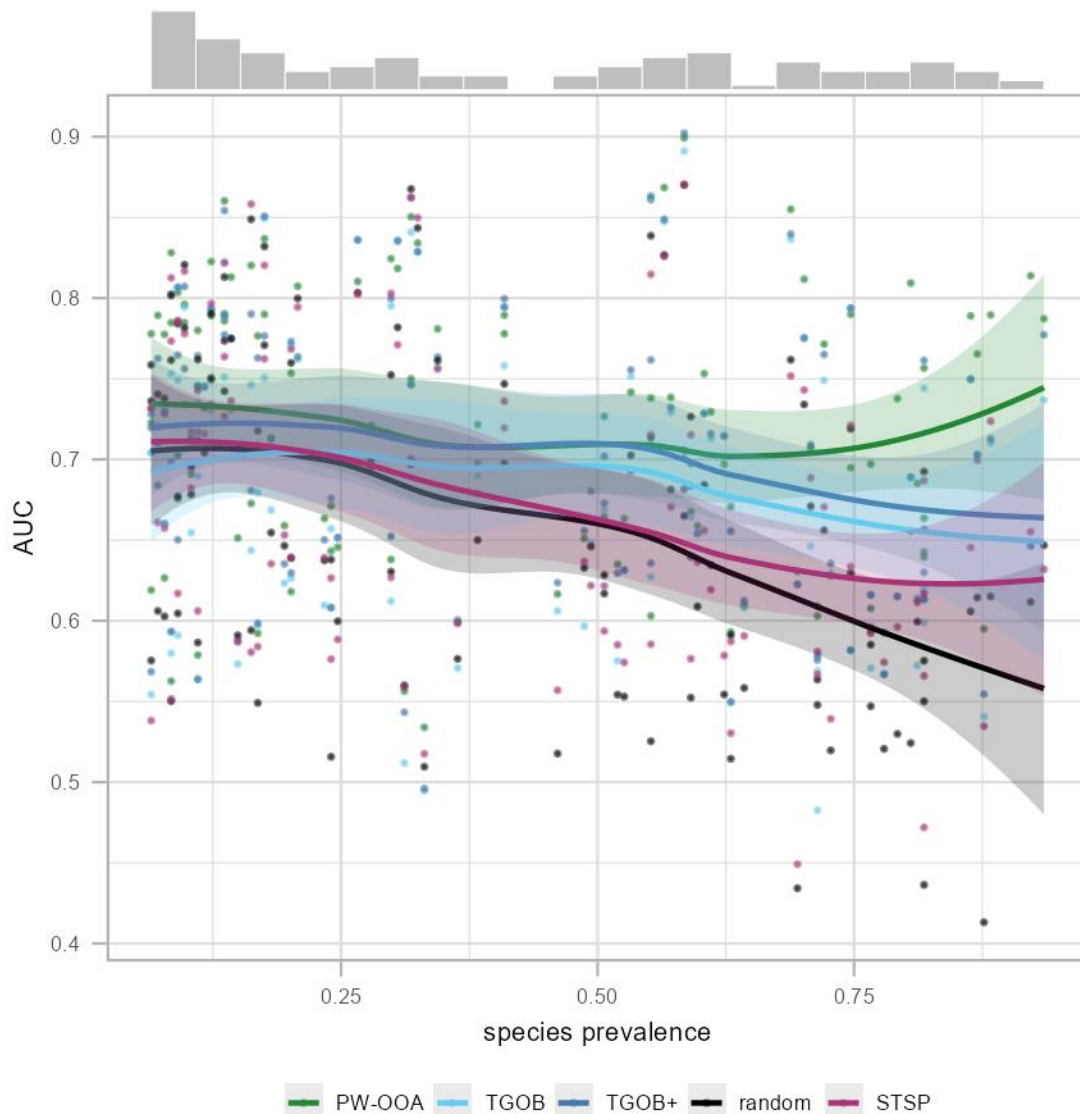
327

328 *Figure S23. The mean AUC (x-axis) and COR (y-axis) of the five presence and/or background simulation*
 329 *methods calculated from 108 bird species using Maxnet. The bars (around each model) represent*
 330 *standard errors that reflect the variation among all species. Spatial thinning of species presences (STSP),*
 331 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 332 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

333

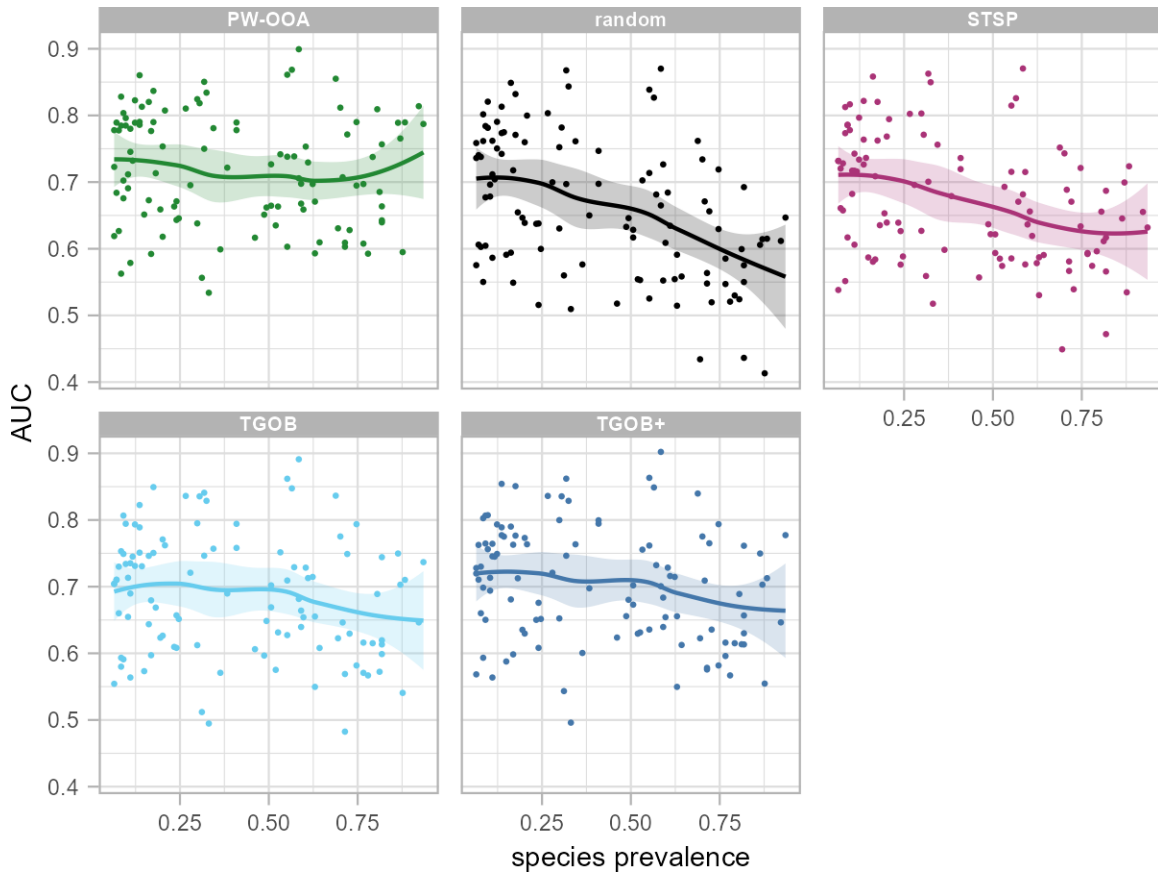


334
 335 *Figure S24. Comparison of SDM prediction performance of 108 bird species differing in species*
 336 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 337 *measured as AUC improvement over the random method (AUC (diff) = difference between the STSP,*
 338 *TGOB, TGOB+, or PW-OOA versus the random method) using Maxnet. Locally estimated scatterplot*
 339 *smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally*
 340 *weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 341 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 342 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 343 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 344 *(PW-OOA).*
 345



346
347
348
349
350
351
352
353
354
355

Figure S25. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as AUC using Maxnet. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



356
357
358
359
360
361
362
363
364

Figure S26. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as AUC using Maxnet. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).

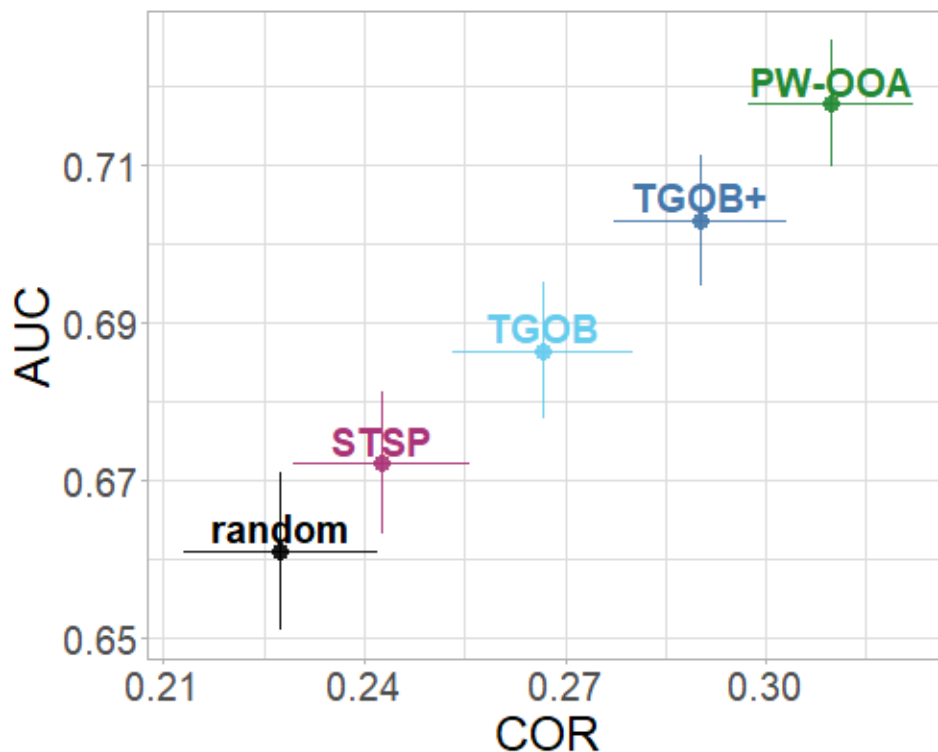
365 COR

366

367 *Table S12. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using Maxnet in*
 368 *terms of prediction performance (COR) for the 108 species of interest. The values represent the number*
 369 *of species for which the model in the row significantly outperformed the model in the column (Unpaired*
 370 *Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the mean (COR)*
 371 *of the respective method across all species. Spatial thinning of species presences (STSP), Target Group*
 372 *Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths)*
 373 *and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	87	86	78	57	0.310
TGOB+	84	80	52	0.290	31
TGOB	68	59	0.267	0	19
STSP	46	0.243	36	17	12
random	0.228	27	35	18	10

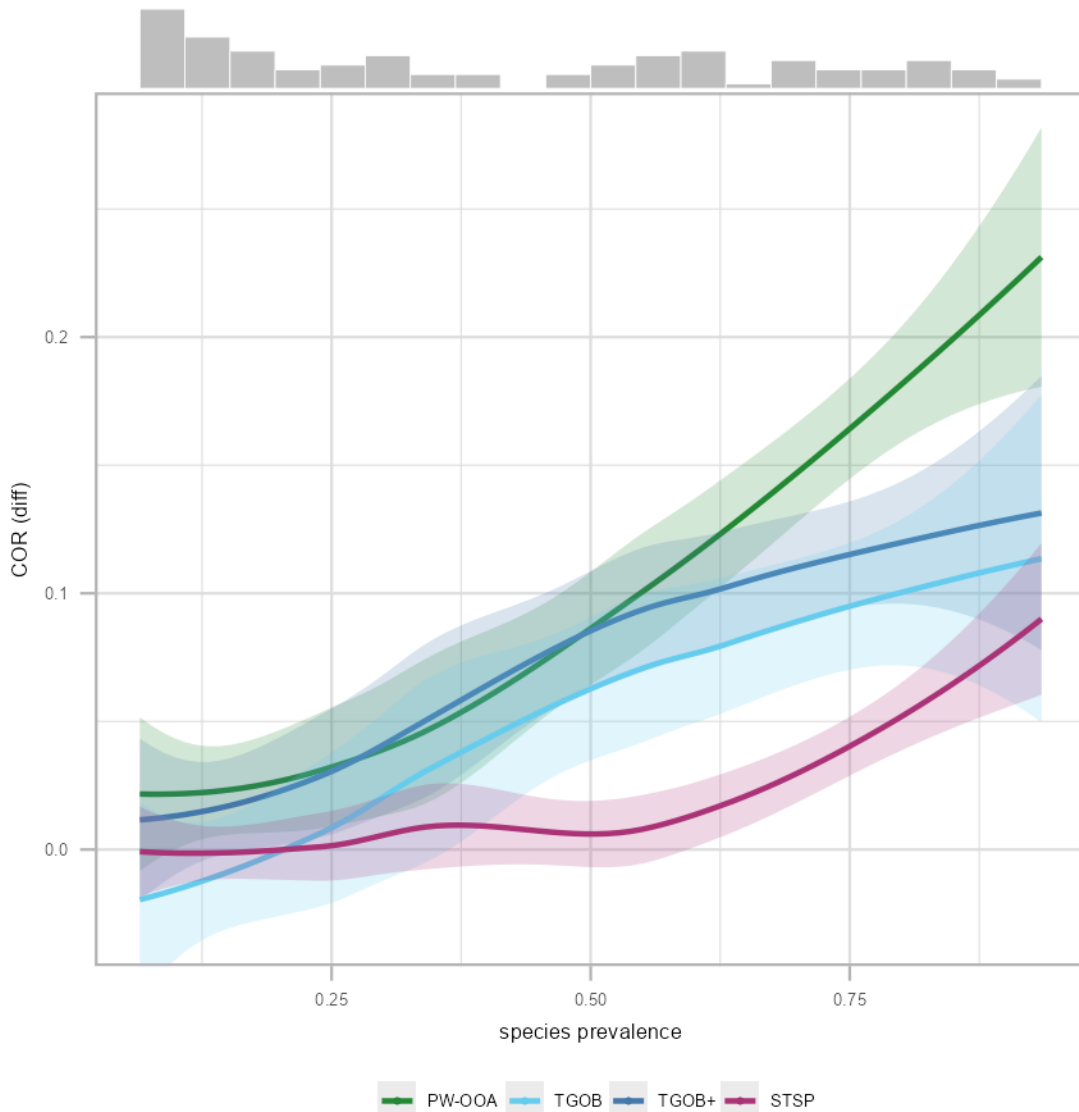
374



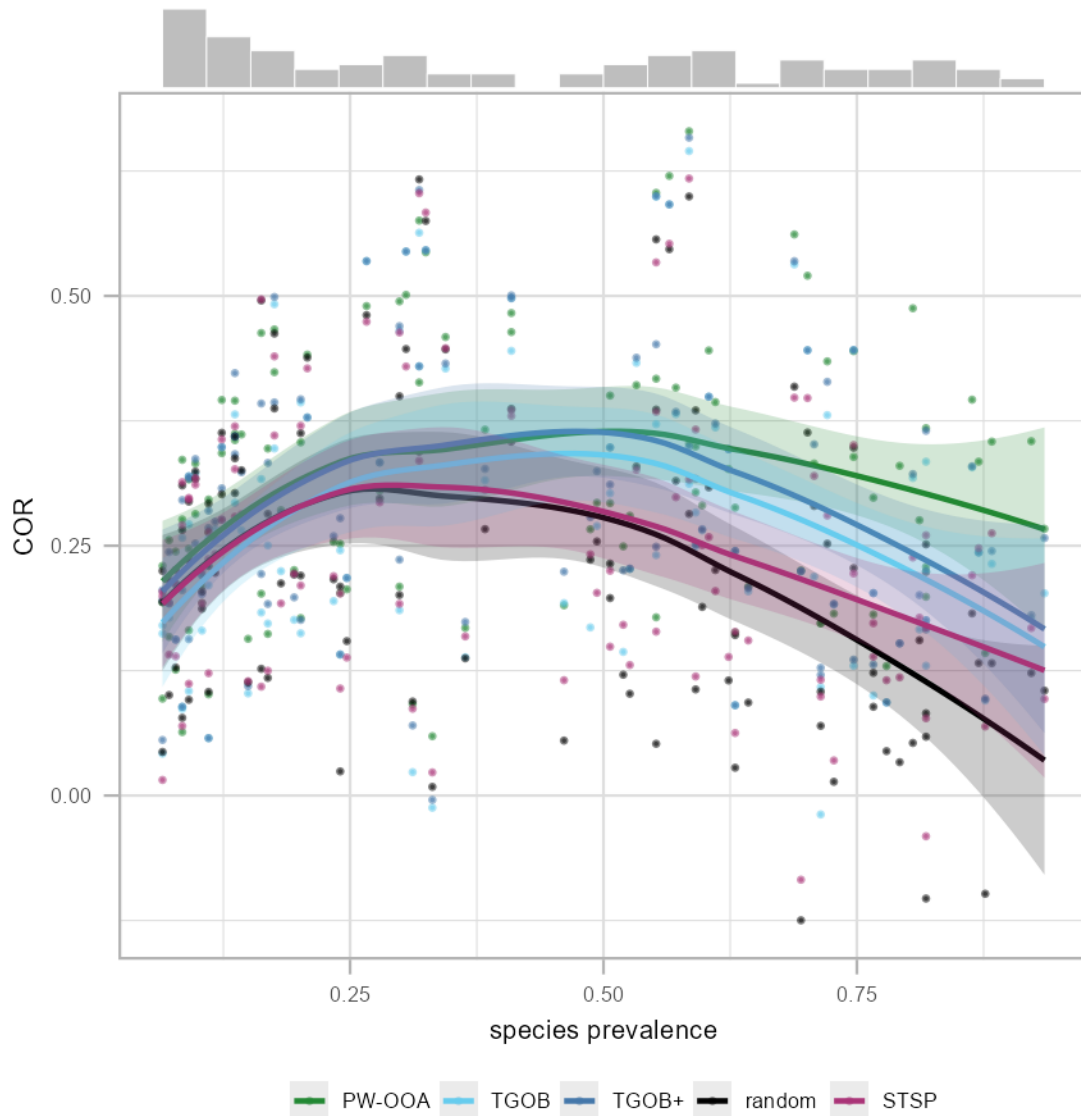
375

376 *Figure S27. The mean COR (x-axis) and AUC (y-axis) of the five presence and/or background simulation*
 377 *methods calculated from 108 bird species using Maxnet. The bars (around each model) represent*
 378 *standard errors that reflect the variation among all species. Spatial thinning of species presences (STSP),*
 379 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 380 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

381

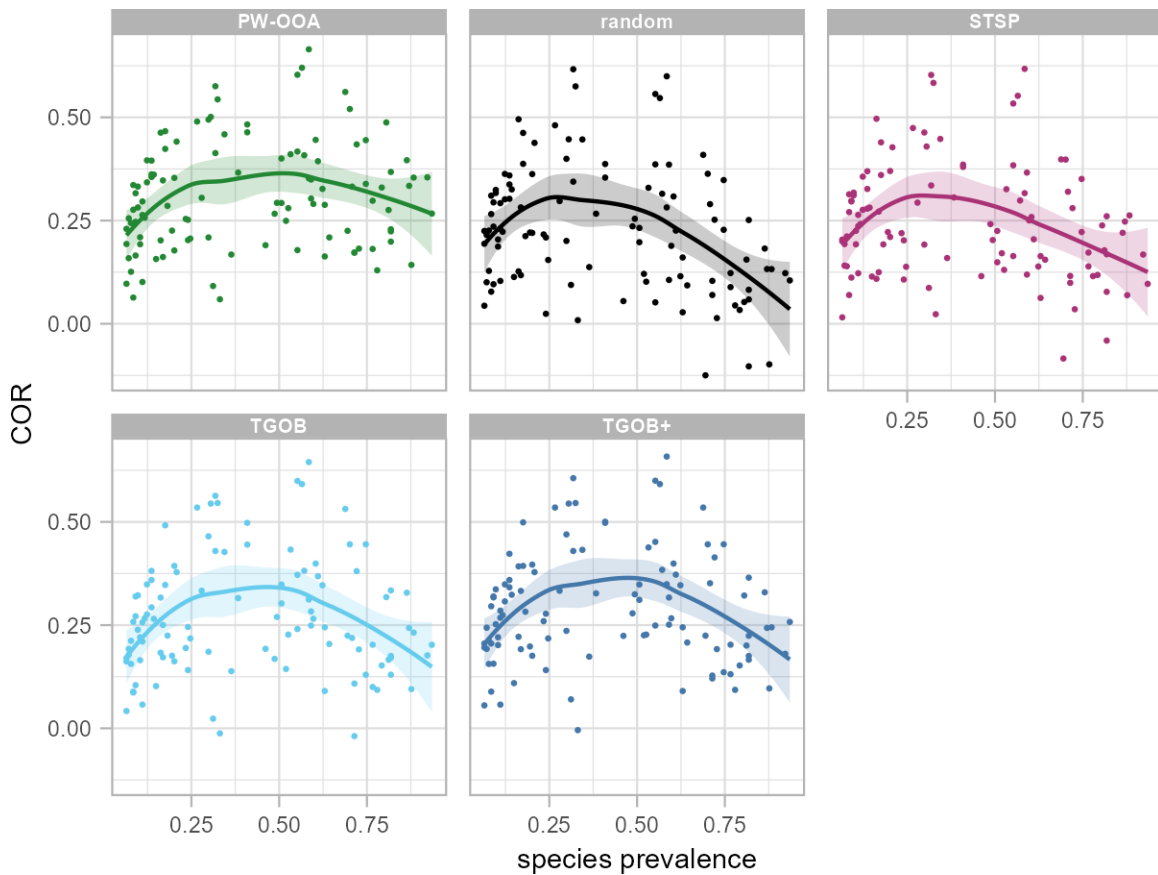


382
 383 *Figure S28. Comparison of SDM prediction performance of 108 bird species differing in species*
 384 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 385 *measured as COR improvement over the random method (COR (diff) = difference between the STSP,*
 386 *TGOB, TGOB+, or PW-OOA versus the random method) using Maxnet. Locally estimated scatterplot*
 387 *smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally*
 388 *weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 389 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 390 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 391 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 392 *(PW-OOA).*
 393



394
395
396
397
398
399
400
401
402
403

Figure S29. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as COR using Maxnet. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



404
 405 *Figure S30. Comparison of SDM prediction performance of 108 bird species differing in species*
 406 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 407 *measured as COR using Maxnet. Locally estimated scatterplot smoothing (LOESS) regression lines, a*
 408 *non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown*
 409 *with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of species*
 410 *presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 411 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 412

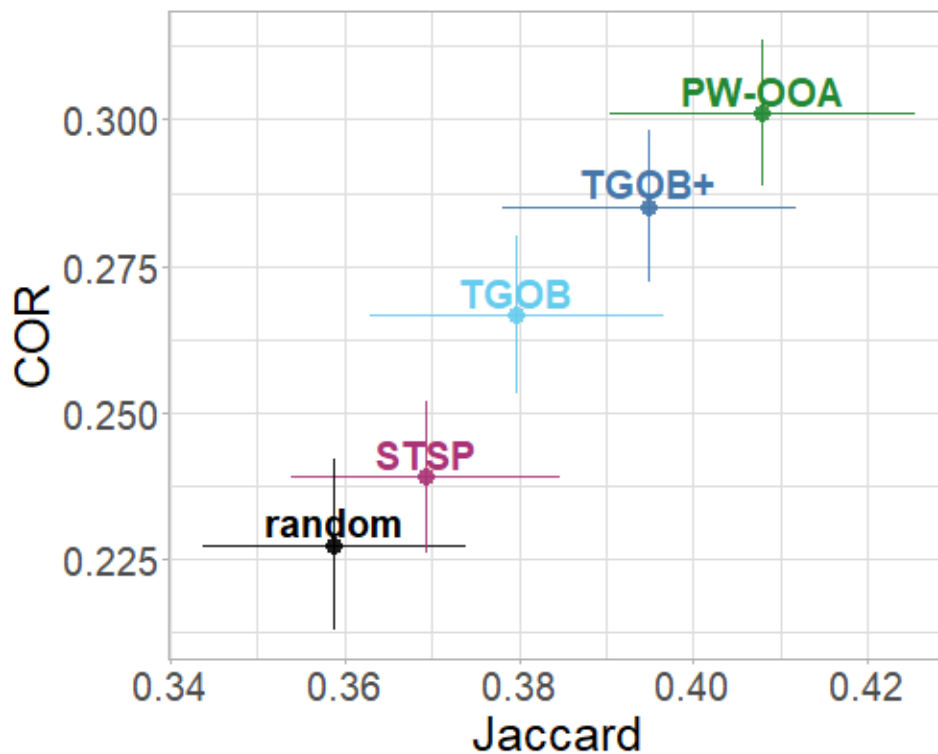
413 Jaccard

414

415 *Table S13. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using Maxnet in*
 416 *terms of prediction performance (Jaccard) for the 108 species of interest. The values represent the*
 417 *number of species for which the model in the row significantly outperformed the model in the column*
 418 *(Unpaired Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the*
 419 *mean (Jaccard) of the respective method across all species. Spatial thinning of species presences*
 420 *(STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel*
 421 *smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	92	84	78	52	0.408
TGOB+	87	68	51	0.395	25
TGOB	65	50	0.380	0	12
STSP	44	0.369	39	19	13
random	0.359	20	31	13	8

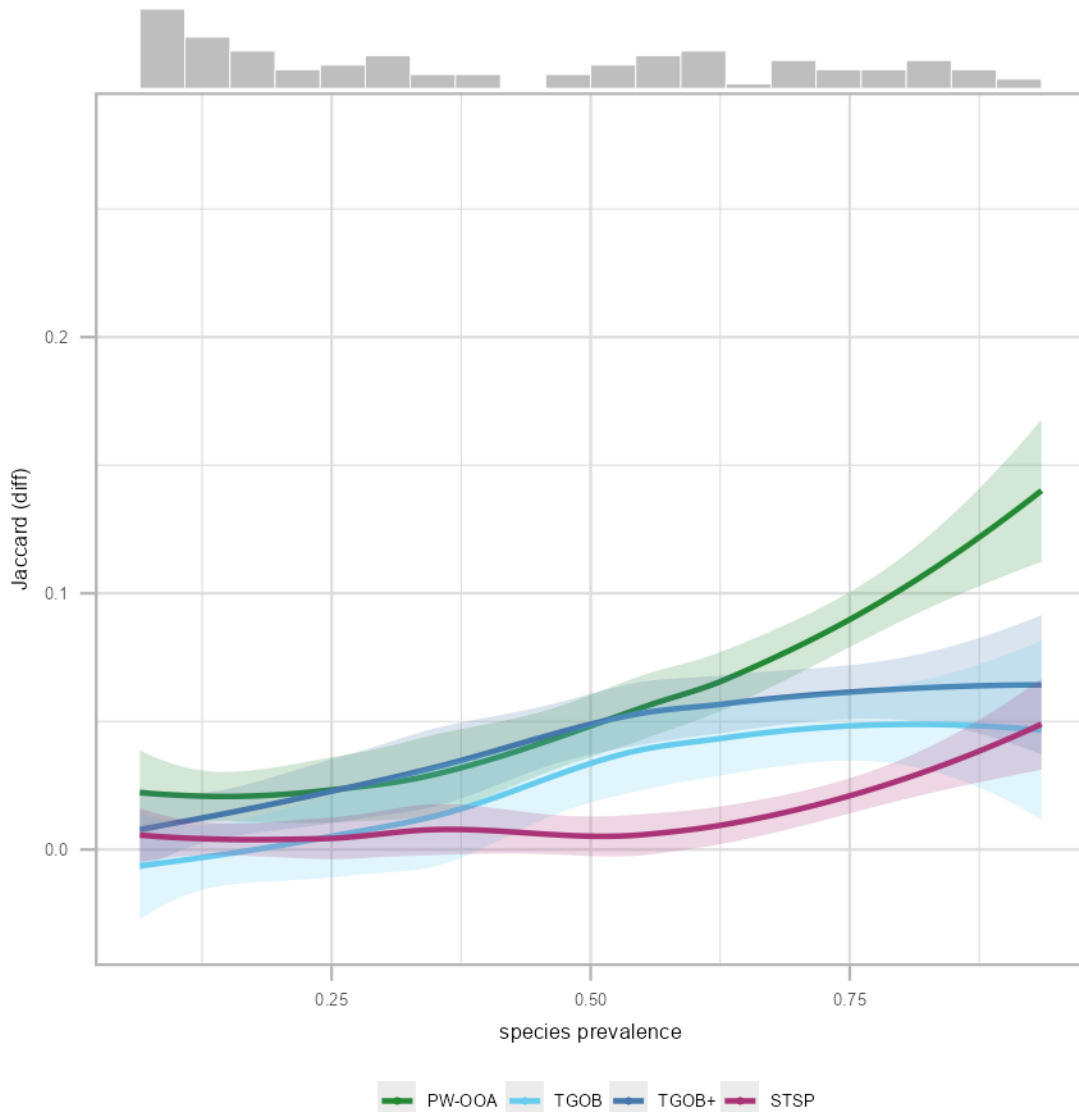
422



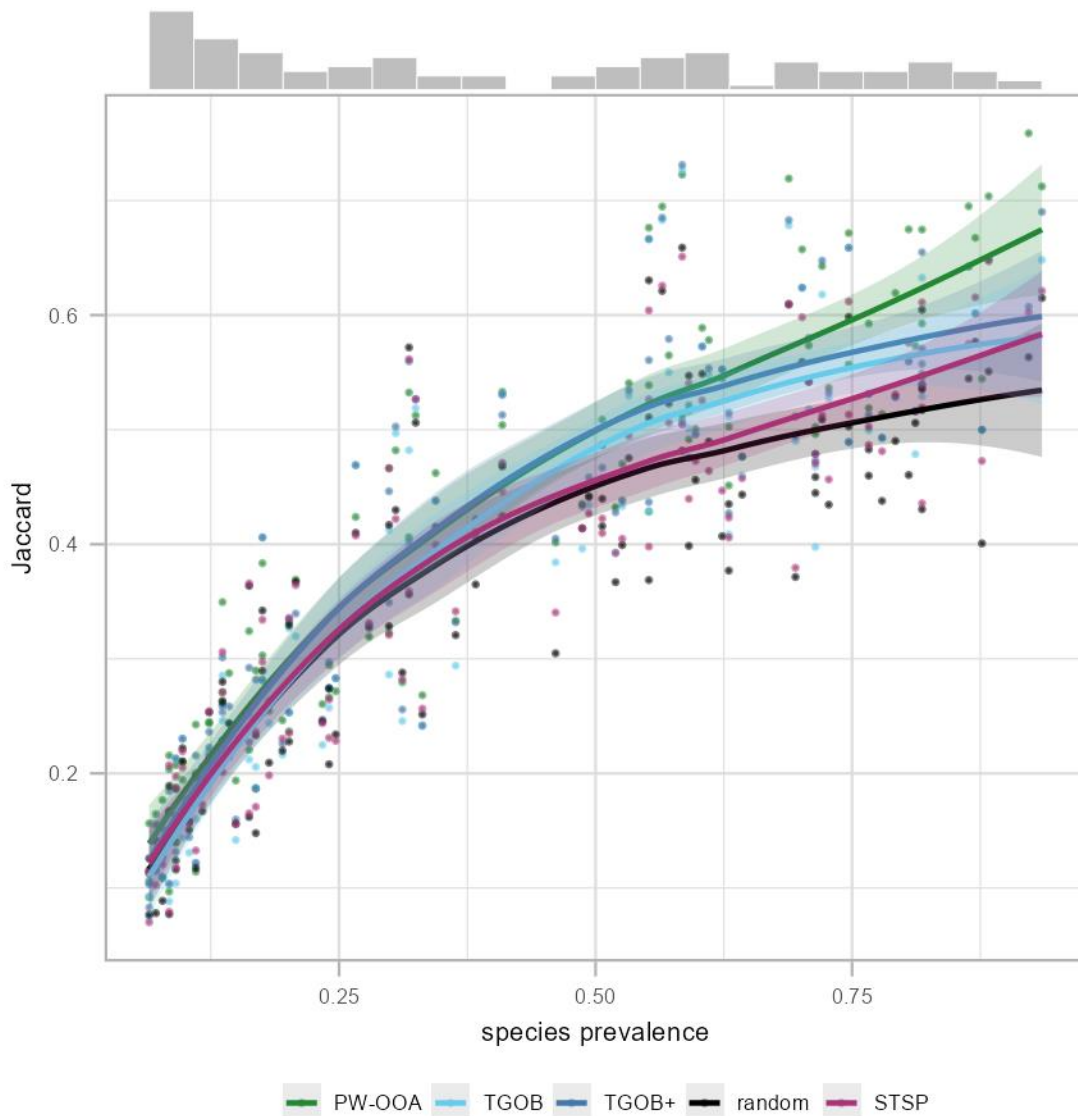
423

424 *Figure S31. The mean Jaccard (x-axis) and COR (y-axis) of the five presence and/or background*
 425 *simulation methods calculated from 108 bird species using Maxnet. The bars (around each model)*
 426 *represent standard errors that reflect the variation among all species. Spatial thinning of species*
 427 *presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 428 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

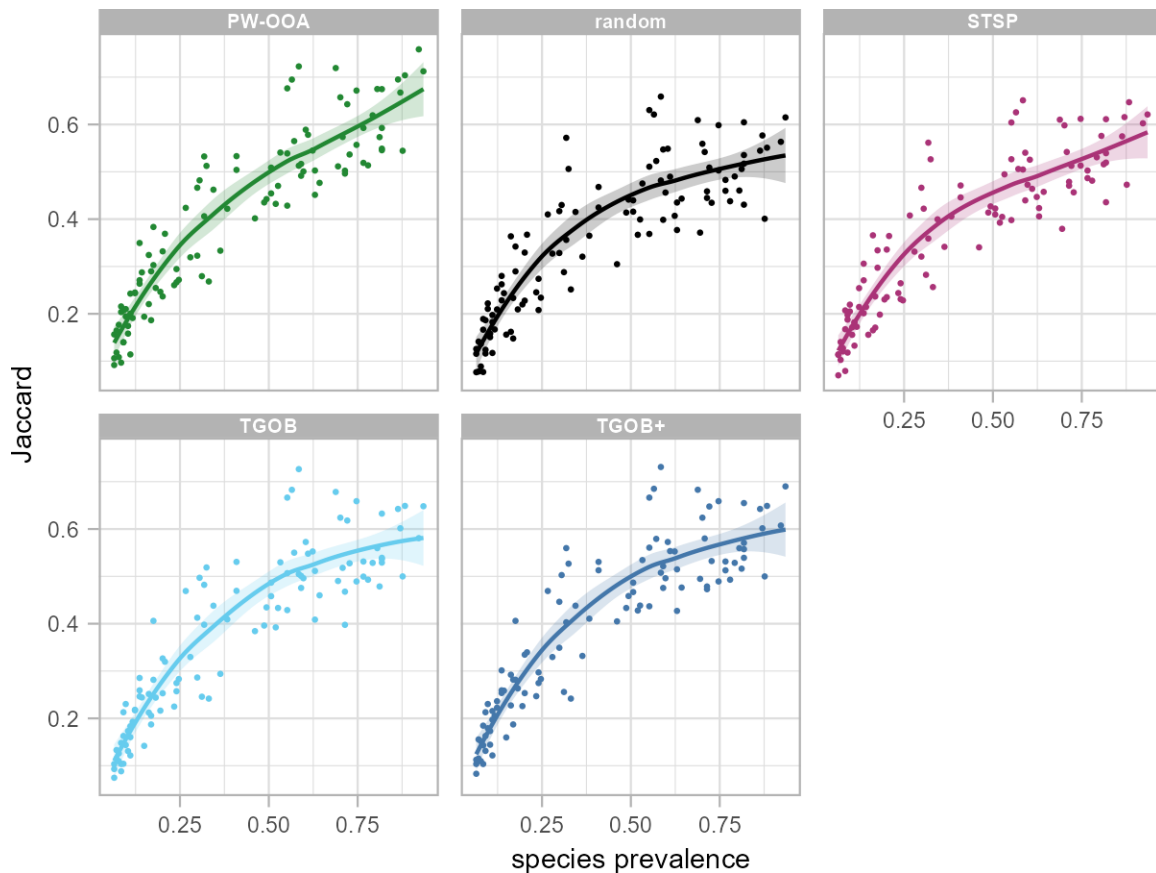
429



430
 431 *Figure S32. Comparison of SDM prediction performance of 108 bird species differing in species*
 432 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 433 *measured as Jaccard improvement over the random method (Jaccard (diff) = difference between the*
 434 *STSP, TGOB, TGOB+, or PW-OOA versus the random method) using Maxnet. Locally estimated*
 435 *scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using*
 436 *locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 437 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 438 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 439 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 440 *(PW-OOA).*
 441



442
 443 *Figure S33. Comparison of SDM prediction performance of 108 bird species differing in species*
 444 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 445 *measured as Jaccard using Maxnet. Locally estimated scatterplot smoothing (LOESS) regression lines, a*
 446 *non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown*
 447 *with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram*
 448 *showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group*
 449 *Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths)*
 450 *and presence-weighted observer-oriented approach (PW-OOA).*
 451



452
453
454
455
456
457
458
459
460

Figure S34. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as Jaccard using Maxnet. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).

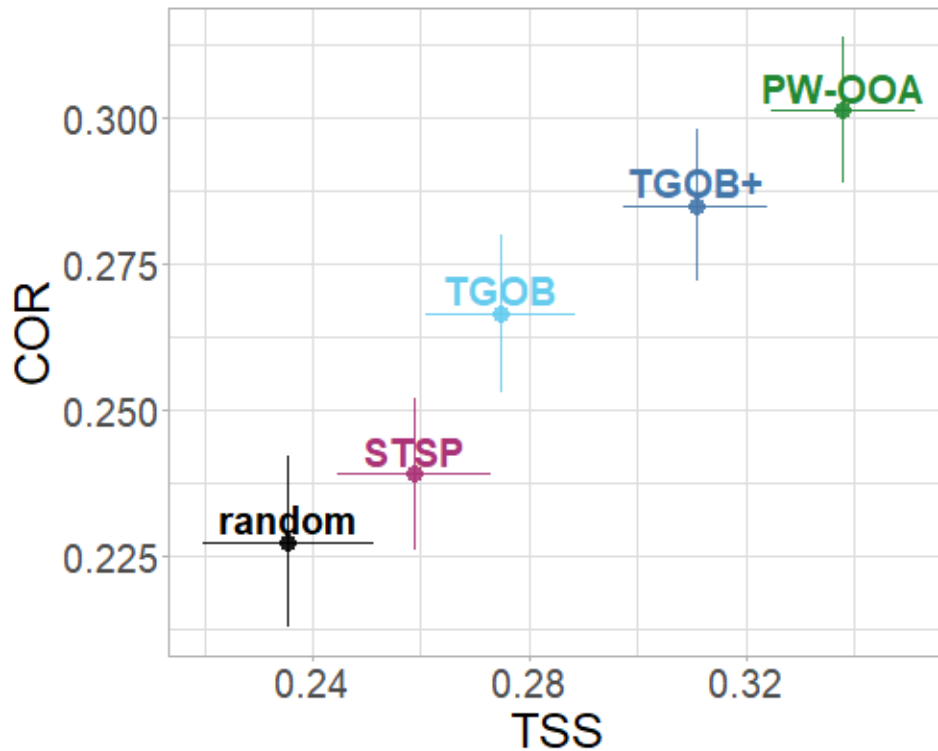
461 TSS

462

463 *Table S14. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using Maxnet in*
 464 *terms of prediction performance (TSS) for the 108 species of interest. The values represent the number*
 465 *of species for which the model in the row significantly outperformed the model in the column (Unpaired*
 466 *Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the mean (TSS) of*
 467 *the respective method across all species. Spatial thinning of species presences (STSP), Target Group*
 468 *Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths)*
 469 *and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	92	84	78	53	0.338
TGOB+	86	68	51	0.311	25
TGOB	65	50	0.275	0	12
STSP	44	0.259	39	19	13
random	0.236	20	31	13	8

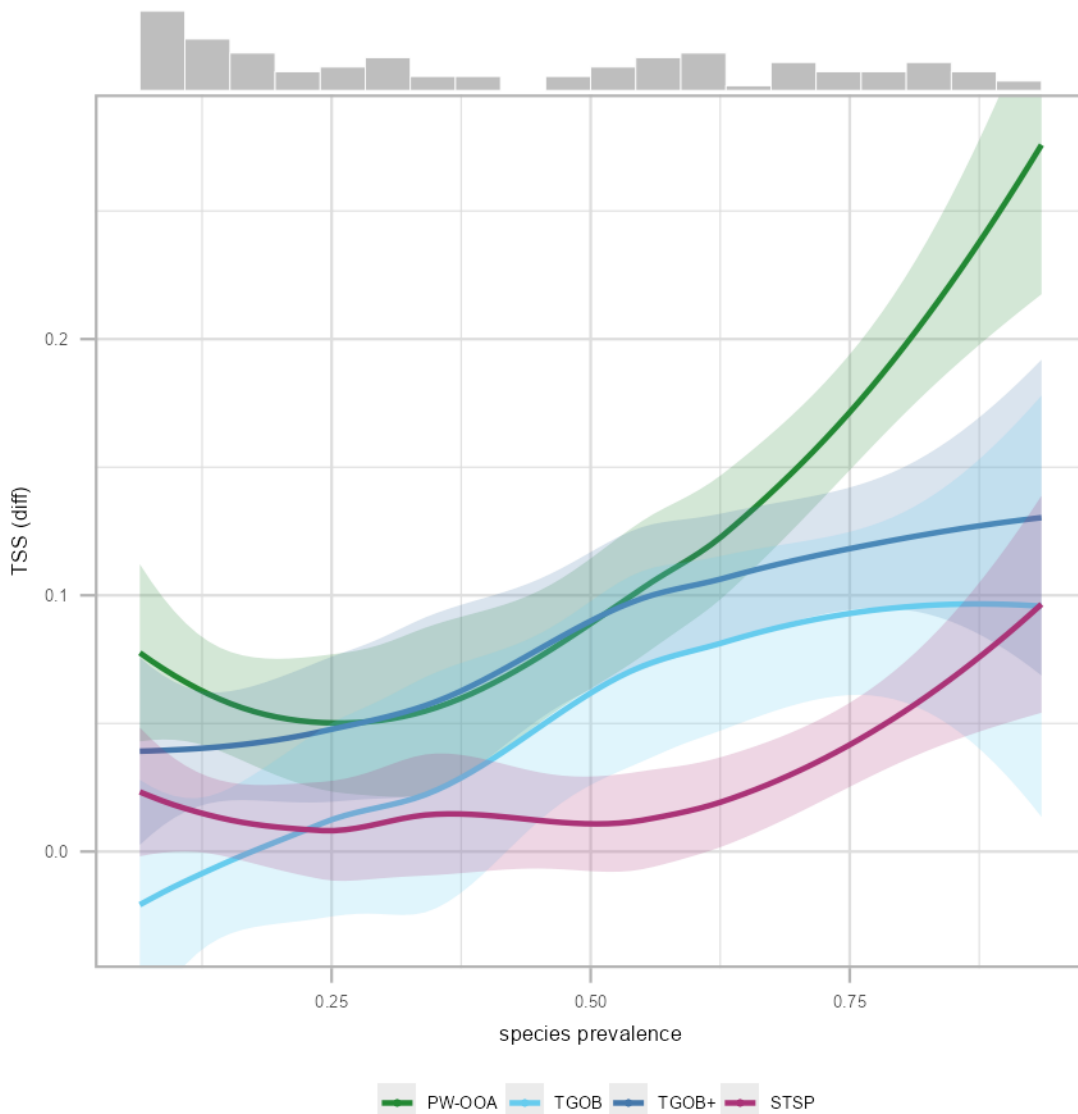
470



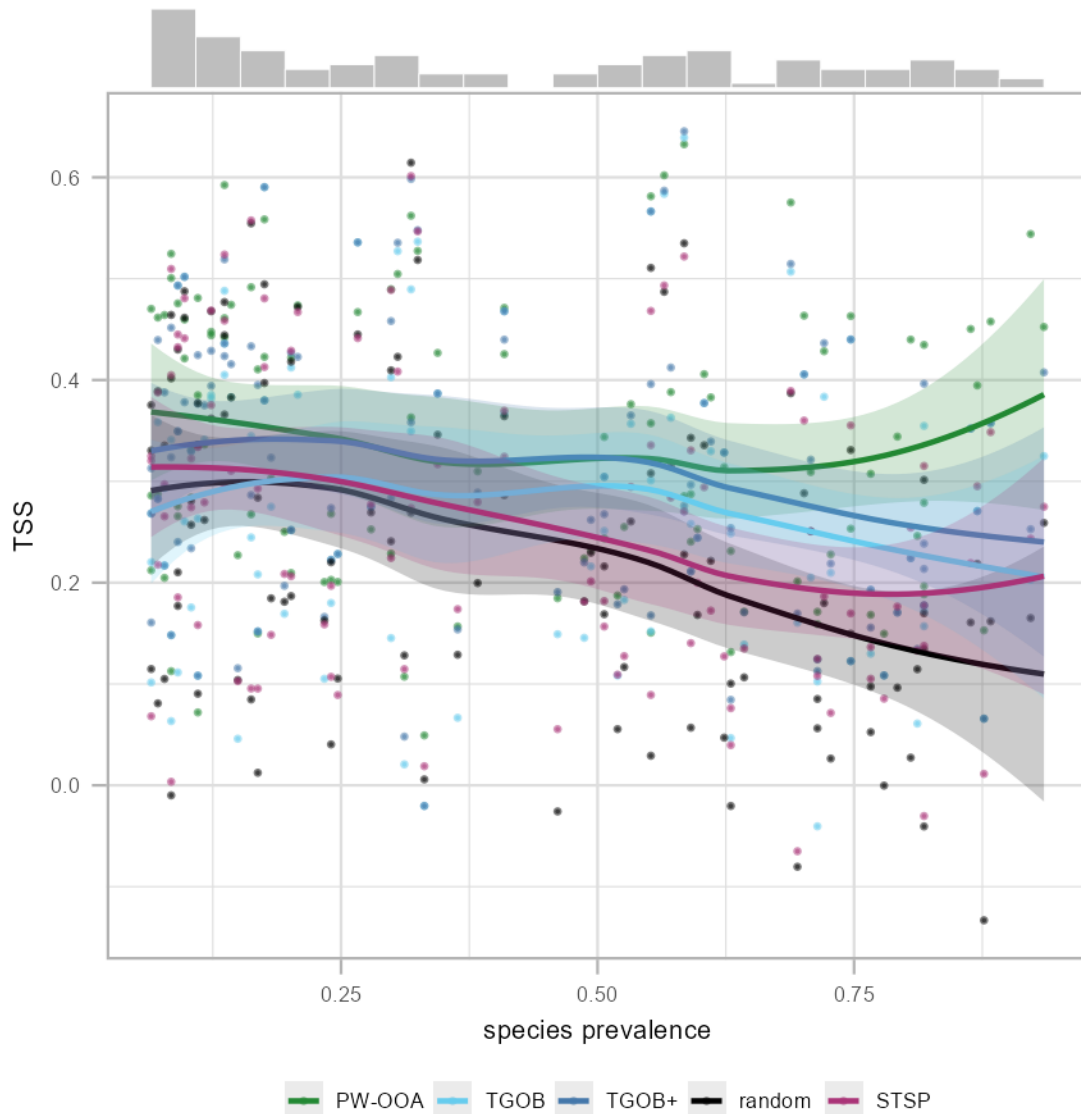
471

472 *Figure S35. The mean TSS (x-axis) and COR (y-axis) of the five presence and/or background simulation*
 473 *methods calculated from 108 bird species using Maxnet. The bars (around each model) represent*
 474 *standard errors that reflect the variation among all species. Spatial thinning of species presences (STSP),*
 475 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 476 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

477

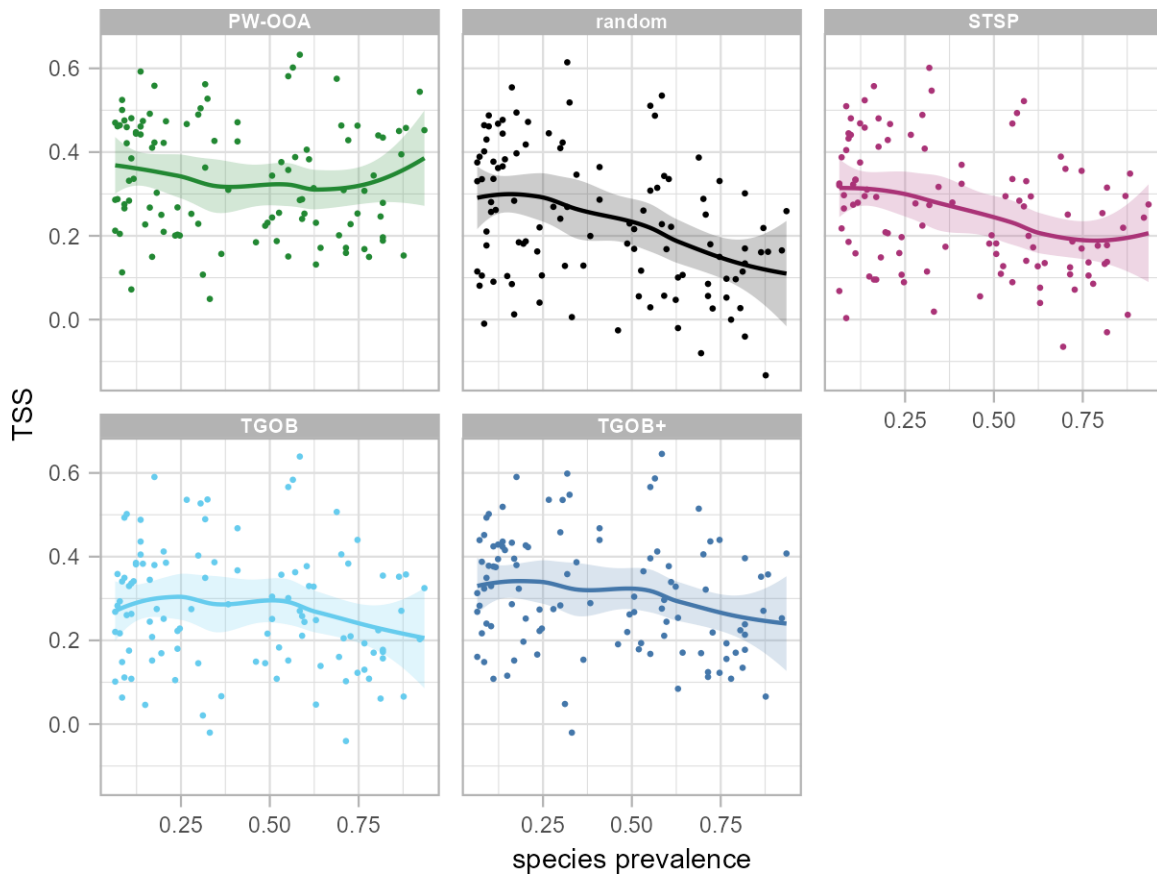


478
 479 *Figure S36. Comparison of SDM prediction performance of 108 bird species differing in species*
 480 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 481 *measured as TSS improvement over the random method ($TSS (diff) = \text{difference between the STSP,}$*
 482 *TGOB, TGOB+, or PW-OOA versus the random method) using Maxnet. Locally estimated scatterplot*
 483 *smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally*
 484 *weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 485 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 486 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 487 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 488 *(PW-OOA).*
 489



490
491
492
493
494
495
496
497
498
499

Figure S37. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as TSS using Maxnet. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



500
501
502
503
504
505
506
507
508

Figure S38. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as TSS using Maxnet. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).

509 Random forest

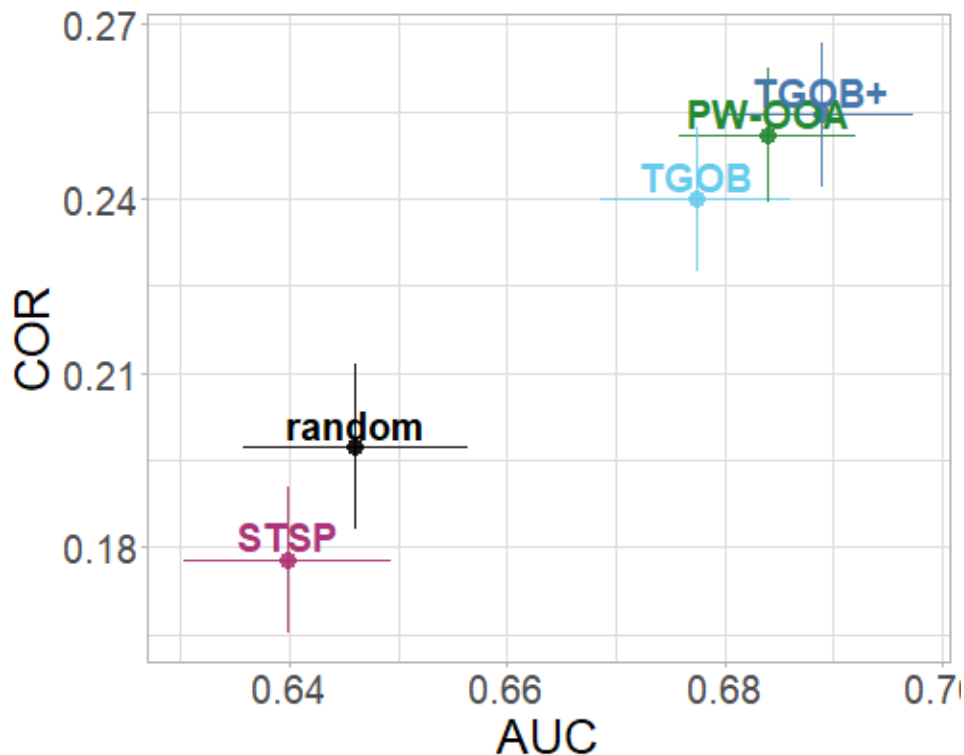
510 Random Forest (default settings) failed to deliver a gradual improvement in average prediction
 511 performance for the progressions: random to STSP, and TGOB+ to PW-OOA. This outcome contrasted
 512 with the trends observed for MaxNet and GLM. Notably, STSP's performance was even lower than the
 513 'baseline' random method, and PW-OOA's performance was lower than TGOB+.

514 AUC

515
 516 *Table S15. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using Random*
 517 *forest in terms of prediction performance (AUC) for the 108 species of interest. The values represent the*
 518 *number of species for which the model in the row significantly outperformed the model in the column*
 519 *(Unpaired Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the*
 520 *mean (AUC) of the respective method across all species. Spatial thinning of species presences (STSP),*
 521 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 522 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

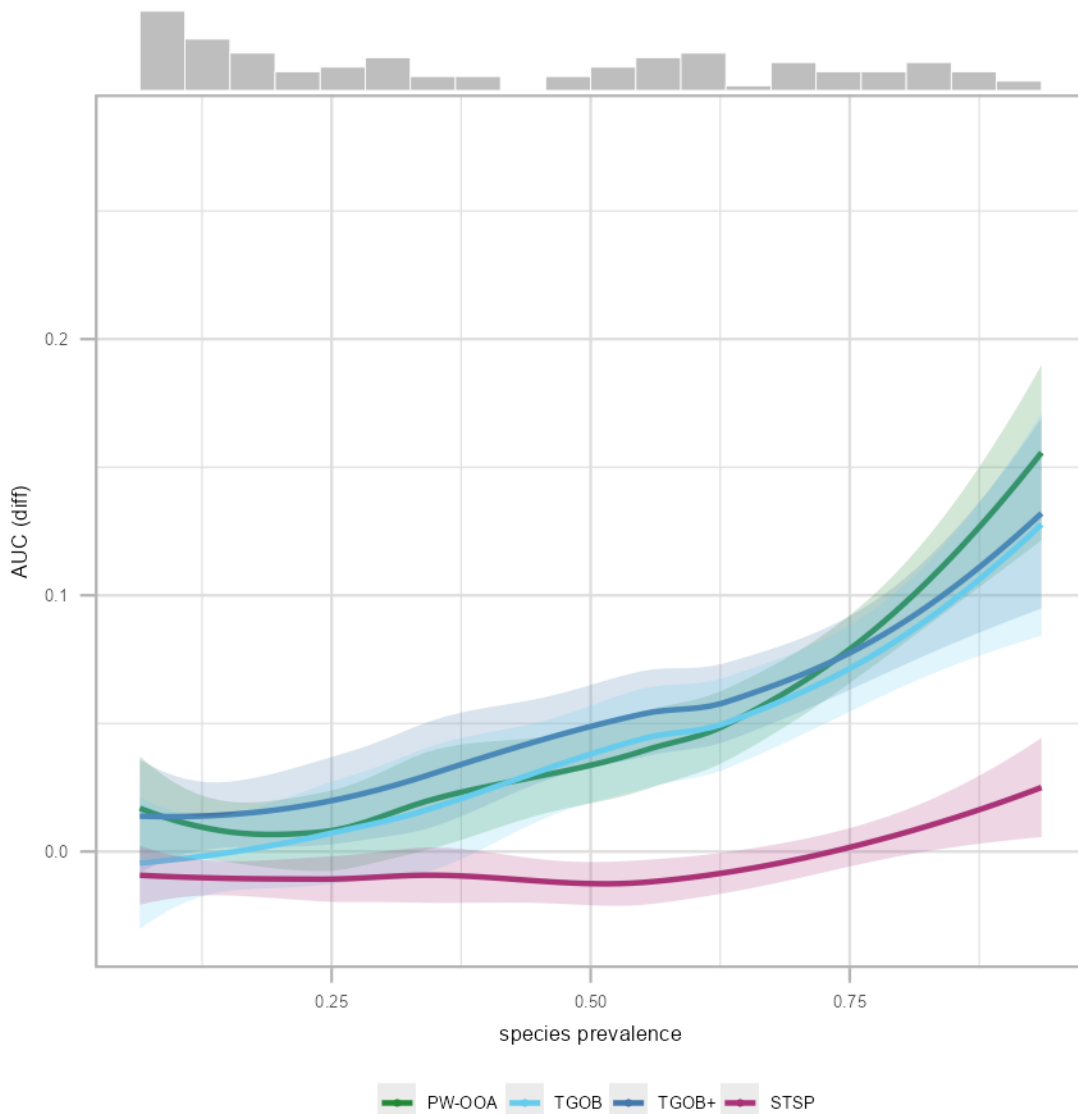
	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	72	84	40	25	0.684
TGOB+	79	88	38	0.689	48
TGOB	68	78	0.677	0	36
STSP	18	0.640	17	10	10
random	0.646	56	25	14	10

523

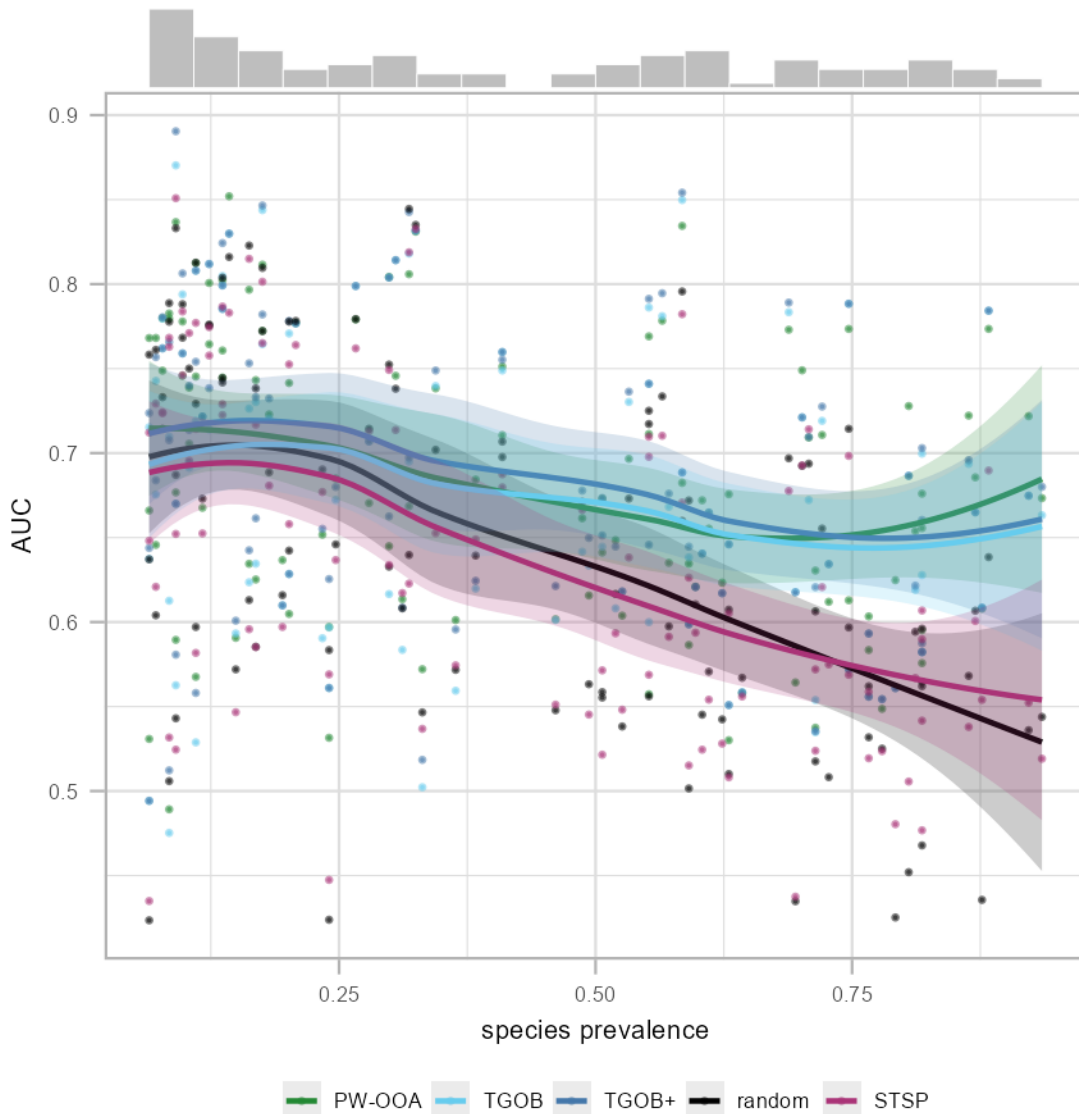


524
 525 *Figure S39. The mean AUC (x-axis) and COR (y-axis) of the five presence and/or background simulation*
 526 *methods calculated from 108 bird species using Random forest. The bars (around each model) represent*

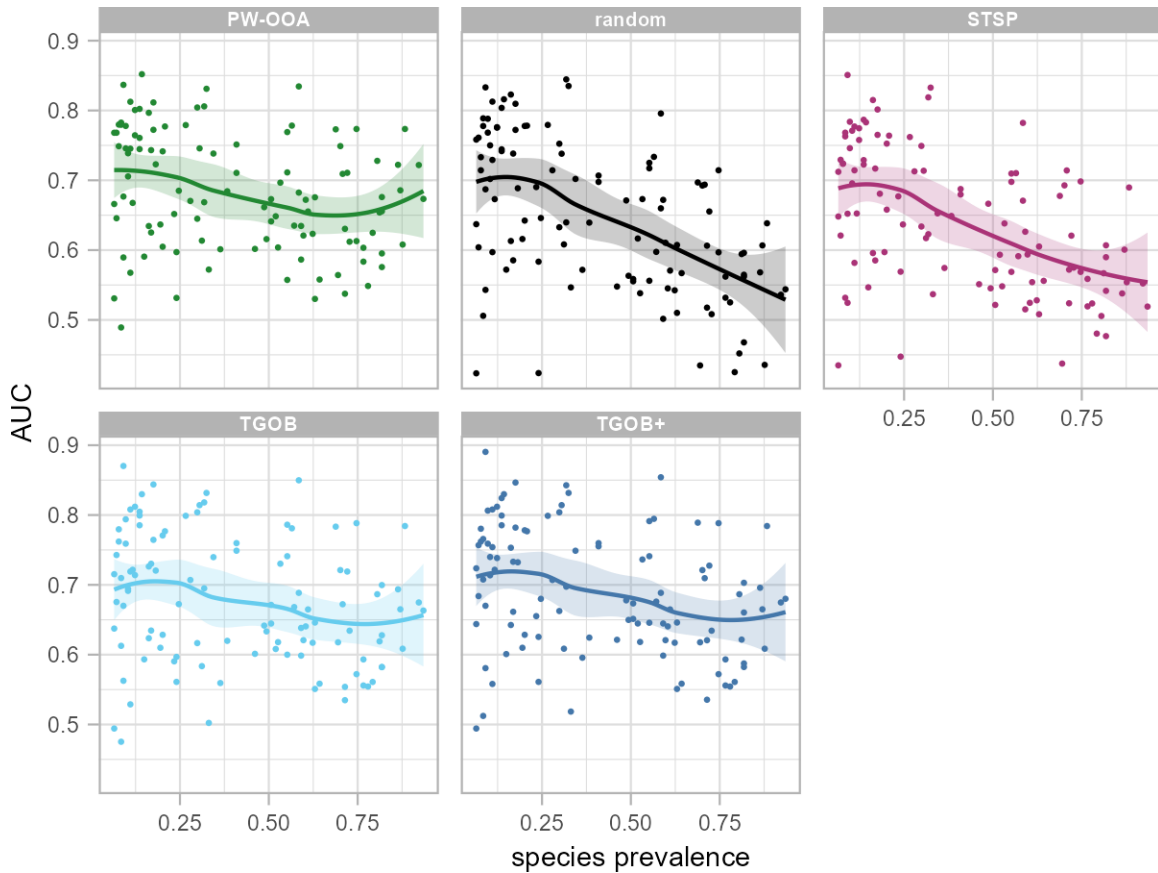
527 *standard errors that reflect the variation among all species. Spatial thinning of species presences (STSP),*
 528 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 529 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 530



531
532 *Figure S40. Comparison of SDM prediction performance of 108 bird species differing in species*
533 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
534 *measured as AUC improvement over the random method (AUC (diff) = difference between the STSP,*
535 *TGOB, TGOB+, or PW-OOA versus the random method) using Random forest. Locally estimated*
536 *scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using*
537 *locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
538 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
539 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
540 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
541 *(PW-OOA).*
542



543
 544 *Figure S41. Comparison of SDM prediction performance of 108 bird species differing in species*
 545 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 546 *measured as AUC using Random forest. Locally estimated scatterplot smoothing (LOESS) regression*
 547 *lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are*
 548 *shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a*
 549 *histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP),*
 550 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 551 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 552



553
 554 *Figure S42. Comparison of SDM prediction performance of 108 bird species differing in species*
 555 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 556 *measured as AUC using Random forest. Locally estimated scatterplot smoothing (LOESS) regression*
 557 *lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are*
 558 *shown with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of*
 559 *species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 560 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 561

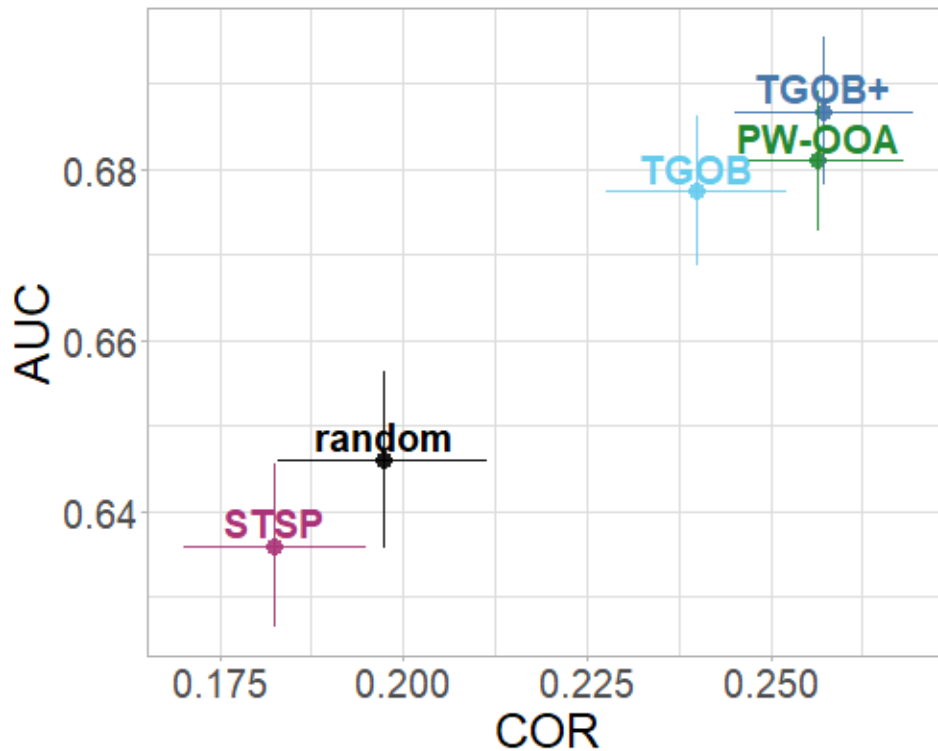
562 COR

563

564 *Table S16. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using Random*
 565 *forest in terms of prediction performance (COR) for the 108 species of interest. The values represent the*
 566 *number of species for which the model in the row significantly outperformed the model in the column*
 567 *(Unpaired Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the*
 568 *mean (COR) of the respective method across all species. Spatial thinning of species presences (STSP),*
 569 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 570 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	73	87	50	33	0.257
TGOB+	83	87	34	0.257	35
TGOB	68	76	0.240	0	24
STSP	17	0.183	19	12	9
random	0.197	59	23	16	14

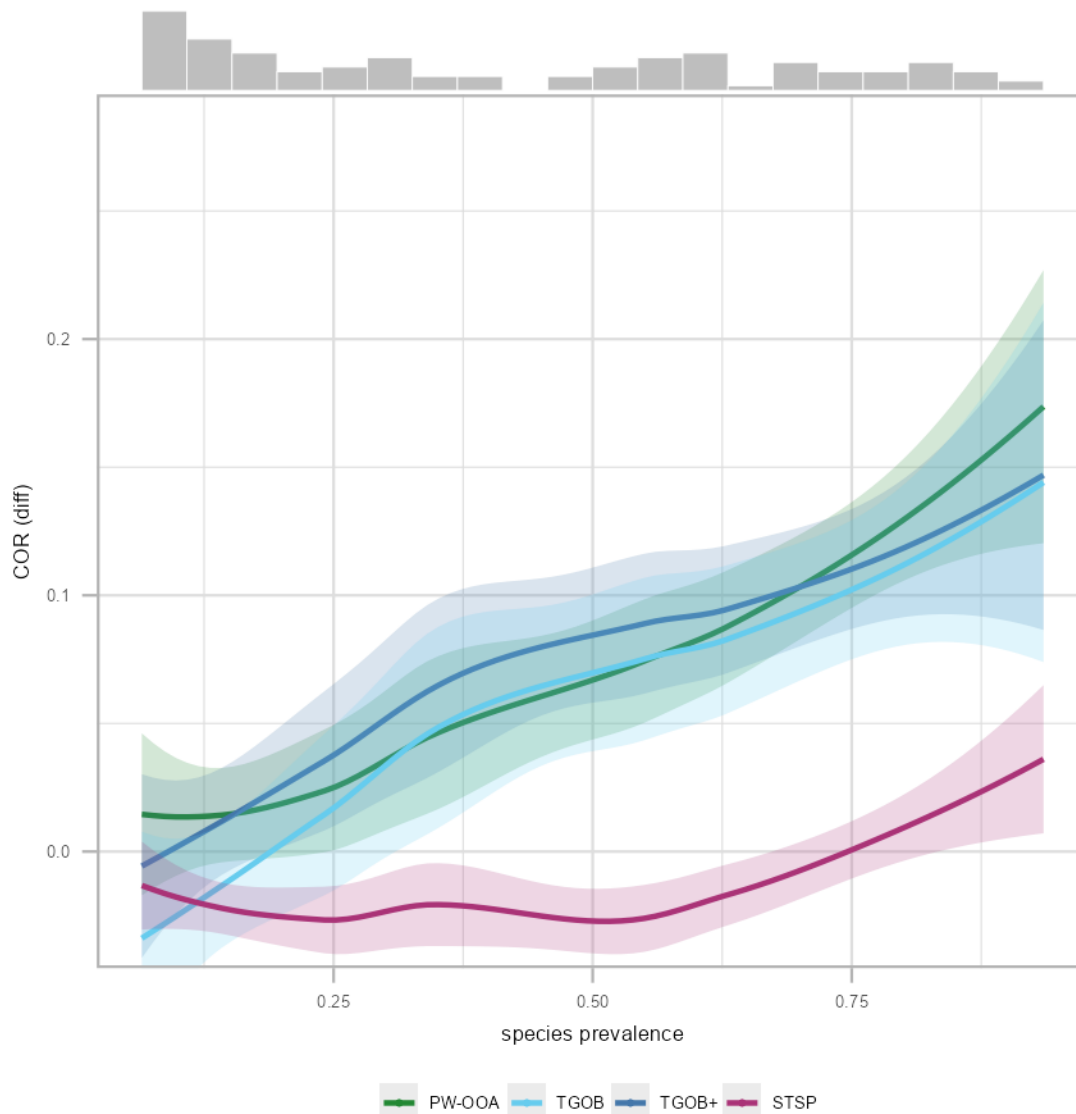
571



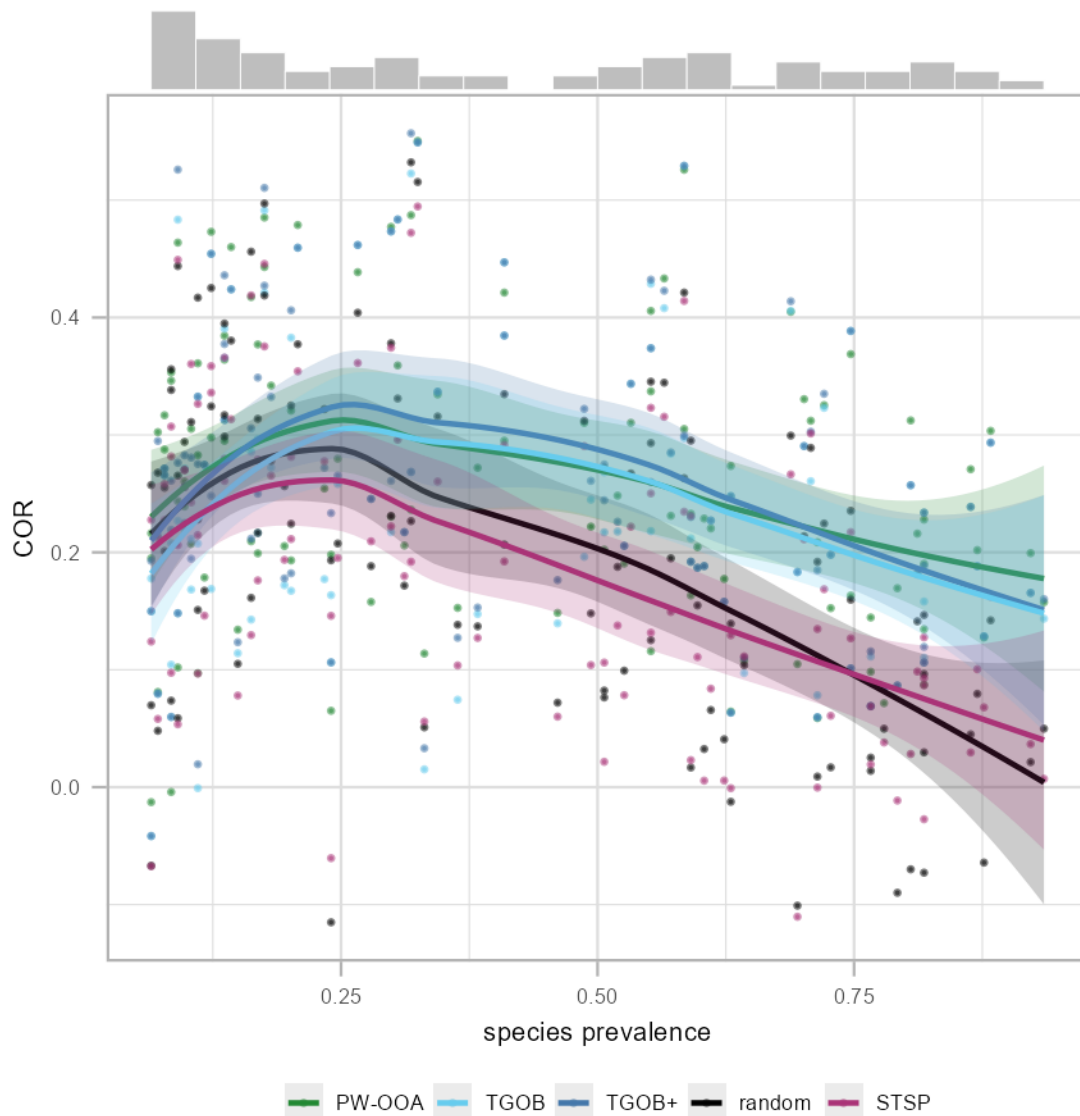
572

573 *Figure S43. The mean COR (x-axis) and AUC (y-axis) of the five presence and/or background simulation*
 574 *methods calculated from 108 bird species using Random forest. The bars (around each model) represent*
 575 *standard errors that reflect the variation among all species. Spatial thinning of species presences (STSP),*
 576 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 577 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

578

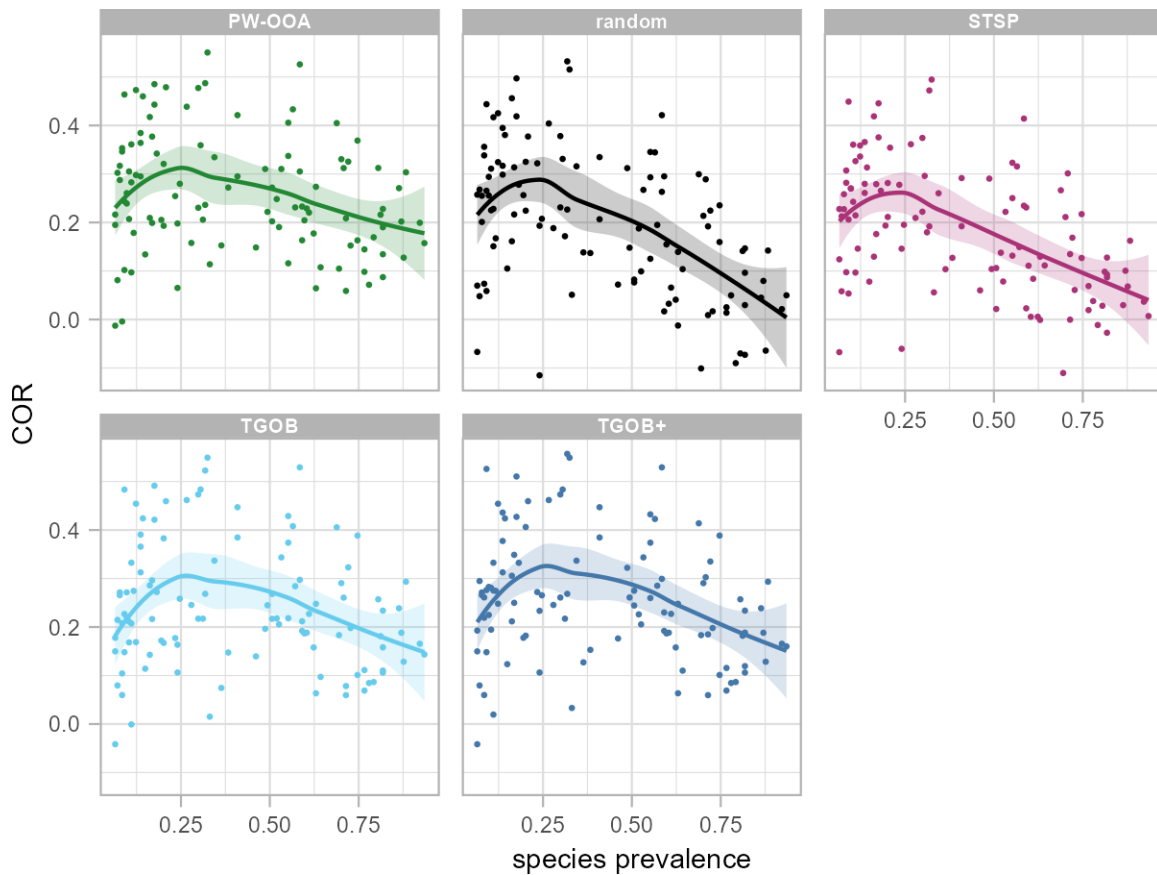


579
 580 *Figure S44. Comparison of SDM prediction performance of 108 bird species differing in species*
 581 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 582 *measured as COR improvement over the random method (COR (diff) = difference between the STSP,*
 583 *TGOB, TGOB+, or PW-OOA versus the random method) using Random forest. Locally estimated*
 584 *scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using*
 585 *locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 586 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 587 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 588 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 589 *(PW-OOA).*
 590



591
592
593
594
595
596
597
598
599
600

Figure S45. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as COR using Random forest. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).



601
 602 *Figure S46. Comparison of SDM prediction performance of 108 bird species differing in species*
 603 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 604 *measured as COR using Random forest. Locally estimated scatterplot smoothing (LOESS) regression*
 605 *lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are*
 606 *shown with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of*
 607 *species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 608 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 609

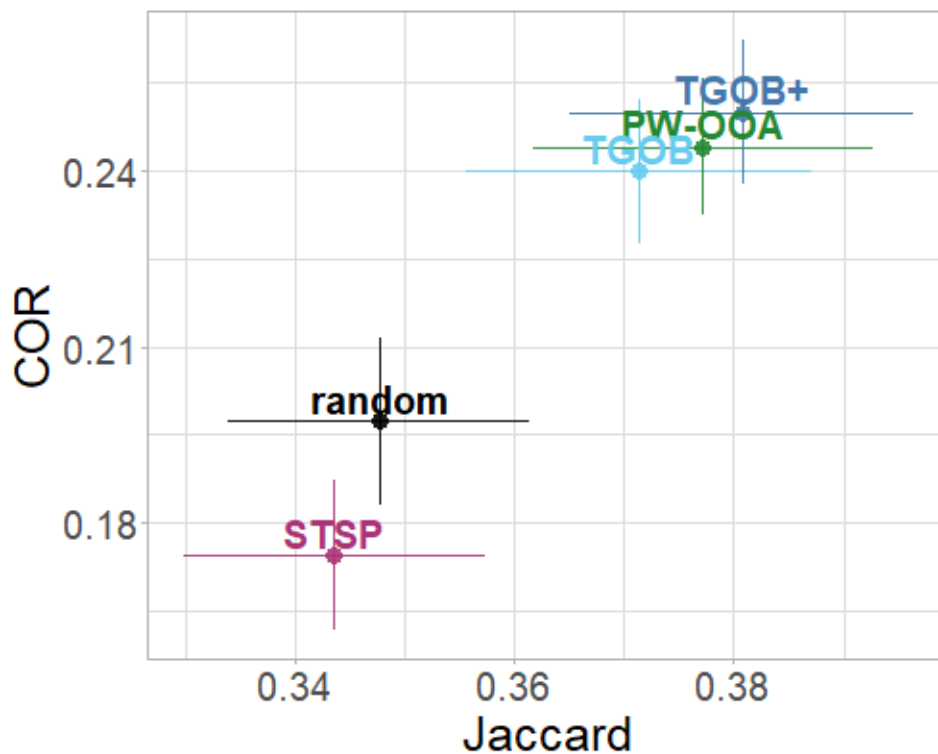
610 Jaccard

611

612 *Table S17. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using Random*
 613 *forest in terms of prediction performance (Jaccard) for the 108 species of interest. The values represent*
 614 *the number of species for which the model in the row significantly outperformed the model in the column*
 615 *(Unpaired Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the*
 616 *mean (Jaccard) of the respective method across all species. Spatial thinning of species presences*
 617 *(STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel*
 618 *smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	73	80	47	24	0.377
TGOB+	72	82	37	0.381	39
TGOB	59	73	0.371	0	29
STSP	14	0.343	21	10	8
random	0.348	44	22	14	11

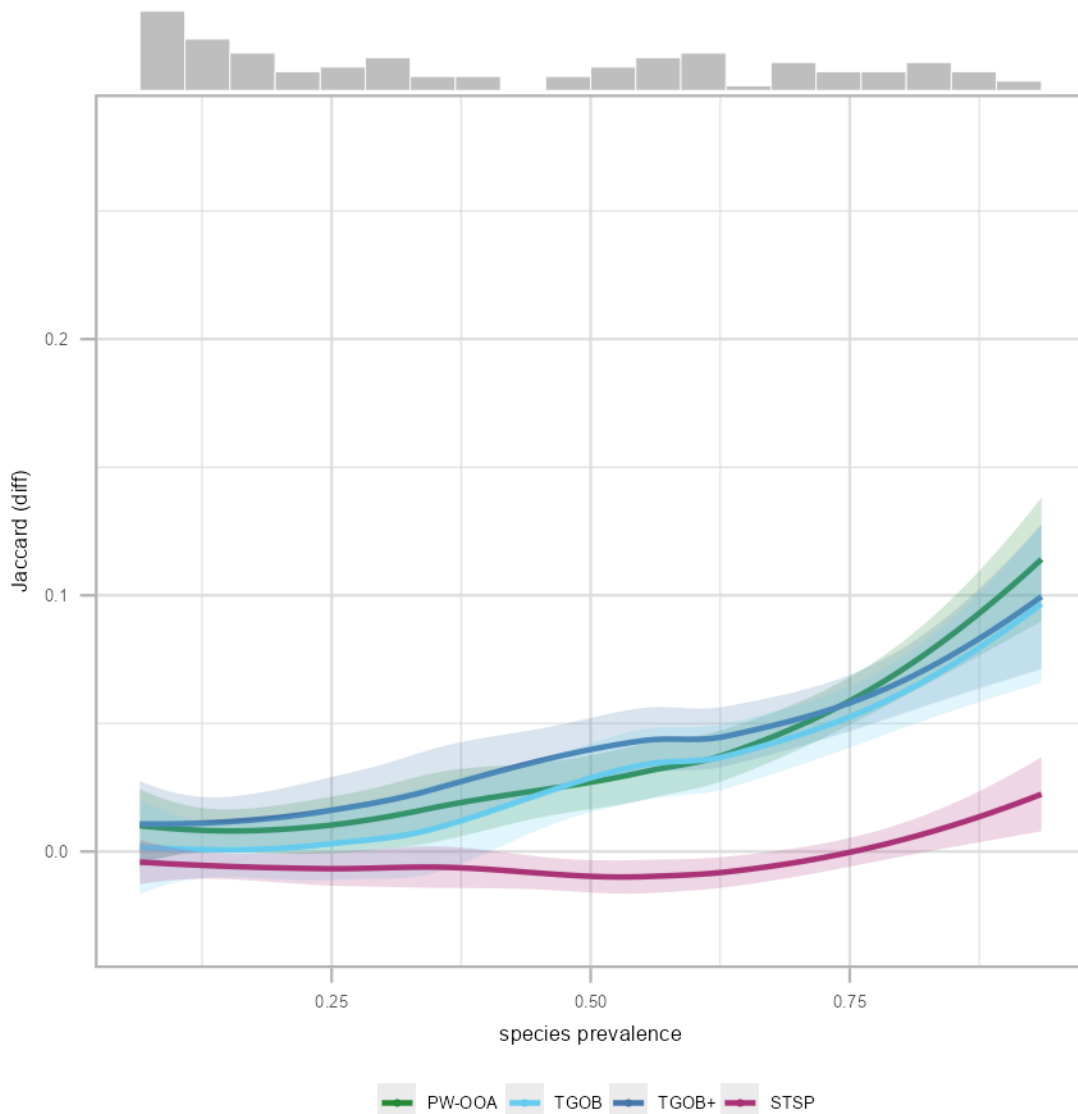
619



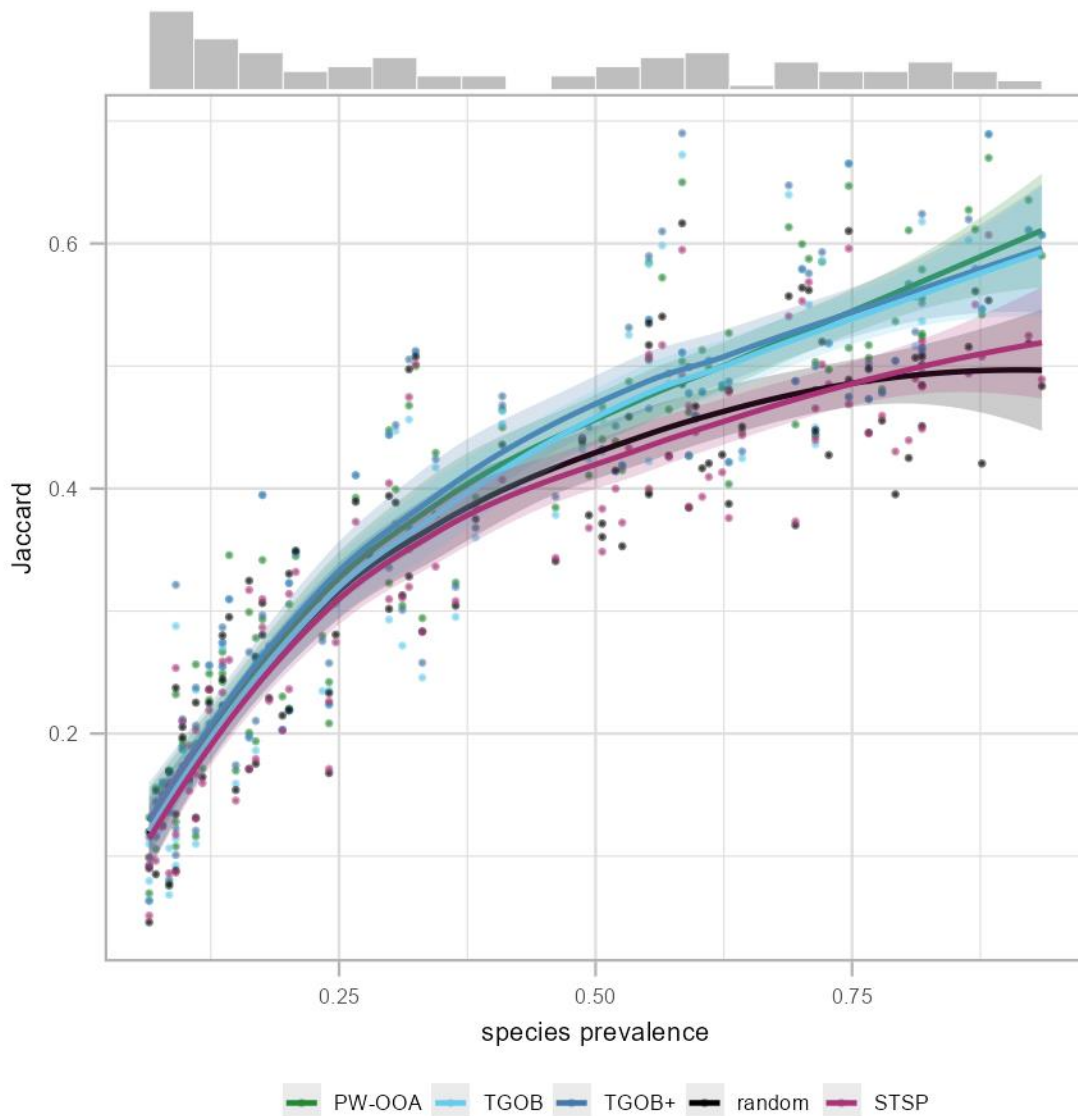
620

621 *Figure S47. The mean Jaccard (x-axis) and COR (y-axis) of the five presence and/or background*
 622 *simulation methods calculated from 108 bird species using Random forest. The bars (around each model)*
 623 *represent standard errors that reflect the variation among all species. Spatial thinning of species*
 624 *presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 625 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

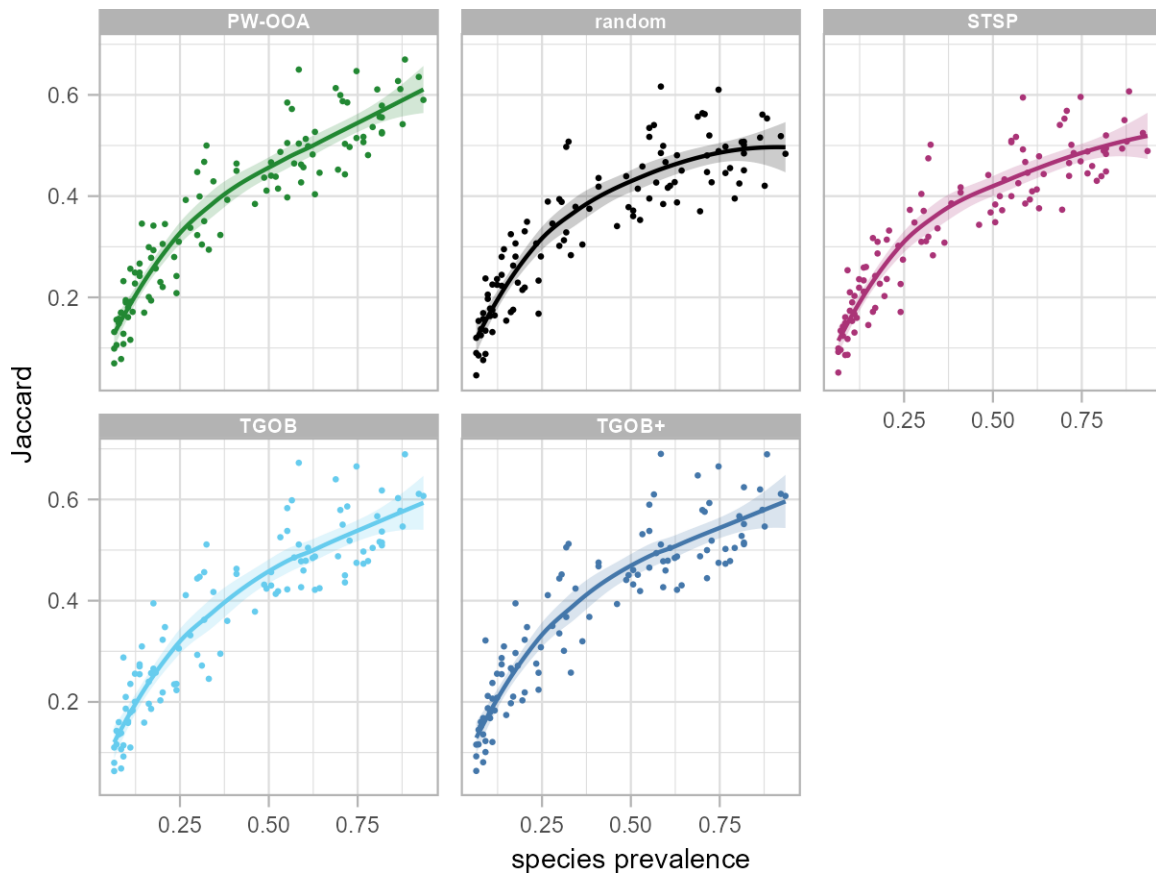
626



627
 628 *Figure S48. Comparison of SDM prediction performance of 108 bird species differing in species*
 629 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 630 *measured as Jaccard improvement over the random method (Jaccard (diff) = difference between the*
 631 *STSP, TGOB, TGOB+, or PW-OOA versus the random method) using Random forest. Locally estimated*
 632 *scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using*
 633 *locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 634 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 635 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 636 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 637 *(PW-OOA).*
 638



639
 640 *Figure S49. Comparison of SDM prediction performance of 108 bird species differing in species*
 641 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 642 *measured as Jaccard using Random forest. Locally estimated scatterplot smoothing (LOESS) regression*
 643 *lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are*
 644 *shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a*
 645 *histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP),*
 646 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 647 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 648



649
 650 *Figure S50. Comparison of SDM prediction performance of 108 bird species differing in species*
 651 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 652 *measured as Jaccard using Random forest. Locally estimated scatterplot smoothing (LOESS) regression*
 653 *lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are*
 654 *shown with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of*
 655 *species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 656 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 657

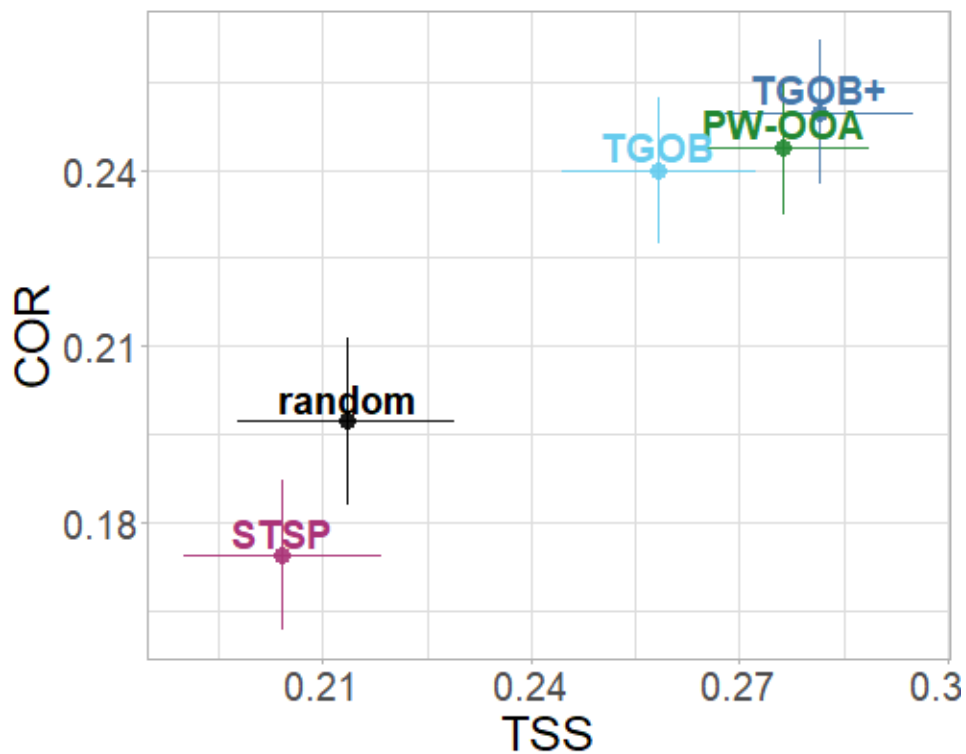
658 TSS

659

660 *Table S18. Comparison of all five methods (random, STSP, TGOB, TGOB+, PW-OOA) using Random*
 661 *forest in terms of prediction performance (TSS) for the 108 species of interest. The values represent the*
 662 *number of species for which the model in the row significantly outperformed the model in the column*
 663 *(Unpaired Two-Samples Wilcoxon Test, p-value < 0.05). The bold numbers on the diagonal show the*
 664 *mean (TSS) of the respective method across all species. Spatial thinning of species presences (STSP),*
 665 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 666 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

	random	STSP	TGOB	TGOB+	PW-OOA
PW-OOA	73	80	47	24	0.276
TGOB+	72	82	36	0.281	39
TGOB	59	73	0.258	0	29
STSP	14	0.204	21	11	8
random	0.213	44	22	14	11

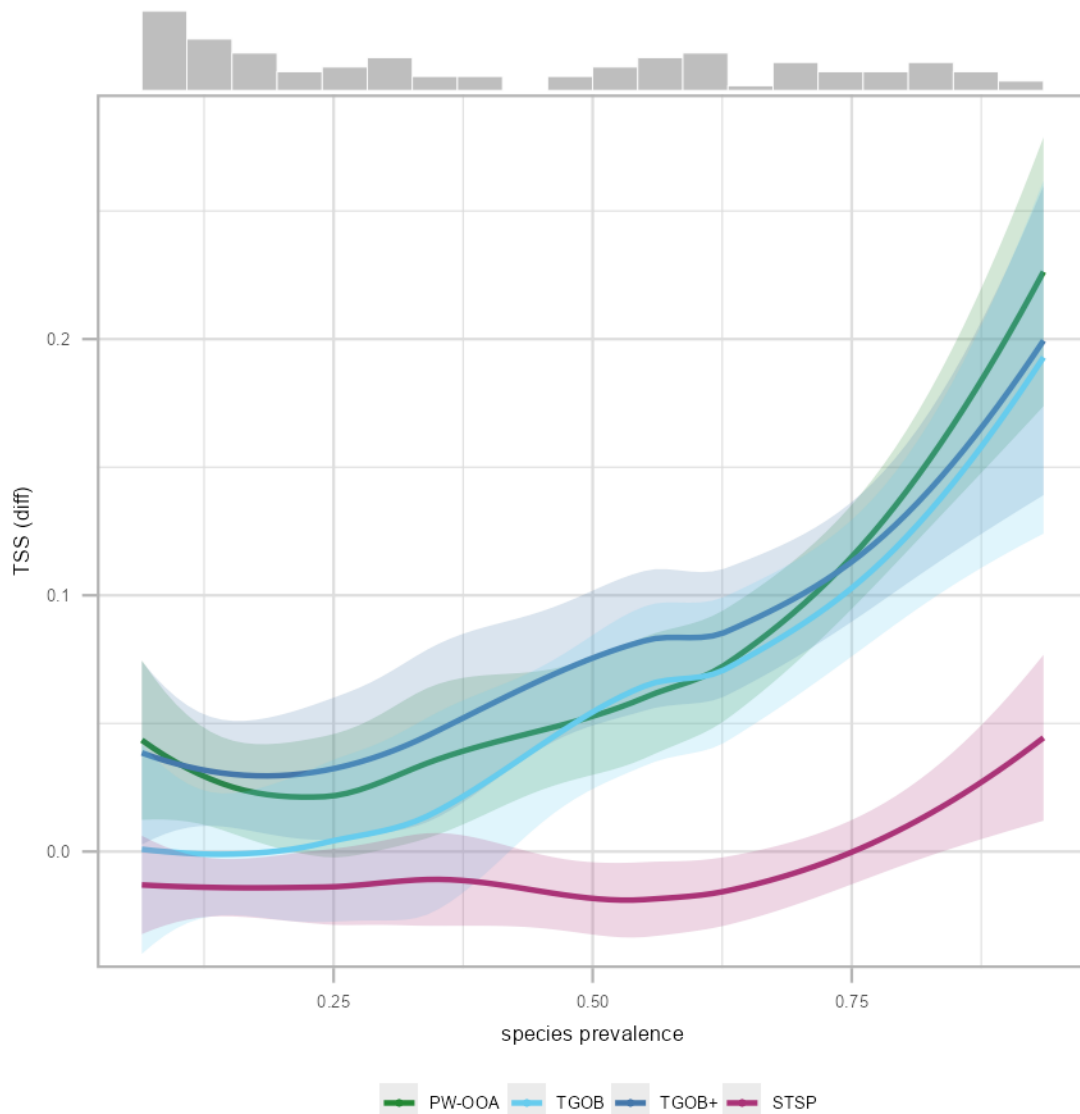
667



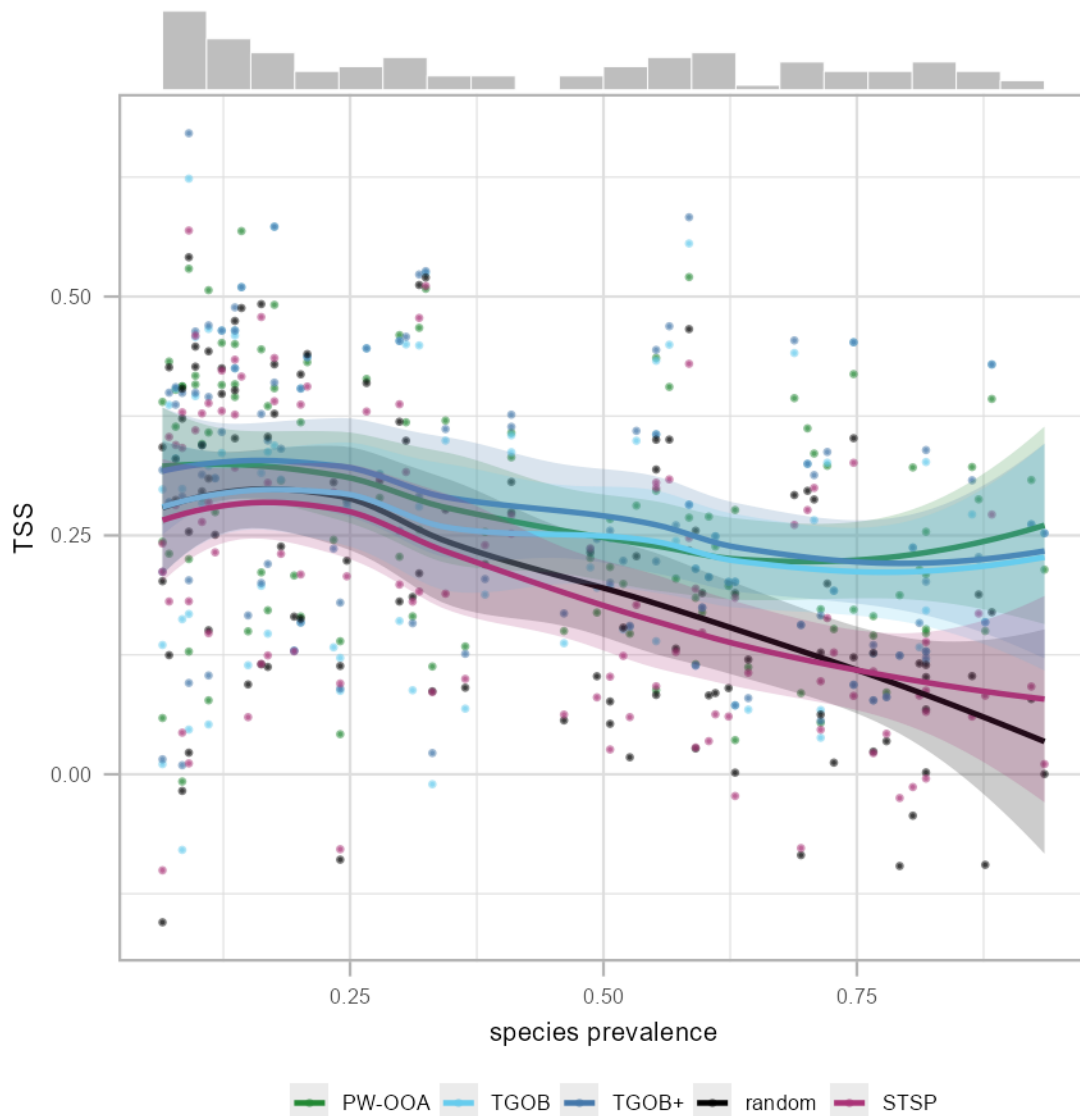
668

669 *Figure S51. The mean TSS (x-axis) and COR (y-axis) of the five presence and/or background simulation*
 670 *methods calculated from 108 bird species using Random forest. The bars (around each model) represent*
 671 *standard errors that reflect the variation among all species. Spatial thinning of species presences (STSP),*
 672 *Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing*
 673 *bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*

674

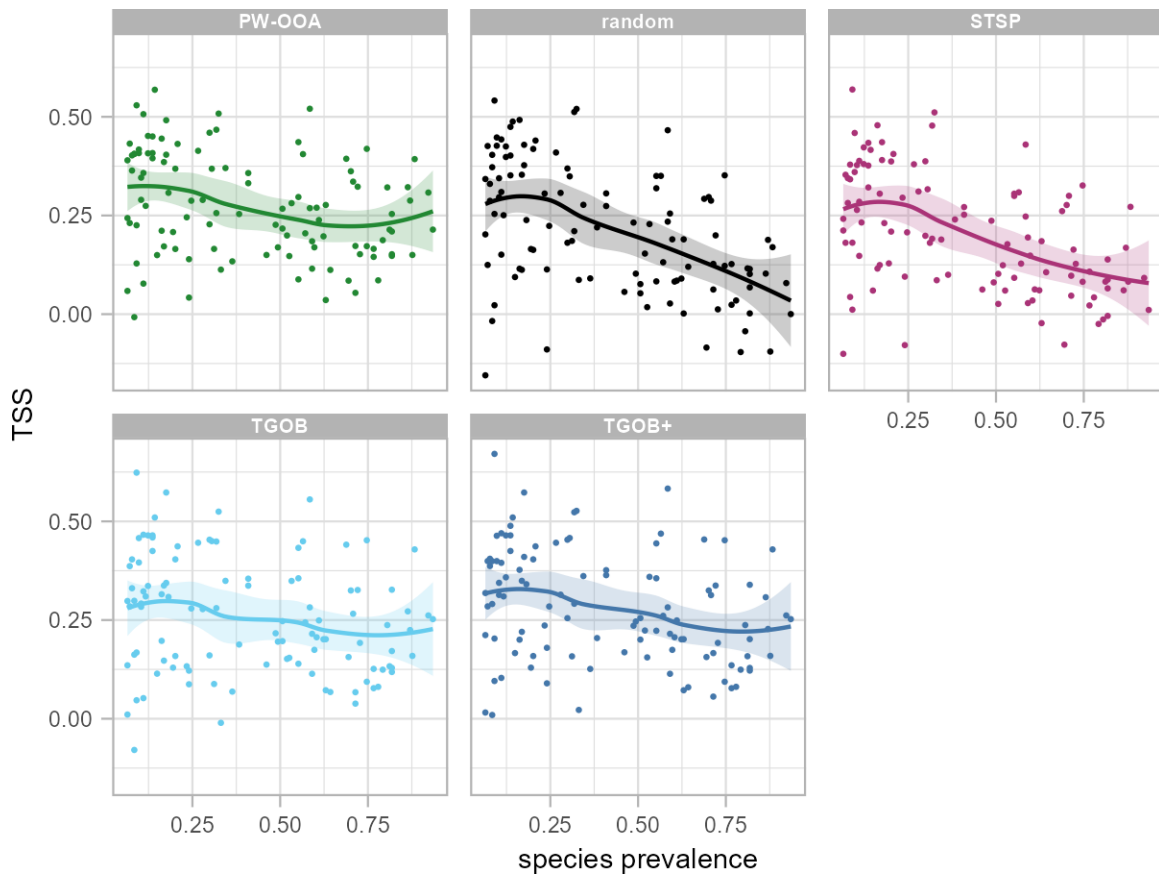


675
 676 *Figure S52. Comparison of SDM prediction performance of 108 bird species differing in species*
 677 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 678 *measured as TSS improvement over the random method (TSS (diff) = difference between the STSP,*
 679 *TGOB, TGOB+, or PW-OOA versus the random method) using Random forest. Locally estimated*
 680 *scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using*
 681 *locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS*
 682 *regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial*
 683 *thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned*
 684 *up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach*
 685 *(PW-OOA).*
 686



687
688
689
690
691
692
693
694
695
696

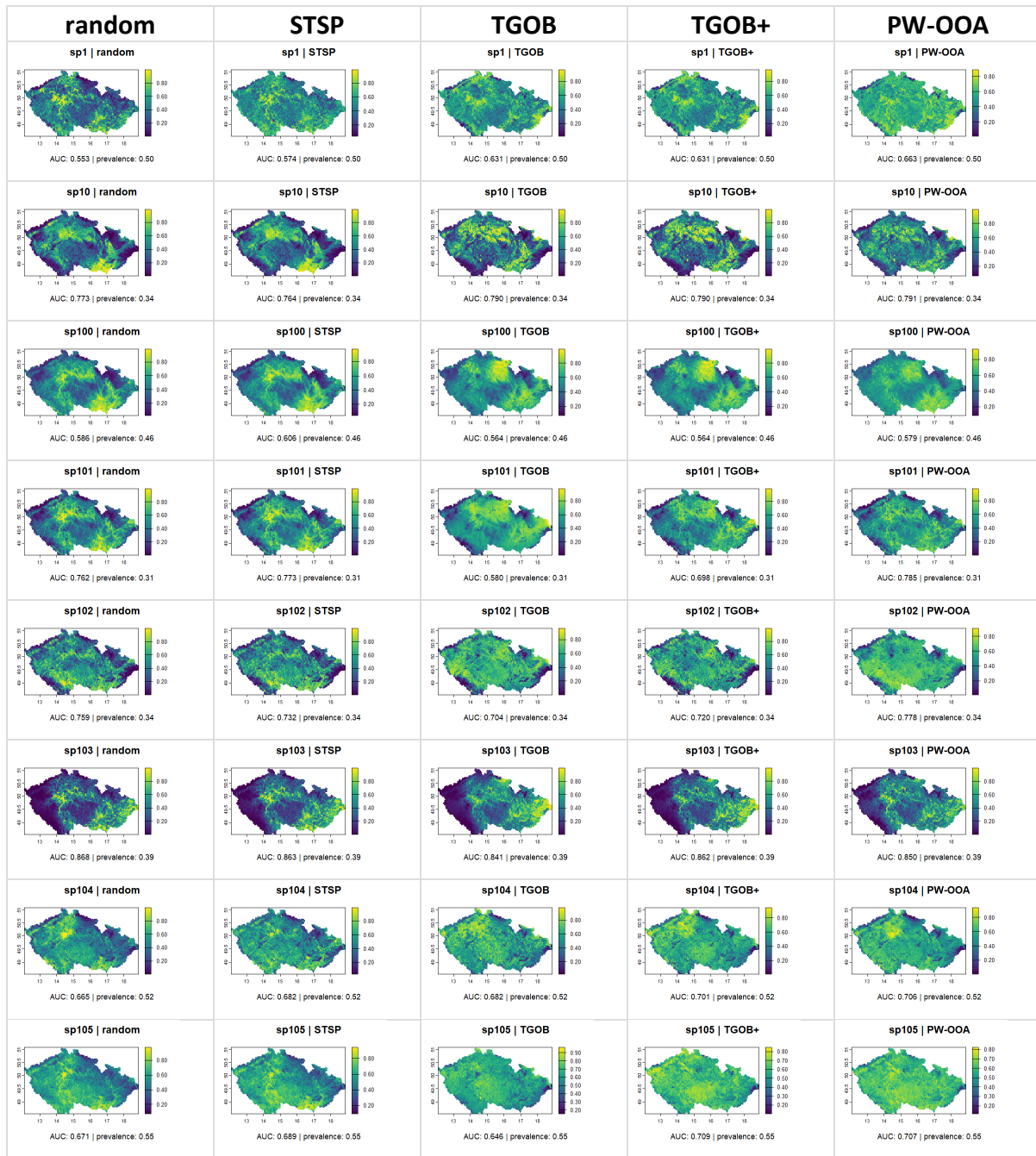
Figure S53. Comparison of SDM prediction performance of 108 bird species differing in species prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels) measured as TSS using Random forest. Locally estimated scatterplot smoothing (LOESS) regression lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are shown with 95% confidence bands. Compares all the LOESS regression lines in one overlay, with a histogram showing the count of species across 20 bins. Spatial thinning of species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).

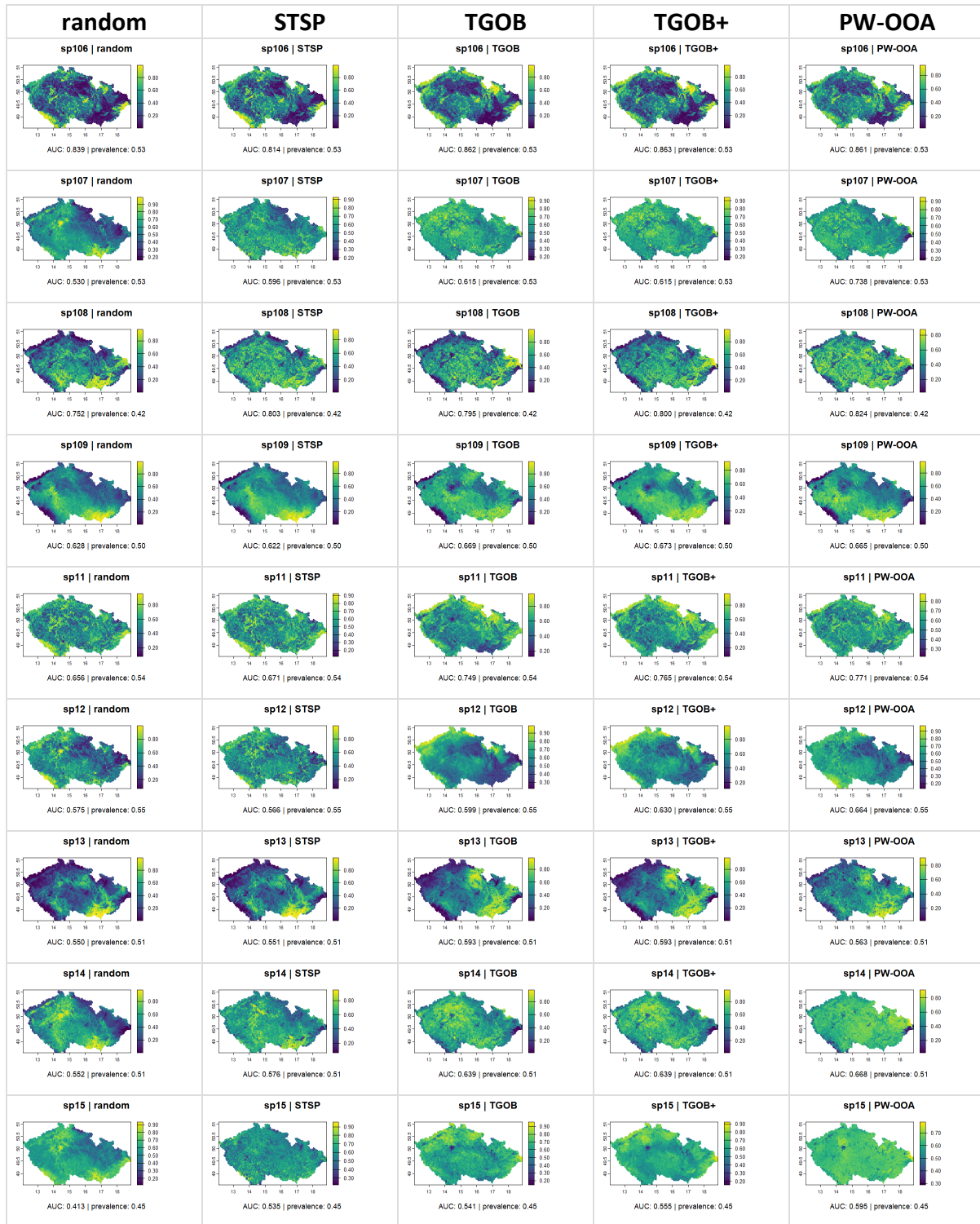


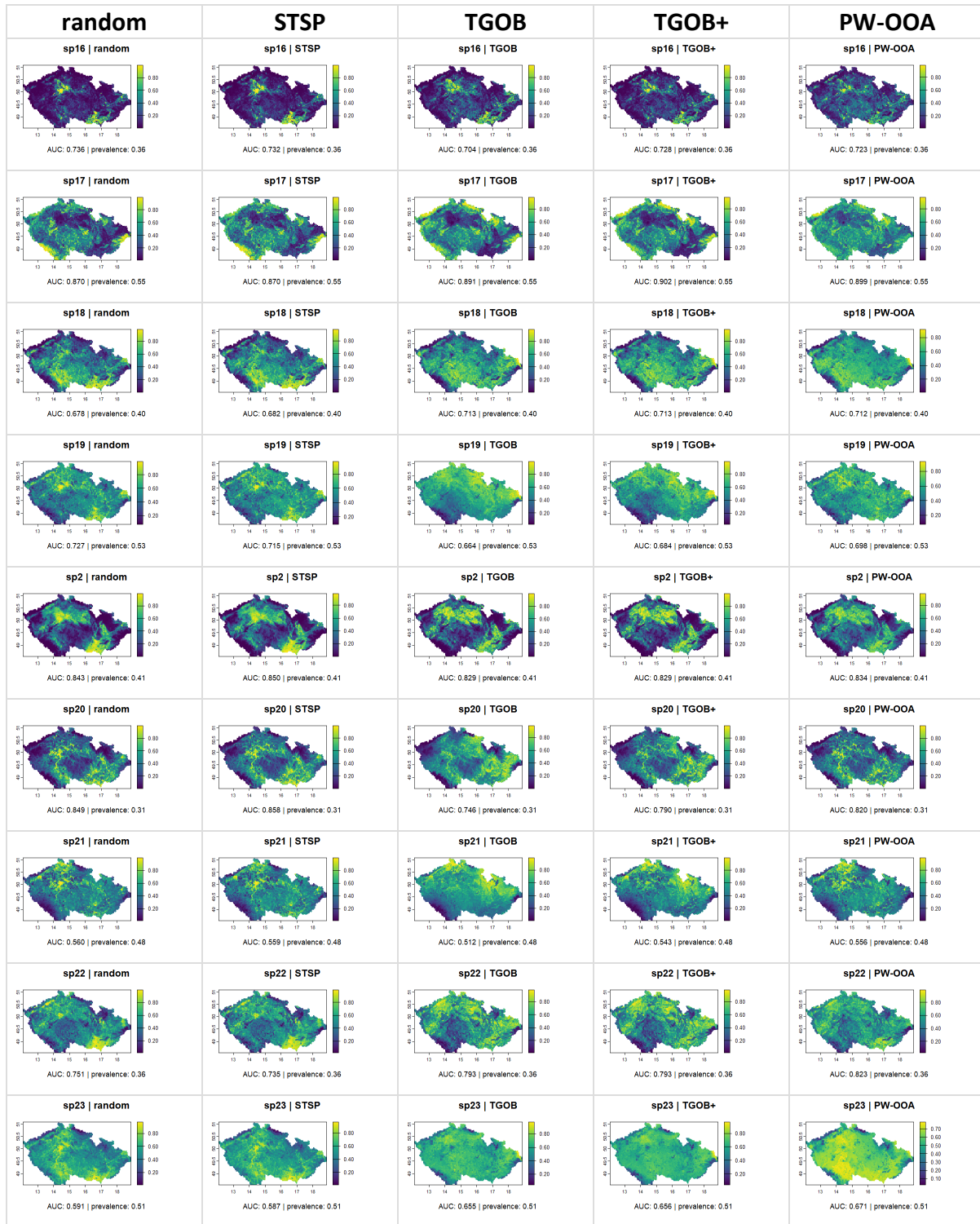
697
 698 *Figure S54. Comparison of SDM prediction performance of 108 bird species differing in species*
 699 *prevalence (the ratio of the number of pixel with species presence to the total number of 154 PAD pixels)*
 700 *measured as TSS using Random forest. Locally estimated scatterplot smoothing (LOESS) regression*
 701 *lines, a non-parametric method that fits a smooth curve using locally weighted polynomial regression, are*
 702 *shown with 95% confidence bands. Compares all the LOESS regression lines in grid. Spatial thinning of*
 703 *species presences (STSP), Target Group Occurrences Background (TGOB), TGOB+ (tuned up TGOB by*
 704 *adjusting kernel smoothing bandwidths) and presence-weighted observer-oriented approach (PW-OOA).*
 705

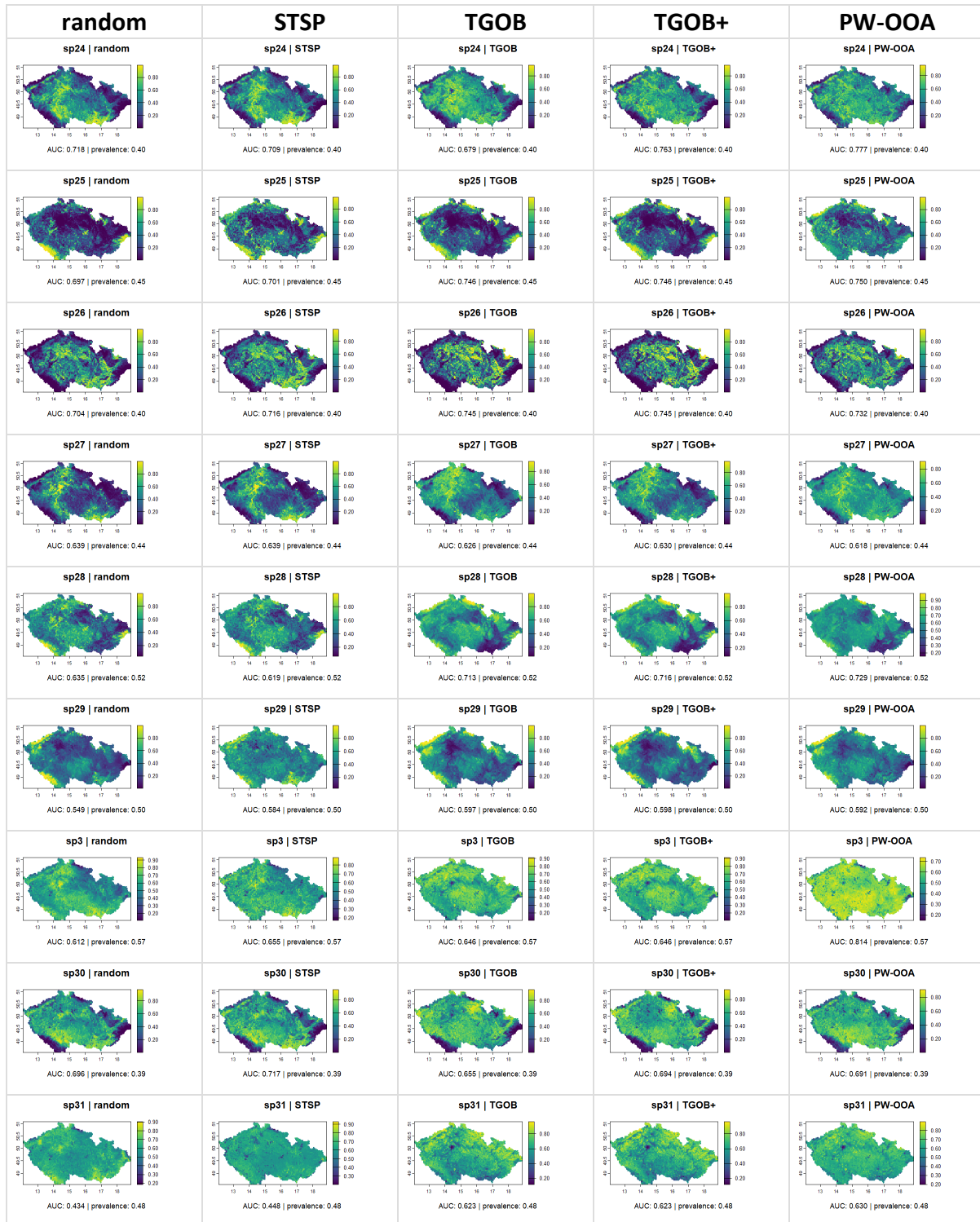
706 Results (part C)

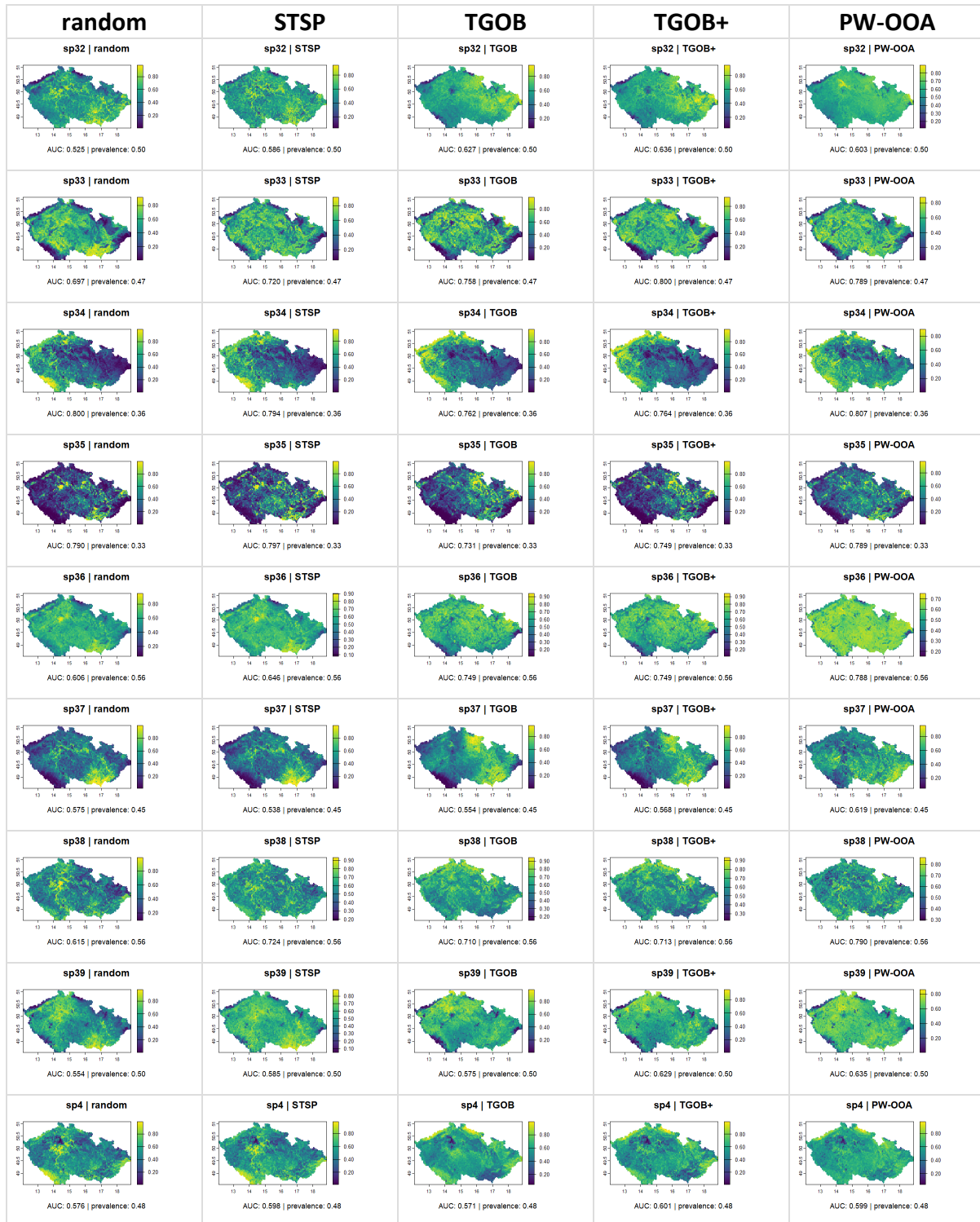
707 *Table S19. Visual comparison of predicted habitat suitability maps generated by five different approaches*
 708 *(random, STSP, TGOB, TGOB+ and PW-OOA) using Maxnet. The best-performing model for each*
 709 *approach was selected based on evaluation against the independent dataset (PAD).*
 710

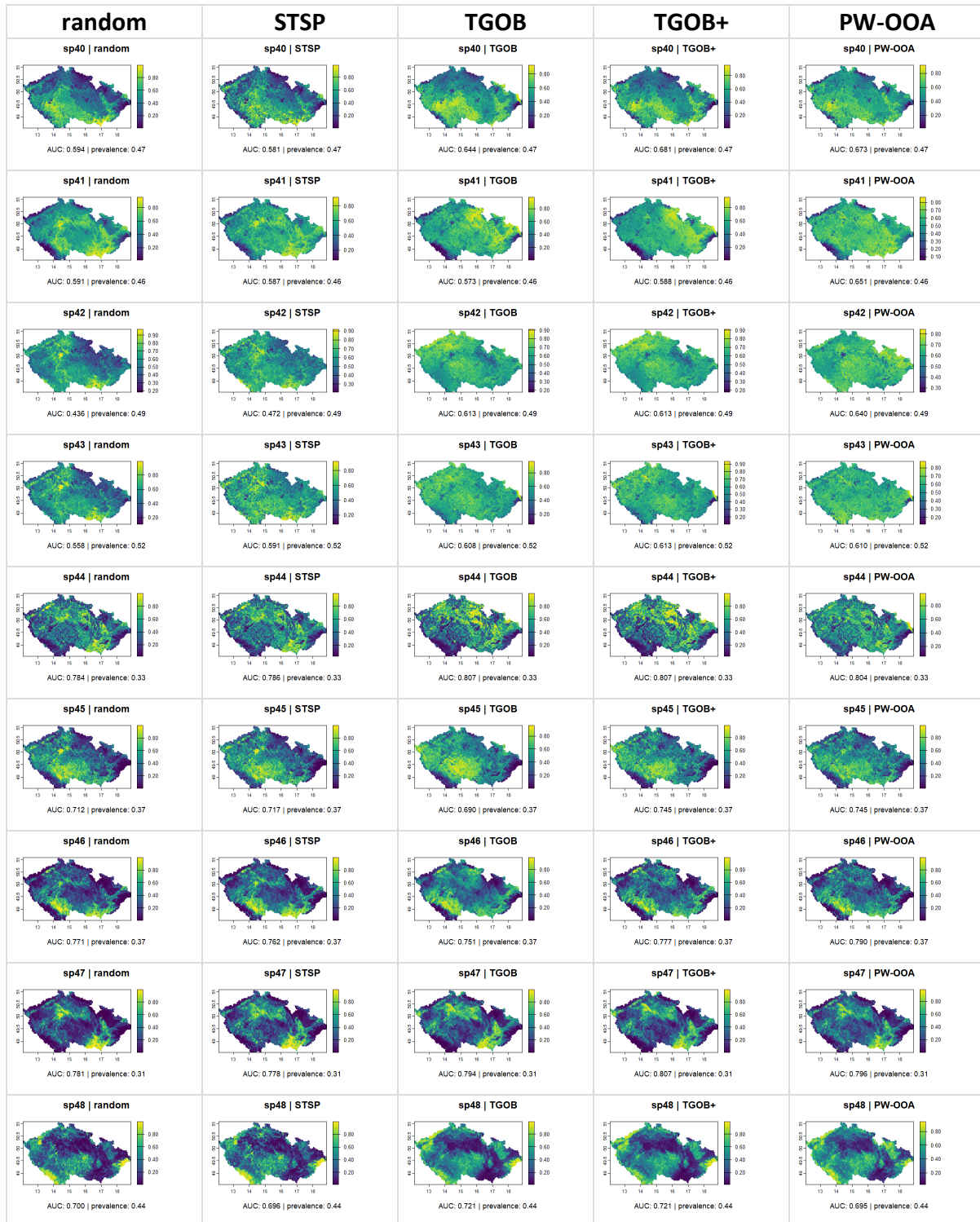


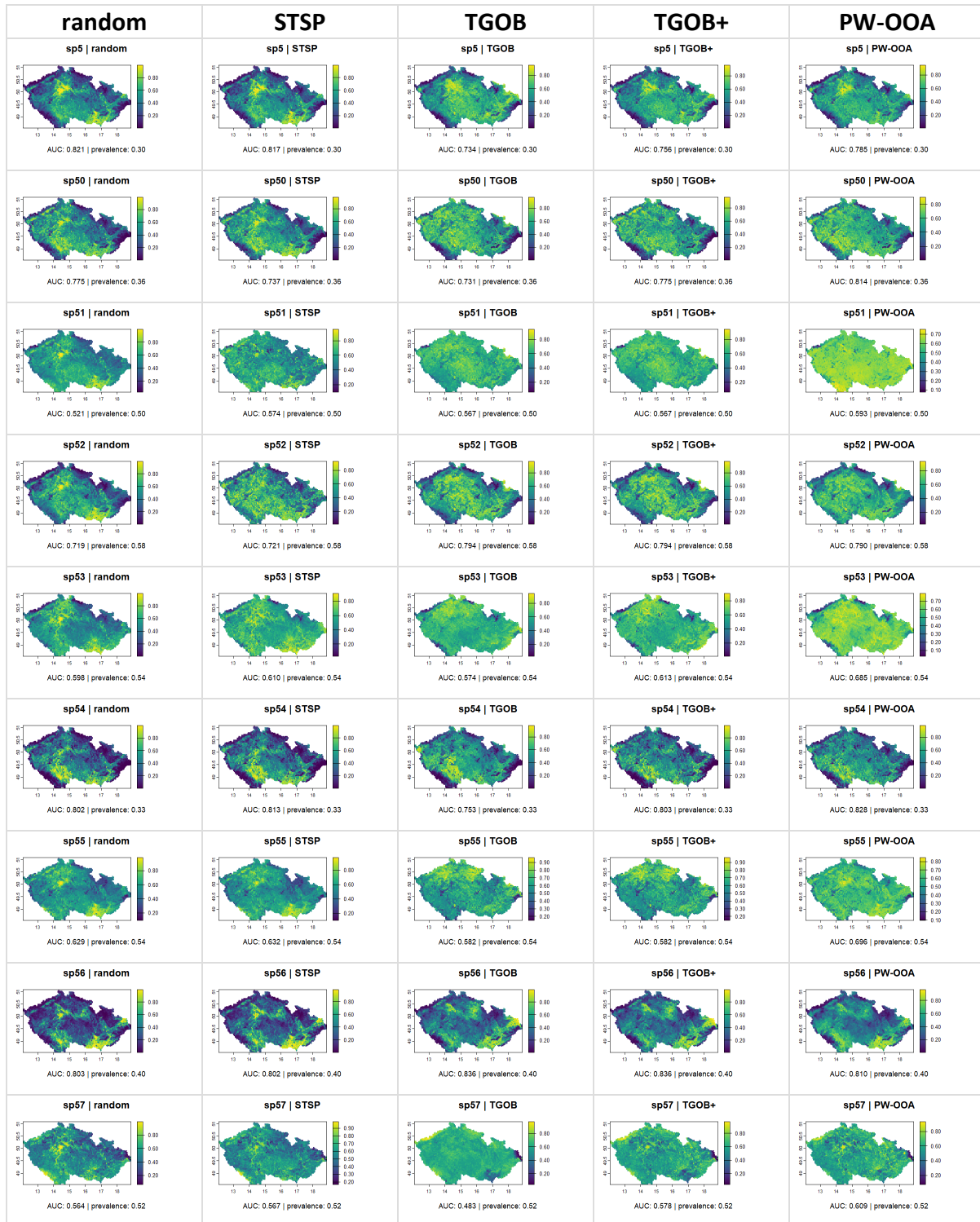


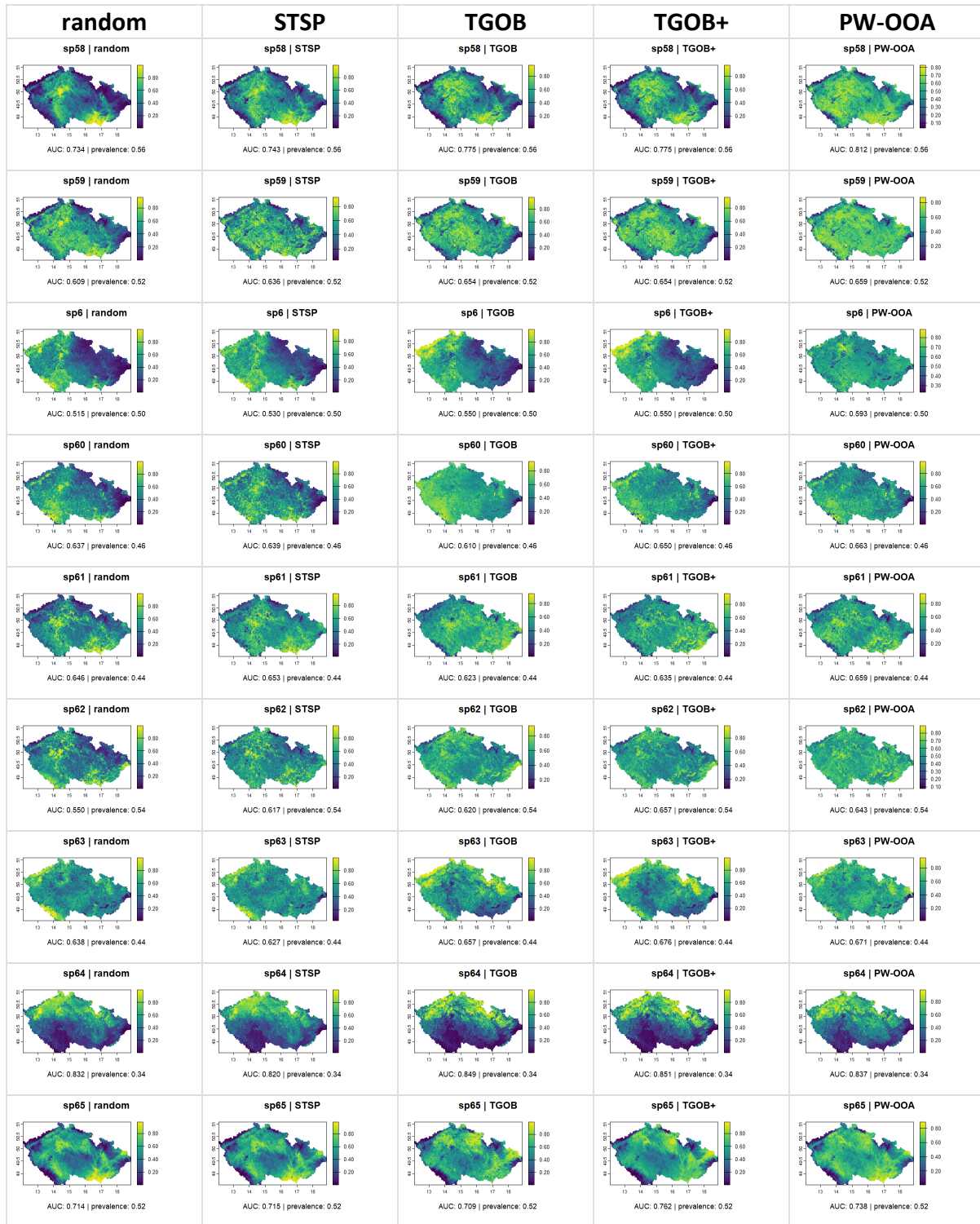


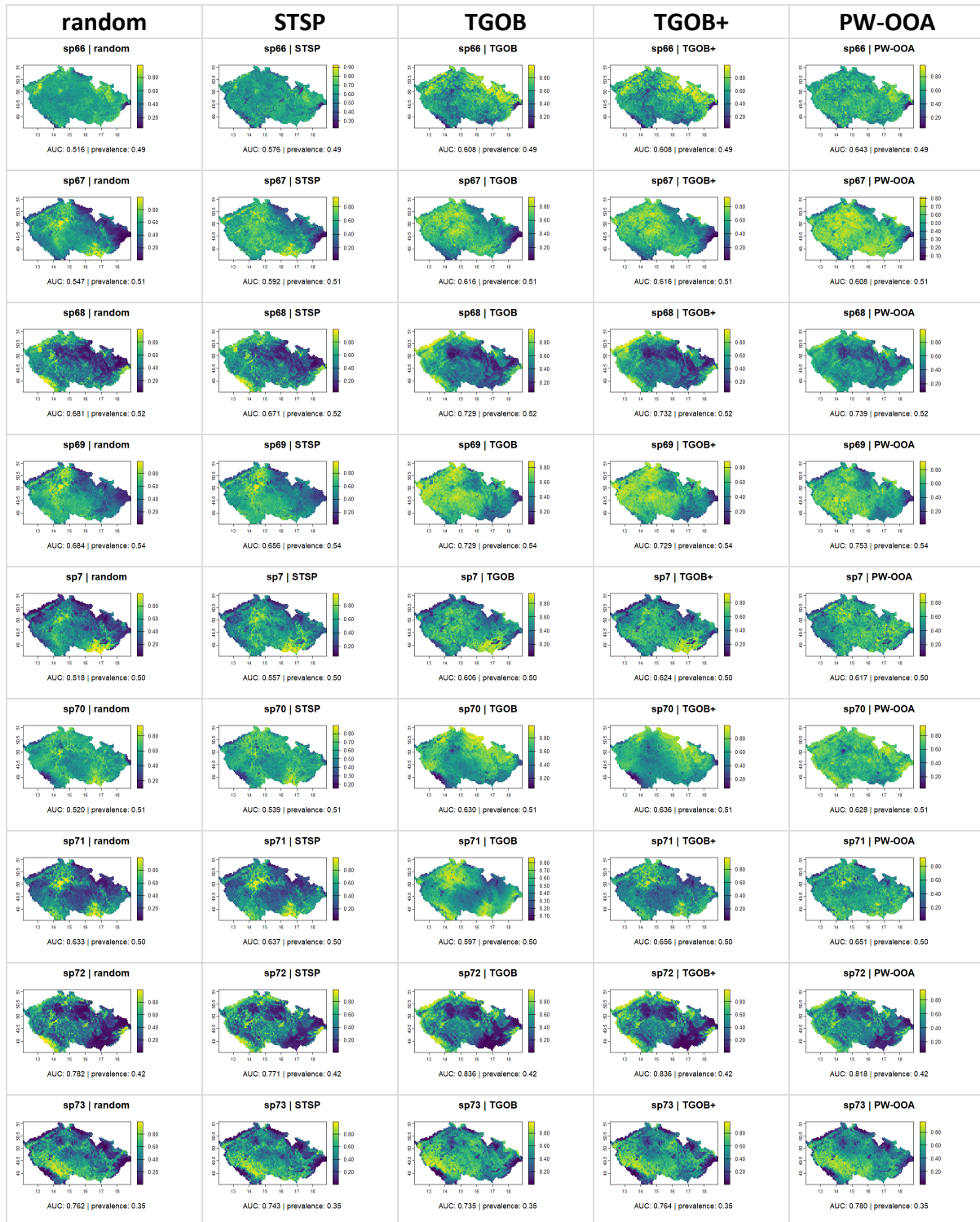


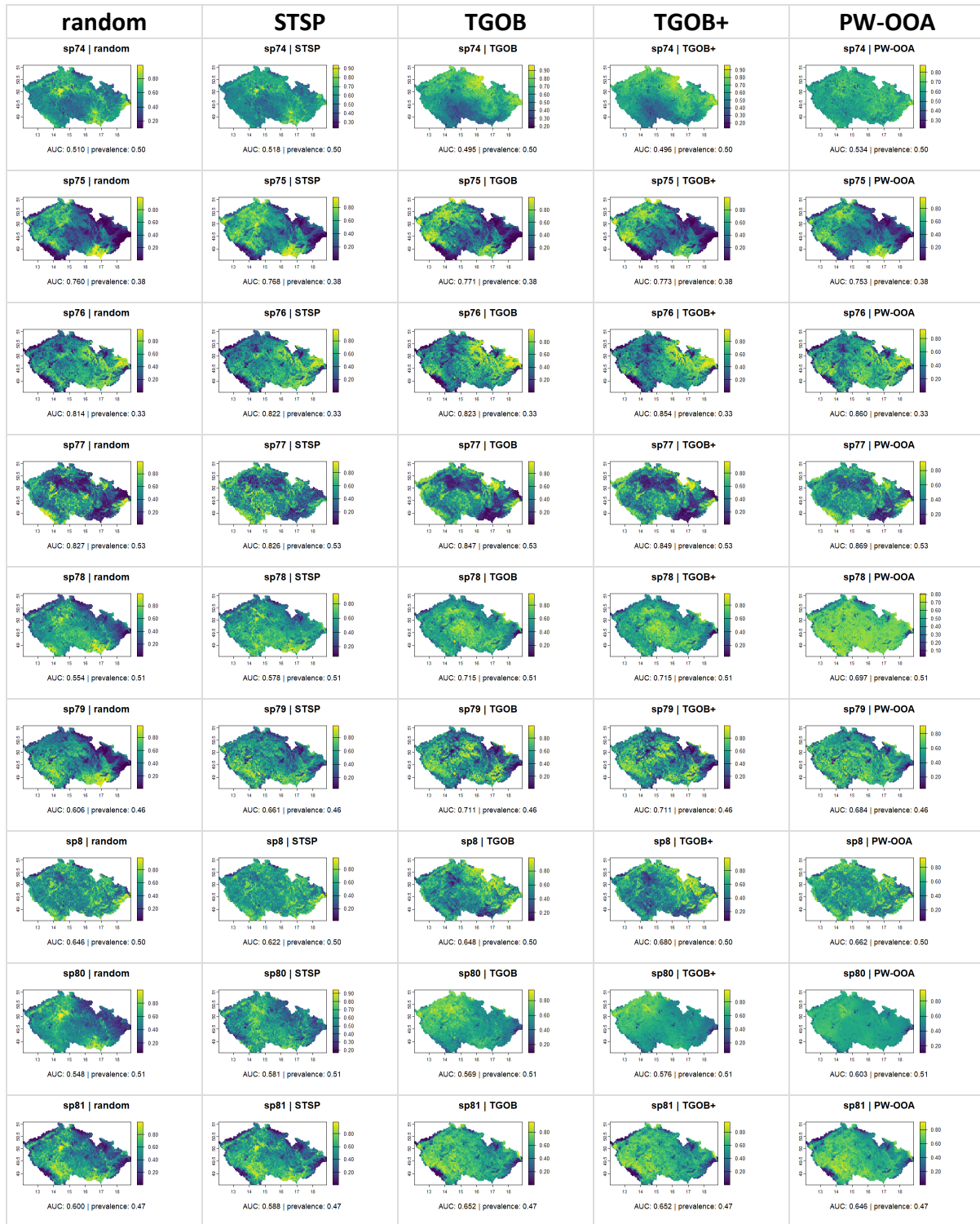


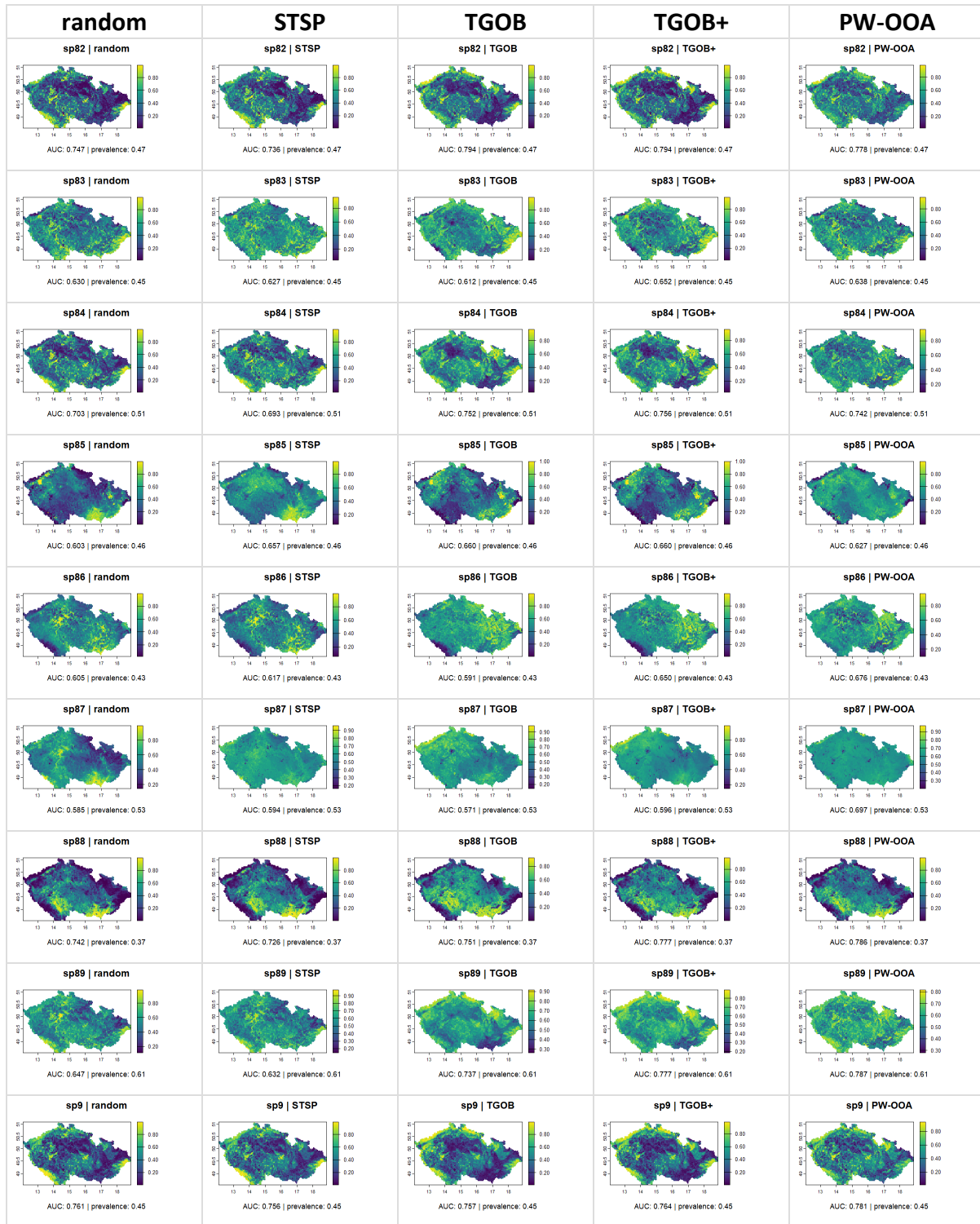


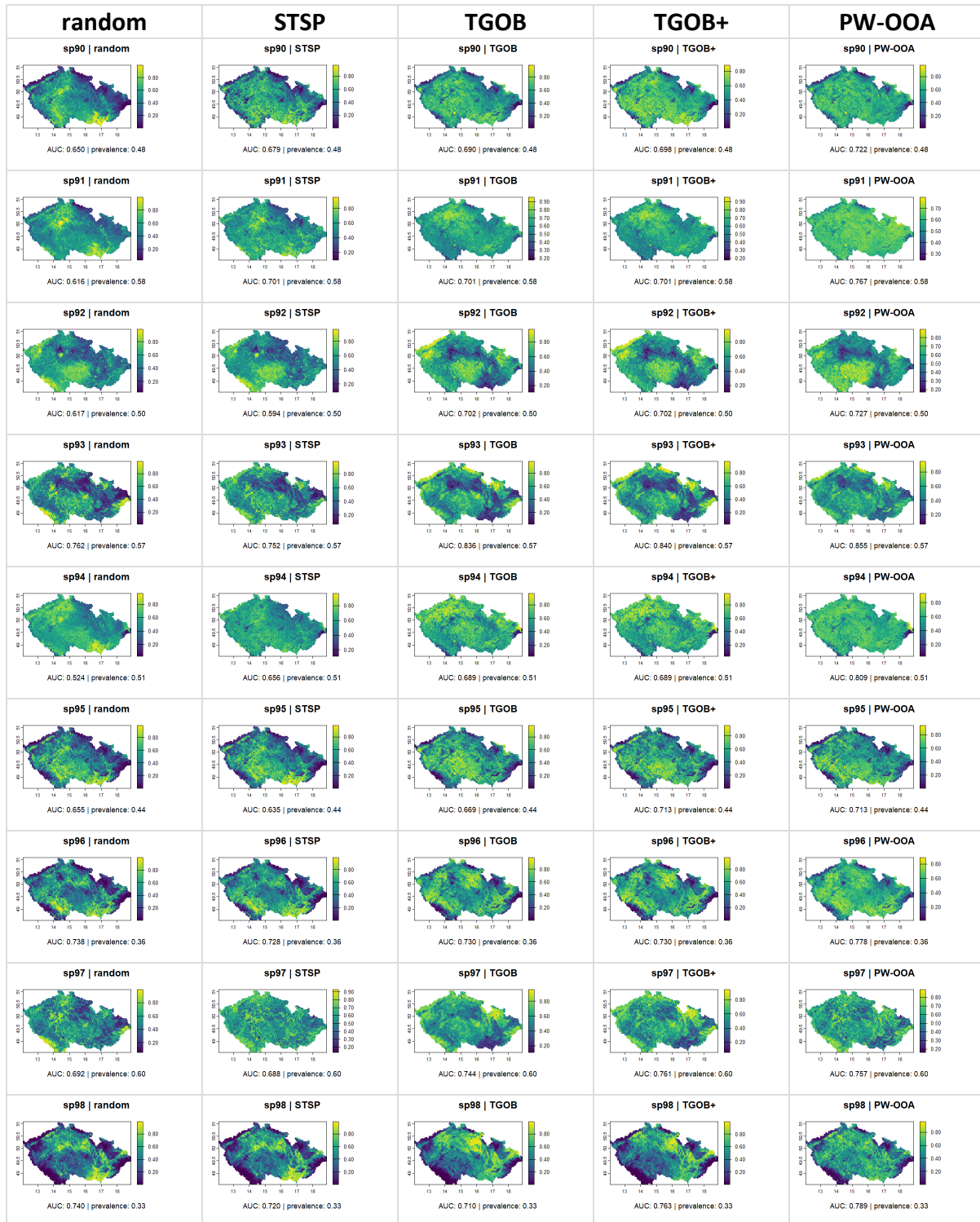


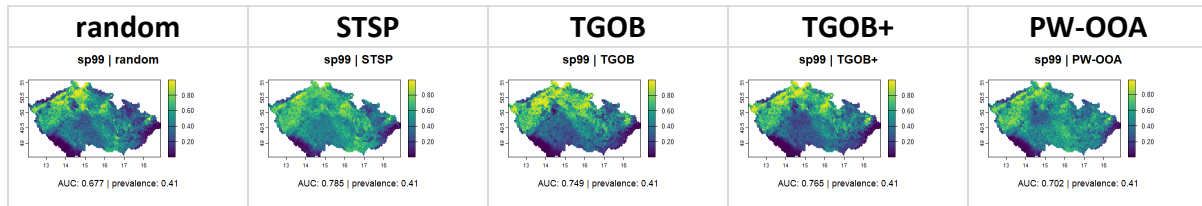












711