Supplementary Information Two (S2) for 'Modelling heterogeneity in the classification process in multi-species distribution models can improve predictive performance.'

Kwaku Peprah Adjei[1,2], Anders Finstad[2,3], Wouter Koch[1,4], Robert B. O'Hara[1,2]

[1]Norwegian University of Science and Technology

[2]Center for Biodiversity Dynamics, NTNU

[3]Department of Natural History, Norwegian University of Science and Technology

[4]Norwegian Biodiversity Information Center

Table 1: Summary of sample sizes for cross-tabulation between each true-state and reported state across the 200 simulations (with standard error in paranthesis) for each of the three simulation scenarios (correlation, full and reduced model) described in section 2.4 in the main paper. The simulation study used two true-states (represented by the first value before the comma in the header of columns three to eight) and three reported states (represented by the second value after the comma in the header of columns three to eight). The number of misclassified samples were decreased by adding a factor of 6 to ω_{0jk} for all $j = k$.

Model	Factor	1,1	2,1	1,2	2,2	1,3	2,3
correlation model	0	226 (8.93)	236 (11.34)	72 (7.47)	184 (9.91)	24 (5.04)	57 (6.56)
correlation model	6	319.5 (9.78)	9 (2.66)	4 (1.9)	465 (176.31)	1 (177.32)	1 (0.89)
full model	0	209 (8.82)	200.5 (9.89)	104 (6.89)	205 (10.61)	28 (5.04)	53 (6.87)
full model	6	325 (9.86)	5.5 (2.33)	14 (3.19)	453 (100.58)	1 (102.24)	0 (0.79)
reduced model	0	249 (10.75)	194 (10.2)	55 (6.95)	195 (11.35)	34 (5.59)	72 (7.13)
reduced model	6	338 (9.16)	1 (1.19)	0 (0.49)	1 (225.79)	450.5 (227.63)	1 (0.65)

Table 2: Summary of classifications of the gull species - common (*Larus canus*), herring (*Larus argentatus*), great black-backed (*Larus marinus*) and lesser black-backed (*Larus fuscus*) - in the training dataset.

True-state	common	great black-backed	herring	lesser black-backed	other
common	775	0	7	0	2
great black-backed	0	254	1	1	0
herring	0	1	846	0	4
lesser black-backed	0	2	1	282	0

Table 3: Summary of classifications of the gull species - common (*Larus canus*), herring (*Larus argentatus*), great black-backed (*Larus marinus*) and lesser black-backed (*Larus fuscus*) - in the validation dataset.

True-state	common	great black-backed	herring	lesser black-backed	other
common	139	0	4	0	2
great black-backed	0	43	0	2	0
herring	0	0	161	0	1
lesser black-backed	0	1	0	34	0

Table 4: Summary variable selection probability from the simulation studies (with standard errors in parenthesis). The first column refers to the simulation method (full, reduced and correlation), the second column indicated whether we decreased the number of misclassified samples (add 6 to the ω_{0jk} for all $j = k$ and baseline refers to using the original true model parameter values) and the next six columns refer to the study scenario type used in this study.

simMethod	trueMisclassIncrease	constant	fixed-covariate	fixed-intercov	intercept	main	variable
correlation	Baseline	0.4999(0.0023)	0.6606(0.0193)	0.6638(0.0112)	0.4998(0.0023)	0.5002(0.0025)	0.6666(0.0045)
correlation	Decreased	0.4997(0.0023)	0.4313(0.0505)	0.484(0.0634)	0.4998(0.0023)	0.4996(0.0023)	0.6663(0.0051)
full	Baseline	0.5001(0.0025)	0.6172(0.0665)	0.5872(0.0807)	0.4997(0.0023)	0.5(0.0024)	0.6666(0.0045)
full	Decreased	0.4999(0.0024)	0.4285(0.0623)	0.5125(0.0857)	0.4996(0.0023)	0.4997(0.0024)	0.6671(0.0048)
reduced	Baseline	0.5002(0.0022)	0.3539(0.0272)	0.3529(0.0233)	0.5001(0.0024)	0.5002(0.0024)	0.3357(0.0083)
reduced	Decreased	0.5001(0.0022)	0.3879(0.0081)	0.4765(0.0236)	0.4996(0.0024)	0.5001(0.0024)	0.359(0.0124)

Table 5: Summary of ecological process paramaters from the simulation studies (with standard errors in parenthesis). The first column refers to the simulation method (full, reduced and correlation), the second column indicated whether we decreased the number of misclassified samples (add 6 to the ω_{0jk} for all $j = k$ and baseline refers to using the original true model parameter values) and the next six columns refer to the study scenario type used in this study.

simMethod	Factor	Parameters	constant	fixed-covariate	fixed-intercov	intercept	main	variable
correlation	Baseline	β_{01}	-0.0382(0.1388)	-0.0395(0.1388)	-0.0366(0.1382)	-0.037(0.1389)	-0.0429(0.1416)	-0.0205(0.13)
correlation	Baseline	β_{11}	0.1008(0.3413)	0.107(0.3424)	0.0951(0.3403)	0.1006(0.3403)	0.1255(0.3455)	0.0881(0.3244)
correlation	Baseline	β_{12}	-0.0559(0.2125)	-0.0662(0.2122)	-0.0599(0.212)	-0.0549(0.2119)	-0.0778(0.2745)	-0.0455(0.196)
correlation	Decreased	β_{01}	-0.0301(0.1272)	-0.0277(0.1264)	-0.0261(0.1264)	-0.0281(0.127)	-0.0339(0.128)	-0.0221(0.1263)
correlation	Decreased	β_{11}	0.0899(0.3061)	0.0856(0.3038)	0.0752(0.3024)	0.0865(0.305)	0.1097(0.3055)	0.0906(0.2982)
correlation	Decreased	β_{12}	-0.0484(0.1889)	-0.046(0.1873)	-0.0405(0.1864)	-0.0464(0.1882)	-0.0686(0.2437)	-0.049(0.1859)
full	Baseline	β_{01}	-0.0025(0.1347)	-0.0033(0.1348)	-4e-04(0.1344)	-0.0019(0.1349)	-0.0086(0.1362)	0.0019(0.1335)
full	Baseline	β_{11}	0.0905(0.3131)	0.0929(0.3134)	0.088(0.312)	0.0897(0.3137)	0.1119(0.3143)	0.097(0.3029)
full	Baseline	β_{12}	-0.047(0.1906)	-0.0487(0.1894)	-0.047(0.189)	-0.0464(0.1899)	-0.0587(0.1904)	-0.0506(0.1848)
full	Decreased	β_{01}	0.0081(0.1217)	0.0081(0.1215)	0.0087(0.121)	0.0074(0.1215)	0.0028(0.122)	0.0062(0.1214)
full	Decreased	β_{11}	0.0702(0.3057)	0.0636(0.3043)	0.0544(0.3045)	0.0645(0.3072)	0.0837(0.3073)	0.0897(0.3006)
full	Decreased	β_{12}	-0.0355(0.1767)	-0.0323(0.1767)	-0.0278(0.1763)	-0.0328(0.1771)	-0.0449(0.1786)	-0.0411(0.1738)
reduced	Baseline	β_{01}	-0.0235(0.1308)	-0.0231(0.1305)	-0.0157(0.1327)	-0.0231(0.1306)	-0.0287(0.1324)	-0.0241(0.1305)
reduced	Baseline	β_{11}	0.1172(0.3359)	0.1162(0.3335)	0.1125(0.3356)	0.1178(0.3341)	0.1367(0.3387)	0.1174(0.3332)
reduced	Baseline	β_{12}	-0.0436(0.2163)	-0.0429(0.2157)	-0.041(0.2172)	-0.0438(0.2151)	-0.0542(0.2174)	-0.044(0.2155)
reduced	Decreased	β_{01}	-0.0192(0.1191)	-0.0223(0.1206)	-0.0172(0.1198)	-0.0223(0.1204)	-0.0241(0.1208)	-0.0226(0.1198)

(continued)

simMethod	Factor	Parameters	constant	fixed-covariate	fixed-intercov	intercept	main	variable
reduced	Decreased	β_{11}	0.0966(0.3114)	0.112(0.314)	0.0926(0.3126)	0.1127(0.3138)	0.1137(0.313)	0.1131(0.3132)
reduced	Decreased	β_{12}	-0.0389(0.1909)	-0.0468(0.1921)	-0.0373(0.1914)	-0.047(0.1919)	-0.0479(0.1917)	-0.0475(0.1928)

Table 6: Summary of bias in observation model parameters from the simulation studies (with standard errors in parenthesis). The first column refers to the simulation method (full, reduced and correlation), the second column indicated whether we decreased the number of misclassified samples (add 6 to the ω_{0jk} for all $j = k$ and baseline refers to using the original true model parameter values) and the next six columns refer to the study scenario type used in this study.

simMethod	Factor	Parameters	constant	fixed-covariate	fixed-intercov	intercept	main	variable
correlation	Baseline	ω_{011}	-1.9989(0.0081)	0.1636(0.1633)	-0.4152(0.0972)	0.1411(0.1592)	-2.002(0.0077)	-0.1395(0.1901)
correlation	Baseline	ω_{021}	-1.0007(0.0078)	0.3844(0.1191)	0.7436(0.1141)	0.3888(0.1162)	-1.0001(0.0075)	-0.0254(0.1441)
correlation	Baseline	ω_{012}	-0.4988(0.0088)	0.5122(0.1986)	-1e-04(0.1448)	0.4906(0.1961)	-0.5023(0.008)	-0.1118(0.2555)
correlation	Baseline	ω_{022}	-1.0001(0.0081)	0.1375(0.1181)	0.5848(0.0972)	0.142(0.1142)	-1.0004(0.0074)	-0.0236(0.1305)
correlation	Baseline	ω_{111}	-2.9983(0.0083)	-2.0409(0.1864)	-2.1866(0.1348)	-2.9992(0.0082)	-3.0006(0.0085)	-0.3725(0.328)
correlation	Baseline	ω_{121}	1.0015(0.0082)	NA(NA)	NA(NA)	1.0002(0.0073)	0.9991(0.0081)	-0.0317(0.1827)
correlation	Baseline	ω_{112}	1.0008(0.0084)	NA(NA)	NA(NA)	1.0002(0.008)	1.001(0.0083)	-0.2041(0.3174)
correlation	Baseline	ω_{122}	-0.9988(0.0086)	-1.1123(0.1019)	-1.0628(0.1331)	-0.9991(0.0074)	-0.9997(0.0083)	0.0063(0.1859)
correlation	Decreased	ω_{011}	-8.0014(0.0083)	-3.5599(0.1409)	-2.3126(0.2386)	-3.5711(0.1496)	-8.0021(0.0079)	-3.4046(0.122)
correlation	Decreased	ω_{021}	-1.0006(0.0078)	-0.3858(0.2633)	0.5529(0.3247)	-0.4001(0.2647)	-0.9992(0.0093)	-1.5889(0.3124)
correlation	Decreased	ω_{012}	-0.5001(0.0074)	-0.614(0.3736)	0.3021(0.4817)	-0.6242(0.3787)	-0.4995(0.0079)	-1.1944(0.2275)
correlation	Decreased	ω_{022}	-6.9985(0.0086)	-2.3738(0.1198)	-1.3126(0.2386)	-2.3893(0.1228)	-7.0015(0.0087)	-2.0769(0.133)
correlation	Decreased	ω_{111}	-2.9991(0.0086)	-2.8256(0.1866)	-2.6638(0.3819)	-2.9998(0.0078)	-3.0002(0.0076)	-2.1691(0.2124)
correlation	Decreased	ω_{121}	1.0009(0.008)	NA(NA)	NA(NA)	0.9998(0.0077)	0.9991(0.0083)	-0.4226(0.2554)
correlation	Decreased	ω_{112}	0.9991(0.0079)	NA(NA)	NA(NA)	1(0.0081)	1.0012(0.0074)	-0.208(0.4353)
correlation	Decreased	ω_{122}	-0.9985(0.008)	-1.0171(0.0641)	-0.9539(0.175)	-1.0006(0.0073)	-0.9989(0.008)	-0.7554(0.1592)

(continued)

simMethod	Factor	Parameters	constant	fixed-covariate	fixed-intercov	intercept	main	variable
full	Baseline	ω_{011}	-1.9998(0.0082)	0.0102(0.1695)	-0.3755(0.1028)	-0.1118(0.1623)	-2.0012(0.0084)	-0.1229(0.2067)
full	Baseline	ω_{021}	-1.0006(0.0084)	0.2892(0.1346)	0.5403(0.123)	0.2801(0.1342)	-0.9991(0.0072)	-0.0183(0.1505)
full	Baseline	ω_{012}	-0.4984(0.0083)	0.8067(0.1814)	0.4686(0.1212)	0.6856(0.1805)	-0.501(0.0086)	-0.1003(0.2397)
full	Baseline	ω_{022}	-0.9991(0.0088)	0.3157(0.134)	0.6245(0.1028)	0.3064(0.1335)	-1.0007(0.0083)	-0.0015(0.1532)
full	Baseline	ω_{111}	-2.9985(0.0079)	-2.468(0.1851)	-2.613(0.1697)	-3.0002(0.0074)	-3.0009(0.0082)	-0.36(0.2698)
full	Baseline	ω_{121}	1.0012(0.0081)	NA(NA)	NA(NA)	0.9998(0.0086)	0.9997(0.0077)	0.0069(0.1792)
full	Baseline	ω_{112}	1.0016(0.008)	NA(NA)	NA(NA)	1.0006(0.008)	1.0015(0.0076)	-0.1365(0.2745)
full	Baseline	ω_{122}	-0.9984(0.0077)	-0.9617(0.0959)	-0.9622(0.1104)	-1.0005(0.0074)	-1.0004(0.0079)	0.0131(0.1838)
full	Decreased	ω_{011}	-8.0006(0.0078)	-3.8273(0.1177)	-2.5454(0.1905)	-3.8677(0.1545)	-8.0003(0.0085)	-3.2998(0.1149)
full	Decreased	ω_{021}	-1.002(0.0072)	-0.8604(0.3543)	-0.2522(0.4011)	-0.8759(0.3542)	-1.0006(0.0085)	-1.5137(0.2893)
full	Decreased	ω_{012}	-0.4993(0.0083)	0.4697(0.2179)	1.6485(0.2733)	0.4329(0.2447)	-0.5004(0.0077)	-1.3927(0.2446)
full	Decreased	ω_{022}	-6.999(0.0079)	-2.2977(0.14)	-1.5454(0.1905)	-2.3177(0.1457)	-7.0024(0.0086)	-2.1148(0.1388)
full	Decreased	ω_{111}	-2.9992(0.0081)	-2.8562(0.1734)	-2.5179(0.4431)	-2.9999(0.0085)	-3(0.0081)	-2.0628(0.2436)
full	Decreased	ω_{121}	1.0009(0.0084)	NA(NA)	NA(NA)	0.999(0.0076)	0.9989(0.0087)	-0.1072(0.3298)
full	Decreased	ω_{112}	1.0005(0.0089)	NA(NA)	NA(NA)	1.0013(0.0085)	1.0005(0.0088)	-0.859(0.2838)
full	Decreased	ω_{122}	-0.9982(0.0084)	-0.9559(0.0753)	-0.8919(0.1719)	-1.0006(0.0079)	-1(0.008)	-0.6688(0.1461)
reduced	Baseline	ω_{011}	-1.9997(0.0082)	-0.0505(0.1502)	-0.5643(0.0867)	-0.052(0.1502)	-2.0017(0.0072)	-0.0528(0.1498)
reduced	Baseline	ω_{021}	-1.0017(0.0089)	-0.0152(0.1052)	0.3316(0.1001)	-0.0159(0.1042)	-0.999(0.0079)	-0.015(0.1049)

(continued)

simMethod	Factor	Parameters	constant	fixed-covariate	fixed-intercov	intercept	main	variable
reduced	Baseline	ω_{012}	-0.4991(0.0081)	-0.07(0.1834)	-0.5059(0.142)	-0.0717(0.1839)	-0.501(0.0078)	-0.0727(0.183)
reduced	Baseline	ω_{022}	-0.9998(0.0082)	-0.0111(0.1148)	0.4357(0.0867)	-0.0117(0.1138)	-1.0008(0.0079)	-0.0109(0.1141)
reduced	Baseline	ω_{111}	-2.9987(0.0081)	-2.9995(0.0374)	-3.0006(0.0207)	-3.0001(0.0076)	-2.9995(0.0082)	-2.9986(0.0165)
reduced	Baseline	ω_{121}	1.0013(0.0086)	NA(NA)	NA(NA)	0.9996(0.0089)	0.9995(0.0081)	1(0.0164)
reduced	Baseline	ω_{112}	1.0011(0.0082)	NA(NA)	NA(NA)	0.999(0.0081)	1.0023(0.0076)	1(0.0168)
reduced	Baseline	ω_{122}	-0.9978(0.0079)	-0.999(0.0229)	-1.0009(0.029)	-0.9992(0.008)	-1.0007(0.0086)	-1.0012(0.0172)
reduced	Decreased	ω_{011}	-7.9997(0.008)	-3.2048(0.0822)	0.479(2.417)	-3.2145(0.081)	-8.0011(0.0082)	-3.2093(0.0816)
reduced	Decreased	ω_{021}	-1.0013(0.0084)	-1.8062(0.3513)	-0.6055(0.5142)	-1.8103(0.3499)	-1.001(0.0089)	-1.8181(0.3488)
reduced	Decreased	ω_{012}	-0.5004(0.0082)	-1.5674(0.1783)	-0.6053(0.3139)	-1.5711(0.1817)	-0.5013(0.0079)	-1.5762(0.1784)
reduced	Decreased	ω_{022}	-7.001(0.0088)	-2.0711(0.1343)	1.479(2.417)	-2.081(0.1345)	-7.0008(0.0077)	-2.0744(0.134)
reduced	Decreased	ω_{111}	-2.9997(0.0082)	-2.9964(0.0209)	-2.9966(0.1032)	-3.0003(0.0074)	-3.0016(0.0076)	-2.9943(0.0178)
reduced	Decreased	ω_{121}	0.9998(0.0085)	NA(NA)	NA(NA)	1.0003(0.0077)	1.0009(0.0083)	0.9955(0.0382)
reduced	Decreased	ω_{112}	0.9995(0.0082)	NA(NA)	NA(NA)	0.9992(0.0077)	1.0008(0.0079)	1.0015(0.0225)
reduced	Decreased	ω_{122}	-0.9991(0.008)	-0.9952(0.0246)	-0.9925(0.1476)	-0.9996(0.0078)	-0.9991(0.0077)	-1.0008(0.0213)