Ecography

E4823
McPherson, J. M. and Jetz, W. 2007. Effects of species' ecology on the accuracy of distribution models. – Ecography 30: 135–151.

Appendix

We provide two tables, three figures and a programming script as supplementary materials. The tables and figures correspond to tables and figures in the main text, but present results obtained at quarter-degree rather than half-degree resolution.

Table S1. Effects of individual ecological traits on model accuracy at quarter-degree resolution. Shown are coefficient estimates (± SE) for the intercept and ecological trait as well as the model's coefficient of determination (r^2). Log-likelihood ratio tests were used to determine whether a trait's influence was significant. Significance is indicated with a single asterisk (*) when $p < 0.05$ or two asterisks (**) when $p < 0.01$; n = 1,092. Parameter estimates in bold remained significant (at global $\alpha = 0.05$) after Holm's adjustment for multiple comparisons.

		Logistic regression			Autologistic regression		
Response	Predictor	Intercept	Parameter	r^2	Intercept	Parameter	r^2
Sensitivity							
	body mass (log)	2.29 ± 0.24	–0.20 ± 0.12	0.03	2.43 ± 0.26	–0.21 ± 0.13	0.03
	conservation status	2.12 ± 0.33	0.20 ± 0.33	0.00	2.14 ± 0.32	0.09 ± 0.32	0.00
	endemism	2.09 ± 0.12	0.28 ± 0.12 *	0.06	2.23 ± 0.13	0.30 ± 0.13 *	0.06
	global range size (log)	5.94 ± 1.14	**–0.61 ± 0.17** **	0.15	6.08 ± 1.20	**–0.61 ± 0.18** **	0.14
	habitat structure						
	open	1.92 ± 0.09	–0.08 ± 0.11	0.05	2.04 ± 0.10	–0.07 ± 0.12	0.03
	complex		0.13 ± 0.06			0.11 ± 0.07	
	habitat tolerance	2.28 ± 0.20	–0.16 ± 0.08 *	0.05	2.37 ± 0.20	–0.14 ± 0.08 *	0.04
	migratory behaviour	1.83 ± 0.10	–0.24 ± 0.10 *	0.07	1.95 ± 0.10	–0.25 ± 0.10 *	0.07
	sub-resolution habitat	1.92 ± 0.09	–0.06 ± 0.09	0.00	2.04 ± 0.10	–0.06 ± 0.10	0.00
	trophic rank	2.35 ± 0.33	–0.22 ± 0.16	0.02	2.45 ± 0.34	–0.21 ± 0.17	0.02
	wetland affinity	1.75 ± 0.10	**–0.32 ± 0.10** **	0.11	1.89 ± 0.11	–0.29 ± 0.11 **	0.08
Specificity							
	body mass (log)	1.97 ± 0.22	–0.19 ± 0.11	0.03	2.04 ± 0.23	–0.14 ± 0.12	0.01
	conservation status	1.71 ± 0.27	0.09 ± 0.27	0.00	1.92 ± 0.30	0.13 ± 0.30	0.00
	endemism	1.71 ± 0.10	0.16 ± 0.10	0.02	1.87 ± 0.11	0.15 ± 0.11	0.02
	global range size (log)	5.62 ± 1.02	**–0.61 ± 0.15** **	0.17	4.89 ± 1.05	**–0.47 ± 0.16** **	0.11
	habitat structure						
	open	1.60 ± 0.08	0.01 ± 0.10	0.03	1.78 ± 0.09	–0.03 ± 0.11	0.02
	complex		0.10 ± 0.06			0.08 ± 0.06	
	habitat tolerance	1.95 ± 0.18	–0.15 ± 0.07 *	0.05	2.06 ± 0.18	–0.12 ± 0.07	0.03
	migratory behaviour	1.53 ± 0.09	–0.23 ± 0.09 *	0.07	1.69 ± 0.09	–0.25 ± 0.09 **	0.08
	sub-resolution habitat	1.62 ± 0.08	–0.03 ± 0.08	0.00	1.78 ± 0.09	–0.06 ± 0.09	0.01
	trophic rank	2.17 ± 0.29	–0.29 ± 0.15 *	0.04	2.28 ± 0.31	–0.26 ± 0.15	0.03
	wetland affinity	1.51 ± 0.10	–0.20 ± 0.10 *	0.04	1.68 ± 0.10	–0.19 ± 0.10	0.04
AUC							
	body mass (log)	2.55 ± 0.27	–0.20 ± 0.13	0.03	2.76 ± 0.30	–0.18 ± 0.15	0.02
	conservation status	2.26 ± 0.34	0.07 ± 0.34	0.00	2.49 ± 0.37	0.08 ± 0.37	0.00
	endemism	2.33 ± 0.13	0.25 ± 0.13 *	0.05	2.55 ± 0.15	0.23 ± 0.15	0.04
	global range size (log)	6.08 ± 1.26	**–0.59 ± 0.19** **	0.15	5.66 ± 1.36	–0.49 ± 0.20 *	0.09
	habitat structure						
	open	2.17 ± 0.10	–0.06 ± 0.13	0.03	2.41 ± 0.11	–0.07 ± 0.14	0.02
	complex		0.10 ± 0.07			0.09 ± 0.07	
	habitat tolerance	2.52 ± 0.22	–0.15 ± 0.08	0.04	2.69 ± 0.24	–0.13 ± 0.09	0.02
	migratory behaviour	2.08 ± 0.11	–0.28 ± 0.11 **	0.09	2.31 ± 0.12	–0.30 ± 0.12 *	0.09
	sub-resolution habitat	2.17 ± 0.10	–0.06 ± 0.10	0.01	2.40 ± 0.11	–0.08 ± 0.11	0.01
	trophic rank	2.79 ± 0.36	–0.32 ± 0.18	0.04	3.06 ± 0.40	–0.33 ± 0.19	0.04
	wetland affinity	2.03 ± 0.12	–0.28 ± 0.12 *	0.08	2.27 ± 0.13	–0.27 ± 0.13 *	0.06

Table S2. Parsimonious generalised linear models (GLMs) describing the influence of ecological characteristics on each of three measures of accuracy in logistic and autologistic distribution models built at quarter-degree resolution. Shown are coefficient estimates (± SE) for each GLM's intercept and applicable predictors, as well as the model's coefficient of determination (r^2).

Response →	Logistic regression models			Autologistic regression models		
↓ Predictors	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC
Intercept	5.08 ± 1.21	5.13 ± 1.05	5.37 ± 1.29	5.47 ± 1.23	4.26 ± 1.07	4.85 ± 1.38
Global range size	–0.49 ± 0.18	–0.54 ± 0.16	–0.49 ± 0.19	–0.53 ± 0.18	–0.39 ± 0.16	–0.38 ± 0.21
Migratory behaviour		–0.15 ± 0.09	–0.20 ± 0.11	–0.17 ± 0.11	–0.19 ± 0.10	–0.24 ± 0.12
Wetland affinity	–0.21 ± 0.11					
r^2	0.19	0.20	0.19	0.17	0.15	0.15

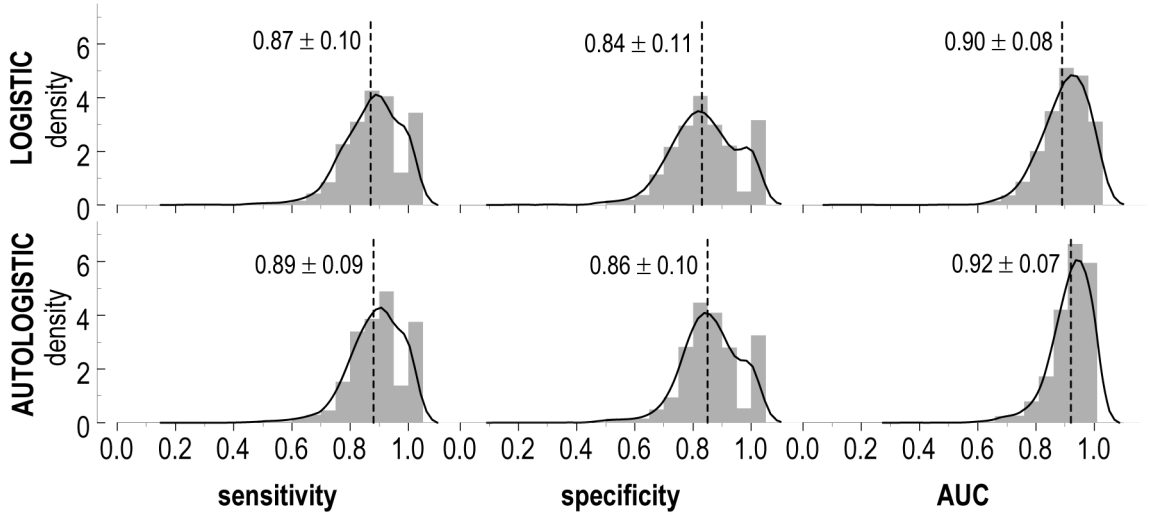

Fig. S1. The accuracy achieved by logistic distribution models (top panels) and autologistic distribution models (bottom panels) at quarter-degree resolution. Accuracy was measured as sensitivity (left), specificity (middle) and AUC (right). Density plots show the spread of accuracy values obtained among models for 1092 species; mean accuracy (± one standard deviation) is indicated by a vertical line.

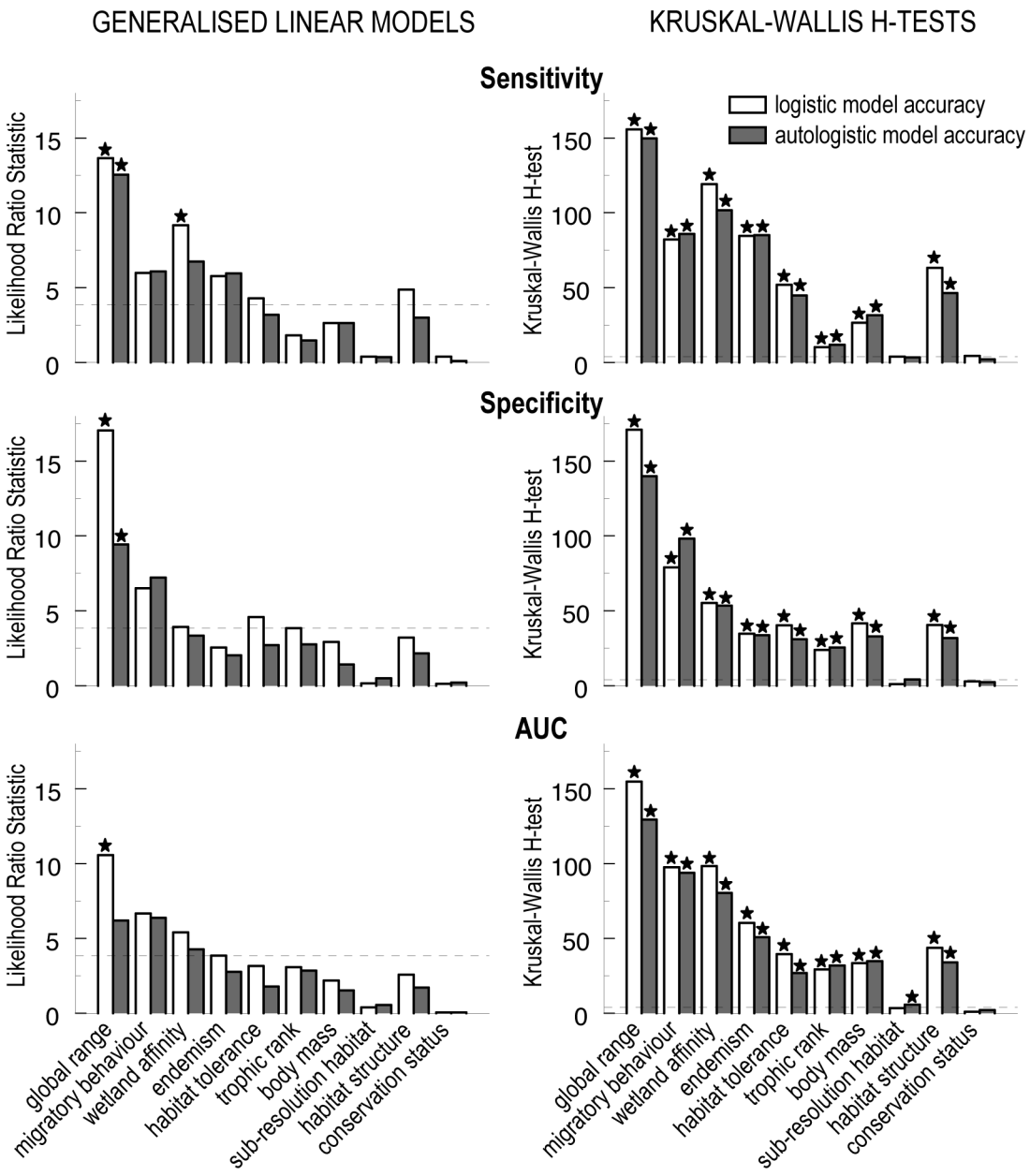

Fig. S2. The relative importance of ecological traits in influencing the accuracy (sensitivity, specificity or AUC, as indicated) of species distribution models built at quarter-degree resolution. Relative importance was judged using: 1) each traits' likelihood ratio statistics in single-trait (left) and multi-trait (not shown) generalised linear models (GLMs), measuring the change in model deviance attributable to the trait; and 2) the test statistic of Kruskal-Wallis H-tests (right). Both statistics are chi-square distributed and were significant (at test-specific $\alpha = 0.05$) only if they surpassed a value of 3.84, demarcated by the dashed horizontal line. (For habitat structure, a three-level factor-variable, the significance cut-off was 5.99, not shown.) Stars above bars indicate which trait's impacts retained significance (at global $\alpha = 0.05$) after Holm's adjustments for multiple comparisons.

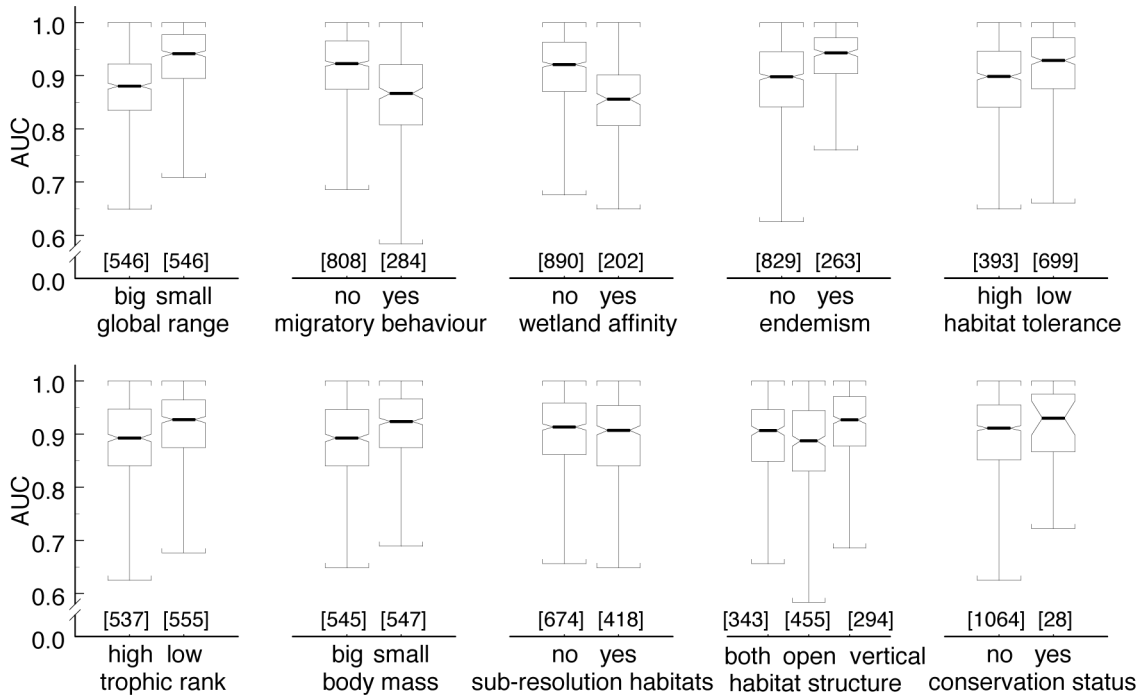

Fig. S3. Box plots illustrating the impact of ecological traits on model accuracy at quarter-degree resolution. Impacts were highly similar across models types (logistic vs autologistic) and accuracy metrics (AUC, sensitivity, and specificity). We therefore show patterns for only AUC in logistic models. Boxes delimit the inter-quartile range, with girdles at the median and notches to indicate the median's 95% confidence intervals. Whiskers show the spread of data up to 1.5 times the inter-quartile range

Text S1. Programming script

Provided below are a set of S-Plus functions, custom-written by JM McPherson, to compute autologistic distribution models. The main function is alogis; it in turn calls the other functions provided (neighbour, block, add.leftcol and add.rightcol). By default, alogis uses an ordinary logistic regression as prior, but other priors can be supplied. It then iterates through a Gibbs sampler, updating blocks of independent cells in turn. These blocks are created using the function *block*. The blocking pattern will depend on the neighbourhood structure used. As set up currently, the neighbourhood structure corresponds to the Queen's case (eight neighbouring cells), and neighbours are identified by function neighbour. Invalid neighbours (that were not sampled, contain no data, or fall outside the study region) are ignored in calculating the autocovariate. To ensure that this is the case, the data matrix must be buffered on all sides. This happens automatically for the northern and southern edge of the study area; for west and east, add.leftcol and add.rightcol ensure a buffer.

Programming script: http://www.oikos.ekol.lu.se/appendixdown/S-Programming-Script_alogis.txt