

Supplementary materials for: Negative impacts of future forest succession on three threatened butterfly species

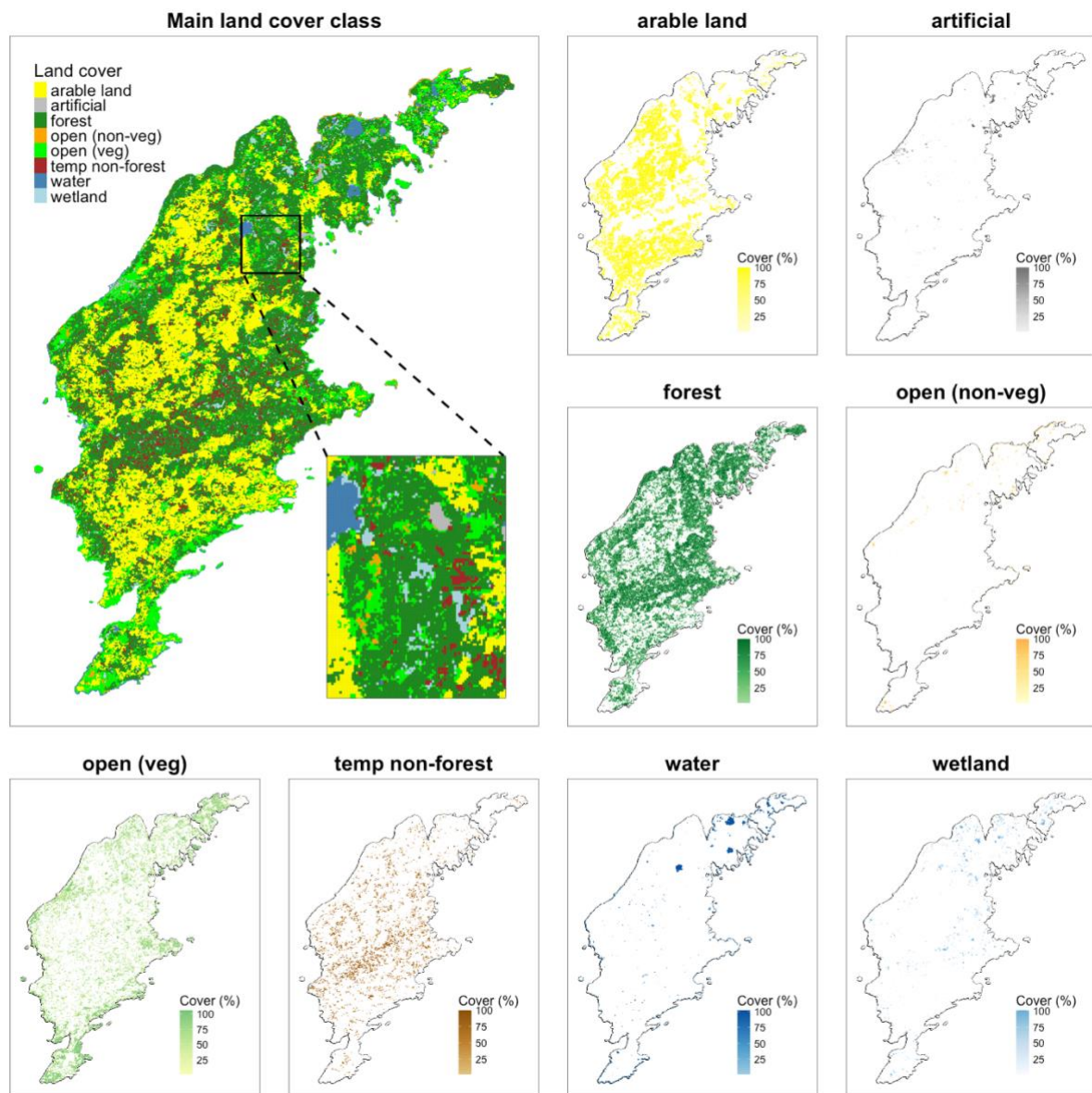


Fig S1. Maps showing the environmental characteristics across the island of Gotland and in the study area. The maps are based on NMD data for which land cover classes has been grouped into eight higher categories: arable land (nmd3), artificial land (nmd51, nmd52, nmd53), forest (nmd111, nmd112, nmd113, nmd114, nmd115, nmd116, nmd117, nmd121, nmd122, nmd123, nmd124, nmd125, nmd126), open non-vegetated land (nmd41), open vegetated land (nmd42), temporary non-forested land (nmd118, nmd128), water (nmd61, nmd62), and wetland (nmd2). The smaller maps show each of the eight grouped land cover categories, shade indicates land cover coverage in %. The larger top left map shows the dominating land cover category in each of the hectare grids, colours indicate the land cover category.

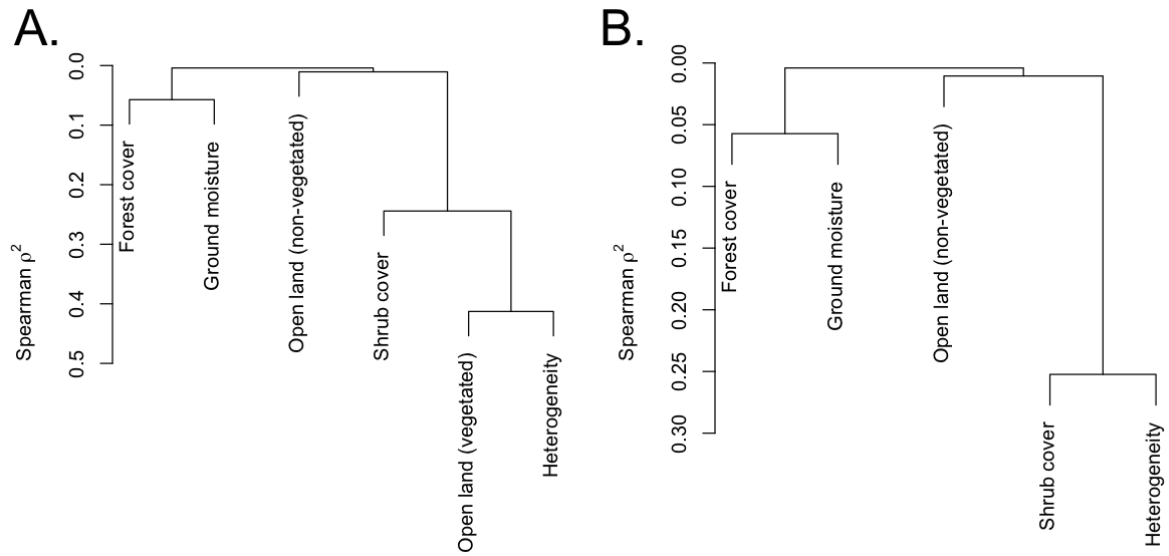


Fig. S2 Results from cluster analysis based on Spearman correlation coefficients. A) shows the correlation coefficients for all the six environmental variables (forest cover, shrub cover, coverage of vegetated open land, coverage of non-vegetated open land, habitat heterogeneity (based on Simpson diversity estimates), and ground moisture) that were initially included in the Varclus analysis (Hmisc package; Harrell, 2020). B) Shows the five variables (all but coverage of vegetated open land) that were retained for the downstream analyses based on an Inclusion threshold of 0.3 (Graham, 2003).

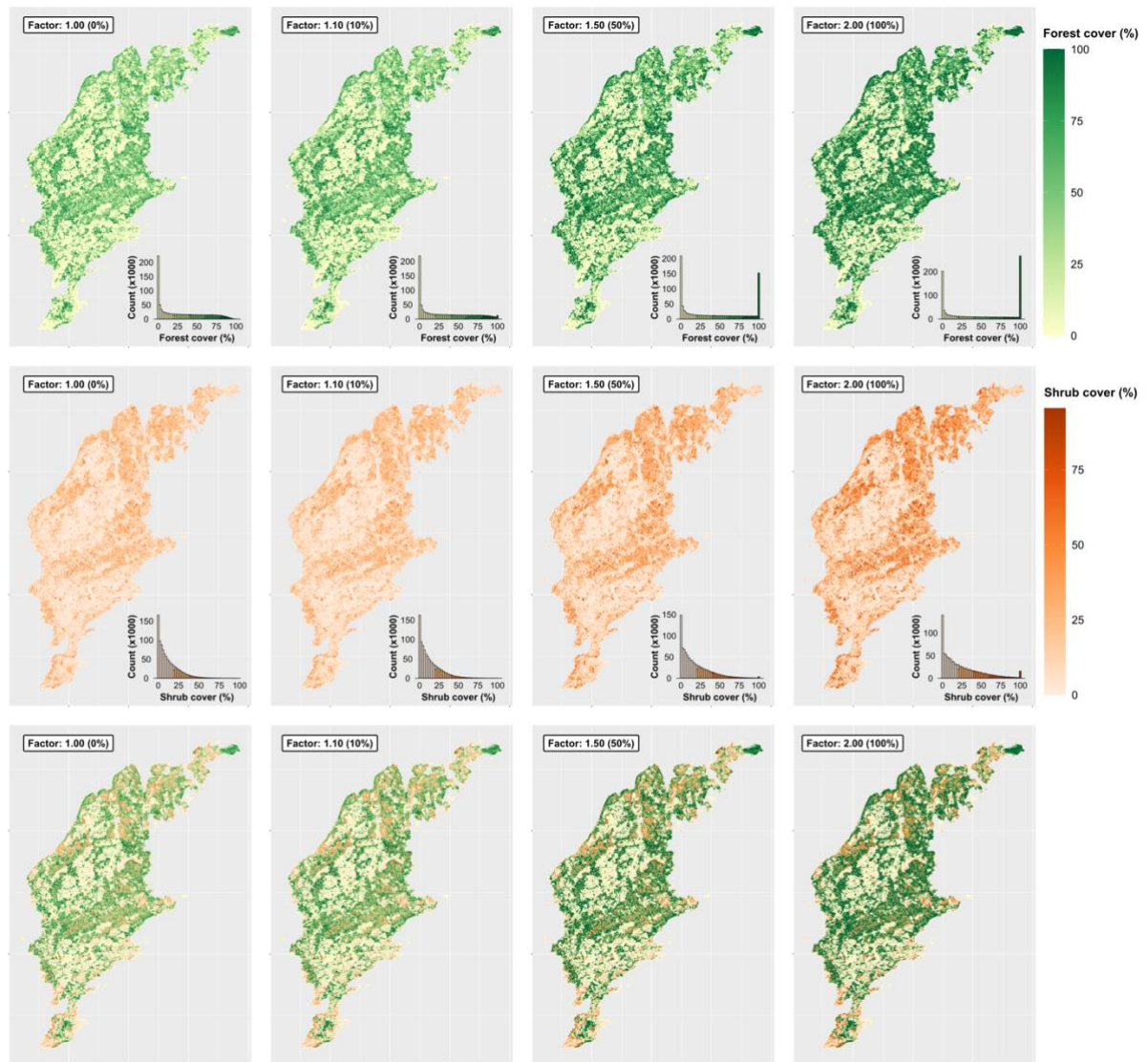


Fig. S3 Potential future land cover scenarios on the island of Gotland given continued forest succession and/or shrub overgrowing. Rows show scenarios for different environmental variables: forest succession (forest cover increase, top row), shrub overgrowing (shrub cover increase, middle row), and both forest succession and shrub overgrowing (bottom row). Panels represent different levels of increase: 0 % increase (current coverage, panel one), 10 % increase (panel two), 50 % increase (panel three), and 100 % increase (panel four). Colours indicate the coverage of forest (green) and shrubs (orange) – the darker the shade, the higher the coverage. The maps show the coverage across the study area, and the histograms show the frequency of different levels of coverage.

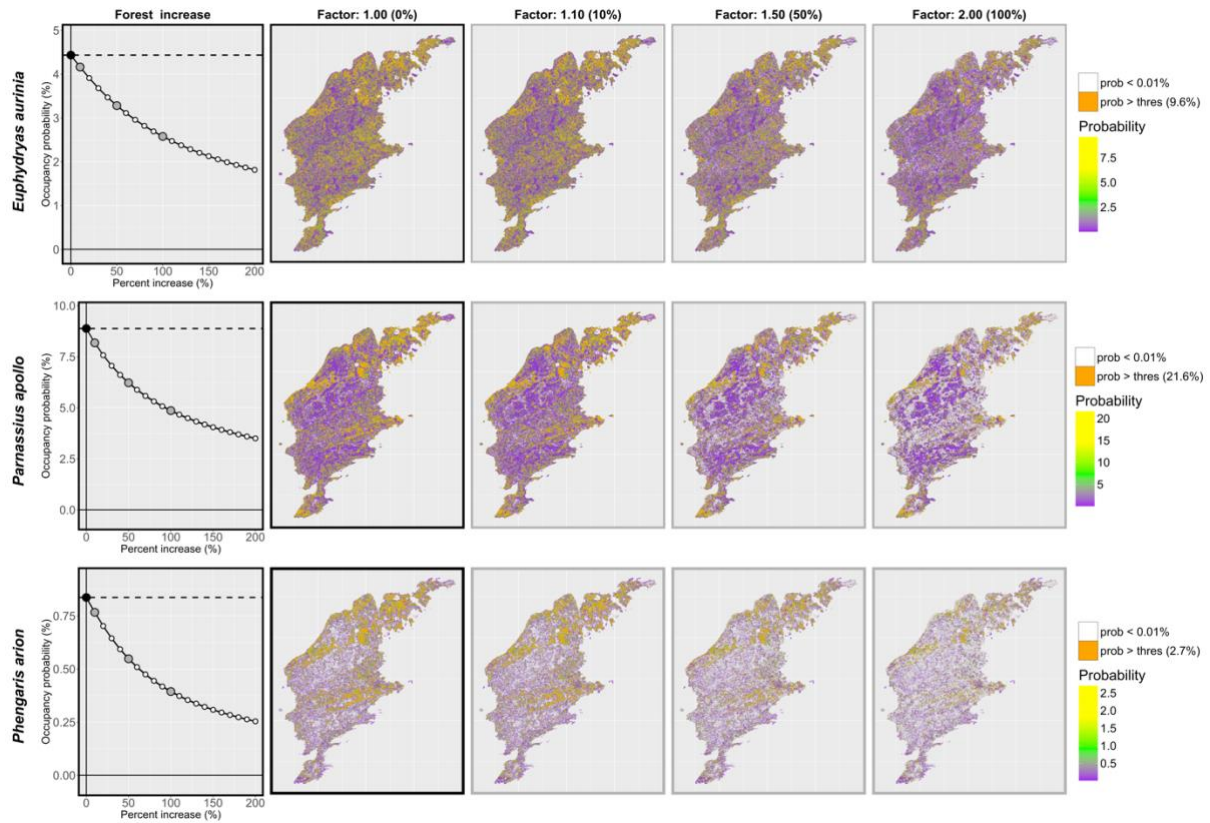


Fig. S4 Predicted occupancy probabilities for the three study species given different scenarios of land cover change with continued forest succession. The figure shows predicted probabilities for the Marsh Fritillary (*Euphydryas aurinia*, top row), Apollo (*Parnassius apollo*, middle row), and Large Blue (*Phengaris arion*, bottom row) across the island of Gotland based on ecological niche modelling. The leftmost panel in each row illustrates changes in butterfly occupancy probability in different scenarios of continued forest succession (forest cover increase) ranging from a 0 % increase (current land cover) to a 200 % increase in 10 % increments. Dashed lines represent the "no-difference-line", circles represent the mean occupancy-- probability, black lines depict fitted lines with 95 % CI. Filled circles represent the scenarios visualised in the maps - black circles the current land cover, and grey the three scenarios of continued forest succession (panel three: 10 % increase, panel four: 50 % increase, and panel five: 100 % increase), visualised in the top row of **Fig. S3**. Colours in the maps indicate the probability of butterfly occupancy (in %), values less than 0.01 % are indicated in white, and values exceeding the species-specific threshold are indicated in orange. Note that the scale for all three species differ.

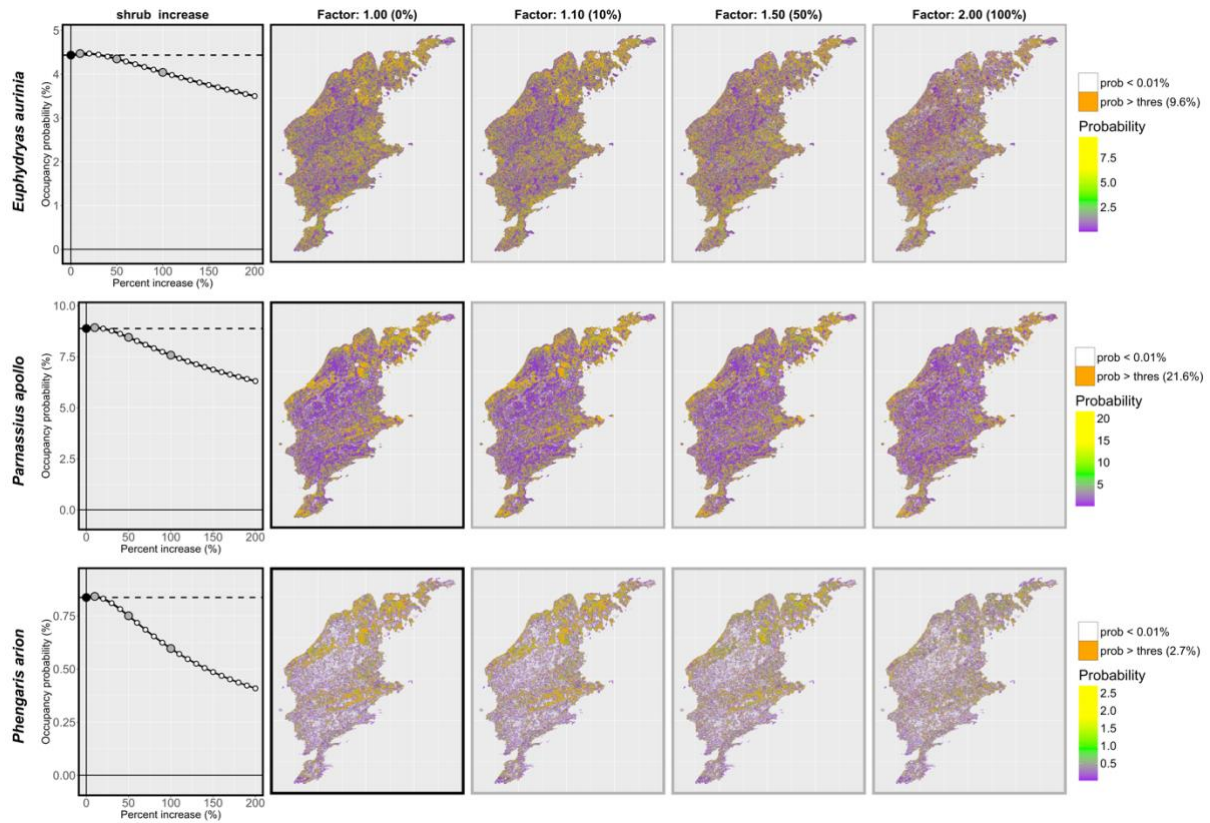


Fig. S5 Predicted occupancy probabilities for the three study species given different scenarios of land cover change with continued shrub overgrowing. The figure shows predicted probabilities for the Marsh Fritillary (*Euphydryas aurinia*, top row), Apollo (*Parnassius apollo*, middle row), and Large Blue (*Phengaris arion*, bottom row) across the island of Gotland based on ecological niche modelling. The leftmost panel in each row illustrates changes in butterfly occupancy probability in different scenarios of continued shrub overgrowing (shrubs cover increase) ranging from a 0 % increase (current land cover) to a 200 % increase in 10 % increments. Dashed lines represent the "no-difference-line", circles represent the mean occupancy-- probability, black lines depict fitted lines with 95 % CI. Filled circles represent the scenarios visualised in the maps to the right - black circles the current land cover, and grey the three scenarios of continued forest succession (panel three: 10 % increase, panel four: 50 % increase, and panel five: 100 % increase), visualised in the middle row of **Fig. S3**. Colours in the maps indicate the probability of butterfly occupancy (in %), values less than 0.01 % are indicated by white, and values exceeding the species-specific threshold are indicated by orange. Note that the scale for all three species differ.

Table S1 Results from model comparison of models testing for associations between butterfly occupancy and environmental variables. The table presents the results of model comparisons for each of the three study species: the Marsh Fritillary (*Euphydryas aurinia*), Apollo (*Parnassius apollo*), and Large Blue (*Phengaris arion*). Each comparison included some or all of five different generalised linear models (GLMs): the null model, full model, optimised ('best') model, linear model, and optimised ('best') linear model. Occupancy (presence/absence) was introduced as a binomial, presence/absence (1/0) response variable for all models. The full model and linear model included all the five environmental variables that passed the Varclus analysis (forest cover, shrub cover, non-vegetated open land, habitat heterogeneity, and ground moisture; **Fig. S2**) - the full model included both linear and squared terms and the linear model included only linear terms. The optimised models were species-specific and constructed based on the variable selection procedures (for details, see Materials and Methods).

Model	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	P-value
null aurinia	1	3395.52	3402.31	-1696.76	3393.52			
full aurinia	11	2705.51	2780.24	-1341.75	2683.51	3.43	2	
best aurinia	9	2704.93	2766.08	-1343.47	2686.93	290.55	3	<0.001
linear aurinia	6	2989.48	3030.24	-1488.74	2977.48	0.24	1	
best linear aurinia	5	2987.71	3021.68	-1488.86	2977.71	415.8	4	<0.001
null apollo	1	6009.08	6015.85	-3003.54	6007.08			
full apollo	11	3912.37	3986.82	-1945.19	3890.37	475.59	5	<0.001
linear apollo	6	4377.96	4418.57	-2182.98	4365.96	1641.12	5	<0.001
null arion	1	1258.39	1265.16	-628.2	1256.39			
full arion	11	1077.27	1151.74	-527.64	1055.27	0.87	2	
best arion	9	1074.14	1135.07	-528.07	1056.14	82.68	3	<0.001
linear arion	6	1150.82	1191.44	-569.41	1138.82	1.67	2	
best linear arion	4	1148.49	1175.57	-570.24	1140.49	115.91	3	<0.001

Df = degrees of freedom, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, logLik = log-Likelihood, Chisq = Chi-square, Chi Df = degrees of freedom for the Chi-square test

Table S2 Results from the optimal threshold performance evaluation. The table presents a comprehensive comparison of the performance of seven threshold methods for the ENMs for the three study species: the Marsh Fritillary (*Euphydryas aurinia*), Apollo (*Parnassius apollo*), and Large Blue (*Phengaris arion*). The methods are: 1) 'Default': threshold fixed at 0.5, 2) 'Sens=Spec': the threshold where sensitivity equals specificity, 3) 'MaxSens+Spec': the threshold that maximises the sum of sensitivity and specificity, 4) 'MaxKappa': the threshold that maximises Cohen's Kappa, a measure of classification accuracy, 5) 'MaxPCC': the threshold that maximises PCC (percent correctly classified), 6) 'PredPrev=Obs': the threshold where the predicted prevalence equals the observed prevalence, and 7) 'ObsPrev': the observed prevalence of each species in the dataset. Assignment accuracy: the percentage of correctly assigned grids (true positives and negatives) based on the field survey data. Sensitivity: the ability of the model to correctly identify true positives. Specificity: the ability of the model to correctly identify true negatives. AUC: area under the curve. TSS: true skill statistic (sensitivity + specificity -1).

Method	Species	Default	Sens=Spec	MaxSens+Spec	MaxKappa	MaxPCC	PredPrev=Obs	ObsPrev
Assignment accuracy	<i>aurinia</i>	0.93	0.75	0.76	0.90	0.93	0.91	0.71
	<i>apollo</i>	0.86	0.81	0.81	0.85	0.87	0.86	0.77
	<i>arion</i>	0.98	0.75	0.71	0.91	0.98	0.96	0.66
Sensitivity	<i>aurinia</i>	0.00	0.75	0.79	0.49	0.10	0.36	0.81
	<i>apollo</i>	0.40	0.81	0.86	0.69	0.41	0.59	0.86
	<i>arion</i>	0.00	0.77	0.84	0.40	0.01	0.07	0.84
Specificity	<i>aurinia</i>	1.00	0.75	0.75	0.93	0.99	0.95	0.70
	<i>apollo</i>	0.96	0.81	0.80	0.89	0.96	0.91	0.76
	<i>arion</i>	1.00	0.75	0.71	0.92	1.00	0.98	0.66
AUC	<i>aurinia</i>	0.84	0.84	0.84	0.84	0.84	0.84	0.84
	<i>apollo</i>	0.89	0.89	0.89	0.89	0.89	0.89	0.89
	<i>arion</i>	0.82	0.82	0.82	0.82	0.82	0.82	0.82
TSS	<i>aurinia</i>	0.00	0.50	0.54	0.42	0.09	0.31	0.51
	<i>apollo</i>	0.36	0.62	0.66	0.58	0.37	0.50	0.62
	<i>arion</i>	0.00	0.52	0.55	0.32	0.01	0.05	0.50

Table S3 Species-specific relationships of butterfly occupancy with environmental variables. The table presents the results (output from ‘Anova’ function in the car package) from the best-fit species-specific GLMs for associations between butterfly occupancy and environmental variables. A separate GLM was run for each of the three study species: the Marsh Fritillary (*Euphydryas aurinia*), Apollo (*Parnassius apollo*), and Large Blue (*Phengaris arion*), in which some or all of the five environmental variables (forest cover, shrub cover, coverage of non-vegetated open land, habitat heterogeneity, and ground moisture) were included, based on the variable selection procedures (for details see Materials and Methods). Occupancy (presence/absence) was introduced as a binomial, presence/absence (1/0) response variable, and each surveyed hectare-grid contributed one observation ($N_{\text{aurinia}} = 6,594$; $N_{\text{apollo}} = 6,425$; $N_{\text{arion}} = 6,437$). For more details and parameter estimates, see **Table 1**.

Species	Predictor	Chisq	Df	P-value	
<i>Euphydryas aurinia</i>	Forest cover, lin. and sq.	61.97	2	< 0.001	***
	Shrub cover, lin. and sq.	102.82	2	< 0.001	***
	Heterogeneity, lin. and sq.	16.37	2	< 0.001	***
	Ground moisture, lin. and sq.	137.24	2	< 0.001	***
<i>Parnassius apollo</i>	Forest cover, lin. and sq.	315.97	2	< 0.001	***
	Shrub cover, lin. and sq.	192.68	2	< 0.001	***
	Open land, lin. and sq.	116.79	2	< 0.001	***
	Heterogeneity, lin. and sq.	17.26	2	< 0.001	***
	Ground moisture, lin. and sq.	198.67	2	< 0.001	***
<i>Phengaris arion</i>	Forest cover, lin. and sq.	19.01	2	< 0.001	***
	Shrub cover, lin. and sq.	14.71	2	< 0.001	***
	Heterogeneity, lin. and sq.	4.39	2	0.11	
	Ground moisture, lin. and sq.	23.79	2	< 0.001	***

Chisq = Chi-square (X²), Df = degrees of freedom. Asterisks indicate statistical significance,

° 0.05 < P < 0.1, * P < 0.05, ** P < 0.01, and *** P < 0.001

References

- Graham, MH (2003) Confronting multicollinearity in ecological multiple regression. *Journal* 84:Issue 2809-2815 <https://www.jstor.org/stable/3449952>
- Harrell, FEJ (2020) Hmisc: Harrell miscellaneous. R package version 4.4-2. <https://CRAN.R-project.org/package=Hmisc>