

Ecography

ECOG-04886

Kass, J. M., Anderson, R. P., Espinosa-Lucas, A., Juárez-Jaimes, V., Martínez-Salas, E., Botello, F., Tavera, G., Flores-Martínez, J. J. and Sánchez-Cordero, V. 2019. Biotic predictors with phenological information improve range estimates for migrating monarch butterflies in Mexico. – Ecography doi: 10.1111/ecog.04886

Supplementary material

Appendix 1: Supplemental Information

1. Data preparation of occurrences

Data for both plants and monarchs were prepared for modeling in similar ways. To reduce the effects of sampling bias due to spatially clustered localities (Veloz 2009), we spatially thinned each species' occurrence records by 10 km using the R package `spThin` (Aiello-Lammens et al. 2015). We ran the thinning algorithm for 100 iterations in order to maximize the number of records retained. To avoid including model training areas inaccessible to species because of dispersal constraints (Peterson et al. 2011), and also to include only those areas proximal to sampled records, we defined study extents for all species by delineating minimum convex polygons (MCPs) around occurrence localities buffered by 100 km. For monarchs, the CONANP occurrence localities alone were used to delineate the MCP, as we had high confidence that this dataset accurately represented the migrating eastern population. We sampled 25,000 background points within MCPs—more points than are typically sampled— as inadequate background sampling can lead to artifactually truncated environmental responses (Guevara et al. 2018).

2. SDM methodology details: plants and monarchs

We tuned SDMs using Maxent v3.4.1 (Phillips et al. 2017) over a range of model settings with the R package `ENMeval` v0.3 (Muscarella et al. 2014). Maxent performs internal variable selection using a form of complexity penalization called regularization: feature classes control the various shapes allowed for the modeled response, affecting how complex the response can be, while increasing the regularization multiplier enforces simpler models with fewer parameters (Phillips and Dudík 2008, Merow et al. 2013). We explored combinations of simple and complex feature classes: linear (L), quadratic (Q), and hinge (H), resulting in the combinations L, LQ, H, and LQH. We specifically chose to exclude the feature classes “product” and “threshold” from the analysis, as their omission results in models that are easier to interpret, or do not show better performance than hinge features, respectively (Phillips et al. 2017). We also explored a range of regularization multipliers that increasingly penalized complexity of fit: the values 1 through 5 with a step value of 0.5, resulting in 40 candidate models for each species (4 combinations of feature classes x 10 regularization multipliers).

We selected optimal model settings for plants based on measures of discrimination and omission on spatially withheld occurrence data instead of relying on information criteria, as we were most interested in optimizing model performance for plants and not in comparing between models representing different hypotheses. Spatial cross validation reduces the effects of spatial autocorrelation in the occurrence data, which avoids overly optimistic model performance due to spatial dependence between localities that can occur through random cross validation (Roberts et al. 2017).

We used 4-fold spatial partitions delineated by longitude and latitude positioned to balance the number of localities in each fold (“block” partition method; Muscarella et al. 2014). We iteratively built models for each combination of feature class and regularization multiplier on the occurrence and background values from three of the four folds and predicted the occurrences from the withheld fold, then averaged model performance over the four folds. Before selecting models with optimal settings, we ensured that overly complex models would not be selected by first filtering out candidate models for which AIC cannot be calculated (i.e., those that had more non-zero coefficients than occurrence localities). We evaluated two performance statistics that rate performance on withheld data sequentially to select models. We first selected the models with the lowest average omission rate using the 10-percentile training presence (OR10), which is the percentage of test occurrences with predicted suitabilities below the 10 percentile of training values. When multiple models had identical OR10, we broke ties by choosing the model with the highest average test AUC (area under the curve of the receiving operator characteristic; Fielding and

Bell 1997) over the withheld folds (AUC_{test}), a standard threshold-independent measure of discriminatory ability on withheld data for SDMs (Peterson et al. 2011). Although valid criticism exists concerning the use of AUC for rating model accuracy and comparing values between studies with different extents, variables, or species (Lobo et al. 2008), we use AUC here to compare between models built with the same data and extents that share low omission rates in order to select model settings that lead to high relative performance. We evaluated monarchs SDMs based solely on AICc, as we had the task of comparing between models with different numbers of parameters (see Appendix A5).

3. Species richness estimates for plants

For each plant species, the model with optimal settings was used to predict suitability values across the respective species-specific study extent, resulting in model prediction rasters for each species. We made model predictions using the cloglog output for Maxent v3.4.1, a scaling of the raw Maxent output (for which direct comparisons of species predictions with different study extents cannot be made; Phillips and Dudík 2008) that preserves rank and estimates probability of presence with values between 0 and 1. Transforming to a probability scale enabled us to combine SDM prediction rasters from multiple species and thus estimate species richness over a shared extent (Ferrier and Guisan 2006). Calabrese et al. (2014) found that summing continuous SDM predictions instead of thresholded presence/absence rasters results in more accurate estimates of site-level species richness. Before summing, we masked each prediction raster by the species-specific study extent to exclude predicted areas far from observed data (even if they were highly suitable according to the model).

4. Low-abundance plant species removal analysis

As plant species with low abundance throughout their ranges could potentially bias the richness estimates towards areas that do not necessarily have a high plant population density (which should be more closely associated with monarch suitability than plant species richness, per se), we removed the SDM predictions of low-abundance species before creating the biotic variables and ran a separate set of models to see if this would affect our results. All the tree species were retained, but we removed five *Asclepias* spp. for this comparison: *A. circinalis*, *A. elata*, *A. fournieri*, *A. sperryi*, and *A. tuberosa*.

This analysis resulted in small changes to the stationary models, but not the monthly, as none of the species removed are listed as flowering for September through November in our phenology database. With the exception of the combined stationary model for September, which had the same AICc score as the combined monthly model, the results remained the same as the original analysis (Tbl. S3).

5. Rationale for model selection of monarch SDMs by AICc

Although performance metrics such as AUC can be used to rate relative performance over suites of models with different settings (Radosavljevic and Anderson 2014), comparing models with differing numbers of predictor variables is difficult because complex and overfit models often have artifactually inflated accuracy scores. Although some ecological modeling studies enforce variable reduction to equalize variable numbers across models (Bateman et al. 2012), others use information criteria, particularly AIC, to select models in these cases (Johnson and Omland 2004). SDMs selected by AIC may be more robust to sampling bias and have been shown to be simpler when compared to models selected via a cross-validation approach (Galante et al. 2018). Therefore, information criteria may be a preferable choice to cross validation when comparing between competing sets of predictor variables, though selected models should also be evaluated for their performance on withheld data to confirm their accuracy. Therefore, of the candidate monarch SDMs for each variable set per month, we selected the model with the lowest AICc value (Warren and Seifert 2011).

6. Null model methods

Raes & ter Steege (2007) pioneered the null model approach for SDMs, which involves building models based on localities randomly sampled across the study extent, plotting a null distribution of an evaluation statistic, and then conducting a one-tailed t-test for significance. Regarding evaluation of null SDMs,

Raes & ter Steege (2007) used AUC calculated on the training localities instead of withheld data, and later studies improved on this by using random cross validation (Beale et al. 2008). Bohl et al. (2019) proposed instead evaluating null models on the same withheld data as were employed for testing the real models. We followed this latter approach but made a novel modification to evaluate null models using $k - 1$ spatial block cross validation (using $k = 4$ folds rather than 2). To begin, we assigned the same spatial folds applied to the monarch occurrence localities to every grid cell in the study extent. For each iteration, we randomly sampled n localities across the training folds, where n is equal to the total number of real occurrence localities in these folds, and then evaluated the model on the real monarch occurrences in the withheld fold (also using the same background values). Per month, we built null SDMs for each variable set using the same model settings as those chosen as optimal for the real models. We calculated AUC_{test} and average OR10 for each null model. We repeated this process 1000 times, resulting in distributions of 1000 null test statistics per month/variable set. Finally, we compared the real model evaluation statistics to those of the null distributions to determine significance with $\alpha = 0.05$.

7. Plant SDM results

The following descriptions of Maxent model settings will use a letter-number notation for feature class and regularization multiplier: e.g., linear, quadratic, and hinge feature classes with regularization multiplier 2.5 will be notated as LQH2.5. The number of Maxent model parameters with non-zero coefficients (i.e., lambda weights as described in Phillips and Dudík 2008) will be referred to as "parameters". The simplest models were for *A. circinalis* (LQ2, 6 parameters) and *Juniperus monticola* (LQ4, 4 parameters), while the most complex models were for *A. tuberosa* (H1.5, 147 parameters) and *Taxodium mucronatum* (LQH2, 56 parameters). The models that omitted the fewest test localities were for *A. similis* (OR10 = 0.062) and *Pinus devoniana* (OR10 = 0.028), and those that omitted the most were for *A. fournieri* (0.292) and *Quercus rugosa* (0.268). The average OR10 across all plant species was 0.150 for *Asclepias* spp. and 0.128 for roosting trees. As AUC cannot be directly compared for models trained on different study extents (Jiménez-Valverde 2012), we do not report these statistics here.

8. Null model results

For AUC_{test} , the September models had significantly high scores at $\alpha = 0.05$ (above the 95th percentile of the null distribution), and October combined monthly, November biotic monthly, and November combined monthly were just below this threshold (Fig. S5). All the combined monthly models performed quite well for AUC_{test} compared to the null distributions. In comparison, no models selected via AICc had significantly low scores for average OR10 (i.e., performing better than the lowest 5th percentile), and only October biotic monthly had a candidate model that met this criterion (although October abiotic and October combined monthly were just above the margin). Unlike for AUC_{test} , the combined monthly models performed poorly for average OR10: although the September and October models had scores lower than the 50th percentile (indicating they performed better than 50% of null models, but not 95%), the November model performed worse than this (Fig. S6).

References

- Aiello-Lammens, M. E. et al. 2015. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. - *Ecography* 38: 541–545.
- Bateman, B. L. et al. 2012. Biotic interactions influence the projected distribution of a specialist mammal under climate change. - *Divers. Distrib.* 18: 861–872.
- Beale, C. M. et al. 2008. Opening the climate envelope reveals no macroscale associations with climate in European birds. - *Proc. Natl. Acad. Sci.* 105: 14908–14912.
- Bohl, C. et al. 2019. A new null model approach to quantify performance and significance for ecological niche models of species distributions. - *J. Biogeogr.* 46: 1101–1111.
- Calabrese, J. M. et al. 2014. Stacking species distribution models and adjusting bias by linking them to

- macroecological models. - *Glob. Ecol. Biogeogr.* 23: 99–112.
- Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. - *J. Appl. Ecol.* 43: 393–404.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. - *Environ. Conserv.* 24: 38–49.
- Galante, P. J. et al. 2018. The challenge of modeling niches and distributions for data-poor species: a comprehensive approach to model complexity. - *Ecography* 41: 726–736.
- Guevara, L. et al. 2018. Toward ecologically realistic predictions of species distributions: A cross-time example from tropical montane cloud forests. - *Glob. Chang. Biol.* 24: 1511–1522.
- Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. - *Glob. Ecol. Biogeogr.* 21: 498–507.
- Johnson, J. B. and Omland, K. S. 2004. Model selection in ecology and evolution. - *Trends Ecol. Evol.* 19: 101–108.
- Lobo, J. M. et al. 2008. AUC: A misleading measure of the performance of predictive distribution models. - *Glob. Ecol. Biogeogr.* 17: 145–151.
- Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. - *Ecography* 36: 1058–1069.
- Muscarella, R. et al. 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. - *Methods Ecol. Evol.* 5: 1198–1205.
- Peterson, A. T. et al. 2011. *Ecological niches and geographic distributions (MPB-49)*. - Princeton University Press.
- Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. - *Ecography* 31: 161–175.
- Phillips, S. J. et al. 2017. Opening the black box: an open-source release of Maxent. - *Ecography* 40: 887–893.
- Radosavljevic, A. and Anderson, R. P. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. - *J. Biogeogr.* 41: 629–643.
- Raes, N. and ter Steege, H. 2007. A null-model for significance testing of presence-only species distribution models. - *Ecography* 30: 727–736.
- Roberts, D. R. et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. - *Ecography* 40: 913–929.
- Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. - *J. Biogeogr.* 36: 2290–2299.
- Warren, D. L. and Seifert, S. N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. - *Ecol. Appl.* 21: 335–342.

Appendix 2: Supplementary Tables and Figures

Table A1. All plant species (nectar plants and roosting trees) initially considered for analysis, with sample size of occurrence localities before and after spatial thinning by 10 km. Those with sample sizes >15 for occurrences in Mexico after thinning (light gray) were retained for analysis.

species	no. samples	no. thinned samples
<i>Asclepias asperula</i>	483	282
<i>A. auriculata</i>	246	91
<i>A. brachystephana</i>	296	127
<i>A. circinalis</i>	52	18
<i>A. coulteri</i>	83	22
<i>A. curassavica</i>	3776	1062
<i>A. elata</i>	114	54
<i>A. engelmanniana</i>	132	
<i>A. fournieri</i>	105	27
<i>A. glaucescens</i>	676	175
<i>A. jaliscana</i>	326	94
<i>A. linaria</i>	2104	691
<i>A. mexicana</i>	190	52
<i>A. oenotheroides</i>	581	207
<i>A. otarioides</i>	91	47
<i>A. ovata</i>	314	100
<i>A. pellucida</i>	140	47
<i>A. similis</i>	275	64
<i>A. sperryi</i>	59	34
<i>A. subverticillata</i>	463	278
<i>A. texana</i>	35	
<i>A. tuberosa</i>	854	492
<i>A. verticillata</i>	571	
<i>A. viridis</i>	299	
<i>A. virentii</i>	22	
<i>Abies religiosa</i>	160	58
<i>Carya illinoensis</i>	259	136
<i>Cupressus lindleyi</i>	59	33
<i>Juglans hirsuta</i>	19	
<i>Juglans major</i>	221	129
<i>Juglans microcarpa</i>	92	
<i>Juglans mollis</i>	105	44
<i>Juglans pyriformis</i>	48	26
<i>Juniperus deppeana</i>	671	273
<i>Juniperus monticola</i>	101	25
<i>Pinus ayacahuite</i>	199	73
<i>Pinus devoniana</i>	133	34
<i>Pinus hartwegii</i>	526	49
<i>Pinus oocarpa</i>	241	90
<i>Pinus pseudostrobus</i>	628	144
<i>Pinus rudis</i>	40	22
<i>Pinus teocote</i>	331	104
<i>Quercus acutifolia</i>	148	91
<i>Quercus candicans</i>	30	17
<i>Quercus castanea</i>	489	204
<i>Quercus crassifolia</i>	438	182
<i>Quercus laurina</i>	359	155
<i>Quercus obtusata</i>	419	173

<i>Quercus rugosa</i>	622	226
<i>Quercus salicifolia</i>	35	22
<i>Taxodium distichum</i>	286	
<i>Taxodium mucronatum</i>	273	146

Table A2. Maxent SDM settings chosen as optimal for *Asclepias* spp. and trees and associated performance statistics. Settings shown are feature class combinations (features) and regularization multiplier (rm). Statistics shown are AUC calculated on training data (AUC_{train}) and averaged over testing data (AUC_{test}), omission rates for 10 percentile training values (OR10), and the number of non-zero coefficients (nparam).

species	features	rm	AUC _{train}	AUC _{test}	OR10	nparam
<i>Asclepias asperula</i>	L	3	0.850	0.819	0.153	12
<i>A. auriculata</i>	H	4.5	0.943	0.918	0.134	29
<i>A. brachystephana</i>	H	5	0.930	0.884	0.118	33
<i>A. circinalis</i>	LQ	2	0.973	0.968	0.100	6
<i>A. coulteri</i>	LQH	2	0.993	0.966	0.142	15
<i>A. curassavica</i>	LQ	0.5	0.956	0.957	0.130	21
<i>A. elata</i>	L	2.5	0.858	0.766	0.074	11
<i>A. fournieri</i>	L	2	0.916	0.848	0.292	9
<i>A. glaucescens</i>	LQ	3	0.964	0.957	0.115	11
<i>A. jaliscana</i>	L	5	0.902	0.866	0.235	7
<i>A. linaria</i>	L	5	0.914	0.900	0.145	8
<i>A. mexicana</i>	L	2	0.953	0.924	0.212	9
<i>A. oenotheroides</i>	L	2.5	0.888	0.867	0.140	14
<i>A. otarioides</i>	L	0.5	0.949	0.922	0.110	13
<i>A. ovata</i>	LQ	0.5	0.953	0.909	0.250	27
<i>A. pellucida</i>	L	5	0.915	0.872	0.106	8
<i>A. similis</i>	LQ	4	0.947	0.935	0.062	12
<i>A. sperryi</i>	L	5	0.906	0.875	0.149	7
<i>A. subverticillata</i>	H	4.5	0.917	0.890	0.166	69
<i>A. tuberosa</i>	H	1.5	0.916	0.864	0.173	147
<i>Abies religiosa</i>	LQ	5	0.969	0.953	0.106	8
<i>Carya illinoensis</i>	L	0.5	0.851	0.752	0.213	19
<i>Cupressus lindleyi</i>	LQ	4.5	0.946	0.933	0.094	8
<i>Juglans major</i>	H	5	0.954	0.951	0.086	62
<i>Juglans mollis</i>	LQ	5	0.960	0.928	0.114	11
<i>Juglans pyriformis</i>	LQ	1	0.956	0.932	0.036	11
<i>Juniperus deppeana</i>	LQ	4.5	0.940	0.914	0.129	12
<i>Juniperus monticola</i>	LQ	4	0.984	0.978	0.113	4
<i>Pinus ayacahuite</i>	L	3	0.962	0.942	0.150	11
<i>P. devoniana</i>	LQ	3.5	0.975	0.976	0.028	8
<i>P. hartwegii</i>	H	5	0.983	0.980	0.061	11
<i>P. oocarpa</i>	L	2.5	0.915	0.852	0.218	10
<i>P. pseudostrobus</i>	L	4	0.902	0.870	0.132	12
<i>P. rudis</i>	H	3.5	0.922	0.913	0.092	14
<i>P. teocote</i>	L	5	0.934	0.915	0.183	9
<i>Quercus acutifolia</i>	L	4.5	0.925	0.919	0.090	7
<i>Q. candicans</i>	LQH	3.5	0.989	0.979	0.125	6
<i>Q. castanea</i>	LQ	5	0.952	0.941	0.127	8
<i>Q. crassifolia</i>	LQ	5	0.940	0.892	0.241	12
<i>Q. laurina</i>	LQ	4.5	0.953	0.941	0.163	10
<i>Q. obtusata</i>	H	5	0.943	0.932	0.139	27
<i>Q. rugosa</i>	L	2.5	0.942	0.896	0.268	15
<i>Q. salicifolia</i>	LQH	2	0.951	0.931	0.092	12
<i>Taxodium mucronatum</i>	LQH	2	0.964	0.947	0.075	56

Table A3. Maxent SDM settings chosen as optimal for monarchs by month and variable group and associated performance statistics, after removal of low-abundance *Asclepias* SDM predictions from the estimated richness variable. Only the stationary versions for the biotic and combined models are shown, as the monthly versions did not include these species to begin with. Settings shown are feature class combinations (features) and regularization multiplier (rm). Statistics shown are AUC calculated on training data (AUC_{train}) and averaged over testing data (AUC_{test}), omission rates for 10 percentile (OR10) training values, delta AICc (based on the lowest AICc in Tbl. 2), and the number of non-zero coefficients (nparam).

month	group	features	rm	AUC_{train}	AUC_{test}	OR10	delta AICc	nparam
September	biotic stationary	L	0.5	0.693	0.700	0.109	5	2
	combined stationary	L	1	0.715	0.729	0.109	0	4
October	biotic stationary	LQ	0.5	0.568	0.574	0.178	23	4
	combined stationary	LQ	0.5	0.637	0.641	0.070	8	9
November	biotic stationary	L	1	0.654	0.644	0.198	41	1
	combined stationary	LQH	2	0.767	0.635	0.233	24	21

Table A4. Null model result summaries for abiotic, biotic monthly, and combined monthly models, reported by evaluation statistic (AUC_{test} , OR10).

statistic	month	group	min	50%	95%	max
AUC _{test}	September	abiotic	0.285	0.513	0.655	0.745
		biotic monthly	0.245	0.506	0.693	0.747
		combined monthly	0.264	0.508	0.659	0.765
	October	abiotic	0.277	0.493	0.630	0.729
		biotic monthly	0.279	0.492	0.643	0.735
		combined monthly	0.277	0.492	0.622	0.733
	November	abiotic	0.306	0.509	0.621	0.703
		biotic monthly	0.246	0.493	0.681	0.756
		combined monthly	0.241	0.501	0.671	0.773
OR10	September	abiotic	0	0.156	0.375	0.560
		biotic monthly	0	0.156	0.422	0.562
		combined monthly	0	0.188	0.432	0.672
	October	abiotic	0	0.152	0.358	0.517
		biotic monthly	0	0.189	0.387	0.533
		combined monthly	0	0.206	0.418	0.606
	November	abiotic	0	0.172	0.328	0.509
		biotic monthly	0	0.172	0.371	0.603
		combined monthly	0	0.172	0.405	0.638

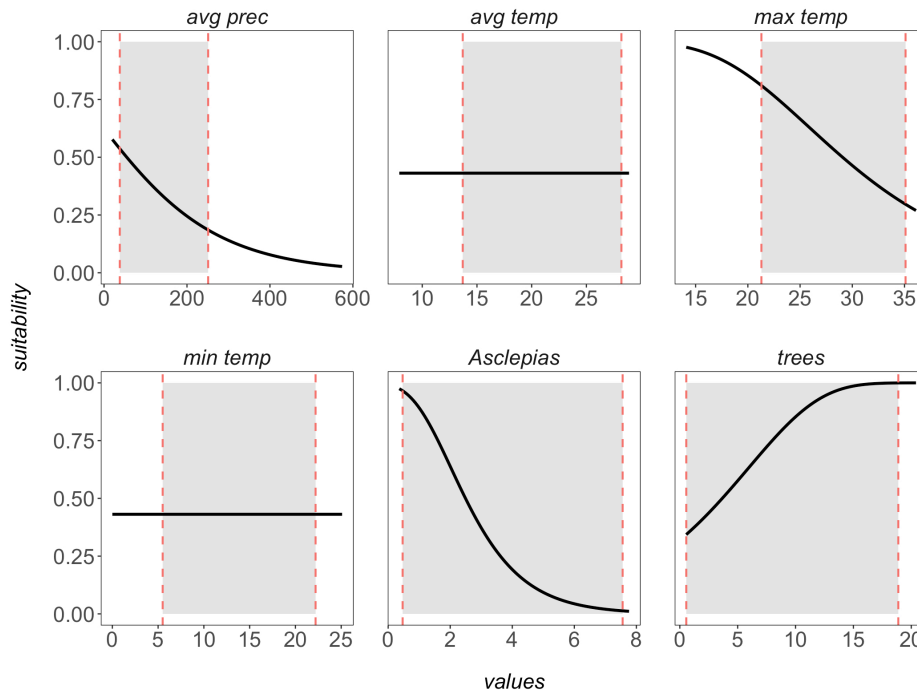


Figure A1. Response curves for the September combined monthly model. Dotted red lines delineate the minimum and maximum values in the occurrence data used for model training, and the gray box represents the range of each variable represented in the occurrence data.

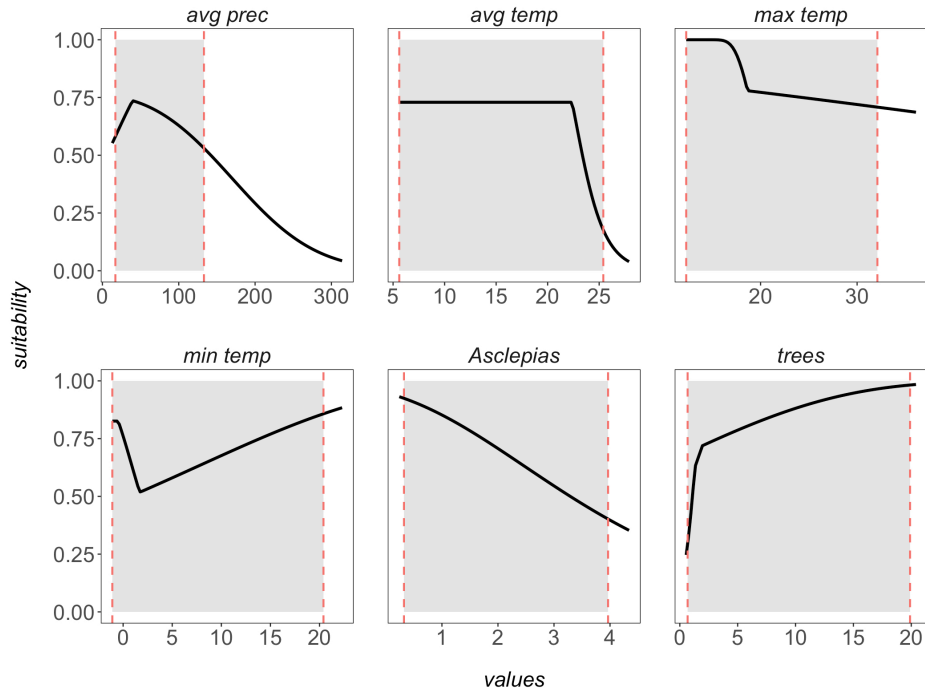


Figure A2. Response curves for the October combined monthly model. Dotted red lines delineate the minimum and maximum values in the occurrence data used for model training, and the gray box represents the range of each variable represented in the occurrence data.

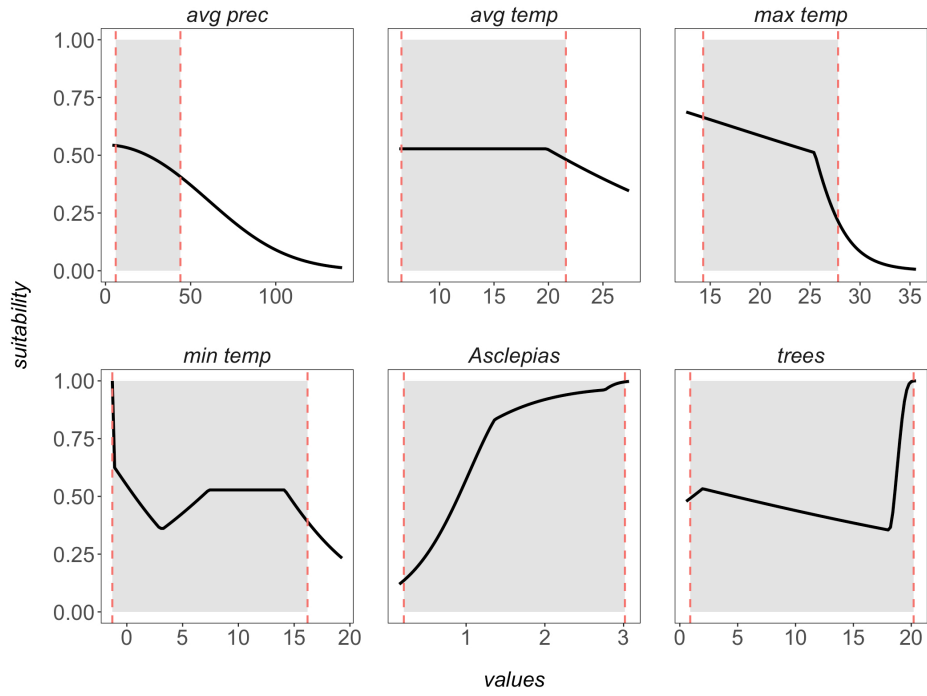


Figure A3. Response curves for the November combined monthly model. Dotted red lines delineate the minimum and maximum values in the occurrence data used for model training, and the gray box represents the range of each variable represented in the occurrence data.

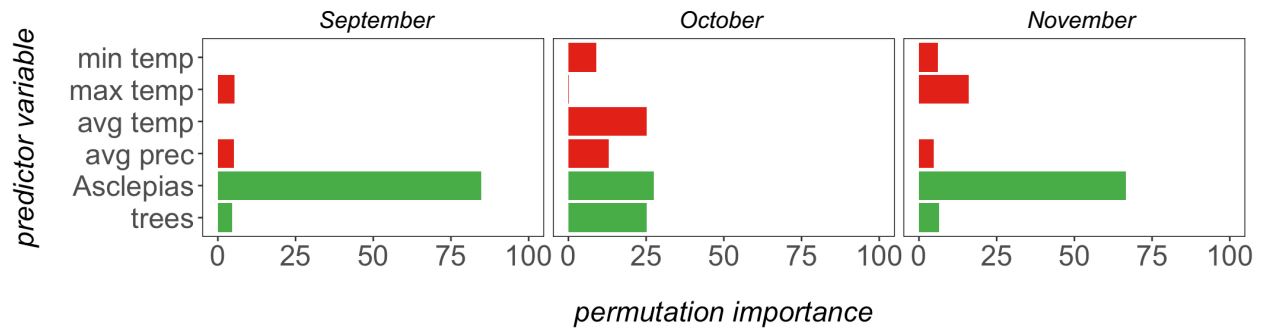


Figure A4. Permutation importance percentages for predictor variables in monarch combined monthly SDMs with combined abiotic (red) and biotic (green) variables that considered phenology.

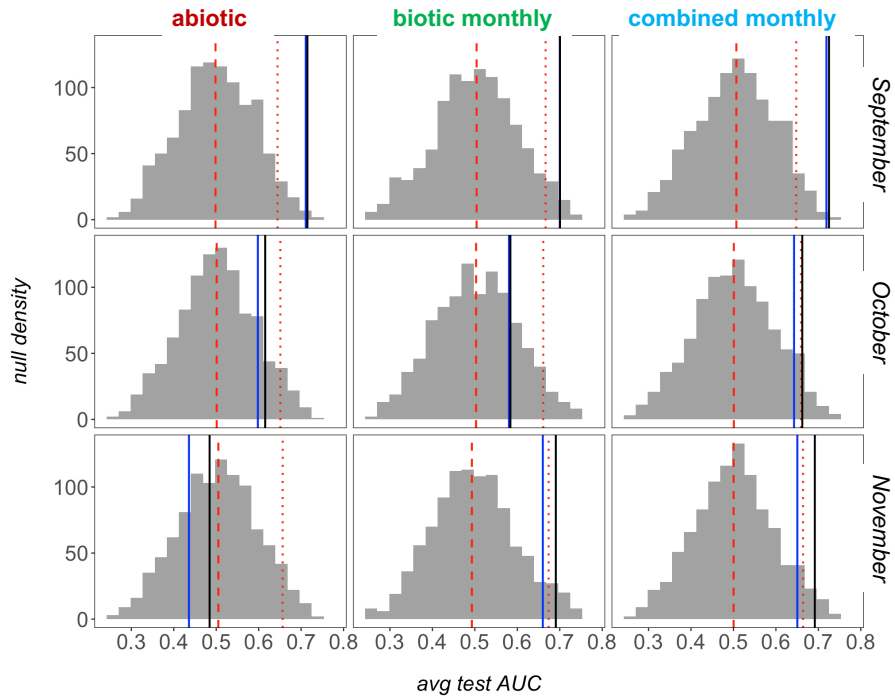


Figure A5. Null model results for average AUC_{test} using the same settings and spatial blocks as the real models. Shown are the 50th percentile (dashed red line) and 95th percentile (dotted red line) of the null distribution, along with the real model value (solid blue line) and the value of the best performing model across all explored settings (solid black line).

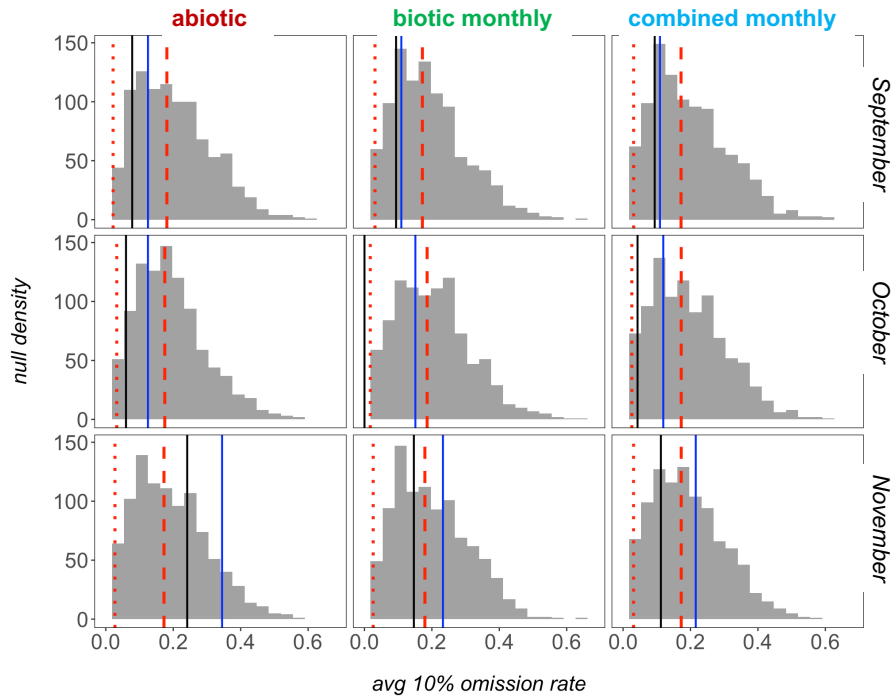


Figure A6. Null model results for average OR10 using the same settings and spatial blocks as the real models. Shown are the 50th percentile (dashed red line) and 95th percentile (dotted red line) of the null distribution, along with the real model value (solid blue line) and the value of the best performing model across all explored settings (solid black line).

Appendix 3: R function for running null species distribution models for multiple specified Maxent settings using spatial block partitioning

Note: The functionality in this script, which was used for the null model analysis in this study, has since been incorporated into the nullENM package (<https://github.com/ndimhypervol/nullENM>), referenced in Bohl et al. (2019).

```
# Run null species distribution models for multiple specified Maxent settings
# with spatial block partitioning
#
# @param occs Data frame of species occurrences (longitude, latitude)
# @param bg Data frame of background coordinates (longitude, latitude)
# @param envs RasterStack of environmental variables
# @param optAll List of rows in ENMevaluate object results attribute
# corresponding to the optimal model settings per variable group; these
# settings will be used to make null species distribution models
# @param samps Numeric for number of repetitions
# @return List of 1) test AUC values, 2) 10 percentile omission rate values,
# 3) the thresholds used to calculate the omission rates
# @note Based on code from randomNull() from dismo package v1.1-4
# @export

nullSDMs <- function(occs, bg, envs, optAll, samps) {
  # divide all grid cells in study extent into same partition groups
  # as the real occurrence data
  envs.xy <- raster::rasterToPoints(envs[[1]])
  envs.parts <- ENMeval::get.block(occ=occs, bg.coords=envs.xy[,1:2])
  # extract the cell numbers for the raster
  envs.xy.cellNum <- raster::extract(envs[[1]], envs.xy[,1:2], cellnumbers=TRUE)
  # make new rasters that leave out (LO) spatial fold
  envs.grp <- envs[[1]][[1]]
  envs.grp[envs.xy.cellNum[,1]] <- envs.parts$bg.grp
  envsLO1 <- envsLO2 <- envsLO3 <- envsLO4 <- envs.grp
  envsLO1[envsLO1 == 1] <- NA
  envsLO2[envsLO2 == 2] <- NA
  envsLO3[envsLO3 == 3] <- NA
  envsLO4[envsLO4 == 4] <- NA
  envsLO <- stack(envsLO1, envsLO2, envsLO3, envsLO4)

  # get the partition groups for occurrences and background used to train the real model
  obs.parts <- ENMeval::get.block(occ=occs, bg.coords=bg)
  # find how many occs are in each group
  occ.grp.tbl <- table(obs.parts$occ.grp)
  # list to hold the average test AUCs for each variable group
  aucTestAvg <- replicate(length(envs), numeric(samps), simplify=FALSE)
  names(aucTestAvg) <- names(envs)
  orTestAvg <- replicate(length(envs), numeric(samps), simplify=FALSE)
  names(orTestAvg) <- names(envs)
  thrs <- replicate(length(envs), matrix(nrow=4, ncol=samps), simplify=FALSE)
  # list of maxent arguments for each variable group for maxent.jar
  mxArgsAll <- lapply(optAll, function(x) ENMeval::make.args(x$rm, x$features)[[1]])

  # iterate null models
  for(i in 1:samps) {
    # list to hold test AUC for all 4 folds for each variable group
    aucs <- replicate(length(envs), numeric(4), simplify=FALSE)
```

```

or10s <- replicate(length(envs), numeric(4), simplify=FALSE)
thr <- replicate(length(envs), numeric(4), simplify=FALSE)
# perform cross validation on 4 spatial folds
for(k in 1:4) {
  # assign training occurrences as all folds but k
  null.occs.train <- dismo::randomPoints(envsLO[[k]], sum(occ.grp.tbl[-k]))
  # assign training background the same as above (subset the data frame of variable values)
  bg.train <- bg[obs.parts$bg.grp != k,]
  # assign testing occurrences as the real occurrences in the k fold
  occs.test <- occs[obs.parts$occ.grp == k,]
  # make vector of 0's and 1's to differentiate occurrences and background for maxent()
  p <- c(rep(1, nrow(null.occs.train)), rep(0, nrow(bg.train)))
  # repeat for each variable group with same randomly sampled occurrences
  for(e in 1:length(envs)) {
    # extract variable values for training occurrences and testing occurrences
    occs.train.vals <- raster::extract(envs[[e]], null.occs.train, df=TRUE)[-1]
    occs.test.vals <- raster::extract(envs[[e]], occs.test, df=TRUE)[-1]
    # get variable values for all background points
    bg.vals <- raster::extract(envs[[e]], bg, df=TRUE)[-1]
    bg.train.vals <- bg.vals[obs.parts$bg.grp != k,]
    # rbind together training data
    x <- data.frame(rbind(occs.train.vals, bg.train.vals))
    # run model with same fc and regm combination as the optimized real model
    m <- maxent(x, p, args = mxArgsAll[[e]])
    # evaluate model on testing data and full background, and record auc
    aucs[[e]][k] <- evaluate(occs.test.vals, bg.vals, m)@auc

    # get model predictions for training and testing points
    p.train <- dismo::predict(m, occs.train.vals)
    p.test <- dismo::predict(m, occs.test.vals)

    # figure out 90% of total no. of training records
    n90 <- ceiling(nrow(occs.train.vals) * 0.9)
    # calculate 10 percentile omission rate
    train.thr.10 <- rev(sort(p.train))[n90]
    thr[[e]][k] <- train.thr.10
    or10s[[e]][k] <- mean(p.test < train.thr.10)
  }
}

# get mean of test AUCs for each variable group for this iteration
for(v in 1:length(envs)) {
  aucTestAvgs[[v]][i] <- mean(aucs[[v]])
  orTestAvgs[[v]][i] <- mean(or10s[[v]])
  thrs[[v]][i] <- thr[[v]]
}

message("-", appendLF = FALSE)
if(i %% 50 == 0) {
  message(" ", i)
  flush.console()
}
}

message("Null models complete.")

```

```
flush.console()
return(list(aucs=aucTestAvg, ors=orTestAvg, thr=thr))
}
```