Ecography

ECOG-03986
Qiao, H., Feng, X., Escobar, L. E., Peterson, A. T., Soberón, J., Zhu, G. and Papeş, M. 2018. An evaluation of transferability of ecological niche models. – Ecography doi: 10.1111/ecog.03986

Supplementary material

Appendix 1.

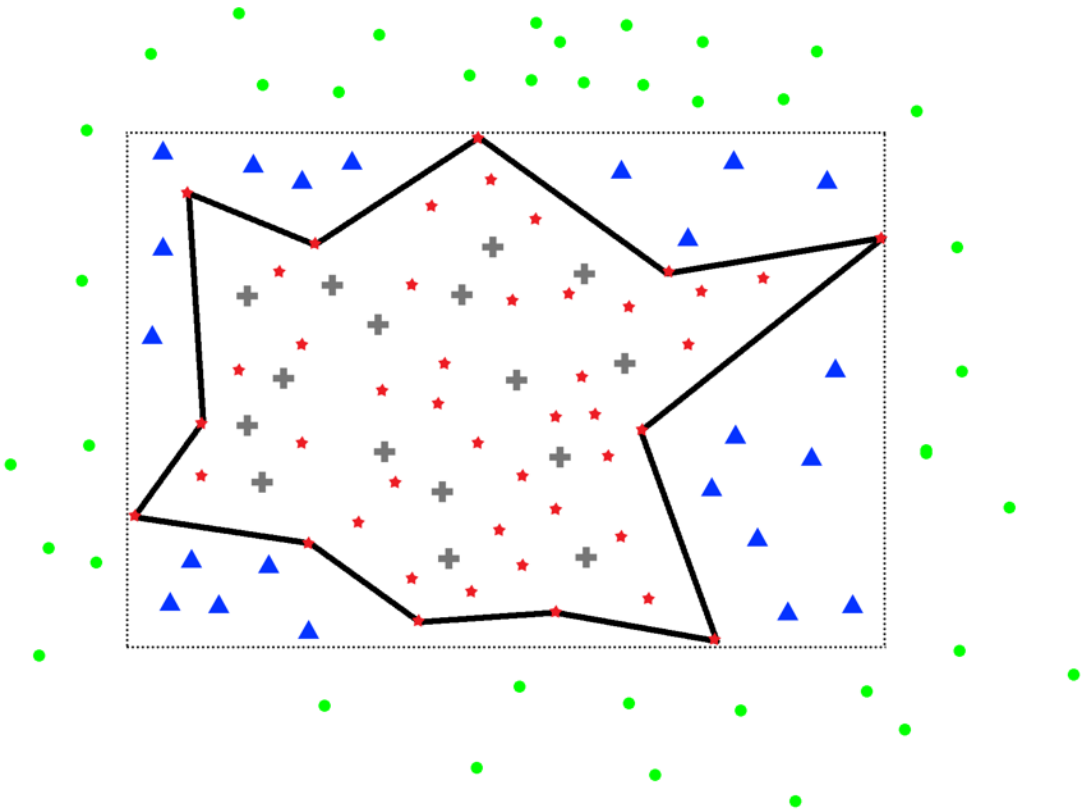


Figure A1. Illustration of types of evaluation data, classified by environmental novelty. The red stars represent the calibration data (presences and background pixels) in a two-dimensional environmental space. The black polygon (which would be a hull in three-dimensional space) represents a conservative estimate of the boundaries of the calibration data; the dotted rectangle represents a broader estimate that shows range limits for the variables. The gray crosses represent the category of overlapping (evaluation data within the polygon enclosing the calibration data). The blue triangles represent the category of novel-combination (evaluation data outside the polygon but inside the rectangle). The green points represent the category of novel (evaluation data outside the environmental limits of the calibration data).

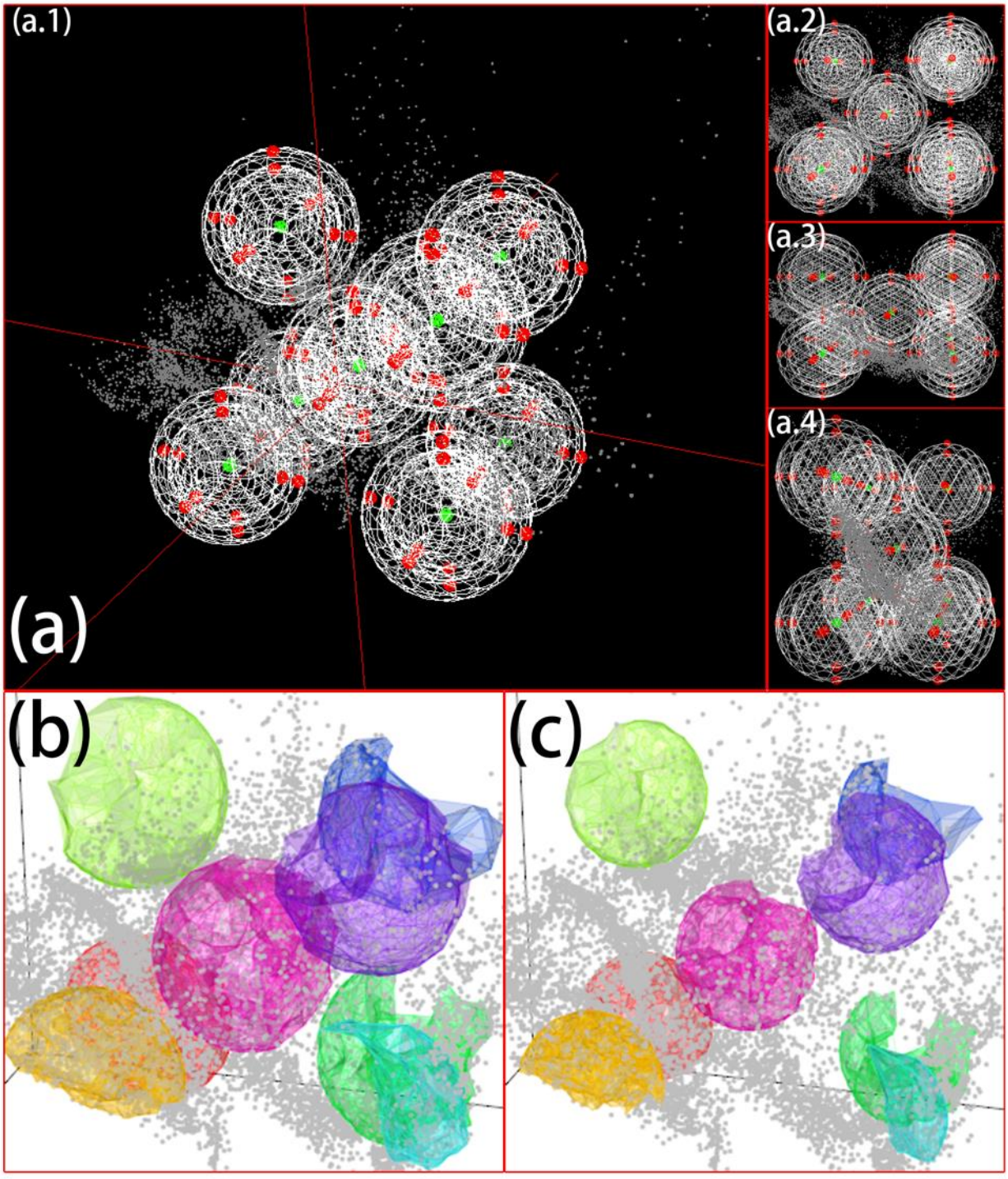
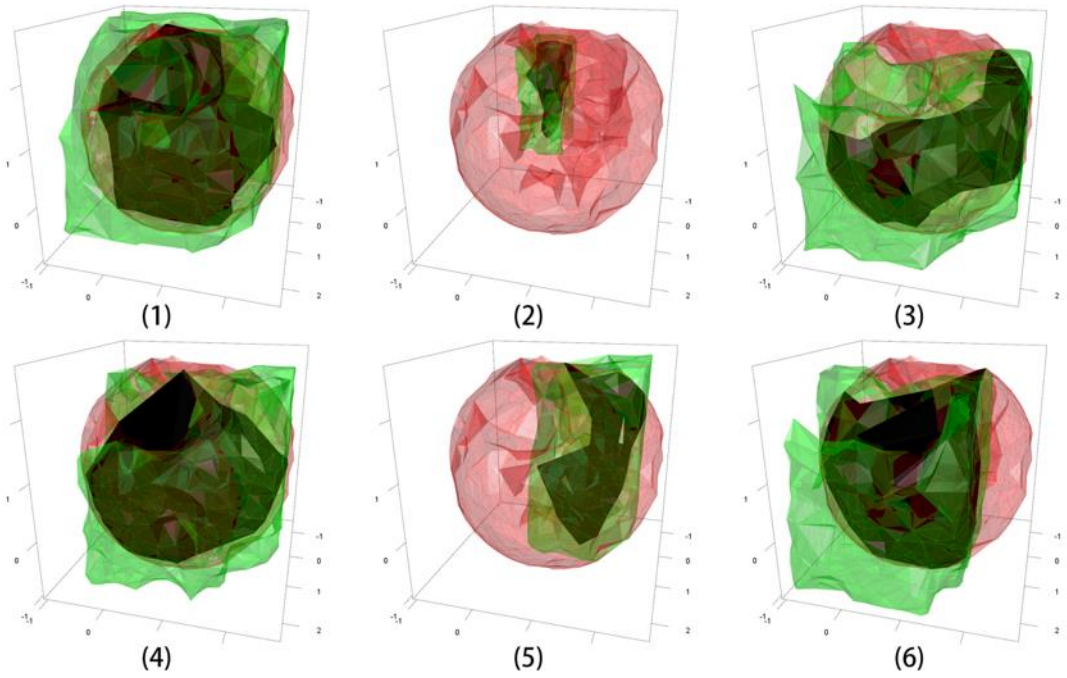
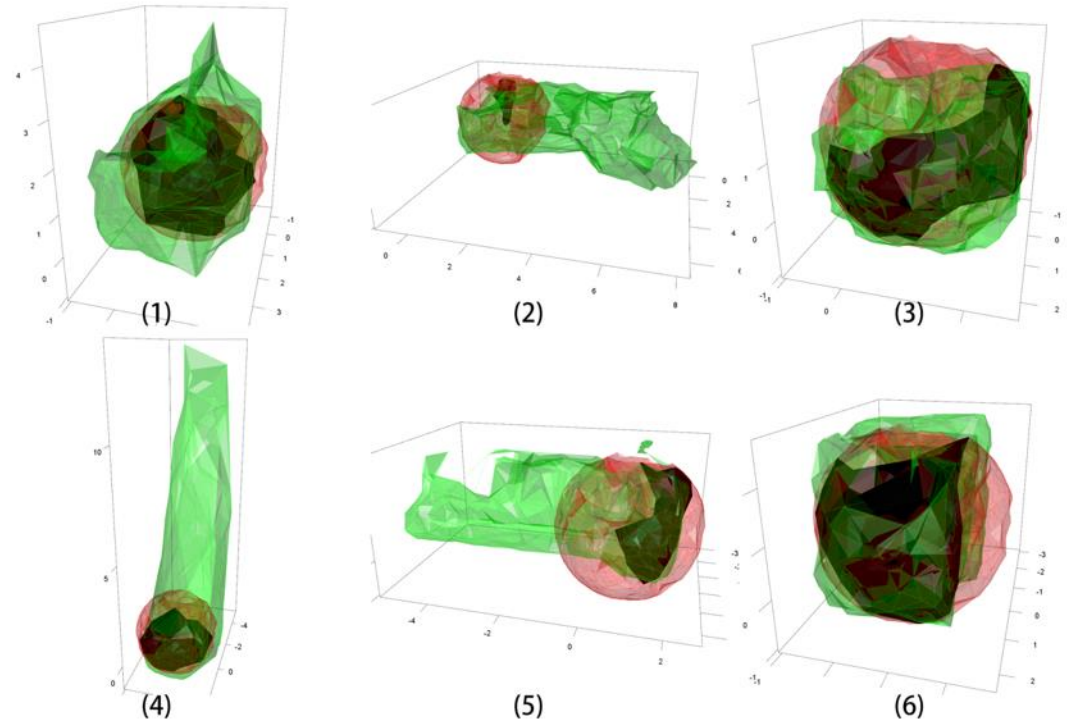


Figure A2. Sixteen virtual species visualized in environmental space. The three axes of the environmental space are PC1, PC2, and PC3. The original defined niches are shown in panel **a**

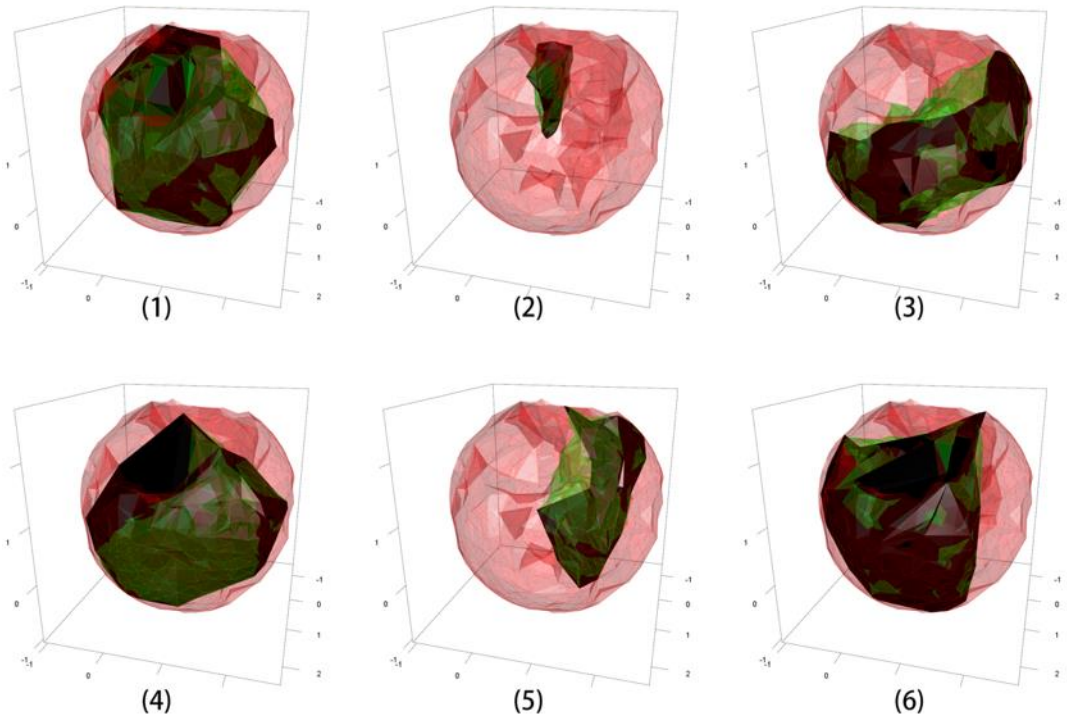
using 3D perspective (a.1), x-y perspective (a.2), x-z perspective (a.3) and y-z perspective (a.4). The niches are different in their centroid (green dot) and radius (different size spheres presented by the white lines with red dots). Because of the unequal transformation between environmental space and geographic space, the defined niches are only partly represented in the corresponding geographic spaces (not shown); see panels **b** and **c**.



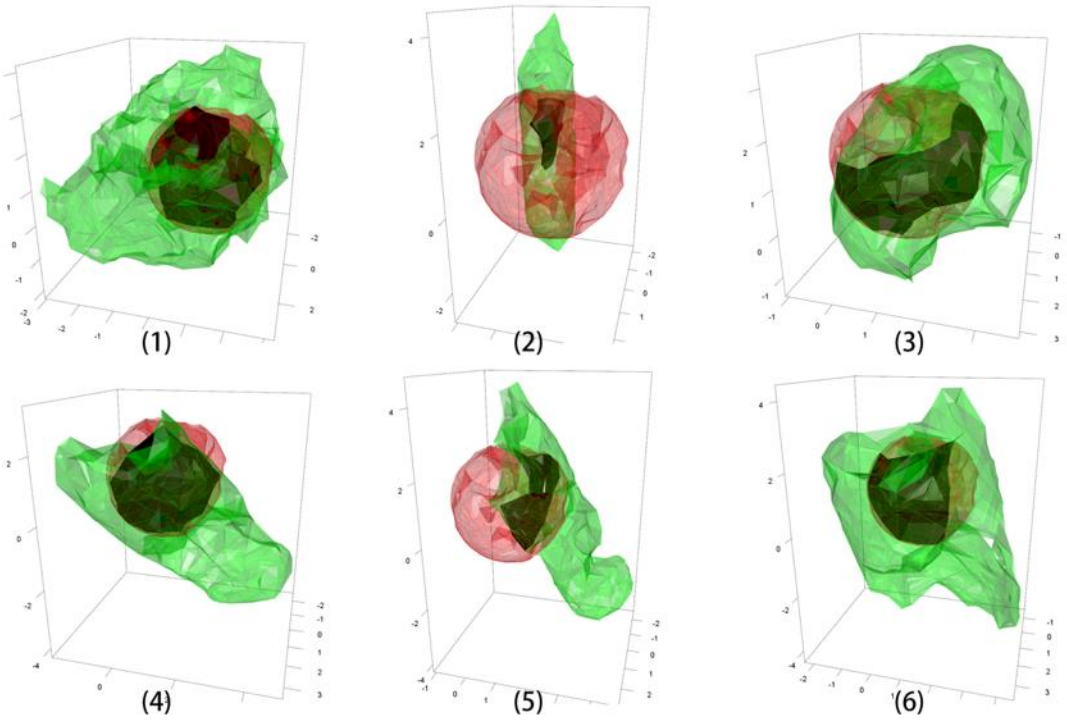
BIOCLIM



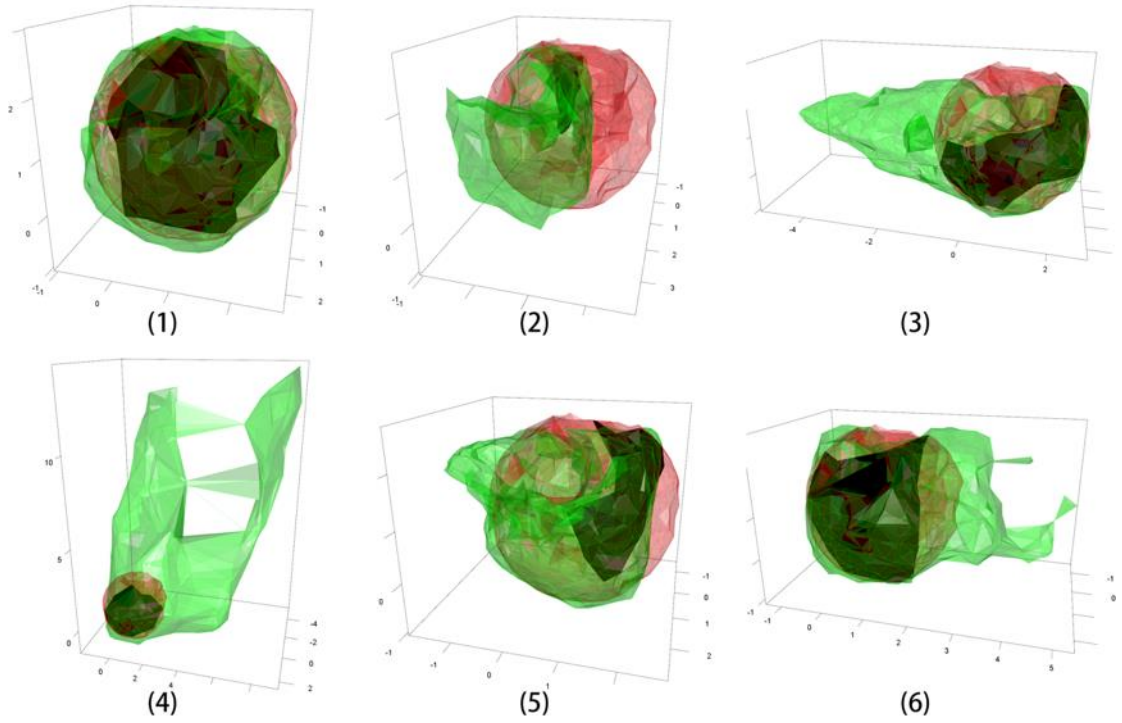
BRT



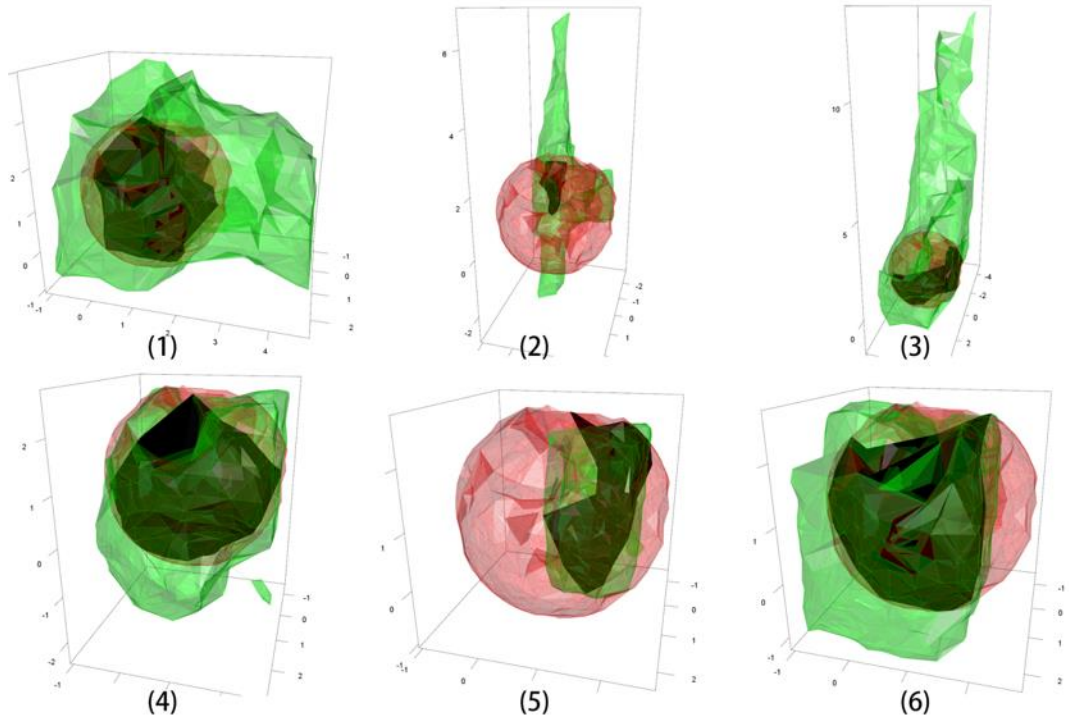
CONVEXHULL



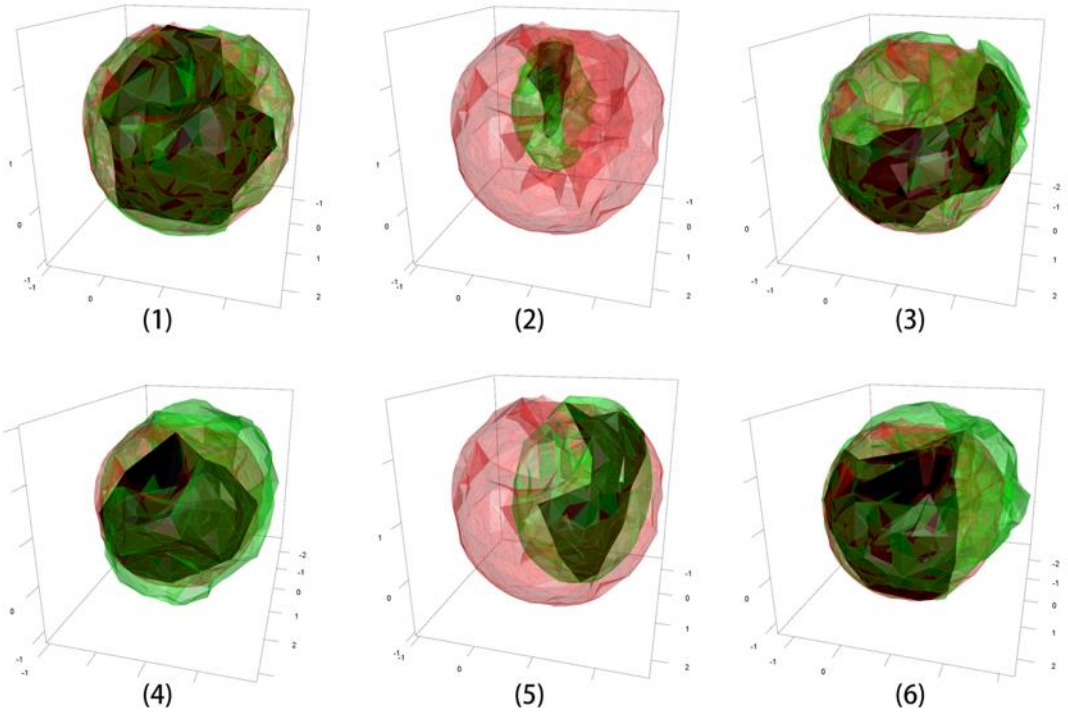
ENFA



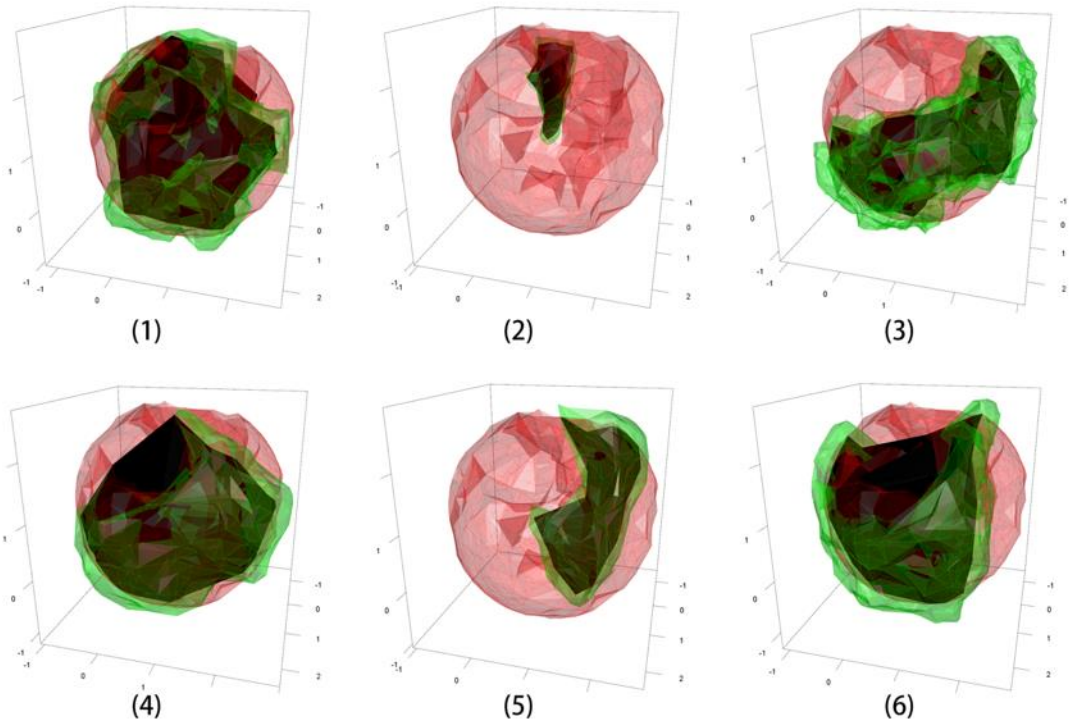
GAM



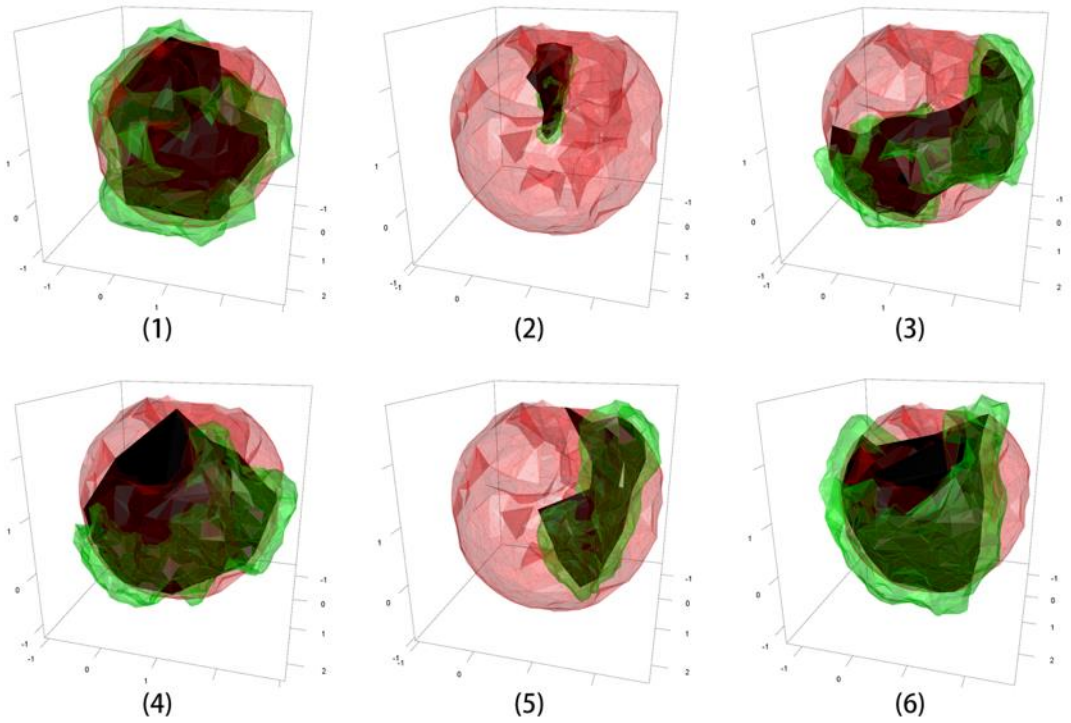
GARP



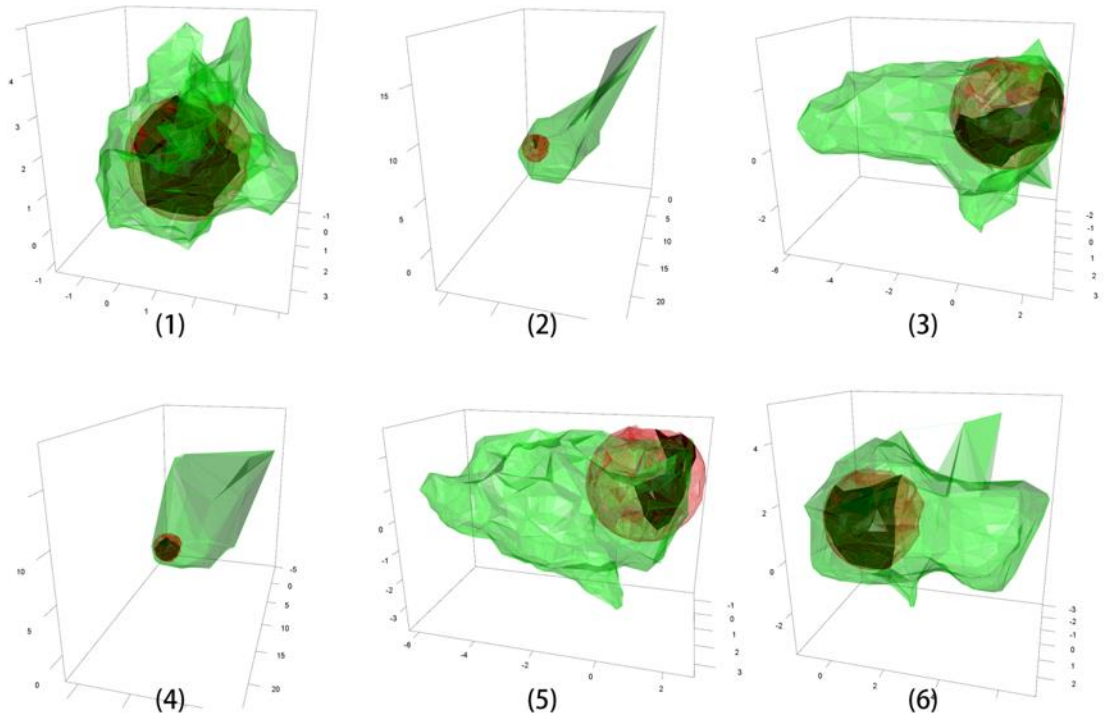
GLM



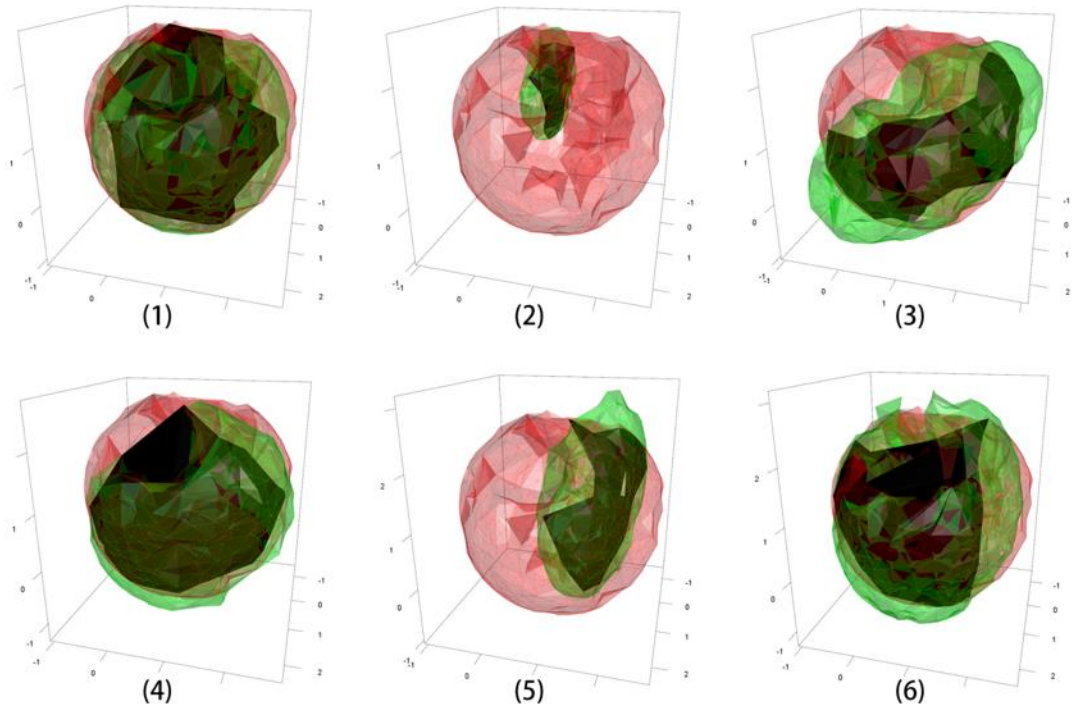
KDE



MA



Maxent



MVE

Figure A3. Representations in PCA space of existing fundamental niche **(N***$_F$; red hull) of a virtual species (the wide niche whose centroid is closest to the origin of the environmental space), the portion of **N***$_F$ used in model calibration (**N***$_{Fx}$; black hull), and the estimated niche ($\hat{\mathbf{N}}$; green hull) by algorithms considered in this study. The study area (Eurasia) is divided into six geographic regions. Each panel (1 to 6) contains different portions of **N***$_F$, represented by different **N***$_{Fx}$ or samples from calibration regions 1 to 6.

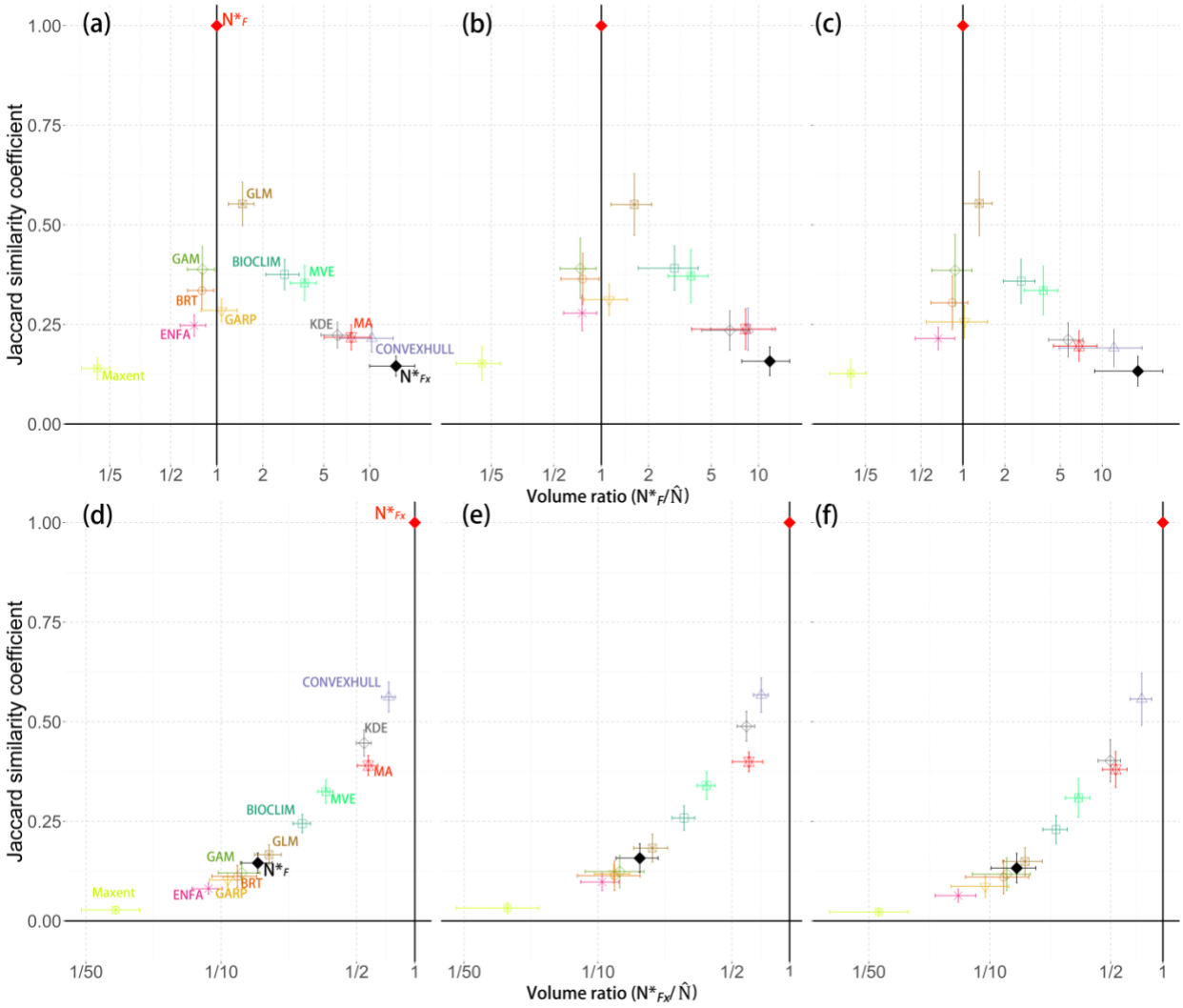


Figure A4. Volume ratio and Jaccard similarity between 3D hulls of the existing fundamental niche (**N***$_F$) a sample of the existing fundamental niche (**N***$_{Fx}$), and the niche estimated from the sample ($\hat{\mathbf{N}}$) in Experiment 1 for different niche configurations. In panels **(a, b and c)**, **N***$_F$ is used as a reference for **N***$_{Fx}$ and $\hat{\mathbf{N}}$ when calculating the volume ratio and Jaccard similarity. In panels **(d, e and f) N***$_{Fx}$ is used as a reference. The panels **a and d** are the results from all the niches, **b and e** are from wide niches and **c and f** are from the narrow niches. Volume ratio represents the geometric volume ratio between the reference 3D hull and target 3D hull, and the Jaccard similarity coefficient measures the geometric similarity between the target 3D hull and reference 3D hull. The error bars represent the 95% confidence intervals.

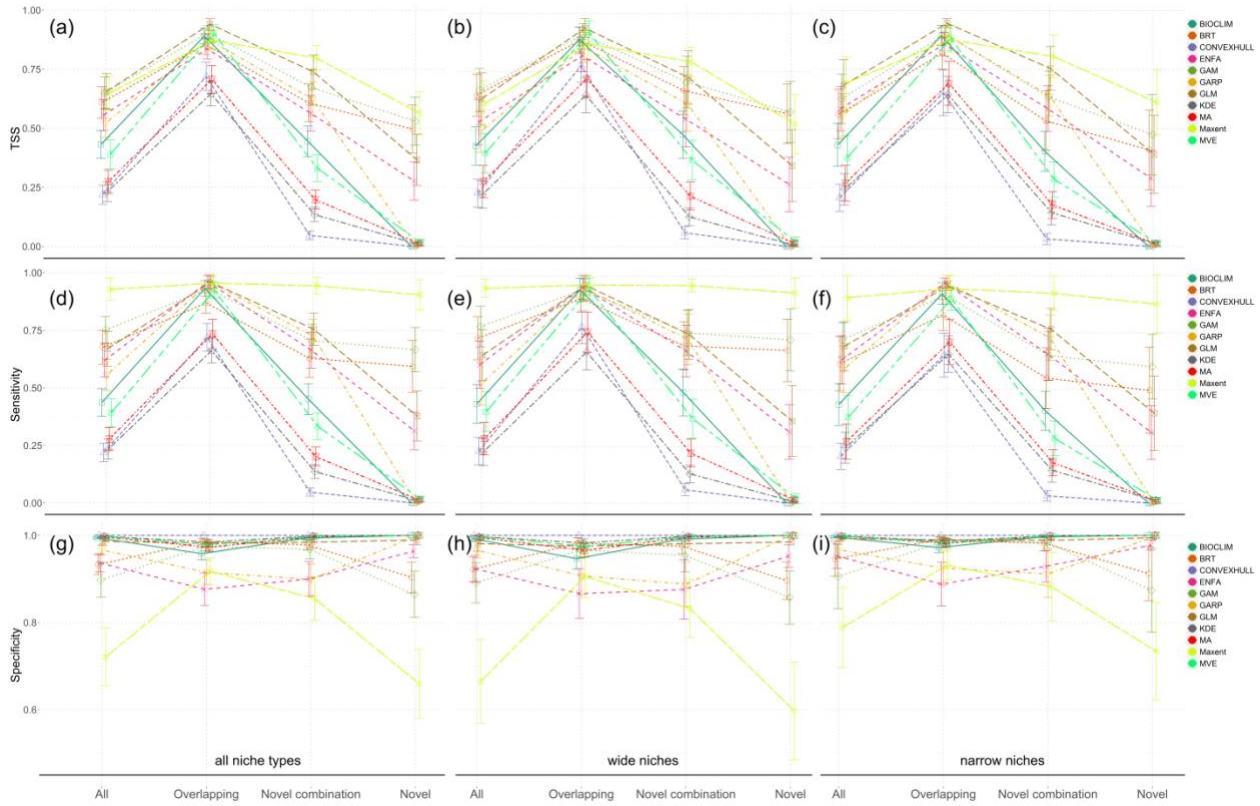


Figure A5. Model evaluation results for Experiment 1. The values of (**a-c**) true skills statistic (TSS), (**d-f**) sensitivity, and (**g-i**) specificity are differentiated by the category of data used for model evaluation. The panels (**a, d, g**) show the results of all the niche configurations, which are the same as Fig.5. The panels (**b, e, h**) show the results of wide niches, and the narrow niches' result are shown in panels (**e, f, i**). "All" represents all evaluation data, "overlapping" represents evaluation data within the 3D hull of the calibration data, "novel combinations" represents evaluation data outside the 3D hull but within the range of the calibration data, and "novel" represents evaluation data outside the range of the calibration data. The error bars represent the 95% confidence intervals.

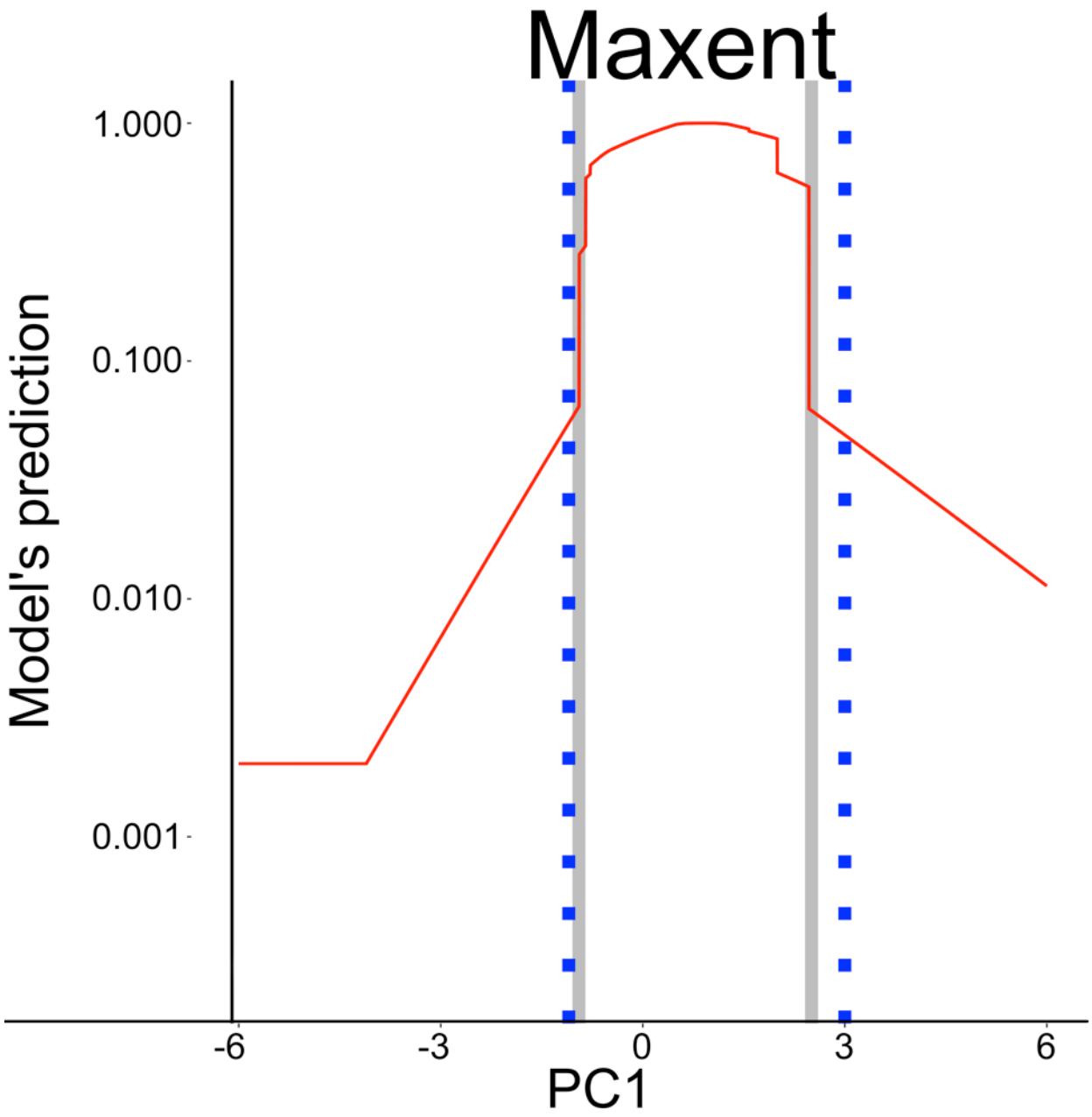


Figure A6. Illustration of extrapolation strategy used by Maxent. The blue dotted line represents the limits of $\mathbf{N}^*_F$ along one environmental variable (PC1) and the gray line represents that of portion of the niche used in model calibration ($\mathbf{N}^*_{Fx}$). The red line represents the prediction by Maxent. The *y* axis in the panel was log transformed to better exemplify the clamping function.

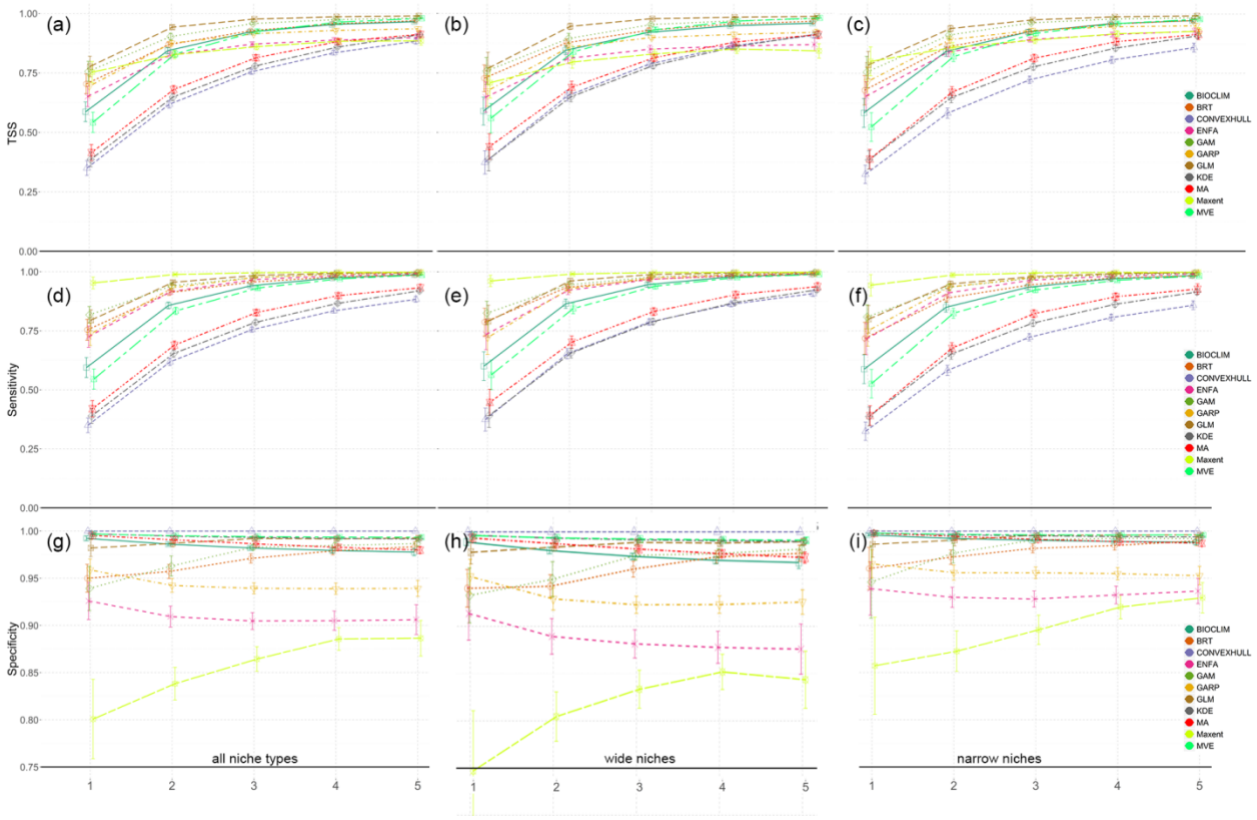


Figure A7. Trends in model evaluation metrics, true skill statistic (TSS), sensitivity, and specificity, aligned with the number of regions used in calibrating models in Experiment 2. The error bars represent the 95% confidence intervals.

65 **Table A1.** Summary of the algorithms used in this study, with relevant parameter settings.
66

Full name	Acronym	Technique	Data input	Implementation	Parameters	Key citation
BIOCLIM	BIOCLIM	Box	Presence-only	*dismo* package (version 1.1-1) in R (version 3.1.2)	N/A	(Busby, 1991)
Ecological Niche Factor Analysis	ENFA	Ellipsoid	Presence-only	openModeller version 1.5.0	10000 background points and 5 retries	(Hirzel *et al.*, 2002)
Convex hull	CONVEXHULL	Convex hull	Presence-only	NicheA (version 3.0)	N/A	(Guisan & Zimmermann, 2000; Qiao *et al.*,
Minimum-volume ellipsoids	MVE	Ellipsoid	Presence-only	NicheA (version 3.0)	N/A	(Van Aelst & Rousseeuw, 2009; Qiao *et al.*,
Generalized linear models	GLM	Logistic regression	Presence and pseudo-absence	base package in R (version 3.1.2)	quadratic function, binomial with logit link, AIC for stepwise selection	(McCullagh & Nelder, 1989)
Generalized additive models	GAM	Logistic regression	Presence and pseudo-absence	*mgcv* package (version 1.8-7) in R (version 3.1.2)	logit link with outer and newton optimizers	(Hastie & Tibshirani, 1990)
Boosted regression trees	BRT	Complex regression	Presence and pseudo-absence	*dismo* package (version 1.1-1) in R (version 3.1.2)	Bernoulli distribution; complexity 5, step size 50, maximum trees size 10,000, and 5 folds cross-validation	(Elith *et al.*, 2008)
Genetic algorithm for rule-set prediction	GARP	Genetic algorithm	Presence and background points	openModeller (version 1.5.0)	10,000 background points, population size 50, and resamples 2500	(Stockwell, 1999)
Maximum entropy modelling	Maxent	Maximum entropy theoty	Presence and background points	*dismo* package (version 1.1-1) in R (version 3.1.2)	default feature and regularization with clamping	(Elith *et al.*, 2006)
Kernel density estimation	KDE	Kernel density	Presence-only	hypervolume package in R	default bandwidth divided by 2	(Blonder *et al.*, 2014; Qiao *et al.*, 2017)

		estimation		(version 3.1.2)		
Marble algorithm	MA	DBScan cluster algorithm	Presence-only	dbscan package (version 1.1-1) in R	epsilon: 5% quantile of all point-pair distances; minimum points: 1% of all calibration points	(Qiao *et al.*, 2015)

67
68

References

Blonder, B., Lamanna, C., Violle, C. and Enquist, B.J. 2014. The n-dimensional hypervolume.– Global Ecol. Biogeogr. 23: 595-609.

Busby, J. 1991. BIOCLIM - a bioclimate analysis and prediction system.– Plant Protection Quarterly 6: 8-9.

Elith, J., Leathwick, J.R. and Hastie, T. 2008. A working guide to boosted regression trees.– J. Anim. Ecol. 77: 802-813.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. and Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data.– Ecography 29: 129-151.

Guisan, A. and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology.– Ecol. Model. 135: 147-186.

Hastie, T. and Tibshirani, R. 1990. Generalized Additive Models.- Chapman and Hall.

Hirzel, A.H., Hausser, J., Chessel, D. and Perrin, N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data?– Ecology 83: 2027-2036.

McCullagh, P. and Nelder, J.A. 1989. Generalized Linear Models.- Chapman and Hall.

Qiao, H., Lin, C., Jiang, Z. and Ji, L. 2015. Marble algorithm: a solution to estimating ecological niches from presence-only records.– Scientific Reports 5: 14232-14232.

Qiao, H., Escobar, L.E., Saupe, E.E., Ji, L. and Soberón, J. 2017. Using the KDE method to model ecological niches: a response to Blonder et al . (2017).– Global Ecology & Biogeography doi: 10.1111/geb.12610.

Qiao, H., Peterson, A.T., Campbell, L.P., Soberón, J., Ji, L. and Escobar, L.E. 2016. NicheA: creating virtual species and ecological niches in multivariate environmental scenarios.– Ecography 39: 805-813.

Stockwell, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction.– International Journal of Geographical Information Science 13: 143-158.

Van Aelst, S. and Rousseeuw, P. 2009. Minimum volume ellipsoid.– Wiley Interdisciplinary Reviews: Computational Statistics 1: 71-82.