Ecography

ECOG-04627
Feng, X., Liang, Y., Gallardo, B. and Papeş, M. 2019. Physiology in ecological niche modeling: using zebra mussel's upper thermal tolerance to refine model predictions through Bayesian analysis. – Ecography doi: 10.1111/ecog.04627

Supplementary material

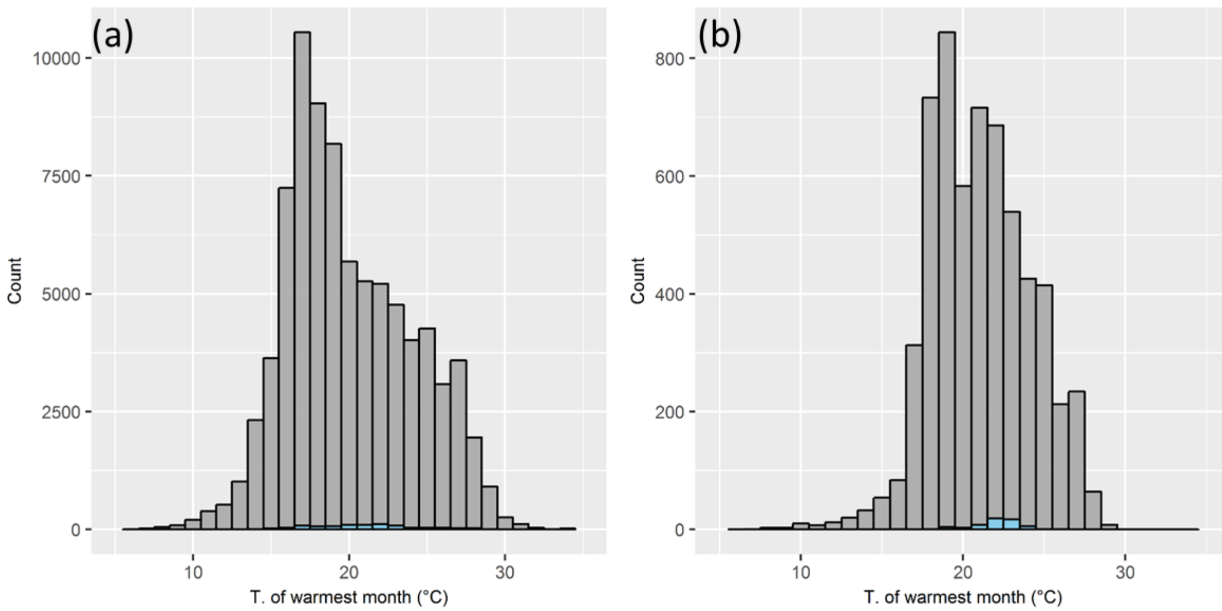

Figure A1. Histograms of environmental conditions of zebra mussel's occurrences (blue) and background points (gray) used in global (panel a) and native (panel b) models.

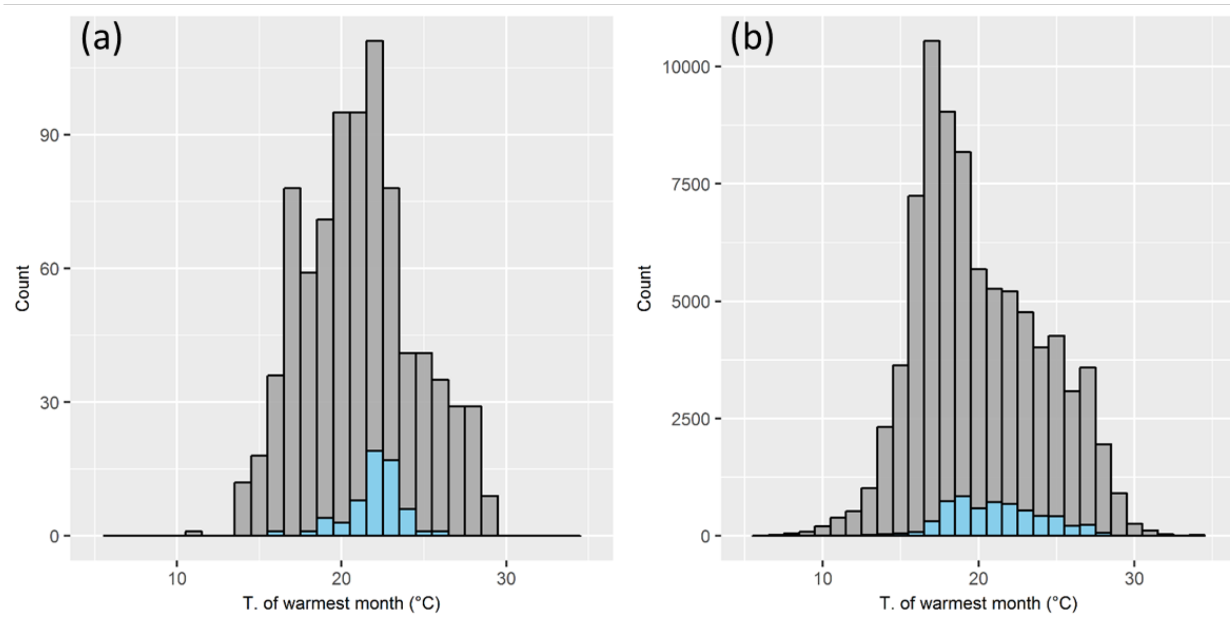

Figure A2. Histograms of environmental conditions of global (gray) and native (blue) occurrences (panel a) and background points (panel b).

Appendix A1. Details of the implementations of classic ecological niche models (ENMs).

To test whether plateau model estimations are similar to established ENM techniques, we included four commonly used ENM algorithms: boosted regression tree (BRT; Friedman et al. 2000, Leathwick et al. 2006), generalized linear model (GLM; McCullagh 1984, Wintle et al. 2005), generalized additive model (GAM; Hastie and Tibshirani 1990, Lehmann et al. 2002), and maximum entropy (Maxent; Phillips et al. 2004, Elith et al. 2006). GLM and GAM are representative of regression-based algorithms and BRT and Maxent are representative of machine learning algorithms.

As comparing performance of algorithms was not the goal of this study, we generally followed the parameter settings from previous studies that have shown reliable performance (Feng and Papeş 2017). Specifically, we generated BRT models in R package dismo (Hijmans et al. 2017) using Bernoulli distribution and learning rate of 0.001, tree complexity of 5, step size of 50, and maximum trees size of 10,000, with five-fold cross-validation (Ridgeway 2006, Elith et al. 2008). We trained GLM in R using binomial with a logit link function and quadratic interaction and adopted the Akaike's information criterion stepwise selection (McCullagh and Nelder 1989). We performed the GAM experiments in R package mgcv, using logit link with outer, newton optimizers and default degree of freedom (-1) that triggers an internal generalized cross-validation to optimize the actual effective degree of freedom (Wood 2011). We assigned equal weights to training presences and absences (total weight of presences equals that of absences) for BRT, GLM, and GAM algorithms (Barbet-Massin et al. 2012). Finally, we ran MaxEnt in R package dismo (Hijmans et al. 2017), with default regularization multiplier and autofeatures option that determines the combination and transformation of variables (i.e., features) based on the number of training presences (Phillips and Dudík 2008). Potentially, MaxEnt parameters could be tuned to find a better fit for a modeling objective [e.g. using ENMeval (Muscarella et al. 2014) or kuenm (Cobos et al. 2019)]. Because our objective is to use MaxEnt as an example of machine learning algorithm to be compared with other regression based models, we here used the default settings to mimic common practices in the applications of MaxEnt studies.

References:

Barbet-Massin, M. et al. 2012. Selecting pseudo-absences for species distribution models: how, where and how many?– Methods Ecol. Evol. 3: 327-338.

Cobos, M.E. et al. 2019. kuenm: an R package for detailed development of ecological niche models using Maxent.– 1-15.

Elith, J. et al. 2008. A working guide to boosted regression trees.– J. Anim. Ecol. 77: 802-813.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data.– Ecography 29: 129-151.

Feng, X. and Papeş, M. 2017. Can incomplete knowledge of species' physiology facilitate ecological niche modelling? A case study with virtual species.– Divers. Distrib. 23: 1157-1168.

Friedman, J. et al. 2000. Additive logistic regression: a statistical view of boosting.– The Annals of Statistics 28: 337-407.

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

Hijmans, R.J. et al. 2017. *dismo*: species distribution modeling –In: R package 1.1-4 <http://CRAN.R-project.org/package=dismo>.

Leathwick, J.R. et al. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees.– Mar. Ecol. Prog. Ser. 321: 267-281.

Lehmann, A. et al. 2002. GRASP: generalized regression analysis and spatial prediction.– Ecol. Model. 157: 189-207.

McCullagh, P. 1984. Generalized linear models.– European Journal of Operational Research 16: 285-292.

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

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.

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. (2004) A maximum entropy approach to species distribution modeling. In: *Proceedings of the Twenty-First International Conference on Machine Learning*, pp. 655–662. ACM press, Banff, Alberta, Canada.

Ridgeway, G. (2006) *Generalized boosted regression models. Documentation on the R Package 'gbm', version 1·5–7*. Retrieved from: http://www.i-pensieri.com/gregr/gbm.shtml (accessed September 2015).

Wintle, B.A. et al. 2005. Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW.– Austral Ecol. 30: 719-738.

Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.– J. Roy. Stat. Soc. Ser. B. (Stat. Method.) 73: 3-36.

Appendix A2. BUGS code for plateau models.

```
# default plateau model without informative priors
model{
	for(i in 1:N) {
		y[i] ~ dbern(p[i])
		y_left[i] <- a1 + b1 * x.clim[i]
		y_right[i] <- a2 + b2 * x.clim[i]
		y_min[i] <- min(y_left[i],y_right[i])
		y_plateau[i] <- min(beta0, y_min[i])
		logit(p[i]) <- y_plateau[i]
	}
	a1 ~ dnorm(0, 0.01)
	b1 ~ dnorm(0, 0.01) I(0,) #dgamma(0.1,0.1)
	a2 ~ dnorm(0, 0.01)
	b2_pov ~ dnorm(0, 0.01) I(0,) #dgamma(0.1,0.1)
	b2 <- -1*b2_pov
	beta0 ~ dnorm(0,0.01)
	}
```

a plateau model using upper thermal limit as informative priors that guide the right segment through a desired point ($Model_{point}$)

```
model{
	for(i in 1:N) {
		y[i] ~ dbern(p[i])
		y_left[i] <- a1 + b1 * x.clim[i]
		y_right[i] <- a2 + b2 * x.clim[i]
		y_min[i] <- min(y_left[i],y_right[i])
		y_plateau[i] <- min(beta0, y_min[i])
		logit(p[i]) <- y_plateau[i]
	}
	a1 ~ dnorm(0, 0.01)
	b1 ~ dnorm(0, 0.01) I(0,)
	a2_mean <- -7.217852  -b2*30
	a2_prec <- 1/0.007217852
	a2 ~ dnorm(a2_mean, a2_prec)
	b2_mean <- abs(0)
	b2_prec <- 1/100
	b2_pov ~ dnorm(b2_mean ,b2_prec  )   I(0,)
	b2 <- -1*b2_pov
	beta0 ~ dnorm(0,0.01)
	}
```

a plateau model using survival data as informative priors that guide the slope of the right segment ($Model_{survival}$)

```
model{

        for(i in 1:N) {

                y[i] ~ dbern(p[i])

                y_left[i] <- a1 + b1 * x.clim[i]

                y_right[i] <- a2 + b2 * x.clim[i]

                y_min[i] <- min(y_left[i],y_right[i])

                y_plateau[i] <- min(beta0, y_min[i])

                logit(p[i]) <- y_plateau[i]

        }

        a1 ~ dnorm(0, 0.01)

        b1 ~ dnorm(0, 0.01) I(0,)

        a2_mean <- 0

        a2_prec <- 1/100

        a2 ~ dnorm(a2_mean, a2_prec)

        b2_mean <- abs(-1.073427)

        b2_prec <- 1/0.001073427

        b2_pov ~ dnorm(b2_mean ,b2_prec  )   I(0,)

        b2 <- -1*b2_pov

        beta0 ~ dnorm(0,0.01)

        }
```

a plateau model using upper thermal limit and survival data as informative priors that guide the right segment through a desired point with a particular slope ($Model_{point\text{-}survival}$)

```
model{

	for(i in 1:N) {

		y[i] ~ dbern(p[i])

		y_left[i] <- a1 + b1 * x.clim[i]

		y_right[i] <- a2 + b2 * x.clim[i]

		y_min[i] <- min(y_left[i],y_right[i])

		y_plateau[i] <- min(beta0, y_min[i])

		logit(p[i]) <- y_plateau[i]

	}

	a1 ~ dnorm(0, 0.01)

	b1 ~ dnorm(0, 0.01) I(0,)

	a2_mean <- -7.217852  -b2*30

	a2_prec <- 1/0.007217852

	a2 ~ dnorm(a2_mean, a2_prec)

	b2_mean <- abs(-1.073427)

	b2_prec <- 1/0.001073427

	b2_pov ~ dnorm(b2_mean ,b2_prec  )   I(0,)

	b2 <- -1*b2_pov

	beta0 ~ dnorm(0,0.01)

	}
```