Article S3. Parameter optimization and analyses for MaxEnt modelling

MaxEnt parameter optimization

Recent studies have called for careful parameter optimization to balance accuracy and complexity in SDM models (Morales et al. 2017). Our goal was to fully optimize our MaxEnt modelling process and the following steps were designed for each modelling scenario.

We focused on two aspects of MaxEnt parameterization (Phillips & Dudík 2008): 1) Feature classes including linear (L), quadratic (Q), product (P), hinge (H) and threshold (T), where combinations of FCs determine how organism-to-environment responses are quantified and more FCs increase model complexity (Elith et al. 2011). 2) Regularization multiplier (RM) for global model complexity tuning, with higher RM values favoring less-complex models. For these two parameters, We generated a pool of 80 parameter combinations from 8 FC sets (L, H, LQ, LQP, LQH, LQPT, LQPH, LQPHT) and 10 RMs (from 1.0 to 5.5 in 0.5 increments).

We then used a stepwise backward selection framework to remove variables that contributed less than a given percentage to the model. Higher cut-offs are expected to retain fewer variables and produce simpler models, thus further reducing model complexity. For the first round of modelling, all 40 variables were used and those contributing less than the cut-off were eliminated. The remaining variables were pooled for the next round of modelling. This process was repeated until all variables contributed above the given cut-off (final variables). We tested three cut-off values 1%, 3%, and 5% and presented results from 3% models in the main text.

Finally, for each cut-off, a cross-validation model (10-fold) for each parameter combination was generated using the final variables. We ranked the 80 validated models from each cut-off by the Akaike Information Criterion corrected for small sample size (AICc. Hurvich & Tsai 1989) and the top model ($\Delta_{AICc} = 0$) was considered as the optimized model. This process produced one optimized model for each cut-off.

Following summarizes our effort on MaxEnt model optimization. Raising the cut-off reduced the model complexity by including fewer variables, with the effect more pronounced in the smaller Rideau region (Table 1). Varying the cut-off changed little for model FC and RM, although for the Rideau region models with the 1% cut-off, three FCs (LQP) were included for the combined dataset, compared to two FCs (LQ) for the remaining models (Table 1). While higher cut-offs lead to lower training and testing AUC values, the overall AUC values were over 0.90 for all models implying satisfactory model performance (Swets 1988). The optimized models had less overfitting with slightly lower training and testing AUC values compared to models generated using default MaxEnt settings (i.e., LQPTH1.0). We decided that models from 3% cut-off balanced model complexity and prediction accuracy[1]. Thus, our optimized models included 1) *Linear* and *Quadratic* features for models in Rideau region with RM = 1.0; and 2) *Linear*, *Quadratic*, *Product*, and *Threshold* features for models in southern Ontario region RM = 2.5. For results from models using 1% and 5% cut-offs see Article S4.

Overall, our optimization protocol reduced model overfitting by including fewer FCs, increasing RM for overfitting, and removing non-significant environmental

[1] However, there is no consensus on what cut-off value is best, although 1% is common. In our case, higher cut-off values lowered the model complexity by reducing the number of variables, rather than altering its FCs or RM. We recommend exploring the effects of different criteria during MaxEnt optimization when using stepwise selection.

variables while maintaining high prediction accuracy. A less complex and more generalizable model not only means greater transferability to areas where the focal species is not yet known, or to past or future climate scenarios (Wenger & Olden 2012), but is also more biologically and ecologically interpretable.

Comparing MaxEnt predictions using KL Divergence

MaxEnt raw output, relative occurrence rate (ROR) is essentially a probability distribution across the modelling space. For example, in our Rideau modelling region, there are 44,639 grid cells (1 km^2) and MaxEnt predicts probability of presence for each cell and values in all cells add up to 1. A widely used metric to measure differences between probability distributions is Kullback-Leibler divergence (KL Divergence, D_{KL}) introduced by Kullback & Leibler (1951). KL divergence asymmetrically measures the distance from one discrete probability distribution *P* to another *Q* (*P* and *Q* are defined on the same probability space, χ):

$$D_{KL}(P \parallel Q) = \sum_{x \in \chi} P(x) \log \left(\frac{P(x)}{Q(x)}\right)$$

Similarly, distance from *Q* to *P* is given by:

$$D_{KL}(Q \parallel P) = \sum_{x \in \chi} Q(x) \log \left(\frac{Q(x)}{P(x)}\right)$$

The mean of these two distances, also known as symmetrised KL divergence, was used in this study. Since KL divergence is a relative measurement of entropy, it does not quantify as a statistical metric. Our workaround is described as followed. In this study, the final MaxEnt models were validated using k-fold (k = 10) cross-validation, where 10 replicate models were built each with 10% randomly occurrences omitted from the original

dataset. The final MaxEnt model is essential the average of these 10 replicate models. Let *A* and *B* denote the raw ROR output of two MaxEnt models, each with 10 replicate models ($A_{1\text{-}10}$ and $B_{1\text{-}10}$). The average distance among the replicate models ***within*** each model can then be calculated as the average pair-wise KL divergence among the replicate models:

$$Mean\ D_{KL}\ within(A) = \frac{1}{90} \times \sum_{i,j \in [1,10]}^{i \neq j} D_{KL}(A_i \parallel A_j)$$

$$Mean\ D_{KL}\ within(B) = \frac{1}{90} \times \sum_{i,j \in [1,10]}^{i \neq j} D_{KL}(B_i \parallel B_j)$$

Similarly, the average distance among the replicate models ***between*** each model can be calculated as:

$$Mean\ D_{KL}\ betweem(AB) = \frac{1}{200} \times \sum_{i,j \in [1,10]} D_{KL}(A_i \parallel B_j)$$

These means can then be statistically compared, where we used Kruskal Wallis test for this study. If the two distribution A and B are of little divergence from each other, one can expect both Mean D_{KL} within A and B to be similar to the Mean D_{KL} between A and B. Conversely, if both Mean DKL within A and B are smaller than Mean DKL between A and B, we considered the two probability distributions A and B are significantly departing from each other. In both cases, we should not observe the Mean D_{KL} within A and B to be significantly different from each other.

Collinearity of environmental variables

Spatial autocorrelation among environmental variables used for niche modelling is inevitable. Whether one should address this spatial collinearity before or after MaxEnt

modelling is debated (Merow et al. 2013). We did not seek to reduce collinearity *a priori* for two reasons: 1) There is no consensus on the need for this, as MaxEnt was developed to accommodate autocorrelated variables (Phillips et al. 2006). For example, Feng et al. (2019) showed that model performance was not affected by the collinearity (although transferability to other regions could be compromised); 2) The available collinearity reduction methods have not been adequately tested for MaxEnt. For example, PCA prior to modelling effectively eliminates collinearity at the cost of rendering variables uninterpretable, as the abstracted multivariate dimensions have no real-world meaning.

Instead, we performed PCA post hoc for variables from the optimized MaxEnt models to interpret the relation between musk turtle distribution and environmental gradients. The retained variables fell into three categories: 1) Thermal conditions (Annual mean temperature, Mean temperature warmest quarter); 2) Aquatic environmental characteristics (Waterbody area proportions, Total shoreline); and 3) Elevation and precipitation (Mean elevation, Precipitation warmest quarter). Mean elevation and Precipitation warmest quarter are grouped according to the PCA results although the correlation between them was relatively low (r_s = 0.41).

References

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, and Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17:43-57. 10.1111/j.1472-4642.2010.00725.x

Feng X, Park DS, Liang Y, Pandey R, and Papes M. 2019. Collinearity in ecological niche modeling: Confusions and challenges. *Ecology and Evolution* 9:10365-10376. 10.1002/ece3.5555

Hurvich CM, and Tsai C-L. 1989. Regression and time series model selection in small samples. *Biometrika* 76:297-307. 10.1093/biomet/76.2.297

Kullback S, and Leibler RA. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22:79-86. 10.1214/aoms/1177729694

Merow C, Smith MJ, and Silander Jr JA. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36:1058-1069. 10.1111/j.1600-0587.2013.07872.x

Morales NS, Fernández IC, and Baca-González V. 2017. MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. *PeerJ* 5:e3093. 10.7717/peerj.3093

Phillips SJ, Anderson RP, and Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259. 10.1016/j.ecolmodel.2005.03.026

Phillips SJ, and Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161-175. 10.1111/j.0906-7590.2008.5203.x

Swets J. 1988. Measuring the accuracy of diagnostic systems. *Science* 240:1285-1293. 10.1126/science.3287615

Wenger SJ, and Olden JD. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* 3:260-267. 10.1111/j.2041-210X.2011.00170.x