Predictor complexity and feature selection affect Maxent model transferability: Evidence from global freshwater invasive species

Bi Wei Low1,2 | Yiwen Zeng1 | Heok Hui Tan2 | Darren C.J. Yeo1,2

Abstract

Aim: Ecological niche models (ENMs) are widely used to address urgent real-world problems such as climate change effects or invasive species; however, the generality of models when projected through space and/or time, that is transferability, remains a key challenge. Here, we explored the effects of complex predictors and feature selection on ENM transferability in a widely employed algorithm, Maxent, using five globally invasive freshwater species as case studies.

Location: Global.

Methods: We modelled the global distributions of five notorious freshwater invasive species (African sharptooth catfish *Clarias gariepinus*, Mozambique tilapia *Oreochromis mossambicus*, American bullfrog *Lithobates catesbeianus*, red swamp crayfish *Procambarus clarkii*, and Australian redclaw crayfish *Cherax quadricarinatus*), using three predictor datasets of varying complexities derived from two commonly used climatic data sources (WorldClim and IPCC) and three methods of model tuning that differentially incorporated feature selection. Spatially explicit transferability assessments were then conducted using a suite of evaluation metrics previously used to quantify Maxent model performance.

Results: We show that in the absence of detailed biological knowledge of focal species, simpler predictor datasets produce models that are more accurate than those calibrated using comprehensive "bioclimatic" datasets. Additionally, we find that tuning models for both optimal regularization parameters as well as feature-class combinations led to the greatest increases in transferability and geographic niche conservatism. Results indicate a tenuous link between model transferability and Akaike's information criterion corrected for small sample sizes (AICc), suggesting that the indiscriminate use of AICc as an estimate of model parsimony may lead to erratic model performance.

Main conclusions: Our findings demonstrate that methodological considerations can drastically affect the reliability of spatial and possibly temporal projections, which
has severe implications when ENMs are used to infer species’ niches, and quantify ecological or evolutionary change across impacted landscapes.

**KEYWORDS**

aquatic invasions, bioclimatic variables, ecological niche model, feature selection, Maxent, model tuning, niche conservatism, predictor selection, species distribution model, transferability

1 | INTRODUCTION

Ecological niche models (ENMs; also known as environmental niche models or species distribution models) are tools that use species occurrence records in conjunction with environmental data to infer species’ niches and predict the relative suitability of habitats across geographic landscapes (Elith et al., 2011; Radosavljevic & Anderson, 2014; Warren & Seifert, 2011). They have been widely used to answer increasingly complex questions in various fields such as ecology and biogeography (e.g., Glor & Warren, 2010), epidemiology (e.g., Cardoso-Leite et al., 2014), conservation biology (e.g., Warren et al., 2014) and invasion biology (e.g., Rodda et al., 2011). Of the numerous ENM techniques available, the most extensively used algorithm is Maxent (Phillips et al., 2006), a presence-only method based on the principle of maximum entropy. Its popularity has been attributed to its high predictive accuracy compared to other modelling methods (Elith et al., 2006), robustness to small sample sizes (Pearson et al., 2007), wide range of customizable settings, and ease of use (Elith et al., 2006; Merow et al., 2013). Despite its flexibility in model construction, and perhaps in part owing to its ease of use, the algorithm is often not applied optimally by users relying on default settings without consideration for data treatment and model parameterization—despite such practices being deemed inappropriate (Merow et al., 2013).

A key challenge in ecological niche modelling is one of transferability, that is the generality of models when transferred and projected into broad unsampled environments across time and space (Peterson et al., 2007). Transferability typically involves two components—interpolation (model predictions within the environmental constraints encountered during model calibration) and extrapolation (predictions outside of environmental ranges encountered during model calibration) (Elith & Leathwick, 2009). These have far-reaching implications for many conservation applications such as forecasting biological invasions (e.g., Rodda et al., 2011; Rödder et al., 2009), the identification of priority conservation areas for rare or threatened species (e.g., Breiner et al., 2015), quantification of climate change risks (e.g., Warren et al., 2014), and elucidating ecological processes driving species diversification (e.g., Acevedo et al., 2014). Consequently, uncertain model transferability can hinder efforts to address urgent real-world problems or lead to misleading conclusions (Merow et al., 2013). This issue arises because presence-only or presence-background ENMs (such as Maxent) typically quantify realized species–environment niche relationships using species occurrence data that may be in part influenced by dispersal limitations and biotic interactions, whereas accurate model projections into unsampled regions require reasonable estimations of fundamental-niche relationships based on realized niche information (Soberón & Peterson, 2005; Peterson et al., 2011). Projections are therefore only robust if modelled species–environment relationships are aligned with biological expectations, and are simple enough to allow for accurate extrapolations (Rodda et al., 2011).

In recent years, a growing number of studies have investigated approaches to improve ENM transferability. These include the selection of appropriate modelling algorithms (e.g., Breiner et al., 2015; Qiao et al., 2019), methods to reduce geographic sampling bias in occurrence datasets (Fourcade et al., 2014), delimitation of appropriate background extents for model calibration (Barve et al., 2011), tuning of parameters for optimal model complexity (Moreno-Amat et al., 2015; Radosavljevic & Anderson, 2014; Warren & Seifert, 2011), and comparisons of different model evaluation techniques (Muscarella et al., 2014; Rödder & Engler, 2011; Warren & Seifert, 2011).

One aspect that has received less attention, however, is the selection of environmental predictors, despite predictor selection being fundamental to any modelling effort (Araújo & Guisan, 2006). A common practice in many ENM studies is to use “standard” sets of environmental predictors (Rödder et al., 2009), such as bioclimatic variables secondarily derived from annual or monthly means in temperature or precipitation to reflect biologically relevant energy and water balances (Xu & Hutchinson, 2013), although this may lead to overfitted models if not tuned for appropriate levels of model complexity (Moreno-Amat et al., 2015; Warren & Seifert, 2011; Warren et al., 2014). In contrast, using a subset of predictors based on the known physiology or ecology of species can produce ENMs that project more accurately into novel environments (Rödder et al., 2009; Zeng et al., 2016; Petitpierre, Broennimann, Kueffer, Daehler, & Guisan, 2017), although such biological information is often unavailable. Alternatively, heuristic approaches in predictor selection and reduction (e.g., stepwise removal of variables using variable contribution scores; Zeng et al., 2016) may also improve model transferability, though the efficacy of such procedures remains highly dependent on the initial set of starting variables.

It has been suggested that the indiscriminate use of complex composite bioclimatic variables (secondarily calculated from annual or monthly means in temperature or rainfall) without a priori ecological hypotheses may explain poor predictivity in certain cases where models are transferred across geographic regions (Peterson & Nakazawa, 2008), although this has sometimes also been attributed
to shifts in species’ niches (Broennimann et al., 2007). Composite variables are widely utilized in many ENM studies; in particular, they comprise the majority (13 out of 19 variables) of the WorldClim bioclimatic dataset (Hijmans et al., 2005). Nonetheless, their overall impact on model transferability is not well understood.

Maxent can construct simple to highly complex, nonlinear species–environment relationships using various mathematical transformations of variables (termed features). Six feature classes are available—linear, quadratic, product, hinge, threshold and categorical. To reduce overfitting, Maxent uses a regularization procedure (L1-regularization) to balance model fit with complexity, by penalizing models based on the magnitude of their coefficients (Phillips et al., 2006). By default, the program uses all feature classes in model fitting (contingent on a sufficient number of species occurrences) and a relatively low regularization multiplier of 1, although this can potentially lead to models that are overparameterized (Elith et al., 2011; Merow et al., 2013). Many recent studies (e.g., Moreno-Amat et al., 2015; Radosavljevic & Anderson, 2014; Warren & Seifert, 2011) have demonstrated the importance of species-specific tuning of regularization parameters for optimal model complexity, though few have examined the effects of fitted feature forms on model transferability. While the use of appropriate regularization parameters can reduce the need for feature selection, it may be advisable to constrain the feature classes used in model fitting to produce models that are more interpretable and transferable (Merow et al., 2013).

In recent years, the Akaike's information criterion correction for small sample sizes (AICc) has been increasingly used as an evaluation tool to tune Maxent parameters. Models tuned using AICc have been shown to generally be more parsimonious and less overfitted, while better estimating habitat suitability and relative importance of predictor variables (Warren & Seifert, 2011). While AICc may serve as a useful alternative to traditional cross-validation model evaluation, it is calculated on unpartitioned species occurrence datasets and therefore does not truly infer model transferability (Warren & Seifert, 2011; Warren et al., 2014). There has been evidence suggesting that AICc-tuned models may occasionally exhibit poorer performance in some cross-validation transferability assessments (e.g., slightly lower discriminatory ability on withheld testing data; Muscarella et al., 2014), although this may be confounded by geographic sampling bias present within training datasets (Galante et al., 2018). Additionally, a recent study utilizing simulated species (Velasco & González-Salazar, 2019) showed weak correspondence between AICc and geographical predictive accuracy. Nonetheless, the overall trend between model parsimony and transferability has yet to be clarified using empirical species datasets.

Here, we modelled the global distributions of five notorious freshwater invasive species (African sharptooth catfish *Clarias gariepinus*, Mozambique tilapia *Oreochromis mossambicus*, American bullfrog *Lithobates catesbeianus*, red swamp crayfish *Procambarus clarkii*, and Australian redclaw crayfish *Cherax quadricarinatus*), using three predictor datasets of varying complexities derived from two commonly used climatic data sources (WorldClim and IPCC) and three methods of model tuning that differentially incorporated feature selection. We then conducted spatially explicit transferability validations using a suite of evaluation metrics previously used to quantify Maxent model performance (Hirzel et al., 2006; Warren et al., 2010; Peterson et al., 2011; Warren & Seifert, 2011; Radosavljevic & Anderson, 2014). Our five focal species were chosen as they are (a) invaders with long histories of introductions and establishments (Ficetola et al., 2007; Lodge et al., 2012; Lowe et al., 2000; Palaoro et al., 2013; Weyl et al., 2016), and thus more likely to be at equilibrium with the environment in non-native regions, and (b) possibly already occupying a large proportion of their fundamental niches across native and invasive distributions (Araújo & Pearson, 2005; Václavík & Meentemeyer, 2009), thereby making them prime candidates for transferability assessments.

Our main objective is to understand how methodological considerations can impact model transferability, and our results serve to guide future modelling efforts and improve ENM specification techniques. To this end, we specifically addressed the following questions: (a) Does the inclusion of complex bioclimatic predictors influence Maxent model transferability? (b) Does incorporating feature selection in model tuning produce Maxent models that are more transferable? And (c) Is model parsimony (estimated using AICc) a good indicator of model transferability?

## 2 | METHODS

### 2.1 | Species occurrences and accessible areas

We collated occurrence data for the five focal species from published literature (e.g., Both et al., 2011; Russell et al., 2012; Snovsky & Gall, 2011; Teugels, 1986; Torres & Álvarez, 2012; Zeng et al., 2016; Zeng & Yeo, 2018), online databases (e.g., Global Biodiversity Information Facility (GBIF) (http://www.gbif.org), Global Invasive Species Database (GISD) (http://www.iucnisd.org/gisd)), and natural history museum collections (e.g., Smithsonian National Museum of Natural History, Lee Kong Chian Natural History Museum; Appendix S1). These comprised both native occurrences as well as records of non-native establishments. A species is considered established at a locality if there is evidence of a self-sustaining breeding population, or if the reported locality is situated within a region where the species is already known to be well-established. In total, we collated 545 occurrence records (412 native and 133 established) for *C. gariepinus*, 2,158 records (1,211 native and 947 established) for *O. mossambicus*, 9,639 records (5,316 native and 4,323 established) for *L. catesbeianus*, 1,240 records (457 native and 783 established) for *P. clarkii*, and 159 records (98 native and 61 established) for *C. quadricarinatus*.

Datasets of species occurrences often exhibit strong geographic bias, whereby certain areas (e.g., near busy roads or towns) tend to be more heavily sampled than others. This bias can lead to the pseudo-replication of points clustered around accessible areas and the over-representation of certain environmental conditions.
during model fitting, therefore negatively impacting ENM transferability (Boria et al., 2014; Kramer-Schadt et al., 2013). To reduce the effect of sampling bias, spatial filtering (Boria et al., 2014; Fourcade et al., 2014) was carried out on the occurrences datasets. While Maxent, by default, removes duplicate occurrences within each predictor raster grid cell, occurrences were further thinned using the thin function in the R package “spThin” (Aiello-Lammens et al., 2015), which utilizes a randomization algorithm to return the highest number of occurrences for a specified thinning distance. We chose a thinning distance of 50 km, as visual examinations of model outputs during preliminary runs showed that predictions were generally overfitted on training datasets at lower thinning distances (e.g., 20 km), whereas higher thinning distances (e.g., 100 km) often produced underfitted models lacking in spatial resolution. Following the spatial-filtering procedure, our final datasets comprised 346 occurrence records (245 native and 101 established) for C. gariepinus, 419 records (134 native and 285 established) for O. mossambicus, 808 records (505 native and 303 established) for L. catesbeianus, 354 records (115 native and 239 established) for P. clarkii, and 87 records (54 native and 33 established) for C. quadricarinatus.

Maxent fits models by contrasting predictor variables at known occurrence points against random “background” localities in the study region (Phillips & Dudík, 2008). Conceptually, the background should comprise areas accessible to a species over relevant time periods (Barve et al., 2011), though delimiting its extent can be challenging for invasive species outside their native distributions as population boundaries might still be expanding (Elith et al., 2010). Native-range background extents were defined using a conventional minimum convex polygon around native occurrences with an additional 4° buffer (Rodda et al., 2011). On the other hand, the areas accessible to non-native populations were defined using distance buffers estimating maximum dispersal distance around known invasive occurrences (Barve et al., 2011; Figure S2.1). Maximum dispersal distances were calculated as the average natural and/or human-mediated dispersal rates for a species (Appendix S3 for details) multiplied by the number of years since first introduction/record at a locality. Native- and invasive-range background extents were concatenated and 10,000 background points randomly drawn for model calibration.

2.2 Environmental datasets

ENMs built for terrestrial organisms often utilize climatic data to directly infer species–environment relationships, as many key physiological processes (e.g., photosynthesis, evapotranspiration) are directly affected by air temperature and rainfall (Austin, 2002). The use of such variables for modelling freshwater species may, however, be counterintuitive, given that their distributions are typically limited by instream conditions (e.g., hydrology, water temperature). While high-resolution instream data are currently unavailable on a global scale, there has been evidence (e.g., McGarvey et al., 2018) demonstrating that climatic variables, in conjunction with physical habitat data, can serve as reasonable surrogates for instream variables, given that these distal variables are often causally linked to proximal drivers of freshwater species distributions (e.g., air temperature directly affecting water temperature; altitude and slope directly influencing channel morphology) (Zeng et al., 2016).

To this end, global climatic layers were obtained from the WorldClim bioclimatic dataset version 1.4 (Hijmans et al., 2005) (http://www.worldclim.org/bioclim) and the Intergovernmental Panel on Climate Change (IPCC) CRU high-resolution climate dataset version 2.1 (Mitchell & Jones, 2005) (http://www.ipcc-data.org). The WorldClim dataset was downloaded at a spatial resolution of 0.0416° (~5 km at the equator), whereas IPCC layers were downloaded at a native resolution of 0.5° (~55 km at the equator) and were artificially resampled to 0.0416° resolution to match the WorldClim layers.

Three climatic predictor datasets were prepared for comparison:

1. BC1: a dataset that comprised all 19 WorldClim bioclimatic variables. This comprised variables describing annual trends in temperature and precipitation, climatic seasonality, and extreme conditions that might limit species distributions.

2. BC2: a dataset using a reduced subset of the BC1 dataset, excluding complex composite variables (secondarily derived from annual or monthly means in temperature or precipitation; Table S4.1). This dataset comprised annual mean temperature (bio1), maximum temperature of warmest month (bio5), minimum temperature of coldest month (bio6), annual precipitation (bio12), precipitation of wettest month (bio13), and precipitation of driest month (bio14).

3. IPCC: a dataset using all nine IPCC variables, comprising mean annual temperature (tmp), diurnal temperature range (dtr), mean daily maximum temperature (tmax), mean daily minimum temperature (tmin), ground frost frequency (frs), cloud cover (cld), mean annual precipitation (pre), wet days (wet), and vapour pressure (vap).

Additionally, in all three datasets, we included additional variables summarizing topographic and streamline characteristics, namely altitude, slope and total upstream area (Lehner & Grill, 2013). These variables have been previously shown to be informative in modelling freshwater species distributions, and served as indicators of habitat integrity (e.g., presence of refugia), hydrologic regimes (e.g., flow velocity) and channel properties (e.g., size of water bodies, stream order) (McGarvey et al., 2018).

While some studies recommend the removal of highly collinear predictor variables prior to analyses (e.g., Dormann et al., 2013), we did not do so because removing correlated variables may confound direct comparisons of our predictor datasets across species. This is given that interspecific differences in collinearity patterns (Appendix S7) would necessitate the removal of different numbers of correlated variables across species, leading to varying numbers of final predictors utilized per species for each predictor dataset. Additionally, in the absence of detailed knowledge of the biology of our study species, the selection of which collinear variables to remove is not straightforward and can be a source of additional error.
Recent studies (e.g., Feng et al., 2019) have shown that Maxent is highly robust to the effects of predictor collinearity during model calibration. Nonetheless, to better understand how predictor datasets can differentially influence model transferability, we quantified predictor collinearity and collinearity shifts across geographic regions by calculating pairwise Pearson's correlation coefficients (Pearson's r) between predictor variables, and differences in knot correlations between native- and invasive-range background extents, using the raster.cor.plot function in the R package “EnMTools” (Warren et al., 2010). Additionally, we quantified the degree of environmental similarity between training localities (native and invasive occurrences) and entire background extents using multivariate environmental similarity surface (MESS) maps (Elith et al., 2010), to compare the extent of extrapolation (i.e., environmental novelty) between predictor datasets.

### 2.3 Model fitting and tuning

Maxent can model complex species responses via transformations of predictor variables (i.e., features). Six feature classes (FCs) are available—linear (the untransformed variable itself), quadratic (square of the variable), product (product of two variables), threshold (a “step” function generating a different constant function above a threshold or knot; equivalent to a piecewise constant spline), hinge (similar to threshold features but generates a linear function above the knot; equivalent to a piecewise linear spline), and categorical (Elith et al., 2011). By default, Maxent determines the FCs allowed in model calibration based on the number of occurrence points, though this has the potential to create complex fitted functions that might overfit the training data and impact model transferability (Merow et al., 2013). Additionally, Maxent uses a procedure known as L1-regularization to select individual features that contribute most to model fit while penalizing for excess parameters (Elith et al., 2011; Merow et al., 2013). The default regularization parameters in Maxent were based on tuning experiments conducted on a large number of species across multiple taxonomic groups (Phillips & Dudík, 2008), though recent studies have shown that default settings can result in overfitted models for many applications (Moreno-Amat et al., 2015; Radosavljevic & Anderson, 2014; Warren & Seifert, 2011; Warren et al., 2014).

For each combination of species and environmental dataset, models were built in Maxent version 3.3.3 k using default settings and three methods of tuning for appropriate regularization multipliers (RM) while differentially constraining the types of FCs allowed in model fits:

1. “Default”: Models were built using current Maxent defaults—regularization multiplier (RM) of 1 and using linear, quadratic, hinge and product features (LQHP) for models calibrated using > 80 occurrence points (Phillips et al., 2017).
2. “RM-only”: Models were built using default FCs (LQHP) and RM values ranging from 0 to 5 with increments of 0.5. Following the methodology of Radosavljevic and Anderson (2014), we used low overfitting as the primary criteria, and high discriminatory ability as a secondary criteria to select for optimal model parameters. Specifically, from each set of initial candidate models constructed per species/predictor dataset combination, we retained the model that minimized the difference between training and testing AUC (AUCDIFF) and reduced 10% omission rates (OR10) to the lowest observed value (or close to it) (primary criteria), while also having an evaluation AUC value (AUCTEST) that was close to the highest observed value (secondary criterion).
3. “RM + Hinge”: Models were built using only hinge features and RM values ranging from 0 to 5 with increments of 0.5. Models built using only hinge features tend to produce relatively smooth fitted response curves, similar to a generalized additive model (GAM) (Elith et al., 2010, 2011). We selected for optimal regularization parameters following the methodology of Radosavljevic and Anderson (2014), with low AUCDIFF and OR10 as the primary criteria and high AUCTEST as the secondary criterion.
4. “Full-tuning”: Models were built using RM values ranging from 0 to 5 with increments of 0.5, and seven FC combinations—linear only (L); linear and quadratic (LQ); hinge only (H); hinge and quadratic (HQ); hinge and product (HQP); and linear, quadratic, hinge and product features (LQHP; Maxent default). We selected for the optimal combination of RM and FCs following the methodology of Radosavljevic and Anderson (2014), with low AUCDIFF and OR10 as the primary criteria and high AUCTEST as the secondary criterion.

Following model tuning, we retained four final selected models ("Default," "RM-only," "RM + Hinge," "Full-tuning") for each species/predictor dataset combination. All subsequent downstream comparisons (e.g., between predictor datasets and tuning methods) and analyses were only based on final selected models.

### 2.4 Evaluation of model performance

Model transferability and overall performance were assessed using seven evaluation metrics that reflected aspects of discriminatory ability (AUCTEST), overfitting (AUCDIFF), omission errors (OR10, ORMIN), niche conservatism (Schoener’s D), predictive accuracy (Boyce Index) and overall parsimony (AICc; Table 1 for details). AICc scores were calculated on full unpartitioned species occurrence datasets (Warren & Seifert, 2011; Warren et al., 2014), whereas AUCTEST, AUCDIFF, OR10, ORMIN and Boyce Index scores were calculated using masked geographically structured cross-validations, which are transferability assessments in the strictest sense as calibration and testing are carried out using occurrences from different geographic regions (Radosavljevic & Anderson, 2014). Species occurrences were first partitioned into two spatially explicit bins—(a) native-range records and (b) invasive-range records—and each bin was in turn used to calibrate a model while the other was withheld for testing. Schoener’s D was calculated between geographic
predictions of the two spatial folds, with a high value indicating that models separately calibrated using each occurrence bin (native- or invasive-range records) resulted in similar geographic predictions (Schoener, 1968; Warren et al., 2010). “Clamping” was performed during model projections into testing regions, which treats environmental conditions not encountered during model training as though they were at the limits of the training range. This holds fitted species responses at constant probabilities outside of training conditions, therefore restricting model extrapolations when projecting into novel environments (Elith et al., 2011). Geographically structured cross-validations, as well as calculations of metric scores for $\text{AUC}_{\text{TEST}}$, $\text{AUC}_{\text{DIFF}}$, OR$_{10}$, OR$_{\text{MIN}}$, and AICc, were performed using the ENMeval function ("user" partitioning option) in the R package “ENMeval” (Muscarella et al., 2014), whereas metric scores for Schoener’s D and Boyce Index were calculated using the nicheOverlap function in the R package “dismo” (Hijmans et al., 2017) and the ecospat.boyce function in the R package “ecospat” (Di Cola et al., 2017), respectively.

| Evaluation metric | Description |
|-------------------|-------------|
| **Model transferability** | |
| $\text{AUC}_{\text{TEST}}$ | The area under the receiver operating characteristic curve (AUC) calculated using withheld testing data during cross-validation (Peterson et al., 2011). A higher value indicates a model’s increased ability to discriminate between conditions at withheld testing points and background points. |
| $\text{AUC}_{\text{DIFF}}$ | The difference between AUC values calculated using training points ($\text{AUC}_{\text{TRAIN}}$) and withheld testing points ($\text{AUC}_{\text{TEST}}$) during cross-validation (Warren & Seifert, 2011). A higher value indicates that a model is overfitted on training data such that it performs poorly when evaluated on independent testing data. |
| OR$_{10}$ | The threshold-dependent 10% training omission rate, defined as the percentage of withheld testing points with Maxent suitability scores lower than that of the 10% of training points with the lowest predicted suitability scores (Peterson et al., 2011; Radosavljevic & Anderson, 2014). A higher value indicates a model’s reduced ability to identify suitable habitats in testing regions. |
| OR$_{\text{MIN}}$ | The threshold-dependent minimum training presence omission rate, defined as the percentage of withheld testing points with Maxent suitability scores lower than the training point with the lowest predicted suitability score (Peterson et al., 2011; Radosavljevic & Anderson, 2014). A higher value indicates a model’s reduced ability to identify suitable habitats in testing regions, though this metric may overestimate model performance if some training occurrences are drawn from sink populations. |
| **Schoener’s D** | The degree of similarity between geographic predictions of sub-models calibrated separately using native- and invasive-range occurrences, and projected across accessible areas (background extents) for each species (Schoener, 1968; Warren et al., 2010). A higher value indicates a greater degree of similarity between geographic predictions of the spatial folds (i.e., greater geographic niche conservatism), but does not reflect model fit (e.g., a high value will be obtained if native- and invasive-range sub-models both grossly under- or overpredict suitability). |
| **Boyce Index** | The Continuous Boyce Index, which measures concordance between predicted habitat suitability gradients and the distribution of withheld testing points (Hirzel et al., 2006). This metric requires only presence data and is threshold-independent. A higher value indicates that model predictions are consistent with the observed distribution of presences in testing regions. |
| **Model parsimony** | |
| AICc | The sample-size-adjusted Akaike’s information criterion as adapted for ENMs (Warren & Seifert, 2011), which measures model overall goodness-of-fit while penalizing for excessive complexity. AICc is calculated using the full unpartitioned dataset and does not indicate model transferability. The model with the lowest AICc value ($\Delta\text{AICc} = 0$) is considered the most parsimonious out of a set of models fitted using the same occurrence dataset. |
To ensure that model performance was directly comparable across multiple species, we standardized metric scores (with the exception of AICc) of all selected models retained for a particular species (12 selected models per species, i.e., 3 predictor datasets × 4 tuning methods) by scaling them by their standard deviation and centring them around their per-species means, using the formula: \[ \text{standardized } X = \frac{X - \text{mean}(X)}{\text{SD}(X)} \], where \( X \) is a metric, and \( \text{mean}(X) \) and \( \text{SD}(X) \) are the mean value and standard deviation, respectively, of \( X \) across all selected models for a particular species. A standardized score of 0 indicates average scores per species. Solid black lines indicate median values, whereas boxes encompass the interquartile range. Whiskers extend to data extremes, excluding outliers. Outliers are indicated by open circles. Higher AUCTEST, Schoener’s D and Boyce Index scores, and lower AUCDIFF OR_{10} and OR_{MIN} scores indicate better model performance.

**FIGURE 1** The effect of environmental predictor datasets on Maxent model (a) discriminatory ability, (b) degree of overfitting, (c–d) omission error rates, (e) geographic niche conservatism and (f) predictive accuracy. Only final selected models were used for comparisons. Statistically significant pairwise differences, based on post hoc Tukey comparisons following mixed-effects ANOVA, are marked with different letters. To ensure that model performance was directly comparable across multiple species, we standardized metric scores of all selected models for a particular species by scaling them by their standard deviation and centring them around their per-species means, using the formula: \[ \text{standardized } X = \frac{X - \text{mean}(X)}{\text{SD}(X)} \], where \( X \) is a metric, and \( \text{mean}(X) \) and \( \text{SD}(X) \) are the mean value and standard deviation, respectively, of \( X \) across all selected models for a particular species. A standardized score of 0 indicates average scores per species. Solid black lines indicate median values, whereas boxes encompass the interquartile range. Whiskers extend to data extremes, excluding outliers. Outliers are indicated by open circles. Higher AUCTEST, Schoener’s D and Boyce Index scores, and lower AUCDIFF OR_{10} and OR_{MIN} scores indicate better model performance.

3 | RESULTS

Model calibration using different predictor datasets and tuning methods across all five focal species resulted in a final set of 60 selected models—45 models tuned for optimal Maxent parameters (RM and/or FCs) and 15 models built using default Maxent settings (Table S5.2). Full predictions (calibrated on unpartitioned occurrence datasets) of selected models can be found in Figures S6.2–S6.6.

3.1 | Effects of predictor dataset

The effects of predictor dataset on model transferability are shown in Figure 1 (top panels of Figures S9.13–S9.17 in Supporting
Information for results of individual species. Compared to models built using all 19 WorldClim bioclimatic variables (BC1), models calibrated on the BC2 and IPCC datasets had higher discriminatory ability (higher AUC_{TEST}), less overfitting (lower AUC_{DIFF}), lower omission errors (lower OR_{10} and OR_{MIN}) and a greater degree of correspondence across native and invasive ranges (higher Schoener’s D). BC1 models had lower predictivity accuracy (lower Boyce Index) as compared to IPCC models, though there were no significant differences in either with BC2 models. Both BC2 and IPCC models had similar levels of performance across all transferability metrics, whereas model parsimony (AICc ranks) did not differ substantially across predictor datasets (Figure S8.12). Mixed-effects ANOVA showed that differences were statistically significant for AUC_{TEST} (p < .01), AUC_{DIFF} (p < .001), OR_{10} (p < .001), OR_{MIN} (p < .01), Schoener’s D (p < .001) and Boyce Index (p < .01), but not for AICc (p = .733).

Analyses of predictor correlations showed that the BC1 dataset had the highest number of collinear variables among the three predictor datasets (31 variable pairs with Pearson’s r > .7), followed by the IPCC dataset (11 variable pairs with Pearson’s r > .7) and the BC2 dataset (4 variable pairs with Pearson’s r > .7; Figure 2; Figures S7.7–S7.11 in Supporting Information for correlation analyses of individual species). The BC1 dataset also exhibited the greatest dissimilarity in predictor correlations between native and invasive ranges (27 variable pairs with cross-geographic change in Pearson’s r > .3),
followed by the IPCC dataset (8 variable pairs with cross-geographic change in Pearson’s $r > .3$) and the BC2 dataset (2 variable pairs with cross-geographic change in Pearson’s $r > .3$).

The MESS analyses (Figures S11.23–S11.27) showed that, in general, the BC1 variables exhibited the highest degree of environmental novelty (cf. BC2 and IPCC variables) when comparing between conditions across entire background extents and those at training localities. On the other hand, the BC2 dataset consistently showed the lowest degree of environmental extrapolation across all five focal species, whereas the IPCC variables were somewhat intermediate between BC1 and BC2 datasets.

### 3.2 Effects of tuning method

Model tuning tended to select for RM values greater than the Maxent default of 1 and fewer FCs than Maxent defaults. Across all five focal species, the “Full-tuning” method selected for seven models built using only linear features (BC2 and IPCC models for *O. mossambicus*; BC1, BC2 and IPCC models for *L. catesbeianus*; BC2 model for *C. quadricarinatus*; IPCC model for *P. clarkii*), seven models built using only hinge features (BC1, BC2 and IPCC models for *C. gariepinus*, BC1 and IPCC models for *C. quadricarinatus*; BC1 and BC2 models for *P. clarkii*), and one model built using hinge, quadratic and product features (BC1 model for *O. mossambicus*; Table S5.2). The effects of model-tuning method on model transferability are shown in Figure 3 (bottom panels of Figures S9.13–S9.17 in Supporting Information for results of individual species). All model-tuning methods led to higher discriminatory ability (higher $\text{AUC}_{\text{TEST}}$), less overfitting (lower $\text{AUC}_{\text{DIFF}}$), and lower 10% omission errors (lower $\text{OR}_{10}$), as compared to “Default” models. Incorporating some form of feature selection in the tuning process (“RM + Hinge” and “Full-tuning”) further increased geographic niche conservatism (higher Schoener’s D) and parsimony (lower AICc ranks; Figure S8.12), as compared to models calibrated using default FCs (“Default” and “RM-only”). Additionally, “Full-tuning” models had less overfitting (lower $\text{AUC}_{\text{DIFF}}$) and lower 10% omission rates (lower $\text{OR}_{10}$) compared to all other model-tuning methods. In contrast, model tuning had no significant effect on minimum training presence omission rates ($\text{OR}_{\text{MIN}}$) and predictive accuracy (Boyce Index). Mixed-effects ANOVA showed that...
3.3 | Relationship between AICc and model transferability

We found that model parsimony (as indicated by AICc) was not a significant predictor of model discrimination ($\text{AUC}_{\text{TEST}}$, $p = .170$), minimum training presence omission error rates ($\text{OR}_{\text{MIN}}$, $p = .526$) and predictive accuracy (Boyce Index, $p = .100$). On the other hand, models with lower AICc (i.e., more parsimonious models) generally exhibited less overfitting (lower $\text{AUC}_{\text{DIFF}}$, $p < .05$), lower 10% omission rates (lower $\text{OR}_{10}$, $p < .05$) and greater correspondence across geographic regions (higher Schoener’s D, $p < .001$), although none of these relationships were particularly strong (absolute Pearson’s $r = .26$–.45; Figure 4). In contrast, pairwise comparisons of metrics indicating model transferability ($\text{AUC}_{\text{TEST}}$, $\text{AUC}_{\text{DIFF}}$, $\text{OR}_{10}$, $\text{OR}_{\text{MIN}}$, Schoener’s D and Boyce Index) found highly concordant relationships across all comparisons ($p < .001$ in all cases; absolute Pearson’s $r = .41$–.93).

4 | DISCUSSION

4.1 | Predictor complexity

Compared to the models fitted using simpler predictor datasets (BC2 and IPCC), models built using all 19 WorldClim bioclimatic variables (BC1), a common practice in many ENM studies (Warren et al., 2014), were generally overfitted on training data and performed poorly when transferred to testing regions. Poor model transferability has sometimes been attributed to shifts in species’ fundamental or realized niches (Broennimann et al., 2007), although the improved performances of our BC2 and IPCC models suggest that predictor

---

**FIGURE 4** Pairwise relationships between model transferability ($\text{AUC}_{\text{TEST}}$, $\text{AUC}_{\text{DIFF}}$, $\text{OR}_{10}$, $\text{OR}_{\text{MIN}}$, Schoener’s D and Boyce Index) and parsimony (AICc rank) for all five focal species. Upper diagonals show pairwise Pearson’s correlation coefficient (Pearson’s $r$) between metrics. Only final selected models were used for comparisons. To ensure that model performance was directly comparable across multiple species, we standardized metric scores (with the exception of AICc) of all selected models for a particular species by scaling them by their standard deviation and centring them around their per-species means, using the formula: $\frac{X - \text{mean}(X)}{\text{SD}(X)}$, where $X$ is a metric, and mean($X$) and $\text{SD}(X)$ are the mean value and standard deviation, respectively, of $X$ across all selected models for a particular species. AICc values were standardized by calculating their within-species relative rankings (i.e., the model with the lowest AICc per species ($\Delta\text{AICc} = 0$) is given an AICc rank of 1). Higher $\text{AUC}_{\text{TEST}}$, Schoener’s D and Boyce Index scores, and lower $\text{AUC}_{\text{DIFF}}$, $\text{OR}_{10}$, $\text{OR}_{\text{MIN}}$ and AICc rank scores indicate better model performance.
choice may be the more parsimonious explanation for the low predictive ability of BC1 models (Peterson & Nakazawa, 2008).

Visual examinations of species response curves (Figures S10.18–10.22) and MESS maps (Figures S11.23–S11.27) showed that the BC1 dataset resulted in more complex modelled responses (with sharper changes in response curves), as well as higher levels of environmental discordance (i.e., novelty), across native and invasive ranges. Consequently, these led to species–environment relationships that were overfitted on training datasets and extrapolated poorly onto testing data. In contrast, the simpler BC2 and IPCC variables exhibited higher environmental similarity between training and testing areas, leading to species–environment relationships that were more transferable across regions. Additionally, correlation analyses showed that the BC1 variables exhibited the highest degree of collinearity, as well as the most pronounced collinearity shifts between native and invasive ranges, among all three predictor datasets. While Maxent appears to be robust to predictor collinearity when correlation patterns do not substantially change across regions, model transferability has been shown to be negatively impacted by environmental novelty and collinearity shifts (Feng et al., 2019; Qiao et al., 2019). While not implying that complex predictors are biologically uninformative, the chances of environmental novelty and collinearity shifts are higher when using all 19 WorldClim variables in model calibration as compared to simpler datasets. Extraneous predictors, if correlated with causal variables limiting species distributions but truly having no effect on species responses, may be erroneously treated as important by the modelling algorithm and lead to non-predictivity when projected into non-analogue (novel) environments where correlation patterns differ (Rödder et al., 2009; Warren et al., 2014).

Comparisons of models constructed using the two simpler datasets showed that IPCC and BC2 models exhibited very similar overall performance across all evaluation metrics, and both correctly inferred habitat suitability in known areas of invasion where BC1 models did not (Figures S6.2–S6.6). This finding was unexpected, given the coarser native spatial resolution of IPCC variables (0.5°) as compared to BC2 variables (0.0416°). Our results corroborated previous studies highlighting that ENMs are generally insensitive to the effects of changes in predictor grain size (e.g., Guisan et al., 2007), although our spatial-filtering procedure (thinning to 50-km resolution) also likely ensured that few occurrence points would fall into the same original IPCC grid cells. Nonetheless, we expect that negative impacts of increased predictor grain sizes on ENM performance may become more pronounced when fine-scale environmental gradients or dispersal barriers shape species distributions, such as in the case of island endemics (Fourcade et al., 2014).

### 4.2 | Constraining fitted functions during model tuning

All model-tuning methods generally selected for higher RM values and led to improved model performance across all evaluation metrics, as compared to Maxent defaults. Our results corroborate those of previous studies showing that Maxent model performance frequently peaks at RM values higher than the default of 1, which is often insufficient to reduce model overparameterization and overfitting to acceptable levels (Elith et al., 2010; Radosavljevic & Anderson, 2014; Warren & Seifert, 2011). Additionally, while we expect the choice of FCs selected during model tuning to differ between models built using different sets of predictor variables, our results showed that the “Full-tuning” method produced models that were largely (93% of all “Full-tuning” models) built using only hinge or linear features (the latter which are special cases of hinge features; Phillips & Dudík, 2008), and exhibited the highest degrees of model transferability. It is not a new concept that models with overly complex species responses to environmental gradients will be overfitted on idiosyncrasies within input datasets and thus perform poorly when extrapolated beyond training regions (Elith et al., 2006; Warren et al., 2014), although fitted functions that are overly simple (i.e., underfitted) will also lead to predictions of low discriminatory ability (Barry & Elith, 2006). Allowing for feature flexibility during model fitting and using quantitative measures to select for optimal parameters appear to produce smooth responses (e.g., from hinge or linear features) that focus on the strongest ecological patterns while excluding spurious fits (Elith et al., 2010).

Nonetheless, we highlight caution when interpreting model projections into novel environmental space. Visual examination of our predictive outputs showed that linear-only “Full-tuning” models had the occasional tendency to overpredict habitat suitability in certain regions known to be suboptimal for our focal species, such as deserts and near latitudinal extremes (Figures S6.2–S6.6). By definition, linear features will be unable to characterize certain nonlinear aspects of species physiologies such as thermal tolerances (Merow et al., 2013), and can result in highly asymmetrical response curves predicting high suitability at one end of background environmental ranges (Figures S10.18–10.22). While model clamping (i.e., holding predicted suitability at constant beyond the limits of training environments) can restrict model extrapolations, it can still lead to highly erroneous predictions when environmental conditions in prediction regions are highly dissimilar to those in training regions (i.e., extremely novel conditions, so requiring a lot of extrapolations). We therefore emphasize that when model transference into unsampled regions is required, it will be necessary to quantify the degree of environmental novelty in prediction regions (e.g., by using multivariate environmental similarity surface maps) (Elith et al., 2010, 2011), as well as examine model outputs (e.g., predictive maps, response curves) for biological plausibility (Guevara et al., 2018).

### 4.3 | Relationship between AICc and model transferability

We did not find any significant relationship between AICc and AUC\textsubscript{TEST} or Boyce Index, although there were significant but weak effects on AUC\textsubscript{DIFF} OR\textsubscript{10} and Schoener’s D. In contrast, pairwise
comparisons of all metrics calculated solely on cross-validation data found highly concordant relationships. Taken together, these suggest that models chosen via AICc may have a higher tendency (cf. models selected using cross-validation metrics) to underfit on training data such that they overpredict wide areas of suitability in testing regions (hence low overfitting and greater niche overlap between native-/invasive-range sub-models), thereby biasing predictions towards higher errors of commission (i.e., predicting suitability in unsuitable areas).

While AICc has been shown to be useful in restricting models to a reasonable range of complexity (Moreno-Amat et al., 2015; Warren & Seifert, 2011), it can bias model selection towards certain feature types owing to its use of non-zero Maxent parameters as an estimate of degrees of freedom (Hastie et al., 2009). As tens to hundreds of hinge or threshold features can be fitted on a single covariate (as compared to a single linear or quadratic feature per predictor), models relying more heavily on these FCs will be expected to have considerably inflated parameter counts and therefore be excessively penalized by AICc (Elith et al., 2011; Merow et al., 2013). Consequently, model selection using AICc may produce models with insufficient parameters and/or overly simple response curves that do not reflect complex species-environment relationships. Additionally, AICc is calculated on the full unpartitioned occurrence dataset and does not account for spatial differences in environmental niche space (Rodda et al., 2011), and may therefore perform poorly at identifying important predictors when strong correlations exist and differ across study regions (Warren et al., 2014).

Our results corroborate a recent simulation study (Velasco & González-Salazar, 2019) demonstrating poor correspondence between AICc and geographical predictive accuracy. While their study found no correlation between AICc and six threshold-dependent evaluation metrics across nine simulated species, we found weak but significant relationships between model parsimony, and the degree of overfitting and geographic niche conservatism. We postulate that these differences may arise from discrepancies in species characteristics (i.e., virtual vs. empirical species), data partitioning methods (random partitioning vs. masked geographically structured cross-validations) or performance metrics (threshold-dependent vs. threshold-independent metrics) applied. Nonetheless, given that both simulated and empirical studies showed a tenuous link between AICc and other model evaluation metrics, we suggest that model parsimony may not be the optimal model-selection criterion when requiring projections into unsampled environments.

### 4.4 Implications for the ecological niche modeller

Our results demonstrate that methodological considerations in predictor choice and model parameterization can have substantial impacts on Maxent model transferability, and highlight the inherent uncertainties in many applications of ENMs. Nevertheless, given the urgent need to address conservation issues such as the impacts of climate change or biological invasions (e.g., Rodda et al., 2011; Warren et al., 2014), imprecise ENMs are often the best tools available to quantify ecological or evolutionary change in impacted landscapes and can help guide conservation interventions (Warren, 2012).

Our comparisons between different predictor datasets suggest that, in the absence of detailed knowledge of the physiology and ecology of study species or if modelling large numbers of species simultaneously (e.g., community-level studies), simpler “standard” environmental datasets (e.g., our BC2 or IPCC datasets) may be the preferable default in place of comprehensive but overly complex datasets (e.g., all 19 WorldClim variables). The datasets utilized in this study aside, other global environmental data sources, such as the normalized difference vegetation index (NDVI) dataset (Tucker, 1979), have been shown to perform well in some cases (Peterson & Nakazawa, 2008) and may be worth consideration to highlight areas of agreement and quantify uncertainty in modelling outputs. Heuristic methods of predictor selection (e.g., Warren et al., 2014; Zeng et al., 2016) may provide additional utility in situations where using “standard” datasets is the only practical option, although we emphasize that these techniques have to be applied cautiously as they often remove extraneous predictors based on heuristically defined variable contribution scores. Such methods may be more robust against spurious model fits if initialized from a condensed set of biologically plausible predictors.

We also found that incorporating feature selection into model tuning can substantially improve model transferability as the type of feature functions limiting species distributions will be expected to differ from species to species, and using Maxent defaults (LQHP) may introduce noise that overfit onto idiosyncrasies or errors within occurrence datasets (e.g., georeferencing errors, sampling bias). We therefore recommend that feature selection be considered an integral part of the model-tuning process though we emphasize that detailed examinations of model outputs, on top of quantitative measures of model performance, will always be necessary to determine biological plausibility.

Our results also indicate a somewhat tenuous relationship between AICc and model transferability, suggesting that the indiscriminate use of AICc in model selection may lead to erratic model performance. Cross-validation may generally be preferable because it allows for straightforward evaluations of model transferability (arguably the primary objective for most ENM applications), and quantitative measures of predictive accuracy and uncertainty. “Traditional” measures of model performance, such as presence-background AUC or threshold-dependent omission rates, have previously been criticized as providing misleading indications of model fits (Merow et al., 2013), although we found high correspondences with the presence-only, threshold-independent Boyce Index. Nonetheless, further empirical testing using simulated or real species across a wide range of taxonomic groups and habitats will be needed to determine the generality of our observations.

Lastly, while shifts in species’ realized ecological niches undoubtedly occur when species undergo range shifts (e.g., Broennimann et al., 2007), we highlight that inferences of niche divergence or conservatism (i.e., effect sizes of niche shifts) can be
sensitive to choices in environmental predictors and model parameters. For example, while we made no prior assumptions of niche conservatism for our five focal freshwater invasive species, models exhibited varying levels of cross-geographic niche similarity dependent on combination of predictor dataset and tuning method (which in the most extreme case can range from low niche overlap (Schoener’s D = 0.2–0.4) to very high niche overlap (Schoener’s D = 0.8–1.0), even within a single species; niche overlap categories following Rödder and Engler (2011)). Conclusions of realized niche shifts based on arbitrary sets of predictors or model parameters need to be examined in the context of ecological processes and fundamental-niche relationships that limit species distributions (Rödder et al., 2009). ENM users should therefore prioritize understanding the underlying biotic and abiotic mechanisms limiting the survival and dispersal of their study species in order to make reliable inferences of ecological niches.

ACKNOWLEDGEMENTS
We thank Dan Warren, Rudolf Meier, Ryan Chisholm and Chong Kwek Yan for valuable insights and comments on early drafts of the manuscript. This study was supported by an AcRF Tier 1 Grant from the Singapore Ministry of Education (National University of Singapore Grant Number R-154-000-633-112) and the Ah Meng Memorial Conservation Fund (National University of Singapore Grant Number R-154-000-617-720). BW Low was supported by additional funding from the Lee Kong Chian Natural History Museum, National University of Singapore. Y Zeng acknowledges support from the National Research Foundation (NRF) Singapore under its Commonwealth Research Fellowship Grant (Grant Number NRF-CSC-ICFC2017-05).

DATA AVAILABILITY STATEMENT
Data sources for occurrence localities and details on dispersal rates for our five study species can be found in Appendices S1 and S3, respectively. Climatic layers were obtained from the WorldClim bioclimatic dataset version 1.4 (Hijmans et al., 2005) (https://www.worldclim.org/) and the Intergovernmental Panel on Climate Change (IPCC) CRU high-resolution climate dataset version 2.1 (Mitchell & Jones, 2005) (http://www.ipcc-data.org). Aggregated occurrence datasets, background points and R codes for model tuning/evaluation are available at: https://doi.org/10.5061/dryad.ttdz08kww.

ORCID
Bi Wei Low  
https://orcid.org/0000-0002-2549-4938
Yiwen Zeng  
https://orcid.org/0000-0001-7787-314X
Heok Hui Tan  
https://orcid.org/0000-0003-2227-8866
Darren C.J. Yeo  
https://orcid.org/0000-0003-2669-4349

REFERENCES
Acevedo, P., Melo-Ferreira, J., Real, R., & Alves, P. C. (2014). Evidence for niche similarities in the allopatric sister species Lepus castroviejoi and Lepus corsicanus. Journal of Biogeography, 41, 977–986.
Aielo-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). sptThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography, 38, 541–545. https://doi.org/10.1111/ecog.01132
Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33, 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x
Araújo, M. B., & Pearson, R. G. (2005). Equilibrium of species’ distributions with climate. Ecography, 28, 693–695. https://doi.org/10.1111/j.2005.0906-7590.04253.x
Austin, M. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. Ecological Modelling, 157, 101–118. https://doi.org/10.1016/S0304-3800(02)00205-3
Barry, S., & Elith, J. (2006). Error and uncertainty in habitat models. Journal of Applied Ecology, 43, 413–423. https://doi.org/10.1111/j.1365-2664.2006.01136.x
Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modelling and species distribution modelling. Ecological Modelling, 222, 1810–1819.
Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological Modelling, 275, 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012
Both, C., Lingnau, R., Santos-Jr, A., Madalozzo, B., Lima, L. P., & Grant, T. (2011). Widespread occurrence of the American bullfrog, Lithobates catesbeianus (Shaw, 1802) (Anura: Ranidae), in Brazil. South American Journal of Herpetology, 6, 127–134.
Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensemble of small models. Methods in Ecology and Evolution, 6, 1210–1218.
Broennimann, O., Trier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T., & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. Ecology Letters, 10, 701–709. https://doi.org/10.1111/j.1461-0248.2007.01060.x
Cardoso-Leite, R., Vilarinho, A. C., Novaes, M. C., Tonetto, A. F., Vilardi, G. C., & Guillermo-Ferreira, R. (2014). Recent and future environmental suitability to dengue fever in Brazil using species distribution model. Transactions of the Royal Society of Tropical Medicine & Hygiene, 108, 99–104. https://doi.org/10.1093/trstmh/trt115
Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D’Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N., & Guisan, A. (2017). eco-spat: An R package to support spatial analyses and modelling of species niches and distributions. Ecography, 40, 774–787. https://doi.org/10.1111/ecog.02671
Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marqués, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x
Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettemann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., ... E. Zimmermann, N. (2006). Novel methods improve prediction of species’ distributions from occurrence data. Ecography, 29, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x
Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution, 1*, 330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and predictions across space and time. *Annual Review of Ecology, Evolution, and Systematics, 40*, 677–697.

Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions, 17*, 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Feng, X., Park, D. S., Liang, Y., Pandey, R., & Papes, M. (2019). Collinearity in ecological niche modeling: Confusions and challenges. *Ecology and Evolution, 9*, 10365–10376. https://doi.org/10.1002/ece3.5555

Ficetola, G. R., Thuiller, W., & Mlaud, C. (2007). Prediction and validation of the potential global distribution of a problematic alien invasive species—the American bullfrog. *Diversity and Distributions, 13*, 476–485.

Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS One, 9*, e97122. https://doi.org/10.1371/journal.pone.0097122

Galante, P. J., Alade, B., Muscarella, R., Jansa, S. A., Goodman, S. M., & Anderson, R. P. (2018). The challenge of modeling niches and distributions for data-poor species: A comprehensive approach to model complexity. *Ecography, 41*, 726–736. https://doi.org/10.1111/ecog.02909

Glor, R. E., & Warren, D. (2010). Testing ecological explanations for biogeographic boundaries. *Evolution, 65*, 673–683. https://doi.org/10.1111/j.1558-5646.2010.01177.x

Guevara, L., Gerstner, B. E., Kass, J. M., & Anderson, R. P. (2018). Toward ecologically realistic predictions of species distributions: A cross-time example from tropical montane cloud forests. *Global Change Biology, 24*, 1511–1522. https://doi.org/10.1111/gcb.13992

Guisan, A., Graham, C. H., Elith, J., & Huettman, F. & NCEAS Species Distribution Modelling Group (2007). Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions, 13*, 332–340. https://doi.org/10.1111/j.1472-4642.2007.00342.x

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction (2nd ed.). Springer series in statistics, (p. 1–745). New York, NY: Springer-Verlag.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology, 25*, 1965-1978. https://doi.org/10.1002/joc.1276

Hijmans, R. J., Phillips, S. J., Leathwick, J. R., & Elith, J. (2017). dismo: Species distribution modelling. https://cran.r-project.org/package=dismo.

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presence. *Ecological Modelling, 199*, 142–152. https://doi.org/10.1016/j.ecolmodel.2005.05.017

Hoathorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*, 346–363. https://doi.org/10.1002/bimj.200810425

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinefelder, V., Stillfried, M., Heckmann, I., Scharf, A. K., Augeri, D. M., Cheyne, S. M., Hearn, A. J., Ross, J., Macdonald, D. W., Mathai, J., Eaton, J., Marshall, A. J., Semladi, G., Rustam, R., ... Wilting, A. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions, 19*, 1366–1379. https://doi.org/10.1111/ddi.12096

Lehner, B., & Grill, G. (2013). Global river hydrology and network routing: Baseline data and new approaches to study the world’s large river systems. *Hydrological Processes, 27*, 2171–2186.

Lodge, D. M., Deines, A., Gherardi, F., Yeo, D. C. J., Arcella, T., Badridge, A. K., Barnes, M. A., Chadderton, W. L., Feder, J. L., Gantz, C. A., Howard, G. W., Jerde, C. L., Peters, B. W., Peters, J. A., Sargent, L. W., Turner, C. R., Wittmann, M. E., & Zeng, Y. (2012). Global introductions of crayfishes: Evaluating the impact of species invasions on ecosystem services. *Annual Review of Ecology, Evolution, and Systematics, 43*, 449–472. https://doi.org/10.1146/annurev-ecolsys-111511-103919

Lowe, S., Browne, M., Boudjelas, S., & de Poorter, M. (2000). 100 of the world’s worst invasive alien species: A selection from the Global Invasive Species Database. (p. 1–12). Auckland, New Zealand: Invasive Species Specialist Group, Species Survival Commission, World Conservation Union. https://www.iucn.org/content/100-worlds-worst-invasive-alien-species-a-selection-global-invasive-species-database

McGarvey, D. J., Menon, M., Woods, T., Tassone, S., Reese, J., Vergamini, M., & Kellogg, E. (2018). On the use of climatic covariates in aquatic species distribution models: Are we at risk of throwing out the baby with the bath water? *Ecography, 41*, 695–712.

Merow, C., Smith, M. J., & Silander, J. A. Jr (2013). A practical guide to MaxEnt for modeling species’ distributions: What it does, and why inputs and settings matter. *Ecography, 36*, 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Mitchell, T. D., & Jones, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology, 25*, 693–712. https://doi.org/10.1002/joc.1181

Moreno-Amat, E., Mateo, R. G., Nieto-Lugilde, D., Morueta-Home, N., Svenning, J.-C., & García-Amorena, I. (2015). Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. *Ecological Modelling, 312*, 308–317. https://doi.org/10.1016/j.ecolmodel.2015.05.035

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods in Ecology and Evolution, 5*, 1198–1205.

Palaoro, A. V., Dalosto, M. M., Costa, G. C., & Santos, S. (2013). Niche conservatism and the potential for the crayfish Procambarus clarkii to invade South America. *Freshwater Biology, 58*, 1379–1391.

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography, 34*, 102–117.

Peterson, A. T., & Nakazawa, Y. (2008). Environmental data sets matter in ecological niche modelling: An example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecology and Biogeography, 17*, 135–144.

Peterson, A. T., Papes, M., & Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography, 30*, 550–560. https://doi.org/10.1111/j.0906-7590.2007.05102.x

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martinez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). Ecological niches and geographic distributions, Monographs in population biology 49, (p. 1–328). Princeton: Princeton University Press.

Phillips, S. J., Anderson, R. P., Dudik, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography, 40*, 887–893. https://doi.org/10.1111/ecog.03049

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026
Phillips, S. J., & Dudik, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, 31, 161–175. https://doi.org/10.1111/j.0906-7590.2008.05203.x

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2019). nlme: Linear and nonlinear mixed effects models. https://cran.r-project.org/package=nlme.

Qiao, H., Feng, X., Escobar, L. E., Peterson, A. T., Soberón, J., Zhu, G., & Papes, M. (2019). An evaluation of transferability of ecological niche models. *Ecography*, 42, 521–534. https://doi.org/10.1111/ecog.03986

Radosavljevic, A., & Anderson, R. P. (2014). Making better MAXENT model predictions. *Ecological Modelling*, 27, 847–857. https://doi.org/10.1016/j.ecolmodel.2014.10.014

Russell, D. J., Thuesen, P. A., & Thomson, F. E. (2012). A review of the environmental niche modelling framework. *Methods in Ecology and Evolution*, 3, 73–78. https://doi.org/10.1111/j.2041-210X.2011.00012.x

Rödder, D., & Engler, J. O. (2011). Quantitative metrics of overlaps in Grinnellian niches: Advances and possible drawbacks. *Global Ecology and Biogeography*, 20, 915–927. https://doi.org/10.1111/j.1466-8238.2011.00659.x

Rödder, D., Schmidtlein, S., Veith, M., & Lötters, S. (2009). Alien invasive snail turtle in unexplored habitat: A matter of niche shift or of predictors studied? *PLoS One*, 4, e7843.

Russell, D. J., Jhun, R. A., & Thomson, F. E. (2012). A review of the biology, ecology, distribution and control of Mozambique tilapia, Oreochromis mossambicus (Peters, 1852) (Pisces: Cichlidae) with particular emphasis on invasive Australian populations. *Reviews in Fish Biology and Fisheries*, 22, 533–554. https://doi.org/10.1007/s11160-011-9249-z

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1, 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

Schoener, T. W. (1968). The *Anolis* lizards of Bimini: Resource partitioning in a complex fauna. *Ecology*, 49, 704–726. https://doi.org/10.2307/1935534

Snovsky, G., & Galil, B. S. (2011). The Australian redclaw crayfish Cherax quadricarinatus (von Martens, 1868) (Crustacea: Decapoda: Parastacidae) in the Sea of Galilee, Israel. *Aquatine Invasions*, 6, 29–31. https://doi.org/10.1391/ai.2011.6.51.007

Soberón, J., & Peterson, A. T. (2006). Interpretation of models of fundamental ecological niches and species’ distributional areas. *Biodiversity Informatics*, 2, 1–10. https://doi.org/10.17161/bi.v2i010

Teugels, G. G. (1986). A systematic revision of the African species of the genus *Clarias* (Pisces: *Clariidae*). *Annales du Musee Royal de l’Afrique Centrale (Zoologie)*, 247, 1–199.

Torres, E., & Álvarez, F. (2012). Genetic variation in native and introduced populations of the red swamp crayfish Procambarus clarkii (Girard, 1852) (Crustacea, Decapoda, Cambaridae) in Mexico and Costa Rica. *Aquatic Invasions*, 7, 235–241. https://doi.org/10.3391/ai.2012.7.2.009

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Václavík, T., & Meentemeyer, R. K. (2009). Invasive species distribution modeling (ISDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling*, 220, 3248–3258. https://doi.org/10.1016/j.ecolmodel.2009.08.013

Velasco, J. A., & González-Salazar, C. (2019). Akaikie information criterion should not be a “test” of geographical prediction accuracy in ecological niche modelling. *Ecological Informatics*, 51, 25–32. https://doi.org/10.1016/j.ecoinf.2019.02.005

Warren, D. L. (2012). In defense of ‘niche modeling’. *Trends in Ecology & Evolution*, 27, 497–500. https://doi.org/10.1016/j.tree.2012.03.010

Warren, D. L., Gior, R. E., & Turelli, M. (2010). ENMTools: A toolbox for comparative studies of environmental niche models. *Ecography*, 33, 607–611. https://doi.org/10.1111/j.1600-0587.2009.06412.x

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21, 335–342. https://doi.org/10.1890/10-1171.1

Warren, D. L., Wright, A. N., Seifert, S. N., & Shaffer, B. H. (2014). Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Diversity and Distribution*, 20, 334–343. https://doi.org/10.1111/ddi.12160

Weyl, O. L. F., Daga, V. S., Ellender, B. R., & Vitule, J. R. S. (2016). A review of *Clarias gariepinus* invasions in Brazil and South Africa. *Journal of Fish Biology*, 89, 386–402.

Xu, T., & Hutchinson, M. F. (2013). New developments and applications in the ANUCLIM spatial climatic and bioclimatic modelling package. *Environmental Modelling & Software*, 40, 267–279. https://doi.org/10.1016/j.envsoft.2012.10.003

Zeng, Y., Low, B. W., & Yeo, D. C. J. (2016). Novel methods to select environmental variables in MaxEnt: A case study using invasive crayfish. *Ecological Modelling*, 341, 5–13. https://doi.org/10.1016/j.ecolmodel.2016.09.019

Zeng, Y., & Yeo, D. C. J. (2018). Assessing the aggregated risk of invasive crayfish and climate change to freshwater crabs: A Southeast Asian case study. *Biological Conservation*, 223, 58–67. https://doi.org/10.1016/j.biocon.2018.04.033

**Biosketch**

BW Low is part of the Freshwater & Invasion Biology Laboratory, National University of Singapore (NUS) (current affiliation: Lee Kong Chian Natural History Museum, NUS). His research interests include the ecology and taxonomy of Southeast Asian freshwater fishes, biological invasions, freshwater community ecology, and ecological niche modelling.

BWL and YZ developed the ideas, designed the study methodology and conducted data collections. BWL analysed the data and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

**Supporting Information**

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wei Low B, Zeng Y, Hui Tan H, Yeo DCJ. Predictor complexity and feature selection affect Maxent model transferability: Evidence from global freshwater invasive species. *Divers Distrib.* 2021;27:497–511. https://doi.org/10.1111/ddi.13211