Received: 18 December 2019 | Revised: 19 August 2020 | Accepted: 25 August 2020

DOI: 10.1111/ddi.13161

BIODIVERSITY RESEARCH

Diversity and Distributions WILEY

Predicting range expansion of invasive species: Pitfalls and best practices for obtaining biologically realistic projections

Thomas A. Lake | Ryan D. Briscoe Runquist | David A. Moeller

Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, USA

Correspondence
Thomas A. Lake and David A. Moeller, Department of Plant and Microbial Biology, University of Minnesota, 1479 Gortner Ave, St. Paul, MN 55108, USA.
Emails: Lakex055@umn.edu; Moeller@umn.edu

Funding information
Legislative-Citizen Commission on Minnesota Resources, Grant/Award Number: CON000000051140

Editor: Céline Bellard

Abstract

Aim: Species distribution models (SDMs) are widely used to forecast potential range expansion of invasive species. However, invasive species occurrence datasets often have spatial biases that may violate key SDM assumptions. In this study, we examined alternative methods of spatial bias correction and multiple methods for model evaluation for seven invasive plant species.

Location: North America.

Taxon: Common Tansy (*Tanacetum vulgare*), Wild Parsnip (*Pastinaca sativa*), Leafy Spurge (*Euphorbia virgata*), Common Teasel (*Dipsacus fullonum*), Brown Knapweed (*Centaurea jacea*), Black Swallowwort (*Vincetoxicum nigrum*) and Dalmatian Toadflax (*Linaria dalmatica*).

Methods: We employed bias-correction measures for both occurrence sampling and background sampling inputs in a factorial design for Maxent resulting in six potential models for each species. We evaluated our models for complexity, model fit and using commonly employed evaluation metrics: AUC, partial AUC, the continuous Boyce index and sensitivity. We then developed a structured process for model selection.

Results: Models developed without occurrence or background bias correction often were overly complex and did not transfer well to expanding range fronts. Conversely, models that employed occurrence and/or background bias-correction measures were less complex, had better AICc scores and had greater projection into incipient areas. These simpler models were also more likely to be selected when evaluated using a process that integrated multiple evaluation metrics. We found that invasion history (e.g. established versus incipient) was associated with the effectiveness of spatial bias correction techniques.

Main Conclusions: While challenges exist in building climate-based correlative species distribution models for invasive species, we found that methods relying on maximizing AUC performed poorly for invasive species. We advocate for the use of multiple and diverse metrics for model evaluation. Users of species distribution models need to incorporate explicit consideration of model discrimination, model fit and model complexity into their decision-making processes if they are to build biologically realistic models.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

KEYWORDS
Boyce index, distribution shifts, environmental niche model, exotic species, gaussian kernel distributions, Maxent, range expansion, sensitivity, spatial bias, transferability

1 | INTRODUCTION

The geographic ranges of invasive species are shifting, both because they are often in disequilibrium with the environment (i.e. limited range filling) and because climate change has altered the distribution of suitable habitat. Early detection and prevention increasingly rely on species distribution models (SDMs) that forecast suitable habitats where surveillance and eradication efforts can be focused (Lodge et al., 2016; Marbuah, Gren, & McKie, 2014). However, invasive species pose unique challenges that stymie the model building, evaluation and selection process, as invasive species commonly violate underlying model assumptions and contain multiple sources of spatial bias (Elith, Kearney, & Phillips, 2010). Although a variety of approaches have been used to correct these biases, it is poorly understood how alternative methods perform on a common set of invasive species.

Building SDMs of invasive species have been challenging for two main reasons. First, invasive species inherently violate the model assumption that species are in equilibrium with the environment and occupy all regions suitable for their persistence. Second, modelled relationships are correlative and only hold over the range of climate variables included in the model-building process; therefore, projecting into uninvaded or incipient regions outside of the current invasive range may be problematic (Elith et al., 2010; Elith & Leathwick, 2009; Hastie & Fithian, 2013; Lodge et al., 2016). Additionally, niche shifts are common between native and invaded ranges, which may lead to differences in the underlying environmental relationships in different regions of the global range (Le Maitre, Thuiller, & Schonegevel, 2008). Many researchers have attempted to address model violations (Broennimann et al., 2007; Fourcade, Besnard, & Secondi, 2018; Mainali et al., 2015; Petitpierre et al., 2012). Nevertheless, these methods have not completely addressed spatial biases in invasive species SDMs.

Uneven species sampling is a common source of bias that can impede accurate modelling (Bradley, Blumenthal, Wilcove, & Ziska, 2010; Elith & Leathwick, 2009; Fourcade, Engler, Rödder, & Secondi, 2014; Merckx, Steyaert, Vanreusel, Vincx, & Vanaverbeke, 2011). Sampling bias can arise if species are challenging to detect or not a priority for detection, which results in occupied regions having no documented occurrence records. In addition, spatial biases unique to invasive species may further complicate the modelling approach, such as recurrent eradication efforts, highly uneven sampling programs across political units, introduction into unsuitable habitats (Elith et al., 2010; Lodge et al., 2016) and introduction via stochastic or anthropogenic long-distance dispersal events (e.g. ship ballast) (von der Lippe & Kowarik, 2007). These biases manifest in oversampling in the established portion of the invaded range and undersampling in regions where the invasive species is incipient, leading to uneven prevalence across the invaded range. Identifying the sources of spatial bias in invasive species, and understanding how they interact with different modelling methods, may help to build more robust and reliable SDMs.

Among the many tools available for building SDMs, Maxent is the most widely used, as it is found to perform well for most species (Hao, Elith, Guillera-Arroita, & Lahoz-Monfort, 2019). Maxent contrasts environments of occurrence points with those sampled from background locations to determine which combinations of variables best predict occurrences (Elith et al., 2010, 2011; Merow, Smith, & Silander, 2013). One key assumption in SDMs is that sample prevalence (the frequency of sampled sites in the total study area) accurately represents species prevalence (the frequency of species over the total study area). Maxent uses a constant species prevalence value across the species' range, despite the fact that many species vary in prevalence geographically or have been documented with differing levels of sampling effort in different regions. While a number of methods have been developed to mitigate the potential effects of spatially biased occurrence records, two methods have shown promise: downsampling occurrence points and modifying the sampling distribution of background points (Kramer-Schadt et al., 2013; Phillips & Dudík, 2008).

Downsampling addresses spatial bias by thinning dense clusters of occurrences where oversampling may have occurred (Fourcade et al., 2014). In model building, this effectively reduces the environmental importance placed on oversampled regions (Fourcade et al., 2014). Alternatively, one can modify the distribution of background point sampling to match the spatial bias that exists in the occurrence points (Barbet-Massin, Jiguet, Hélène Albert, & Thuiller, 2012; Kramer-Schadt et al., 2013; Phillips & Dudík, 2008). Background sampling with a Gaussian Kernel Density (GKD) distribution can be used to sample a greater proportion of background points in areas of highly sampled occurrence points (Elith et al., 2010). GKD sampling draws more background points in areas with more occurrences, which can enable greater discriminating power in areas that are heavily sampled.

Model discrimination metrics, such as those relying on a confusion matrix, are commonly used to evaluate model performance. The area under the curve (AUC) is most frequently used due to its threshold-independence, but it has been criticized due to its equal weighting of commission and omission errors and dependence on prevalence (Leroy et al., 2018). Alternative metrics such as the partial AUC and sensitivity (TP/TP + FN) have been proposed in order to mitigate problems caused by assuming background points in presence-only modelling are errors of commission (Allouche, Tsoar, & Kadmon, 2006; Boyce, Vernier, Nielsen, & Schmiegelow, 2002; Jiménez-Valverde et al., 2011; Lobo, Jiménez-Valverde, & Real, 2008; Merckx et al., 2011; Radosavljevic & Anderson, 2014). Others have suggested that the Continuous Boyce Index (CBI) may be useful as a

spatially explicit metric to account for correlation between a models' prediction and occurrences (Boyce et al., 2002; Hirzel, Le Lay, Helfer, Randin, & Guisan, 2006). Last, model complexity should be considered in the model evaluation process; complex models tend to be overfit (i.e. high performance on training datasets, low performance on testing datasets) and often lack generalization (Yates et al., 2018; Brun, Thuiller, Chauvier, Pellissier & Wüest et al., 2019). Model complexity may be assessed by recording the number of model parameters or using information criteria, such as AICc, which balances model gains with penalties for added parameters (Akaike, 1974).

For invasive species, model selection needs to explicitly consider transferability because one goal of invasive species SDMs is to transfer predictions into incipient or unoccupied habitats (Morey & Venette, 2020; Wenger & Olden, 2012; Yates et al., 2018). One approach partitions cross-validation datasets using a spatially explicit block method, where the modelled area is divided into geographic regions that contain roughly the same number of occurrences (Wenger & Olden, 2012; Yates et al., 2018). Modellers can then estimate performance based on spatially explicit cross-validation and the overall model performance is a useful indicator of how well SDMs transfer between geographic regions (Finnoff, Shogren, Leung, & Lodge, 2007; Lodge et al., 2016; Yates et al., 2018). Additionally, the coefficient of variation (CoV) among cross-validated data partitions may provide insight into how well SDMs generalize among geographic regions, where greater CoV indicates that model performance varies depending on the geographic region.

For this study, we developed SDMs for seven invasive species with different invasion histories. We refer to these models as SDMs throughout the text, as we use climate variables to determine potential areas of occupancy, but do not attempt to model the niche globally. Our seven invasive species are of special interest in prevention and eradication efforts in the Upper Midwest of the United States. Our seven species represent three different invasion histories: (a) long established, widespread species that are likely in equilibrium with the environment, (b) incipient species with recent range expansion into the Upper Midwest that are diffusely distributed across the continental United States, and (c) incipient species with recent range expansion into the Upper Midwest that have irregular and clustered distributions. Using these species, we explored differences in model performance, model variability and model complexity and examined if these model characteristics interacted with different invasion history (e.g. established versus incipient invasions). The goal of the study was to determine how to build and select SDMs to make biologically sound predictions about habitat suitability that can be used by land managers for decision-making, especially for regions that may be challenging to model but are of utmost interest for surveillance and eradication efforts. We used a factorial design to systematically explore three general themes that affect the model building and selection process: (a) how did methods that reduce the effects of spatial bias affect measures of model complexity and amount of predicted suitable habitat? (b) how were model performance and variability affected by differences in model inputs? and (c) how did model selection affect projections of habitat suitability?

2 | METHODS

2.1 | Study species

We modelled seven terrestrial invasive plant species in North America with different invasion histories. Common Tansy (*Tanacetum vulgare* L.), Wild Parsnip (*Pastinaca sativa* L.) and Leafy Spurge (*Euphorbia virgata* L.) are historically established and common species throughout North America and the Upper Midwest. Common Teasel (*Dipsacus fullonum* L.) and Brown Knapweed (*Centaurea jacea* L.) are diffuse throughout North America but incipient to the Upper Midwest. Black Swallowwort (*Vincetoxicum nigrum* (L.) Moench 1802) and Dalmatian Toadflax (*Linaria dalmatica* L.) have clustered distributions throughout North America and are likewise incipient to the Upper Midwest. For detailed information on the species' native and invasive distributions, ecology and invasion history, see Table S1. All species presented in this paper are native to Europe and portions of western Asia, are considered invasive across the continental United States and are of special concern in the Upper Midwest (Minnesota, Wisconsin, Illinois, Iowa, North Dakota and South Dakota).

2.2 | Data sources

2.2.1 | Spatial data

We downloaded occurrence records from two publicly available databases: (a) Global Biodiversity Information Facility (GBIF; www.gbif.org), and (b) Early Detection & Distribution Mapping System (EDDMapS; www.eddmaps.org) (Table S2). We processed occurrence records to remove duplicate, erroneous and/or imprecise (e.g. gridded) coordinates. Data points with less than 0.1 decimal degrees of spatial precision were removed. Spatial analyses occurred in the R statistical environment (v3.5.1; R Core Team, 2018) using the Geospatial Data Abstraction Library ("rgdal"; Bivand, Keitt, & Rowlingson, 2019), Geometry Engine Open Source ("rgeos"; Bivand & Rundel, 2015), Spatial Points ("sp"; Pebesma & Bivand., 2005) and Raster packages ("raster" & vanEtten, 2012).

2.2.2 | Spatial data processing

Datasets were trimmed to include only the invaded range of North America. We used occurrence records from only the invaded range to capture unique niche shifts that may have occurred during invasion (Broennimann et al., 2007). Additionally, we produced a set of models containing the occurrences from both the native and invaded range, the results of which did not appreciably change modelling outcomes (See Supplementary Materials). To reduce pseudoreplication among the invaded-only occurrence records, we resampled occurrences to a baseline resolution of 1 km, the resolution of the climate data (*hereafter*: filtered datasets).

2.2.3 | Environmental data

We used bioclimatic variables available from WorldClim [30 arcsecond (~1 km) resolution; http://worldclim.org/version2] because they provide biologically meaningful axes of climate variation and are commonly used in SDMs (Table S3; Fick & Hijmans, 2017; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). We selected 3–5 bioclimatic variables per species that have been shown to encapsulate relevant axes of climate variation that affect plant establishment and growth as well as enhance model transferability for invasive species (Petitpierre, Broenniman, Kueffer, Daehler, & Guisan, 2017). We further eliminated variables that had >0.7 correlation across the modelled environmental space (Table S3).

2.3 | Species distribution modelling

2.3.1 | Occurrence record sampling and background point selection

For each species, we developed two occurrence datasets and three background datasets as inputs for model construction in a 2 × 3 factorial design. For the occurrence datasets, our filtered datasets were downsampled to one occurrence per 1 km raster grid cell to match the resolution of the environmental data. We generated an additional dataset downsampled to one occurrence per 20 km raster grid cell, which effectively reduced prevalence in very densely sampled clusters without removing overall patterns of landscape occurrence. We then generated three different forms of background point selection: randomly selected background points and two different Gaussian kernel density distributions (GKD). For the random background points, we used a polygon object with 200 km buffers centred on occurrence points and dissolving overlaps. We then selected the default 10,000 background points from the polygon. GKD background point selection aligns the background point density more closely to the density of occurrence points, so we generated two GKD background point datasets for both occurrence datasets. The probability of choosing a background point was based on a surface generated from summing Gaussian kernels (with dispersion parameter λ) at each occurrence point. For the GKD background points, we varied the dispersion parameter λ to create both highly clustered background (λ = 1) and moderately clustered background (λ = 3) points. All manipulations were conducted using ArcGIS Pro 2.3 (ESRI, 2019).

2.3.2 | Model construction

We constructed SDMs using Maxent v. 3.4.1 (Phillips, Anderson, Dudík, Schapire, & Blair, 2017) implemented through the ENMEval package (Muscarella et al., 2014). We developed six Maxent models per species using each combination of downsampled occurrence data (1 and 20 km) and GKD background point selection (random; λ = 1: highly clustered; λ = 3: moderately clustered). For all models, we set the regularization beta multiplier = 2 and allowed for all feature classes. In preliminary model building, we tested four (n = 4) beta multipliers from 1.0 to 4.0 with 1.0 increments. Differences among beta multipliers were minimal. Models were projected in North America using the "project" function in package "rmaxent" (Baumgartner & Wilson, 2019) and visualized using the "rasterVis v0.45" package (Lamigueriro & Hijmans, 2018).

2.3.3 | Model evaluation

We evaluated models by partitioning occurrences into four geographically structured blocks containing an approximately equal number of occurrence records. We assessed performance with fourfold cross-validation using three blocks as training data and retaining one block for testing. For each model, we calculated AUC (Area under the curve), the partial AUC (pAUC) ratio, the Continuous Boyce Index (CBI) and sensitivity (True Positives/ (True Positives + False Negatives)). AUC characterizes model prediction and discrimination ability relative to random chance (Phillips & Dudík, 2008), with values ranging from 0 to 1 where higher values indicating greater model performance and 0.5 is indicative of model predictive discrimination that is approximately random. The use of AUC has received criticism due to the equal weighting of omission and commission error and dependence on prevalence (Lobo et al., 2008). As commission error and true negatives are difficult to estimate in presence-only SDMs and may be of particular issue for invasive species, we explored model evaluation metrics that less reliant on commission error. We calculated the pAUC as the area under the portion of the curve (Peterson, Papeş, & Soberón, 2008) where omission errors are less than or equal to 10%. To assess whether pAUC was larger and significantly different from random expectation, we calculated the pAUC ratio, defined as the pAUC divided by the area under the 1:1 line that indicates random expectation for the same portion of the AUC curve. The pAUC ratio ranges from 0 to 2 and has a random expectation of 1 with values above 1 indicating better than random performance (Escobar, Qiao, Cabello, & Townsend Peterson, 2018). pAUC ratios were calculated and their significance assessed using the 'pROC' function in the package "NicheToolbox" (Osorio-Olivera, 2016) with 50% random withholding and 500 bootstrap iterations. Additionally, we calculated CBI, which measures the correlation between occurrence probability in the data and predicted probability (Boyce et al., 2002; Hirzel et al., 2006). CBI ranges from −1 to 1, where positive values indicate an accurate model that correlates with the data, zero indicates no difference from random, and negative values indicate a negative correlation. Sensitivity quantifies the correctly predicted positive fraction of occurrences and was evaluated at the threshold of 0.5 for all models.

2.3.4 | Model predicted area

To quantify how occurrence point downsampling and GKD background points affected the quantity of predicted suitable habitat, we

computed the area of suitable habitat across North America using thresholds varying from 0.01 to 0.99 at 0.01 intervals. We then plotted the relationship between proportion of suitable habitat against threshold value for each projection (*hereafter*: "threshold-area curves"). We tested for differences in the effects of model input modifications on projected suitable area at a 50% threshold (i.e. total area of predicted habitat suitability at or above 0.5; $N = 36$ models) with ANOVA where species grouping (e.g. common and historically established, incipient and diffuse, and incipient and clustered), occurrence dataset (filtered or downsampled), and background dataset (random, moderately clustered or highly clustered) and all second-order interaction terms were included as fixed, independent variables and area predicted to be suitable habitat at a 50% probability of occurrence threshold as the dependent variable.

2.3.5 | Assessing model performance, variability and complexity

For each model, we differentiated between a model's performance (mean cross-validated AUC, mean cross-validated pAUC ratio, mean cross-validated sensitivity and mean cross-validated CBI), variability among the four cross-validation datasets (AUC coefficient of variation and sensitivity coefficient of variation), and complexity (number of model parameters and AICc—although we note that AICc does not strictly assess only model complexity and instead balances model complexity with model fit). For performance, a higher mean cross-validated statistic indicated greater performance. We also calculated the coefficient of variation for AUC and sensitivity among the four cross-validation datasets, where models with greater variability exhibited a higher coefficient of variation among data partitions. To quantify complexity, we report the number of model parameters and AICc values. Maxent model parameters are transformations of environmental covariates that allow complex relationships to be modelled (Elith et al., 2010; Merow et al., 2013). More complex models had a greater number of parameters and generally higher AICc scores. We tested for significant differences in model performance (mean cross-validated AUC, mean cross-validated pAUC ratio, mean cross-validated sensitivity and mean cross-validated CBI), variability (AUC coefficient of variation and sensitivity coefficient of variation) and complexity (number of parameters) using ANOVA by including species, occurrence dataset (1 or 20 km) and background sampling ($\lambda = 0$: random; $\lambda = 1$: highly clustered; $\lambda = 3$: moderately clustered) and all second-order interactions as independent variables. We corrected for multiple tests using a sequential Bonferroni (Holm) correction.

2.3.6 | Structured decision process for model selection

We designed a model selection process based on measures of model performance, variability and complexity (Figure 1). We used this process to assess each of the six models per species and choose the model that optimally balances performance, variability and complexity. For each performance, variability and complexity score, we ranked each model and assigned it a scalar value from one to six, where higher values indicate better evaluation metrics. For example, one model per species with the highest mean cross-validated AUC received the highest score of six, and models with lower AUC performance received descending scores. Similarly, models with the lowest coefficient of variation received the highest score of six. For model complexity, we assigned two scores, where models with the fewest parameters received a score of six, and the lowest AICc score also received the highest score of six. Once all model metrics were scored, the model with the highest overall summed score was selected as the optimum model among the six for each species.

3 | RESULTS

3.1 | Theme 1: How did methods that reduce the effects of spatial bias affect measures of model complexity and amount of predicted suitable habitat?

3.1.1 | Model complexity

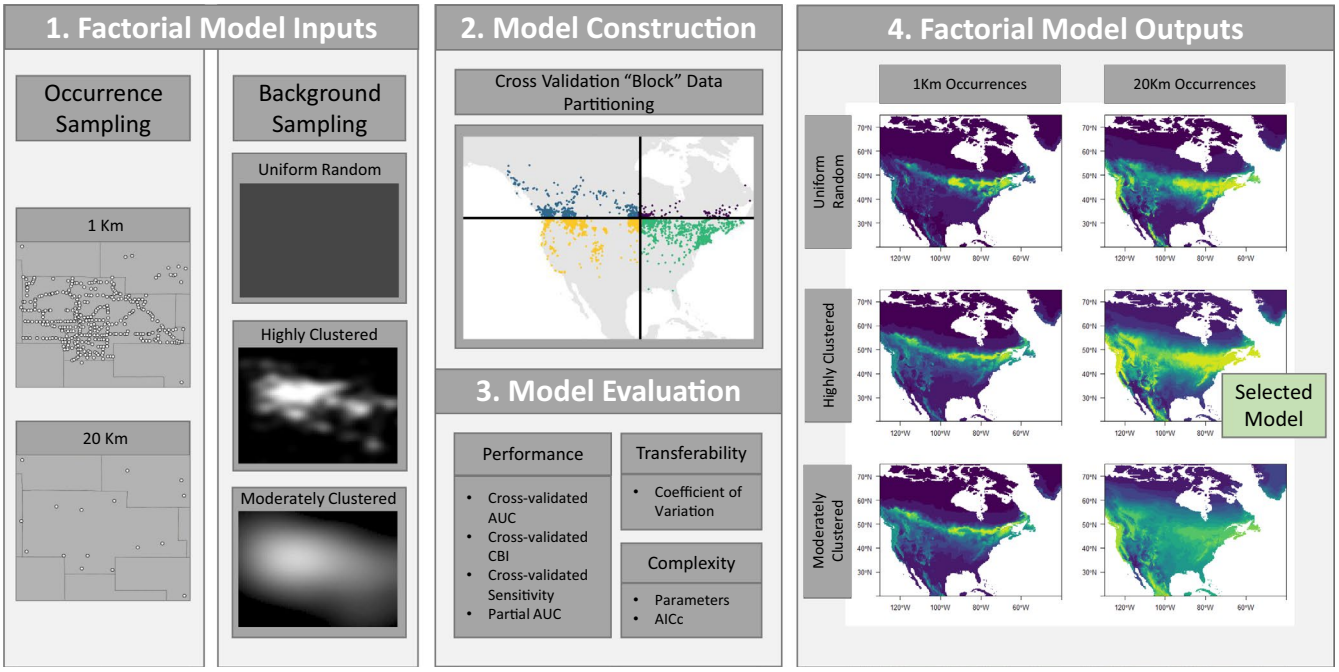
Models constructed with modified background sampling and/or occurrence data downsampling were less complex. This was most apparent for number of model features; across all species, models with GKD (clustered) background sampling had fewer features relative to random background sampling ($F_{df=2,25} = 16.81$, $p < .001$; Figure 2; Table S9), suggesting the latter are highly parameterized and more complex. Similarly, models with 20 km occurrence downsampling had fewer model features ($F_{df=2,12} = 71.07$, $p < .001$; Figure 2; Table S9). Additionally, these models had lower AICc scores (Table S15), suggesting models with 20 km occurrence downsampling are more likely to balance model fit with model complexity.

3.1.2 | Response curves

Upon visual inspection, model response curves for all species were smoother in models constructed with occurrence downsampling and GKD background sampling (Figures S8–S14).

3.1.3 | Threshold-area curves

Regardless of the occurrence dataset used (1 km—filtered; 20 km—downsampled), models using random background sampling predicted substantially less suitable habitat than models that used background points sampled using a GKD of 1 or 3 (highly or moderately clustered sampling) at nearly every possible threshold value (Figure 3). Among the differing invasion histories, there were distinct patterns in their



5. Model Selection by Ranking Model Evaluation Metrics

Inputs			Performance				Transferability		Complexity		Sum
Species	Occurrence Sampling	Background Sampling	AUC	Partial ROC Ratio	Continuous Boyce Index	Sensitivity	AUC Coefficient of Variation	Sensitivity Coefficient of Variation	AICc	Paramters	Rank
Tansy	Twenty	Moderately Clustered	0.560	1.501	0.370	0.621	0.157	0.630	51891.8	33	-
		Highly Clustered	0.546	1.735	0.870	0.830	0.093	0.141	49175.0	41	-
		Random	0.718	1.781	0.728	0.786	0.109	0.155	47551.1	86	-
	One	Moderately Clustered	0.721	1.272	0.716	0.683	0.211	0.434	231358.4	94	-
		Highly Clustered	0.746	1.252	0.750	0.713	0.194	0.394	229724.9	77	-
		Random	0.832	1.228	0.643	0.805	0.154	0.303	216222.9	97	-
Tansy	Twenty	Moderately Clustered	2	4	1	1	3	1	4	6	22
		Highly Clustered	1	5	6	6	6	6	5	5	40
		Random	3	6	4	4	5	5	6	3	36
	One	Moderately Clustered	4	3	3	2	1	2	1	2	18
		Highly Clustered	5	2	5	3	2	3	2	4	26
		Random	6	1	2	5	4	4	3	1	26

FIGURE 1 Model construction, evaluation and structured decision-making process flowchart. A total of 36 models for six species were constructed and evaluated, resulting in one selected model per species

threshold-area curves. For the common and historically established species (Common Tansy, Wild Parsnip, and Leafy Spurge), downsampling (to 20 km) led to substantially greater predicted habitat suitability across all thresholds, whereas GKD background sampling had minimal effect. For incipient and diffuse species (Brown Knapweed and Common Teasel), models with occurrence downsampling or GKD background sampling led to similar increases in predicted habitat suitability across all thresholds. For incipient and clustered species (Black Swallowwort and Dalmatian Toadflax), the effect was opposite to the common and historically established species: GKD background sampling led to greater predicted habitat suitability but occurrence downsampling had minimal effects (Figure 3).

When suitable habitat was assessed using a threshold probability of occurrence of 50% (i.e. all area with predicted probability of occurrence greater than or equal to 50%), models using GKD background sampling (highly or moderately clustered) had greater predicted suitable habitat area relative to models built using random background sampling ($F_{df=2,28} = 6.58$, $p = .005$; Table S10). Further, using that same threshold, models using occurrence downsampling (to 20 km) had significantly greater predicted suitable habitat area relative to models using the filtered (downsampling only to 1 km) occurrence datasets ($F_{df=2,28} = 10.49$, $p = .003$; Table S10).

3.2 | Theme 2: How were model performance and model variability affected by differences in model inputs?

3.2.1 | Performance

Downsampling occurrence points to 20 km increased model performance when measured by sensitivity and pAUC ratio (sensitivity: $F_{1,12} = 5.38$, $p = .039$; pAUC ratio: $F_{1,12} = 91.7$, $p < .001$; Tables S11 and S12). Conversely, AUC scores were greater for 1 km filtered datasets (AUC: $F_{1,12} = 79.9$, $p < .001$; Tables S11 and S12). The effects of background sampling design were more consistent

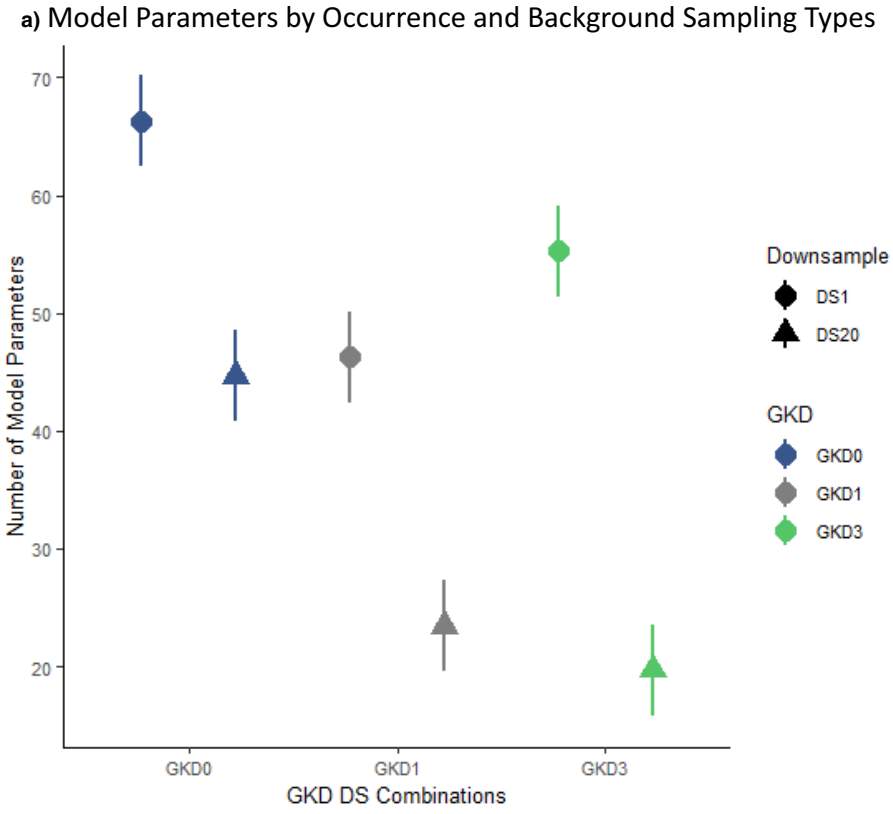


FIGURE 2 Interaction plot of occurrence downsampling by background point clustering effects for number of model parameters. Models were constructed with either 1 km filtered datasets (circles) or 20 km downsampled datasets (triangles) and either random (GKD 0, blue), highly clustered (GKD 1, grey) or moderately clustered (GKD 3, green) background sampling

across performance metrics. On average, AUC, CBI and sensitivity were higher for models that used random background sampling relative to models using either GKD background sampling (AUC: $F_{df=2,12} = 218.64$, $p < .001$; CBI: $F_{df=2,12} = 42.86$, $p < .001$; sensitivity: $F_{df=2,12} = 15.38$, $p < .001$; Tables S11 and S13). Among the models that used GKD background sampling, species differed in which model had better evaluation metrics (Species*Background Sampling: AUC: $F_{df=12,12} = 6.07$, $p = .002$; CBI: $F_{df=12,12} = 3.65$, $p = .017$; Sens: $F_{df=12,12} = 4.56$, $p = .007$; Table S11).

3.2.2 | Variability

The coefficients of variation (among the four cross-validation datasets) for both AUC and sensitivity were significantly greater when models were built using either GKD background sampling relative to random sampling (AUC: $F_{2,12} = 9.49$, $p = .003$; sensitivity: $F_{2,12} = 20.07$, $p < .001$; Table S14).

3.3 | Theme 3: How did model selection affect projections of habitat suitability, especially in currently unoccupied areas?

3.3.1 | Model projections

The results for theme 1 and 2 were agnostic to model selection and were meant to explore common patterns in model outputs generated by changes in data modifications. Here, we focus on the models that were retained following our structured decision process (Figures 1 and 4). Six of seven selected models used the 20 km occurrence downsampling and three of seven models used GKD background sampling (Figure 4; Tables S15 and S16). Overall, projections of selected models using current climate aligned well with current species distributions both across the continent and in the region of interest (Figures S15–S21).

For the common and historically established species (Common Tansy, Wild Parsnip, and Leafy Spurge), selected models used 20 km occurrence downsampling and either highly clustered (GKD $\lambda = 1$) or random background sampling (Table S16). Occurrence downsampling reduced bias associated with uneven sampling effort at local scales. The projected suitable but unoccupied area primarily represents range filling, rather than range expansion.

For incipient but diffusely distributed species (Brown Knapweed and Common Teasel), the selected models used 20 km occurrence downsampling and highly clustered (GKD $\lambda = 1$) background sampling (Table S16). These occurrence and background point modifications reduced bias associated with uneven sampling effort at both local and regional levels. The projected suitable but unoccupied area primarily represented range expansion, not range filling.

For incipient but highly clustered species (Black Swallowwort and Dalmatian Toadflax), there were no clear patterns (Table S16). For Dalmatian Toadflax, the selected model was built using the 1 km filtered dataset and random background sampling, suggesting this dataset may not have readily identifiable spatial bias patterns in the occurrence records. Conversely, for Black Swallowwort, the

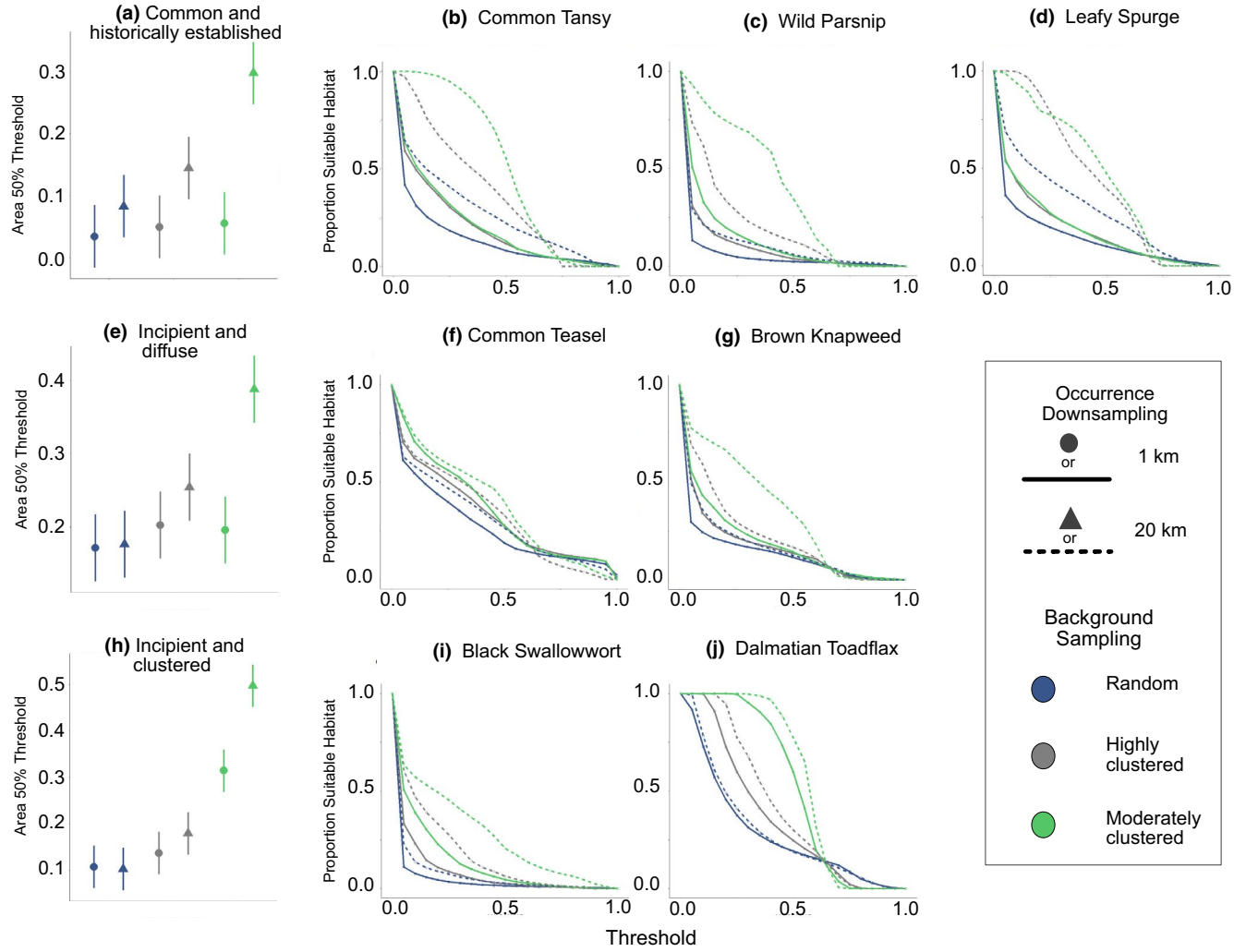


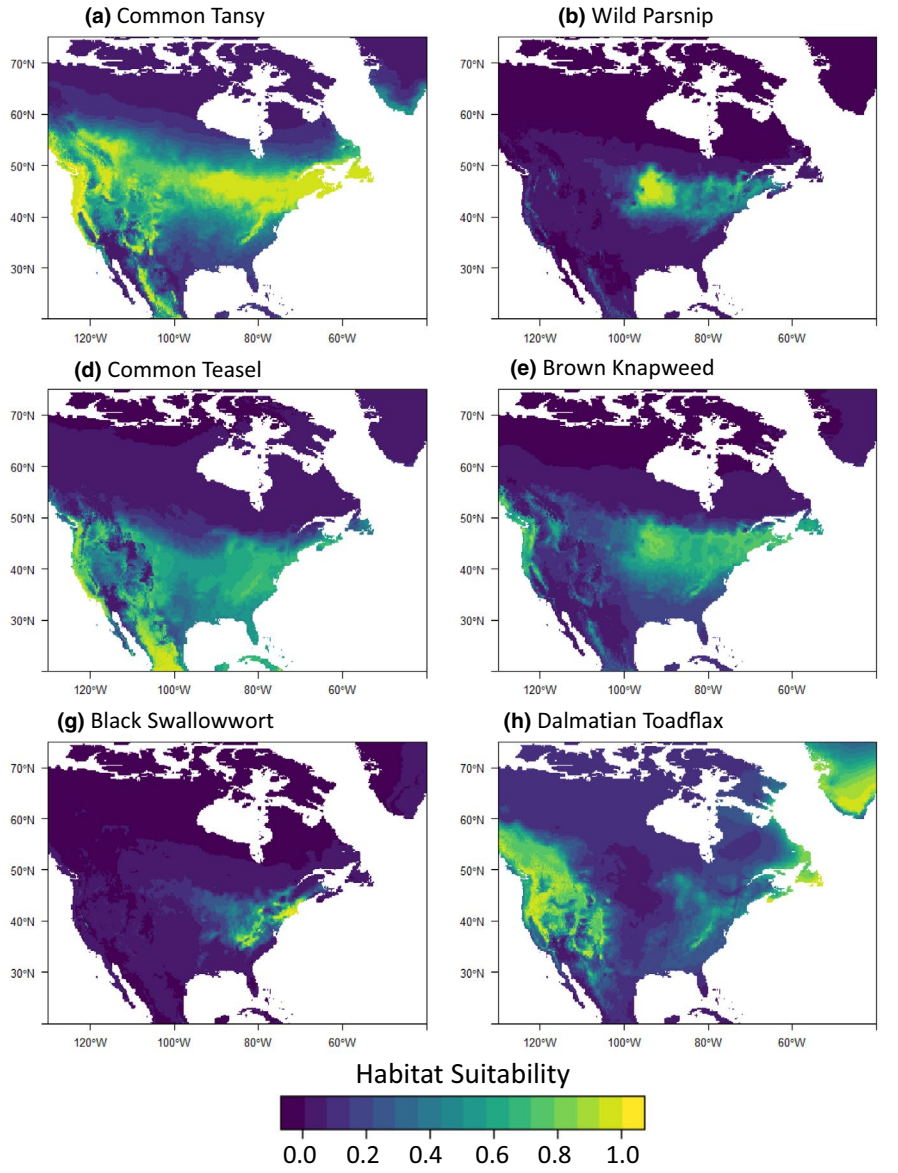
FIGURE 3 Species threshold-area curves and proportion of predicted suitable habitat at 50% threshold. For each species and species grouping: (a) common and historically established species including (b) Common Tansy, (c) Wild Parsnip, (d) Leafy Spurge, (e) incipient and diffuse species including (f) Common Teasel, (g) Brown Knapweed, and incipient and clustered species including (i) Black Swallowwort, (j) Dalmatian Toadflax. The curves depict the relationships between proportion of predicted suitable habitat (*y*-axis) and the threshold used to determine suitability (*x*-axis). Each threshold-area curve panel has six curves that depend on the combination of occurrence dataset used (1 km filtered dataset = solid lines; 20 km downsampled dataset = dashed lines) and background sampling (random = blue; highly clustered/GKD 1 = grey; moderately clustered/GKD 3 = green). In the left column, the interaction between GKD and downsampling is shown at the 50% threshold; error bars represent ± 1 *SE*

selected model was built using 20 km occurrence downsampling and highly clustered background sampling. The suitable but unoccupied area for both species primarily represented range expansion into the Upper and Central Midwest.

4 | DISCUSSION

Building climate-based SDMs that can successfully predict suitable habitat for invasive species remains a fundamental challenge despite its widespread adoption for management purposes. We implemented a factorial design of data modifications to comprehensively evaluate and reduce spatial bias for multiple invasive species. Our results suggest that models constructed with downsampled occurrence data and GKD background sampling had fewer parameters (i.e. were less complex) and were more regularized. We found that model performance and model complexity metrics were often not in agreement and that models that were overly complex often failed to predict occurrences in a region at the advancing front of geographic ranges (low transferability). Using our full matrix of models, we designed a decision-making process that considered model complexity, performance and geographic variability for multiple model evaluation metrics. Overall, our results suggest that careful consideration of spatial bias in model inputs and evaluation criteria are essential for identifying SDMs that have the capacity to predict suitable habitat at expanding range fronts of invasive species.

Our approach is uncommon in that we used a factorial design to independently assess the effects of different data modifications on



(f)

Selected Model Inputs		
Species	Downsample	GKD
Common Tansy	Twenty Kilometers	Highly Clustered
Wild Parsnip	Twenty Kilometers	Random
Leafy Spurge	Twenty Kilometers	Random
Common Teasel	Twenty Kilometers	Highly Clustered
Brown Knapweed	Twenty Kilometers	Highly Clustered
Dalmatian Toadflax	One Kilometer	Random
Black Swallowwort	Twenty Kilometers	Random

(i)

Selected Model Outputs	
Species	Predicted Area at 50% Threshold
Common Tansy	7,864,528 Square Kilometers
Wild Parsnip	1,403,520 Square Kilometers
Leafy Spurge	6,051,648 Square Kilometers
Common Teasel	7,816,712 Square Kilometers
Brown Knapweed	4,010,008 Square Kilometers
Dalmatian Toadflax	4,541,832 Square Kilometers
Black Swallowwort	608,880 Square Kilometers

FIGURE 4 Projections of habitat suitability for set of selected models for each of the seven species. (a) Common Tansy, (b) Wild Parsnip, (c) Leafy Spurge, (d) Common Teasel, (e) Brown Knapweed, (g) Black Swallowwort and (h) Dalmatian Toadflax. Darker purple/blue colour values indicate low predicted habitat suitability and brighter yellow/green colour values indicate high predicted habitat suitability. Occurrence and background sampling used to construct models are summarized in (f), and amount of predicted suitable habitat in square kilometres at the 50% threshold are summarized in (i)

the capacity of models to predict invasive species distributions and transferability to expanding range fronts. Both occurrence downsampling and GKD background selection had significant and independent effects on model complexity (e.g. number of model parameters) and the amount of projected suitable habitat. The independent effects of both modifications likely occurred because they correct for spatial bias at differing scales: downsampling of occurrence data reduces bias introduced by local oversampling whereas modification of background sampling reduces the effects of spatial biases at continental scales (e.g. incomplete sampling in some regions and/or dispersal lags). These modifications also had different effects depending on species invasion history and the spatial biases that may potentially be associated with invasion histories. For common and historically established species, downsampling had much larger effects; whereas, for the incipient and clustered species, GKD background selection had much larger effects. For the incipient and diffuse species, both had roughly equivalent effects.

There has been growing recognition that model input for SDMs may require modification to mitigate the effects of spatial bias. For both native and invasive species, downsampling of occurrence data (e.g. Briscoe Runquist, Lake, Tiffin, & Moeller, 2019; Kramer-Schadt et al., 2013) and background bias correction (Clements et al., 2012; Kramer-Schadt et al., 2013; Phillips et al., 2009; Syfert, Smith, & Coomes, 2013) have been implemented separately. However, systematic investigation of the interaction between the two techniques and exploration of multiple levels of each (e.g. downsampling at

different spatial scales) is uncommon. Sampling bias in occurrence records may be particularly pronounced in invasive species because of the geographically uneven sampling and the relatively shortened timeframes over which they have been tracked. Therefore, when building invasive species SDMs, incorporation of downsampling and background point manipulation should be explored for all species. While occurrence and background data manipulation may improve the quality of models and their projections of habitat suitability, reducing the number of input records or manipulating background sampling to achieve a balanced design may reduce model performance measures (Hernandez, Graham, Master, & Albert, 2006). Due to this known tradeoff in improving model quality and decreasing model performance, the most commonly used metric to evaluate model performance (test AUC) may be misleading (Lobo et al., 2008; Warren & Seifert, 2011). Overall, maximizing AUC will often not result in SDMs that are successful at predicting suitable but unoccupied territory. Rather, they reflect success in predicting occurrences in the oldest regions of the invaded range where records are numerous and dense across the landscape. These SDMs also often have sharp transitions between habitat predicted to be highly suitable and highly unsuitable, which indicates that the model is likely overly complex and not appropriately regularized (Figures S15–S21).

One of the overarching patterns for our set of selected models was that they consistently had low AUC scores despite high pAUC ratio, CBI and sensitivity scores. Additionally, AUC did not agree well with model fit statistics, such as AICc (which balanced model complexity with model information); this result is common across model-building efforts (Lobo et al., 2008). AUC favoured models built using standard datasets (i.e. filtered to 1 km with random background points); however, these models generally performed poorly when evaluated using alternative methods. Perhaps because of this, we found no obvious relationship between maximizing AUC and model selection. AUC equally weights errors of omission and commission (Lobo et al., 2008). When AUC fails, it often does so because validation statistics assume errors of omission and commission are well characterized (Oreskes, Shrader-Frechette, & Belitz, 1994). Because background points are not true absences, true species prevalence and true omission and commission errors are not well captured, and AUC may be misleading (Leroy et al., 2018). Indeed, a recent simulation study of AUC found that discrimination accuracy metrics were only informative about model quality when the models were simple, and the input data closely mimicked the true ecological niche (e.g. nearly complete range filling where background points were more similar to absences; Warren, Matzke, & Iglesias, 2019). In our models, we did see a pattern that AUC more likely to partially agree with other metrics for our common, historically established species, which are the most likely to have filled much of their available ecological niche in their invaded range.

Due to the known challenges of AUC, we advocate that modellers use multiple evaluation metrics and integrate that information during model selection. In particular, modellers should incorporate metrics that do not equally weight commission and omission errors such as pAUC ratio and sensitivity (Anderson, Lew, & Peterson, 2003; Lobo et al., 2008; Peterson, 2006). The pAUC ratio limits evaluation of models to regions of the AUC curve with low omission (Peterson, 2006). Likewise, sensitivity evaluates only the proportion of true positives (the omission rate) and does not rely on commission errors. CBI is valuable because it is spatially explicit and useful for incorporating knowledge about variation in prevalence across a landscape, but still may be prone to errors of commission. In the end, because all metrics have potentially complementary pitfalls, we used a structured process to incorporate information from them all during model selection.

4.1 | Invasive species distributions

Our selected models predicted variable but increased habitat suitability at and beyond species' range margins (in the Upper Midwest) under current climate. For common and historically established species (Common Tansy, Wild Parsnip and Leafy Spurge), predicted suitable habitat mostly reflected range infilling, with some geographic expansion to the north and west. For the other four species, all showed potential for range expansion to the north and west. Similarly, other invasive species have also been shown to have greater potential for range infilling and northward expansion (e.g. Jarnevich & Stohlgren, 2009; Mainali et al., 2015). Similarly, non-native plant species of North America occupied a smaller fraction of their potential range than native species, suggesting greater potential for range infilling (Bradley, Early, & Sorte, 2014). Last, non-native species have also been found to have potential for east-west range expansion, which was also common among our species (Bradley et al., 2014).

5 | CONCLUSIONS

Invasive species test the limits of SDM capabilities. Navigating these challenges requires developing and selecting models using data modifications to understand model behaviour and guard against common model pitfalls. It is important that SDMs are evaluated holistically with a diverse set of model diagnostics to avoid selecting and relying on models that do not generalize beyond the oldest regions of invaded ranges and cannot predict suitable habitat in incipient areas of invasion. Our study suggests that modifications of occurrence and background data should be implemented when modelling invasive species to minimize model overfit and spatial biases. Finally, interpretation of these models as the description of a species' ecological niche should be done with caution and include additional sources of evidence beyond SDMs.

ACKNOWLEDGEMENTS

We thank R. Venette for thoughtful discussion and comments and Z. Radford and K. Wilson for help gathering occurrence record data. Funding for this project was provided by the Minnesota Invasive Terrestrial Plants and Pests Center through the Environment and

Natural Resources Trust Fund as recommended by the Legislative-Citizen Commission on Minnesota Resources (LCCMR).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ddi.13161.

DATA AVAILABILITY STATEMENT

All data will be made publicly available through Dryad (https://doi.org/10.5061/dryad.73n5tb2v8) the Data Repository for the University of Minnesota (https://doi.org/10.13020/z71w-jx69).

ORCID

Thomas A. Lake https://orcid.org/0000-0002-8836-5164
Ryan D. Briscoe Runquist https://orcid.org/0000-0001-7160-9110
David A. Moeller https://orcid.org/0000-0002-6202-9912

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Allouche, O., Tsoar, A., & Kadmon, K. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Anderson, R. P., Lew, D., & Peterson, A. T. (2003). Evaluating predictive models of species' distributions: Criteria for selecting optimal models. *Ecological Modelling*, *162*, 211–232. https://doi.org/10.1016/S0304-3800(02)00349-6

Barbet-Massin, M., Jiguet, F., Hélène Albert, C., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, *3*, 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Baumgartner, J., & Wilson, P. (2019). *Rmaxent: Tools for working with Maxent in R. R package version 0.803.9000*. Retrieved from https://github.com/johnbaums/rmaxent

Bivand, R., Keitt, T., & Rowlingson, B. (2019). *Rgdal: bindings for the "geospatial"data abstraction library. R package version 1.2-16*. in press.

Bivand, R., & Rundel, C. (2015). *rgeos: Interface to Geometry Engine—Open Source (GEOS). R package version 3–19*. in press.

Boyce, M., Vernier, P., Nielsen, S., & Schmiegelow, F. (2002). Evaluating resource selection functions. *Ecological Modelling*, *157*, 281–300. https://doi.org/10.1016/S0304-3800(02)00200-4

Bradley, B. A., Blumenthal, D. M., Wilcove, D. S., & Ziska, L. H. (2010). Predicting plant invasions in an era of global change. *Trends in Ecology and Evolution*, *25*, 310–318. https://doi.org/10.1016/j.tree.2009.12.003

Bradley, B. A., Early, R., & Sorte, C. J. B. (2014). Space to invade? Comparative range infilling and potential range of invasive and native plants. *Global Ecology and Biogeography*, *24*, 348–359. https://doi.org/10.1111/geb.12275

Briscoe Runquist, R. D., Lake, T., Tiffin, P., & Moeller, D. A. (2019). Species distribution models throughout the invasion history of *Palmer amaranth* predict regions at risk of future invasion and reveal challenges with modeling rapidly shifting geographic ranges. *Scientific Reports*, *9*, 2426.

Broennimann, O., Treier, 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

Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R. O., Wang, Z., & Zimmermann, N. E. (2019). Model complexity affects species distribution projections under climate change. *Journal of Biogeography*, *47*, 130–142. https://doi.org/10.1111/jbi.13734

Clements, G. R., Rayan, D. M., Aziz, S. A., Kawanishi, K., Traeholt, C., Magintan, D., … Tingley, R. (2012). Predicting the distribution of the Asian tapir in Peninsular Malaysia using maximum entropy modeling. *Integrative Zoology*, *7*, 400–406.

Douglass, C. H., Weston, L. A., & DiTommaso, A. (2009). Black and Pale Swallow-Wort (*Vincetoxicum nigrum* and *V. rossicum*): The biology and ecology of two perennial, exotic and invasive vines. In Inderjit (ed.), *Management of invasive weeds. invading nature – Springer series in invasion ecology* (Vol. *5*, pp. 261–277). Dordrecht, the Netherlands: Springer.

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 prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith, J., Phillips, S., Hastie, T., Dudík, M., En Chee, Y., & Yates, C. (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

Escobar, L. E., Qiao, H., Cabello, J., & Townsend Peterson, A. (2018). Ecological niche modeling re-examined: A case study with the Darwin's Fox. *Ecology and Evolution*, *8*(10), 4757–4770. https://doi.org/10.1002/ece3.4014

ESRI. (2019). *ArcGIS Pro: Release 2.3*. Redlands, CA: Environmental Systems Research Institute.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302–4315. https://doi.org/10.1002/joc.5086

Finnoff, D., Shogren, J. F., Leung, B., & Lodge, D. (2007). Take a risk: Preferring prevention over control of biological invaders. *Ecological Economics*, *62*, 216–222. https://doi.org/10.1016/j.ecolecon.2006.03.025

Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, *27*, 245–256. https://doi.org/10.1111/geb.12684

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

Hao, T., Elith, J., Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, *25*, 839–852. https://doi.org/10.1111/ddi.12892

Hastie, T., & Fithian, W. (2013). Inference from presence-only data; the ongoing controversy. *Ecography*, *36*, 864–867. https://doi.org/10.1111/j.1600-0587.2013.00321.x

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, *29*, 733–785. https://doi.org/10.1111/j.0906-7590.2006.04700.x

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

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*, 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Jarnevich, C. S., & Stohlgren, T. J. (2009). Near term climate projections for invasive species distributions. *Biological Invasions*, *11*, 1373–1379. https://doi.org/10.1007/s10530-008-9345-8

Jiménez-Valverde, A., Peterson, A. T., Soberón, J., Overton, J. M., Aragón, P., & Lobo, J. M. (2011). Use of niche models in invasive species risk assessments. *Biological Invasions*, *13*, 2785–2797. https://doi.org/10.1007/s10530-011-9963-4

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., … 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

Lamigueriro, O. P., & Hijmans, R. (2018). *rasterVis. R package version 0.45*. in press.

Le Maitre, D. C., Thuiller, W., & Schonegevel, L. (2008). Developing an approach to defining the potential distributions of invasive plant species: A case study of Hakea species in South Africa. *Global Ecology and Biogeography*, *17*(5), 569–584.

Leroy, B., Delsol, R., Hugueny, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. (2018). Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, *45*(9), 1994–2002.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*, 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

Lodge, D. M., Simonin, P. W., Burgiel, S. W., Keller, R. P., Bossenbroek, J. M., Jerde, C. L., … Zhang, H. (2016). Risk analysis and bioeconomics of invasive species to inform policy and management. *Annual Review of Environment and Resources*, *41*, 453–488. https://doi.org/10.1146/annurev-environ-110615-085532

Mainali, K. P., Warren, D. L., Dhileepan, K., McConnachie, A., Strathie, L., Hassan, G., … Parmesan, C. (2015). Projecting future expansion of invasive species: Comparing and improving methodologies for species distribution modeling. *Global Change Biology*, *21*, 4464–4480. https://doi.org/10.1111/gcb.13038

Marbuah, G., Gren, I., & McKie, B. (2014). Economics of Harmful Invasive Species: A Review. *Diversity*, *6*, 500–523. https://doi.org/10.3390/d6030500

Merckx, B., Steyaert, M., Vanreusel, A., Vincx, M., & Vanaverbeke, J. (2011). Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. *Ecological Modelling*, *222*, 588–597. https://doi.org/10.1016/j.ecolmodel.2010.11.016

Merow, C., Smith, M. J., & Silander, J. A. (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

Morey, A. C., & Venette, R. C. (2020). Minimizing risk and maximizing spatial transferability: Challenges in constructing a useful model of potential suitability for an invasive insect. *Annals of the Entomological Society of America*, *113*(2), 100–113. https://doi.org/10.1093/aesa/saz049

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.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*, 641–646. https://doi.org/10.1126/science.263.5147.641

Osorio-Olivera, L. (2016). *NicheToolbox: A web tool for exploratory data analysis and niche modeling*. Retrieved from http://shiny.conabio.gob.mx:3838/nichetoolb2/

Pebesma, E., & Bivand, R. S. (2005). S classes and methods for spatial data: The sp package. *R News*, *5*, 9–13.

Peterson, A. T. (2006). Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, *3*, 59–72. https://doi.org/10.17161/bi.v3i0.29

Peterson, A. T., Papeş, M., & Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, *213*(1), 63–72.

Petitpierre, B., Broenniman, O., Kueffer, C., Daehler, C., & Guisan, A. (2017). Selecting predictors to maximize the transferability of species distribution models: Lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, *26*, 275–287. https://doi.org/10.1111/geb.12530

Petitpierre, B., Kueffer, C., Broenniman, O., Randin, C., Daehler, C., & Guisan, A. (2012). Climatic niche shifts are rare among terrestrial plant invaders. *Science*, *335*, 1344–1348. https://doi.org/10.1126/science.1215933

Phillips, S. J., Anderson, R. P., Dudík, 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., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, *31*, 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*, 181–197. https://doi.org/10.1890/07-2153.1

Phillips, S. J., Dudík, M., & Schapire, R. E. (2018). *Maxent software for modeling species niches and distributions (Version 3.4.1)* [Internet]. Retrieved from http://biodiversityinformatics.amnh.org/open_source/maxent/

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, *41*, 629–643.

Syfert, M. M., Smith, M. J., & Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One*, *8*, e0055158.

von der Lippe, M., & Kowarik, I. (2007). Long-distance dispersal of plants by vehicles as a driver of plant invasions. *Conservation Biology*, *21*, 986–996. https://doi.org/10.1111/j.1523-1739.2007.00722.x

Warren, D. L., Matzke, N. J., & Iglesias, T. L. (2019). Evaluating species distribution models with discrimination accuracy is uninformative for many applications. *bioRxiv*, 684399.

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

Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, *3*, 260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., … Sequeira, A. M. M. (2018). Outstanding Challenges in the transferability of ecological models. *Trends in Ecology and Evolution*, *33*, 790–802. https://doi.org/10.1016/j.tree.2018.08.001

BIOSKETCH

Thomas Lake is a Ph.D. student with Dr. D.A. Moeller in the Plant and Microbial Biology Program at the University of Minnesota studying the biogeography and landscape genomics of invasive species. D.A. Moeller is an Associate Professor at UMN whose research focuses on the ecological and evolutionary drivers of geographic range limits, plant mating system evolution and plant speciation. R.D. Briscoe Runquist is a Postdoctoral Research Associate with D.A.M., whose research focuses on the biogeography and landscape genomics of invasive plants.

Author contributions: All authors conceived of the work. T.L. performed analyses. T.L. and R.D.B.R. led writing of the manuscript. All authors contributed to substantial editing of the manuscript.

SUPPORTING INFORMATION

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

How to cite this article: Lake TA, Briscoe Runquist RD, Moeller DA. Predicting range expansion of invasive species: Pitfalls and best practices for obtaining biologically realistic projections. *Divers Distrib*. 2020;26:1767–1779. https://doi.org/10.1111/ddi.13161