Received: 11 May 2020 | Accepted: 10 February 2021

DOI: 10.1111/jbi.14105

Journal of Biogeography WILEY

RESEARCH ARTICLE

Improving predictions of range expansion for invasive species using joint species distribution models and surrogate co-occurring species

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

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

Correspondence
Ryan D. Briscoe Runquist, University of Minnesota, Plant and Microbial Biology Department, 1479 Gortner Ave, 140 Gortner Lab, St. Paul, MN 55108, USA.
Email: rbriscoe@umn.edu

Funding information
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

Handling Editor: Daniel Chapman

Abstract

Aims: Species distribution models (SDMs) are often used to forecast potential distributions of important invasive or rare species. However, situations where models could be the most valuable ecologically or economically, such as for predicting invasion risk, often pose the greatest challenges to SDM building. These challenges include non-equilibrium range expansion, low or uneven prevalence, and projecting distributions into environments that are non-analogous to the environments used for model building.

Location: Minnesota, USA.

Taxon: *Cardamine impatiens* (Narrowleaf Bittercress), *Celastrus orbiculatus* (Oriental Bittersweet), and *Humulus japonicus* (Japanese Hops).

Methods: We took a novel approach to build robust species distribution models of invasive species using occurrence-environment correlations between invasive species and co-occurring native community members. The correlations were obtained from a joint species distribution model (JSDM) of a densely sampled database of 10,336 MN plant communities from across the state of Minnesota, USA. Positively and negatively associated native species were incorporated into the model as surrogate presences and pseudoabsences (weighted by their environmental correlations) along with invasive species occurrences records (surrogate SDMs).

Results: Surrogate models performed better than traditional SDMs in predicting occurrences along the northern invasion margin (outside the training area). Both types of models had similarly high cross-validation metrics in the area of training. Surrogate models also predicted greater range expansion beyond the current geographic range.

Main conclusions: These results demonstrate that modelers can take advantage of detailed community data to develop SDMs that leverage surrogate native species as phytometers of environments beyond the current area of occupancy. The additional information in surrogate models generates highly effective predictions of invasive species along expanding range margins.

KEYWORDS

boosted regression trees, ecological niche models, gjam, incipient invasive species, range margin

1 | INTRODUCTION

Invasive species are reshaping terrestrial and aquatic habitats and causing substantial economic losses worldwide (Gren et al., 2009; Marbuah et al., 2014; Pimentel et al., 2005). To most effectively manage invasive species, scientists need to predict geographic areas at risk of invasion under current and future climate scenarios. However, developing accurate and highly predictive distribution models has proven challenging (Elith et al., 2010; Fourcade et al., 2018; Gallien et al., 2010; Lodge et al., 2016).

Traditional methods used to assess invasion risk rely on modeling the responses of individual species to environmental factors (Lodge et al. 2016). These species distribution models (SDMs) are typically correlative models that relate known presence data (and possibly absence data) with environmental data and extrapolate those relationships into regions where land managers are interested in determining risk of invasion (Broennimann et al., 2007; Elith et al., 2010; Elith & Leathwick, 2009). However, SDMs often fail to predict the range expansion of invasive species because they violate key assumptions of SDMs (e.g., that the modeled species is at range and environmental equilibrium; Elith et al., 2006; Elith et al., 2010; Journé et al., 2019). For example, SDMs of the rapidly-spreading noxious weed, Palmer Amaranth, built using a time series of historical invasion records could not reliably predict its current distribution (Briscoe Runquist et al., 2019).

Projecting outside of geographic ranges may be particularly problematic because regions of incipient invasion often contain suitable but unoccupied environments that are not included as such in model building (Broennimann et al., 2007, 2012; Elith et al., 2010; Guisan et al., 2014). Thus, the models lack information on what additional, unoccupied environments might be suitable (Bradley et al., 2010; Elith et al. 2010). Incorporating experimental information on how an invasive species would respond to novel environments would greatly improve model predictions (Gallien et al., 2012; Merow et al., 2017; Singer et al., 2016). However, such field experiments are rarely feasible with invasive species and are not easily extended to large numbers of species. Thus, new approaches are needed that leverage additional information on potential environmental suitability beyond the current range and incorporate them into SDMs (Gallien et al., 2010, 2012; Larson et al., 2014; Lodge et al. 2016). In this study, we take a novel approach that borrows occurrence-environment relationships from native species via Joint Species Distribution Models (JSDMs; Norberg et al., 2019).

Our approach involves borrowing information from the distributions of native species that commonly co-occur with a focal invasive species and incorporating those occurrence–environment relationships into model building. This approach uses co-occurring species as barometers of potential suitability beyond the invaded range. It may be especially advantageous in regions where invasive species prevalence is low, such as incipient areas of invasion or for species that are difficult to detect. Thus far, two different types of approaches have been taken to borrow predictive power from other species. First, information on the mean environmental response of a collection of species (some common and some rare) has been used to aid in forecasting the environmental responses of individual species (e.g., Hui et al., 2013; Larson et al., 2014). Second, the distributions of associated or interacting species have been incorporated by including them directly as predictive model variables or during the model selection process (e.g., competitors, mutualists, or related species: Araújo & Luoto, 2007; Baselga & Araújo, 2009; Pellissier et al., 2010; Mod et al., 2015; Molloy et al., 2017; Qiao et al., 2017; Mäkinen & Vahatalo, 2018). An alternative approach would incorporate weighted occurrences of surrogate species into SDMs, which has several potential advantages. Our approach explicitly evaluates which co-occurring (established) species share similar environmental responses. This may be particularly valuable when co-occurring species are from the same functional group as the focal species (e.g., herbaceous understory plants) because they likely experience the environment in similar ways and their environmental responses will potentially generate an integrated environmental response. In addition, native or common species are more likely to be at range equilibrium and therefore conform better to modeling assumptions (Elith et al., 2010; Norberg et al., 2019). However, which native species to include as potential surrogates and how to best incorporate their distributional information remains an open question that we address by leveraging joint species distribution models.

To build SDMs involving co-occurring species, we need to be able to identify native and established species that share environmental affinities and can be included as surrogate presence/absence data. We took a new approach that leveraged Joint Species Distribution Models (JSDMs) to identify surrogate species and calculate weightings that can be incorporated into SDMs. JSDMs are a relatively new class of methods (e.g., Clark et al., 2017; Harris, 2015; Pollock et al., 2014) that encompass a broad collection of approaches, but all incorporate community co-occurrence patterns and environmental data within a unified framework. Integration of both sets of data allows for the estimation of species' responses to the environment as well as the correlation among species' shared environmental responses. Although useful for predicting large geographic and environmental community patterns, they are not always successful at predicting individual species distributions, particularly for rare species or species with low prevalence and may not be able to directly make projections of invasive species distributions (Norberg et al. 2019; Zhang et al., 2018; Zurell et al., 2018). However, for invasive species, JSDMs can provide information about species' occurrence–environment correlations and thereby identify species that could be phytometers of invasion potential. The JSDM method developed by Clark et al. (2017) provides an additional advantage because it can incorporate multiple sources of data and explicitly calculates species' environmental response correlations on the scale of the input data within the JSDM framework. We used JSDMs to select and weight the occurrences of surrogate species in our SDMs.

In this study, we built SDMs for three plant species invasive to North America, *Celastrus orbiculatus* (Oriental bittersweet), *Humulus japonicus* (Japanese hops), and *Cardamine impatiens* (Narrowleaf bittercress), all of which have range boundaries that have recently expanded north and west into Minnesota, U.S.A. We leveraged a large dataset on 10,336 plant communities distributed across Minnesota to determine which plant species were most commonly associated with our focal invasive species and the extent of their shared environmental responses. We then built SDMs that included information about the locations and environmental similarities of surrogate species as well as the invasive occurrence records and compared them to traditional SDMs built using only invasive species occurrence records. In doing this, we were interested in answering the following questions:

a. Do JSDM-informed surrogate models improve predictions of climate suitability at range margins and provide a viable method for correlative invasive species modeling?
b. How does the prediction of invasion risk change when incorporating information on surrogate species?

TABLE 1 Traditional and surrogate Species Distribution Model inputs for (A) *Cardamine impatiens*, (B) *Celastrus orbiculatus*, and (C) *Humulus japonicus*

Species	MN occurrence records	Surrogate Positives (similarity weight/occurrence records)	Surrogate Negatives (similarity weight/occurrence records)	Trad. model pres/pseudoabs	Surrogate model pres/pseudoabs
Cardamine impatiens Narrowleaf Bittercress	43	*Cryptotaenia canadensis* (0.97/384) *Parthenocissus inserta* (0.94/787)	*Aralia nudicaulis* (−0.67/3174)	188/180	1214/3174
Celastrus orbiculatus Oriental Bittersweet	368	*Caulophyllum thalictroides* (0.97/536), *Smilacina racemosa* (0.95/880)	*Salix discolor* (−0.65/692) *Populus tremuloides* (−0.49/2203)	1363/1360	1784/2895
Humulus japonicus Japanese Hops	141	*Ribes missouriense* (0.99/313) *Euphorbia corollota* (0.98/89)	*Humulus lupulus* (−0.91/36) *Carex pensylvanica* (−0.09/1787)	378/560	543/1823

2 | MATERIALS AND METHODS

2.1 | Species descriptions

We modeled the distributions of three flowering plant species: *Celastrus orbiculatus* (Celastraceae), *Humulus japonicus* (Cannabaceae), and *Cardamine impatiens* (Brassicaceae). All three species are Eurasian in origin and were introduced to New England, U.S.A., in the late 1800's and have since spread westward around the Great Lakes into the Central and Upper Midwest (Figure S1& S2). The invaded ranges for all three have recently begun to encroach on the southern and eastern portions of Minnesota (especially in the last ca. 10–15 years; See Table S1 for more details on each species' distribution, natural history, and ecological impacts).

2.2 | Data sources

2.2.1 | Occurrence record data

We gathered invasive species' occurrence records (presence-only) from two publicly available databases: (1) Global Biodiversity Information Facility (GBIF; www.gbif.org; Table S2), and (2) Early Detection & Distribution Mapping System (EDDMapS; www.eddmaps.org). We processed occurrence records to remove duplicate, erroneous, and/or imprecise coordinates (<0.1 decimal degrees). In order to fairly and directly compare traditional and JSDM-informed surrogate SDMs (described in detail below), we built and tested models using two basic occurrence datasets. We built traditional SDMs using occurrence data drawn from the Midwest (Table 1; projections from models based on North American and Global models were functionally similar: Figure S3). We built JSDMs and surrogate SDMs using data from Minnesota (Table 1; GBIF/EDDMaps invasive occurrence records and DNR relevé dataset – see below). After model building, we validated traditional and surrogate SDMs using data from the Midwest. In this study, the

Midwest was defined by a bounding box (−98°W, −80°W, 40°N, 53°N), and included Minnesota, Iowa, Wisconsin, Michigan, northern Illinois, northern Indiana, and northern Ohio in the United States as well as portions of Ontario and Manitoba Provinces in Canada (Figure S2).

All spatial analyses on occurrences were performed in R (R Core Team v3.5.1) using the Geospatial Data Abstraction Library ('rgdal'; Bivand et al., 2017), Geometry Engine Open Source ('rgeos'; Bivand & Rundel, 2015), Spatial Points ('sp'; Pebesma & Bivand, 2005), and Raster packages ('raster'; Hijmans & vanEtten, 2012).

2.2.2 | Environmental data

We used three climate variables derived from the WorldClim v. 2 dataset (30 second resolution: https://www.worldclim.org/data/worldclim21.html; Fick & Hijmans, 2017), which function as basic constraints on plant growth (Chapman et al., 2019). Mean temperature of the warmest quarter (Bio10, °C) reflects the temperature of the growing season. Minimum temperature of the coldest month (Bio6, °C) indicates cold exposure and tolerance. Climate Moisture Index (CMI) reflects potential drought stress and was calculated as the natural logarithm of one plus the ratio of annual precipitation (Bio12) to potential evapotranspiration (Zomer et al. 2008; http://www.cgiar-csi.org) (Figuress. S4-S6).

2.3 | Occurrence record sampling for traditional model building

Building traditional SDMs with generalized boosted models (GBMs; see methods below for more details) required presence and absence data. For each species, we downsampled all occurrence records to the same resolution as the environmental data (~1 km; Figure 1A-C). Because there was no true absence dataset for any of these species, we generated pseudoabsence datasets that were approximately equal in number to the number of presence points (Figure 1A-C). We created pseudoabsence datasets by randomly sampling half of the

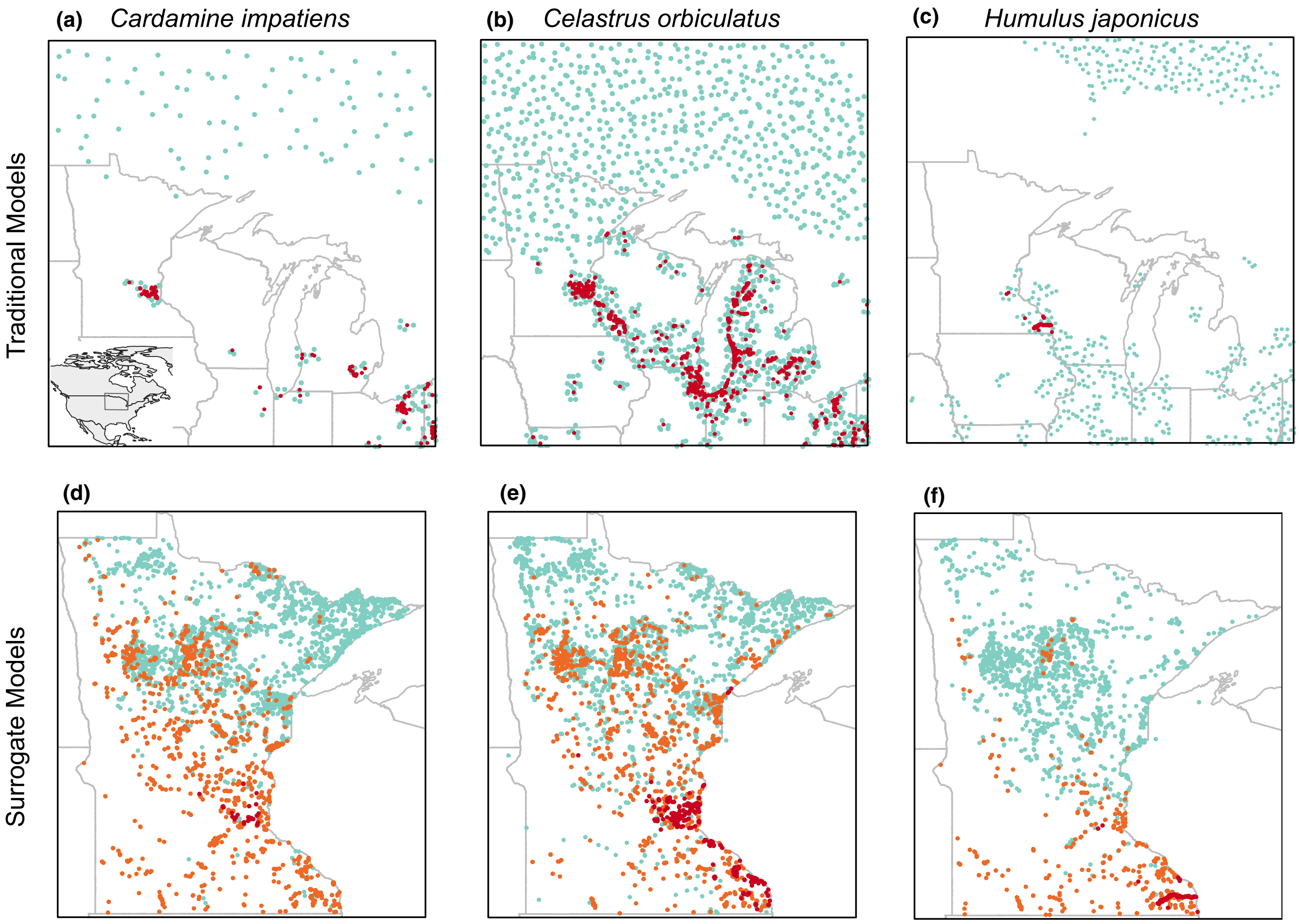

FIGURE 1 Species occurrence records and pseudoabsences for (top row) the Midwest and (bottom row) Minnesota. Columns represent data used for each of the focal invasive species (left: *C. impatiens*, center: *C. orbiculatus*, right: *H. japonicus*). Top row: Presences (red) and pseudoabsences (teal) in the Midwest were used to build traditional models and validate all models. Bottom row: Invasive species presences (red), surrogate presences from the relevé dataset (orange), and surrogate absences from the relevé dataset (teal) are shown. These were used to build surrogate SDM models. Inset map of North America in the lower corner of panel A shows the region of the Upper Midwest within the outlined bounding box [Colour figure can be viewed at wileyonlinelibrary.com]

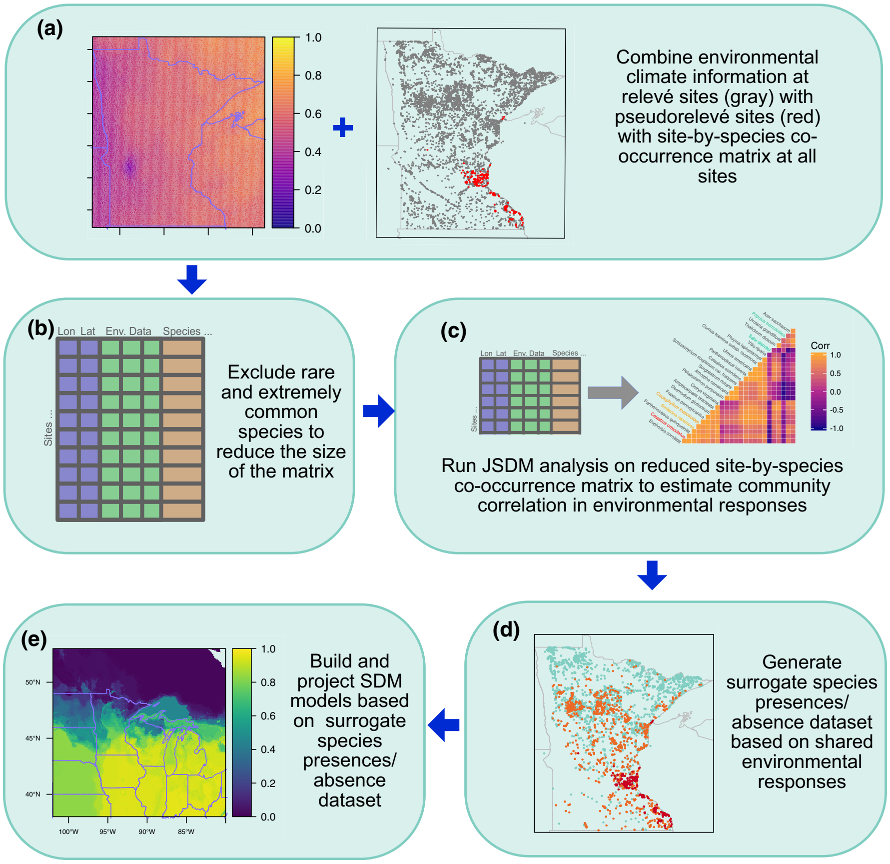

FIGURE 2 Flowchart of methods for selecting native species for incorporation into surrogate models. The example shown is for *C. orbiculatus*. (A) Combine climate information for relevé/pseudorelevé sites with a site-by-species co-occurrence matrix. (B) Reduce the size of the spatially and climatically explicit site-by-species co-occurrence by excluding rare or extremely common species. (C) Use the reduced matrix to run JSDM and obtain correlation of similarity in climatic environmental responses (e-matrix) for all of the community members included. (D) Choose surrogate species that are most and least similar in integrated environmental response. Invasive species presences shown in red, surrogate presences shown in orange, and surrogate absences shown in teal. (E) Incorporate surrogate species and their environmental response correlation coefficient as a weighting factor into SDM framework and project models to estimate climate suitability. Map pixels represent a relative metric of climate suitability, where 1.0 is the most suitable climate. Values above 0.9 are considered high suitability, 0.6–0.9 moderate suitability, and below 0.6 poor suitability [Colour figure can be viewed at wileyonlinelibrary.com]

pseudoabsences from potentially available (i.e., areas of potential dispersal) but unoccupied habitat (30 km buffers centered on each known occurrence point with dissolved overlaps). The second half were sampled from habitat deemed unsuitable (i.e., the value of climate variables fell outside of the 99.9% quantiles for global occurrences; Table S3 & Figure S7). Sampling pseudoabsences using this method increases the range of environments available for model fitting and improves the realism and accuracy of models (Chapman et al., 2019).

2.4 | Identification of surrogate presences and absences using joint species distribution models (JSDMs)

2.4.1 | Native community data

The Minnesota Department of Natural Resources (DNR) has collected data on 10,336 MN plant communities (2829 species) using a standardized relevé technique beginning in 1971 (DNR 2013). Relevés are a type of plant survey that characterizes the habitat and catalogues all of the community members at a site. Relevés often use a set of nested and increasingly widened quadrats in order to capture the majority of species in a community. We used relevé data collected from 1971 to 2016, which provided an extensive spatial dataset on plant species presence/absence and co-occurrence (Figures S8).

2.4.2 | Species co-occurrence matrices

We built a georeferenced site by species co-occurrence matrix (presence/absence) using the DNR relevé database. Because many of our invasive species occurrences fell outside of relevés, for each invasive species occurrence site, we generated pseudorelevés using information on community composition from the nearest relevé (Figure S9). Most invasive species occurrence records were within 0–5 km of a relevé site (Figure S10). Minnesota climate gradients are gradual with little to no change over 5–10 km distances and plant taxa. To limit pseudoreplication in our co-occurrence matrix, real relevés could not be resampled more than five times. We generated separate site by species co-occurrence matrices for each invasive species; each matrix contained information from all DNR-assessed relevés plus the pseudorelevés for the focal invasive species (Figures 1 & 2A,B).

For computational purposes, we reduced the size of our species co-occurrence matrices by eliminating very rare or common species, both of which fail to provide useful information for these analyses. As a first pass, we used the pseudorelevé data for each species separately to filter species lists. In the area of incipient invasion in Southeastern MN, (bounding box: −97.25°W, −89.45°W, 43.45°N, 49.4°N), we compiled and sorted the pseudorelevé species and eliminated all species that were rare (< 4% of sites) or extremely common (> 40%). After this elimination step, we retained approximately 50 community associates present in ca. 4–30% of the pseudorelevés. We then re-assessed potential community associates for each invasive species by conducting field surveys of plant communities in 10–15 sites/species. At each site, we laid out a 10x10 meter quadrat and cataloged all species that were >1% cover at 3 different habitat stratifications: ground-level, mid-story, and upper-story/canopy. We used this information to aid in further reducing each dataset by eliminating community associates that did not occur in relevant habitat (e.g., plants limited to bogs or marshes). For *C. impatiens*, we retained 23 species, all of which were herbaceous forbs and grasses. For *C. orbiculatus* and *H. japonicus*, we randomly sampled 25 species (Figure 2B; Table S4). Preliminary models run using all 50 species had similar results, but computational time was much greater and less predictable (results not shown).

We assessed whether JSDM results were affected by (1) our methods for creating pseudorelevés and (2) the choice of co-occurring community members. To examine these issues, we used a native species as a case study. *Smilacina racemosa* is a common herbaceous forb found throughout MN and was included as a community member in all of the invasive species JSDMs. We performed JSDMs using both relevé and pseudorelevé data, and we used alternative species co-occurrence matrices. We found that estimates of environmental response coefficients and their correlation with responses of co-occurring species were robust to the occurrence technique (relevé or pseudorelevé) and species co-occurrence matrix. These findings supported our use of the pseudorelevé method for estimating environmental responses and their covariance with co-occurring species (see supplement for greater detail on the case study).

2.4.3 | Identification of surrogate species using JSDMs

We used the R package, 'gjam', to perform JSDMs (Clark et al., 2017). We used the same three environmental variables as used for traditional SDMs. Gjam models required data on species taxonomy (plants), data structure (presence/absence), and collection effort. The 'effort' parameter allowed the model to incorporate information on relative differences among sampling efforts. We assumed that all sites from the DNR survey were completely and evenly sampled; therefore, we equalized the effort to 1 in all site by species combinations, excepting for occurrence records of our invasive species. We set sampling effort for these occurrences to 0 due to their history of recent invasion in MN. We also tried 0.01 and 0.1 with similar effects (results not shown). To streamline model computation, we used dimension reduction (N [total number of response groups modeled] = 10 and r [flexibility of parameters for groups] = 8) to reduce the number of covariance parameters estimated (Clark et al., 2017). We ran all gjam models for 100,000 generations, with 1500 generations of burn-in. For all models, we visually inspected MCMC chains for convergence (Figure S15).

For each of the potential community associates, the model estimates response parameters for environmental predictors, computes a correlation matrix of environmental responses (*henceforth*: e-matrix; Figure 2C & Figure S16–S17), correlations among species due to other environmental factors, and a number of community metrics. We retained the environmental predictors and e-matrix from each model and used it to identify surrogate species. We focused exclusively on the e-matrix because we wanted to retain the suite of native species that showed the most or least similar environmental responses after accounting for other potential covariance among species co-occurrence records. From each e-matrix, we recorded the 3 species with the highest correlation coefficients and 1–2 species with the lowest correlation coefficients (Table 1).

For each invasive species, we built a new JSDM-informed presence/pseudoabsence dataset (Figures 1D,E & 2D). In these datasets, occurrence records for the invasive species plus positively correlated surrogates were included as presences, and negatively correlated surrogates were included as absences. We retained the absolute value of the correlation coefficient as a weighting factor for SDMs; invasive species records were given a weight equal to 1. If more than one occurrence record (invasive or surrogate species) was found in the same location, we only retained the record for the species with the largest correlation coefficient [i.e., invasive species ($r = 1$) > surrogate presences ($1 > r > 0$) > surrogate absences ($r < 0$)]. Because Generalized Boosted Models (GBMs) function best when presence and absence records are close to balanced, we computed the average of all occurrence records weighted by their correlation coefficient. If the average over all presence/absence records weighted by their correlation was very different from zero, we added an additional surrogate species to the absence data (*C. orbiculatus* and *H. japonicus*) to balance the dataset.

2.5 | SDM generation with generalized boosted models (GBMs)

2.5.1 | Model Construction

We used our six presence/absence datasets to build GBM models (two datasets per species: (1) traditional presence/pseudoabsence

TABLE 2 Model parameters and evaluation metrics of traditional and surrogate SDMs

Species	Mean best number of trees of trad. model	Mean best number of trees of surrogate model	Traditional Model Mean total deviance/ Mean residual deviance/ cv deviance	Surrogate Model Mean total deviance/ Mean residual deviance/ cv deviance	AUC of trad. model[a] (train/ cv)	AUC of surrogate model[a] (train/ cv)	Midwest AUC/ Midwest pAUC of trad. model	Midwest AUC/ Midwest pAUC of surrogate. model	Midwest Sensitivity of trad. model	Midwest Sensitivity of Surrogate Model
Cardamine impatiens Narrowleaf Bittercress	5950	5400	1.386/ 0.683/ 0.857 ± 0.05	0.974/ 0.537/ 0.555 ± 0.01	0.93 0.86 ± 0.02	0.91 0.90 ± 0.00	0.93/ 0.81	0.88/ 0.79	0.94	0.99
Celastrus orbiculatus Oriental Bittersweet	6300	7500	1.386/ 0.66/ 0.761 ± 0.02	0.964/ 0.637/ 0.657 ± 0.02	0.94 0.90 ± 0.01	0.86 0.85 ± 0.02	0.93/ 0.81	0.76/ 0.78	0.92	1.0
Humulus japonicus Japanese Hops	5850	8250	1.348/ 0.836/ 0.922 ± 0.01	0.355/ 0.109/ 0.127 ± 0.01	0.89 0.85 ± 0.01	0.97 0.96 ± 0.00	0.80/ 0.72	0.66/ 0.68	0.41	1.0

[a] AUC values from 'gbm' model objects and based on model presence/pseudoabsence data: traditional models include invasive species presences and pseudoabsences drawn from available and unavailable habitats, surrogate models include invasive species presences and positive and negative surrogate occurrences.

dataset, (2) surrogate presence/absence dataset). We selected GBM model parameters based on practices outlined in Elith et al. (2008). We aimed for well-regularized models (i.e., not overly complex) with a learning rate that reached an asymptote in model deviance within 10,000 trees. We built models using the 'gbm.step' function, which relies on the 'dismo' and 'gbm' packages (Greenwell et al., 2019; Hijmans et al., 2017). Because we were interested in comparing traditional models with JSDM-informed, surrogate models (*hereafter*: surrogate models), we kept model building parameters as similar as possible across all models generated. For all models built, we used tree complexity 2 (i.e., number of nodes allowed in the tree), bag fraction of 0.8, and a learning rate of 0.001 (*C. impatiens*, *H. japonicus*, *C. orbiculatus* surrogate model) or 0.005 (*C. orbiculatus* traditional model, which did not reach an asymptote in model deviance with the smaller learning rate). This ensured that models were built with between 1000–10,000 trees (Table 2). Models were evaluated within the 'gbm' function framework using cross-validation (cv) with 10 cross-validation datasets.

2.5.2 | Model evaluation

To evaluate models from the 'gbm' model objects, we collected information on model deviance and AUC for training and cross-validation data. These metrics were calculated over the areas used to build the models (i.e., the Midwest for traditional models and Minnesota for the surrogate models). Surrogate model evaluations from the model objects included presence/absence information on the invasive species as well as the surrogate species. Model deviance is a measure of the goodness-of-fit of the model to the data (Elith et al., 2008). AUC is the area under the Receiver Operator Curve (graph of the true positive rate versus the false positive rate) and quantifies the discrimination ability of a model. It is commonly used to evaluate SDMs and values greater than 0.8 often indicate models with sufficient performance. It should be noted, however, that AUC has been shown to be misleading in many instances (e.g., Journé et al. 2019; Lobo et al., 2008; Warren et al., 2019), especially for non-equilibrium species, such as invasive species, because errors of omission and commission are equally weighted (Townsend Peterson et al., 2008).

To obtain independent validation metrics, we also evaluated all models for AUC, corrected partial AUC (pAUC), and sensitivity over the Midwest region using the presence/pseudoabsence dataset used for traditional model building. We chose to validate all models over this area because it included the expanding northern range front of all of the invasive species and represents the regions where we are most interested in making projections. These evaluations allowed us to make a fair comparison of the predictive and discriminatory value between traditional and surrogate models. We used partial AUC (pAUC) in addition to AUC because AUC equally weights errors of omission and commission. However, commission errors in presence-only models are calculated using pseudoabsences or background points, which do not represent true negatives (Townsend Peterson et al., 2008). Corrected partial AUC is derived from the ROC curve (as in AUC) but focuses on the region of the curve where sensitivity is greater than a specified predictive value. This approach allows for evaluation in the range of the ROC where omission errors are low and avoids placing equal importance on commission errors. For pAUC values to be comparable across models, they are corrected using the expected null model prediction (area under the ROC 1:1 line for the same portion of the ROC curve) and standardized to the interval [0.5, 1], with 0.5 indicating no discriminatory power and 1 indicating perfect discriminatory power (Robin et al., 2011; Townsend Peterson et al., 2008). We chose to evaluate pAUC at sensitivities greater than or equal to 0.95. AUC and pAUC estimates were made using the 'pROC' package (Robin et al., 2011). Sensitivity quantifies the correctly predicted positive fraction of occurrences and may be particularly valuable for evaluating the predictive capabilities of invasive species models (Radosavljevic & Anderson, 2014). We evaluated point estimates of sensitivity at a threshold of 0.5. Additionally, to provide a more synthetic evaluation, we also calculated and visualized the change in sensitivity across all threshold values (*hereafter*: sensitivity–threshold curves).

Models were projected using the 'predict' function in the 'dismo' package (Hijmans et al., 2017) and visualized using the 'rasterVis v0.45' package (Lamigueiro & Hijmans, 2019).

3 | RESULTS

3.1 | Surrogate SDMs indicated a broader environmental niche for invasive species

Traditional and surrogate models converged for all three species. Models were built with between 5400–8250 trees (Table 2). Both traditional and surrogate models had high training AUC scores (0.86 or greater; Table 2). When evaluating models using cross validation in the area that the model was built (i.e., the Midwest for traditional models and MN with surrogate occurrences for surrogate models),

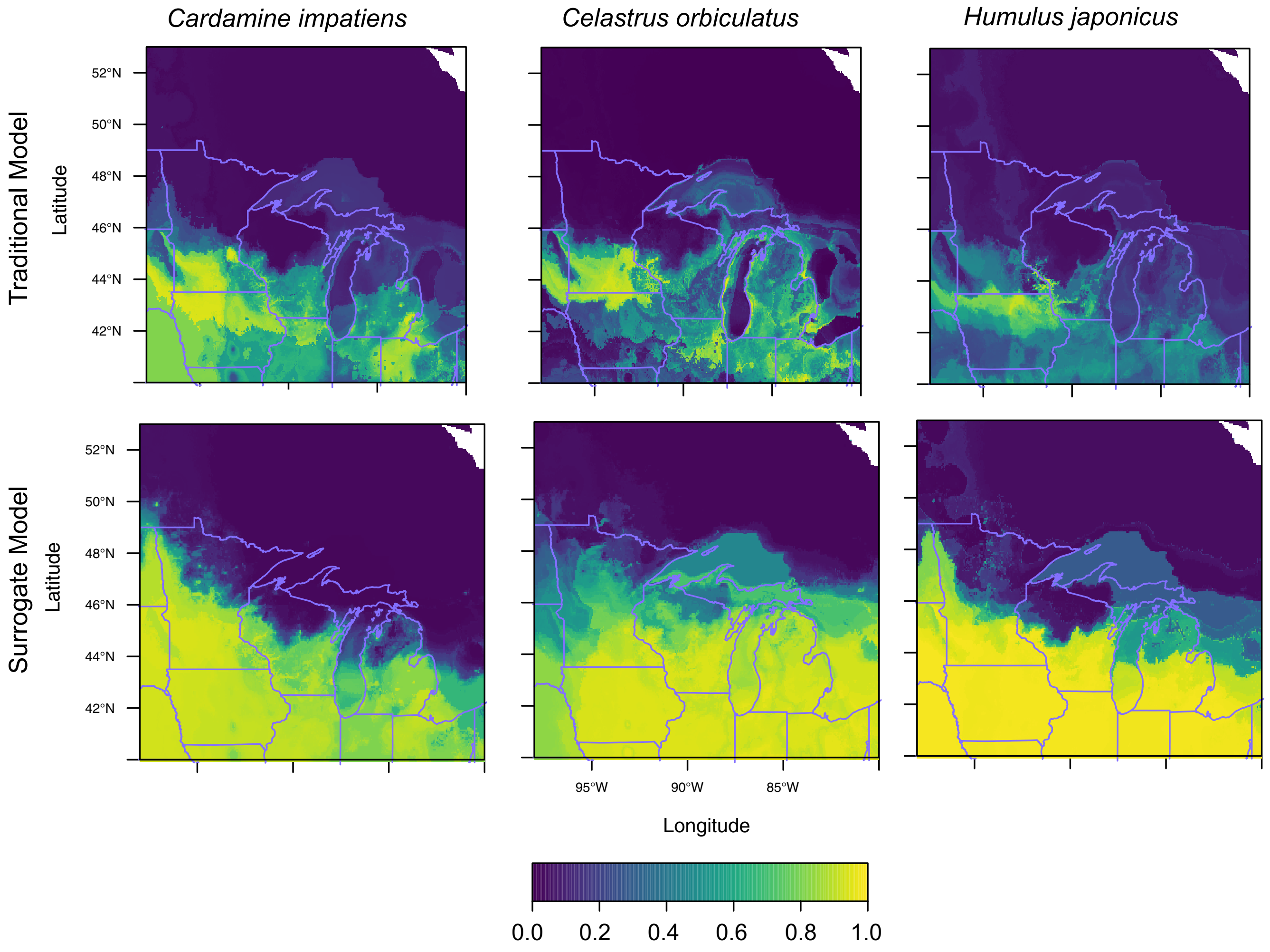

FIGURE 3 Projections of climate suitability from traditional and surrogate SDMs in MN and surrounding area for *C. impatiens* (left column), *C. orbiculatus* (center column), and *H. japonicus* (right column). Maps in the top row result from traditional SDM projections. Maps in the bottom row result from surrogate SDM projections. Map pixels represent a relative metric of climate suitability, where 1.0 is the most suitable climate. Values above 0.9 are considered high suitability, 0.6–0.9 moderate suitability, and below 0.6 poor suitability [Colour figure can be viewed at wileyonlinelibrary.com]

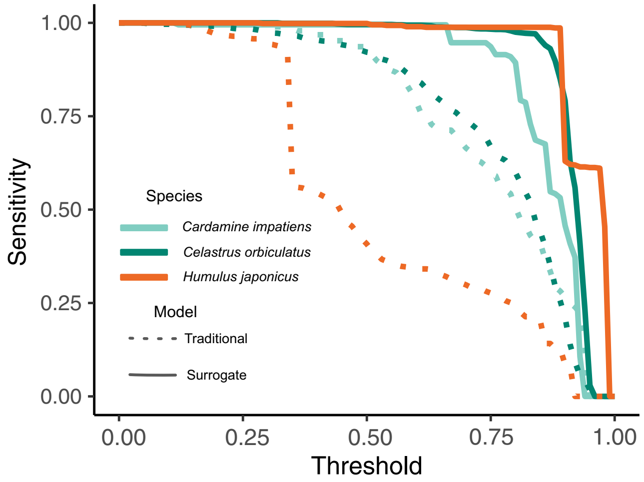

FIGURE 4 Sensitivity–threshold curves of occurrence data from states nearby Minnesota. The curves show how sensitivity changes across thresholds from 0 to 1 for the three species surveyed: *C. impatiens* (light teal), *C. orbiculatus* (teal), and *H. japonicus* (orange). Traditional models represented by dashed lines and surrogate models represented by solid lines [Colour figure can be viewed at wileyonlinelibrary.com]

traditional and surrogate models both performed well; traditional models had slightly higher AUCs (0.85–0.90) than surrogate models (0.85–0.96; Table 2). Surrogate models had lower estimates of deviance, residual deviance, and cross-validation deviance (Table 2).

Response curves from surrogate models had a slightly larger niche breadth along all measured environmental axes. Additionally, the gradients of fitted functions for surrogate models were less sharp when transitioning from less suitable to more suitable environmental conditions (Figure S18). For most species and models, mean temperature of the warmest quarter had the most relative influence (~18%–96%; Figure S18), followed by minimum temperature of the coldest month (~1%–72%; Figure S18). In most models, CMI had the lowest relative importance (~3%–11%; Figure S18).

3.2 | Surrogate SDMs had greater predictive power and similar discrimination and accuracy across the expanding range of the Upper Midwest

When evaluating models with a common dataset: invasive species presences and pseudoabsences from the Midwest (i.e., the datasets used to build traditional models; Figure 1A-C), we found that surrogate models had similar accuracy as traditional models. For AUC, traditional model metrics were generally higher than the surrogate model metrics; AUC was 0.93 versus 0.88 for *C. impatiens*, 0.93 versus 0.78 for *C. orbiculatus*, and 0.80 versus 0.66 for *H. japonicus* (Figure S19). However, when using pAUC, which focuses on the region of the ROC with low omission rates, the accuracy of models was more even between model types. pAUC for traditional models versus surrogate models was 0.81 versus 0.79 for *C. impatiens*, 0.81 versus 0.78 for *C. orbiculatus*, and 0.72 versus 0.68 for *H. japonicus* (Figure S19).

Surrogate models had greater sensitivities than traditional models in the Midwest, especially along the northern range margin of Upper Midwestern states (Figure 3; Table 2). When evaluated at a threshold of 0.5, sensitivities increased between traditional and surrogate models from 0.94 to 0.99 for *C. impatiens*, 0.92 to 1.0 for *C. orbiculatus*, and 0.41 to 1.0 for *H. japonicus* (Table 2). Differences in sensitivities for each of the species were mainly driven by increased suitability at the expanding range boundaries, which included more occurrence points found at this leading edge (Figure 3 & Figure S20).

Sensitivity–threshold curves (Figure 4) showed that in the Midwest sensitivity decreased more quickly at low threshold values for traditional models relative to surrogate models. This pattern is apparent for all three species and especially noticeable for *H. japonicus*. This pattern indicated that traditional models had a greater proportion of true occurrence records in areas that were predicted to be of low suitability. The pattern persisted across all threshold values <0.9. At threshold values >0.9, surrogate models and traditional models were most similar. This indicated that traditional models were more likely to project very high climate suitability in areas that were already densely invaded (Figures 3, 4, & Figure S19).

3.3 | Surrogate SDMs project greater northern range expansion

Across the Midwest, surrogate models had a ca. 1.5 to 6-fold increase in area of suitable climate (0.5 threshold) relative to traditional models. The regions projected to be of highest suitability in traditional SDMs were primarily focused on the areas that have already been invaded. Projections from surrogate models indicated high climate suitability across many regions in the Upper Midwest (Figure 3). The proportion of suitable climate in Midwest increased between traditional and surrogate models from 26% to 40% for *C. impatiens*, 18% to 56% for *C. orbiculatus*, and 7% to 42% for *H. japonicus*. Projections for surrogate models predicted range filling for all species compared to traditional models and also increased predicted northern range extents in the Upper Midwest. The northern extent of each species projected range increased in the surrogate compared to traditional models: by 3° latitude for *C. impatiens* (46°N to 49°N), by 2° latitude for *C. orbiculatus* (46°N to 48°N), and by 4° latitude for *H. japonicus* (45 °N to 49°N; Figure 3).

4 | DISCUSSION

Building SDMs for invasive species has proven challenging, often leaving biogeographers and decision-makers with incomplete or uninformative models of invasion risk. Although mechanistic and process-based models incorporate more information about species' biology (Merow et al., 2014, 2017; Singer et al., 2016), the data required are often costly to obtain and not feasible for most species

(Singer et al., 2016). In this study, we pursued an alternative strategy that leveraged the power of other species that are both (1) tightly associated with the focal species occurrences and (2) have broader geographic ranges. Unsurprisingly, traditional and surrogate SDMs performed similarly in the regions where models were trained; however, surrogate models were superior in predicting occurrences outside of the region where the model was built. Surrogate models also predicted substantially more suitable areas beyond the range margin compared to traditional SDMs (1.5- to 6-fold more). This information provides actionable information for detection and eradication efforts as well as prioritizing surveillance for species that are likely to establish (https://www.cbd.int/invasive/done.shtml).

4.1 | Surrogate models had higher sensitivity outside of focal area where models were built

As might be expected, traditional and surrogate SDMs performed similarly for AUC and pAUC given the ability of both models to predict occurrences as well as regions of highly unsuitable climate. However, surrogate models had higher sensitivities and were superior in predicting occurrences at the expanding northern range, outside of the area where it was built. The difference was greatest for *C. orbiculatus* and *H. japonicus*, and less pronounced for *C. impatiens*. The differences among species were likely due to the idiosyncrasies of the occurrence data in the states surveyed (i.e., some species had more occurrence records along range edges). Nonetheless, increased sensitivities were a general outcome of surrogate models and were due to in-filling of projected suitable areas along the northern invasion fronts and northward extensions of projected suitable area for all of the United States (Figure 3 & S20). The increased sensitivities indicated that surrogate models had the capacity to identify climates of high suitability that were missed by traditional models. The identification of this additional suitable climate greatly increased the areas of projected risk across the region of interest, the Upper Midwest.

4.2 | Surrogate models included more potential suitable climates and predicted greater areas at risk of invasion

Surrogate model projections identified more suitable climates at the northern range margins that was at risk of invasion. This result suggests that surrogate models identified a greater range of potentially suitable environments for model building and projection by leveraging the occurrence records of native species with similar climatic environmental responses. For invasive species, one of the main challenges for predicting range expansion is that environments at the expanding range margin (and beyond) cannot be included in model building because they have few (or no) occurrences. Those beyond-range environments challenge model projections because species–environment relationships are not truly known (Broennimann et al., 2007; Elith et al., 2010; Guisan et al., 2014; Mesgaran et al., 2014). Although we specifically included pseudoabsences in our traditional SDMs from regions that we assumed to be highly unsuitable (Chapman et al., 2019), there remains a large portion of environmental variable space where species–climate relationships are uncharacterized. Extrapolations of models to environments outside the range edge may still result in either over- or under-prediction of climate suitability, and in the case of our models, may have been associated with under-prediction for many areas in MN and the Upper Midwest. Inclusion of surrogate species greatly increased information about potential species–climate relationships in climatically analogous environments along the northern range margin for all three species. Beyond surrogate SDMs, mechanistic models are the most common approach for projecting into novel environments, such as those beyond range limits (Guisan et al., 2014). Thus, including surrogate species records helped to minimize potential pitfalls inherent in predicting range expansion.

4.3 | Methodological considerations for producing surrogate SDMs

The choice of surrogate species may impact the performance of surrogate models and is an open area for future research. We used the two most environmentally similar species and one or two most environmentally dissimilar species as surrogate presence and absence data. These species were drawn from a set of approximately 25 community members that we included in the JSDM analyses for each invasive species. We restricted our community datasets for two reasons, (1) to reduce computational resources required and (2) to limit the inclusion of extremely common and extremely rare species. Rare or common species often do not have strong environmental signals that can be leveraged for surrogate models; however, measures of commonness and rarity will depend on the scale over which the model is built (e.g., in our case, the state of MN).

Our choice of surrogate species was based on how similar the environmental responses of the invasive species were to native species. Previous approaches that leveraged multiple species in SDMs relied on either detailed information about a species' biotic interactions (e.g., Araújo & Luoto, 2007; Baselga & Araújo, 2009; Mäkinen & Vahatalo, 2018; Mod et al., 2015; Molloy et al., 2017; Pellissier et al., 2010) or incorporated aggregated metrics of species' environmental responses to correct individual SDMs (e.g., Hui et al., 2013; Larson et al., 2014). Although these methods produced more realistic projections, each had limitations that can make them challenging to apply to invasive species (e.g., modelers rarely have detailed knowledge of biotic interactions). We extended this general approach for invasive species by including distributional information on environmentally similar native or common species that function as phytometers of suitable and unsuitable climates. This also allowed us to explicitly include more potential environments that were positively and negatively associated with our species.

We took two different approaches to narrowing the initial species lists from approximately 50 to 25. For *C. impatiens*, we limited our community list to other forest herbaceous species. For *C.*

orbiculatus and *H. japonicus*, we randomly selected species from the preliminary lists because, as vines, they occur in a range of habitats. For our three focal species as well as our native case study species, both procedures produced similar results. Additionally, in early preliminary analyses, we tried different and/or enlarged species lists, which resulted in qualitatively similar results (not shown). We suggest that for other invasive species, modelers explore a variety of species lists to check the robustness of the surrogates used.

Concomitant with the choice of community associates, it is also important to consider the spatial resolution over which environmental similarity is ascertained. We included communities from across MN because that was the region over which we were most interested in producing reliable projections. In preliminary analyses, we also ran JSDMs based on relevés from only a portion of MN. Our results were similar, and species' environmental responses did not differ substantially. However, over larger spatial scales, the e-matrix may change in important ways that could affect species selection. Therefore, we suggest testing multiple spatial resolutions and/or limiting the spatial resolution to a focal region rather than a continent. This will likely help ensure that the environmental relationships used to choose surrogate species hold over the areas of interest for prediction. Additionally, if researchers intend to use SDMs for prediction, they should be evaluated using spatially explicit cross-validation (Roberts et al., 2017).

Our surrogate models were informed by an extensive relevé dataset (MN DNR), which comprises a dense sample of plant communities from across our area of interest. However, modelers may not have access to datasets of this completeness for many geographic regions. Therefore, it may be necessary to modify our methods and use individual species occurrence records from databases such as GBIF to build pseudo-co-occurrence datasets at the spatial scale of the environmental data. As herbarium and museum records become increasingly available (e.g., GBIF: www.gbif.org, iNaturalist: www.inaturalist.org; Birds of Fennoscandia: Lindström et al., 2015; Butterflies for the new millennium: Asher et al., 2001), methods of constructing pseudo-co-occurrence datasets should become increasingly possible.

5 | Conclusions

Using community and co-occurrence information provides a powerful and cost-effective tool for improving the predictions of correlative SDMs. Surrogate models effectively projected suitable habitat beyond ranges by borrowing information on occurrence–environment relationships from co-occurring native species. In particular, surrogate models better predicted occurrences in other parts of the current range outside of the region where the model was built. To use this approach, modelers may be able to leverage already available co-occurrence datasets or create datasets using occurrence records for this process.

ACKNOWLEDGEMENTS

We thank Rob Venette for early discussions about ideas in this manuscript as well as Kady Wilson and Zack Radford for early help in gathering invasive species occurrence records and field surveys. We thank Andy Holdsworth, Laura Van Riper, and the MN Department of Natural Resources for providing the community relevé 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). No permits were required for this research.

DATA AVAILABILITY STATEMENT

All data are publicly available. R scripts, processed occurrence records, and co-occurrence matrices that can be used to run our analyses are available at https://doi.org/10.13020/93r8-zk72. All invasive species occurrence records were obtained through GBIF (see Supplementary Materials for DOI) and EDDMaps. Raw relevé and species co-occurrence data are available upon request from the Minnesota Department of Natural Resources.

ORCID

Ryan D. Briscoe Runquist https://orcid.org/0000-0001-7160-9110

Thomas A. Lake https://orcid.org/0000-0002-8836-5164

David A. Moeller https://orcid.org/0000-0002-6202-9912

REFERENCES

Araújo, M. B., & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography: A Journal of Macroecology*, *16*(6), 743–753.

Asher, J., Warren, M., & Fox, R. (2001). *The Millennium atlas of butterflies in Britain and Ireland*. Oxford University Press.

Baselga, A., & Araújo, M. B. (2009). Individualistic vs community modelling of species distributions under climate change. *Ecography*, *32*(1), 55–65.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., & Rouault, E. (2017). *Rgdal: bindings for the "geospatial"data abstraction library. R package version 1* (pp. 2–16).

Bivand, R., & Rundel, C. (2015). rgeos: Interface to Geometry Engine—Open Source (GEOS). R package version 0.3–19. See https://cran.R-Project.Org/package=Rgeos.

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 & Evolution*, *25*(5), 310–318.

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*(1), 2426.

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N. E., Graham, C. H., & Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography: A Journal of Macroecology*, *21*(4), 481–497.

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*(8), 701–709.

Chapman, D., Pescott, O. L., Roy, H. E., & Tanner, R. (2019). Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. *Journal of Biogeography*, *46*(5), 1029–1040.

Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*, *87*(1), 34–56. https://doi.org/10.1002/ecm.1241.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, 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*(2), 129–151.

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution / British Ecological Society*, *1*(4), 330–342.

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*(1), 677–697.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *The Journal of Animal Ecology*, *77*(4), 802–813.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315.

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: A Journal of Macroecology*, *27*(2), 245–256.

Gallien, L., Douzet, R., Pratte, S., Zimmermann, N. E., & Thuiller, W. (2012). Invasive species distribution models - how violating the equilibrium assumption can create new insights: Beyond the equilibrium assumption of SDMs. *Global Ecology and Biogeography: A Journal of Macroecology*, *21*(11), 1126–1136.

Gallien, L., Münkemüller, T., Albert, C. H., Boulangeat, I., & Thuiller, W. (2010). Predicting potential distributions of invasive species: Where to go from here? *Diversity and Distributions*, *16*(3), 331–342.

Greenwell, B., Boehmke, B., & Cunningham, J. (2019). *Package "gbm". Generalized boosted regression models (Version, 2.1.5), CRAN*. https://github.com/gbm-developers/gbm.

Gren, I.-M., Isacs, L., & Carlsson, M. (2009). Costs of alien invasive species in Sweden. *Ambio*, *38*(3), 135–140.

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: Insights from biological invasions. *Trends in Ecology & Evolution*, *29*(5), 260–269.

Harris, D. J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution / British Ecological Society*, *6*(4), 465–473.

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). *Package "dismo": Species distribution modeling (Version 1.1-4)*. https://cran.r-project.org/web/packages/dismo/dismo.pdf.

Hijmans, R. J., & vanEtten, J. (2012). *raster: Geographic analysis and modelling with raster data. R package vzrersion 20-08*. https://rspatial.org/raster.

Hui, F. K. C., Warton, D. I., Foster, S. D., & Dunstan, P. K. (2013). To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models. *Ecology*, *94*(9), 1913–1919.

Journé, V., Barnagaud, J.-Y., Bernard, C., Crochet, P.-A., & Morin, X. (2019). Correlative climatic niche models predict real and virtual species distributions equally well. *Ecology*, https://doi.org/10.1002/ecy.2912.

Lamigueiro, O. P., & Hijmans, R. (2019). *Package "rasterVis": Visualization methods for raster data (Version 0.47)*. https://github.com/oscarperpinan/rastervis/issues.

Larson, E. R., Gallagher, R. V., Beaumont, L. J., & Olden, J. D. (2014). Generalized "avatar" niche shifts improve distribution models for invasive species. *Diversity and Distributions*, *20*(11), 1296–1306. https://doi.org/10.1111/ddi.12233.

Lindström, Å., Green, M., Husby, M., Kålås, J. A., & Lehikoinen, A. (2015). Large-scale monitoring of waders on their boreal and arctic breeding grounds in northern Europe. *Ardea*, *103*(1), 3–15. https://doi.org/10.5253/arde.v103i1.a1.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography: A Journal of Macroecology*, *17*(2), 145–151.

Lodge, D. M., Simonin, P. W., Burgiel, S. W., Keller, R. P., Bossenbroek, J. M., Jerde, C. L., Kramer, A. M., Rutherford, E. S., Barnes, M. A., Wittmann, M. E., Chadderton, W. L., Apriesnig, J. L., Beletsky, D., Cooke, R. M., Drake, J. M., Egan, S. P., Finnoff, D. C., Gantz, C. A., Grey, E. K., … Zhang, H. (2016). Risk analysis and bioeconomics of invasive species to inform policy and management. *Annual Review of Environment and Resources*, *41*(1), 453–488.

Mäkinen, J., & Vanhatalo, J. (2018). Hierarchical Bayesian model reveals the distributional shifts of Arctic marine mammals. *Diversity and Distributions*, *24*(10), 1381–1394.

Marbuah, G., Gren, I.-M., & McKie, B. (2014). Economics of harmful invasive species: A review. *Diversity*, *6*(3), 500–523.

Merow, C., Bois, S. T., Allen, J. M., Xie, Y., & Silander, J. A. Jr (2017). Climate change both facilitates and inhibits invasive plant ranges in New England. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(16), E3276–E3284.

Merow, C., Dahlgren, J. P., Metcalf, C. J. E., Childs, D. Z., Evans, M. E. K., Jongejans, E., Record, S., Rees, M., Salguero-Gómez, R., & McMahon, S. M. (2014). Advancing population ecology with integral projection models: A practical guide. *Methods in Ecology and Evolution / British Ecological Society*, *5*(2), 99–110.

Mesgaran, M. B., Cousens, R. D., & Webber, B. L. (2014). Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity & Distributions*, *20*(10), 1147–1159.

Mod, H. K., le Roux, P. C., Guisan, A., & Luoto, M. (2015). Biotic interactions boost spatial models of species richness. *Ecography*, *38*(9), 913–921.

Molloy, S. W., Davis, R. A., Dunlop, J. A., & van Etten, E. (2017). Applying surrogate species presences to correct sample bias in species distribution models: a case study using the Pilbara population of the Northern Quoll. *Nature Conservation*, *18*, 25.

Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., … Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, *89*(3), 18.

Pebesma, E., & Bivand, R. S. (2005). S classes and methods for spatial data: the sp package. *R News*, *5*(2), 9–13.

Pellissier, L., Bråthen, K. A., Pottier, J., Randin, C. F., Vittoz, P., Dubuis, A., Yoccoz, N. G., Alm, T., Zimmermann, N. E., & Guisan, A. (2010). Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, *33*(6), 1004–1014.

Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics: the Journal of the International Society for Ecological Economics*, *52*(3), 273–288.

Pollock, L. J., Tingley, R., & Morris, W. K. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution / British Ecological Society*, *5*, 397–406.

Qiao, H., Peterson, A. T., Ji, L., & Hu, J. (2017). Using data from related species to overcome spatial sampling bias and associated limitations in ecological niche modelling. *Methods in Ecology and Evolution*, *8*, 1804–1812.

R Core Team, & Others (2013). *R: A language and environment for statistical computing (Version 3.3.2) [Computer software]*. ftp://ftp.uvigo.es/CRAN/web/packages/dplR/vignettes/intro-dplR.pdf.

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, *41*(4), 629–643. https://doi.org/10.1111/jbi.12227.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77.

Singer, A., Johst, K., Banitz, T., Fowler, M. S., Groeneveld, J., Gutiérrez, A. G., Hartig, F., Krug, R. M., Liess, M., Matlack, G., Meyer, K. M., Pe'er, G., Radchuk, V., Voinopol-Sassu, A.-J., & Travis, J. M. J. (2016). Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? *Ecological Modelling*, *326*, 63–74.

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

Warren, D. L., Matzke, N. J., & Iglesias, T. L. (2019). Evaluating species distribution models with discrimination accuracy is uninformative for many applications. In *bioRxiv*, 684399. doi: https://doi.org/10.1101/684399.

Zhang, C., Chen, Y., Xu, B., Xue, Y., & Ren, Y. (2018). Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography*, *41*(11), 1876–1887.

Zomer, R. J., Trabucco, A., Bossio, D. A., van Straaten, O., & Verchot, L. V. (2008). Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agric. Ecosystems and Environment*, *126*, 67–80.

Zurell, D., Pollock, L. J., & Thuiller, W. (2018). Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? *Ecography (Vol.*, *41*(11), 1812–1819. https://doi.org/10.1111/ecog.03315.

BIOSKETCH

Dr. R. D. Briscoe Runquist is a Postdoctoral Research Associate with Dr. D. A. Moeller; her research focuses on the biogeography and landscape genomics of invasive plants. Thomas Lake is a Ph.D. student with D.A.M. 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.

Authors contribution: RDBR and DAM conceived of this study. RDBR and TAL gathered data and ran the analysis. RDBR wrote the initial draft of the manuscript. All authors contributed substantially to editing the manuscript.

SUPPORTING INFORMATION

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

How to cite this article: Briscoe Runquist RD, Lake TA, Moeller DA. Improving predictions of range expansion for invasive species using joint species distribution models and surrogate co-occurring species. *J Biogeogr*. 2021;48:1693–1705. https://doi.org/10.1111/jbi.14105