Received: 11 June 2024 | Revised: 29 October 2024 | Accepted: 26 November 2024

DOI: 10.1002/ecs2.70138

ARTICLE

ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

Spatial differences in predicted *Phalaris arundinacea* (reed canarygrass) occurrence in floodplain forest understories

John T. Delaney | **M. Van Appledorn** | **N. R. De Jager** | **K. L. Bouska** | **J. J. Rohweder**

U.S. Geological Survey, La Crosse, Wisconsin, USA

Correspondence
John T. Delaney
Email: jdelaney@usgs.gov

Funding information
U.S. Army Corps of Engineers' Upper Mississippi River Restoration Program

Handling Editor: Natalie A. Griffiths

Abstract

Reed canarygrass (*Phalaris arundinacea* L.) is one of the most common invaders of floodplains and wetlands in North America. In the Upper Mississippi River floodplain, invasion by reed canarygrass in forest understories can inhibit forest regeneration when gaps form in the overstory. Understanding the distribution of reed canarygrass in forest understories is essential for effective management and control. We used an ensemble of species distribution models including Bayesian additive regression trees, boosted trees, and random forest algorithms to predict habitat suitability for reed canarygrass in forest understories across the Upper Mississippi River floodplain (~41,000 ha). Data from forest inventory study plots with reed canarygrass presence and absence were combined with 10 hypothesized environmental predictors of reed canarygrass invasion. We applied three approaches to better understand and incorporate the influence of spatial autocorrelation among our predictor variables, including random cross-validation, spatial cross-validation, and spatial cross-validation with Euclidean distance fields. Flood frequency, distance to contiguous floodplain, distance to forest edge, and distance to invaded wet meadow were among the most important environmental predictors across the three algorithms. Generally, the mean probability of reed canarygrass presence decreased with increasing flood depth, distance to contiguous floodplain, distance to invaded wet meadow, forest cover, and forest height, while relationships with other predictors were more variable. The ensemble of the three models (i.e., the average prediction) was used to map and summarize potential reed canary grass habitat suitability across the landscape. The maps generated quantified the habitat suitability for reed canarygrass and areas of agreement among the models in forest understories across the floodplain. This information can be used to better understand the extent of invasion, prioritize restoration efforts, and develop further research.

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.
Published 2024. This article is a U.S. Government work and is in the public domain in the USA. *Ecosphere* published by Wiley Periodicals LLC on behalf of The Ecological Society of America.

KEYWORDS
ensemble modeling, Euclidean distance fields, floodplain forest invasion, habitat suitability modeling, *Phalaris arundinacea*, reed canarygrass, spatial autocorrelation, spatial cross-validation, species distribution modeling

INTRODUCTION

Invasive species are increasing in prevalence in ecosystems worldwide as a result of translocation (Hulme, 2009) and anthropogenic alterations to the environment (van der Wal et al., 2008). Management of invasive species can often be costly, requiring prioritization of the most impacted areas or areas with the greatest potential for control and restoration success (Januchowski-Hartley et al., 2018). It can be difficult to inventory-invaded locations and quantify the extent of the invasion across broad spatial scales. Habitat suitability modeling offers a useful tool for predicting habitat suitability for species across space using a selection of environmental predictors (Elith & Leathwick, 2009; Guisan & Zimmermann, 2000; Miller, 2010). Outputs from habitat suitability models can be used by managers to prioritize management efforts across the landscape (Sofaer et al., 2019).

Phalaris arundinacea L. (reed canarygrass, hereafter RCG) is a cool-season perennial grass that has been suggested as a biological model for studying plant invasions (Lavergne & Molofsky, 2004) and is considered one of the most pervasive plant invaders in wet areas in North America (Lavergne & Molofsky, 2006). RCG has been planted as a forage crop since the mid-1800s and has been used for soil stabilization/remediation, wastewater treatment, and as a bioenergy crop (Lavergne & Molofsky, 2004). RCG's broad distribution can be partially attributed to its ability to persist across a wide range of environmental conditions and disturbances, including flooding (Kercher & Zedler, 2004a, 2004b), drought (Wilkins & Hughes, 1932), and grazing (Hillhouse et al., 2010). Although RCG is native to midwestern North America, it has been widely accepted that the native genotypes have been lost through displacement by, or hybridization with, Eurasian genotypes and cultivars (Merigliano & Lesica, 1998). Recent research suggests, however, that many populations within the midwestern United States may be of native origin with little hybridization with non-native genotypes or cultivars (Noyszewski et al., 2021). Native genotypes are still considered highly invasive and the status of many populations is unknown (Anderson et al., 2021). Regardless of its status, RCG has increased in prevalence in ecosystems across North America in recent decades, resulting in changes in ecological function (De Jager et al., 2015; De Jager, Swanson, et al., 2019; Kreiling et al., 2015; Swanson et al., 2017), complications with ecosystem restoration (Adams & Galatowitsch, 2005; Matthews et al., 2020; Stiles et al., 2008; Thomsen et al., 2012), and loss of biodiversity (Annen et al., 2008; Kirsch & Gray, 2017; Price et al., 2020; Spyreas et al., 2010; Weilhoefer et al., 2017; Werner & Zedler, 2002).

Several factors contribute to the invasiveness of RCG. RCG reproduces through outcrossed seeds and vegetatively through rhizomes and tillers (Lavergne & Molofsky, 2006). RCG is more tolerant of inundation than many other native herbaceous species, but is susceptible to prolonged inundation (greater than 3 days/week under 9–10 weeks of cyclic inundation; Rice & Pinkerton, 1993) and greater depth of submergence (>0.85 m in late spring; Jenkins et al., 2008). Flood events could potentially facilitate dispersal to new sites (Elliott et al., 2019), but prolonged inundation could inhibit seed germination and survival (Coops & van der Velde, 1995). RCG can survive high rates of sedimentation (Chen et al., 2014), which can facilitate invasion by reducing microtopographic relief, species richness, and canopy cover (Kercher & Zedler, 2004a; Maurer et al., 2003; Werner & Zedler, 2002). RCG flourishes under high nutrient conditions through increasing aboveground growth and rhizome spread (Galatowitsch et al., 1999; Kercher & Zedler, 2004a; Maurer & Zedler, 2002), and nutrient levels can influence its competitive ability with other plants (Green & Galatowitsch, 2002; Perry et al., 2004). Additionally, legacies of past soil disturbance can influence the distribution of RCG (Bart et al., 2011). While these factors promote the spread of RCG, low light availability can be a constraint on RCG establishment and growth (Adams et al., 2011; Lindig-Cisneros & Zedler, 2002; Perry & Galatowitsch, 2004).

Although RCG is limited by light availability in forest understories, when gaps in the forest overstory form, RCG can take advantage of increased light conditions and form dense stands that inhibit natural forest regeneration (De Jager et al., 2013; Knutson & Klaas, 1998; Thomsen et al., 2012). RCG dominance can be further promoted by the presence of white-tailed deer (*Odocoileus virginianus*), which can slow forest regeneration through herbivory of woody vegetation (Cogger et al., 2014; De Jager et al., 2013). RCG monocultures can persist without management intervention or disturbance by limiting

light availability for other plant species and increasing nitrogen availability (Swanson et al., 2017). This is especially problematic in floodplain systems because they often lack shade-tolerant tree species and can have poorly developed woody understories. Over the long term, as overstory trees reach their longevity, the presence of RCG in the forest understory has the potential to cause a shift from forest cover to wet meadow monocultures (Bouska et al., 2020).

In the floodplain of the Upper Mississippi River (UMR), vegetation has been described as existing in two alternate regimes, a dynamic floodplain vegetation mosaic comprising native wet meadows and forests of varying ages and structures versus a wet meadow monoculture of invasive plants, including RCG (Bouska et al., 2020). Vegetation communities in the floodplain are structured largely by inundation regimes with herbaceous communities found in areas with the lengthiest flooding, followed by floodplain forest at intermediate durations, and more diverse forest communities in areas experiencing the shortest periods of flooding (De Jager et al., 2016). The floodplain forests of the UMR are less diverse than they were prior to Euroamerican settlement and have become dominated by flood-tolerant tree species (Romano, 2010), which are reaching their longevity and are susceptible to mortality due to increased inundation and other factors (De Jager, Van Appledorn, et al., 2019). Patterns of inundation have changed since a series of locks and dams were constructed in the 1930s, and annual discharges on the UMR have increased since 1940 (Van Appledorn, 2022), potentially due to changes in climate, land use, or both (Frans et al., 2013). Regional climate change analyses indicate that changes in precipitation and discharge are likely to continue, although the direction and magnitude are uncertain (Byun et al., 2019; Neri et al., 2020).

It is important to understand which areas may be more susceptible to RCG invasion following forest loss to design effective monitoring programs, target restoration activities, and to predict future changes in RCG invasion in the UMR. When invasive species dominate the canopy, it may be possible to map the distribution using aerial or satellite imagery (Bradley, 2014; Walsh, 2018), including RCG (Bernthal & Willis, 2004; De Jager et al., 2017; Narumalani et al., 2009). However, when invasion is at early stages or the species is covered by a canopy, detection can be more difficult (Joshi et al., 2004). Habitat suitability modeling is one approach to mapping the potential distribution of an invasive species by combining species occurrence data with ecologically relevant environmental variables to predict species occurrence across the landscape (Miller, 2010). There are many machine learning algorithms available to create predictive models (Gobeyn et al., 2019; Kaky et al., 2020; Li & Wang, 2013) that are well suited for studying ecological systems because they can learn complex relationships and do not require strict assumptions (Thessen, 2016). A popular approach in habitat suitability modeling includes combining several algorithms together in an ensemble to account for the variability in predictions from individual models, improve predictive performance, and reduce individual model bias (Grenouillet et al., 2011; Hao et al., 2019; Marmion et al., 2009).

Spatial autocorrelation is a common concern with ecological modeling (Dormann et al., 2007; Koenig, 1999; Miller, 2012). Typically, classification models are constructed using cross-validation folds where a random subset of sites are held back while training the model, and performance is evaluated on how well the trained model predicts the held-out set of sites for each fold. However, if there is considerable spatial autocorrelation in the dataset, the resulting performance measures can be overly optimistic and not an accurate representation of how the model might perform when predicting habitat suitability in new areas (Gazis & Greinert, 2021; Hao et al., 2020; Ploton et al., 2020; Roberts et al., 2017; Valavi et al., 2023). Several approaches for spatially explicit cross-validation have been developed (Roberts et al., 2017). Additionally, standard classification models do not account for the inherent spatial structure within a dataset and treat observations as independent. There are several methods for incorporating spatial information into machine learning modeling such as the inclusion of Euclidean distance fields as spatial predictors to improve model performance and account for spatial autocorrelation (Behrens et al., 2018).

In this study, we used forest inventory monitoring data combined with ecologically relevant predictors from spatially explicit datasets to construct habitat suitability models to better understand the distribution of RCG during the early stages of invasion in the floodplain forest understory of the UMR. Our goal was to answer the following research questions: (1) Which approaches for estimating and accounting for spatial autocorrelation are most appropriate for our dataset? (2) Could we predict the habitat suitability of RCG in forest understories? (3) What was the spatial extent of predicted understory habitat suitability of RCG? (4) Which factors influenced habitat suitability for RCG? Kettenring et al. (2019) modeled the distribution of RCG across riparian areas of the Columbia and Missouri River basins, but to our knowledge, ours is the first attempt to model RCG habitat suitability that focuses on floodplain forest understories specifically.

METHODS

The UMR (i.e., the portion of the Mississippi River upstream from the confluence with the Ohio River) comprises a series of navigation pools created by 29 lock and dams constructed in the 1930s to manage river levels for navigation. The dams do not control discharge but serve to maintain navigable water levels during low discharge conditions. During high discharge conditions, the dam gates are raised, allowing the floodplain to become inundated. For this study, we focus on forested areas of the UMR floodplain that spans Navigation Pools 3–13, starting below Lock and Dam 2 and ending at Lock and Dam 13 (Figure 1). Annual average temperature ranged from 7.8°C at Lock and Dam 2 to 9.9°C at Lock and Dam 13 and annual precipitation ranged from 846.6 mm at Lock and Dam 2 to 870.5 mm at Lock and Dam 13 based on 30-year climate normal from 1991 to 2020 (NOAA National Centers for Environmental Information, 2023).

Presence ($n = 1977$) and absence ($n = 2425$) records of RCG in forest understories were obtained from the U.S. Army Corps of Engineers' Phase II Forest Inventory dataset (U.S. Army Corps of Engineers-St. Paul District 2022). The forest inventory was conducted on U.S. Army Corps of Engineers fee title lands, which are co-managed with the U.S. Fish and Wildlife Service as part of the

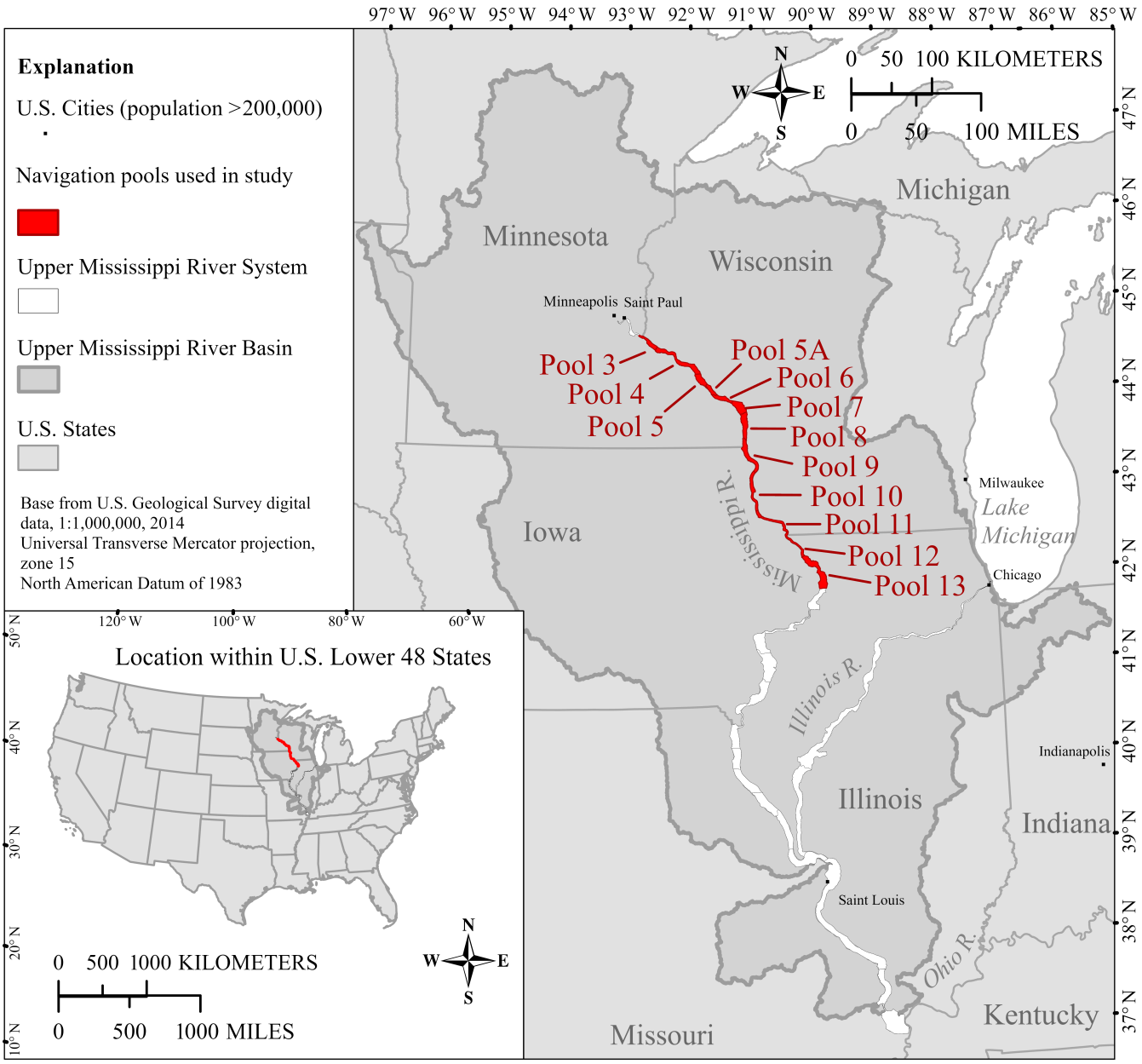


FIGURE 1 Location of Navigation Pools 3–13 of the Upper Mississippi River.

Upper Mississippi River National Wildlife and Fish Refuge. Forest inventory sampling in our study area occurred between 2008 and 2019 at sampling points spaced 100.6 m apart and arranged on a grid. At each sampling point, RCG was recorded within a circular plot (5.1-m radius) if it was at least the third most dominant herbaceous species in the plot or if it was reported as one of the top five notable species in the plot (Meier, 2017). Guideline for recording notable species was determined by the following priority ranking: invasive woody, invasive herbaceous, previously undocumented woody species, and herbaceous species outside of typical floodplain forest vegetation. We assumed that RCG was recorded if RCG was present within the plot because RCG falls second in the list of priority categories (invasive herbaceous), and 90% of plots recorded four or fewer notable species.

Machine learning algorithms learn the relationships between the target variable (in our case a binary classification of RCG presence or absence) and a set of variables referred to as predictors, and once the relationships have been learned, the algorithms use the values of the predictors to predict the target variable where observations are lacking. We obtained several spatially continuous datasets that cover our study area (Table 1) to serve as environmental predictors. We used only spatially continuous datasets in our analysis rather than finer scale plot-level data because we sought to predict RCG distribution across all forested areas of the floodplain. We focused on only forested areas because: (1) our presence/absence observations only contained plots from forested areas; (2) the relevance to management concerns related to the potential for RCG invasion to inhibit forest regeneration (Bouska et al., 2020); and (3) RCG in unforested areas has already been mapped by interpreting aerial imagery (De Jager et al., 2017). Although metrics derived from temperature and precipitation are frequently used in species distribution models, we did not include those variables in our analysis because of the relatively small study area that has very little difference in precipitation and temperature and because of the wide geographical distribution of RCG that is found across temperate areas of five continents (Galatowitsch et al., 1999). Additionally, although digital elevation products are often used in habitat suitability modeling, we did not include elevation as a predictor in our model because elevation increased as river mile increased and there were likely a number of factors that were correlated with that gradient that would have confounded ecological interpretation of an elevation variable. Instead, we used metrics related to the long-term inundation regime, which were dependent upon elevation relative to water levels and thus more ecologically informative.

Inundation metrics came from the "Floodplain Inundation Attribute Rasters: Mississippi & Illinois Rivers" dataset (Van Appledorn et al., 2018) and included average event number during the growing season, median event duration during the growing season, and median of the maximum depth of submergence during the growing season over a 40-year period (1972–2011). These inundation metrics exhibited nonlinear relationships among each other, indicating complex interactions of topography, hydrology, and anthropogenic factors across the UMR (Van Appledorn et al., 2021). The inundation rasters were 4-m cells and were used as a template to rasterize all of the following vector datasets to derive the metrics to ensure all layers and outputs were at the same spatial scale and extent.

Historical pre-lock and dam landcover was obtained from the "1890s Mississippi River Commission Surveys Landcover Data: Mississippi River" dataset (LTRM, 2000). Landcover from the "2010 Upper Mississippi River Restoration (UMRR) Program Land Cover Data: Mississippi & Illinois Rivers" dataset (Ruhser et al., 2015) was used to represent landcover at the time of forest inventory sampling. Landcover classifications were interpreted from aerial photography and validated with field surveys according to the methods of Dieck et al. (2015). There were five forest cover classes associated with different flooding conditions, with *Salix* communities, *Populus* communities, and floodplain forests in wetter conditions, lowland forest in areas with moderate flooding, and upland forests in infrequently flooded places (De Jager et al., 2016).

Distance to forest edge rasters was calculated using the Euclidean distance transformation (Shih & Wu, 2004) from the "whitebox" package (Lindsay, 2016) in R (R Core Team, 2020). Wet meadow and wet meadow shrub areas classified as invaded were obtained from the "2010 *Phalaris arundinacea* (Reed canarygrass) mapped locations within Pools 2–13 of the Upper Mississippi River System" dataset (Hoy et al., 2017). Least cost distance to the nearest upstream invaded wet meadow/shrub area was calculated for each forested pixel. The nearest distance between the edge of an island and the contiguous floodplain (i.e., not surrounded by water) was also derived from the 2010 landcover dataset, with any contiguous floodplain areas assigned 0 using ArcGIS version 10.8.1 (ESRI, 2020).

For each raster, we coarsened the 4-m cells to 12-m cells by averaging to ensure that the raster cells had a similar area as the circular sampling plots (5.1-m radius). For each pool, the rasters for each of the predictors were combined in a raster stack, and sampling points were intersected with the stack to obtain predictor values for each sampling location using the "raster" package

TABLE 1 List of environmental predictors used in the model along with reasoning for their inclusion and data sources.

Predictor	Reasoning for inclusion	Source
Flood depth (ft; median of the maximum depth of submergence during the growing season)	Reed canarygrass may be favored under shallower inundation and inhibited under deeper inundation.	Floodplain inundation attribute rasters: Mississippi and Illinois rivers. https://doi.org/10.5066/F7VD6XRT
Flood duration (median event duration during the growing season)	Shorter inundation durations may favor reed canarygrass over other species while longer durations may inhibit reed canarygrass.	Floodplain inundation attribute rasters: Mississippi and Illinois rivers. https://doi.org/10.5066/F7VD6XRT
Flood frequency (average growing season event no.)	The frequency of inundation events could be an important factor influencing reed canarygrass with flood events providing opportunities for propagule transport to new areas along with reed canarygrass's tolerance for periodic inundation.	Floodplain inundation attribute rasters: Mississippi and Illinois rivers. https://doi.org/10.5066/F7VD6XRT
Distance to contiguous floodplain (m)	More isolated islands may have lower deer browsing or lower reed canarygrass propagule pressure.	Derived from 2010 Upper Mississippi River Restoration (UMRR) Program Landcover Data: Mississippi & Illinois Rivers. https://doi.org/10.5066/F77942QN
Distance to forest edge (m)	Edges of forest may be more susceptible to invasion than interiors.	Derived from 2010 Upper Mississippi River Restoration (UMRR) Program Landcover Data: Mississippi & Illinois Rivers. https://doi.org/10.5066/F77942QN
Distance to invaded wet meadow (m)	Wet meadows invaded with reed canarygrass may provide a source of seed/root fragments.	2010 *Phalaris arundinacea* (Reed canarygrass) mapped locations within Pools 2–13 of the Upper Mississippi River System. https://doi.org/10.5066/F7JW8C4X
Forest cover class (%)	Less forest cover may increase light availability in the ground layer, promoting reed canarygrass establishment and growth.	2010 Upper Mississippi River Restoration (UMRR) Program Landcover Data: Mississippi & Illinois Rivers. https://doi.org/10.5066/F77942QN
Forest height class (m)	More mature (greater height) forests may be more resilient to invasion.	2010 Upper Mississippi River Restoration (UMRR) Program Landcover Data: Mississippi & Illinois Rivers. https://doi.org/10.5066/F77942QN
2010 forest community (type)	Different forest communities may be more or less susceptible to reed canarygrass invasion.	2010 Upper Mississippi River Restoration (UMRR) Program Landcover Data: Mississippi & Illinois Rivers. https://doi.org/10.5066/F77942QN
1890s landcover (type)	Land use legacy effects could influence the invasion of reed canarygrass.	1890's Mississippi River Commission Surveys Land Cover Data: Mississippi River. https://doi.org/10.5066/P9NDNA8Y

Note: 1890s landcover dataset was based on a landcover survey conducted in the late 1880s and early 1900s to represent potential land use legacies; the floodplain inundation attribute rasters were 40-year summaries of inundation patterns from 1972 to 2011 included to characterize the long-term hydrologic regime; and all remaining predictors were derived from landcover classifications from 2010 aerial imagery. For comparison, reed canarygrass presence/absence observations were made between 2008 and 2019. Reasoning for inclusion is based upon relationships from the literature, which were outlined in the introduction.

(Hijmans, 2022) in R. Although we did not intend to project our models into new regions or times, we did have some areas of the floodplain with sparse coverage of the training observations, so we performed a Multivariate Environmental Similarity Surface test using the modEvA package (Barbosa et al., 2013) in R, which indicated conditions in 97.6% of our study area were captured in the training dataset. Because highly correlated predictor variables can cause issues with interpretation, we checked for correlations and found none of the environmental predictors were highly correlated (absolute Pearson correlation coefficient <0.27; Appendix S1: Figure S1).

To set up and run our classification models, we used the tidy modeling framework (Kuhn & Silge, 2022) using

the "tidymodels" (Kuhn & Wickham, 2020) meta-package in R. We employed three tree-based classification algorithms using functions from the "parsnip" package (Kuhn & Vaughan, 2022) (Table 2). We used tree-based machine learning algorithms because of their ability to handle complex data structures that are often present in ecological datasets. For 1890s landcover, infrequent ($n < 100$ plots) landcover types were combined into an "other" category, retaining forest, grassland, forbs and shrubs, and agriculture. Our boosted trees algorithm (i.e., XGBoost) could not handle categorical variables, so the landcover types were converted to dummy variables (i.e., each category was assigned to a new column with zero if that category was not present and one if the location fell into that category) for that algorithm. The boosted trees and random forest algorithms were tuned with 1000 trees and the BART algorithm tuned with 200 trees. Each algorithm had a unique set of tuning parameters listed in Table 2. We used the "workflowsets" package (Kuhn, 2022) in R to combine all the model workflows and tune the model parameters. For each algorithm, we selected the best performing tuning parameter set using area under the receiver operating characteristic curve (AUC). AUC values range from 0.5 to 1, with 0.5 being no better than random discrimination and 1 representing perfect discrimination.

Using the full dataset of presence/absence and the predictor variables, we performed a test for spatial autocorrelation in our outcome variable using the joint-count statistic and found our dataset had significant spatial autocorrelation ($p = 0.01$). A significant result indicates a need to account for spatial autocorrelation in estimates of performance (Dormann et al., 2007). We addressed spatial autocorrelation in our dataset using spatial 10-fold cross-validation, where samples used for training and evaluation came from different locations (blocks) for each fold. The appropriate block size (4254 m) to address spatial autocorrelation in our dataset was determined by the "blockCV" package (Valavi et al., 2019) and the "spatialsample" package (Silge & Mahoney, 2022) was used to create the spatial 10-fold cross-validation at the specified block size (Appendix S1: Figure S2). Although spatial cross-validation is now a preferred method when autocorrelation is present within a dataset, we do also present the performance metrics of the commonly used random cross-validation approach to compare it with the influence of spatial cross-validation on performance metrics. We did not use the models created with random cross-validation to make predictions outside of the training dataset.

Additionally, we tested the use of Euclidean distance fields as spatial predictors to address concerns about spatial autocorrelation and improve each model's ability to predict RCG habitat suitability in locations that were not used to train the models. The Euclidean distance fields consisted of seven spatial predictors generated from a raster that covered the extent of the study area, each with cell values that included longitude coordinate, latitude coordinate, distance from upper left corner, distance from upper right corner, distance from lower left corner, distance from lower right corner, and distance from the center of the raster. For each distance field, each cell within the raster has a value of the Euclidean distance from the cell at the location specified (e.g., upper left corner), resulting in a circular pattern of increasing distance that radiates from the specified cell. We evaluated the performance of these three approaches for comparison (i.e., random 10-fold cross-validation, spatial 10-fold cross-validation, and spatial 10-fold cross-validation with Euclidean distance fields).

Because different algorithms learn from the training dataset in different ways, combining their predictions in an ensemble can sometimes help improve prediction accuracy. To increase performance and accuracy of our predictions, we employed a model-ensembling approach without weighting by calculating the mean prediction probability from the model outputs. Predictions were only made for raster cells that did not have any missing values for one or more of the raster layers, which resulted in maps that only cover forested floodplain areas that are not perennially covered with water and were not missing values in any of the predictors at that location.

For presence/absence classification, we used a threshold that minimizes the distance between ROC

TABLE 2 Classification algorithms.

Algorithm	Engine	Tuned parameters
Random forest	ranger	min_n (Minimal Node Size), mtry (No. Randomly Selected Predictors)
Boosted trees	xgboost	learn_rate (Learning Rate), loss_reduction (Minimum Loss Reduction), min_n, mtry, sample_size (Proportion Observations Sampled), tree_depth (Tree Depth)
BART (Bayesian additive regression trees)	dbarts	prior_outcome_range (Prior for Outcome Range), prior_terminal_node_coef (Terminal Node Prior Coefficient), prior_terminal_node_expo (Terminal Node Prior Exponent)

Note: The names of the engine and tuning parameters are those used in the "parsnip" package (Kuhn & Vaughan, 2022) in R (R Core Team, 2020).

(receiver operating characteristic) curve and the upper left corner of the plot (0,1) calculated using the "PresenceAbsence" package (Freeman & Moisen, 2008) in R for each of the models individually including the ensemble. The thresholds were used to calculate accuracy (the number of correct predictions divided by the total number of predictions) and the true skill statistic (TSS) for each model. TSS is calculated from the sum of sensitivity (true positive rate) and specificity (true negative rate) minus 1, and ranges from −1 to +1, where 1 indicates perfect agreement and 0 or less indicates performance is no better than random (Allouche et al., 2006).

The "DALEXtra" package (Maksymiuk et al., 2020) in R was used to evaluate the explanatory power of individual predictors in the models. Variable importance scores for the predictor variables were calculated for each model measured by one minus AUC loss after permutations. The variable importance scores represent the drop in performance when each predictor is randomized (permuted), with greater values indicating greater importance. Accumulated local effects plots that depict the conditional effect of each predictor on the prediction outcome were created for each variable, and all the models were plotted together for comparison.

RESULTS

Performance evaluation

Performance measured by accuracy, AUC, and TSS varied among models and cross-validation approaches (Figure 2). Performance was lower for all models when spatial block-cross-validation was used (Spatial CV), but inclusion of Euclidean distance fields (Spatial CV + EDF) improved performance. Performance did not differ greatly among algorithms when using either of the spatial block-cross-validation approaches. For all following results, only the spatial blocking with Euclidean distance fields approach is reported because it was the best performing method of the two approaches (Spatial CV + EDF and Spatial CV) that accounted for the spatial autocorrelation in our dataset, and thus provided more

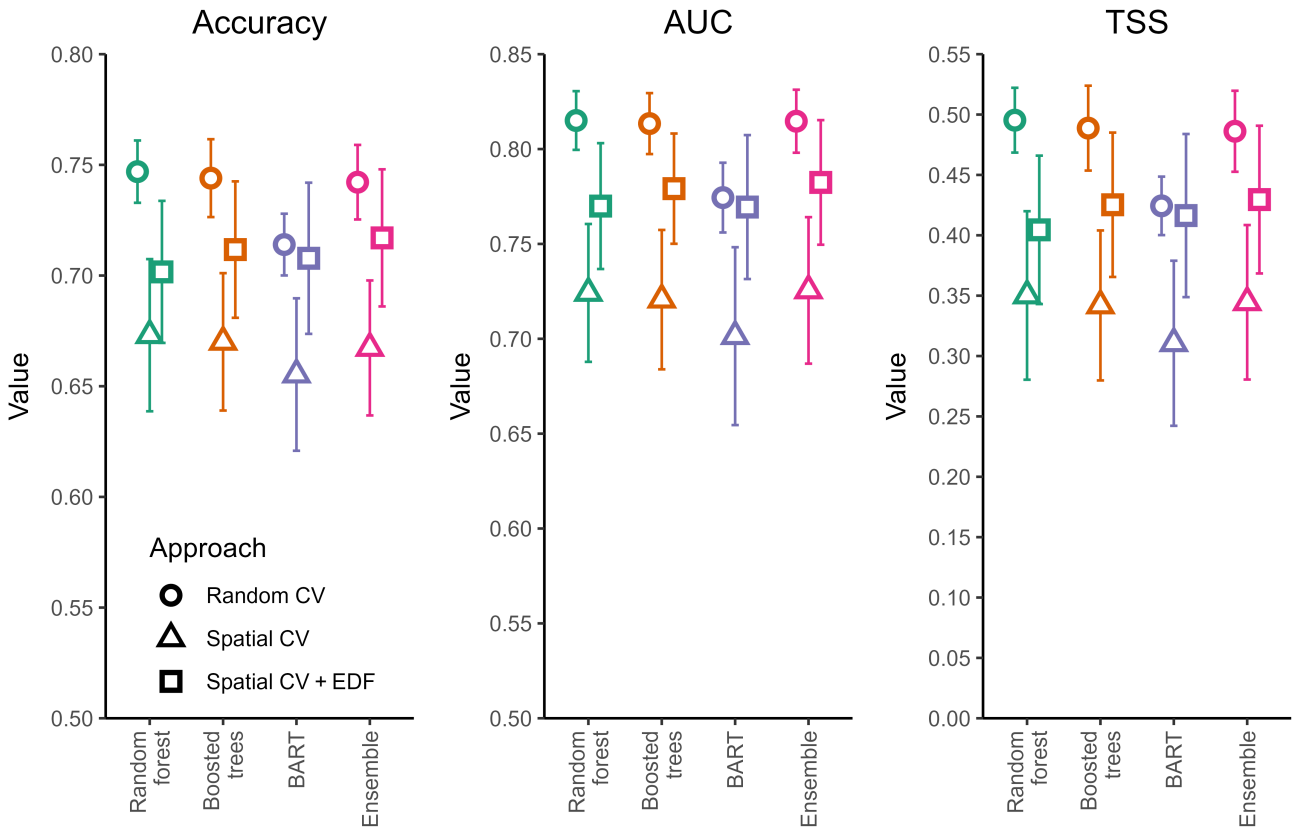


FIGURE 2 Performance metrics for each model and the ensemble for accuracy, area under the receiver operating characteristic curve (AUC), and the true skill statistic (TSS). Models include boosted trees, random forest, Bayesian additive regression trees (BART), and the ensemble of all three models. Results are presented for each approach: Random 10-fold cross-validation (Random CV), spatial 10-fold cross-validation (Spatial CV), and spatial 10-fold cross-validation with Euclidean distance fields (Spatial CV + EDF). Error bars around the mean of performance metrics represent the 90% confidence bounds across the cross-validation folds.

realistic estimates when predicting new locations within the floodplain.

Extent of RCG predicted in forest understories

The distribution of mean and standard deviation (SD) of prediction probabilities among the three models varied by navigation pool (Figure 3). Navigation Pools 3, 6, and 10 had some of the lowest densities of prediction probabilities (mean probability <0.22), and Navigation Pools 4, 5, 8, and 9 had some of the highest (mean probability >0.45). Pool 3 had greater agreement among the five models indicated by lower SDs (mean SD = 0.044). Prediction probabilities were classified as absent and present using the following thresholds: random forest = 0.47, boosted trees = 0.42, BART = 0.46, and ensemble = 0.46. Overall, 34% of forest understory raster cells in the UMR were predicted as having RCG present, with differences among navigation pools. Navigation Pools 3, 6, and 10 were predicted to have the lowest percentage of forest understory raster cells having habitat suited for RCG invasion (<10%), while Navigation Pools 5, 8, and 9 were predicted to have the highest (>66%).

Mapped predictions show the spatial distribution of multimodel mean prediction probabilities, multimodel SD prediction probabilities, and agreement of presence and absence predictions among the three models (see, e.g., Navigation Pool 8 in Figure 4). Ensemble prediction maps for each of the pools are shown in Appendix S1: Figures S4–S15. The prediction rasters for each model are publicly accessible in a data release (Delaney et al., 2023).

Model interpretation

Variable importance for each of the predictors showed differences among algorithms, due to the unique approach each algorithm uses to learn patterns in the data, along with some similarities among algorithms (Figure 5). Random forest and boosted trees showed the greatest similarity in variable importance. Flood depth, flood duration, flood frequency, distance to contiguous floodplain, distance to forest edge, distance to invaded wet meadow, and forest cover class tended to be of greater

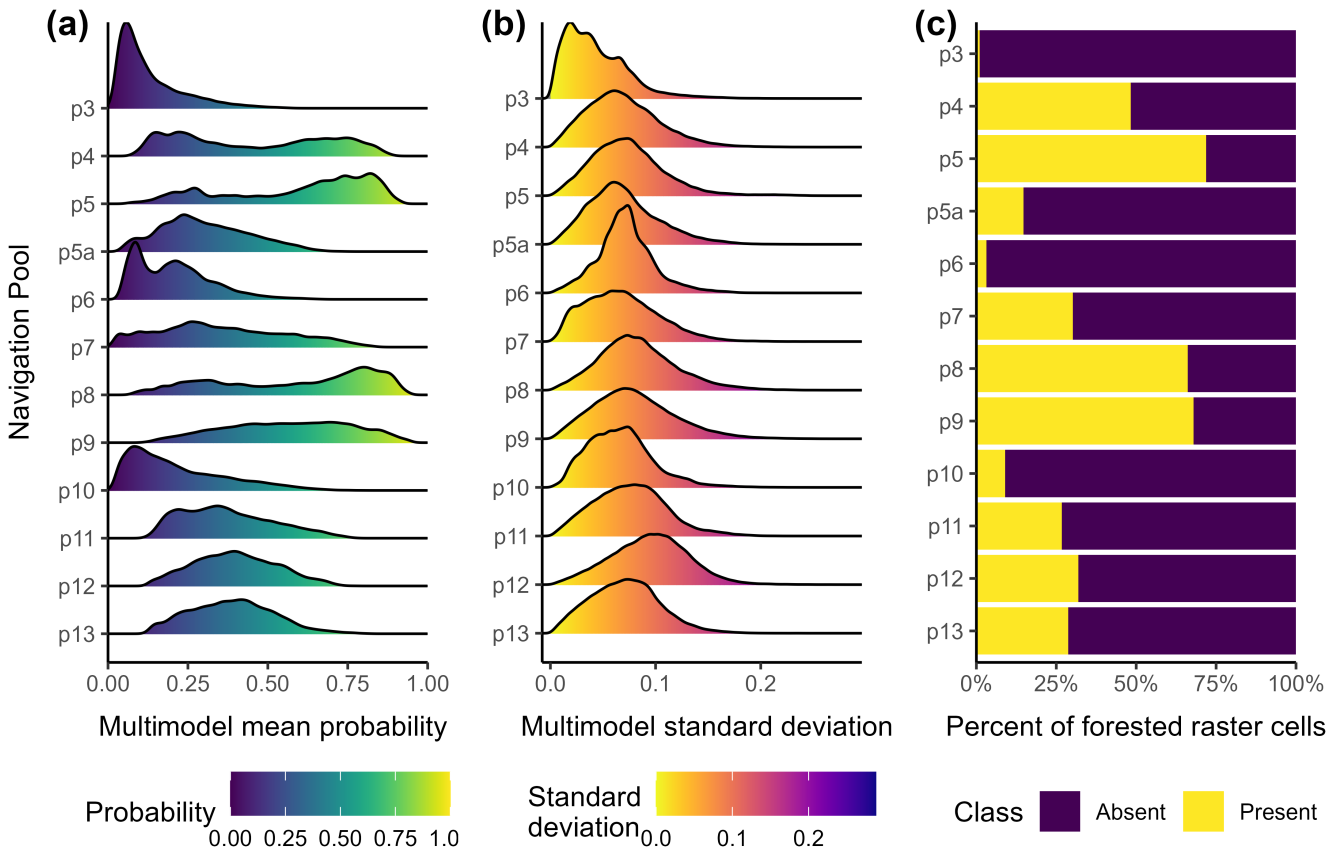


FIGURE 3 (a) Ensemble mean probabilities of reed canarygrass presence, (b) stardard deviation (SD) of mean probability, and (c) predicted classification by navigation pool. Ensemble prediction probabilities ranged from 0.012 to 0.96 and SDs ranged from 0.000038 to 0.29.

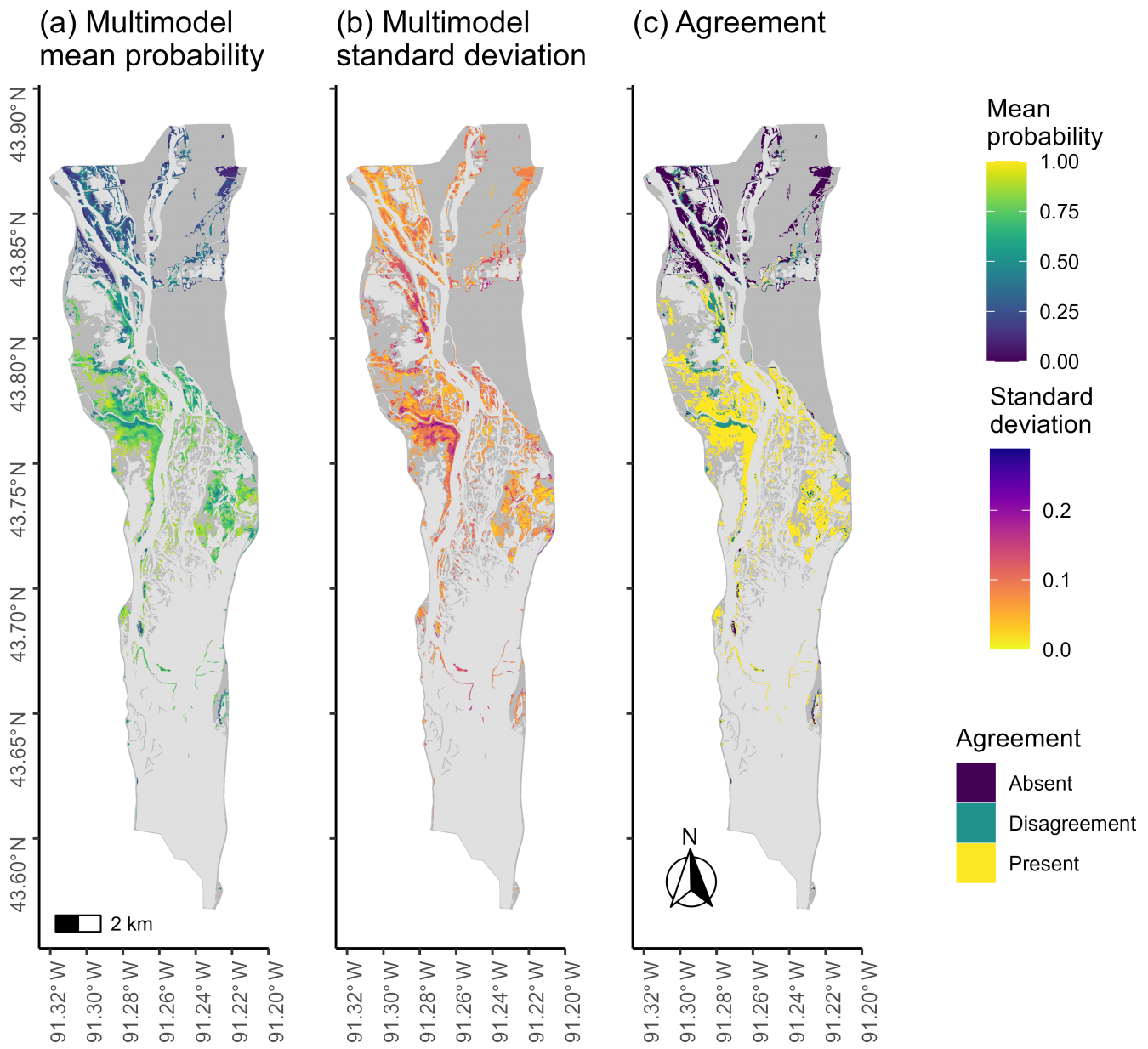


FIGURE 4 Example maps of Navigation Pool 8. Mapped areas include all forested areas within the floodplain, with areas of other landcover types in dark gray and water areas in light gray. Panels show (a) the multimodel mean probability of reed canarygrass presence, (b) the multimodel standard deviation, and (c) areas where all models agree on presence, absence, or where there is disagreement among the three models. Ensemble prediction maps for each of the pools are shown in Appendix S1: Figures S4–S15.

importance, and 1890s landcover, 2010 forest community, and forest height were of lesser importance. Generally, the environmental predictors were more important than the Euclidean distance fields with the exception of the BART algorithm where the distance from the upper right corner, distance from the center, and distance from the lower right corner were among the most important predictors.

Accumulated local effects plots showed similar trends among the algorithms for each of the predictors (Figure 6). The probability of RCG occurrence decreased with increasing canopy cover, forest height, distance to invaded wet meadows, and flood depth. Distance to forest edge and distance to contiguous floodplain showed a u-shaped relationship where average prediction decreased as distance increased, but then began to increase. Flood frequency and flood duration showed more complex relationships. The probability of RCG prediction was lower if the site was classified as forest or agricultural in 1890s for all models, and probability was greater if the site was classified as forbs/brush, grassland, or other in 1890s. As for 2010 forest community type, mean probability was lower for upland and lowland forest categories.

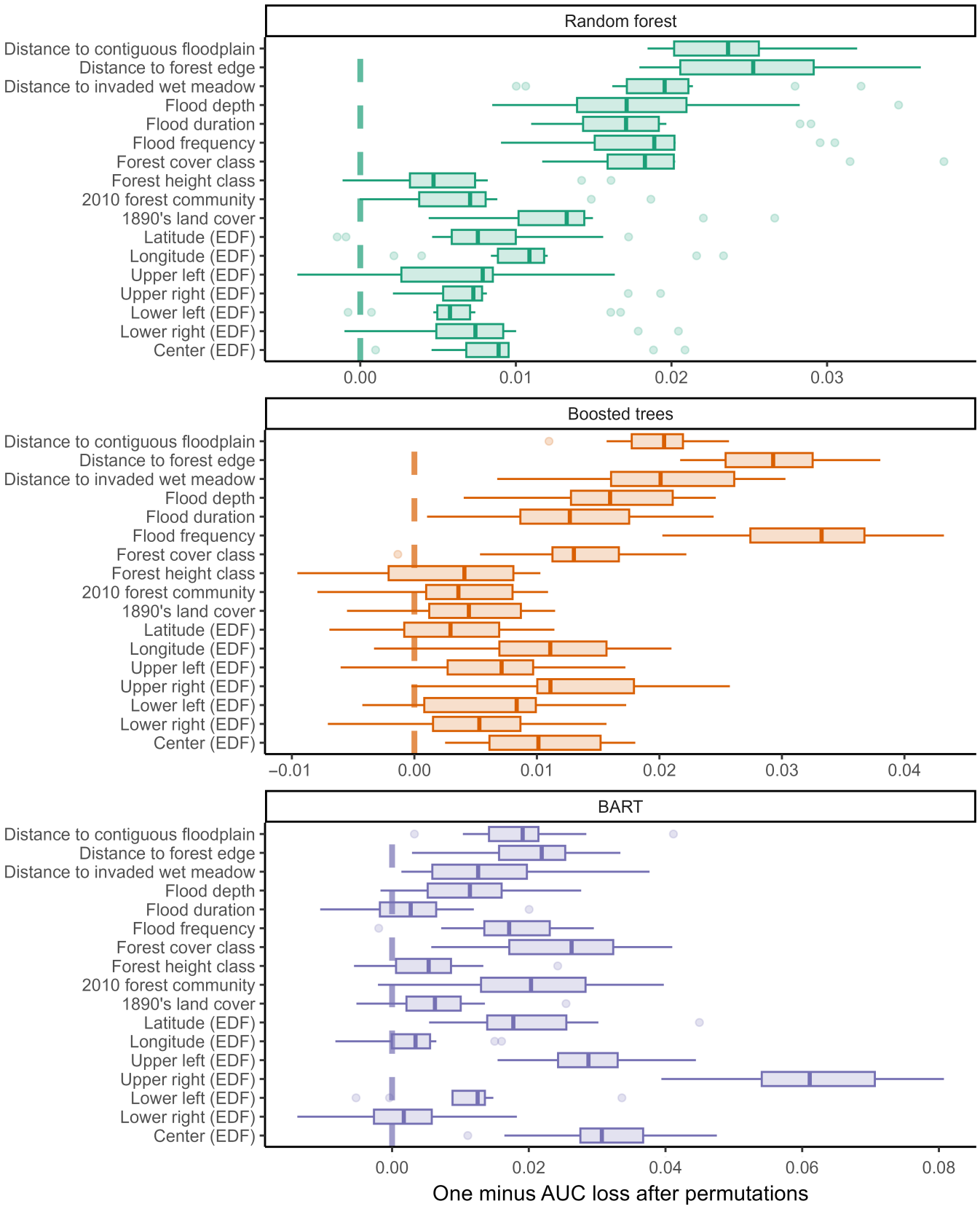


FIGURE 5 Variable importance plots for each model measured by one minus area under the receiver operating characteristic curve (AUC) loss after 10 permutations (i.e., the drop in performance when each predictor is randomized). Within the boxplots, the middle vertical line depicts the median, the edges of the box are the first and third quantiles, the whiskers extend up to 1.5 times the inter-quartile range, and outlying points are depicted individually. The vertical dashed lines show the point of zero loss.

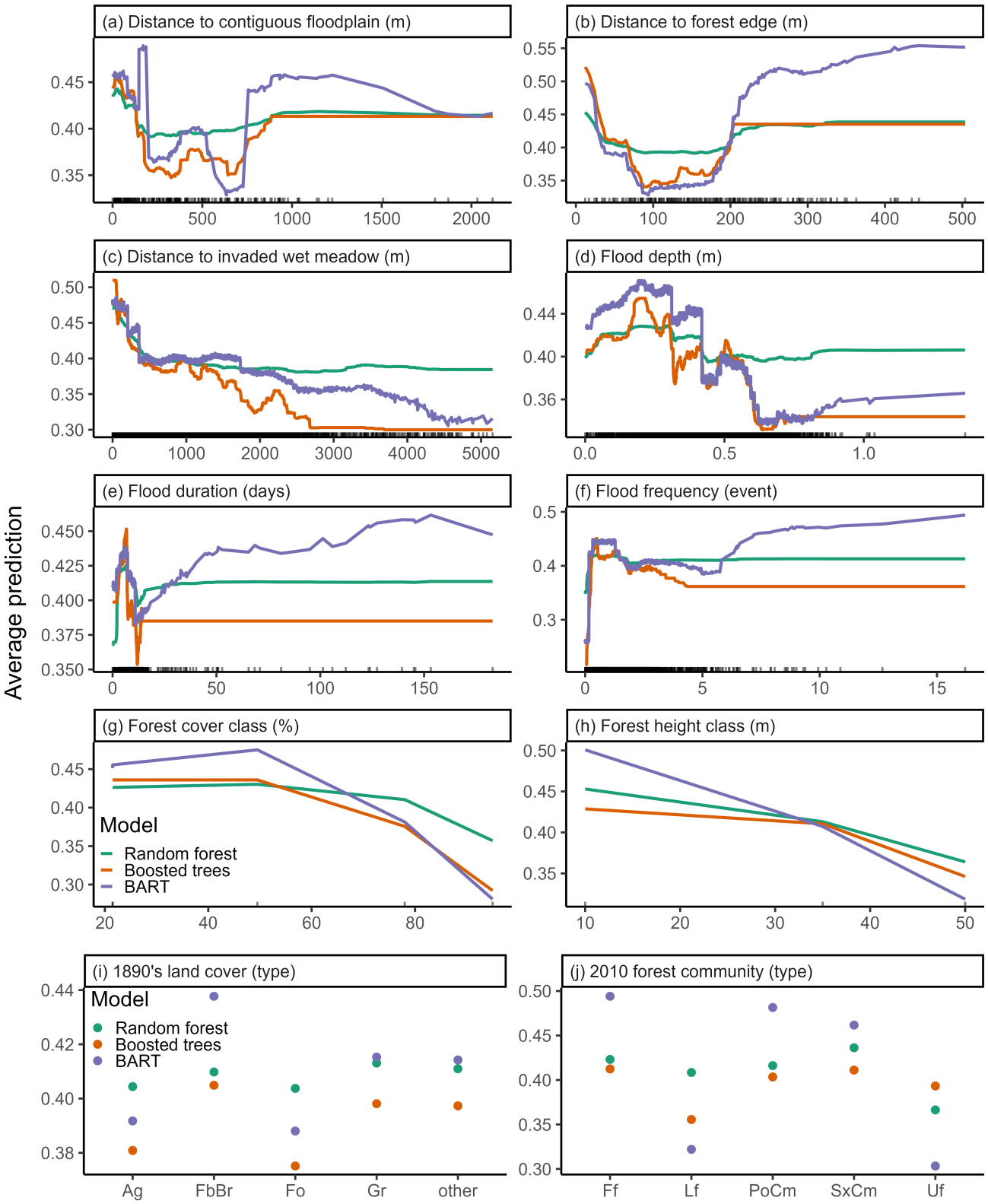


FIGURE 6 Accumulated local effects plots for each of the predictor variables. Models are plotted together to aid in the comparison of the general relationships. Lines are depicted for continuous variables and points for categorical variables. Lines and points are colored by model. Units are indicated in the title of each panel. Rug plots depict the frequency of observations for each continuous variable (a–h). There are no rug plots for the categorical variables (i, j). Areas with fewer observations may indicate lower confidence in average prediction values. To further aid in interpretation, histograms for each environmental predictor are shown in Appendix S1: Figure S3. Categories for 1890s landcover include agriculture (Ag), forbs/brush (FbBr), forest (Fo), grassland (Gr), and other. Categories for 2010 landcover include floodplain forest (Ff), lowland forest (Lf), *Populus* community (PoCm), *Salix* community (SxCm), and upland forest (Uf).

DISCUSSION

Using an ensemble habitat suitability modeling approach, we were able to estimate the habitat suitability of RCG in forest understories from Navigation Pools 3–13. Our performance was similar to Kettenring et al. (2019) who used species distribution models for RCG across riparian areas of the Columbia and Missouri River Basins over a much larger area with greater environmental variability than our study. Performance metrics can vary widely across species and regions (Valavi et al., 2022). The users of outputs from habitat suitability modeling should consider whether the performance is adequate for each specific application. For instance, a model that gets RCG predictions right around 70% of the time as ours did might be useful for prioritizing areas to study the consequences of forest loss or broad-scale estimates of potential invasion across the study area, but may not be suitable to plan budgets for eradication efforts prior to field validation.

Performance evaluations often overlook spatial autocorrelation, which can lead to overestimation of performance (Miller, 2012; Ploton et al., 2020). Indeed, we found that performance was lower when spatial cross-validation was used to account for spatial autocorrelation, as we expected. The inclusion of spatial predictors (Euclidean distance fields) improved the performance of our models when using spatial cross-validation. Behrens et al. (2018) first proposed Euclidean distance fields in machine learning (EDM) and showed improved performance in two soil mapping case studies. To our knowledge, ours is the first attempt to utilize this new method for predicting habitat suitability.

Information on the historical prevalence of RCG within the UMR is limited. One study showed RCG increased substantially between 1975 and 1995 in Navigation Pool 8 (Tyser et al., 2001). De Jager et al. (2017) used aerial imagery from 2010 to quantify wet meadows and wet meadow shrub areas dominated (i.e., >80% cover) by RCG. That study found Navigation Pools 3, 4, 7, 8, and 9 had the highest proportion (>0.5) of invaded wet meadow/shrub areas (i.e., areas outside of the forest understories). We used the invaded wet meadow/shrub classifications from that study as one of our predictors in our model by quantifying the upstream distance to the nearest invaded wet meadow/shrub. In our study, Navigation Pools 4, 5, 8, and 9 had the greatest proportion of forest understories predicted to have habitat suited for RCG invasion. A partial explanation for the disparity between the forested and unforested areas between these two studies for Navigation Pools 3, 5, and 7 could be related to differences in propagule pressure from unforested RCG patches from upstream pools.

Distance to invaded wet meadows was among the most important predictors in our study, and we saw a decline in probability as the distance from wet meadows dominated by RCG increased. Conversely, Jakubowski et al. (2010) found pastures intentionally planted with RCG were not an important predictor in their model estimating the proportion of watersheds invaded with RCG in southern Wisconsin, USA. White-tailed deer can facilitate and maintain dominance of RCG by herbivory of woody and herbaceous native vegetation (Cogger et al., 2014; De Jager et al., 2013; Kellogg & Bridgham, 2004). We used distance to contiguous floodplain as a proxy for deer herbivory pressure, with the assumption that the farther away from the contiguous floodplain an island is the lower deer herbivory, and thus less RCG it may have. We found that the prediction probabilities did decrease with increasing distance from the contiguous floodplain. However, RCG occurrence probabilities increased at greater distances, which may be a result of fewer observations at greater distances.

Once RCG propagules reach a forest understory site, low light availability may limit establishment and spread (Adams et al., 2011; Lindig-Cisneros & Zedler, 2002; Perry & Galatowitsch, 2004). Forest edges may have greater light availability, and clones of RCG are able to expand into heavily shaded areas by providing nutrients through clonal subsidy (Maurer & Zedler, 2002). In our study, occurrence probabilities decreased in areas with greater tree cover and canopy height, and areas near the edge of the forest were more likely to be predicted to have RCG. Canopy height was of lesser importance relative to canopy cover and distance to forest edge across all models. Interestingly, probabilities related to distance to forest edge showed a strong nonlinear response, where probabilities decreased as distance from forest edge increased, as we expected, but increased at greater distances. This unexpected relationship could indicate that samples taken from forest interiors were not representative due to fewer observations at distances greater than 175 m (3.4% of observations), or that there was some aspect of forest interiors that make them more suitable for RCG that was not included in our suite of environmental predictors.

The three flood metrics (depth, duration, and frequency) were also important predictors for the random forest and boosted trees models. Flooding may facilitate the transport of seeds, roots, and shoots from invaded wet meadows to new areas. We did observe the probability of RCG presence increases with increasing flood frequency, but only to a point. Inundation depth can be an important factor influencing RCG (Coops et al., 1996; Miller & Zedler, 2003), and Jenkins et al. (2008) observed reductions in RCG cover in response to managed flooding greater than 0.85-m depth. In our study, all models showed the probability of RCG presence declining at

depths (median of the maximum depth of submergence across all flood events during the growing season from 1972 to 2011) >0.3 m. The duration of inundation can also influence RCG (Miller & Zedler, 2003; Rice & Pinkerton, 1993). For all models, prediction probabilities increased with increasing flood duration, but then decreased, which indicates that greater flood duration may benefit RCG but inhibit it at greater durations. Charles et al. (2023) observed decreases in RCG dominance in vegetation plots with prolonged inundation, with RCG being replaced with species more tolerant of inundation.

The importance of the inundation metrics in predicting RCG presence in forest understories has relevance to future climate and land use change. Over the past few decades, changes in streamflow resulting from changes in precipitation and land use in the region have been observed (Neri et al., 2019; Slater & Villarini, 2017), and annual discharges on the Upper Mississippi River System have increased since 1940 (Van Appledorn, 2022), resulting in increasingly wetter floodplain conditions. If the trend of increasing discharge and flood duration continues as regional climate change analyses indicate they may (Byun et al., 2019; Neri et al., 2020), the distribution of RCG in floodplain forest may change or rapid conversion from forest- to RCG-dominated wet meadows may occur due to flood-related tree mortality. Additionally, several of our predictors were derived from landcover types, which could change in the future because the inundation regime is important for structuring floodplain forest communities (De Jager et al., 2016; De Jager, Van Appledorn, et al., 2019).

The information generated from this study could aid resource managers in determining which management approach to take at different locations in the floodplain (Bouska et al., 2022). Given the effort required to effectively control RCG (Lavergne & Molofsky, 2004), it is impractical that resource managers resist RCG invasion everywhere in the UMR. Managers could identify areas with high probability of RCG invasion to evaluate and target for control and restoration efforts such as treatment with herbicide, mowing, or planting of native herbaceous species (Miller-Adamany et al., 2019) where appropriate. Alternatively, managers may decide to accept RCG in areas with high probability of RCG invasion, particularly where RCG is already dominant, and focus resources elsewhere. Areas with moderate RCG invasion probabilities may be targets to resist RCG by enhancing forest resources. De Jager et al. (2024) further discuss potential management options for maintaining forest cover where RCG is most likely to be a driver of forest loss under two potential future hydrologic scenarios. Based upon our results and previous research discussed above, management efforts that maintain forest cover and longevity by managing species and age structure, control RCG in nearby wet meadows, and reduce deer herbivory of forest understory communities could help to limit invasion by RCG. Locations vulnerable to transitioning to RCG-dominant wet meadows could be identified by overlaying the maps produced in this study with areas of recent forest loss (due to windfall, extended inundation, emerald ash borer [*Agrilus planipennis*], etc.) to determine the likelihood that newly unforested areas contain RCG, which may limit the ability of the forest to regenerate. Additionally, maps identifying sites predicted to have RCG present or absent across the UMR could be used to identify locations for future field studies or adaptive management efforts.

As with many habitat suitability modeling studies, we were limited by the availability of spatially continuous datasets that could serve as ecologically relevant predictors. While we were able to gather datasets that represent important ecological influences on the distribution of RCG with reasonable temporal correspondence with the period of sampling, predictions could potentially be improved with additional predictors, more refined data layers, and data sets that are more closely aligned with the period of sampling. The inundation metrics used in this study are summarized over a 40-year period and are reflective of long-term patterns that structure ecological communities (De Jager et al., 2018; Van Appledorn et al., 2021). Shorter term (e.g., within the previous 1–2 growing seasons) extreme events (e.g., long period of inundation) could affect RCG either positively or negatively, thus influencing its distribution on the landscape. Additionally, an assumption of a habitat suitability model is that the species is at equilibrium with its environment (Guisan & Thuiller, 2005). This assumption may be violated for invasive species if they have not yet had time to spread to all suitable habitats. The UMR is well within the historic range, as well as what may be considered the invaded range, of RCG. Determination of whether RCG has reached equilibrium within the UMR floodplain would require information on the distribution from several time periods (Foster et al., 2022), which is currently unavailable. Price et al. (2018) did document increases in RCG in wetlands across Illinois from 1997 to 2015, however.

The maps we produced in this study are temporally static, but future research could identify areas where RCG is projected to increase, or possibly decrease over time. Doing so would require greater temporal representation of true presence/absence data than we had for our study. While the predictors used in this study were derived from remotely sensed spatial datasets sometimes with coarse estimates or categories, more detailed

field-collected measurements (e.g., forest composition, basal area, understory cover) may reveal additional information on RCG invasion, management, and restoration of floodplain forests. Although not examined in our models, it is well-recognized that an additional component of predicting the vulnerability of sites to forest loss and conversion to wet meadows is the existing species and age structure of forest areas with and without RCG. Stands that lack a mid-story and understory and contain RCG have a higher likelihood of converting to wet meadows. Finally, climate change may alter hydrologic dynamics (Byun et al., 2019) which would affect floodplain forest ecosystems, and future research could focus on how the distribution of RCG could change due to alterations in inundation regimes and changes in floodplain forest community and structure.

CONCLUSION

Understanding the spatial distribution of invasive species is important for effective management. We were able to characterize RCG distribution in floodplain forests throughout the UMR using an ensemble modeling approach. By combining outputs from different algorithms and accounting for spatial autocorrelation, we provided conservative estimates of the likelihood of RCG presence and the levels of model agreement. Relationships of each of the predictors with the average prediction probability aligned with our reasoning for selecting them based on published literature. The importance of inundation metrics in our models indicates that potential changes to the inundation regime in the UMR could alter the distribution of RCG in forest understories in the future. Information produced in this study may be useful for managers to prioritize areas for conservation and restoration of floodplain forests across the UMR and provide information on mechanisms influential to RCG invasion more broadly.

ACKNOWLEDGMENTS

Funding for this research was provided by the U.S. Army Corps of Engineers' (USACE) Upper Mississippi River Restoration Program. Reed canarygrass presence/absence observations were collected during Phase II Forest Inventory with USACE Operations funds for Environmental Stewardship and through the Upper Mississippi River Restoration Program. We would like to thank three anonymous reviewers for their helpful reviews that improved this manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The reed canarygrass presence/absence locations along with the prediction rasters generated for this analysis (Delaney et al., 2023) are available from the USGS ScienceBase-Catalog: https://doi.org/10.5066/P9KBFBHW. The floodplain inundation rasters (Van Appledorn et al., 2018) are available from the USGS ScienceBase-Catalog: https://doi.org/10.5066/F7VD6XRT. The 2010 landcover data (U.S. Army Corps of Engineers' Upper Mississippi River Restoration [UMRR] Program Long Term Resource Monitoring [LTRM] element, 2017) are available from the USGS ScienceBase-Catalog: https://doi.org/10.5066/F77942QN. The reed canarygrass dominance in wet meadows dataset (Hoy et al., 2017) is available from the USGS ScienceBase-Catalog: https://doi.org/10.5066/F7JW8C4X. The 1890s land cover data (U.S. Army Corps of Engineers' Upper Mississippi River Restoration [UMRR] Program Long Term Resource Monitoring [LTRM] Element, 2000) are available from the USGS ScienceBase-Catalog: https://doi.org/10.5066/P9NDNA8Y.

ORCID

John T. Delaney https://orcid.org/0000-0003-1038-0265
M. Van Appledorn https://orcid.org/0000-0002-8029-0014
N. R. De Jager https://orcid.org/0000-0002-6649-4125
K. L. Bouska https://orcid.org/0000-0002-4115-2313
J. J. Rohweder https://orcid.org/0000-0001-5131-9773

REFERENCES

Adams, C. R., and S. M. Galatowitsch. 2005. "*Phalaris arundinacea* (Reed Canary Grass): Rapid Growth and Growth Pattern in Conditions Approximating Newly Restored Wetlands." *Ecoscience* 12: 569–573.

Adams, C. R., P. J. Kauth, and J. W. Sorenson. 2011. "Assessing Competition between Reed Canary Grass (*Phalaris arundinacea*) and Swamp White Oak (*Quercus bicolor*)." *Ecological Restoration* 29: 332–38.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. "Assessing the Accuracy of Species Distribution Models: Prevalence, Kappa and the True Skill Statistic (TSS): Assessing the Accuracy of Distribution Models." *Journal of Applied Ecology* 43: 1223–32.

Anderson, N. O., A. G. Smith, A. K. Noyszewski, E. Ito, D. Dalbotten, and H. Pellerin. 2021. "Management and Control Issues for Native, Invasive Species (Reed Canarygrass): Evaluating Philosophical, Management, and Legislative Issues." *HortTechnology* 31: 354–366.

Annen, C. A., E. M. Kirsch, and R. W. Tyser. 2008. "Reed Canarygrass Invasions Alter Succession Patterns and May

Reduce Habitat Quality in Wet Meadows." *Ecological Restoration* 26: 190–93.

Barbosa, A. M., R. Real, A. Román Muñoz, and J. A. Brown. 2013. "New Measures for Assessing Model Equilibrium and Prediction Mismatch in Species Distribution Models." *Diversity and Distributions* 19: 1333–38.

Bart, D., M. Simon, Q. Carpenter, and S. Graham. 2011. "Historical Land Use and Plant-Community Variability in a Wisconsin Calcareous Fen." *Rhodora* 113: 160–186.

Behrens, T., K. Schmidt, R. A. Viscarra Rossel, P. Gries, T. Scholten, and R. A. MacMillan. 2018. "Spatial Modelling with Euclidean Distance Fields and Machine Learning." *European Journal of Soil Science* 69: 757–770.

Bernthal, T. W., and K. G. Willis. 2004. *Using LANDSAT 7 Imagery to Map Invasive Reed Canary Grass (Phalaris arundinacea): A Landscape Level Wetland Monitoring Methodology, Final Report to the U.S. Environmental Protection Agency Region V*. Madison, Wisconsin: Wisconsin Department of Natural Resources.

Bouska, K. L., N. R. De Jager, and J. N. Houser. 2022. "Resisting-Accepting-Directing: Ecosystem Management Guided by an Ecological Resilience Assessment." *Environmental Management* 70: 381–400.

Bouska, K. L., J. N. Houser, N. R. De Jager, D. C. Drake, S. F. Collins, D. K. Gibson-Reinemer, and M. A. Thomsen. 2020. "Conceptualizing Alternate Regimes in a Large Floodplain-River Ecosystem: Water Clarity, Invasive Fish, and Floodplain Vegetation." *Journal of Environmental Management* 264: 110516.

Bradley, B. A. 2014. "Remote Detection of Invasive Plants: A Review of Spectral, Textural and Phenological Approaches." *Biological Invasions* 16: 1411–25.

Byun, K., C.-M. Chiu, and A. F. Hamlet. 2019. "Effects of 21st Century Climate Change on Seasonal Flow Regimes and Hydrologic Extremes over the Midwest and Great Lakes Region of the US." *Science of the Total Environment* 650: 1261–77.

Charles, B., M. Chase, G. Pociask, R. Bhattarai, and J. W. Matthews. 2023. "Restored Wetlands Are Greatly Influenced by Hydrology and Non-Native Plant Invasion." *Wetlands Ecology and Management* 31: 129–142.

Chen, X., Z. Deng, Y. Xie, F. Li, Z. Hou, X. Li, and Y. F. Li. 2014. "Effects of Sediment Burial Disturbance on the Vegetative Propagation of *Phalaris arundinacea* with Different Shoot Statuses." *Aquatic Ecology* 48: 409–416.

Cogger, B. J., N. R. De Jager, M. Thomsen, and C. R. Adams. 2014. "Winter Browse Selection by White-Tailed Deer and Implications for Bottomland Forest Restoration in the Upper Mississippi River Valley, USA." *Natural Areas Journal* 34: 144–153.

Coops, H., F. W. B. van den Brink, and G. van der Velde. 1996. "Growth and Morphological Responses of Four Helophyte Species in an Experimental Water-Depth Gradient." *Aquatic Botany* 54: 11–24.

Coops, H., and G. van der Velde. 1995. "Seed Dispersal, Germination and Seedling Growth of Six Helophyte Species in Relation to Water-Level Zonation." *Freshwater Biology* 34: 13–20.

De Jager, N. R., B. J. Cogger, and M. A. Thomsen. 2013. "Interactive Effects of Flooding and Deer (*Odocoileus virginianus*) Browsing on Floodplain Forest Recruitment." *Forest Ecology and Management* 303: 11–19.

De Jager, N. R., J. T. Rogala, J. J. Rohweder, M. Van Appledorn, K. L. Bouska, J. N. Houser, and K. J. Jankowski. 2018. *Indicators of Ecosystem Structure and Function for the Upper Mississippi River System*. Reston, Virginia: U.S. Geological Survey. Open-File Report. 134 pp.

De Jager, N. R., J. J. Rohweder, and E. E. Hoy. 2017. *Mapping areas invaded by Phalaris arundinacea in Navigation Pools 2–13 of the Upper Mississippi River*. La Crosse, WI: U.S. Army Corps of Engineers' Upper Mississippi River Program, U.S. Geological Survey.

De Jager, N. R., J. J. Rohweder, M. Van Appledorn, E. Hlavacek, and A. Meier. 2024. "Identifying Conditions where Reed Canarygrass (*Phalaris arundinacea*) Functions as a Driver of Forest Loss in the Upper Mississippi River Floodplain under Different Hydrological Scenarios." *Wetlands Ecology and Management* 32: 153–170.

De Jager, N. R., J. J. Rohweder, Y. Yin, and E. Hoy. 2016. "The Upper Mississippi River Floodscape: Spatial Patterns of Flood Inundation and Associated Plant Community Distributions." *Applied Vegetation Science* 19: 164–172.

De Jager, N. R., W. Swanson, D. L. Hernández, J. Reich, R. Erickson, and E. A. Strauss. 2019. "Effects of Flood Inundation, Invasion by *Phalaris arundinacea*, and Nitrogen Enrichment on Extracellular Enzyme Activity in an Upper Mississippi River Floodplain Forest." *Wetlands Ecology and Management* 27: 443–454.

De Jager, N. R., W. Swanson, E. A. Strauss, M. Thomsen, and Y. Yin. 2015. "Flood Pulse Effects on Nitrification in a Floodplain Forest Impacted by Herbivory, Invasion, and Restoration." *Wetlands Ecology and Management* 23: 1067–81.

De Jager, N. R., M. Van Appledorn, T. J. Fox, J. J. Rohweder, L. J. Guyon, A. R. Meier, R. J. Cosgriff, and B. J. Vandermyde. 2019. "Spatially Explicit Modelling of Floodplain Forest Succession: Interactions among Flood Inundation, Forest Successional Processes, and Other Disturbances in the Upper Mississippi River Floodplain, USA." *Ecological Modelling* 405: 15–32.

Delaney, J. T., M. Van Appledorn, and J. J. Rohweder. 2023. *Estimates of Habitat Suitability of Reed Canarygrass (Phalaris arundinacea) in Upper Mississippi River Floodplain Forest Understories*. Reston, Virginia: U.S. Geological Survey data release. https://doi.org/10.5066/P9KBFBHW.

Dieck, J. J., J. Ruhser, E. Hoy, and L. R. Robinson. 2015. *General Classification Handbook for Floodplain Vegetation in Large River Systems (ver. 2.0, November 2015)*. Reston, Virginia: U.S. Geological Survey. Techniques and Methods, Book 2, Chap. A1, 51 pp.

Dormann, C. F., J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, et al. 2007. "Methods to Account for Spatial Autocorrelation in the Analysis of Species Distributional Data: A Review." *Ecography* 30: 609–628.

Elith, J., and J. R. Leathwick. 2009. "Species Distribution Models: Ecological Explanation and Prediction across Space and Time." *Annual Review of Ecology, Evolution, and Systematics* 40: 677–697.

Elliott, S. H., D. D. Tullos, and C. Walter. 2019. "Physical Modeling of the Feedbacks between a Patch of Flexible Reed

Canarygrass (*Phalaris arundinacea*), Wake Hydraulics, and Downstream Deposition." *Environmental Fluid Mechanics* 19: 255–277.

ESRI (Environmental Systems Research Institute). 2020. *ArcGIS Desktop Version 10.8.1*. Redlands, CA: ESRI (Environmental Systems Research Institute).

Foster, S. L., H. M. Kharouba, and T. W. Smith. 2022. "Testing the Assumption of Environmental Equilibrium in an Invasive Plant Species over a 130 Year History." *Ecography* 2022: e12933.

Frans, C., E. Istanbulluoglu, V. Mishra, F. Munoz-Arriola, and D. P. Lettenmaier. 2013. "Are Climatic or Land Cover Changes the Dominant Cause of Runoff Trends in the Upper Mississippi River Basin?" *Geophysical Research Letters* 40: 1104–10.

Freeman, E. A., and G. Moisen. 2008. "PresenceAbsence: An R Package for Presence Absence Analysis." *Journal of Statistical Software* 23: 1–31.

Galatowitsch, S. M., N. O. Anderson, and P. D. Ascher. 1999. "Invasiveness in Wetland Plants in Temperate North America." *Wetlands* 19: 733–755.

Gazis, I.-Z., and J. Greinert. 2021. "Importance of Spatial Autocorrelation in Machine Learning Modeling of Polymetallic Nodules, Model Uncertainty and Transferability at Local Scale." *Minerals* 11: 1172.

Gobeyn, S., A. M. Mouton, A. F. Cord, A. Kaim, M. Volk, and P. L. M. Goethals. 2019. "Evolutionary Algorithms for Species Distribution Modelling: A Review in the Context of Machine Learning." *Ecological Modelling* 392: 179–195.

Green, E. K., and S. M. Galatowitsch. 2002. "Effects of *Phalaris arundinacea* and Nitrate-N Addition on the Establishment of Wetland Plant Communities." *Journal of Applied Ecology* 39: 134–144.

Grenouillet, G., L. Buisson, N. Casajus, and S. Lek. 2011. "Ensemble Modelling of Species Distribution: The Effects of Geographical and Environmental Ranges." *Ecography* 34: 9–17.

Guisan, A., and W. Thuiller. 2005. "Predicting Species Distribution: Offering More than Simple Habitat Models." *Ecology Letters* 8: 993–1009.

Guisan, A., and N. E. Zimmermann. 2000. "Predictive Habitat Distribution Models in Ecology." *Ecological Modelling* 135: 147–186.

Hao, T., J. Elith, G. Guillera-Arroita, and J. J. Lahoz-Monfort. 2019. "A Review of Evidence about Use and Performance of Species Distribution Modelling Ensembles like BIOMOD." *Diversity and Distributions* 25: 839–852.

Hao, T., J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Arroita. 2020. "Testing whether Ensemble Modelling Is Advantageous for Maximizing Predictive Performance of Species Distribution Models." *Ecography* 43: 549–558.

Hijmans, R. 2022. "raster: Geographic Data Analysis and Modeling." R Package Version 3.5-15. https://CRAN.R-project.org/package=raster.

Hillhouse, H. L., S. J. Tunnell, and J. Stubbendieck. 2010. "Spring Grazing Impacts on the Vegetation of Reed Canarygrass–Invaded Wetlands." *Rangeland Ecology & Management* 63: 581–87.

Hoy, E. E., N. R. DeJager, and J. Rohweder. 2017. *2010 Phalaris arundinacea (Reed canarygrass) Mapped Locations Within Pools 2-13 of the Upper Mississippi River System*. Reston, Virginia: U.S. Geological Survey data release. https://doi.org/10.5066/F7JW8C4X.

Hulme, P. E. 2009. "Trade, Transport and Trouble: Managing Invasive Species Pathways in an Era of Globalization." *Journal of Applied Ecology* 46: 10–18.

Jakubowski, A. R., M. D. Casler, and R. D. Jackson. 2010. "Landscape Context Predicts Reed Canarygrass Invasion: Implications for Management." *Wetlands* 30: 685–692.

Januchowski-Hartley, S. R., V. M. Adams, and V. Hermoso. 2018. "The Need for Spatially Explicit Quantification of Benefits in Invasive-Species Management." *Conservation Biology* 32: 287–293.

Jenkins, N. J., J. A. Yeakley, and E. M. Stewart. 2008. "First-Year Responses to Managed Flooding of Lower Columbia River Bottomland Vegetation Dominated by *Phalaris arundinacea*." *Wetlands* 28: 1018–27.

Joshi, C., J. de Leeuw, and I. C. van Duren. 2004. "Remote Sensing and GIS Applications for Mapping and Spatial Modelling of Invasive Species." In *Proceedings of the XXth Congress of the International Society for Photogrammetry and Remote Sensing, Istanbul*, 669–677.

Kaky, E., V. Nolan, A. Alatawi, and F. Gilbert. 2020. "A Comparison between Ensemble and MaxEnt Species Distribution Modelling Approaches for Conservation: A Case Study with Egyptian Medicinal Plants." *Ecological Informatics* 60: 101150.

Kellogg, C. H., and S. D. Bridgham. 2004. "Disturbance, Herbivory, and Propagule Dispersal Control Dominance of an Invasive Grass." *Biological Invasions* 6: 319–329.

Kercher, S. M., and J. B. Zedler. 2004a. "Multiple Disturbances Accelerate Invasion of Reed Canary Grass (*Phalaris arundinacea* L.) in a Mesocosm Study." *Oecologia* 138: 455–464.

Kercher, S. M., and J. B. Zedler. 2004b. "Flood Tolerance in Wetland Angiosperms: A Comparison of Invasive and Noninvasive Species." *Aquatic Botany* 80: 89–102.

Kettenring, K. M., D. R. Menuz, and K. E. Mock. 2019. "The Nativity and Distribution of the Cryptic Invader *Phalaris arundinacea* (Reed Canarygrass) in Riparian Areas of the Columbia and Missouri River Basins." *Wetlands* 39: 55–66.

Kirsch, E. M., and B. R. Gray. 2017. "Differences in Breeding Bird Assemblages Related to Reed Canary Grass Cover and Forest Structure on the Upper Mississippi River." *Journal of Fish and Wildlife Management* 8: 260–271.

Knutson, M. G., and E. E. Klaas. 1998. "Floodplain Forest Loss and Changes in Forest Community Composition and Structure in the Upper Mississippi River: A Wildlife Habitat at Risk." *Natural Areas Journal* 18: 138–150.

Koenig, W. D. 1999. "Spatial Autocorrelation of Ecological Phenomena." *Trends in Ecology & Evolution* 14: 22–26.

Kreiling, R. M., N. R. De Jager, W. Swanson, E. A. Strauss, and M. Thomsen. 2015. "Effects of Flooding on Ion Exchange Rates in an Upper Mississippi River Floodplain Forest Impacted by Herbivory, Invasion, and Restoration." *Wetlands* 35: 1005–12.

Kuhn, M. 2022. "workflowsets: Create a Collection of "tidymodels" Workflows." R Package Version 0.2.1. https://CRAN.R-project.org/package=workflowsets.

Kuhn, M., and J. Silge. 2022. *Tidy Modeling with R*. Sebastopol, CA: O'Reilly.

Kuhn, M., and D. Vaughan. 2022. "parsnip: A Common API to Modeling and Analysis Functions." https://github.com/tidymodels/parsnip, https://parsnip.tidymodels.org/.

Kuhn, M., and H. Wickham. 2020. "tidymodels: A Collection of Packages for Modeling and Machine Learning using Tidyverse Principles." https://www.tidymodels.org.

Lavergne, S., and J. Molofsky. 2004. "Reed Canary Grass (*Phalaris arundinacea*) as a Biological Model in the Study of Plant Invasions." *Critical Reviews in Plant Sciences* 23: 415–429.

Lavergne, S., and J. Molofsky. 2006. "Control Strategies for the Invasive Reed Canarygrass (*Phalaris arundinacea* L.) in North American Wetlands: The Need for an Integrated Management Plan." *Natural Areas Journal* 26: 208–214.

Li, X., and Y. Wang. 2013. "Applying Various Algorithms for Species Distribution Modelling." *Integrative Zoology* 8: 124–135.

Lindig-Cisneros, R., and J. B. Zedler. 2002. "*Phalaris arundinacea* Seedling Establishment: Effects of Canopy Complexity in Fen, Mesocosm, and Restoration Experiments." *Canadian Journal of Botany* 80: 617–624.

Lindsay, J. B. 2016. "Whitebox GAT: A Case Study in Geomorphometric Analysis." *Computers & Geosciences* 95: 75–84.

LTRM (U.S. Army Corps of Engineers' Upper Mississippi River Restoration [UMRR] Program Long Term Resource Monitoring [LTRM] Element). 2000. *1890 UMRS Land Cover Land Use*. Reston, Virginia: U.S. Geological Survey data release. https://doi.org/10.5066/P9NDNA8Y.

Maksymiuk, S., A. Gosiewska, and P. Biecek. 2020. "Landscape of R Packages for eXplainable Artificial Intelligence." arXiv: 2009.13248.

Marmion, M., M. Parviainen, M. Luoto, R. K. Heikkinen, and W. Thuiller. 2009. "Evaluation of Consensus Methods in Predictive Species Distribution Modelling." *Diversity and Distributions* 15: 59–69.

Matthews, J. W., S. McIntyre, A. L. Peralta, and C. Rodgers. 2020. "Long-Term Assessment of Alternative Strategies for the Restoration of Floodplain Forest in the Presence of an Invasive Grass, *Phalaris arundinacea*." *Wetlands* 40: 655–665.

Maurer, D. A., R. Lindig-Cisneros, K. J. Werner, S. Kercher, R. Miller, and J. B. Zedler. 2003. "The Replacement of Wetland Vegetation by Reed Canarygrass (*Phalaris arundinacea*)." *Ecological Restoration* 21: 116–19.

Maurer, D. A., and J. B. Zedler. 2002. "Differential Invasion of a Wetland Grass Explained by Tests of Nutrients and Light Availability on Establishment and Clonal Growth." *Oecologia* 131: 279–288.

Meier, A. 2017. "MVP-ES Forest Inventory Phase II Protocol." https://ecos.fws.gov/ServCat/DownloadFile/157228?Reference=105804.

Merigliano, M. F., and P. Lesica. 1998. "Native Status of Reed Canarygrass (*Phalaris arundinacea* L.) in the Inland Northwest, USA." *Natural Areas Journal* 18: 223–230.

Miller, J. 2010. "Species Distribution Modeling." *Geography Compass* 4: 490–509.

Miller, J. A. 2012. "Species Distribution Models: Spatial Autocorrelation and Non-Stationarity." *Progress in Physical Geography: Earth and Environment* 36: 681–692.

Miller, R. C., and J. B. Zedler. 2003. "Responses of Native and Invasive Wetland Plants to Hydroperiod and Water Depth." *Plant Ecology* 167: 57–69.

Miller-Adamany, A., D. Baumann, and M. Thomsen. 2019. "Facilitating Natural Succession in a Heavily Invaded Ecosystem." *Forest Ecology and Management* 444: 235–243.

Narumalani, S., D. R. Mishra, R. Wilson, P. Reece, and A. Kohler. 2009. "Detecting and Mapping Four Invasive Species along the Floodplain of North Platte River, Nebraska." *Weed Technology* 23: 99–107.

Neri, A., G. Villarini, and F. Napolitano. 2020. "Statistically-Based Projected Changes in the Frequency of Flood Events across the U.S. Midwest." *Journal of Hydrology* 584: 124314.

Neri, A., G. Villarini, L. J. Slater, and F. Napolitano. 2019. "On the Statistical Attribution of the Frequency of Flood Events across the U.S. Midwest." *Advances in Water Resources* 127: 225–236.

NOAA National Centers for Environmental Information. 2023. "U.S. Climate Normals (v1.0.1)." National Oceanic and Atmospheric Administration. https://www.ncei.noaa.gov/products/landbased-station/us-climate-normals.

Noyszewski, A. K., N. O. Anderson, A. G. Smith, A. Kilian, D. Dalbotten, E. Ito, A. Timm, et al. 2021. "Riparian Populations of Minnesota Reed Canarygrass (*Phalaris arundinacea*) Are Most Likely Native, Based on SNPs (DArTseqLD)." *Wetlands Ecology and Management* 29: 467–494.

Perry, L. G., and S. M. Galatowitsch. 2004. "The Influence of Light Availability on Competition between *Phalaris arundinacea* and a Native Wetland Sedge." *Plant Ecology* 170: 73–81.

Perry, L. G., S. M. Galatowitsch, and C. J. Rosen. 2004. "Competitive Control of Invasive Vegetation: A Native Wetland Sedge Suppresses *Phalaris arundinacea* in Carbon-Enriched Soil." *Journal of Applied Ecology* 41: 151–162.

Ploton, P., F. Mortier, M. Réjou-Méchain, N. Barbier, N. Picard, V. Rossi, C. Dormann, et al. 2020. "Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models." *Nature Communications* 11: 4540.

Price, E. P. F., G. Spyreas, and J. W. Matthews. 2018. "Biotic Homogenization of Regional Wetland Plant Communities within Short Time-Scales in the Presence of an Aggressive Invader." *Journal of Ecology* 106: 1180–90.

Price, E. P. F., G. Spyreas, and J. W. Matthews. 2020. "Biotic Homogenization of Wetland Vegetation in the Conterminous United States Driven by *Phalaris arundinacea* and Anthropogenic Disturbance." *Landscape Ecology* 35: 779–792.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Rice, J. S., and B. W. Pinkerton. 1993. "Reed Canarygrass Survival under Cyclic Inundation." *Journal of Soil and Water Conservation* 48: 132–35.

Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, et al. 2017. "Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure." *Ecography* 40: 913–929.

Romano, S. P. 2010. "Our Current Understanding of the Upper Mississippi River System Floodplain Forest." *Hydrobiologia* 640: 115–124.

Ruhser, J., J. C. Nelson, L. R. Robinson, E. E. Hoy, and J. L. Hanson. 2015. *2010/11 Land Cover/Use Data for the Upper Mississippi River System*. Reston, Virginia: U.S. Geological Survey data release. https://doi.org/10.5066/F77942QN.

Shih, F. Y., and Y.-T. Wu. 2004. "The Efficient Algorithms for Achieving Euclidean Distance Transformation." *IEEE Transactions on Image Processing* 13: 1078–91.

Silge, J., and M. Mahoney. 2022. "spatialsample: Spatial Resampling Infrastructure." https://gihub.com/tidymodels/spatialsample, https://spatialsample.tidymodels.org.

Slater, L., and G. Villarini. 2017. "Evaluating the Drivers of Seasonal Streamflow in the U.S. Midwest." *Water* 9: 695.

Sofaer, H. R., C. S. Jarnevich, I. S. Pearse, R. L. Smyth, S. Auer, G. L. Cook, T. C. Edwards, et al. 2019. "Development and Delivery of Species Distribution Models to Inform Decision-Making." *Bioscience* 69: 544–557.

Spyreas, G., B. W. Wilm, A. E. Plocher, D. M. Ketzner, J. W. Matthews, J. L. Ellis, and E. J. Heske. 2010. "Biological Consequences of Invasion by Reed Canary Grass (*Phalaris arundinacea*)." *Biological Invasions* 12: 1253–67.

Stiles, C. A., B. Bemis, and J. B. Zedler. 2008. "Evaluating Edaphic Conditions Favoring Reed Canary Grass Invasion in a Restored Native Prairie." *Ecological Restoration* 26: 61–70.

Swanson, W., N. R. De Jager, E. Strauss, and M. Thomsen. 2017. "Effects of Flood Inundation and Invasion by *Phalaris arundinacea* on Nitrogen Cycling in an Upper Mississippi River Floodplain Forest." *Ecohydrology* 10: e1877.

Thessen, A. 2016. "Adoption of Machine Learning Techniques in Ecology and Earth Science." *One Ecosystem* 1: e8621.

Thomsen, M., K. Brownell, M. Groshek, and E. Kirsch. 2012. "Control of Reed Canarygrass Promotes Wetland Herb and Tree Seedling Establishment in an Upper Mississippi River Floodplain Forest." *Wetlands* 32: 543–555.

Tyser, R. W., S. J. Rogers, T. W. Owens, and L. R. Robinson. 2001. "Changes in Backwater Plant Communities from 1975 to 1995 in Navigation Pool 8, Upper Mississippi River." *Regulated Rivers: Research & Management* 17: 117–129.

U.S. Army Corps of Engineers-St. Paul District. 2022. *Phase II Forest Inventory Dataset*. St. Paul, MN.: USACE La Crescent Field Office.

U.S. Army Corps of Engineers' Upper Mississippi River Restoration (UMRR) Program Long Term Resource Monitoring (LTRM) Element. 2017. *UMRR 2010 Land Cover Land Use*. Reston, Virginia: U.S. Geological Survey data release. https://doi.org/10.5066/F77942QN.

Valavi, R., J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Arroita. 2019. "blockCV: An R Package for Generating Spatially or Environmentally Separated Folds for *k*-Fold Cross-Validation of Species Distribution Models." *Methods in Ecology and Evolution* 10: 225–232.

Valavi, R., J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Arroita. 2023. "Flexible Species Distribution Modelling Methods Perform Well on Spatially Separated Testing Data." *Global Ecology and Biogeography* 32: 369–383.

Valavi, R., G. Guillera-Arroita, J. J. Lahoz-Monfort, and J. Elith. 2022. "Predictive Performance of Presence-Only Species Distribution Models: A Benchmark Study with Reproducible Code." *Ecological Monographs* 92: e01486.

Van Appledorn, M. 2022. "Chap B Hydrologic Indicators." In *Ecological Status and Trends of the Upper Mississippi and Illinois Rivers*, edited by J. N. Houser, 38–54. Reston, Virginia: U.S. Geological Survey. Open-File Report 2022-1039.

Van Appledorn, M., N. R. De Jager, and J. J. Rohweder. 2021. "Quantifying and Mapping Inundation Regimes within a Large River-Floodplain Ecosystem for Ecological and Management Applications." *River Research and Applications* 37: 241–255.

Van Appledorn, M., N. R. DeJager, and J. Rohweder. 2018. *UMRS Floodplain Inundation Attribute Rasters*. Reston, Virginia: U.S. Geological Survey data release. https://doi.org/10.5066/F7VD6XRT.

van der Wal, R., A.-M. Truscott, I. S. K. Pearce, L. Cole, M. P. Harris, and S. Wanless. 2008. "Multiple Anthropogenic Changes Cause Biodiversity Loss through Plant Invasion." *Global Change Biology* 14: 1428–36.

Walsh, S. J. 2018. "Multi-Scale Remote Sensing of Introduced and Invasive Species: An Overview of Approaches and Perspectives." In *Understanding Invasive Species in the Galapagos Islands*, edited by M. L. de Torres and C. F. Mena, 143–154. Cham: Springer International Publishing.

Weilhoefer, C. L., D. Williams, I. Nguyen, K. Jakstis, and C. Fischer. 2017. "The Effects of Reed Canary Grass (*Phalaris arundinacea* L.) on Wetland Habitat and Arthropod Community Composition in an Urban Freshwater Wetland." *Wetlands Ecology and Management* 25: 159–175.

Werner, K. J., and J. B. Zedler. 2002. "How Sedge Meadow Soils, Microtopography, and Vegetation Respond to Sedimentation." *Wetlands* 22: 451–466.

Wilkins, F. S., and H. D. Hughes. 1932. "Agronomic Trials with Reed Canary Grass." *Agronomy Journal* 24: 18–28.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Delaney, John T., M. Van Appledorn, N. R. De Jager, K. L. Bouska, and J. J. Rohweder. 2024. "Spatial Differences in Predicted *Phalaris Arundinacea* (reed Canarygrass) Occurrence in Floodplain Forest Understories." *Ecosphere* 15(12): e70138. https://doi.org/10.1002/ecs2.70138