

DISSERTATION

IMPROVING THE DEVELOPMENT, EVALUATION, AND INTERPRETATION
OF SPECIES DISTRIBUTION MODELS

Submitted by

Peder Scott Engelstad

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2026

Doctoral Committee:

Advisor: Jody Vogeler

Co-Advisor: Catherine Jarnevich

Jeffrey Morisette

Ruth Hufbauer

Copyright by Peder Scott Engelstad 2026

All Rights Reserved

ABSTRACT

IMPROVING THE DEVELOPMENT, EVALUATION, AND INTERPRETATION OF SPECIES DISTRIBUTION MODELS

Species distribution models (SDMs) encompass a variety of statistical approaches to model the potential geographic distributions of species based on functional relationships between the values of spatial environmental covariates at a set of focal observations (Elith & Leathwick, 2009; Guisan & Zimmermann, 2000). Under the assumption that a species has had sufficient time to reach equilibrium with its environment, SDMs represent an expression of a species' fundamental environmental niche as formulated by Hutchinson (1957). However, that assumption rarely holds, and thus the distributions characterized by SDMs only represent a partial expression of a species' realized niche. Fully capturing the realized niche means also understanding and incorporating the barriers to dispersal that species face (Barve et al., 2011; Soberón & Peterson, 2005). Implementing movement into SDM workflow has continued to be a difficult task. As a result, models are unconstrained and predict to every available location. For SDMs of invasive species, unconstrained predictions can lead to unrealistic estimates of the amount of suitable habitat due to overestimation (can an organism arrive at the location?) or underestimation (is the species still undergoing range expansion and/or niche infilling?) (Atwater & Barney, 2021; Guisan et al., 2014).

Given the current impacts of invasive species to biodiversity, economic activity, human health, and ecosystem services (Simberloff et al., 2013) and the potential for these impacts to worsen under future climate conditions (Bertelsmeier et al., 2013; Finch et al., 2021), there is

urgency to understand the processes driving invasions while producing ecological research that can also complement and enhance ongoing prevention, mitigation, and eradication efforts. Mapped predictions of potential establishment of invasive species from SDMs offer land managers information that can be used for spatially prioritization and to allocate limited resources within affected or at-risk landscapes (Reaser et al., 2020). To ensure the efficacy and applicability of SDMs as predictive tools across geographies and taxa, ongoing improvements are needed to advance to the capabilities and methodology of SDMs.

The research presented in this dissertation describes how advances in the development, evaluation, and interpretation of SDMs have the potential to improve both ecological understanding and management outcomes. Specifically, the following chapters address three current challenges in species distribution modeling: integrating dispersal constraints, independently validating performance, and understanding the drivers of model predictions. These avenues of investigation demonstrate how SDMs can be enhanced both as tools to explain or identify the potential processes underlying the spatial patterns of distributions or as resources to applied research where predictions serve practical land management and decision-making needs.

Overview

Chapter one addresses the limitations of unconstrained SDM predictions by augmenting mapped outputs with spatial representations of where species could realistically move within the landscape. By combining current and future habitat suitability with invasion pathway-level dispersal information for over 100 freshwater invaders, this research represents a more realistic and pragmatic depiction of spatial invasion risk. The chapter is also a novel, large-scale application

that provides a template for future regional invasion risk assessments focused on supporting management activities like early detection and rapid response.

The utility of SDMs for invasive species management is underpinned by the confidence that they are built with the strongest and most relevant measures of model performance. In chapter two, I focus on enhancing existing spatial sampling techniques to address the challenge of model evaluation. Due to bias in species occurrence data from non-random aggregations of observations, traditional random data-splitting can produce misleading and erroneously optimistic performance metrics. Using the inherent gridded geometry of the hashtag (#) shape, I introduce a novel implementation of spatial blocking that provides a robust and ecologically informed methodology that can evaluate spatial ecological models like SDMs for spatial transferability, a key consideration when predicting suitable habitat for invasive species that continue to spread.

Even a carefully constructed SDM can be subject to the "black box" problem (Molnar et al., 2020) wherein the behavior and decision making behind model predictions becomes more opaque as methods increase in complexity. In chapter three, I explore the ability of Shapley Additive exPlanations to improve model interpretation and look beyond the predicted values alone. I use a focal SDM to better understand how and why a model generates individual predictions, given the spatial environmental covariates present in the model and the biological context of a focal species on the landscape.

By presenting methodological improvements related to the broad-scale integration of invasion pathways, rigorous spatial testing, and out-of-sample importance of model drivers, this work argues that species distribution modeling can move forward as both a tool of explanation and prediction while balancing predictive capacity, spatial transferability, and utility for applied ecological research.

ACKNOWLEDGEMENTS

My sincere gratitude to all those who contributed to this work and supported me through to its completion. First and foremost, I thank my advisors, Catherine Jarnevich and Jody Vogeler, for their trust, patience, expertise, and encouragement over the years. Every conversation with them calmed my anxieties and served as a reminder of their ever-present confidence in me. Catherine and Jody's mentorship cannot be overstated, and I am forever grateful for their guidance through this experience. I am also grateful to my committee members, Frenchy Morisette and Ruth Hufbauer, for their suggestions, thoughtful comments, and support in formal meetings or informal science discussions over a beer (or two). Your insights greatly improved both the quality of this work and helped push me to improve myself as a scientist and researcher. To my "shadow committee" members and longtime colleagues (Ian Pearse, Helen Sofaer, Janet Prevey, Wilfred Previant, Magda Garbowski, Annette Evans, Eve Beaury, Deah Lieurance, Nick Young, Brian Woodward, Sarah Carroll, Aida Cuni-Sanchez, Cini Brown, Richie Eriskson, Paul Evangelista, and Wes Daniel), I cannot thank you enough for all the on/offline discussions, technical assistance, and moral support you've given over the years. Your mentorship, professional assistance, reality checks, life advice, laughter, and consolation have been essential and exceedingly generous given your own busy lives and careers.

To my chapter co-authors that have yet to be thanked (Arden Williams, Cameron Reimer, Demetra Williams, Grace Henderson, Ian Pfingsten, Keana Shadwell, Linnea Fraser, Logan Stratton, Rich Inman, Shelby LeClare), I feel very lucky to have you as collaborators. Thank you for your ongoing participation and your thoughtful contributions. I am also grateful for the technical and program support I received from the Natural Resource Ecology Laboratory, Graduate

Degree Program in Ecology, and the USGS Fort Collins Science Center. Thank you in particular to Colin Talbert, Jamal Spencer, Jennifer Sieracki, Jeremy Livingston, Matt Nielsen, Pavel Hoyeu, Terri Hogan, and Tim Mancuso. Many of the logistics of this research would have been impossible without your support and by funding from US Army Corps of Engineers, the USGS Science Analytics and Synthesis Program, the USGS Invasive Species Program, and the USGS Northeast Climate Adaptation Science Center.

To my dear friends in Colorado, Minnesota, and beyond, thank you for giving me many opportunities to focus on and take breaks from this research. Canoe trips, record nights, dungeons and dragons, and everything in between helped me balance my mental state and constantly refueled my emotional reserves. I thank my parents (Scott and Laura Engelstad) and sister (Linnea Engelstad) for happily embracing my lifelong love of science and absorbing the occasionally manic energy driving my scientific curiosity and enthusiasm for new knowledge. Finally, to my wife, Andrea Engelstad, I love you and I know none of this would have happened without you. There were an uncountable number of moments where you lifted me out of the darkness, told me things were going to be ok, and made me feel like my best self again. You are the best partner I could hope to have and I am so glad to have shared all the ups and downs alongside you.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	v
CHAPTER 1— COMBINING HABITAT SUITABILITY AND SPATIAL SECONDARY PATHWAYS TO ASSESS FRESHWATER INVASION RISK IN THE NORTHEAST UNITED STATES	1
1 Abstract.....	1
2 Introduction.....	2
3 Methods	4
3.1 Focal area and species	4
3.2 Occurrence data	5
3.3 Species distribution modeling.....	6
3.3.1 Background data.....	6
3.3.2 Model building	7
3.4 Invasion pathways	9
3.5 Composite risk score outputs & summaries	10
4 Results.....	11
4.1 SDMs – future suitability predictions.....	11
4.2 Maps of invasion pathways	12
4.3 Composite risk scores	16
5 Discussion.....	16
5.1 Species distribution models	20
5.2 Secondary pathways	20
5.3 Management implications of composite risk.....	22
5.4 Riskiest species.....	23
6 Future Work.....	25
REFERENCES	27

CHAPTER 2 — #MODELEVALUATION: GENERATING SPATIAL TESTING AND CROSS-VALIDATION SPLITS FOR ECOLOGICAL MODELS USING HASHTAG GEOMETRY	34
1 Abstract.....	34
2 Introduction.....	35
3 Methods	39
3.1 Virtual Species.....	39
3.2 Generating and splitting occurrence data	39
3.3 Model building and tuning.....	42
3.4 Model evaluation	43
4 Results.....	43
4.1 Data splitting.....	43

4.2 Model evaluation	44
5 Discussion.....	46
REFERENCES	50
CHAPTER 3 — OPENING THE BLACK BOX: USING LOCAL FEATURE IMPORTANCE TO INTERPRET, DIAGNOSE, AND COMMUNICATE THE PREDICTIONS OF ECOLOGICAL MODELS.....	54
1 Abstract.....	54
2 Introduction.....	55
3 Methods	57
3.1 Focal species.....	57
3.2 Sampling units	58
3.3 Focal model	58
3.4 Shapley calculations	60
3.5 Feature importance comparisons	63
4 Results.....	64
4.1 Local importance values	64
4.2 Comparisons of model and local importance	68
5 Discussion.....	71
5.1 Model to local comparisons.....	72
5.2 Implications for modelers	73
5.3 Implications for land managers	74
6 Future Work.....	75
REFERENCES	77
SYNTHESIS	80
1 Overview.....	80
2 Beyond prediction engines	81
3 Dual aims of applied ecological research	82
4 Invasive species as study taxa.....	83
5 Future directions	83
5.1 INHABIT and open science.....	83
5.2 Invasion science in an age of uncertainty	84
6 Conclusion	85
REFERENCES FOR SYNTHESIS	87
APPENDIX 1- SUPPLEMENTAL INFORMATION FOR CHAPTER 1	89
APPENDIX 2 - SUPPLEMENTAL INFORMATION FOR CHAPTER 2	97

CHAPTER 1— COMBINING HABITAT
SUITABILITY AND SPATIAL SECONDARY PATHWAYS TO
ASSESS FRESHWATER INVASION RISK IN THE NORTHEAST UNITED STATES

1 Abstract

A comprehensive understanding of the potential spread and establishment of freshwater species is essential to informing early detection and rapid response management strategies for lentic and lotic systems. The effective implementation of early detection strategies hinges on optimizing the use of limited resources and prioritizing prevention in high-risk areas. Key inputs for early detection efforts can be derived from spatial information on pathways (how a species could arrive at a location) and habitat suitability (the potential of a species to establish at a location). But pathways or suitability in isolation only provide part of the story of invasion risk and are often limited to small geographic areas, narrow taxonomic scopes, or overly simplified niche estimation. Moreover, climate change, evolving land use patterns, and altered streamflows further complicate the prediction of future invasion dynamics. To address these challenges, we introduce a novel approach that combines current and future habitat suitability from species distribution models with pathways information to generate invasion risk scores for 103 freshwater aquatic invasive species (fishes, plants, and macroinvertebrates). Our approach produces comprehensive, regionally mapped coverage of invasion risk in freshwater systems. Using the Northeast United States as a study area, we summarize geographic trends of risk for more than 8,700 subwatersheds to inform management activities. Our results can be used to help identify geographic hotspots of invasion risk, generate ranked risk watchlists by species, and predict change under future environmental conditions to help anticipate and prevent freshwater species establishment.

2 Introduction

Freshwater aquatic invasive species cause broad ecological and economic impacts and incur annual costs of hundreds of billions of dollars globally (Cuthbert et al., 2021). As a result, there is an ongoing need to perform systematic, large-scale, and spatially explicit assessments of freshwater invasion risk. Ideally, these assessments include ecologically plausible predictions of establishment potential with the taxonomic breadth necessary to generate meaningful data for prevention or early detection and rapid response management activities (e.g., species watchlists). But currently available assessment methods typically require tradeoffs between geographic scope, environmental granularity, and computational efficiency. For example, high-resolution, species-specific predictions require additional modeling expertise and can be more data intensive as spatial grain increases, reducing the overall number of target species or size of assessed area (e.g., Roberts et al., 2024). Simplified spatial methods like climate matching are computationally efficient, facilitating wider taxonomic and geographic coverage (Daniel et al., 2025; Howeth et al., 2025; Kingsbury et al., 2025; Sanders et al., 2023) but are often spatially coarse (less detailed in smaller geographic contexts) and lack more complete habitat information, including edaphic or hydrological characteristics. Regardless of method, a primary goal of spatially explicit risk assessments is to estimate an ecologically plausible distributions that capture the biotic, abiotic, and movement factors relevant to ecological niches and establishment potential of focal species. Assessments meeting this goal have products that are usable by the resource managers charged with addressing the establishment and impact of freshwater invasive species.

To assess the risk of invasive species establishment in support of prevention and early detection strategies, a common approach is to use correlative species distribution models (SDMs) (Reaser et al., 2020), which quantify species-environment relationships to predict habitat suitability, an estimate of establishment and survival potential for a species at a given location

(Elith & Leathwick, 2009; Guisan & Thuiller, 2005). But even thoughtfully developed, species-specific SDMs (or simpler distribution methods like climate matching) can overestimate establishment potential because they do not inherently limit predictions to locations that a species can access through intentional or unintentional transport. As a result, SDMs alone may not be refined enough to support the kind of targeted invasive species management needed for early detection efforts and other invasive species management activities. However, by integrating potential species movement with mapped predictions of suitability, modelers may be able to create more realistic representations of where the accessibility of suitable habitat is influenced by the physical barriers to species dispersal (Barve et al., 2011; Soberón & Peterson, 2005).

One way to characterize movement potential for invasive species is through the use of information on secondary pathways, or the modes of movement (natural or human mediated e.g., hitchhiking on watercraft) for a species within a biogeographic boundary after introduction (Hulme et al., 2008; Pergl et al., 2020). Previous large-scale risk assessments used secondary pathways as scoring components (Daniel et al., 2025; Howeth et al., 2025; Lieurance et al., 2023) but in non-spatial contexts. While a limited number of pathways are inherently spatial (e.g., dispersal kernels; Cook et al., 2019), there is a lack of readily available, continuously mapped geospatial data for secondary pathways. Augmenting mapped suitability with a robust set of spatially explicit secondary pathways could lead to more realistic predictions of where species both arrive and establish, constraining SDM outputs to the information most relevant to freshwater invasive species management. However, the coupling of SDM predictions with pathways data remains uncommon and underexplored in large scale, spatially explicit invasion risk assessments of freshwater species.

In this study, we present a novel, spatially explicit framework for assessing freshwater invasion risk in the Northeast United States. (1) Our approach begins by developing habitat-specific SDMs to estimate current suitability for 103 invasive species, including plants, macroinvertebrates, and fishes, in freshwater systems (e.g., lakes, rivers, streams). (2) We then apply the SDMs under plausible future environmental scenarios to identify areas where invasion risk may shift over time. (3) To improve ecological realism, we develop new, spatially explicit representations of secondary pathways, capturing both natural and human-mediated dispersal routes, and (4) we combine suitability and pathways components into composite risk scores that reflect both establishment potential and movement constraints. (5) We demonstrate how to synthesize the data into management-ready products, including maps and prioritization watchlists that support freshwater invasive species prevention and early detection strategies in local and regional contexts. (6) Finally, we highlight the scalability and transferability of our approach, outlining opportunities for refinement and application to other geographies or taxa.

3 Methods

3.1 Focal area and species

Hundreds of freshwater fishes, macroinvertebrates, and plants are considered invasive within the lower 48 US states (Simpson et al., 2022) and many are expected to expand their ranges northward in response to changing environmental conditions (Evans et al., 2024). However, not all invasive species are anticipated to expand their ranges into or within the Northeast United States under current and future environmental conditions. To ensure the relevance and utility of our modeling and invasion risk assessment efforts to resource managers in the geographic purview of the U.S. Geological Survey's Northeast Climate Adaptation Science Center network (Fig. 1),

we engaged directly with local, state, and federal resource managers and other regional stakeholders. In a series of webinars, we asked participants to identify freshwater invasive species of greatest concern, both currently and under projected future environmental conditions. Suggestions included species native to portions of the region and non-native species from outside of North America and already present in the U.S. But all solicited species for this study have known impacts in freshwater systems in the conterminous U.S. and are generally not considered widely established across the Northeast United States. Our resulting list included an initial 154 species (40 fishes, 18 macroinvertebrates, and 96 plants).

3.2 Occurrence data

We obtained publicly available georeferenced observations in the conterminous United States of our focal plant, macroinvertebrate, and fish species from online biodiversity aggregators and agency databases (Appendix 1). We retained records from 1980-2024, designated as “observation” or “specimen only” collection types (GBIF only) and with low (≤ 30 m) or unknown coordinate uncertainties. We removed potentially erroneous records with *CoordinateCleaner* (Zizka et al., 2019) in R (R Core Team, 2025), including records located near capitals, state/national centroids, biodiversity institutions (e.g., research centers, universities, herbaria, museums, zoos, and botanical gardens), or records with impossible or improbable coordinate combinations. To avoid pseudoreplication, we spatially thinned occurrence records to a single record per lentic polygon or rasterized lotic polyline (e.g., stream reach; see below for further discussion on modeled habitat types). We did not model species with fewer than 50 thinned records (Jarnevich et al., 2026).

3.3 Species distribution modeling

For each species, we adapted a presence-background SDM methodology for terrestrial habitats from Jarnevich et al. (2024) to create separate models for lentic and lotic habitats based on the spatially and categorically delineated data types found in the National Hydrography Dataset (NHD; United States Geological Survey, 2025). For a detailed list of NHD codes used, see Table S1 of Henderson et al (2026a). Separate models allowed us to address the unique features of each habitat type by including distinct structural characteristics (e.g., flow; sedimentation) and known species-habitat associations. We applied models of current suitability within each habitat type to future conditions (see 3.3.2 *Model building*). We built and tuned all models using the *Software for Assisted Habitat Modeling* (Jarnevich et al., 2026; Morisette et al., 2013) and performed all other data preparation, evaluation, and analyses in R (R Core Team, 2025).

3.3.1 Background data

To generate background data, we used a ‘target guild’ approach (Phillips et al., 2009) to account for spatial sampling biases found in observation datasets. We identified all non-native freshwater fish, macroinvertebrate, and plant species found in the United States based on their status in the Nonindigenous Aquatic Species database (U.S. Geological Survey, 2024; nas.er.usgs.gov) and compiled occurrence records (1980-present) for those species by taxa group. We filtered background data with the same methods as our occurrence data (one point per lake or stream reach). For each species, we randomly selected up to 10,000 records from our filtered background data (by taxa group) that fell within polygons representing a 99% isopleth of a kernel density estimate (KDE) built from the focal species’ occurrence records (Jarnevich et al., 2024).

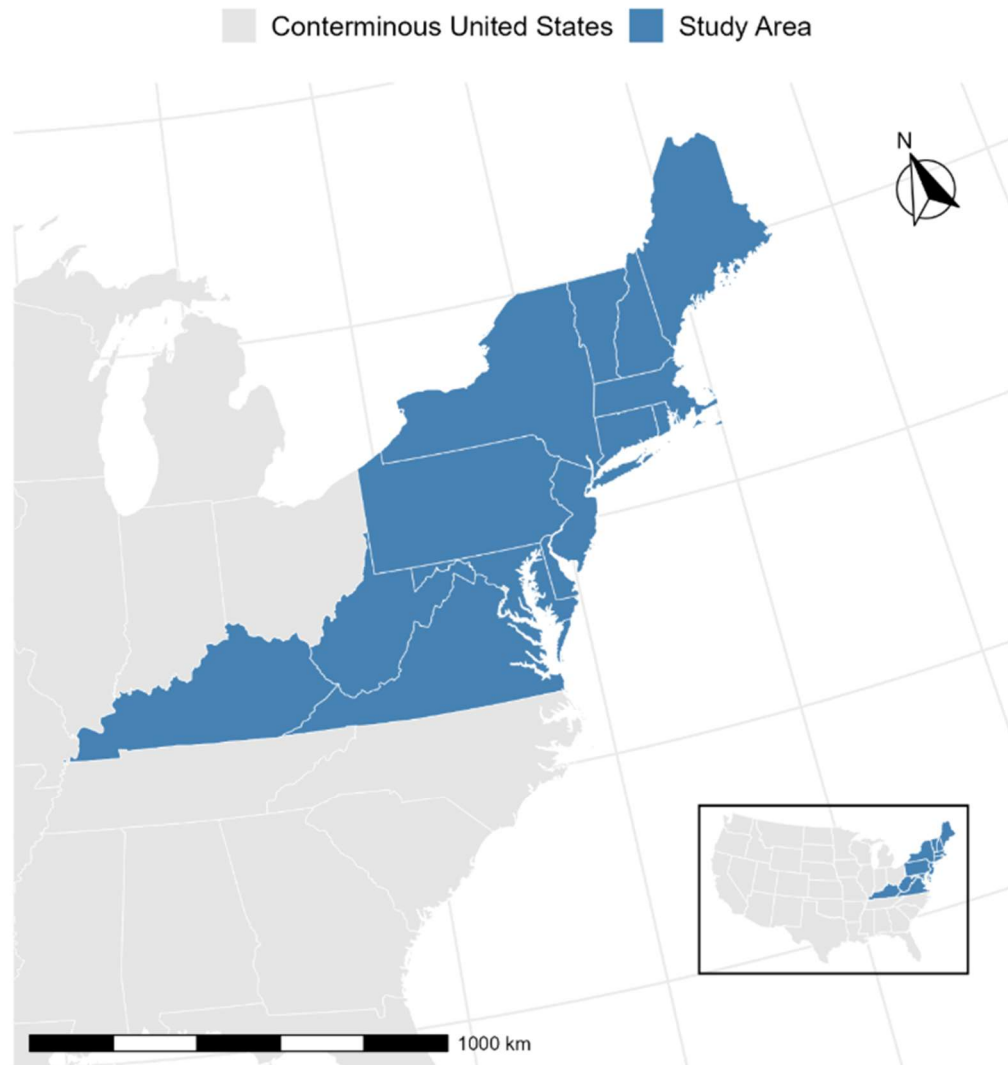


Figure 1. Area of interest in the conterminous Northeast United States. The included states (blue) are Connecticut, Delaware, Kentucky, Maine, Maryland, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Virginia, and West Virginia.

3.3.2 *Model building*

For each species and habitat type (freshwater lentic or lotic system), we divided presence and background data into training (70%) and testing (30%) partitions using non-random spatial splits (hashtag shape; Chapter 2). Within training data, we used additional spatial splits (blocks)

to create internal k-fold cross-validation (CV) splits (max k: 9). We used the remaining withheld test data to evaluate model performance (see below). We did not create a test partition if spatial splitting reduced the number of training data records below 150 or when spatial CV was inappropriate (e.g., highly clustered points). In these instances, we replaced the spatial CV split with a randomized 10-fold CV split.

To create species-specific models, we considered a suite of current and future spatial environmental predictors with continuous coverage of freshwater habitat across the conterminous United States including temperature, atmospheric water, land-use, soil properties, topographic characteristics, and hydrologic flow (Henderson et al., 2026a; 2026b). When pairwise correlation coefficients $\geq |0.7|$ between predictors (Dormann et al., 2013), we selected the predictor most biologically relevant for the target species based on expert knowledge of its life history and ideally maintained a ratio of at least 10:1 observations to predictors (Hosmer & Lemeshow, 2000). When applying models to future conditions, we held the selected predictor list constant apart from climate (single emissions scenario, five global circulation models [GCMs]), land use (single land cover stability scenario) and hydrologic flow (one future scenario). For climate, we used the five GCMs to explore the variation in model prediction for shared socioeconomic pathway 370, which represents a near-term, moderate emissions scenario ending in 2040 (Karger et al., 2020).

For each species and habitat type, we attempted to fit five algorithms: boosted regression trees (Elith et al., 2008), generalized linear models (McCullagh & Nelder, 2019), multivariate adaptive regression splines (Elith & Leathwick, 2007), Maxent (Phillips et al., 2017), and random forests (Breiman, 2001). We investigated potential overfitting for any model with a difference > 0.1 between training and mean CV performance (measured by the area under the receiver-operator curve) or where visual assessment of predictor response curves appeared overly complex. We

tuned hyperparameters as needed to reduce overfitting in these cases. We evaluated individual model performance using the continuous Boyce index (CBI; Hirzel et al., 2006) on training and available testing sets, revisiting any models with CBI scores ≤ 0.50 to determine if they should be dropped from consideration or refit based on expert species knowledge and ecological plausibility.

We used the final, retained models to generate mapped predictions of habitat suitability for current and future conditions across lentic and lotic habitats in the Northeast U.S. at ~ 100 m spatial resolution using the U.S. Geological Survey's high-performance computing resources (Falgout et al., 2023). From the individual model maps, we created weighted mean ensembles for each species by habitat type and within each GCM ($n = 5$) using model CBI scores as weights. We then calculated CBI scores for each ensemble output and again revisited the component models for any ensemble with $\text{CBI} < 0.50$ as above. To better understand the variability of predictions of changing climate, hydrology, and land cover, we also calculated per pixel minimum and maximum values across all individual future suitability model predictions ($\text{max } n = 25$). Finally, to identify areas of potential model extrapolation, we generated a multivariate environmental similarity surface (MESS; Elith et al., 2010) for each species based on the environmental ranges found in model training data. See Jarnevich et al (2026) for a full list of tuning values and evaluation metrics.

3.4 Invasion pathways

We selected five potential biotic and abiotic pathways of secondary spread to capture potential freshwater invasive species movement (Table 1). We calculated pathways scores by subwatershed (HUC12) with the exception of “Anthropogenic connectivity” which counts the number of human-built water movement corridors that cross hydrological subregions (HUC4) resulting in a single value for all nested subwatersheds (HUC12). Each pathway layer is a spatial

proxy of potential movement with normalized values ranging from 0 to 1. For each species, we summed individual layer scores for only the pathways determined to be relevant for each individual species (selected via expert opinion; Appendix 1) and then renormalized this combined value (range: 0 - 1) to create a single, comprehensive pathways score for each focal species.

Table 1. List of secondary pathway spatial layers, their associated data sources, and brief descriptions of what they represent. See production R scripts in Jarnevich et al (2026) for more detailed code used to produce each layer.

Name	Data Sources	Description
Downstream Connectivity	National Waterbody Dataset (NHD; U.S. Geological Survey, 2025)	Distance to nearest occupied subwatershed (HUC12) based on downstream flow. Intended to capture proximity to potential downstream spread through natural waterways.
Boat Ramp Density	Benson et al. (2023)	Ratio of the number of boat ramps to total freshwater area (waterbodies & streams) by subwatershed (HUC12). Intended to represent potential hitchhiking opportunities on boats, trailers, or other equipment used in or near freshwater systems.
Permanent Water	Henderson et al. (2026a; 2026b)	Ratio of permanent water body area (lakes & wetlands) to total subwatershed (HUC12) area. Intended to capture the potential for transport via hitchhiking on or in waterfowl.
Human Transport Risk	Prev��y et al. (2025)	Mean of human transport risk index by subwatershed (HUC12). Intended to capture anthropogenic sources of potential release from captivity, water gardens, aquaculture, or live bait.
Anthropogenic Connectivity	National Hydrography Dataset	Count of human-built water infrastructure (e.g., canals, waterpipes) connections that cross HUC4 boundaries. Intended to capture potential spread via artificial connections.

3.5 Composite risk score outputs & summaries

To create our final spatial representations of invasion risk scores, we first calculated the mean current and future habitat suitability scores by subwatershed and normalized each to a 0-1

scale. For future suitability, this score represented the mean of the five GCM ensembles. We produced our final invasion risk scores by summing the three normalized values (current suitability + future suitability + composite pathways) to provide equal weighting from each of the components. The resulting scores range from 0-3 with zero indicating the lowest potential risk and three indicating the highest potential risk.

To demonstrate one way to summarize species and taxa-level trends, we calculated watchlists of the 10 highest risk scores by species for each U.S. state in the focal study area. We chose to down weight composite risk scores where subwatersheds A) had few predictions due to limited spatial covariate coverage or environmental masking; B) had a high degree of uncertainty as measured by variability (standard deviation) across future predictions. Down weighting these conditions helped address areas of minimal freshwater coverage and areas of high variability among future predictions to create more ecologically realistic watchlists. We recomputed the down weighted composite scores and summarized them by state. See Appendix 1 for an expanded description of the down weighting methodology.

4 Results

4.1 SDMs – future suitability predictions

Of the initial 154 identified species, 104 had sufficient occurrence data to build either lentic models (n=3), lotic models (n=18), or both (n=83). We dropped models with low performance, resulting in at least one well-performing algorithm for 103 species (fishes: 34; macroinvertebrates: 15; plants: 54; Jarnevich et al., 2026). Model performance varied by habitat type and taxa (Appendix 1). Across all taxa, lentic models had greater variation in model performance than lotic models ($p < 0.001$ for fishes and plants) but were

not significantly different across taxa groups (Appendix 1). The minimum and maximum of future suitability (from a potential maximum of 25 models per species) varied across the study region (Fig. 2). The difference between predictions of current and mean future habitat suitability varied in geographic space by taxa group and species. For fishes, we observed increases in suitability across the study region with notable increases in the coastal states of Maryland and Delaware. For plants, we observed increases in suitability across most of the study region without notable concentrated regions of increase or decrease. For macroinvertebrates, we observed more areas of decreased suitability and more isolated areas of increased suitability (Fig. 3).

4.2 Maps of invasion pathways

We created spatial representations of five secondary invasion pathways (Table 1; Fig. 4). The overall number and geographic spread of occurrence records determined the results for our only species-specific pathway (“Downstream Connectivity”; Fig. 4-B). For boat ramp density (Fig. 4-C), we removed coastal watercraft access areas to limit the pathway to areas with adjacent freshwater habitat. The boundaries of subregion (HUC4) units are visible in some of the overall secondary pathways scores, reflecting the high number of landscape modifications (e.g., artificial waterways) typical of large human population centers.

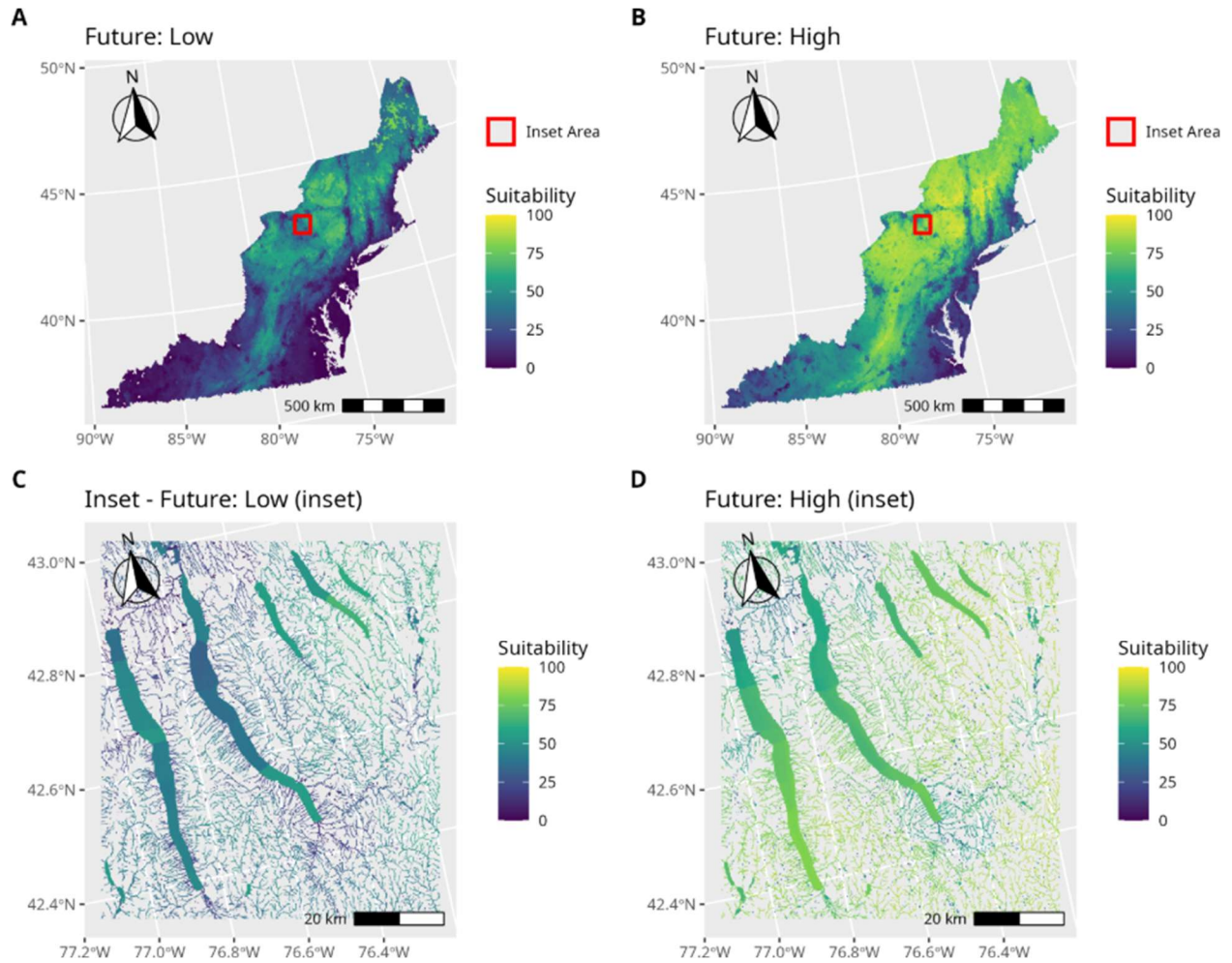


Figure 2. A display of the range of potential future habitat suitability for an example species (brown trout; *Salmo trutta*). Plots A and B show raster data summarized by subwatershed for the lowest (A) and highest (B) potential future suitability predictions across all applications of individual models to future environmental conditions ($n = 25$) across the focal study area of the Northeast United States. Plots C and D show the rasters at the pixel level for an inset area from A and B, highlighting the level of detail available in a local context (here, the Finger Lakes region of New York, USA).

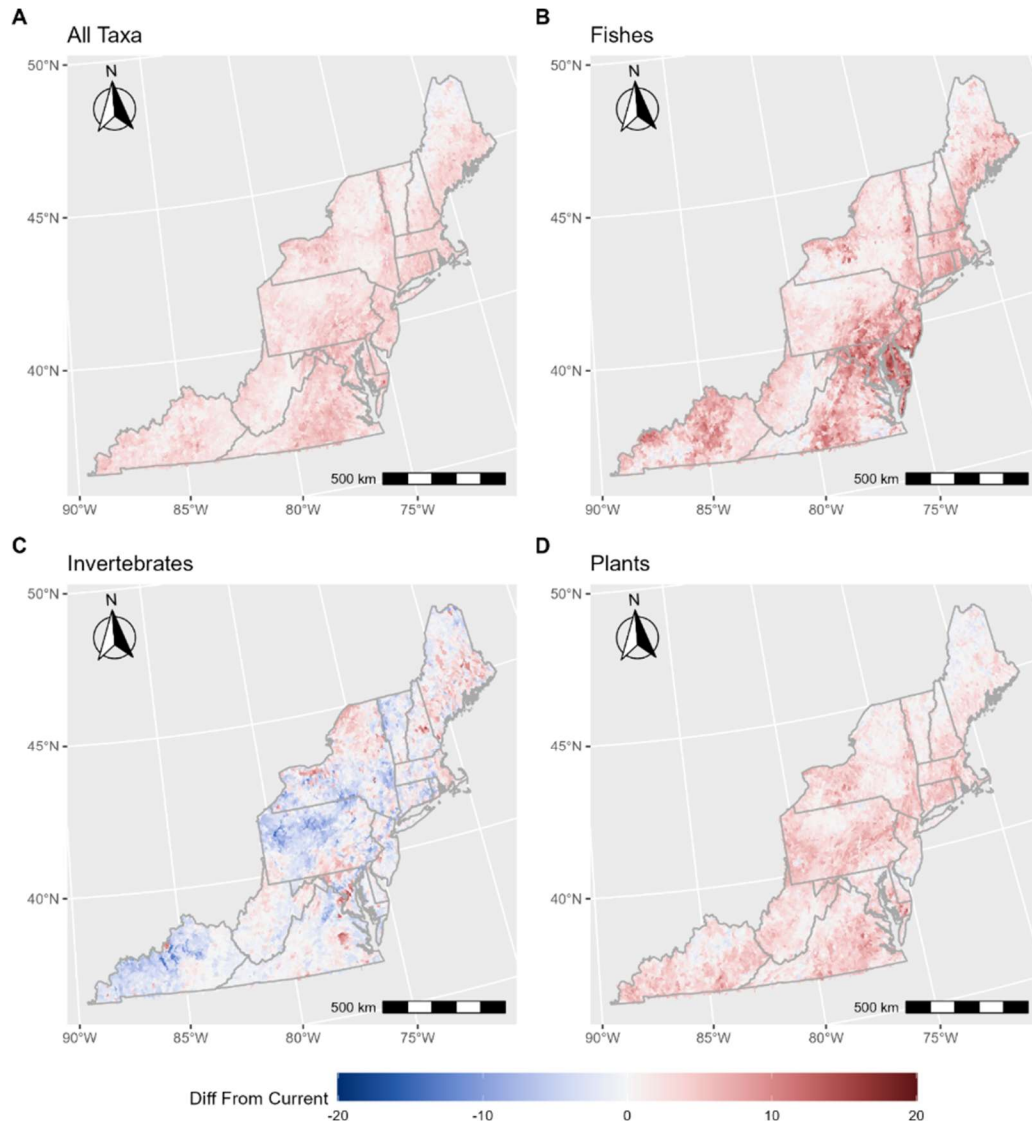


Figure 3. Differences between current and future relative habitat suitability across all study taxa and by taxa group (**A** - fishes, **B** - macroinvertebrates, **C** - plants) summarized by mean value within subwatersheds across the Northeast United States. The original suitability values each range from 0-100 meaning that differenced values can range from -100 to 100. Positive values (red) indicate increased suitability under future environmental scenarios whereas negative values (blue) indicate decreased suitability.

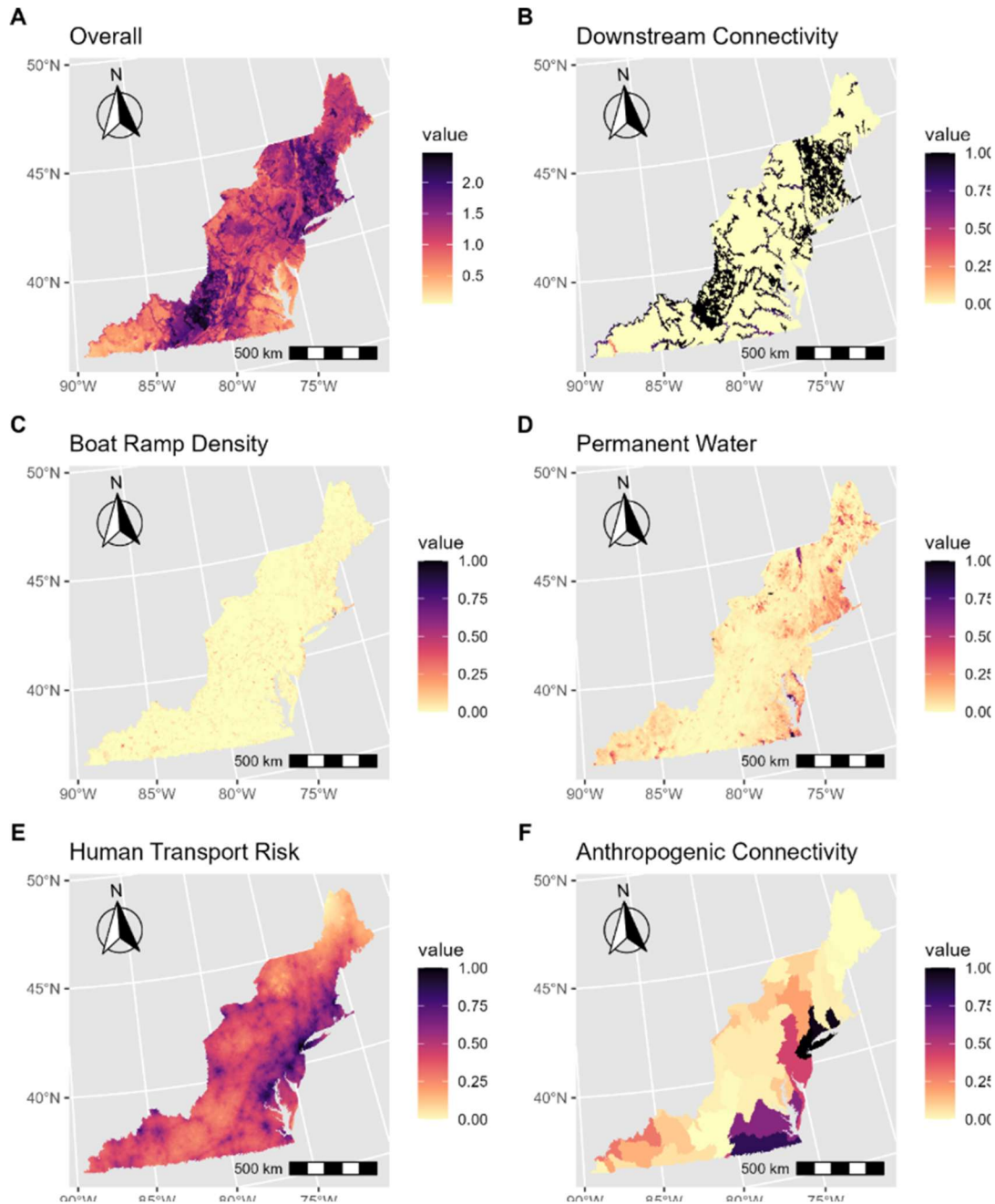


Figure 4. Secondary pathways score components by subwatershed across the study area in the Northeast United States for Japanese knotweed (*Fallopia japonica*). Note that “Anthropogenic Connectivity” (F) is calculated by first counting the number of artificial waterway connections between subregions (HUC4) and then applying a uniform value to all contained subwatersheds (HUC12). All other pathways are calculated solely at the subwatershed level.

4.3 Composite risk scores

Composite risk scores reflected the geographic trends observed in suitability and pathways scores, with higher scores nearer to large urban population centers or transport corridors (Figs. 5-6). Our method for generating watchlists by state highlighted 38 unique species across all top 10 lists (Fig. 7). The watchlists are dominated by plant species (66%; 25 of 38), which are overrepresented relative to the percentage of plant species modeled (52%; 54 of 103). Fishes are underrepresented in the top 10 lists (13%; 5 of 38) relative to the modeled species (33%; 34 of 103), while macroinvertebrates are slightly overrepresented in the watchlists (21%; 8 of 38) relative to the modeled proportion (15%; 15 of 103).

5 Discussion

We present a novel, spatially explicit approach for generating risk assessments of freshwater invasive species by augmenting current and future habitat suitability predictions with maps of potential species movement representing secondary pathways of invasion. This work builds upon previous invasion risk assessment methods in overcoming tradeoffs between taxonomic breadth, geographic coverage, and computational effort by leveraging a wide array of data resources and modeling expertise alongside end user driven species prioritization. The resulting products include mapped assessments of 103 freshwater invasive species across multiple taxa (fishes, macroinvertebrates, plants) that can support local and regional management (including early detection efforts) across the Northeast United States. Shifts in habitat suitability are influenced by changing climate, land use, and hydrology coupled with secondary pathways that allow physical connections or transport potential between invaded and previously uninvaded (but now suitable) habitat. Our approach attempts to capture these spatial factors

to generate assessments of invasion risk that can help managers of freshwater systems better anticipate invasions and efficiently leverage limited resources.

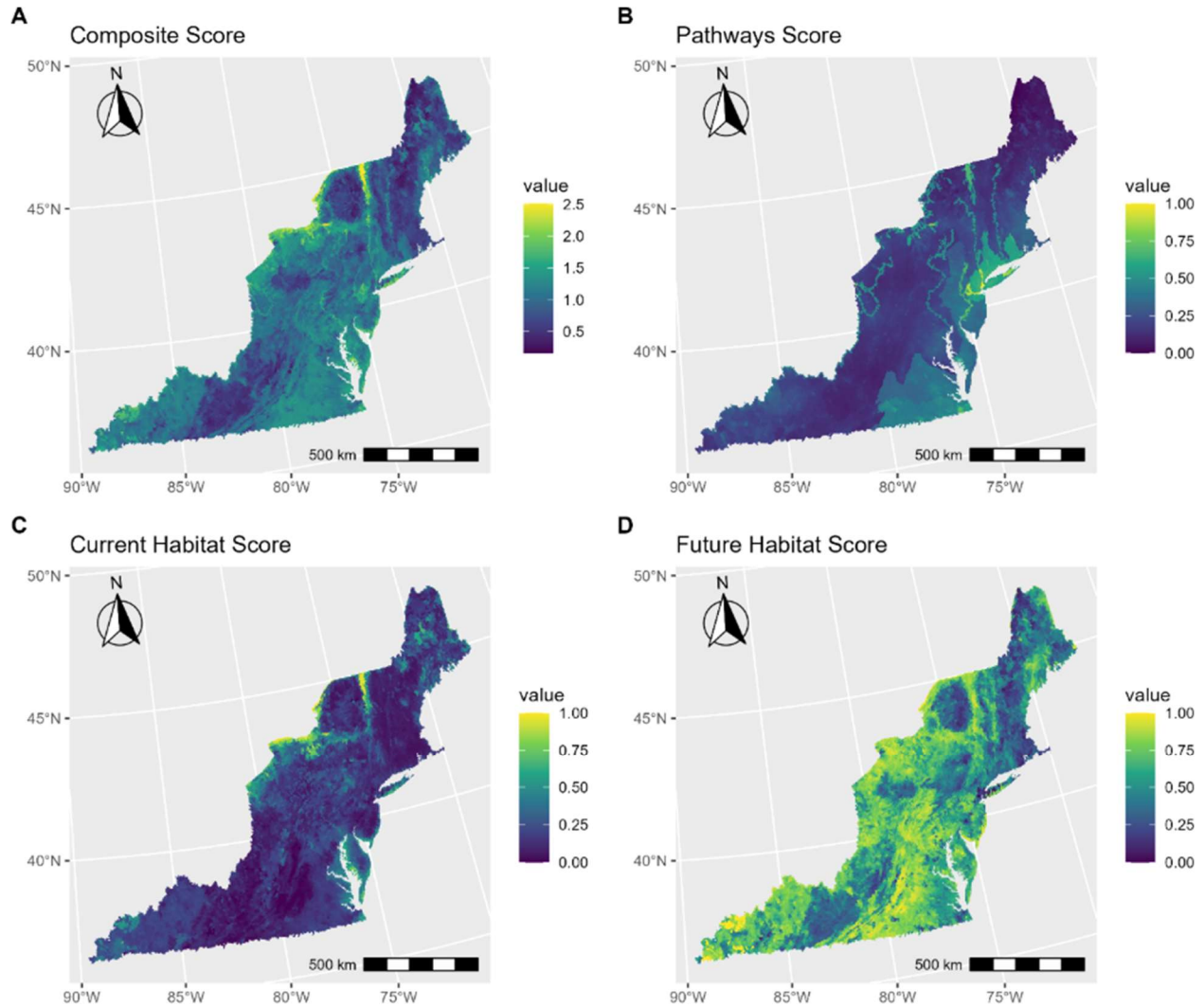


Figure 5. Example composite risk score (A) and components (secondary pathways - B, current habitat suitability - C, future habitat suitability - D) summarized (mean) within subwatersheds across the Northeast United States for European frogbit (*Hydrocharis morsus-ranae*). The scale for A ranges from 0-3, representing the cumulative sum of the components by subwatershed (HUC12). The scale for B-D ranges from 0-1, representing normalized scores.

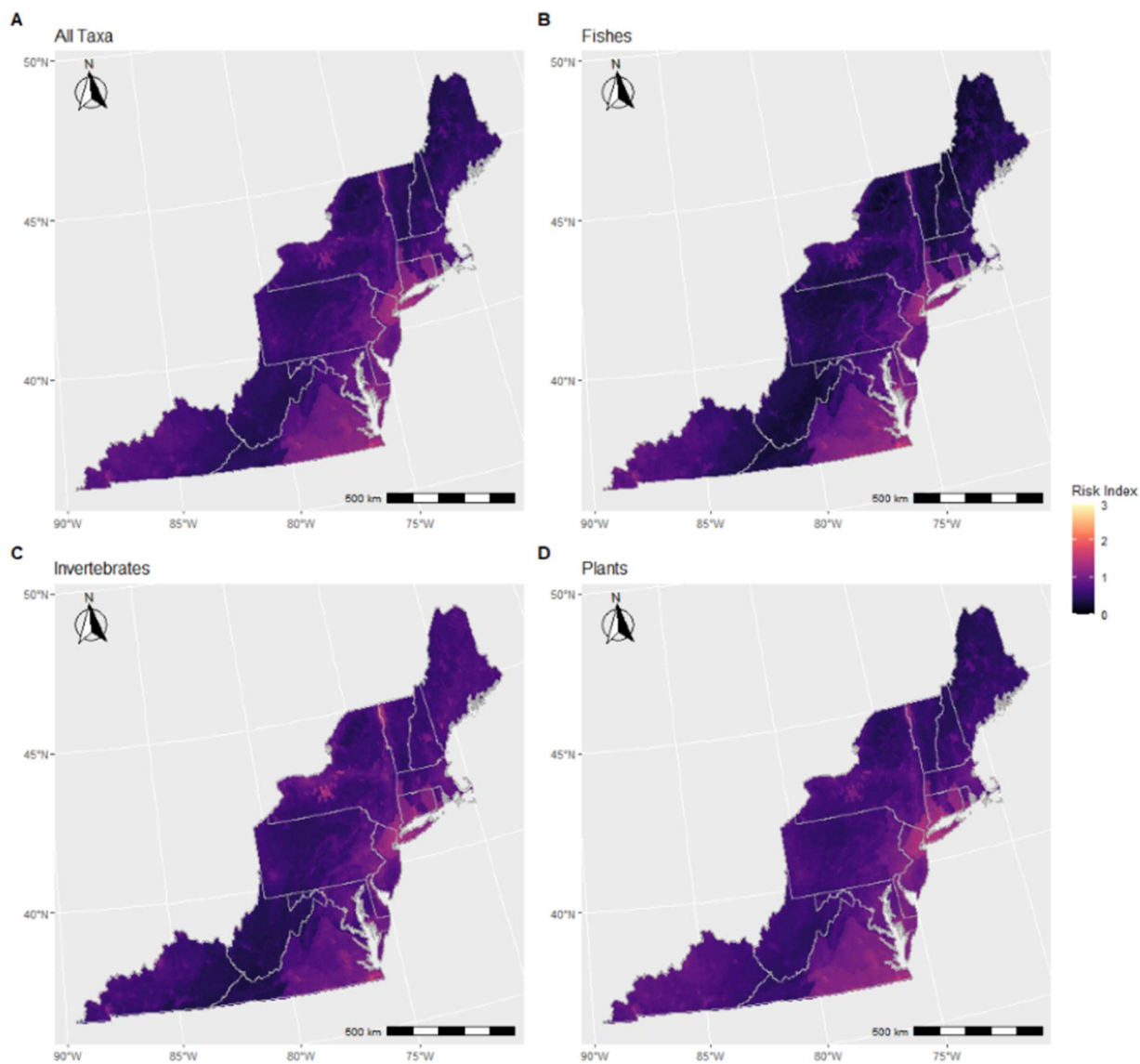


Figure 6. Composite risk scores across all study taxa and by taxa group. Values are summarized as the mean of composite risk (min: 0, max: 3) within subwatersheds across the Northeast United States.



Figure 7. An example ranking of species by median invasion risk within each state in the focal study area and across the Northeast United States. Data have been filtered to down-weight subwatersheds where predictions may be less reliable (e.g., a species with predictions in <10% of the freshwater habitat of a watershed or with highly uncertain future suitability—see Appendix 1). “Alert Species” are those with fewer than five publicly available observations after spatial thinning, indicating a possible lack of establishment and potential for early detection and rapid response. Note that species lists for this study, while informed by resource managers across multiple jurisdictions, should not be considered comprehensive of all potential freshwater invasive species in the Northeast United States.

5.1 Species distribution models

We chose to model lentic and lotic habitats separately due to the differing spatial covariates that best represent habitat characteristics unique to each system. For models of lentic habitat, performance was lower and more variable compared to models of lotic habitat (significantly so for fishes and plants; Appendix 1). The difference in model performance between habitat types may be explained by a combination of data types and thinning methods. To define lentic habitat, we used polygon features from NHD. This meant that our data thinning process was conservative in that it may have removed occurrences falling near but just outside polygon boundaries, potentially missing observations taken further ashore (e.g., near boat ramps or in parking lots). However, for lotic models, we rasterized polylines to the standardized spatial resolution of our environmental predictors (~100 m) due to a lack of consistent or dynamic stream or river width metadata in NHD. As a result, our thinning of occurrence data relative to rasterized lotic data types likely had a higher spatial overlap with records taken near but not in streams and rivers. Alternate thinning methods (e.g., buffered polygons) for future iterations of lentic models may improve model performance through increased data density and refining or broadening niche specification.

5.2 Secondary pathways

The inclusion of mapped secondary pathways is especially useful for invasion risk assessments for three primary reasons: 1) they help refine ecological niche specification; 2) they expand extant uses of pathways knowledge in existing assessment frameworks; 3) they help improve the understanding of range shifting species.

While our methods do not directly include movement data as predictors in the modeling, a rare technique in the SDM literature (Miller & Holloway, 2015; Seaborn et al., 2020), we leverage

our mapped secondary pathways as spatial weights for predictions from SDMs. This post-hoc combination of data types has the potential to improve the ecological realism and practical utility of modeled predictions by narrowing the modeled expression of the realized niche to highlight locations with greater levels of accessibility (Barve et al., 2011).

Second, a variety of invasion risk assessment frameworks specify the understanding of pathways as a way to determine a species' potential to disperse and spread (Canavan et al., 2025). But existing frameworks often only consider pathways in conceptual, non-spatial contexts and therefore exclude meaningful information about how pathways act as geographically dynamic spatial systems (e.g., hydrological networks) that directly influence the distribution of invasive propagules. A handful of innovative, spatially explicit approaches integrate pathways into risk assessments of invasive species but remain limited by geography (Bossenbroek et al., 2007), pathway type (Leung et al., 2006) or species breadth (Cook et al., 2019). We build upon these previous works by expanding geographic and taxonomic breadth and using our five mapped secondary pathways to help constrain predictions of establishment potential (suitability maps). Our outputs add high-utility, geographically detailed information that complements the existing knowledge of resource managers in the Northeast United States.

Third, spatial representations of secondary pathways are especially important when risk assessments attempt to characterize the future risk of range shifting species. Though the availability of highly suitable habitat is expected to increase for invasive species change under future environmental conditions (Bradley et al., 2024), the dispersal capacity of a species will help determine its ability to reach potential locations of high suitability (Sofaer et al., 2018). Previous assessments of range shifting species primarily focused on terrestrial habitats and seldom integrate pathways, leading to assessments that effectively represent unlimited dispersal scenarios (e.g.,

Salva & Bradley, 2023). Our results represent a novel contribution that builds on previous work on range shifts by expanding into freshwater habitats while using mapped pathways to create realistic estimates of dispersal limitations for freshwater invasive species. We are unaware of other studies that integrate dispersal limitations along with predictions of both current and future suitability to assess spatially explicit invasion risk in freshwater systems at an equivalent geographic extent.

5.3 Management implications of composite risk

Our spatially explicit invasion risk assessment approach offers a systematic, repeatable framework that is multi-taxa and jurisdiction-agnostic, providing fine-scale resolution but maintaining geographically broad coverage. This directly supports existing invasive species management activities across the Northeast United States and responds to longstanding calls for unified, science-based assessments that expand both the geographic and temporal scope for evaluations of invasion risk (Bradley et al., 2022; U.S. Environmental Protection Agency, 2008). Importantly, our composite risk scores can be summarized across local, state, or region-wide levels (Fig. 7), and the results of our watchlist approach align with existing risk assessments and state watchlists.

Despite growing interest in incorporating global change dynamics (e.g., climate-driven range shifts) into invasive species management, resource managers still face significant knowledge gaps in quantifying habitat suitability under future conditions (Beaury et al., 2020). Our results help bridge this gap by providing actionable, spatially explicit resources that integrate multiple dimensions of invasion risk: current habitat suitability, projected future suitability, and secondary introduction pathways. These components can be examined individually or combined into a composite score, offering flexibility for differing management goals and environmental contexts.

Our high-resolution spatial outputs (raster maps of current and future suitability) enable managers to quickly identify and prioritize waterbodies where invasion risk is expected to increase for individual species or across multiple taxa (Fig. 3). This prioritization is critical for allocating limited resources efficiently. For example, areas identified as risk hotspots within ecologically sensitive habitats or high-use recreational destinations can be targeted for early detection activities, including environmental DNA (eDNA) sampling, enhanced monitoring, and prevention efforts. Furthermore, these outputs can inform strategic planning for outreach, permitting, and infrastructure design, enabling management actions that are proactive rather than reactive (Van Rees et al., 2022). By integrating multiple dimensions of invasion risk into a single, scalable framework, our approach provides a foundation for adaptive invasive species management under changing environmental conditions while supporting data-driven and forward-looking decision-making, helping agencies anticipate emerging threats and coordinate responses across jurisdictions.

5.4 Riskiest species

Our top ten lists (Fig. 7) highlighted invasive species that may spread into and within the Northeast United States. Here we review several of the most commonly ranked species where their high ecological and economic impacts coincide with their high rankings. Golden clams (*Corbicula fluminea*) are a widely established non-native bivalve across much of the contiguous United States and portions of Northeast states (Benson & Williams, 2021) in almost any perennial stream or waterbody if the system has sufficient calcium (Rollwagen-Bollens et al., 2021). Not only is *C. fluminea* the highest risk regionwide in our example watchlist, it is also number one for four states (Delaware, Kentucky, Maryland, Virginia). Although not as impactful as *Dreissena* species, *C. fluminea* can compete with native bivalves for food and habitat (Ferreira-Rodríguez et al., 2018),

alter sediment and nutrient cycling (Castañeda et al., 2018), and has the potential to clog industrial and municipal water intake pipes (Isom, 1986). The two fish species in the regionwide top ten (*Ambloplites rupestris* and *Salmo trutta*) have been introduced to drainages for new sport fishing opportunities (USGS 2024). Both species are known to compete with native fishes (Jenkins & Burkhead, 1994; Taylor et al., 1984) and reduce the abundance and diversity of native prey (Houde et al., 2017; Hughes & Herlihy, 2012). The last non-plant in the top ten, *Faxonius virilis*, is a crayfish native to parts of the Midwest United States but introduced widely beyond its range through bait bucket and aquarium releases (Kilian et al., 2012). Once established, *F. virilis* can compete with native crayfish (Hale et al., 2016) and alter benthic habitats and vegetation (Chambers et al., 1990; Hanson et al., 1990).

The plants in the regionwide top ten list (Fig. 7) represent introductions from multiple sources: culinary (*Persicaria hydropiper* and *Nasturtium officinale*), forage (*Fallopia japonica*), and ornamental uses (*Myosotis scorpioides*, *Lythrum salicaria*, and *Iris pseudacorus*) (Les, 2018; Mills et al., 1993). Though these plants are ubiquitous in the Northeast United States, they have the potential to spread along available pathways to new habitats by hitchhiking on humans and animals, traveling downstream from established populations, or spread via nursery and grocers' sales (Les, 2018). The plants tend to be highly competitive for resources in streams (Beerling et al., 1994; Blokhina, 2003), produce abundant and dense seeds for long-term establishment (Malecki et al., 1993; Thompson et al., 1987), have detrimental effects on habitats (Stoll et al., 2012; Templer et al., 1998; Walsh & Phelps, 1991), and potentially reduce property, agricultural, and recreational value (Child & Wade, 2000; Les, 2018). Many state resource agencies in the Northeast United States already include the plant species we identified in their state management plans (e.g., <https://www.northeastans.org/index.php/home-page/member-states-and-provinces/>).

Thirteen species in Fig. 7 are native to the U.S. (*Ambloplites rupestris*, *Ameiurus natalis*, *Lepomis gulosus*, *Procambarus clarkii*, *Faxonius rusticus*, *Faxonius virilis*, *Myriophyllum heterophyllum*, *Cabomba caroliniana*, *Utricularia inflata*, *Hydrocotyle ranunculoides*, *Ludwigia peploides*, *Nelumbo lutea*, and *Didymosphenia geminata*). While their native ranges include some of the states in the study range, native establishment coupled with changing habitat conditions and available pathways make them likely to continue spreading with potential, new populations establishing in more northern latitudes. We labeled any species with fewer than five publicly available occurrence records in a state yet ranked among the top ten species by median invasion risk within that state as “alert species.” These species represent potential emergent threats and are considered high-risk for future establishments in that state. For example, the most common “alert species” was *Didymosphenia geminata*, showing up in our lists for Maine, New Hampshire, Vermont, and West Virginia. Species may also have multiple alert statuses depending on the focal state. For example, *Faxonius virilis* is an alert species for New York and Rhode Island but not for Connecticut, Maine, Massachusetts, or New Hampshire. These examples underline the flexibility of our approach and results to anticipate differing management responses for differing levels of risk for a given species in a given management context.

6 Future Work

The methods presented here represent a reproducible workflow that could be adapted to new geographies, taxa, and ecosystems beyond the freshwater systems of the Northeast United States. To continue iterating and improving this methodology, we stress the need for ongoing partnerships and resource sharing (e.g., advanced computing resources or taxa expertise) to generate quantitative invasive species risk assessments that are taxonomically broad, computationally efficient, and geographically comprehensive. For example, we made use of publicly available

observations of species occurrences, and increasing the accessibility of data can reduce the ‘data silo’ effect, filling spatial information gaps and leading to more robust and credible risk assessments. We only explored a single future climate scenario (under multiple GCMs) and revised methods could look for ways of incorporating a larger suite of potential climate projections aligned with our other future projections (streamflow, land use). Finally, the inclusion of additional information on secondary pathways (e.g., potential water connections from simulated flood maps; Pfingsten et al., 2024) and expert elicited species-specific pathway weighting could serve to refine the next generation of freshwater invasive species risk maps in the Northeast United States and beyond.

REFERENCES

- Atwater, D., & Barney, J. N. (2021). Climatic niche shifts in 815 introduced plant species affect their predicted distributions. *Global Ecology and Biogeography*, 30(8), 1671–1684. <https://doi.org/10.1111/geb.13342>
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222(11), 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Beaury, E. M., Fusco, E. J., Jackson, M. R., Laginhas, B. B., Morelli, T. L., Allen, J. M., Pasquarella, V. J., & Bradley, B. A. (2020). Incorporating climate change into invasive species management: Insights from managers. *Biological Invasions*, 22(2), 233–252. <https://doi.org/10.1007/s10530-019-02087-6>
- Beerling, D. J., Bailey, J. P., & Conolly, A. P. (1994). *Fallopia Japonica* (Houtt.) Ronse Decraene. *The Journal of Ecology*, 82(4), 959. <https://doi.org/10.2307/2261459>
- Benson, A. J., Beck, H. J., Neilson, M. E., Daniel, W. M., Devantier, A. L., Cousins, A., King, A. M., Schermerhorn, M., & Tangkulwarodom, S. (2023). *Boat ramp locations in the United States of America* [Dataset]. U.S. Geological Survey data release. <https://doi.org/10.5066/P9ZIRVF0>
- Benson, A. J., & Williams, J. D. (2021). *Review of the Invasive Asian Clam Corbicula spp. (Bivalvia: Cyrenidae) Distribution in North America, 1924–2019* (U.S. Geological Survey Scientific Investigations Report 2021–5001 Nos. 2021–5001; pp. 1–66). <https://doi.org/10.3133/sir20215001>
- Bertelsmeier, C., Luque, G. M., & Courchamp, F. (2013). Increase in Quantity and Quality of Suitable Areas for Invasive Species as Climate Changes. *Conservation Biology*, 27(6), 1458–1467. <https://doi.org/10.1111/cobi.12093>
- Blokhina, O. (2003). Antioxidants, Oxidative Damage and Oxygen Deprivation Stress: A Review. *Annals of Botany*, 91(2), 179–194. <https://doi.org/10.1093/aob/mcf118>
- Bossenbroek, J. M., Johnson, L. E., Peters, B., & Lodge, D. M. (2007). Forecasting the Expansion of Zebra Mussels in the United States. *Conservation Biology*, 21(3), 800–810. <https://doi.org/10.1111/j.1523-1739.2006.00614.x>
- Bradley, B. A., Beaury, E. M., Fusco, E. J., Munro, L., Brown-Lima, C., Coville, W., Kesler, B., Olmstead, N., & Parker, J. (2022). Breaking down barriers to consistent, climate-smart regulation of invasive plants: A case study of US Northeast states. *Ecosphere*, 13(5), e4014. <https://doi.org/10.1002/ecs2.4014>
- Bradley, B. A., Beaury, E. M., Gallardo, B., Ibáñez, I., Jarnevich, C., Morelli, T. L., Sofaer, H. R., Sorte, C. J. B., & Vilà, M. (2024). Observed and Potential Range Shifts of Native and Nonnative Species with Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, 55(1), 23–40. <https://doi.org/10.1146/annurev-ecolsys-102722-013135>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

- Canavan, S., Canavan, K., Kumschick, S., Gordon, D. R., Wilson, J. R. U., & Lieurance, D. (2025). Invasive species risk assessment in practice: Insights from a survey of practitioners. *NeoBiota*, 99, 341–362. <https://doi.org/10.3897/neobiota.99.153010>
- Castañeda, R., Cvetanovska, E., Hamelin, K., Simard, A., & Ricciardi, A. (2018). Distribution, abundance and condition of an invasive bivalve (*Corbicula fluminea*) along an artificial thermal gradient in the St. Lawrence River. *Aquatic Invasions*, 13(3), 379–392. <https://doi.org/10.3391/ai.2018.13.3.06>
- Chambers, P. A., Hanson, J. M., Burke, J. M., & Prepas, E. E. (1990). The impact of the crayfish *Orconectes virilis* on aquatic macrophytes. *Freshwater Biology*, 24(1), 81–91. <https://doi.org/10.1111/j.1365-2427.1990.tb00309.x>
- Child, L. E., & Wade, P. M. (2000). The Japanese Knotweed Manual: The Management and Control of an Invasive Alien Weed (*Fallopia Japonica*) (1st ed). Liverpool University Press.
- Colwell, R. K., & Rangel, T. F. (2009). Hutchinson’s duality: The once and future niche. *Proceedings of the National Academy of Sciences*, 106(supplement_2), 19651–19658. <https://doi.org/10.1073/pnas.0901650106>
- Cook, G., Jarnevich, C., Warden, M., Downing, M., Withrow, J., & Leinwand, I. (2019). Iterative Models for Early Detection of Invasive Species across Spread Pathways. *Forests*, 10(2), 108. <https://doi.org/10.3390/f10020108>
- Cuthbert, R. N., Pattison, Z., Taylor, N. G., Verbrugge, L., Diagne, C., Ahmed, D. A., Leroy, B., Angulo, E., Briski, E., Capinha, C., Catford, J. A., Dalu, T., Essl, F., Gozlan, R. E., Haubrock, P. J., Kourantidou, M., Kramer, A. M., Renault, D., Wasserman, R. J., & Courchamp, F. (2021). Global economic costs of aquatic invasive alien species. *Science of The Total Environment*, 775, 145238. <https://doi.org/10.1016/j.scitotenv.2021.145238>
- Daniel, W. M., Sofaer, H. R., Jarnevich, C. S., Erickson, R. A., DeGregorio, B. A., Engelstad, P. S., Freedman, J., Canavan, S., Dean, E. M., Adams, M. J., Anderson, C. L., Barnett, M., Brey, M. K., Brumm, K. J., Bunting, M. S., Caffrey, E., Cardador, L., Carter, J., Cassey, P., ... Lieurance, D. (2025). Vertebrates in trade that pose high invasion risk to the United States. *Biological Conservation*, 302, 110887. <https://doi.org/10.1016/j.biocon.2024.110887>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Elith, J., & Leathwick, J. (2007). Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions*, 13(3), 265–275. <https://doi.org/10.1111/j.1472-4642.2007.00340.x>
- Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>

- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Evans, A. E., Jarnevich, C. S., Beaury, E. M., Engelstad, P. S., Teich, N. B., LaRoe, J. M., & Bradley, B. A. (2024). Shifting hotspots: Climate change projected to drive contractions and expansions of invasive plant abundance habitats. *Diversity and Distributions*, 30(1), 41–54. <https://doi.org/10.1111/ddi.13787>
- Falgout, J. T., Gordon, J., & Lee, R. L. (2023). *USGS Hovenweep Supercomputer*. <https://doi.org/10.5066/P927BI7R>
- Ferreira-Rodríguez, N., Fandiño, L., Pedreira, A., & Pardo, I. (2018). First evidence of asymmetric competition between the non-native clam *Corbicula fluminea* and the native freshwater mussel *Unio delphinus* during a summer heat wave. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 28(5), 1105–1113. <https://doi.org/10.1002/aqc.2964>
- Finch, D. M., Butler, J. L., Runyon, J. B., Fetting, C. J., Kilkenny, F. F., Jose, S., Frankel, S. J., Cushman, S. A., Cobb, R. C., Dukes, J. S., Hicke, J. A., & Amelon, S. K. (2021). Effects of Climate Change on Invasive Species. In T. M. Poland, T. Patel-Weynand, D. M. Finch, C. F. Miniati, D. C. Hayes, & V. M. Lopez (Eds.), *Invasive Species in Forests and Rangelands of the United States* (pp. 57–83). Springer International Publishing. https://doi.org/10.1007/978-3-030-45367-1_4
- GBIF. (2024). *Occurrence Download* [Dataset]. The Global Biodiversity Information Facility. <https://doi.org/10.15468/DL.U4HCXQ>
- 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. <https://doi.org/10.1016/j.tree.2014.02.009>
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8(9), 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3), 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Hale, P., Wilson, J., Loughman, Z., & Henkanaththegedara, S. (2016). Potential impacts of invasive crayfish on native crayfish: Insights from laboratory experiments. *Aquatic Invasions*, 11(4), 451–458. <https://doi.org/10.3391/ai.2016.11.4.09>
- Hanson, J. M., Chambers, P. A., & Prepas, E. E. (1990). Selective foraging by the crayfish *Orconectes virilis* and its impact on macroinvertebrates. *Freshwater Biology*, 24(1), 69–80. <https://doi.org/10.1111/j.1365-2427.1990.tb00308.x>
- Henderson, G. C., Engelstad, P., Reimer, C. J., Leclare, S. K., Fraser, L. S., Williams, D. A., Shadwell, K. S., Daniel, W. M., Pfingsten, I. A., & Jarnevich, C. S. (2026a). Current and near-future conditions of aquatic spatial data for use in ecological models in the United States. *Knowledge & Management of Aquatic Ecosystems*, (427), 11. <https://doi.org/10.1051/kmae/2026003>
- Henderson, G. C., Williams, D. A., Shadwell, K. S., Reimer, C. J., LeClare, S. K., Fraser, L. S., Engelstad, P. S., & Jarnevich, C. S. (2026b). *Freshwater Habitat Environmental Predictor Layers for Waterbodies and Streams across the Contiguous United States: U.S. Geological Survey data release*. [Dataset]. <https://doi.org/10.5066/P14JDTTJ>

- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (1st ed.). Wiley. <https://doi.org/10.1002/0471722146>
- Houde, A. L. S., Wilson, C. C., & Neff, B. D. (2017). Performance of four salmonids species in competition with Atlantic salmon. *Journal of Great Lakes Research*, 43(1), 211–215. <https://doi.org/10.1016/j.jglr.2016.10.017>
- Howeth, J. G., Amjad, S. A., Gantz, C. A., Mandrak, N. E., Angermeier, P. L., Marchetti, M. P., Olden, J. D., & Lodge, D. M. (2025). Predicting invasiveness of freshwater fishes imported into North America: Regional differences in models and outcomes. *Biological Invasions*, 27(4), 107. <https://doi.org/10.1007/s10530-025-03560-1>
- Hughes, R. M., & Herlihy, A. T. (2012). Patterns in Catch Per Unit Effort of Native Prey Fish and Alien Piscivorous Fish in 7 Pacific Northwest USA Rivers. *Fisheries*, 37(5), 201–211. <https://doi.org/10.1080/03632415.2012.676833>
- Hulme, P. E., Bacher, S., Kenis, M., Klotz, S., Kühn, I., Minchin, D., Nentwig, W., Olenin, S., Panov, V., Pergl, J., Pyšek, P., Roques, A., Sol, D., Solarz, W., & Vilà, M. (2008). Grasping at the routes of biological invasions: A framework for integrating pathways into policy. *Journal of Applied Ecology*, 45(2), 403–414. <https://doi.org/10.1111/j.1365-2664.2007.01442.x>
- Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0), 415–427. <https://doi.org/10.1101/SQB.1957.022.01.039>
- Isom, B. G. (1986). Historical review of Asiatic clam (*Corbicula*) invasion and biofouling of waters and industries in the Americas. *Proceedings of the Second International Corbicula Symposium. American Malacological Bulletin Special Edition, No. 2.*, 1–5.
- Jarnevich, C. S., Engelstad, P. S., Williams, D. A., Shadwell, K. S., Reimer, C. J., Henderson, G. C., Fraser, L. S., LeClare, S. K., Inman, R. D., Pfingsten, I. A., & Daniel, W. M. (2026). *Aqua INHABIT species potential distribution across the contiguous United States* [Dataset]. U.S. Geological Survey. <https://doi.org/10.5066/P13JMOQW>
- Jarnevich, C. S., Engelstad, P., Williams, D., Shadwell, K., Reimer, C., Henderson, G., Prevey, J. S., & Pearse, I. S. (2024). Predicted occurrence and abundance habitat suitability of invasive plants in the contiguous United States: Updates for the INHABIT web tool. *NeoBiota*, 96, 261–278. <https://doi.org/10.3897/neobiota.96.134842>
- Jenkins, R. E., & Burkhead, N. M. (1994). *Freshwater Fishes of Virginia*. American Fisheries Society, Bethesda, MD.
- Karger, D. N., Schmatz, D. R., Dettling, G., & Zimmermann, N. E. (2020). High-resolution monthly precipitation and temperature time series from 2006 to 2100. *Scientific Data*, 7(1), 248. <https://doi.org/10.1038/s41597-020-00587-y>
- Kilian, J. V., Klauda, R. J., Widman, S., Kashiwagi, M., Bourquin, R., Weglein, S., & Schuster, J. (2012). An assessment of a bait industry and angler behavior as a vector of invasive species. *Biological Invasions*, 14(7), 1469–1481. <https://doi.org/10.1007/s10530-012-0173-5>
- Kingsbury, S., Poirier, L., Aucoin, R., Shaw-McDonald, S., Engelstad, P., Therriault, T., & Wilcox, M. (2025). A Horizon Scan watchlist of freshwater non-indigenous species in the maritime provinces of Canada. *Management of Biological Invasions*, 16(3), 699–723. <https://doi.org/10.3391/mbi.2025.16.3.07>

- Les, D. H. (2018). Aquatic dicotyledons of North America: Ecology, life history, and systematics. CRC press, Taylor & Francis group.
- Leung, B., Bossenbroek, J. M., & Lodge, D. M. (2006). Boats, Pathways, and Aquatic Biological Invasions: Estimating Dispersal Potential with Gravity Models. *Biological Invasions*, 8(2), 241–254. <https://doi.org/10.1007/s10530-004-5573-8>
- Lieurance, D., Canavan, S., Behringer, D. C., Kendig, A. E., Minter, C. R., Reisinger, L. S., Romagosa, C. M., Flory, S. L., Lockwood, J. L., Anderson, P. J., Baker, S. M., Bojko, J., Bowers, K. E., Canavan, K., Carruthers, K., Daniel, W. M., Gordon, D. R., Hill, J. E., Howeth, J. G., ... Wanamaker, C. (2023). Identifying invasive species threats, pathways, and impacts to improve biosecurity. *Ecosphere*, 14(12), e4711. <https://doi.org/10.1002/ecs2.4711>
- Malecki, R. A., Blossey, B., Hight, S. D., Schroeder, D., Kok, L. T., & Coulson, J. R. (1993). Biological Control of Purple Loosestrife. *BioScience*, 43(10), 680–686. <https://doi.org/10.2307/1312339>
- McCullagh, P., & Nelder, J. A. (2019). *Generalized Linear Models* (2nd ed.). Routledge. <https://doi.org/10.1201/9780203753736>
- MDDNR. (2024). *Maryland Aquatic Nuisance Species Management Plan*. Maryland Department of Natural Resources. https://dnr.maryland.gov/Invasives/Documents/Maryland_Aquatic_Nuisance_Species_Plan.pdf
- Miller, J. A., & Holloway, P. (2015). Incorporating movement in species distribution models. *Progress in Physical Geography: Earth and Environment*, 39(6), 837–849. <https://doi.org/10.1177/0309133315580890>
- Mills, E. L., Leach, J. H., Carlton, J. T., & Secor, C. L. (1993). Exotic Species in the Great Lakes: A History of Biotic Crises and Anthropogenic Introductions. *Journal of Great Lakes Research*, 19(1), 1–54. [https://doi.org/10.1016/S0380-1330\(93\)71197-1](https://doi.org/10.1016/S0380-1330(93)71197-1)
- Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges. In I. Koprinska, M. Kamp, A. Appice, C. Loglisci, L. Antonie, A. Zimmermann, R. Guidotti, Ö. Özgöbek, R. P. Ribeiro, R. Gavaldà, J. Gama, L. Adilova, Y. Krishnamurthy, P. M. Ferreira, D. Malerba, I. Medeiros, M. Ceci, G. Manco, E. Masciari, ... J. A. Gulla (Eds.), *ECML PKDD 2020 Workshops* (Vol. 1323, pp. 417–431). Springer International Publishing. https://doi.org/10.1007/978-3-030-65965-3_28
- Morisette, J. T., Jarnevich, C. S., Holcombe, T. R., Talbert, C. B., Ignizio, D., Talbert, M. K., Silva, C., Koop, D., Swanson, A., & Young, N. E. (2013). VisTrails SAHM: Visualization and workflow management for species habitat modeling. *Ecography*, 36(2), 129–135. <https://doi.org/10.1111/j.1600-0587.2012.07815.x>
- Pergl, J., Brundu, G., Harrower, C. A., Cardoso, A. C., Genovesi, P., Katsanevakis, S., Lozano, V., Perglová, I., Rabitsch, W., Richards, G., Roques, A., Rorke, S. L., Scalera, R., Schönrogge, K., Stewart, A., Tricarico, E., Tsiamis, K., Vannini, A., Vilà, M., ... Roy, H. E. (2020). Applying the Convention on Biological Diversity Pathway Classification to alien species in Europe. *NeoBiota*, 62, 333–363. <https://doi.org/10.3897/neobiota.62.53796>
- Pfingsten, I. A., Reaver, K. M., Dean, E. M., Neilson, M. E., Chivoiu, B., & Daniel, W. M. (2024). Combining storm flood water level and topography to prioritize inter-basin transfer of non-

- native aquatic species in the United States. *Biological Invasions*, 26(12), 4105–4120. <https://doi.org/10.1007/s10530-024-03430-2>
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, 40(7), 887–893. <https://doi.org/10.1111/ecog.03049>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197. <https://doi.org/10.1890/07-2153.1>
- Prevéy, J. S., Reimer, C. J., Engelstad, P. S., Belamaric, P. N., Hogan, T., LaRoe, J. M., Mumford, C. J., Sieracki, J. L., & Jarnevich, C. S. (2025). *A Site Prioritization Tool for Invasive Species Management: Integrating Diverse Spatial Data to Improve Decision Making*. <https://doi.org/10.2139/ssrn.5342105>
- R Core Team. (2025). *R: A language and environment for statistical computing*. [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reaser, J. K., Burgiel, S. W., Kirkey, J., Brantley, K. A., Veatch, S. D., & Burgos-Rodríguez, J. (2020). The early detection of and rapid response (EDRR) to invasive species: A conceptual framework and federal capacities assessment. *Biological Invasions*, 22(1), 1–19. <https://doi.org/10.1007/s10530-019-02156-w>
- Roberts, C. P., Grant, W. E., Horton, M. L., LaBrie, L. A. P., Peterson, M. R., Rogosch, J. S., & Wang, H. (2024). Balancing ecology and practicality to rank waterbodies for preventative invasive species management. *Ecological Solutions and Evidence*, 5(3), e12362. <https://doi.org/10.1002/2688-8319.12362>
- Rollwagen-Bollens, G., Bolam, B. A., Bollens, S. M., Henricksen, S., Sandison, C., & Zimmerman, J. (2021). Temperature-dependent functional response of the invasive Asian clam, *Corbicula fluminea*, feeding on natural phytoplankton. *Inland Waters*, 11(2), 250–256. <https://doi.org/10.1080/20442041.2020.1843933>
- Salva, J. D., & Bradley, B. A. (2023). High-impact invasive plants expanding into mid-Atlantic states: Identifying priority range-shifting species for monitoring in light of climate change. *Invasive Plant Science and Management*, 16(4), 197–206. <https://doi.org/10.1017/inp.2023.24>
- Sanders, S., Castiglione, C., & Hoff, M. (2023). *Risk Assessment Mapping Program: RAMP, version 5.0*. U.S. Fish and Wildlife Service.
- Seaborn, T. J., Goldberg, C. S., & Crespi, E. J. (2020). Integration of dispersal data into distribution modeling: What have we done and what have we learned? *Frontiers of Biogeography*, 12(4). <https://doi.org/10.21425/F5FBG43130>
- Simberloff, D., Martin, J.-L., Genovesi, P., Maris, V., Wardle, D. A., Aronson, J., Courchamp, F., Galil, B., García-Berthou, E., Pascal, M., Pyšek, P., Sousa, R., Tabacchi, E., & Vilà, M. (2013). Impacts of biological invasions: What's what and the way forward. *Trends in Ecology & Evolution*, 28(1), 58–66. <https://doi.org/10.1016/j.tree.2012.07.013>
- Simpson, A., Fuller, P., Faccenda, K., Evenhuis, N., Matsunaga, J., & Bowser, M. (2022). *United States Register of Introduced and Invasive Species (US-RIIS) (ver. 2.0, November 2022)* [Dataset]. U.S. Geological Survey. <https://doi.org/10.5066/P9KFFTOD>
- Soberón, J., & Peterson, A. T. (2005). Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics*, 2(0). <https://doi.org/10.17161/bi.v2i0.4>

- Sofaer, H. R., Jarnevich, C. S., & Flather, C. H. (2018). Misleading prioritizations from modelling range shifts under climate change. *Global Ecology and Biogeography*, 27(6), 658–666. <https://doi.org/10.1111/geb.12726>
- Stoll, P., Gatzsch, K., Rusterholz, H.-P., & Baur, B. (2012). Response of plant and gastropod species to knotweed invasion. *Basic and Applied Ecology*, 13(3), 232–240. <https://doi.org/10.1016/j.baae.2012.03.004>
- Taylor, J. N., Courtenay, Jr., W. R., & McCann, J. A. (1984). Known impact of exotic fishes in the continental United States. In *Distribution, biology, and management of exotic fish* (pp. 322–373). Johns Hopkins University Press, Baltimore, MD.
- Templer, P., Findlay, S., & Wigand, C. (1998). Sediment chemistry associated with native and non-native emergent macrophytes of a Hudson River marsh ecosystem. *Wetlands*, 18(1), 70–78. <https://doi.org/10.1007/BF03161444>
- Thompson, D. Q., Stucky, R. L., & Thompson, E. B. (1987). *Spread, impact, and control of purple loosestrife (Lythrum salicaria) in North American wetlands* (p. 55). United States Department of the Interior, Fish and Wildlife Service, Washington, DC.
- United States Geological Survey. (2025). *USGS National Hydrography Dataset Plus High Resolution National Release 2* [FileGDB]. U.S. Geological Survey. <https://doi.org/10.5066/P13V7GVY>
- U.S. Environmental Protection Agency. (2008). *Effects of climate change for aquatic invasive species and implications for management and research* (EPA/600/R-08/014). National Center for Environmental Assessment. Available from the National Technical Information Service, Springfield, VA, and online at <http://www.epa.gov/ncea>.
- U.S. Fish and Wildlife Service. (2024). *Basket Clam (Corbicula fluminea) Ecological Risk Screening Summary*. U.S. Fish and Wildlife Service. Available: <https://www.fws.gov/media/ecological-risk-screening-summary-basket-clam-corbicula-fluminea-high-risk>
- U.S. Geological Survey. (2025). *Watershed Boundary Dataset (published 20250107)*. <https://www.usgs.gov/national-hydrography/watershed-boundary-dataset>
- U.S. Geological Survey. (2026). *Nonindigenous Aquatic Species Database*. <https://nas.er.usgs.gov/>
- Van Rees, C. B., Hand, B. K., Carter, S. C., Barger, C., Cline, T. J., Daniel, W., Ferrante, J. A., Gaddis, K., Hunter, M. E., Jarnevich, C. S., McGeoch, M. A., Morissette, J. T., Neilson, M. E., Roy, H. E., Rozance, M. A., Sepulveda, A., Wallace, R. D., Whited, D., Wilcox, T., ... Luikart, G. (2022). A framework to integrate innovations in invasion science for proactive management. *Biological Reviews*, 97(4), 1712–1735. <https://doi.org/10.1111/brv.12859>
- Walsh, J. A., & Phelps, K. (1991). Development and evaluation of a technique for screening watercress (*Rorippa nasturtium-aquaticum*) for resistance to watercress yellow spot virus and crook-root fungus (*Spongospora subterranea* f. *Sp. Nasturtii*). *Plant Pathology*, 40(2), 212–220. <https://doi.org/10.1111/j.1365-3059.1991.tb02369.x>
- Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10(5), 744–751. <https://doi.org/10.1111/2041-210X.13152>

CHAPTER 2 — #MODELEVALUATION: GENERATING SPATIAL TESTING AND CROSS-VALIDATION SPLITS FOR ECOLOGICAL MODELS USING HASHTAG GEOMETRY

1 Abstract

Credible estimates of model performance are critical for assessing an ecological model's utility for decision-making. Best practices from the literature on spatial ecological modeling emphasize the importance of spatially independent data to improve confidence and credibility in the evaluation of model performance. A growing collection of data splitting methods has the potential to improve performance estimation for spatial ecological models built with structured species occurrence data yet these methods rarely generate withheld test data, preventing robust, independent assessments of models. Here, we present a novel data splitting method using hashtag geometry (#) to simplify the implementation of statistical and ecological modeling best practices by generating multiple spatially distinct data splits. Our method creates spatially blocked cross-validation splits for initial model assessment, hyperparameter tuning, and model selection. The method also generates spatially withheld test data separate from training data that is suitable for the evaluation of final model performance. We compare the performance of this method to two other common splitting approaches (spatially buffered leave-one-out cross validation, random sampling) in a virtual species framework. Using cross-validation splits to tune models, we generated more accurate predictions compared to 'true' probabilities of suitability for a virtual species. Additionally, performance of the tuned models on the hashtag's withheld test data were representative of performance on new data beyond the geographic boundary of training and testing splits. Under the idealized conditions of our virtual species, results based on hashtag and random data splitting were similar, and tuning selected the same hyperparameter settings. These results

indicate our new method can help build accurate modeling workflows, through model selection and tuning, while retaining independent testing data through the spatial splitting best practices now common to ecological modeling.

2 Introduction

Spatial prediction is a central objective of many ecological models, including species distribution models (SDMs) that estimate relationships between occurrence data and environmental attributes to predict species' distributions in geographic space (Elith & Leathwick, 2009). SDM model outputs, including maps and zonal statistics, are used to inform conservation and land management decision making (Sofaer et al., 2019). However, a model validated with non-independent data can systematically overestimate the predictive performance of SDMs (Veloz, 2009) and undermine its perceived credibility.

To develop and validate models, the best practices of the statistical and machine learning communities (Hastie et al., 2009; Kuhn & Johnson, 2019; Lever et al., 2016) advocate splitting data into three tiers (Fig. 1). First, data are split between training and testing, with the former containing all the data used to construct models. Within the training data, cross-validation splits create k-fold subsets that are commonly used to refine and select models through hyperparameter tuning and comparative assessments. The test data split remains fully withheld during the model building and cross-validation processes, enabling an independent evaluation of the performance of a final model. In the absence of internal data structures or biases, random or stratified splits would often suffice at the test/train and train/cross-validation data splitting steps, but ecological data almost always contain some form of spatial, temporal, group, or phylogenetic structure (Roberts et al. 2017).

Previous research has repeatedly demonstrated that random splits overestimate SDM performance (Araújo et al., 2005; Bahn & McGill, 2013; Matutini et al., 2021; Ploton et al., 2020; Roberts et al., 2017; Veloz, 2009) but randomly sampled internal cross-validation splits nevertheless remain commonplace (Hao et al., 2019). The ecological literature now provides modelers with many alternatives to random data splitting (e.g., Frick et al., 2022; Milà et al., 2022; Pohjankukka et al., 2017; Valavi et al., 2019). However, although the advantages of final model performance assessments based on independent data are emphasized (Araújo et al., 2005, 2019; Fourcade et al., 2018), the primary focus and application of new data splitting methods continue to be for cross-validation. The use of cross-validation as a substitute form of final model assessment (in the absence of test data) contrasts with statistical and machine learning norms, where cross-validation folds may provide metrics for internal model assessment but are not considered an ideal, independent evaluation of a final model.

There is no universally applicable approach for splitting testing and training or assigning cross-validation folds to account for all biases and structure inherent in model input data. But data splitting approaches can be designed for spatially structured datasets and for the objective of spatial prediction. Data splitting creates a tension between independence and extrapolation – with spatial blocking, larger blocks can be more independent from each other, but each cross-validation fold may not span the full range of environmental conditions (Roberts et al. 2017). Intentionally encompassing environmental gradients in the spatial configuration of data splits can therefore provide realistic assessments of performance within the range of both geographic and environmental conditions. Predictions beyond the training extent or conditions are notoriously suspect due to extrapolation and changing environmental correlations and context (Thuiller et al., 2004; Yates et al., 2018). An open question is whether spatial tuning and spatially independent

final validation could be tailored to improve and measure performance. We propose that spatial blocks may be useful in tuning hyperparameter settings to optimize for performance in spatial prediction within the extent of its training data, and we test whether this approach can also improve performance when transferring models spatially. While using cross-validation blocks for tuning, it is critical to retain an additional, spatially distinct test split for final model evaluation because data used in cross-validation shape the selected model.

Here, we introduce a novel data splitting method that leverages recent advances in spatial-block sampling design in combination with the unique geometry of the hashtag shape (#) to generate spatially independent data partitions for cross-validation and testing. By spatially buffering vertical and horizontal lines of the hashtag, data can be withheld for final model evaluation, while the nine blocks created between the buffered lines are suitable for cross-validation splits. The hashtag shape is well-suited for ecological modeling because it generates a final test split that covers longitudinal and latitudinal gradients representative of climatic and biogeographical patterns. We use a virtual species framework to test whether the hashtag approach (alone or in tandem with tuning on cross-validation splits) can improve spatial prediction and increase the realism of final model evaluation. We compare the hashtag method to random cross-validation and spatially buffered leave-one-out approaches for a common but difficult model application scenario: predictions made outside the study area in which models are initially trained and tested. The hashtag geometry exemplifies how data splitting approaches can be designed for the goal of spatial prediction, with tuning on cross-validation data to improve model performance and withheld final test data for rigorous evaluation.

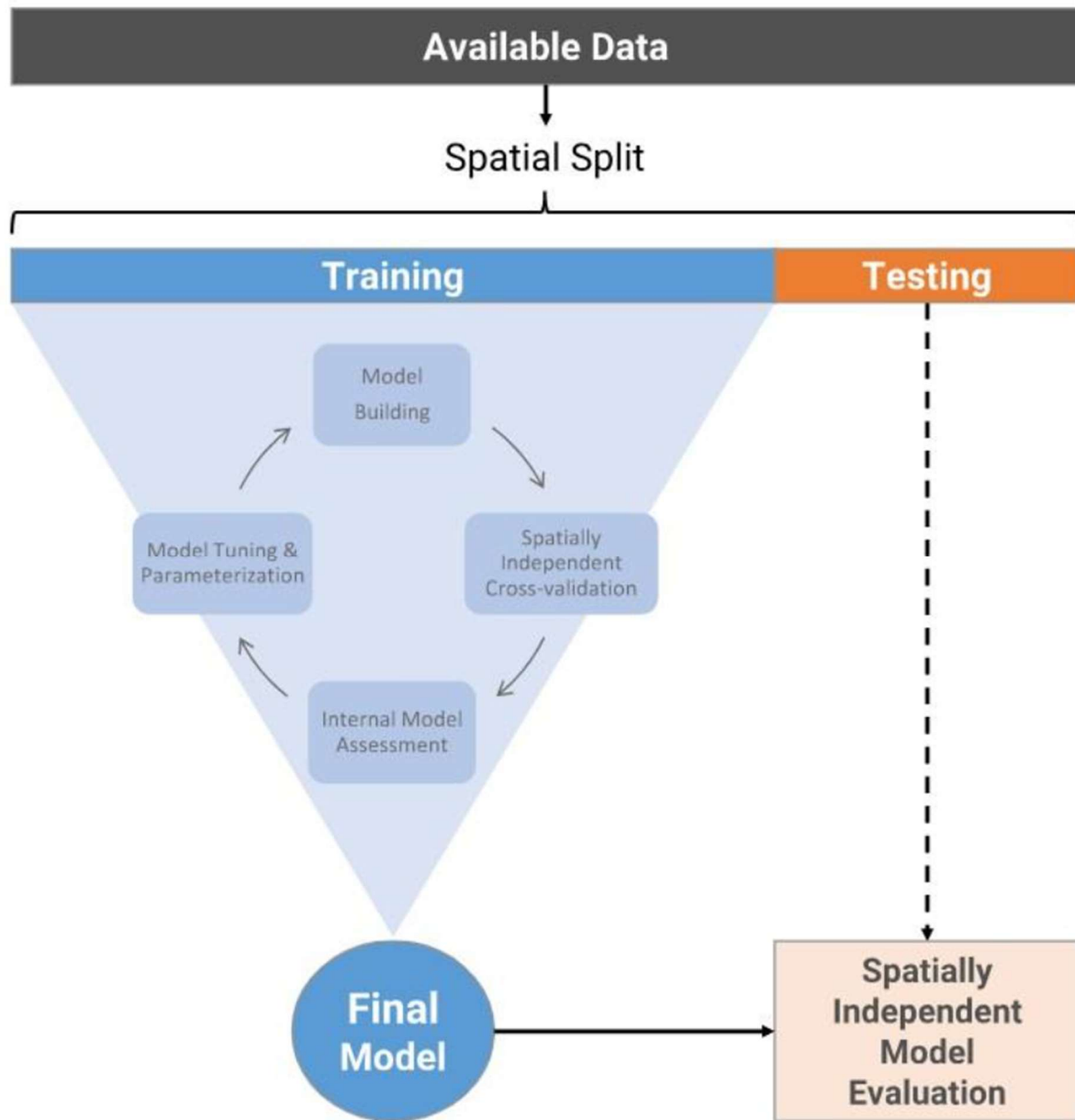


Figure 1. A conceptual diagram illustrating a modeling workflow with training, cross-validation, and testing data. Cross-validation data can be used to tune hyperparameters and select among model structures and algorithms. Spatially independent cross-validation data can be used to develop and tune models to maximize performance at spatial prediction in new areas. A final model can then be evaluated with a separate, spatially independent test dataset.

3 Methods

3.1 Virtual Species

We performed all our analyses in R (R Core Team, 2025; v. 4.4.1), under a virtual species framework (Leroy et al., 2016) using a set of three spatial environmental predictors previously established as important across many SDMs (Williams et al., 2024): mean minimum winter temperature (TMIN; Abatzoglou, 2013), median normalized difference moisture index (NDMI), and a global human modification index (GHM). NDMI is a Landsat-derived, remotely sensed index and, when analyzed with LandTrendr (a Landsat time series trend fitting algorithm; Kennedy et al., 2018), can help describe long-term geographic patterns in landscape water persistence (Halmy et al., 2019). GHM is a composite index reflecting 13 types of anthropogenic stressors on terrestrial habitat (C. M. Kennedy et al., 2019). We defined responses for each predictor: a linear function for GHM to represent linear increases due to human disturbance and Gaussian responses for TMIN and NDMI to represent unimodal climatic tolerances. We combined the predictors using a multiplicative function to generate a “true” continuous suitability surface across the extent of the conterminous United States at a spatial grain of 100 m (Fig. 2-A). From the suitability surface we drew a Bernoulli sample to probabilistically discretize the suitability values into 0 or 1, simulating the presence and absence of a species on the landscape (Fig. 2-B).

3.2 Generating and splitting occurrence data

To generate training/testing data for model building and evaluation, we randomly sampled the binary map ($n = 1,000$ total data points; presences: 295; absences: 705) within a study area selected based on a data boundary polygon from a real-world invasive species (*Vicia cracca*; Jarnevich et al., 2024). Restricting the sampling geography can emulate data collection limitations

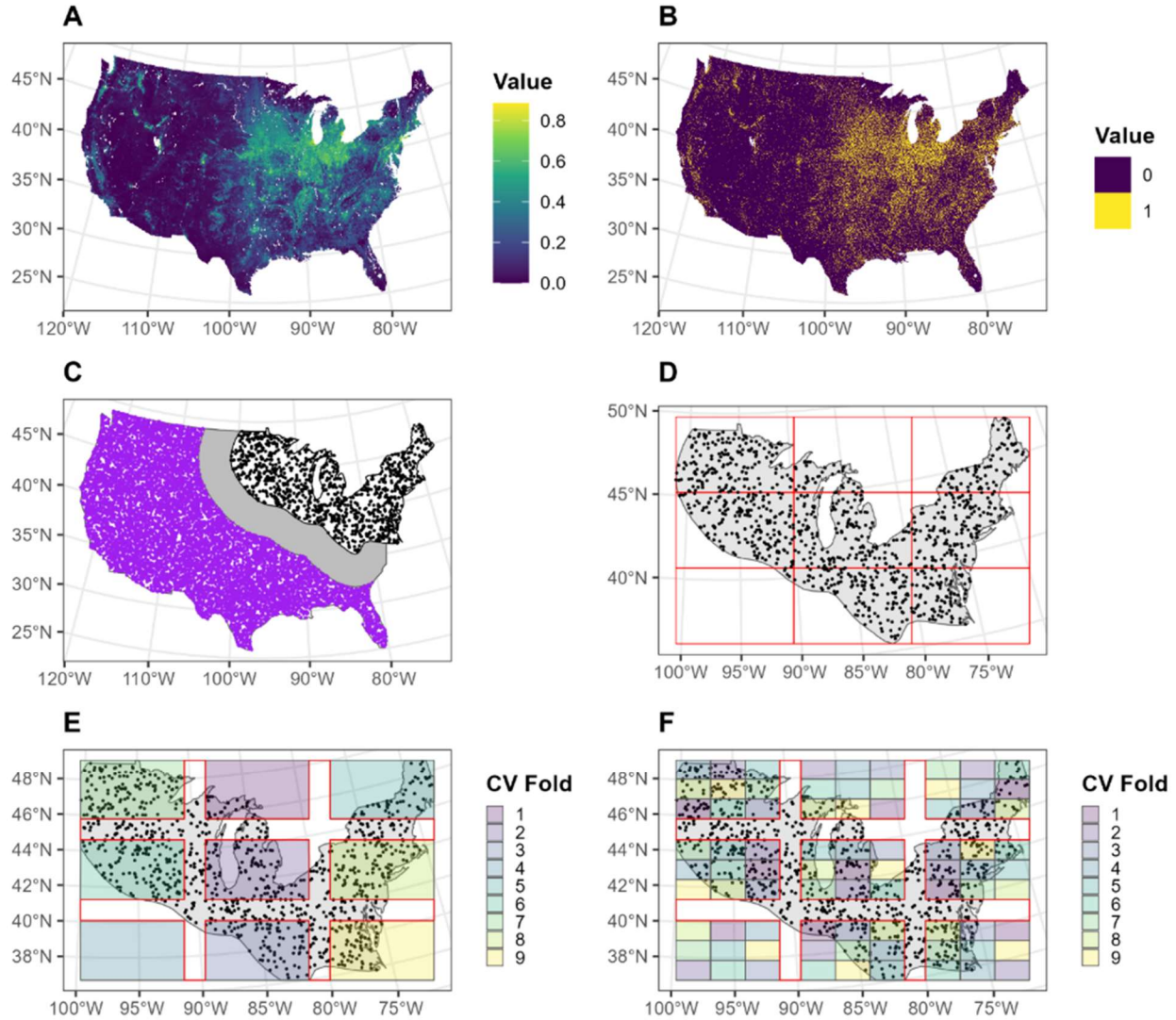


Figure 2. An overview of data generation and hashtag-based data splitting workflow. **(A)** We simulated a virtual species to create a true continuous suitability surface (min: 0.0, max: 1.0) across the conterminous United States. **(B)** From **A**, we created a binary surface (0, 1) using Bernoulli sampling to simulate species presences (yellow) and absences (purple). **(C)** We randomly sampled locations within a defined study area (shown with black points), for model training and internal evaluation. To evaluate the geographic transferability of our models, we created a buffer the size of an estimated range of spatial autocorrelation (shown in solid gray), and sampled external evaluation data, shown in purple, from beyond that range. **(D)** To implement the hashtag method, we divided a bounding box around the study area with an initial hashtag frame as a nine-part grid. **(E)** We buffered the grid from **D** inward to create a hashtag polygon that defines the spatially withheld test data set (outlined in red). Each cross-validation (CV) fold is shown in a different background color. Points for final model testing are those that lie within the red outlines. **(F)** In an optional step, we applied randomly assigned, spatial block cross-validation sub-splitting within non-hashtag areas.

or a distribution extent that does not (yet) include all suitable areas. To evaluate model transferability beyond a region in which models are fit and evaluated, we first buffered the study area boundary polygon by the semivariogram range (~ 402 km) of the model residuals from an initial, exploratory downsampled random forest model (Breiman, 2001; Valavi et al., 2022). With sufficient spatial independence between the study area and the evaluation geography, we randomly sampled new data points ($n = 10,000$) for evaluation of model predictions beyond the spatial buffer but within the contiguous United States (Fig. 2-C).

To split the model building and evaluation data into training, cross-validation, and testing partitions, we used three distinct methods: random (RAND), spatially buffered leave-one-out (LOO; Telford & Birks, 2009), and our new hashtag method (HASH, see below). For RAND, we made an initial split of training (70%) and testing (30%). From the RAND training data, we randomly assigned each data point to one of 10 cross-validation splits. For LOO, we created cross-validation folds for each data point ($k=1,000$) by removing neighboring occurrences from the submodel's training set within a defined radius (402.5 km) based on the previously calculated semivariogram range (Telford & Birks, 2009). For LOO, there was only cross-validation as the method does not generate test data that are withheld from the final model.

For the third method, we introduce hashtag splitting (HASH), which spatially separates data by inwardly buffering nine spatial blocks within a bounding polygon that defines the study area (Fig. 2, D-F). As the buffer increases, a hashtag-shaped polygon grows in the resulting negative space until it overlaps a user-defined test split percentage. The hashtag polygon's geometry allows geographically continuous test data coverage similar to other “test strip” or “spatial strip” methods (Bahn & McGill, 2013; Jarnevich et al., 2022; Sofaer et al., 2022; Wenger & Olden, 2012). The advantage of test strips is their ability to simultaneously capture continuous

environmental variation along both latitudinal (ex. temperature) and longitudinal (ex. aridity in North America) gradients which can minimize extrapolation between training and test splits during model evaluation. Also unique to the hashtag is the joint generation of cross-validation and test spatial regions. Though we do not utilize disjunct distributions in our virtual species example, the HASH method can also be applied to such scenarios, where it will create multiple, unique hashtags and spatial blocks for each disjunct boundary polygon (Appendix 2).

No single data splitting method is universally applicable as each introduces specific tradeoffs between a variety of factors (sample size, number of folds, spatial autocorrelation, extrapolation, computational efficiency). For example, subdividing spatial blocks leads to smaller individual block sizes and hence less spatial independence, but may lead to a better balance in the number of points and more completely capture the range of environmental conditions in each split. Block size and assignment pattern may affect the level of model extrapolation during cross-validation. Therefore, we built two HASH iterations, each with a maximum of nine cross-validation folds and a test data proportion covering 30% of the initial observations. The first iteration (hereafter HASH-NOSUB) retained the original nine blocks while the second iteration subdivided each original block into a further nine blocks, resulting in 81 total, smaller blocks (hereafter HASH-WITHSUB). We measured sample sizes within each cross-validation split for each approach. Lastly, for each splitting method we calculated a multivariate environmental similarity surface values (MESS; Elith et al., 2010) to quantify the level of extrapolation between all training and testing sets and between all training and cross-validation splits.

3.3 Model building and tuning

We built all models in a *tidymodels* workflow (Kuhn & Wickham, 2020) using the lightGBM algorithm (Ke et al., 2017), a highly efficient tree classifier with multiple tunable

hyperparameters. We tuned models with win/loss racing (Kuhn, 2014) available in the `finetune` library in R (Kuhn, 2024) and using cross-validation blocks as resamples to optimize five `lightGBM` hyperparameters: number of leaves, number of boosting rounds, loss reduction, learning rate, and tree depth. Other model hyperparameters were held constant (`ntrees` = 1,000; `mtry` = 2) to reduce tuning dimensionality. We tuned model performance based on the area under the receiver operator curve. Only the HASH and RAND approaches employed tuning with the cross-validation data because they include a separately withheld internal dataset for final model evaluation following tuning.

3.4 Model evaluation

For each method, we measured the performance of the associated model predictions (probability of the presence class) against the values of the continuous suitability surface of the virtual species using root mean squared error (RMSE). All models and methods had predictions to cross-validation splits and to external test data outside the study area (see: Section 3.2 - *Generating and splitting training data*). For RAND and HASH, we also calculated the performance of internal test splits within the study area. Finally, for data beyond the study area, we also calculated RMSE incrementally along a gradient of MESS values representing high to low extrapolation in our spatial model transfer scenario.

4 Results

4.1 Data splitting

Training data for RAND included 700 data points (presences: 294, absences: 496) and 300 for testing (presences: 87, absences: 213) with 10 cross-validation splits (mean presences: 21;

mean absences: 50). Training data for LOO included all 1,000 training data locations with 1,000 cross-validation splits (mean points removed in buffer: 182). Data for both HASH approaches included 684 training data points (presences: 205, absences: 479) and 326 testing points (presences: 85, absences: 216). For HASH-NOSUB, the mean fold sample size was 89 (mean presences: 29; mean absences: 60) with two dropped folds due to a lack of sufficient samples (final splits: 7). For HASH-SUB, the mean split sample size was 76 (mean presences: 23; mean absences: 53).

Environmental dissimilarity between cross-validation folds, training and internal splits, or training and external test data were similar across data splitting methods, as measured by MESS (Fig. 3). For cross-validation, the large block HASH-NOSUB method led to the most extrapolation (i.e., had the lowest MESS values, including some below zero) but for all methods the vast majority of folds were not environmentally dissimilar. Internal test data were rarely dissimilar and median scores were slightly higher than cross-validation medians for RAND or either HASH method. All methods had nearly identical MESS values when comparing training data to external test points; as expected, external test data had lower MESS values than either cross-validation or internal test data. Approximately one third of the external test points had MESS values below zero, indicating environmental extrapolation.

4.2 Model evaluation

Tuning model hyperparameter settings using either HASH or RAND methods led to the same selected hyperparameter values (Appendix 2). RMSE varied across splitting method and whether tuning was performed (Fig. 4) but tuning on cross-validation data (HASH and RAND only) substantially improved performance in both internal and external (geographically transferred) assessments. All untuned models were similar in their performance for cross-

validation median, internal and external test data except for LOO which had better median cross-validation performance relative to its external test performance. Restricting performance measures to the external test points (Appendix 2) with negative MESS values (representing model extrapolation) we found that performance generally decreased as level of extrapolation increased, as expected.

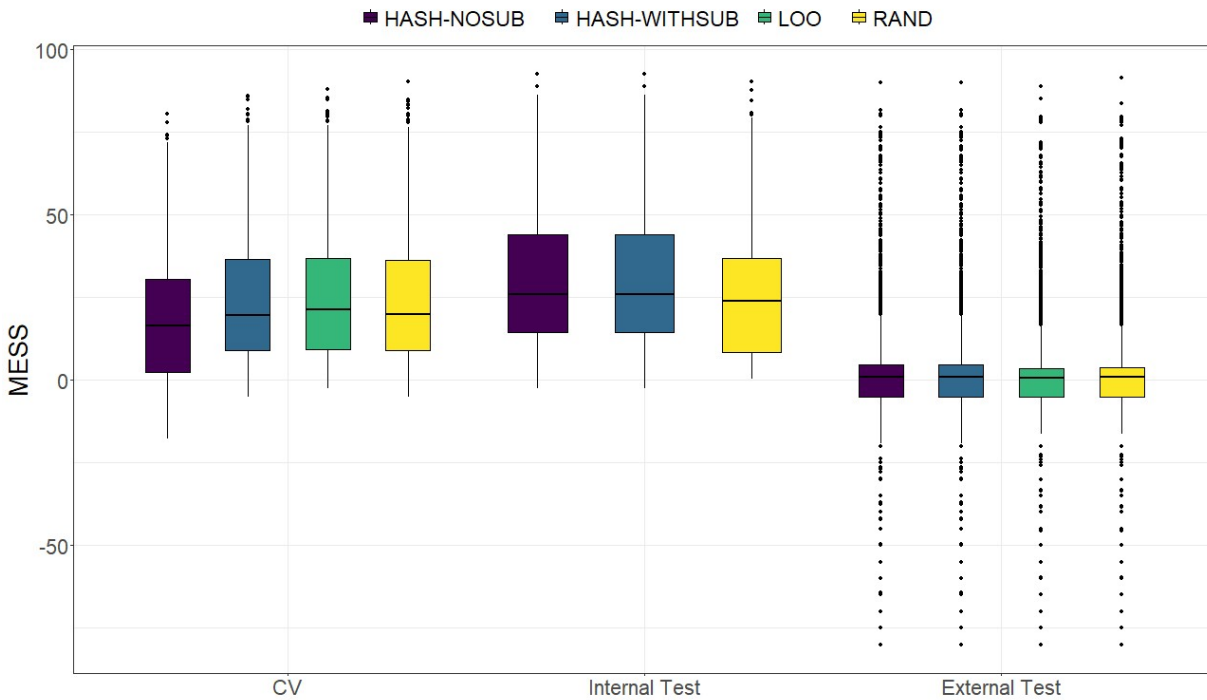


Figure 3. Measurements of the variation in environmental dissimilarity (MESS) between model training data and cross-validation folds (left), internal test data (middle), and external test data (right) for each data splitting method (HASH-NOSUB: hashtag without large spatial blocks; HASH-WITHSUB: hashtag with smaller, subsampled blocks; LOO: buffered leave-one-out; RAND: random). Values below zero represent environmental extrapolation.

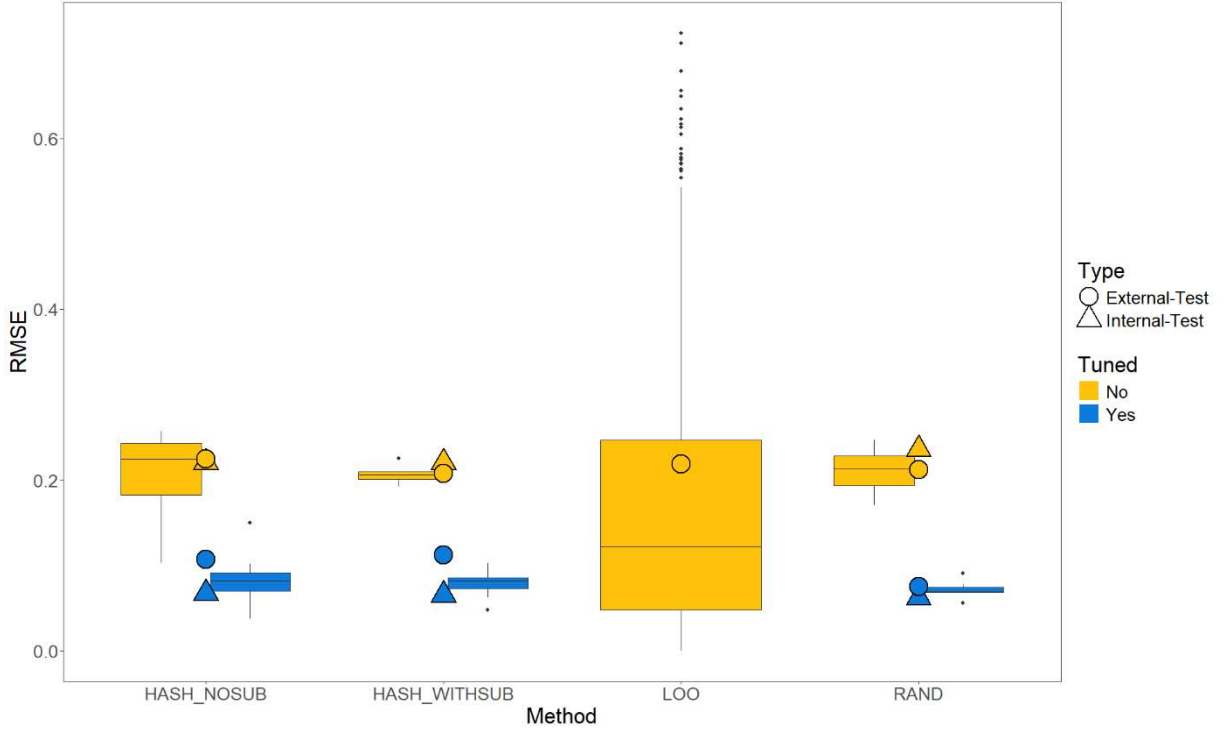


Figure 4. Variation in model performance as measured by root mean square error (RMSE; lower values indicate higher model accuracy) by data splitting method within the study area (“Internal-Test”) and cross-validation (box-and-whisker plots) and when geographically transferred outside the study area (“External-Test”). Performance is derived by comparing the predicted values of tuned and untuned lightGBM models to true probabilities from virtual species. (HASH-NOSUB: hashtag without large spatial blocks; HASH-WITHSUB: hashtag with smaller, subsampled blocks; LOO: buffered leave-one-out; RAND: random).

5 Discussion

We found that our new hashtag method of data splitting could optimize ecological models through hyperparameter tuning similar to existing splitting methods in a machine learning model workflow. Furthermore, the predictive performance of the tuned hashtag method models on spatially withheld test data was a close approximation of the geographically external evaluation in our idealized virtual species simulation. These results highlight the ability of the hashtag method to help address two key challenges in ecological modeling: filling a methodological gap of model

evaluation using withheld test data and improving the accuracy of models intended to be applied beyond a study area.

Using simulated data, we calculated the performance of models built with three data splitting approaches (LOO, random, hashtag) and compared predicted presence-class probabilities to the known probability of occurrence from the virtual species continuous suitability surface. Performance improved when model tuning was performed on a number of lightGBM hyperparameters but tuning did not produce distinct settings between models. As a result, the relative performance of tuned models, though improved, was the same between hashtag and random data splitting approaches. Comparing performance on geographically buffered test data allowed us to emulate a common modeling application: prediction outside a study area. Here, we also explored performance across a gradient of environmental dissimilarity, where performance steadily decreased as environmental conditions became more dissimilar as expected from previous work (Thuiller et al., 2004).

Beyond the matching tuned hyperparameter settings, the similarity of the model results from hashtag and random methods may be attributable to the idealized conditions present in simulated data under our virtual species approach. First, in our study we created models using a random sample of presence and absence points. However, many species distribution models lack absence data and must use presence-only approaches and data that can reflect substantial sampling biases. Additionally, presence-absence models suffer less from sample selection bias compared to presence-background models (Phillips et al., 2009). The generation of our training data set was also not designed to mimic either the systematic sampling bias (Martin et al., 2012) or spatial structure and patterns otherwise common to species occurrence data. Here, we have perfect knowledge of the true level of prevalence (the ratio of presences to total observations) which is

otherwise typically unknown (or in some cases unknowable; Ward et al., 2009). The resulting accuracy of model performance metrics using either real world or simulated data can depend on prevalence (Santika, 2011). Finally, we built the virtual species with a simplified set of environmental variables and trained models on only these variables, which entirely comprise the underlying simulated distribution. Models built with empirical data will typically not be able to encompass all the true drivers of habitat suitability.

Because preparing data for ecological models involves unavoidable tradeoffs between a variety of factors (sample sizes, extrapolation, spatial separation, model transferability, prediction accuracy, computational efficiency), we anticipate that the hashtag method will be flexible to many data and modeling scenarios but will not be optimal or even applicable in all cases. For example, for species that are data-poor, highly clustered, or limited to small spatial extents our method may be more appropriately addressed with methods built for these scenarios (e.g., LOO for small sample sizes). But the global volume of species occurrence locations is increasing exponentially (Heberling et al., 2021) and will eventually necessitate the development of more future-proof data splitting strategies. New methods like the hashtag are less computationally intensive but can still take advantage of increasingly large datasets to hold spatially distinct the information used to tune (cross-validation folds) and evaluate models (testing splits).

Our results provide a demonstration of how formal model tuning based on cross-validation data splits improves performance, which can be evaluated using spatially distinct test data. While further work remains to explore the use of the hashtag method across a variety of taxa and geographies, we anticipate that our approach will provide a flexible data splitting solution that encourages the evaluation of ecological models beyond spatial cross-validation alone by including spatially withheld test data. We focus here on ecological modeling methods (and specifically

SDMs) but integrated spatial data splitting methods (enabling tuning and model selection on cross-validation splits followed by evaluation of the final model on withheld data) can be broadly applied to numerous modeling applications that incorporate spatially structured training data, especially when a goal is model transferability in geographic space.

REFERENCES

- Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1), 121–131. <https://doi.org/10.1002/joc.3413>
- Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaat4858. <https://doi.org/10.1126/sciadv.aat4858>
- Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology*, 11(9), 1504–1513. <https://doi.org/10.1111/j.1365-2486.2005.01000.x>
- Bahn, V., & McGill, B. J. (2013). Testing the predictive performance of distribution models. *Oikos*, 122(3), 321–331. <https://doi.org/10.1111/j.1600-0706.2012.00299.x>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27(2), 245–256. <https://doi.org/10.1111/geb.12684>
- Frick, H., Chow, F., Kuhn, M., Mahoney, M., Silge, J., & Wickham, H. (2022). *rsample: General Resampling Infrastructure* [Computer software]. <https://rsample.tidymodels.org>, <https://github.com/tidymodels/rsample>
- Halmy, M. W. A., Fawzy, M., Ahmed, D. A., Saeed, N. M., & Awad, M. A. (2019). Monitoring and predicting the potential distribution of alien plant species in arid ecosystem using remotely-sensed data. *Remote Sensing Applications: Society and Environment*, 13, 69–84. <https://doi.org/10.1016/j.rsase.2018.10.005>
- Hao, T., Elith, J., Guillera-Aroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, 25(5), 839–852. <https://doi.org/10.1111/ddi.12892>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Model Assessment and Selection. In T. Hastie, R. Tibshirani, & J. Friedman, *The Elements of Statistical Learning* (pp. 219–259). Springer New York. https://doi.org/10.1007/978-0-387-84858-7_7
- Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, 118(6), e2018093118. <https://doi.org/10.1073/pnas.2018093118>
- Jarnevich, C. S., Engelstad, P., Williams, D. A., Shadwell, K. S., Reimer, C. J., Henderson, G. C., Prev  y, J. S., & Pearse, I. S. (2024). *INHABIT species potential distribution across the*

- contiguous United States* (ver. 4.0, June 2024) [Dataset]. U.S. Geological Survey. <https://doi.org/10.5066/P14HNEJF>
- Jarnevich, C. S., Sofaer, H. R., Belamaric, P., & Engelstad, P. (2022). Regional models do not outperform continental models for invasive species. *NeoBiota*, 77, 1–22. <https://doi.org/10.3897/neobiota.77.86364>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S., & Kiesecker, J. (2019). Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Global Change Biology*, 25(3), 811–826. <https://doi.org/10.1111/gcb.14549>
- Kennedy, R. E., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W. B., & Healey, S. (2018). Implementation of the LandTrendr Algorithm on Google Earth Engine. *Remote Sensing*, 10(5), 691. <https://doi.org/10.3390/rs10050691>
- Kuhn, M. (2014). *Futility Analysis in the Cross-Validation of Machine Learning Models* (arXiv:1405.6974). arXiv. <https://doi.org/10.48550/arXiv.1405.6974>
- Kuhn, M. (2024). *finetune: Additional Functions for Model Tuning*. <https://github.com/tidymodels/finetune>
- Kuhn, M., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315108230>
- Kuhn, M., & Wickham, H. (2020). *Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles*. <https://www.tidymodels.org>
- Leroy, B., Meynard, C. N., Bellard, C., & Courchamp, F. (2016). Virtualspecies, an R package to generate virtual species distributions. *Ecography*, 39(6), 599–607. <https://doi.org/10.1111/ecog.01388>
- Lever, J., Krzywinski, M., & Altman, N. (2016). Model selection and overfitting. *Nature Methods*, 13(9), 703–704. <https://doi.org/10.1038/nmeth.3968>
- Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, 10(4), 195–201. <https://doi.org/10.1890/110154>
- Matutini, F., Baudry, J., Pain, G., Sineau, M., & Pithon, J. (2021). How citizen science could improve species distribution models and their independent assessment. *Ecology and Evolution*, 11(7), 3028–3039. <https://doi.org/10.1002/ece3.7210>
- Milà, C., Mateu, J., Pebesma, E., & Meyer, H. (2022). Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. *Methods in Ecology and Evolution*, 2041–210X.13851. <https://doi.org/10.1111/2041-210X.13851>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197. <https://doi.org/10.1890/07-2153.1>
- Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Péliissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping

- models. *Nature Communications*, 11(1), 4540. <https://doi.org/10.1038/s41467-020-18321-y>
- Pohjankukka, J., Pahikkala, T., Nevalainen, P., & Heikkonen, J. (2017). Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science*, 31(10), 2001–2019. <https://doi.org/10.1080/13658816.2017.1346255>
- R Core Team. (2025). *R: A language and environment for statistical computing*. [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, 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. <https://doi.org/10.1111/ecog.02881>
- Santika, T. (2011). Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography*, 20(1), 181–192. <https://doi.org/10.1111/j.1466-8238.2010.00581.x>
- Sofaer, H. R., Jarnevich, C. S., Buchholtz, E. K., Cade, B. S., Abatzoglou, J. T., Aldridge, C. L., Comer, P. J., Manier, D., Parker, L. E., & Heinrichs, J. A. (2022). Potential cheatgrass abundance within lightly invaded areas of the Great Basin. *Landscape Ecology*. <https://doi.org/10.1007/s10980-022-01487-9>
- Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., Edwards, T. C., Guala, G. F., Howard, T. G., Morissette, J. T., & Hamilton, H. (2019). Development and Delivery of Species Distribution Models to Inform Decision-Making. *BioScience*, 69(7), 544–557. <https://doi.org/10.1093/biosci/biz045>
- Telford, R. J., & Birks, H. J. B. (2009). Evaluation of transfer functions in spatially structured environments. *Quaternary Science Reviews*, 28(13–14), 1309–1316. <https://doi.org/10.1016/j.quascirev.2008.12.020>
- Thuiller, W., Brotons, L., Araújo, M. B., & Lavorel, S. (2004). Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, 27(2), 165–172. <https://doi.org/10.1111/j.0906-7590.2004.03673.x>
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Aroita, G. (2019). block CV: 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(2), 225–232. <https://doi.org/10.1111/2041-210X.13107>
- Valavi, R., Guillera-Aroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs*, 92(1). <https://doi.org/10.1002/ecm.1486>
- Veloz, S. D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36(12), 2290–2299. <https://doi.org/10.1111/j.1365-2699.2009.02174.x>
- Ward, G., Hastie, T., Barry, S., Elith, J., & Leathwick, J. R. (2009). Presence-Only Data and the EM Algorithm. *Biometrics*, 65(2), 554–563. <https://doi.org/10.1111/j.1541-0420.2008.01116.x>
- Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3(2), 260–267. <https://doi.org/10.1111/j.2041-210X.2011.00170.x>

- Williams, D. A., Shadwell, K. S., Pearse, I. S., Prevéy, J. S., Engelstad, P., Henderson, G. C., & Jarnevich, C. S. (2024). Predictor importance in habitat suitability models for invasive terrestrial plants. *Diversity and Distributions*, e13906. <https://doi.org/10.1111/ddi.13906>
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*, 33(10), 790–802. <https://doi.org/10.1016/j.tree.2018.08.001>

CHAPTER 3 — OPENING THE BLACK BOX: USING LOCAL FEATURE IMPORTANCE TO INTERPRET, DIAGNOSE, AND COMMUNICATE THE PREDICTIONS OF ECOLOGICAL MODELS¹

1 Abstract

The field of ecological modeling has largely kept pace with, and taken cues from, advances in machine learning. But the adoption of increasingly complex algorithms is steadily making modeling methods less transparent. For ecologists, this lack of visibility into model behavior means that spurious correlations or ecologically implausible logic may remain hidden, endangering model credibility. While the field of interpretable machine learning is rapidly developing methods and tools to improve the interpretation of models and model predictions, surprisingly few new approaches have been adopted into the best practices of species distribution model (SDM) development. Improving the understanding of the drivers of prediction is particularly important for SDMs built to identify the location and amount of potentially suitable habitat for invasive species. As invasive species continue to spread in their non-native range, existing SDMs will often be used to make predictions in new geographic areas. For this reason, SDMs of invasive species may be a useful testing ground to explore the ability of post-hoc analysis methods from interpretable machine learning to draw larger inferences about model behavior beyond the initial modeling domain. But interpretable machine learning is a young and rapidly developing field and the utility of the field's methods for analyzing SDMs remains underexplored. Here, we hypothesized that a focal SDM would use spatial environmental predictor variables (features) differently when predicting to model training data compared to predictions on new data

¹ Anticipated co-authors: Catherine S. Jarnevich, Rich Inman, Ian S. Pearse, Janet Prev  y, Jeffrey Morisette, Jody C. Vogeler

(not used to train the model). To test this, we compared the relative importance of model features within a focal SDM of the invasive annual grass medusahead (*Taeniatherum caput-medusae*) to the application of the focal SDM to nine counties across the western United States. We found partial support for our hypothesis where the correlation structure and distributions of feature importance differed while relative importance and response curves remained stable from model to county-level data. We discuss the implications of our results as they pertain to modelers and model users to address how summaries and maps of local feature importance can enhance the analysis and implementation of SDMs.

2 Introduction

Ecological modeling methods are becoming increasingly complex. More complex modeling methods allow the use of messy, large, non-linear datasets to help explain biological phenomena and make predictions in a wide array of applied ecological contexts. But model complexity also makes the interpretation of model behavior and decision making behind predictions less transparent. This is problematic as model developers and users of model outputs then have fewer ways of distinguishing models that plausibly reflect underlying ecological processes from those built on spurious correlations (Bell & Schlaepfer, 2016; Fourcade et al., 2018). A further darkening of the proverbial “black box” subsequently endangers the credibility of future ecological models if modelers cannot critically examine model drivers and perform robust model validation or interpretation.

A widely used method for assessing how features (also called covariates, predictors or variables) influence model predictions is by measuring changes in model accuracy while randomly re-arranging the values of a focal feature (permutation-based feature importance; Molnar, 2025). This approach is built into machine learning models like random forests (Breiman, 2001) due to

its usefulness in describing feature importance relative to model performance. Because permutation-based methods rely on model accuracy, they cannot be used to compute variable importance on out-of-sample values (new information not included in training data) without comprehensive validation data.

A lack of methods and tools has previously limited the ability of ecological modelers to critically examine feature importance for out-of-sample predictions beyond a handful of model-specific solutions (e.g., Hooper & Kennedy, 2018). The developing area of interpretable machine learning (also called “explainable artificial intelligence”) offers new post-hoc, model-agnostic ways to explore feature importance to help improve the understanding of model behavior when used for prediction on out-of-sample data (Molnar et al., 2020). Despite their common use as prediction tools in applied ecological research, species distribution models (SDMs) are primarily assessed with methods designed to interpret model behavior using in-sample data. Recent work has started demonstrate the utility of interpretable machine learning methods to validate, evaluate, or interpret SDMs (Bourhis et al., 2023; Cha et al., 2021; He et al., 2022; Ryo et al., 2021; Song et al., 2025; Waldock et al., 2024; Zbinden et al., 2026) but more research is needed.

In the task of quantifying the potential establishment of invasive species, SDMs provide information that can inform and enhance land management actions. For preventative efforts, SDMs can be used to anticipate and respond to the expansion of invasives species (Reaser et al., 2020) but these efforts often require continuous predictions on predominantly out-of-sample data. Modelers and managers subsequently need tools to thoroughly assess transferability for SDMs. Medusahead (*Taenitherum caput-medusae*) is an invasive winter annual grass with serious environmental and economic impacts, a long invasion history, and continued spread. Predicting suitability for medusahead in areas where it has not yet become established could be a useful test

case to better understand the use of post-hoc analysis of interpretation methods for SDMs. But the use of interpretable machine learning remains underexplored for SDMs. Specifically, there is a knowledge gap pertaining to the geographic patterns and trends of feature importance for models of invasive species.

To address this knowledge gap, we calculated and mapped feature importance using a model-agnostic interpretable machine learning method (Shapley Additive exPlanations; SHAP) for an SDM of medusahead occurrence. We asked whether SHAP data from model predictions across nine counties across the western United States (hereafter, local feature importance) differed from SHAP data based on predictions of model training data alone (hereafter, model feature importance). We hypothesized that variation in local environmental conditions would be sufficient to cause model and local feature importance to diverge. To test this hypothesis, we compared the correlation structure, relative importances, distributions, and responses of SHAP data across features to scrutinize the differences and similarities between the model and local feature importance values. We discuss the implications of these results for modelers and model users to contextualize the utility of local feature importance for model interpretation and validation.

3 Methods

3.1 Focal species

To investigate feature importance in a specific biological context, we chose medusahead (*Taenitherum caput-medusae*) as a focal study species. Medusahead is an invasive annual grass of management concern across the Western United States (Young, 1992) that invades rangelands, reducing native plant biodiversity (Davies, 2011), providing low quality forage for livestock and wildlife (Dahl & Tisdale, 1975), and altering ecosystem function (Sheley et al., 2008).

Medusahead is commonly reported in areas with hot, dry summers and wetter springs, falls, and winters (Young, 1992) and is consistently noted for its ability to establish in high clay soils (Nafus & Davies, 2014; Sheley et al., 2008; Young, 1992). Despite a 140-year invasion history, publicly available observation data show that medusahead continues to spread in its non-native range (Jarnevich et al., 2021), highlighting the ongoing need for spatially explicit predictions of potential establishment to inform and enhance prevention and mitigation efforts.

3.2 Sampling units

To understand the potential variation of local feature importance in differing environmental contexts, we opportunistically selected nine counties across the western United States (Fig. 1; Table 1) as focal sampling units. The sampled counties are large enough to generate a robust sample (size range: 5,125 km² to 7,736 km²) but small enough to visually explore local ecological features and landscape processes. The samples vary along geographic (latitude and longitude) and ecological gradients (ecoregions; Omernik & Griffith, 2014) and by the level of predicted suitability. Given medusahead's continued spread across the Western United States, we included counties with and without established populations, based on publicly available observations.

3.3 Focal model

To generate feature importance values, we selected an existing generalized linear model (GLM; McCullagh & Nelder, 1989) of medusahead previously used to predict suitability for occurrence across the contiguous United States (Jarnevich et al., 2024). The model reflects SDM best practices by using eleven spatial covariates that are uncorrelated (Dormann et al., 2013) and selected to reflect species-specific factors relevant to the establishment and persistence of

medusahead at a given location (Elith, 2017). Simple, parsimonious models like GLMs provide efficient prediction

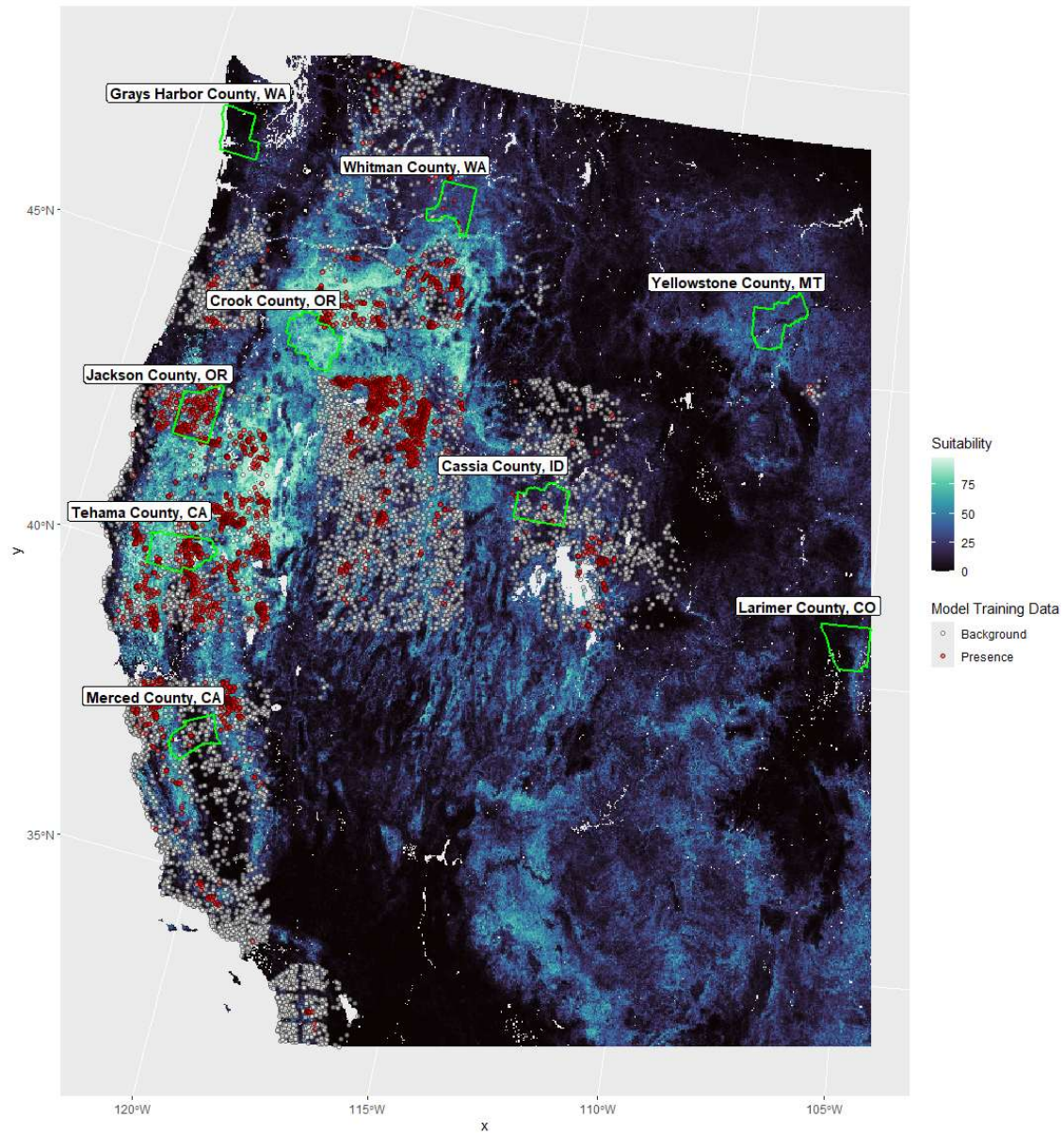


Figure 1. Focal sampling units ($n = 9$ counties) across the Western United States with publicly available observations of medusahead (*Taenitherum caput-medusae*) occurrences and sampled background (pseudoabsence) locations used to fit a species distribution model overlaid on predicted suitability.

Table 1. The sampled counties used to explore local feature importance. Columns **B.** and **C.** display latitude (LAT) and longitude (LON) values from county centroids rounded to the nearest degree. Column **E.** displays the count of spatially thinned presence locations (used to build our focal model) within the county. Column **F.** displays zonal statistics for each county, summarizing the mean and standard deviation (SD) of suitability predicted by the focal model.

A. County Name & State Abbreviation	B. LAT	C. LON	D. Ecoregion (Level II)	E. Number of Observations	F. Suitability Mean $\pm$ SD
Cassia, ID	42	-113	Cold Deserts	9	21.9 $\pm$ 11.4
Crook, OR	44	-120	Western Cordillera	322	65.2 $\pm$ 16.8
Grays Harbor, WA	47	-124	Marine West Coast Forest	0	4.3 $\pm$ 6.2
Jackson, OR	42	-123	Western Cordillera	186	54.5 $\pm$ 17.6
Larimer, CO	40	-105	Western Cordillera	0	10.0 $\pm$ 12.7
Merced, CA	37	-120	Mediterranean California	9	19.2 $\pm$ 21.4
Tehama, CA	40	-122	Mediterranean California	163	64.0 $\pm$ 20.1
Whitman, WA	47	-117	Cold Deserts	16	36.1 $\pm$ 14.7
Yellowstone, MT	46	108	West-Central Semiarid Prairies	0	30.6 $\pm$ 10.6

methods and are thought to be more robust than complex models when transferred in space and time (Coelho et al., 2019; Wenger & Olden, 2012 but see Valavi et al., 2023). While GLMs may not be as complex as other machine learning approaches, the post-hoc model interpretation of local predictor importance for SDMs of invasive species remains critically underexplored.

3.4 Shapley calculations

We measured feature importance using Shapley values (Shapley, 1952) a model-agnostic approach to calculate the marginal contribution of features to model predictions (Molnar, 2025). In practice, Shapley values are usually approximated using estimation methods like Shapley Additive exPlanations (SHAP; Lundberg & Lee, 2017; Štrumbelj & Kononenko, 2014) due to the

computational intractability of evaluating all possible feature combinations as the number of features increases. For SDMs, SHAP values can be understood as a way to quantify the contribution of each feature to the model's predicted suitability at a given location relative to the mean predicted suitability from reference conditions (typically model training data or a small subset thereof). For each feature, positive SHAP values indicate an attempt to increase the prediction and negative SHAP values an attempt to decrease it. The final predicted suitability is then the sum of all SHAP values plus the mean expected suitability from the reference conditions.

We performed all calculations and analyses in R 4.5.2 (R Core Team, 2025). To generate SHAP values at the model and county level, we used *fastshap* (Greenwell, 2026), which implements a user-defined number of Monte Carlo simulations that randomize feature ordering and feature value permutations necessary to converge on an approximation of true Shapley values. All calculations used a consistent reference set of 50 observations from model training data for permutation within 100 simulations (after which SHAP values converged on stable values). For each county, we mapped SHAP values for each feature ($n = 11$; Table 2) to visualize the spatial variation in local feature importance across the landscape. From maps of individual feature values, we created two types of aggregate, categorical maps representing only the features that had the maximum or minimum values at each pixel location. We summarized the counts from these maps to assess the relative proportion of features for minima and maxima across each county. We also calculated the correlation of SHAP values in the model and within each county using Pearson correlation coefficients to identify potential dependencies among features. Finally, we used the categorical maps to identify where the most important local features within counties could be used to further delineate predicted suitability.

Table 2. List of model features. Original resolutions are listed but all features have been resampled to common spatial resolution of ~ 100 m (Jarnevich et al., 2024).

Feature Name	Description (Date Range)	Original Resolution	Data Source	Reference
Bare Ground	Standard deviation of percent cover of bare ground (2000-2016)	250 m	MODIS	DiMiceli et al., 2015
Burn Frequency	The number of years with at least one recorded fire within a pixel (1984-2018).	4000 m	MTBS	Eidenshink et al., 2007
Min Winter Temp	The monthly mean minimum temperature (December through February), averaged across years (1981-2018).	800 m	GridMET	Abatzoglou, 2013
Monthly Temp Range	The mean of monthly temperature ranges (bio2) (1981-2010).	800 m	PRISM	PRISM Group, 2026
NDMI	Median Normalized Difference Moisture Index between April 1st and September 30th across year (1985-2020).	30 m	Landsat Google Earth Engine	Halmy et al., 2019 Kennedy et al., 2018
Soil Clay	Clay content (0-2 μ m) mass fraction at a depth of 0.05 m (2017).	100 m	SoilGrids	Ramcharan et al., 2018
Soil Depth	Mean depth to restriction layer (2017).	100 m	SoilGrids	Ramcharan et al., 2018
Soil Potassium	Potassium content in ppm at depth 0.05 m (2017).	100 m	SoilGrids	Ramcharan et al., 2018
Spring PET	Mean potential water deficit of early spring months (March through May) across years (1981-2018).	4000 m	GridMET	Abatzoglou, 2013
Summer PET	Mean potential water deficit of summer months (June through August) across years (1981-2018).	4000 m	GridMET	Abatzoglou, 2013

Topo Diversity	The variability in moisture and temperature conditions within local habitats (2015).	90 m	Conservation Science Partners	Theobald et al., 2015
----------------	--	------	-------------------------------	-----------------------

3.5 Feature importance comparisons

Ranked feature importance is an efficient, common method to determine the relative impact of features on model predictions. In ecological models, ranked importance can be assessed to determine if the primary drivers of the model are ecologically realistic given the environmental context of the observations and biology of the focal species. Within each county, we assessed importance rankings by calculating the mean absolute value of rescaled importance (range: 0-100). To measure the variation in rank, we took random bootstrap samples ($n = 100$) of pixels within each county using a sample size equivalent to the training data size used in our focal model ($n = 13,000$). Within each sample, we calculated feature ranks using the same method described above.

Mean absolute importance ranks can obscure the fact that the SHAP values can be distributed positively or negatively. Therefore, examining raw SHAP values is critical to interpretation as the sign indicates whether the feature is influencing the model toward higher or lower predicted values. We noted sign changes between mean SHAP values of model and county data for each feature and calculated two-sample Kolmogorov-Smirnov tests for each feature to determine if the distributions were significantly different between model and local data. Our expectation was that both comparisons could highlight possible instances where local environmental conditions changed model prediction behavior.

While ranks and distributions of importance values reveal ordering and variation among features, we also wanted to investigate how feature importance values varied in response to SHAP values. The shape and directionality of feature response curves can be evaluated for potential model overfitting and to confirm the ecological plausibility in the context of focal species. In the

SDM literature, this practice usually involves feature responses to predicted suitability but has rarely been explored where models are used to make predictions on out-of-sample data. To understand the implications of local feature importance on model development, we assessed how well model-based and local response data reflected expectations from existing information on medusahead biology and site preference.

4 Results

4.1 Local importance values

The correlative structure of local importance values varied by county (Fig. 2). Mean absolute correlation values ranged from 0.23 (Yellowstone County) to 0.36 (Whitman County) with high variability ($SD \pm 0.27$ to 0.30). All counties had at least one correlation > 0.7 suggesting that some features in local environmental contexts are not independent and may describe the same ecological process or habitat condition underlying medusahead establishment potential. Absolute correlation values were similar in the model (mean: 0.23, SD : 0.26) but never exceeded 0.4. For maps of the “most important” features in each county, we determined which feature had the highest and lowest value in each gridded pixel to summarize the proportion of the county land area most influenced by each feature toward both positive and negative suitability directions. Among both high and low values, soil potassium had the highest mean proportional area across counties but with high variability (Fig. 3). For some counties, a small number of features were most important. For example, in Merced County, summer PET and NDMI accounted for the vast majority of highest importance values (Fig. 4), but within the county the raw importance values for these features were also highly correlated (0.76; Fig. 2). Across maximum and minimum maps (Fig. 5), we qualitatively observed that changes in the most important feature often mirrored changes in

land cover (e.g., agricultural to undeveloped land) and changes in topographic relief (e.g., mountains to river valleys).

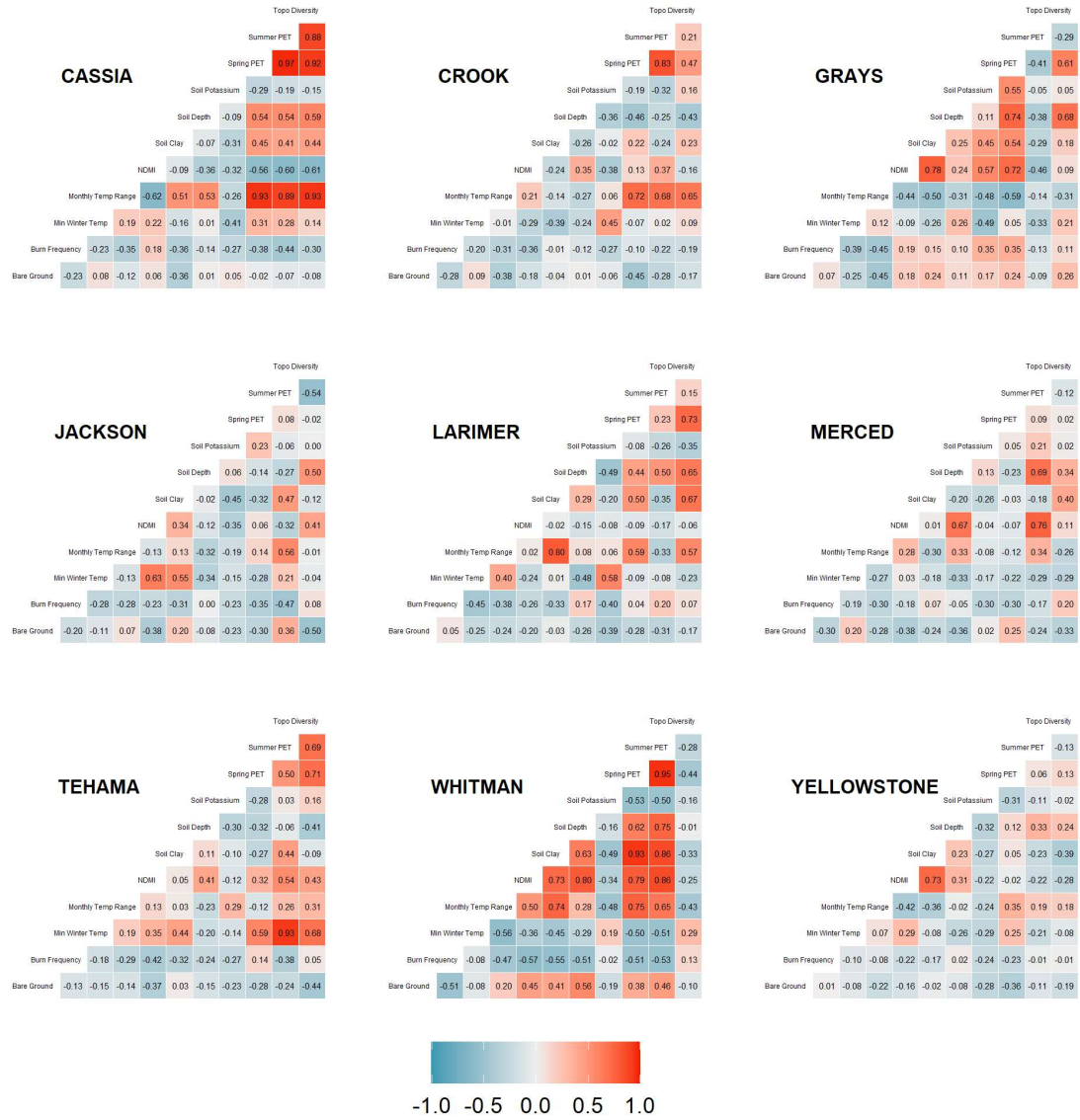


Figure 2. Correlations (Pearson's coefficient) among local feature importance values within counties across the Western United States for a focal model predicting the suitability of occurrence for medusahead.

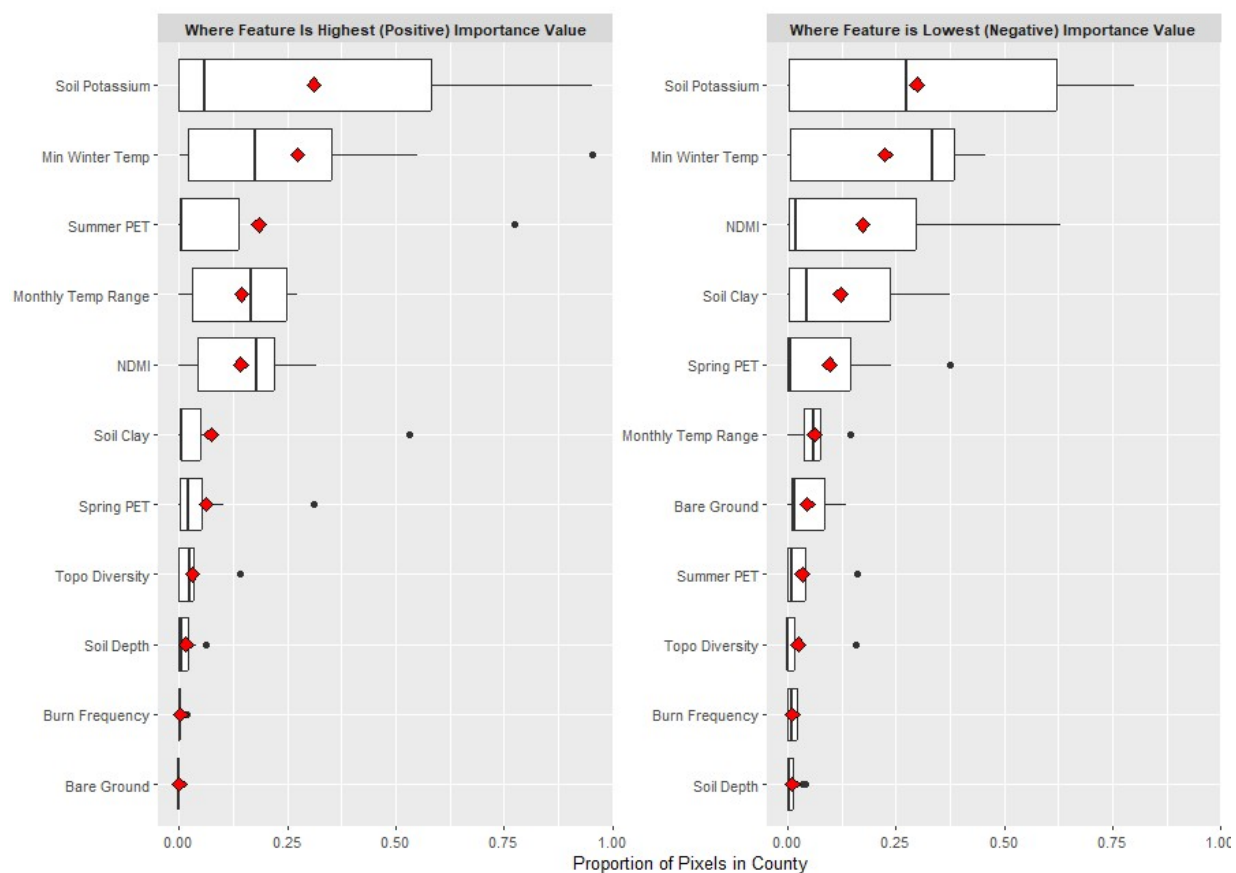


Figure 3. Variation in proportional area of model features when they are either the maximum (highest positive) or minimum (lowest negative) pixel value across counties ($n = 9$). The y-axis is ordered from top to bottom by mean value (red diamonds), and the distribution medians are indicated by a vertical black bar.

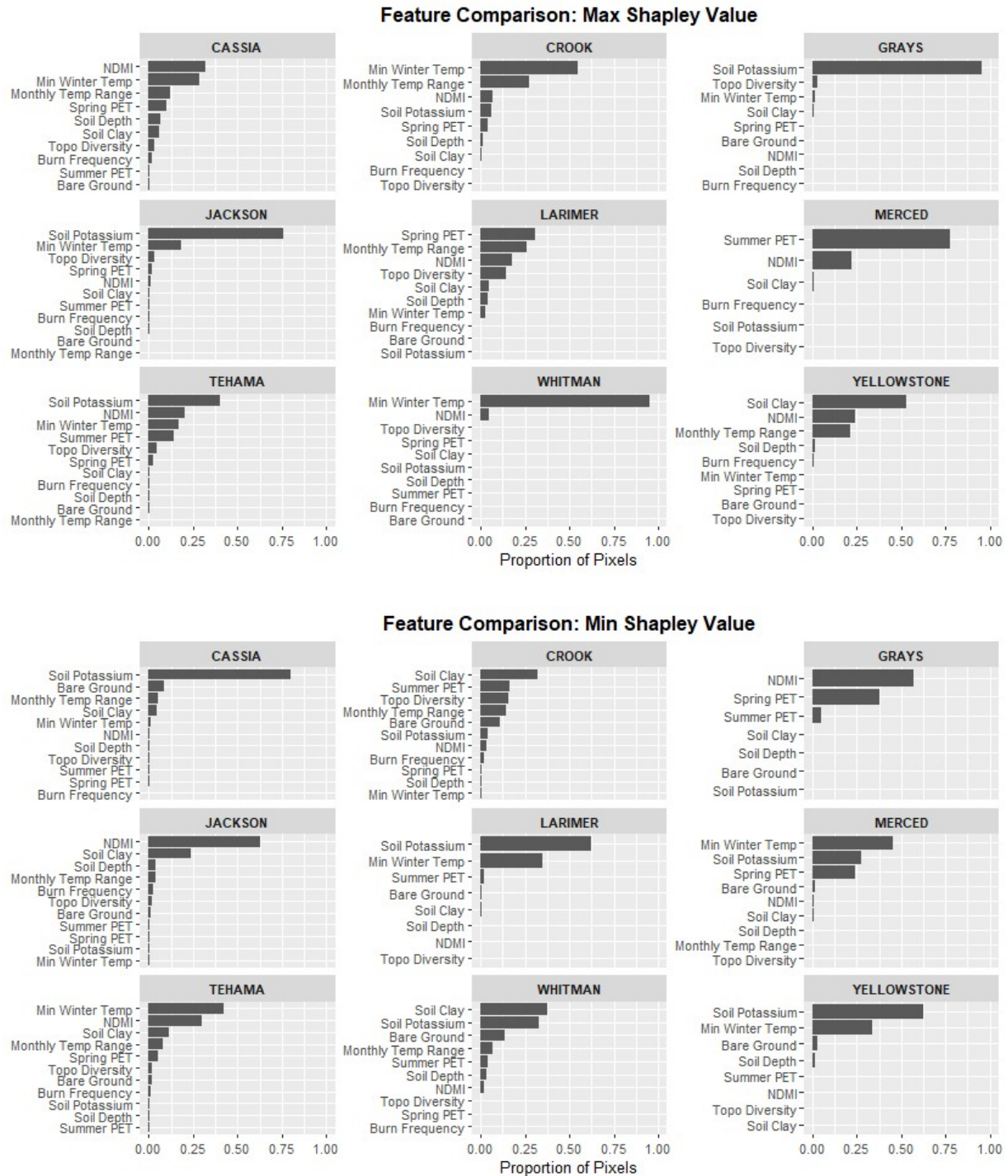


Figure 4. Relative spatial coverage of features (proportion of pixels within a focal county) where they represent the pixel-wise minima or maxima. Max Shapley values indicate where the feature is influencing the model toward a higher suitability prediction. Min Shapley values indicate that the feature is influencing the model toward lower suitability predictions.

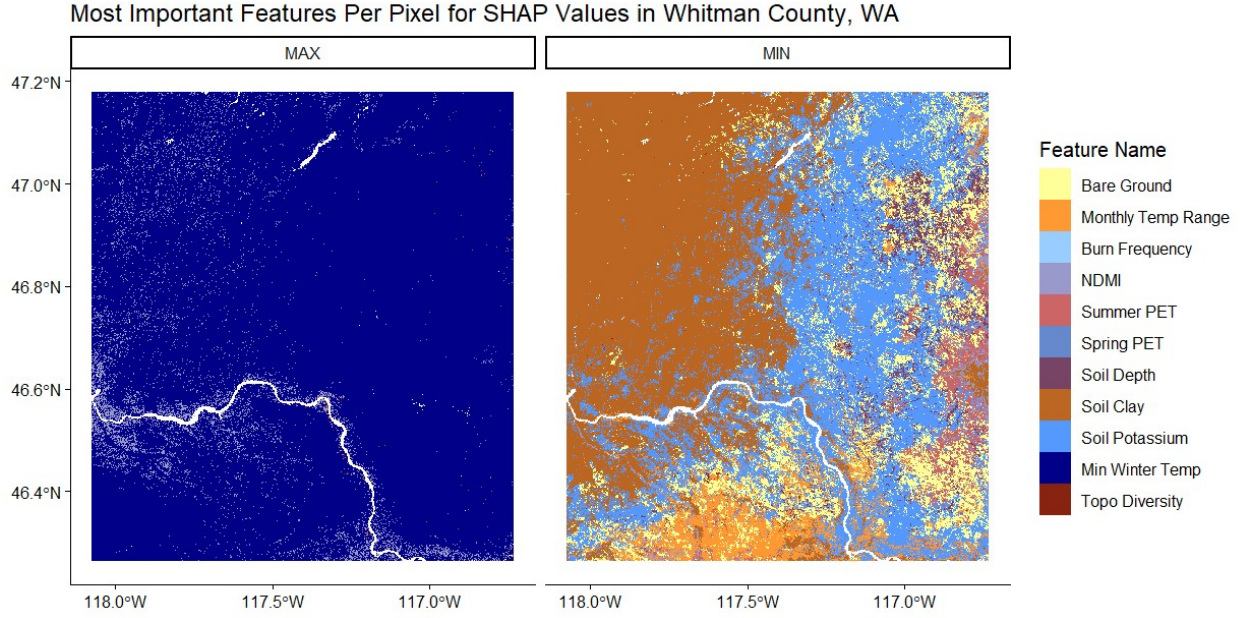


Figure 5. An example of categorical maps displaying the most important features based on per-pixel minimums and maximums for locally mapped SHAP values in Whitman County, Washington, USA.

4.2 Comparisons of model and local importance

The level of agreement between model and bootstrap samples of local data for relative importance varied (R^2 0.45 – 0.85), with model values generally higher than local values (Fig. 6). Despite their small sample size ($n = \sim 13k$), bootstrap sample means of relative importance by feature showed very little variation or difference from overall county means (Fig. 6). In both the model and counties, all raw feature importance values were non-normally distributed ($p < 0.05$; Anderson-Darling normality test). All per feature paired distribution tests indicated significant differences between model and local SHAP values ($p < 0.001$), indicating consistent changes in local feature influence on model prediction. Relatedly, distribution medians often changed signs between model and local data (Fig. 7), indicating shifts in the direction of features influence were common while varying in magnitude. We observed similar shapes and directionality in feature responses of SHAP values between model and local values across all counties (Fig. 8). Divergence

from model responses primarily occurred at SHAP values below zero with larger divergences at local range extremes. We also observed changes in response shape when local values extended beyond the range of model values.

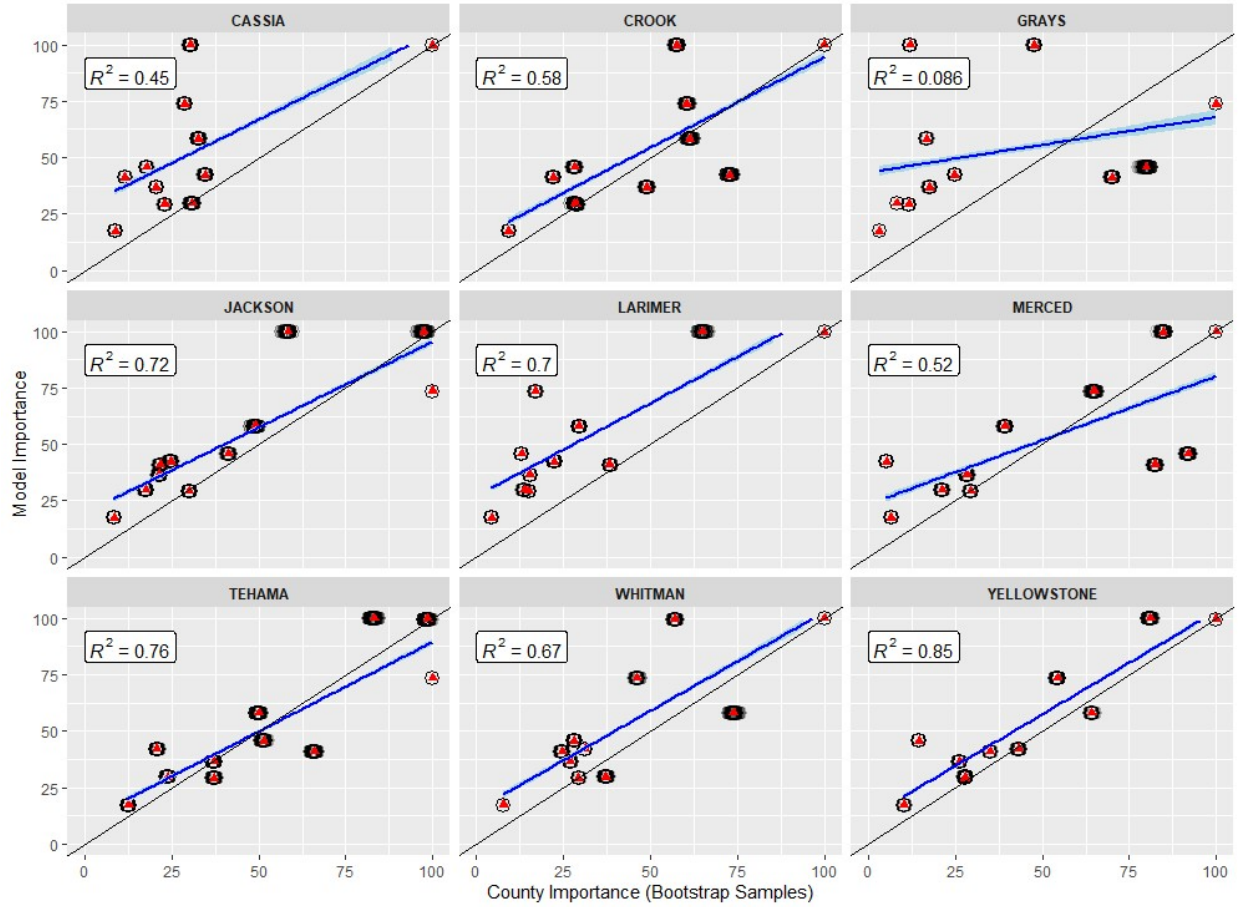


Figure 6. Relationships between relative ranks of model and county-level feature importance. Relative values were obtained by rescaling from mean absolute importance to a range of 0-100 to account for the magnitude of difference between values. Black circles represent mean values from 100 bootstrap samples ($n = 13,000$ in each sample), and the red triangles represent means across entire counties. R^2 values are for linear models using bootstrap samples. The blue lines represent best-fit and the black lines represent a hypothetical 1:1 fit for comparison.

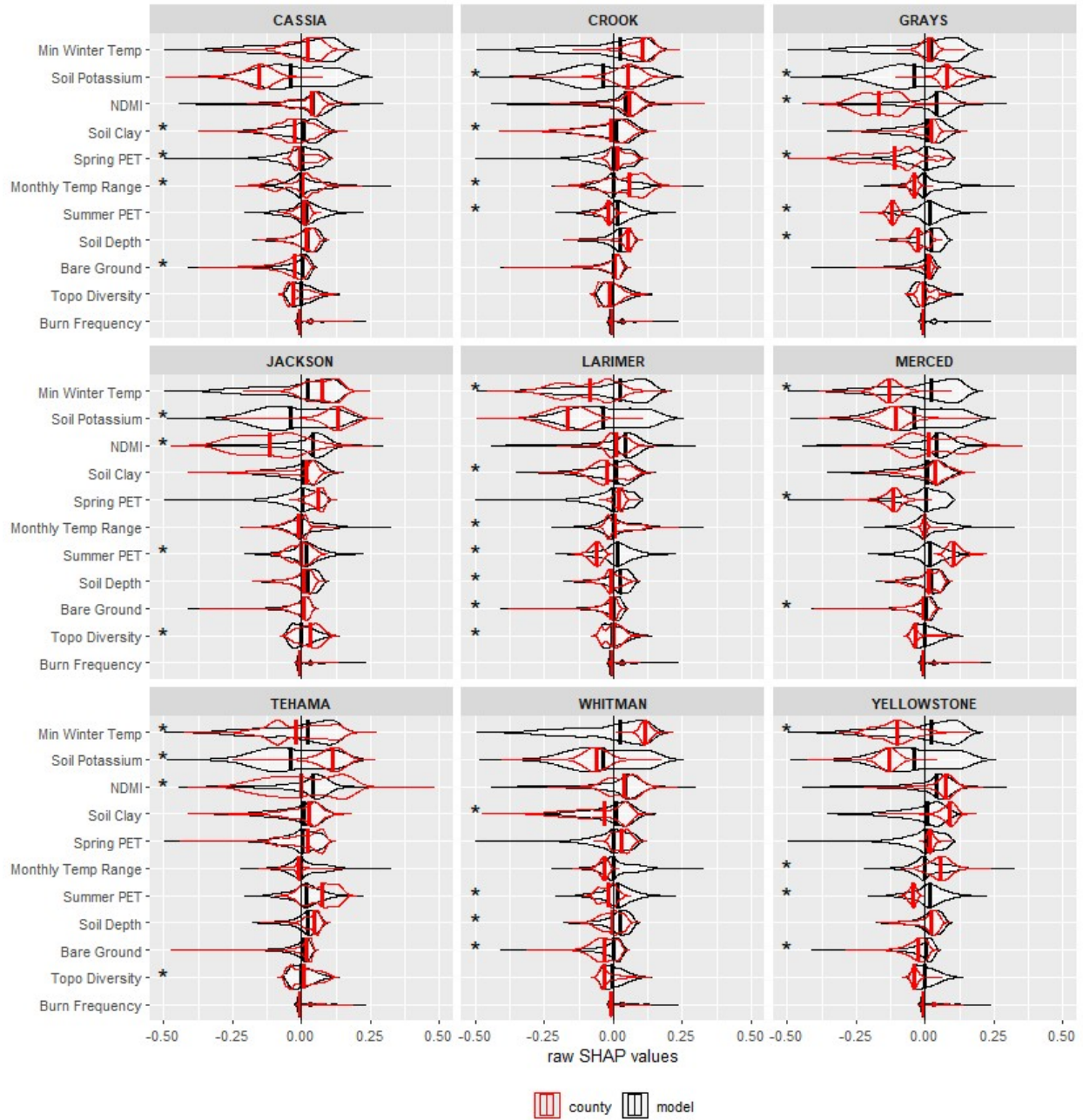


Figure 7. Comparisons of raw SHAP importance value distributions between model and county. Negative values indicate that a feature influences the model toward a lower prediction, and positive values indicate that a feature influences the model toward a higher prediction. Values closer to zero indicate the feature has a minimal effect on model prediction. Feature names on the y-axis are ordered by the rank of mean absolute SHAP values in the model. Distribution medians are indicated by vertical bars and features that flip signs are indicated with a “*.” All paired distributions are significantly different ($p < 0.001$).

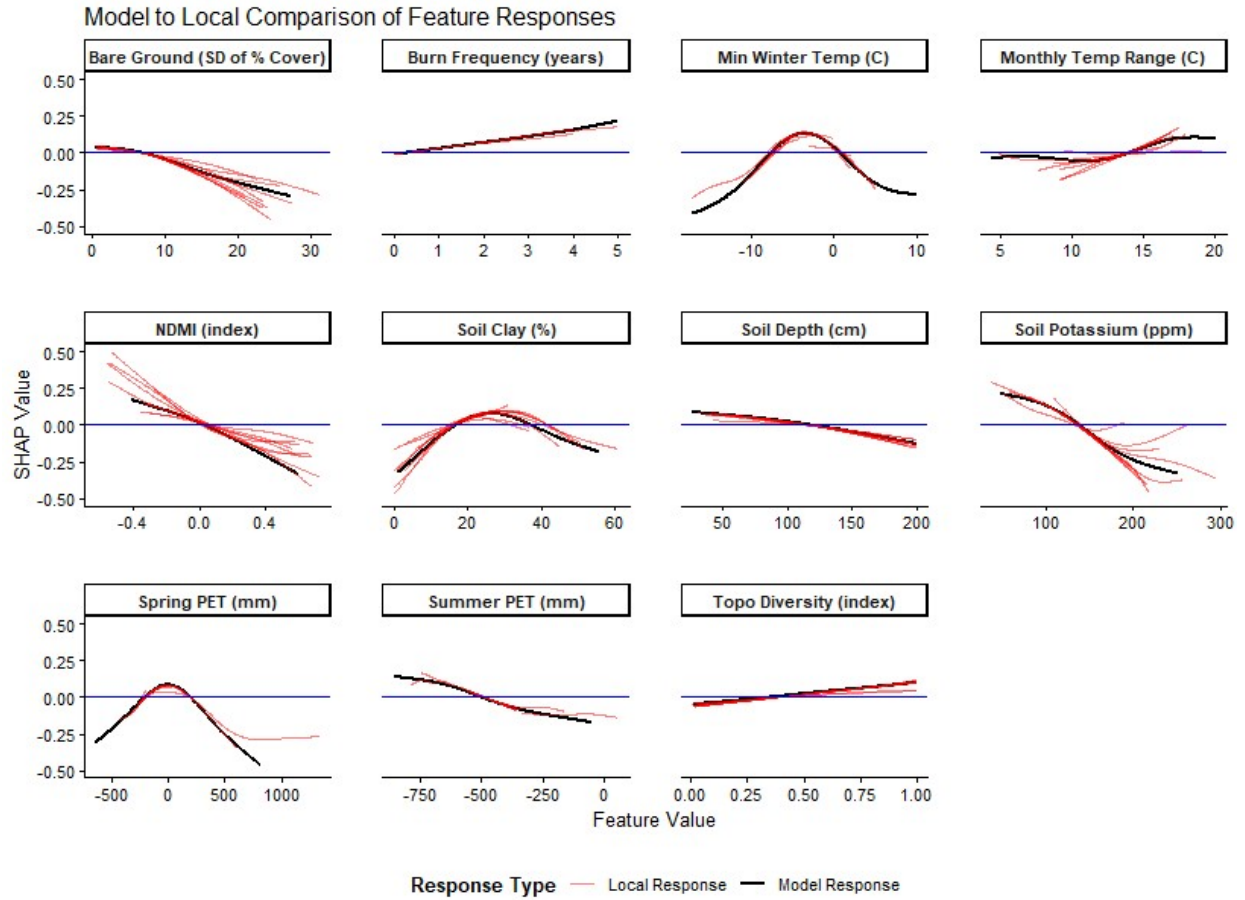


Figure 8. Feature response curves relative to SHAP values. Trendlines generated with Bayesian additive models to show variation in the shape and direction of feature responses across all focal counties (red) compared to model trendlines (black).

5 Discussion

As machine learning methods with vague and opaque decision-making processes are adopted as standard practice for species distribution modeling, it is more necessary than ever to understand how model predictions are generated. In this study, we tested whether Shapley Additive Explanations (SHAP) feature importance values changed between model predictions for a species distribution models of medusahead and predictions within nine counties across the Western United States. We found partial support for our hypothesis that local feature importance would diverge from the model feature importance when looking at local SHAP correlation structure and the

distribution of raw SHAP values between the model and counties across features. In contrast, the overall similarity of model and local data for relative importance rankings and responses of SHAP values across feature gradients provide partial support against our hypothesis. One implication of these results is that modelers may find summaries of feature importance most useful as a part of the interpretation and validation of SDMs intended to be transferable to new data. Alternately, mapped feature importance may better serve model users to further delineate the spatial patterns of suitability for land management activities.

5.1 Model to local comparisons

Most notable among differences between model and local importance were shifts in the distribution of raw SHAP values (Fig. 7) and stronger correlations among local SHAP values (Fig. 2). Both instances of local divergence may be explained by a lack of local environmental variation (driven by spatial autocorrelation of features), especially when feature values derive from environmental data with low spatial resolution (Table 2). Higher resolution features that do contain similar variation and spread to model features and have a higher magnitude change in distribution means may be indicators of locally driven, scale-dependent ecological processes underlying model predictions. Otherwise, shifts in the overall directionality of feature importance may also simply reflect whether a county has higher or lower overall predicted suitability (Table 1).

In contrast, model and local importance demonstrated similar patterns of relative ranks when comparing rescaled mean absolute importance (Fig. 6). One notable exception is the low agreement ($R^2 = 0.086$) in Grays Harbor County which is situated along the Pacific coast in southwest Washington state. One explanation for the low agreement is the high amount of model extrapolation found in Grays County. Extrapolation occurs primarily along the coast where proximity to the ocean could cause stark changes in available atmospheric water that may alter the

value ranges of some locally important features (Spring PET, Summer PET; Fig. 4). Model and local data were also similar in the responses of SHAP values to feature values (Fig. 8), where the shape and direction of local response sometimes provided near exact mirrors of model data. Small deviations from the modeled response could be caused by new combinations of feature values that are not well represented in model training data. Finally, local importance values were more often highly correlated with one another (Fig. 2). This could be a by-product of the limited geographic area of the counties where there is a lack of variation in feature values due to coarse original spatial resolution of features (Table 2) or low overall environmental heterogeneity. Importance values that covary in geographic space may confound the ability to distinguish relative importance between them (Harisena et al., 2021).

5.2 Implications for modelers

There are three primary implications of our results for SDM development related to model interpretation, transferability, and validation. First is that, as an interpretation tool, feature importance with SHAP values can identify where local importance diverges from model importance, at least through comparative measures like distributions or through spatial variation in the maps of the most important features (Fig. 5). Second, local feature importance may be able to serve as a substitute for independent data when determining a model's transferability to new, out-of-sample data. Nowhere is this clearer than in the consistency and similarity of model and local responses (Fig. 8) across a wide array of local environmental conditions. Lastly, our results show that post-hoc analysis methods can be used to validate the ecological plausibility of how a features influence model predictions on new data. Medusahead establishment is widely considered to be more likely at sites with high soil clay and in climates with cooler winters (Dahl & Tisdale, 1975; Nafus & Davies, 2014; Young, 1992). We consider our predictions plausible based on the

responses of SHAP values to the Soil Clay and Min Winter Temp features. For Soil Clay, positive SHAP values predominately fell within the range of 18 to 23% which we consider “high” as they are all above the 50th percentile of values found in the source data for Soil Clay. For Min Winter Temp, positive SHAP values predominantly fell within the range of -1 to -6 C, indicating positive feature influence at cool but not cold winter temperatures.

5.3 Implications for land managers

Our results suggest two primary ways in which mapped feature importance could augment and enhance the use of mapped suitability in the context of invasive species management. First, feature importance provides an additional level of explanation and interpretation to improve the credibility and uptake of suitability information, if the locally important features align with expectations from local ecological knowledge and experience. Second, land managers with limited operational budgets and large jurisdictions will likely need additional ways to effectively leverage mapped suitability from SDMs. On one hand, maps of the most important features (Fig. 5) for maximum SHAP values could provide a secondary spatial filter within areas of high suitability (Fig. 9). Here, land managers could prioritize the landscape by identifying areas where higher suitability values are largely influenced by habitat aspects that change more slowly over time (e.g., edaphic conditions) and may indicate a consistent source of species establishment. Similarly, some maps of the most important features (Fig. 5) for minimum SHAP values show lower predicted suitability driven primarily by climatic features. If climate adaptation planning is a part of land management, this information could help target and prioritize long-term monitoring efforts where future warming may increase suitability.

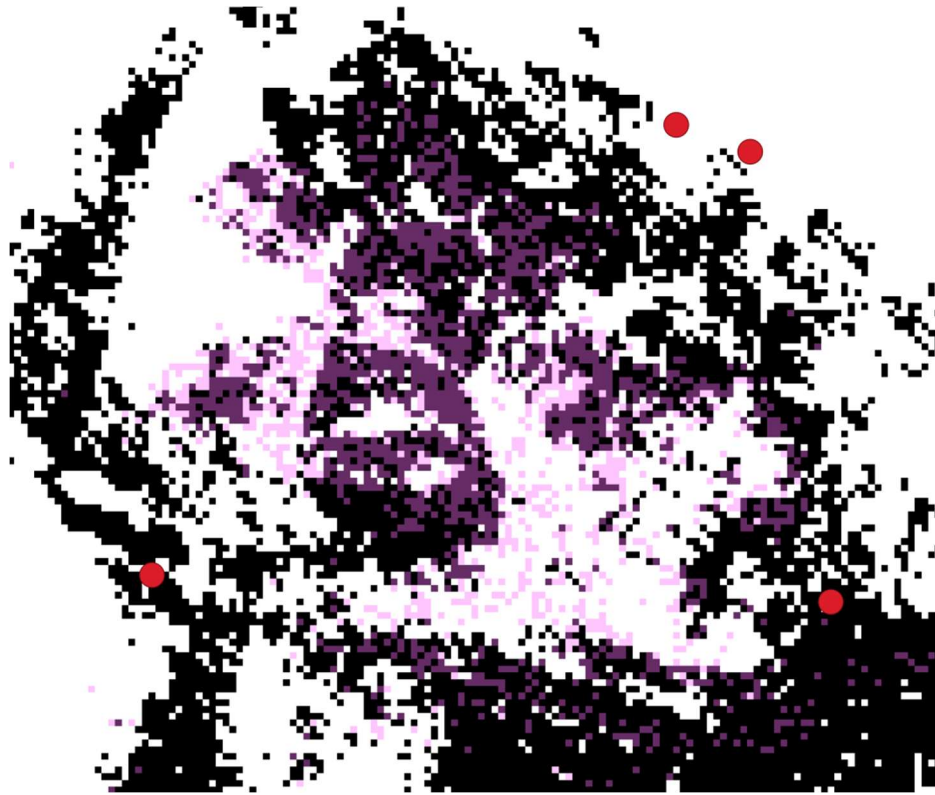


Figure 9. An example landscape from Crook County, OR showing where large clusters of predicted high suitability for medusahead occurrence (black, > 0.75). The purple colors show the spatial distribution of the most locally important model feature (Soil Clay). The darker purple areas where high suitability predictions are driven primarily by Soil Clay.

6 Future Work

The use of IML methods for the analysis, validation, and interpretation of SDMs is an exciting frontier of developing research. The results of this study emphasize the need for continued work on the understanding of model behavior through the lens of local feature importance. For example, we are just beginning to see the first tests of local feature importance as it varies across models and IML methods (Buebos-Esteve & Dagamac, 2025). Though some preliminary work exists at the country level (Waldock et al., 2024), we are not aware of any research generating large-scale feature importance to explore biogeographic trends of model prediction processes across

continental or global extents. Continuous geographic coverage across multi-temporal datasets could also facilitate work on the spatiotemporal trends of model behavior. Finally, the reliability of feature importance values where models are extrapolating remains an open area of future research (Ryo et al., 2021) but could provide an additional way to identify and communicate model uncertainty.

REFERENCES

- Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1), 121–131. <https://doi.org/10.1002/joc.3413>
- Bell, D. M., & Schlaepfer, D. R. (2016). On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological Modelling*, 330, 50–59. <https://doi.org/10.1016/j.ecolmodel.2016.03.012>
- Bourhis, Y., Bell, J. R., Shortall, C. R., Kunin, W. E., & Milne, A. E. (2023). Explainable neural networks for trait-based multispecies distribution modelling—A case study with butterflies and moths. *Methods in Ecology and Evolution*, 14(6), 1531–1542. <https://doi.org/10.1111/2041-210X.14097>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Buebos-Esteve, D. E., & Dagamac, N. H. A. (2025). Evaluating model-agnostic post-hoc methods in explainable artificial intelligence: Augmenting species distribution models. *Biologia Futura*. <https://doi.org/10.1007/s42977-025-00288-w>
- Cha, Y., Shin, J., Go, B., Lee, D.-S., Kim, Y., Kim, T., & Park, Y.-S. (2021). An interpretable machine learning method for supporting ecosystem management: Application to species distribution models of freshwater macroinvertebrates. *Journal of Environmental Management*, 291, 112719. <https://doi.org/10.1016/j.jenvman.2021.112719>
- Coelho, M. T. P., Diniz-Filho, J. A., & Rangel, T. F. (2019). A parsimonious view of the parsimony principle in ecology and evolution. *Ecography*, 42(5), 968–976. <https://doi.org/10.1111/ecog.04228>
- Dahl, B. E., & Tisdale, E. W. (1975). Environmental Factors Related to Medusahead Distribution. *Journal of Range Management*, 28(6), 463. <https://doi.org/10.2307/3897223>
- DiMiceli, C., Carroll, M., Sohlberg, R., Kim, D.-H., Kelly, M., & Townshend, J. (2015). *MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006* [Dataset]. NASA Land Processes Distributed Active Archive Center. <https://doi.org/10.5067/MODIS/MOD44B.006>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z.-L., Quayle, B., & Howard, S. (2007). A Project for Monitoring Trends in Burn Severity. *Fire Ecology*, 3(1), 3–21. <https://doi.org/10.4996/fireecology.0301003>
- Elith, J. (2017). Predicting Distributions of Invasive Species. In A. P. Robinson, T. Walshe, M. A. Burgman, & M. Nunn (Eds.), *Invasive Species* (1st ed., pp. 93–129). Cambridge University Press. <https://doi.org/10.1017/9781139019606.006>
- Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27(2), 245–256. <https://doi.org/10.1111/geb.12684>
- Greenwell, B. (2026). *fastshap: Fast Approximate Shapley Values* (Version R Package version 0.1.1) [Computer software]. <https://bgreenwell.github.io/fastshap/>. <https://github.com/bgreenwell/fastshap>
- Halmy, M. W. A., Fawzy, M., Ahmed, D. A., Saeed, N. M., & Awad, M. A. (2019). Monitoring and predicting the potential distribution of alien plant species in arid ecosystem using remotely-sensed

- data. *Remote Sensing Applications: Society and Environment*, 13, 69–84. <https://doi.org/10.1016/j.rsase.2018.10.005>
- Harisena, N. V., Groen, T. A., Toxopeus, A. G., & Naimi, B. (2021). When is variable importance estimation in species distribution modelling affected by spatial correlation? *Ecography*, 44(5), ecog.05534. <https://doi.org/10.1111/ecog.05534>
- He, B., Zhao, Y., & Mao, W. (2022). Explainable artificial intelligence reveals environmental constraints in seagrass distribution. *Ecological Indicators*, 144, 109523. <https://doi.org/10.1016/j.ecolind.2022.109523>
- Hooper, S., & Kennedy, R. E. (2018). A spatial ensemble approach for broad-area mapping of land surface properties. *Remote Sensing of Environment*, 210, 473–489. <https://doi.org/10.1016/j.rse.2018.03.032>
- Jarnevich, C. S., Engelstad, P., Williams, D., Shadwell, K., Reimer, C., Henderson, G., Prevey, J. S., & Pearse, I. S. (2024). Predicted occurrence and abundance habitat suitability of invasive plants in the contiguous United States: Updates for the INHABIT web tool. *NeoBiota*, 96, 261–278. <https://doi.org/10.3897/neobiota.96.134842>
- Jarnevich, C. S., Sofaer, H. R., & Engelstad, P. (2021). Modelling presence versus abundance for invasive species risk assessment. *Diversity and Distributions*, 27(12), 2454–2464. <https://doi.org/10.1111/ddi.13414>
- Kennedy, R. E., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W. B., & Healey, S. (2018). Implementation of the LandTrendr Algorithm on Google Earth Engine. *Remote Sensing*, 10(5), 691. <https://doi.org/10.3390/rs10050691>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). Routledge. <https://doi.org/10.1201/9780203753736>
- Molnar, C. (2025). *Interpretable machine learning: A guide for making black box models explainable* (3:e upplagan). Christoph Molnar.
- Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges. In I. Koprincka, M. Kamp, A. Appice, C. Loglisci, L. Antonie, A. Zimmermann, R. Guidotti, Ö. Özgöbek, R. P. Ribeiro, R. Gavaldà, J. Gama, L. Adilova, Y. Krishnamurthy, P. M. Ferreira, D. Malerba, I. Medeiros, M. Ceci, G. Manco, E. Masciari, ... J. A. Gulla (Eds.), *ECML PKDD 2020 Workshops* (Vol. 1323, pp. 417–431). Springer International Publishing. https://doi.org/10.1007/978-3-030-65965-3_28
- Nafus, A. M., & Davies, K. W. (2014). Medusahead Ecology and Management: California Annual Grasslands to the Intermountain West. *Invasive Plant Science and Management*, 7(2), 210–221. <https://doi.org/10.1614/IPSM-D-13-00077.1>
- Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. *Environmental Management*, 54(6), 1249–1266. <https://doi.org/10.1007/s00267-014-0364-1>
- PRISM Group. (2026). *PRISM Data Grids*. <https://prism.oregonstate.edu>
- R Core Team. (2025). *R: A language and environment for statistical computing*. [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., & Thompson, J. (2018). Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. *Soil Science Society of America Journal*, 82(1), 186–201. <https://doi.org/10.2136/sssaj2017.04.0122>
- Reaser, J. K., Burgiel, S. W., Kirkey, J., Brantley, K. A., Veatch, S. D., & Burgos-Rodríguez, J. (2020). The early detection of and rapid response (EDRR) to invasive species: A conceptual framework and federal capacities assessment. *Biological Invasions*, 22(1), 1–19. <https://doi.org/10.1007/s10530-019-02156-w>
- Ryo, M., Angelov, B., Mammola, S., Kass, J. M., Benito, B. M., & Hartig, F. (2021). Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography*, 44(2), 199–205. <https://doi.org/10.1111/ecog.05360>
- Shapley, L. S. (1952). *A Value for n-person Games*.

- Sheley, R. L., Bingham, B. S., & Svejcar, T. J. (2008). Crested Wheatgrass Defoliation Intensity and Season on Medusahead Invasion. *Rangeland Ecology & Management*, 61(2), 211–217. <https://doi.org/10.2111/07-062.1>
- Song, L., Frazier, A., Kedron, P., Araujo, D. S. A., Cui, D., Enquist, B., Hinojo-Hinojo, C., Krieger, J. M., Maitner, B. S., Merow, C., Moulatlet, G. M., Nikolopoulos, E. I., & Roehrdanz, P. (2025). *Explainable Artificial Intelligence Reveals Spatially Divergent Effects of Global Change on Mammals*. In Review. <https://doi.org/10.21203/rs.3.rs-7500921/v1>
- Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3), 647–665. <https://doi.org/10.1007/s10115-013-0679-x>
- Theobald, D. M., Harrison-Atlas, D., Monahan, W. B., & Albano, C. M. (2015). Ecologically-Relevant Maps of Landforms and Physiographic Diversity for Climate Adaptation Planning. *PLOS ONE*, 10(12), e0143619. <https://doi.org/10.1371/journal.pone.0143619>
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Aroita, G. (2023). Flexible species distribution modelling methods perform well on spatially separated testing data. *Global Ecology and Biogeography*, 32(3), 369–383. <https://doi.org/10.1111/geb.13639>
- Waldock, C., Wegscheider, B., Josi, D., Calegari, B. B., Brodersen, J., Jardim De Queiroz, L., & Seehausen, O. (2024). Deconstructing the geography of human impacts on species' natural distribution. *Nature Communications*, 15(1), 8852. <https://doi.org/10.1038/s41467-024-52993-0>
- Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3(2), 260–267. <https://doi.org/10.1111/j.2041-210X.2011.00170.x>
- Young, J. A. (1992). ECOLOGY AND MANAGEMENT OF MEDUSAHEAD(TAENIATHERUM CAPUT-MEDUSAE SSP. ASPERUM [SIMK.] MELDERIS). *Great Basin Naturalist*, 52(3), 242–252.
- Zbinden, R., Van Tiel, N., Sumbul, G., Vanalli, C., Kellenberger, B., & Tuia, D. (2026). MASKSDM with Shapley values to improve flexibility, robustness and explainability in species distribution modelling. *Methods in Ecology and Evolution*, 17(1), 188–206. <https://doi.org/10.1111/2041-210X.70200>

SYNTHESIS

1 Overview

Species distribution models (SDMs) are a common statistical approach in ecology to understand and predict species distributions with uses cases ranging from conservation planning and biodiversity monitoring to climate change assessments and invasive species management. Despite their widespread use, SDMs are often treated primarily as prediction engines, with model outputs serving as endpoints rather than opportunities to explore underlying ecological processes. Primarily through the lens of invasive species, this dissertation addressed the challenge of improving the way we develop, evaluate, and interpret SDMs in an attempt to move away from the prediction engine paradigm and toward a better understanding of *how* and *why* SDMs work the way they do.

In Chapter 1, I demonstrated how SDMs can be developed within a quantitative, spatially explicit risk assessment framework to support the management of freshwater aquatic invasive species across broad geographic extents. By augmenting habitat suitability models for 103 invasive species with spatial proxies of secondary invasion pathways, I showed the utility of combining dispersal and movement information with species-specific ecological modeling to generate realistic, management-ready products such as risk maps and watch lists. In Chapter 2, I introduced a novel spatial data-splitting approach designed to simultaneously generate cross-validation splits alongside withheld test data to provide robust evaluations of model performance. This work used a virtual species approach to evaluate the performance of models built with this new splitting technique, especially when transferred beyond the geographic range of training data. Finally, in Chapter 3, I explored the use of a recent feature importance method to explain model predictions and demonstrate how interpretable machine learning approaches (Molnar, 2025) can help connect

model outputs to recognizable environmental processes and improve confidence in model predictions.

Although these chapters focus on different components of the modeling process, they share a common theme: effective species distribution modeling requires a deeper, holistic understanding than prediction accuracy alone can provide. Together, they illustrate how advances in model development, evaluation, and interpretation can contribute to both ecological understanding and practical decision-making.

2 Beyond prediction engines

A central lesson emerging from this work is a reminder that SDMs can be viewed as tools for ecological inquiry as much as tools for prediction (Elith and Leathwick, 2009). Prediction remains an important goal, particularly when addressing urgent ecological and economic challenges such as biological invasions but prediction alone provides limited insight into how SDMs arrived at a decision. Models built with spurious correlations could result in poor transferability in novel locations or future climates (Fourcade et al., 2018; Thuiller, 2004). Conversely, models that accurately capture ecologically meaningful relationships may provide valuable information even when predictions are imperfect (Yates et al., 2018). This perspective becomes particularly important as ecological models continue to increase in complexity. Advances in machine learning have enabled the development of increasingly sophisticated modeling approaches capable of capturing nonlinear relationships and interactions among spatial environmental variables. While these methods are often aimed at improving predictive performance, they can also reduce transparency and make it more difficult to understand the drivers of model behavior. The work presented in this dissertation demonstrates that model evaluation and interpretation should be considered integral components of a more holistic model

development workflow. Understanding where a model succeeds, where it fails, and what environmental relationships drive its predictions help build confidence in its use for both ecological investigation and management applications (Boschetti et al., 2018; Ryo et al., 2021).

3 Dual aims of applied ecological research

The complementarity between methodological development and applied ecological research is another important theme that emerges from this dissertation. While much of this work is applied ecological research motivated by the ongoing environmental and economic impacts of invasive species, the methodological advances presented here contribute directly to the ability of modelers to ask broader ecological questions. Real-world management challenges often expose limitations in existing analytical approaches and create opportunities for methodological innovation (Guillera-Arroita et al., 2015; Sofaer et al., 2019).

Invasive species management provides a particularly useful example of this duality in that predictions often need to be spatially explicit, transferable to a local environmental context, interpretable and believable by end users, and, in the end, actionable. Meeting these needs forces model developers to confront the need to improve uncertainty measures, evaluation metrics, extrapolation assessments, and communication strategies. As demonstrated throughout this dissertation, efforts to improve SDMs in an applied setting frequently generate insights that advance ecological modeling methods more generally.

Reciprocally, improvements in SDM methodology can also benefit resource managers who are increasingly asked to make decisions for uncertain future environmental conditions with limited resources (Guisan et al., 2013; Villero et al., 2017). Accurate predictions help prioritize monitoring and management efforts. Robust evaluation methods improve confidence in model outputs. Interpretability tools help connect model predictions to environmental processes that are

familiar to stakeholders and practitioners. Together, these advances can improve the transparency, credibility, and usefulness of ecological models in decision-making contexts (Sofaer et al., 2019; Young et al., 2020).

4 Invasive species as study taxa

Invasive species remain as valuable study taxa for the continued development of SDMs. Unlike the well-established and comparably stable distributions of native species, invasive species are often actively expanding in their non-native ranges. The result is the creation of a natural experiment where the efficacy of SDMs can be evaluated in anticipating potential habitat or range shifts (Elith, 2017; Elith et al., 2010). The results of these experiments can reveal strengths and weaknesses in modeling approaches that might remain hidden in more static systems or study taxa. For this reason, invasive species represent more than just an important management challenge; they also serve as valuable stress tests to advance spatial ecological modeling.

5 Future directions

5.1 INHABIT and open science

The themes explored throughout this dissertation also inform several directions for future research. One of the most immediate opportunities involves the continued development of the Invasive Species Habitat Tool (INHABIT; Engelstad et al., 2022). The invasion risk framework, evaluation methods, and interpretation approaches developed here directly contribute to the iteration of INHABIT and its continued capacity to provide freely accessible information for scientists, resource managers, and the public. The guiding philosophy behind INHABIT is to widen the window into species distribution modeling through open, transparent, and accessible

science. Rather than treating model outputs as static products, the goal is to create a platform that supports ongoing, co-produced efforts between researchers and practitioners (Engelstad et al, 2022). The development and delivery of spatial ecological models is more effective when it operates as a collaborative process. Ongoing feedback and active communication with end users brings INHABIT in line with broader movements to develop more open, reproducible, and actionable ecological research (Hampton et al., 2015).

An emphasis on collaboration also reflects a broader commitment to co-production in ecological research (Enquist et al., 2017). For example, the management of invasive species often integrates the modeling expertise of researchers with the practical experience of those responsible for implementing management actions. Because resource managers often possess local expert knowledge that may not be captured in model inputs alone, incorporating these perspectives in subsequent model iteration has the potential to improve both model relevance and model performance. The work presented in this dissertation then represents one component of an ongoing partnership among researchers, resource managers, and stakeholders that helps ensure methodological advances remain digestible and relevant.

5.2 Invasion science in an age of uncertainty

The need for a collaborative approach is increasingly important in an era characterized by rapid environmental change, increasing ecological complexity, and substantial uncertainty. Biological invasions continue to expand globally and climate change has the potential to change both invasive and native species distributions (Bradley et al., 2024). Adapting science and management in the face of global change will likely require more than increasingly complex models alone. It will also require a robust understanding of the statistical relationships that drive model predictions and the ability to assess whether or not those relationships reflect ecological

realities. Without such understanding, it becomes difficult to evaluate the credibility of forecasts or identify the conditions or locations under which predictions may fail or be less reliable. The methods developed in this dissertation contribute to the continued development of a deeper model understanding and transparency by emphasizing the importance of robust evaluation and interpretation methods alongside predictive performance.

The importance of this work extends beyond methodological considerations due to the substantial ecological, economic, and human health impacts caused by the ongoing, worldwide spread of non-native species (Haubrock et al., 2025). Recent estimates suggest that the global economic costs of biological invasions exceed hundreds of billions of dollars annually and continue to increase over time (Diagne et al., 2021). Costs are often comparable in scale to those associated with major natural disasters (Turberlin et al., 2023) and invasive species are likely to remain a persistent challenge for the foreseeable future largely a consequence of human activities. Although human movement of goods, organisms, and materials has accelerated the redistribution of species around the globe, we remain the primary agents capable of preventing, detecting, and managing invasions. This dual role creates both a responsibility and an opportunity with advances in SDM methods helping to guide on-the-ground efforts by helping to create effective and useful spatial information that can support informed decision-making.

6 Conclusion

Ultimately, the work presented in this dissertation reflects a broader goal of building spatial ecological models that are not only accurate, but also understandable, transferable, and useful. By combining large-scale invasive species risk assessments, rigorous approaches to model evaluation, and methods for interpreting individual predictions, this research contributes to a more comprehensive understanding of how species distribution models in particular can support both

scientific discovery and practical management. As ecological challenges continue to grow in complexity, the need for such integrated approaches will only become more important.

REFERENCES FOR SYNTHESIS

- Boschetti, F., Hughes, M., Jones, C., & Lozano-Montes, H. (2018). On Decision Makers' Perceptions of What an Ecological Computer Model is, What It Does, and Its Impact on Limiting Model Acceptance. *Sustainability*, 10(8), 2767. <https://doi.org/10.3390/su10082767>
- Bradley, B. A., Beaury, E. M., Gallardo, B., Ibáñez, I., Jarnevich, C., Morelli, T. L., Sofaer, H. R., Sorte, C. J. B., & Vilà, M. (2024). Observed and Potential Range Shifts of Native and Nonnative Species with Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, 55(1), 23–40. <https://doi.org/10.1146/annurev-ecolsys-102722-013135>
- Diagne, C., Leroy, B., Vaissière, A.-C., Gozlan, R. E., Roiz, D., Jarić, I., Salles, J.-M., Bradshaw, C. J. A., & Courchamp, F. (2021). High and rising economic costs of biological invasions worldwide. *Nature*, 592(7855), 571–576. <https://doi.org/10.1038/s41586-021-03405-6>
- Elith, J. (2017). Predicting Distributions of Invasive Species. In A. P. Robinson, T. Walshe, M. A. Burgman, & M. Nunn (Eds.), *Invasive Species* (1st ed., pp. 93–129). Cambridge University Press. <https://doi.org/10.1017/9781139019606.006>
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Engelstad, P., Jarnevich, C. S., Hogan, T., Sofaer, H. R., Pearse, I. S., Sieracki, J. L., Frakes, N., Sullivan, J., Young, N. E., Prevéy, J. S., Belamaric, P., & LaRoe, J. (2022). INHABIT: A web-based decision support tool for invasive plant species habitat visualization and assessment across the contiguous United States. *PLOS ONE*, 17(2), e0263056. <https://doi.org/10.1371/journal.pone.0263056>
- Enquist, C. A., Jackson, S. T., Garfin, G. M., Davis, F. W., Gerber, L. R., Littell, J. A., Tank, J. L., Terando, A. J., Wall, T. U., Halpern, B., Hiers, J. K., Morelli, T. L., McNie, E., Stephenson, N. L., Williamson, M. A., Woodhouse, C. A., Yung, L., Brunson, M. W., Hall, K. R., ... Shaw, M. R. (2017). Foundations of translational ecology. *Frontiers in Ecology and the Environment*, 15(10), 541–550. <https://doi.org/10.1002/fee.1733>
- Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27(2), 245–256. <https://doi.org/10.1111/geb.12684>
- Guillera-Aroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24(3), 276–292. <https://doi.org/10.1111/geb.12268>
- Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., ... Buckley, Y. M. (2013). Predicting species distributions

- for conservation decisions. *Ecology Letters*, 16(12), 1424–1435. <https://doi.org/10.1111/ele.12189>
- Hampton, S. E., Anderson, S. S., Bagby, S. C., Gries, C., Han, X., Hart, E. M., Jones, M. B., Lenhardt, W. C., MacDonald, A., Michener, W. K., Mudge, J., Pourmokhtarian, A., Schildhauer, M. P., Woo, K. H., & Zimmerman, N. (2015). The Tao of open science for ecology. *Ecosphere*, 6(7), 1–13. <https://doi.org/10.1890/ES14-00402.1>
- Haubrock, P. J., Tarkan, A. S., Martín-Forés, I., Katsanevakis, S., Sousa, R., Soto, I., Green, A. J., Kouba, A., Everts, T., Dominguez Almela, V., Belouard, N., Hui, C., Bojko, J., Deklerck, V., Boeraeve, M., Essl, F., & Britton, J. R. (2025). The spread of non-native species. *Biological Reviews*, brv.70121. <https://doi.org/10.1002/brv.70121>
- Molnar, C. (2025). Interpretable machine learning: A guide for making black box models explainable (3:e upplagan). Christoph Molnar.
- Ryo, M., Angelov, B., Mammola, S., Kass, J. M., Benito, B. M., & Hartig, F. (2021). Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography*, 44(2), 199–205. <https://doi.org/10.1111/ecog.05360>
- Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., Edwards, T. C., Guala, G. F., Howard, T. G., Morisette, J. T., & Hamilton, H. (2019). Development and Delivery of Species Distribution Models to Inform Decision-Making. *BioScience*, 69(7), 544–557. <https://doi.org/10.1093/biosci/biz045>
- Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10(12), 2020–2027. <https://doi.org/10.1111/j.1365-2486.2004.00859.x>
- Turbelin, A. J., Cuthbert, R. N., Essl, F., Haubrock, P. J., Ricciardi, A., & Courchamp, F. (2023). Biological invasions are as costly as natural hazards. *Perspectives in Ecology and Conservation*, S2530064423000214. <https://doi.org/10.1016/j.pecon.2023.03.002>
- Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J., & Brotons, L. (2017). Integrating species distribution modelling into decision-making to inform conservation actions. *Biodiversity and Conservation*, 26(2), 251–271. <https://doi.org/10.1007/s10531-016-1243-2>
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*, 33(10), 790–802. <https://doi.org/10.1016/j.tree.2018.08.001>
- Young, N. E., Jarnevich, C. S., Sofaer, H. R., Pearse, I., Sullivan, J., Engelstad, P., & Stohlgren, T. J. (2020). A modeling workflow that balances automation and human intervention to inform invasive plant management decisions at multiple spatial scales. *PLOS ONE*, 15(3), e0229253. <https://doi.org/10.1371/journal.pone.0229253>

APPENDIX 1- SUPPLEMENTAL INFORMATION FOR CHAPTER 1

Table S1. List of selected secondary pathways by species.

Latin Name	Taxa Group	Pathway: Downstream Connectivity	Pathway: Boat Ramp Density	Pathway: Anthropogenic Connectivity	Pathway: Permanent Water	Pathway: Human Transport Risk
<i>Alosa pseudoharengus</i>	Fish	X		X		
<i>Alternanthera philoxeroides</i>	Plants	X	X	X	X	
<i>Ambloplites rupestris</i>	Fish	X		X		
<i>Ameiurus catus</i>	Fish	X		X		
<i>Ameiurus melas</i>	Fish	X		X		
<i>Ameiurus natalis</i>	Fish	X		X		
<i>Arundo donax</i>	Plants	X		X		
<i>Astronotus ocellatus</i>	Fish	X		X		X
<i>Bithynia tentaculata</i>	Invert	X	X	X	X	
<i>Butomus umbellatus</i>	Plants	X	X	X	X	X
<i>Bythotrephes longimanus</i>	Invert	X	X	X	X	
<i>Callitriche stagnalis</i>	Plants	X		X		
<i>Cabomba caroliniana</i>	Plants	X	X	X	X	X
<i>Carassius auratus</i>	Fish	X		X		X
<i>Channa argus</i>	Fish	X		X		
<i>Chrosomus oreas</i>	Fish	X		X		
<i>Cipangopaludina chinensis</i>	Invert	X	X	X	X	
<i>Clarias batrachus</i>	Fish	X		X		X
<i>Colocasia esculenta</i>	Plants	X	X	X		X
<i>Corbicula fluminea</i>	Invert	X	X	X	X	
<i>Ctenopharyngodon idella</i>	Fish	X		X		
<i>Cyperus blepharoleptos</i>	Plants	X	X	X	X	
<i>Cyprinella lutrensis</i>	Fish	X		X		
<i>Cyprinus carpio</i>	Fish	X		X		
<i>Daphnia lumholtzi</i>	Invert	X	X	X	X	
<i>Didymosphenia geminata</i>	Plants	X	X	X	X	

<i>Dorosoma cepedianum</i>	Fish	X		X		
<i>Dorosoma petenense</i>	Fish	X		X		
<i>Dreissena polymorpha</i>	Invert	X	X	X		
<i>Dreissena bugensis</i>	Invert	X	X	X		
<i>Echinogammarus ischnus</i>	Invert	X	X	X	X	
<i>Egeria densa</i>	Plants	X	X	X	X	X
<i>Eichhornia crassipes</i>	Plants	X	X	X	X	X
<i>Fallopia japonica</i>	Plants	X		X		
<i>Faxonius rusticus</i>	Invert	X		X		X
<i>Faxonius virilis</i>	Invert	X		X		X
<i>Gambusia affinis</i>	Fish	X		X		
<i>Gambusia holbrooki</i>	Fish	X		X		
<i>Gymnocephalus cernua</i>	Fish	X		X		
<i>Hydrilla verticillata peregrina</i>	Plants	X	X	X	X	
<i>Hydrilla verticillata verticillata</i>	Plants	X	X	X	X	
<i>Hydrilla verticillata</i>	Plants	X	X	X	X	
<i>Hydrocharis morsus-ranae</i>	Plants	X	X	X	X	X
<i>Hydrocotyle ranunculoides</i>	Plants	X	X	X	X	
<i>Hygrophila polysperma</i>	Plants	X	X	X	X	X
<i>Hypophthalmichthys molitrix</i>	Fish	X		X		
<i>Hypophthalmichthys nobilis</i>	Fish	X		X		
<i>Ictalurus furcatus</i>	Fish	X		X		
<i>Iris pseudacorus</i>	Plants	X	X	X	X	X
<i>Lepomis gulosus</i>	Fish	X		X		
<i>Limnophila sessiliflora</i>	Plants	X	X	X	X	X
<i>Ludwigia grandiflora</i>	Plants	X		X		
<i>Ludwigia hexapetala</i>	Plants	X	X	X	X	
<i>Ludwigia peploides</i>	Plants	X	X	X	X	
<i>Ludwigia peruviana</i>	Plants	X	X	X	X	
<i>Lythrum hyssopifolia</i>	Plants	X		X		
<i>Lythrum salicaria</i>	Plants	X	X	X	X	X

<i>Marsilea quadrifolia</i>	Plants	X		X		X
<i>Melaleuca quinquenervia</i>	Plants	X		X		
<i>Melanoides tuberculata</i>	Invert	X	X	X	X	
<i>Mentha aquatica</i>	Plants	X		X		
<i>Micropterus henshalli</i>	Fish	X		X		
<i>Misgurnus anguillicaudatus</i>	Fish	X		X		X
<i>Monopterus albus</i>	Fish	X		X		
<i>Murdannia keisak</i>	Plants	X	X	X	X	
<i>Mylopharyngodon piceus</i>	Fish	X		X		
<i>Myosotis scorpioides</i>	Plants	X		X		
<i>Myriophyllum aquaticum</i>	Plants	X	X	X	X	X
<i>Myriophyllum heterophyllum</i>	Plants	X	X	X	X	
<i>Myriophyllum sibiricum</i>	Plants	X	X	X	X	
<i>Najas marina</i>	Plants	X	X	X	X	
<i>Najas minor</i>	Plants	X	X	X	X	
<i>Nasturtium officinale</i>	Plants	X	X	X	X	
<i>Nelumbo lutea</i>	Plants	X	X	X	X	X
<i>Neogobius melanostomus</i>	Fish	X		X		
<i>Nitellopsis obtusa</i>	Plants	X	X	X	X	
<i>Nymphoides peltata</i>	Plants	X	X	X		X
<i>Oreochromis aureus</i>	Fish	X		X		X
<i>Oreochromis niloticus</i>	Fish	X		X		X
<i>Panicum repens</i>	Plants	X		X		
<i>Persicaria hydropiper</i>	Plants	X		X		
<i>Phalaris arundinacea</i>	Plants	X		X		
<i>Phragmites australis australis</i>	Plants	X	X	X	X	
<i>Pistia stratiotes</i>	Plants	X	X	X	X	X
<i>Pomacea canaliculata</i>	Invert	X		X		X
<i>Pomacea maculata</i>	Invert	X		X		X
<i>Potamogeton crispus</i>	Plants	X	X	X	X	

<i>Potamopyrgus antipodarum</i>	Invert	X	X	X	X	
<i>Procambarus clarkii</i>	Invert	X		X		X
<i>Proterorhinus semilunaris</i>	Fish	X		X		
<i>Pterygoplichthys disjunctivus</i>	Fish	X		X		X
<i>Pterygoplichthys multiradiatus</i>	Fish	X		X		X
<i>Pylodictis olivaris</i>	Fish	X		X		
<i>Salmo trutta</i>	Fish	X		X		
<i>Salvinia minima</i>	Plants	X	X	X	X	X
<i>Salvinia molesta</i>	Plants	X	X	X	X	X
<i>Tamarix chinensis and ramosissima</i>	Plants	X		X		
<i>Tamarix sp</i>	Plants	X		X		
<i>Trapa bispinosa</i>	Plants	X	X	X	X	
<i>Trapa natans</i>	Plants	X	X	X	X	
<i>Typha domingensis</i>	Plants	X	X	X	X	
<i>Utricularia inflata</i>	Plants	X	X	X	X	
<i>Veronica anagallis-aquatica</i>	Plants	X	X	X	X	

Table S1. List of occurrence data sources. We used records from these sources to build species distribution model training and testing data.

Name	Source
Bureau of Land Management's (BLM) Assessment, Inventory and Management (AIM) and Lotic databases	www.blm.gov/aim
Early Detection and Distribution Mapping System (ED-DMapS)	www.eddmaps.org
Global Biodiversity Information Facility (GBIF)	www.gbif.org (GBIF, 2024a, 2024b)
iMapInvasives	www.imapinvasives.org
LANDFIRE (LF) 2016 Remap Reference Database (LFRDB)	LANDFIRE, 2016
Standardized Plant Community with Introduced Status (SPCIS) database	Petri et al., 2023
National Park Service's National Invasive Species Information Management System (NISIMS)	irma.nps.gov/NISIMS/
Nonindigenous Aquatic Species Database (NAS)	U.S. Geological Survey, 2024
Plant occurrence and abundance data	Bradley et al., 2024

Calculations for watchlist down weighting

We down weighted composite risk scores in some subwatersheds to develop our top ten watchlists (Chapter 1, Fig. 7). To determine which scores to down weight, we first discretized suitability rasters into suitable and unsuitable categories using a 5th percentile threshold rule following Jarnevich et al. (2024). A percentile threshold is calculated by extracting predicted values for all training observations and selecting the predicted value for the observation that would misclassify the set percentage of positive observations. All suitability values in the following equations are summarized by focal species and subwatershed across the study area. To begin, we identified subwatershed with a limited number of current and/or future predictions using the following calculations:

Percent Current Suitable (PCS)

$$\frac{(\text{pixels with stable suitability values} + \text{pixels with decreasing suitability in the future})}{\text{total number of pixels of available freshwater}}$$

Percent Future Suitable (PFS)

$$\frac{(\text{pixels with stable suitability values} + \text{pixels with increasing suitability in the future})}{\text{total number of pixels of available freshwater}}$$

Stable suitability is where pixels are predicted as suitable both currently and in the future. Increasing suitability is when pixels are predicted as unsuitable currently but suitable in the future. Decreasing suitability is when pixels are predicted as currently suitable but unsuitable in future. Available freshwater pixels come from the rasterized values of lakes, wetlands (lentic systems), rivers, and streams (lotic systems) from the National Hydrological Dataset (NHD) polygons and polylines (United States Geological Survey, 2025). For either the PCS or PFS ratio, we flagged

values less than 0.1, down weighted their respective current or future suitability to zero, and re-calculated the final composite score. These locations represent subwatersheds where suitability incompletely describes freshwater invasion risk, creating misleading rankings when aggregating results.

Our second down weighting process identified subwatersheds where future suitability was highly variable across models, which is one measure of uncertainty in predictions. We first classified future uncertainty by comparing maximum and minimum future suitability values across the Generalized Circulation Models using the following calculation:

Future Uncertainty (FU)

$$\frac{(\text{pixels with stable suitability values}_{max} + \text{pixels with increasing suitability values in the future}_{max}) - (\text{pixels with stable suitability values}_{min} + \text{pixels with increasing suitability values in the future}_{min})}{\text{total number of pixels of available freshwater}}$$

To represent the ratio of uncertain future predictions to available freshwater, we then calculated:

Percent Future Uncertain (PFU)

$$\frac{\text{total number of uncertain pixels}}{\text{total number of pixels of available freshwater}}$$

Where $PFU > PFS$ or $PFU > (PFS * 2)$, we down weighted future suitability scores to zero and re-calculated the final composite score. These changes captured areas where most pixels with predictions within a subwatershed had high uncertainty. The table below provides examples of hypothetical values to demonstrate the down weighting logic and resulting changes in final composite scores.

Table S3. Example down weighting logic.

PCS	PFS	PFU	Down Weight Reason	Original Current Suitability Score Max: 1 Min: 0	Original Future Suitability Score Max: 1 Min: 0	Original Secondary Pathways Score Max: 1 Min: 0	New Current Suitability Score Max: 1 Min: 0	New Future Suitability Score Max: 1 Min: 0	Original Composite Score Max: 1 Min: 0	New Composite Score Max: 1 Min: 0	Composite Score Change Max: 1 Min: -1
0.03	0.28	0.12	PCS < 0.1	0.30	0.30	0.30	0.00	unchanged	0.90	0.60	-0.30
0.13	0.08	0.12	PFS < 0.1	1.00	0.50	0.50	unchanged	0.00	2.00	1.50	-0.50
0.13	0.15	0.52	FVP > FSP	0.75	1.00	0.25	unchanged	0.00	2.00	1.00	-1.00
0.13	0.3	0.53	FVP > FSP *2	0.60	0.80	0.90	unchanged	0.00	2.30	1.50	-0.80
0.99	0.28	0.03	None	0.90	0.20	0.60	unchanged	unchanged	1.70	1.70	0.00

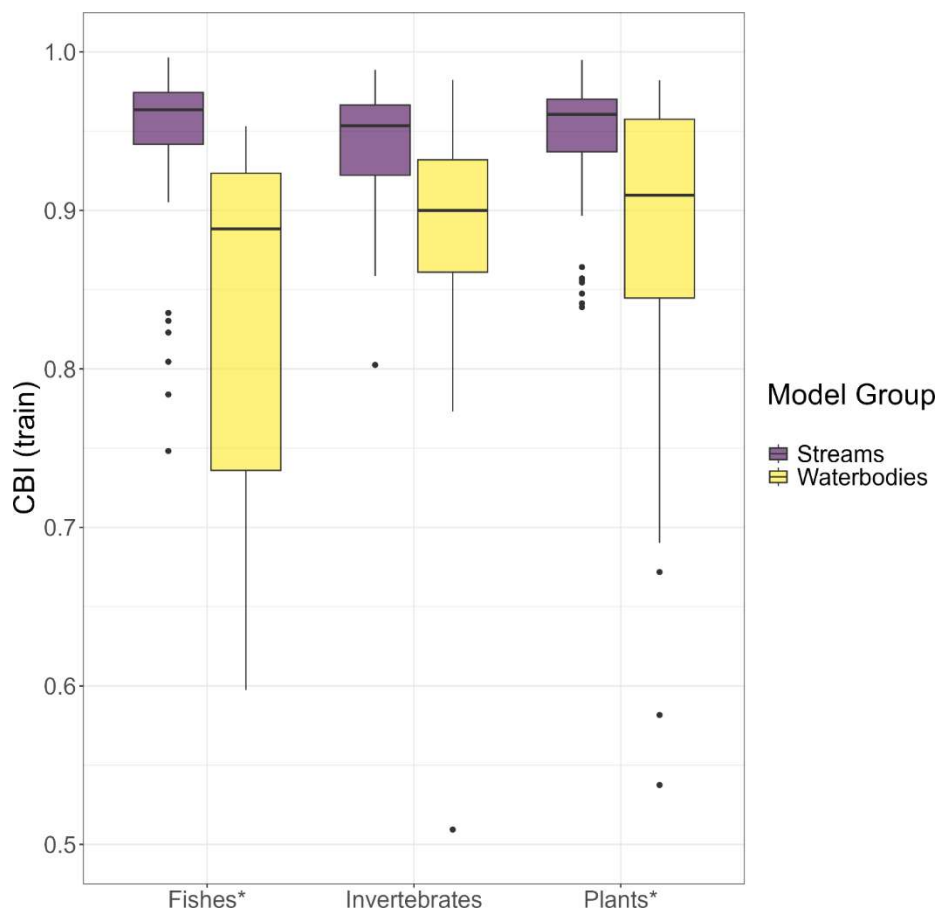


Figure S1. Performance across all models evaluated with the continuous Boyce index (CBI; Hirzel et al., 2006) using model training data against randomly generated comparison points ($n = 10,000$). Results are split by model habitat type (waterbodies or streams) and taxa group (fishes, invertebrates, or plants). Taxa names with * have significant differences ($p < 0.001$) between the distributions of waterbody (lentic) and stream (lotic) CBI values.

References

- Bradley, B. A., Evans, A. E., Jarnevich, C. S., Beaury, E. M., Engelstad, P. S., Teich, N. B., & LaRoe, J. M. (2024). *US non-native plant occurrence and abundance data and distribution maps for Eastern US species with current and future climate* [Data Release]. U.S. Geological Survey. <https://doi.org/10.5066/P14VVRES>
- GBIF. (2024a). *Occurrence Download* [Dataset]. The Global Biodiversity Information Facility. <https://doi.org/10.15468/DL.U4HCXQ>
- GBIF. (2024b). *Occurrence Download* [Dataset]. The Global Biodiversity Information Facility. <https://doi.org/10.15468/DL.VEDXDM>
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>
- Jarnevich, C. S., Engelstad, P., Williams, D., Shadwell, K., Reimer, C., Henderson, G., Prevey, J. S., & Pearse, I. S. (2024). Predicted occurrence and abundance habitat suitability of invasive plants in the contiguous United States: Updates for the INHABIT web tool. *NeoBiota*, 96, 261–278. <https://doi.org/10.3897/neobiota.96.134842>
- LANDFIRE. (2016). *LANDFIRE existing vegetation type* [Online]. U.S. Department of Interior, Geological Survey and U.S. Department of Agriculture. Available: <https://landfire.gov/reference/lfrdb>
- Petri, L., Beaury, E. M., Corbin, J., Peach, K., Sofaer, H., Pearse, I. S., Early, R., Barnett, D. T., Ibáñez, I., Peet, R. K., Schafale, M., Wentworth, T. R., Vanderhorst, J. P., Zaya, D. N., Spyreas, G., & Bradley, B. A. (2023). SPCIS: Standardized Plant Community with Introduced Status database. *Ecology*, 104(3), e3947. <https://doi.org/10.1002/ecy.3947>
- United States Geological Survey. (2025). *USGS National Hydrography Dataset Plus High Resolution National Release 2* [FileGDB]. U.S. Geological Survey. <https://doi.org/10.5066/P13V7GVY>
- U.S. Geological Survey. (2026). *Nonindigenous Aquatic Species Database*. <https://nas.er.usgs.gov/>

APPENDIX 2 - SUPPLEMENTAL INFORMATION FOR CHAPTER 2

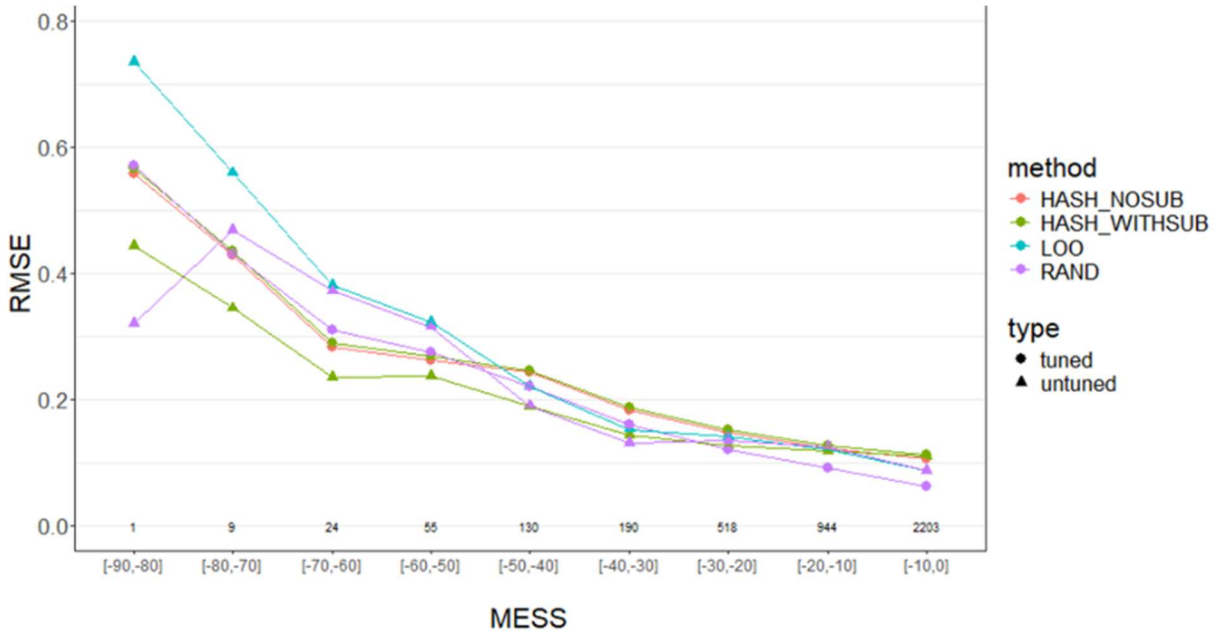


Figure S1. Model performance (root mean square error; RMSE) when comparing predictions to true probabilities of presence for the virtual species along a gradient of environmental similarity between model training data and external test data as measured by multivariate environmental similarity surface values (MESS; Elith et al., 2010). RMSE calculated here in bucketed increments of 10 MESS values (ex. < 0 and ≥ -10). The smaller number above the bucket range shows the mean count of values within the range across methods. The more negative the value bucket range, the more dissimilar the environmental conditions are from those found in the training data. (HASH-NOSUB: hashtag without large spatial blocks; HASH-WITHSUB: hashtag with smaller, subsampled blocks; LOO: buffered leave-one-out; RAND: random).

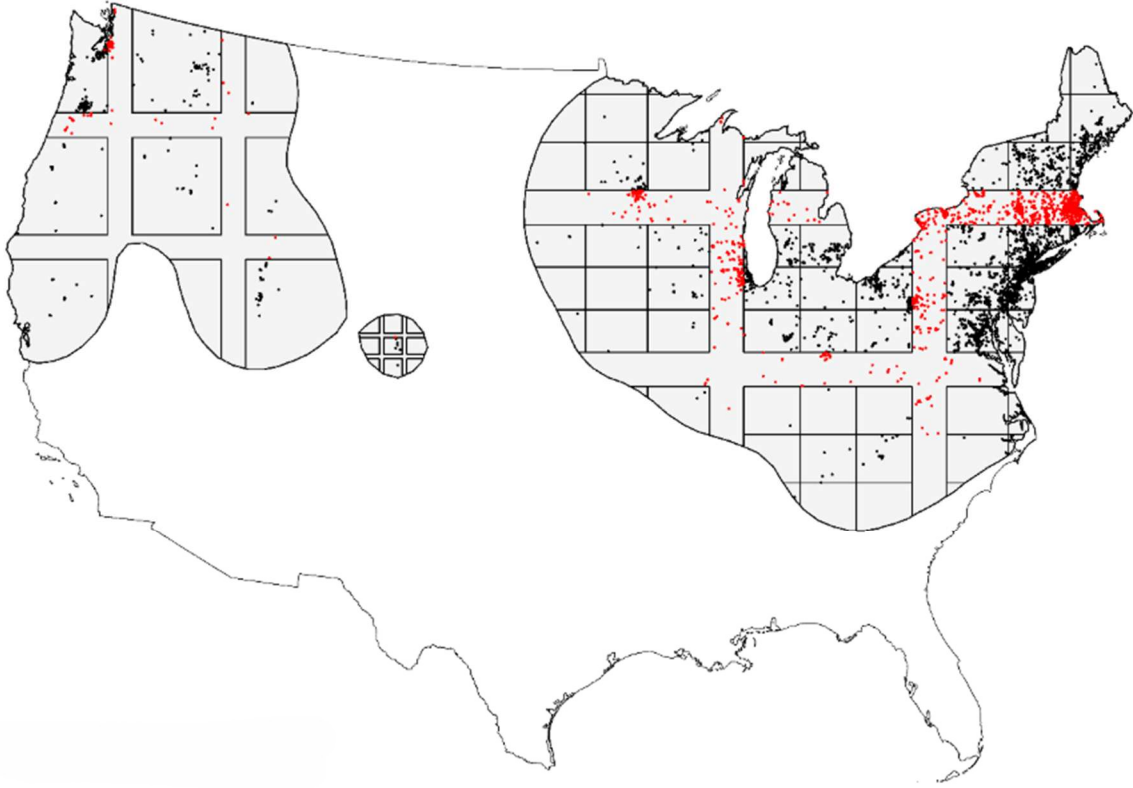


Figure S2. An example of the hashtag splitting method performed across multiple disjunct distributions for a single species in the conterminous United States. Black points represent training data and red points represent withheld test data.

Table S1. Optimized hyperparameter settings for tuned lightGBM models for random (RAND) and hashtag with subsampling (HASH-WITHSUB) and without subsampling (HASH-NOSUB).

Method	stop_iter	learn_rate	tree_depth	loss_reduction	num_leaves
HASH-NOSUB	12	0.233	1	3.48	64
HASH-WITHSUB	12	0.233	1	3.48	64
RAND	12	0.233	1	3.48	64