Species Distribution Modeling for Aquatic Invasive Species

Joshua Naudet [1], Brian Hand [1,2], Leif Howard [2], Gordon Luikart [1,2]

Montana State University/Flathead Lake Biological Station

Abstract

Invasive alien species (IAS) are among the greatest threats to biodiversity, and species distribution modeling is a key resource for proactively managing their impacts and spread. Modeling workflows that distill habitat suitability into useful products for conservation practice are lacking and under-explored. This study demonstrates how the USGS Non-indigenous Aquatic Species (NAS) database can be combined with remotely-sensed environmental data to rapidly and flexibly generate maps of relative IAS spread risk for stakeholders at relevant spatial scales. Ecological niche modeling workflows (species distribution models) for 5 high-priority AIS in North America were created. The sensitivity and performance of this workflow were examined on three different analytical scales. Crucial environmental variables were also identified across the taxa. This study highlights the utility of mobilizing presence data from NAS and NASA remotely-sensed environmental data in invasive species modeling and proactive management and to help managers prioritize efficient use of limited conservation resources.

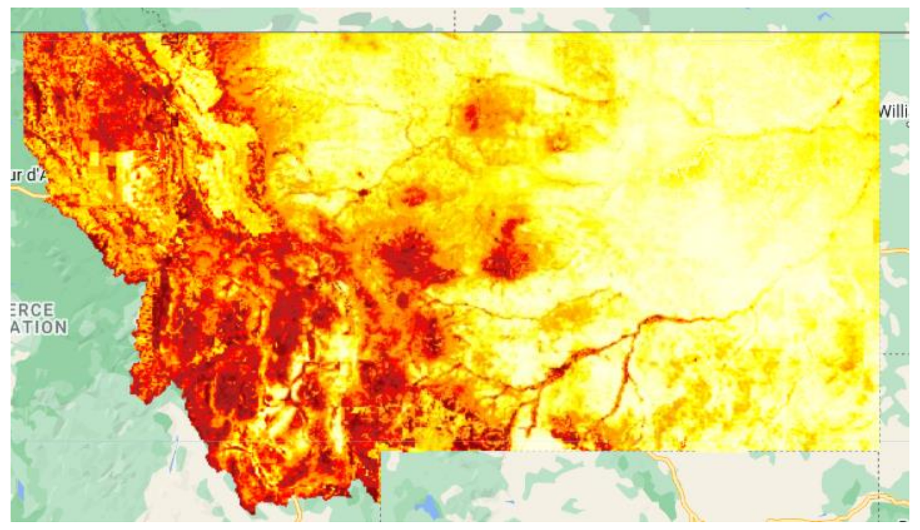

Figure 2: Habitat suitability model for Eastern Brook Trout in Montana.

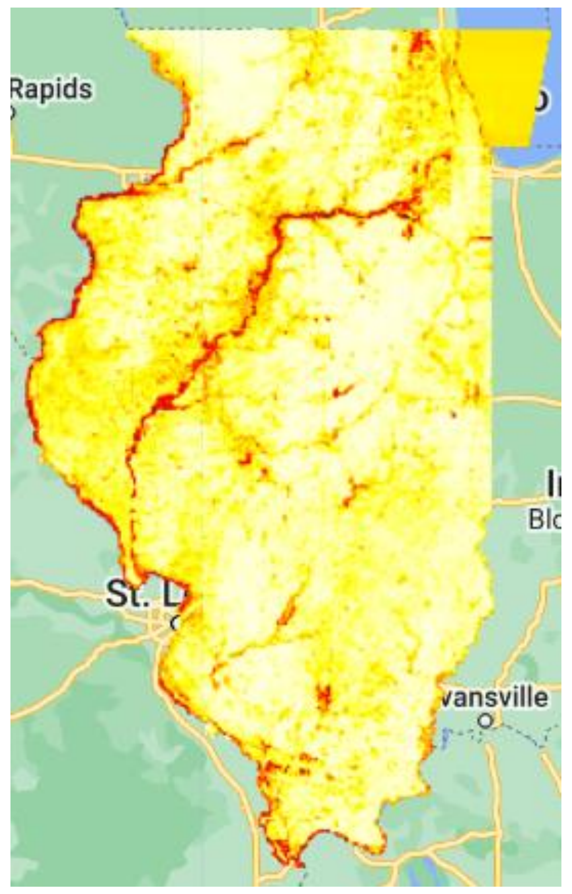

Figure 5: Habitat suitability model for Big Head Carp in Illinois.

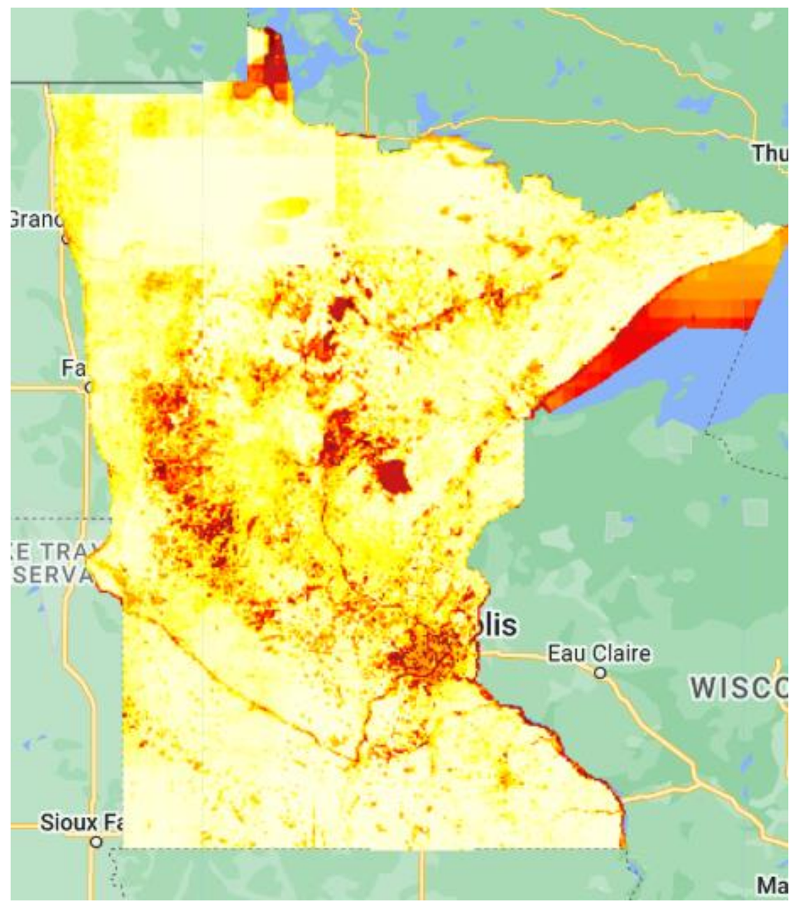

Figure 3: Habitat suitability model for Zebra Mussels in Minnesota.

Habitat Suitability Modeling

MaxEnt was the machine learning tool used for the presence-only habitat suitability model. MaxEnt provides a Google Earth Engine (GEE) classifier which employs the principle of maximum entropy to produce a log-linear model for probability of presence for equilibria organisms (Phillips et al. 2006). MaxEnt was utilized to estimate habitat suitability using occurrence data for all focal taxa. Background points were selected following Phillips et al. (2006). At each aggregation scale, background areas were identified as those with no known occurrences and selected an equivalent or greater number of background points than available occurrence points at that analytical scale. The environmental predictors were utilized to assure the most acceptable efficiency for the model.

For watershed-based analyses, three levels of aggregated Hydrologic Unit Codes (HUC) were observed: regional (multi-state), state, and local (multiple counties within a state). These correspond with the typical management boundaries used for decision-making in containment and monitoring programs.

	500m	2500m	5000m
Eastern Brook Trout (EBT) in Montana	0.13142	0.08503	0.03508
Rainbow Trout (RBT) in Montana	0.076305	0.10801	0.10167
Zebra Mussels (ZM) in Minnesota	0.12216	0.14625	0.16169
Bighead Carp (BHC) in Illinois	0.14686	0.13251	0.15606
Eurasian Milfoil (EMF) in Ohio	0.34448	0.36823	0.44643

Table 1: False negative rates for all 5 taxa at the three studied pixel scales.

The results show that the smaller pixel scales had the lowest false negative error rates in most cases. The EMF results were the only species that were not considered acceptable for running the model at a 70% efficiency, this could be due to the lack of overall data points, as with the other species, there were over 500 points to train the model, while the EMF data had over 100 less points (Table 1).

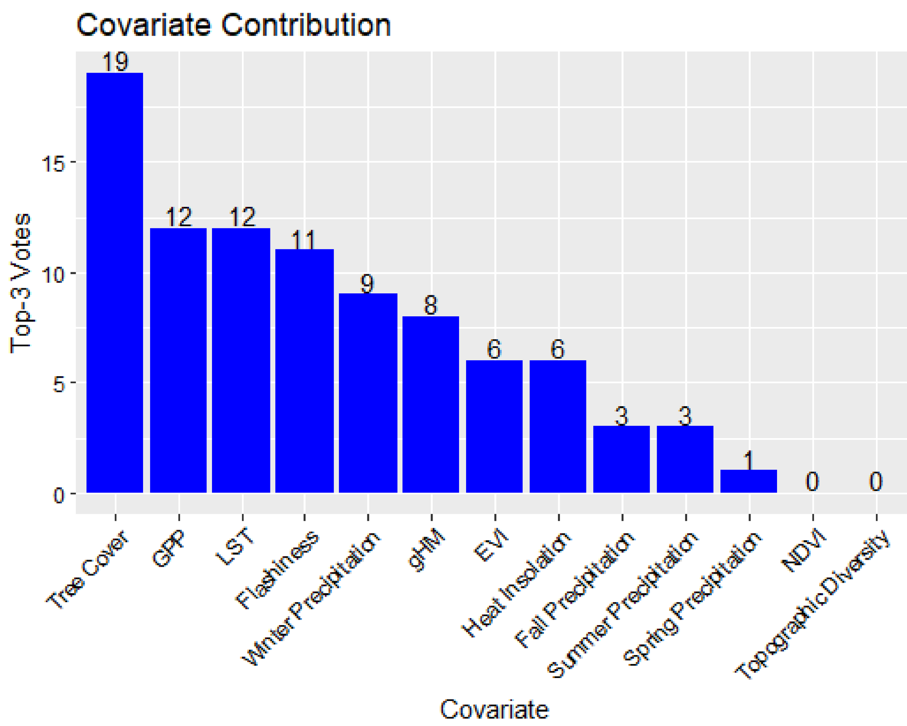

Figure 4: Magnitude of significance for each covariate variable for all studied taxa.

Spatial Environmental Covariates

Environmental data was gathered for use as predictor variables in all occurrence modeling. Environmental covariates were derived from publicly available NASA remote sensing data downloaded through GEE. The suite of predictors was examined in the model because of their known biological relevance across many AIS.

Aggregating covariate data to watershed level can reduce scale mismatch issues between different remotely sensed (Carter et al 2021). For the pixel-based aggregation, data was sampled at a magnitude of 500m, 2,500m, and 5,000m (Figure 1). All the covariate data was averaged over the years 2002-2018.

Significant Covariate Parameters

- Percent Tree Coverage
- Mean General Physical Precipitation (GPP)
- Max Annual Land Surface Temperature (LST)
- Flashiness
- Total Winter Precipitation
- gHM
- Mean Enhanced Vegetation Index (EVI)
- Heat Insolation Load
- Total Fall Precipitation
- Total Summer Precipitation
- Total Spring Precipitation
- Mean Normalized Difference Vegetation Index (NDVI)
- Topographic Diversity

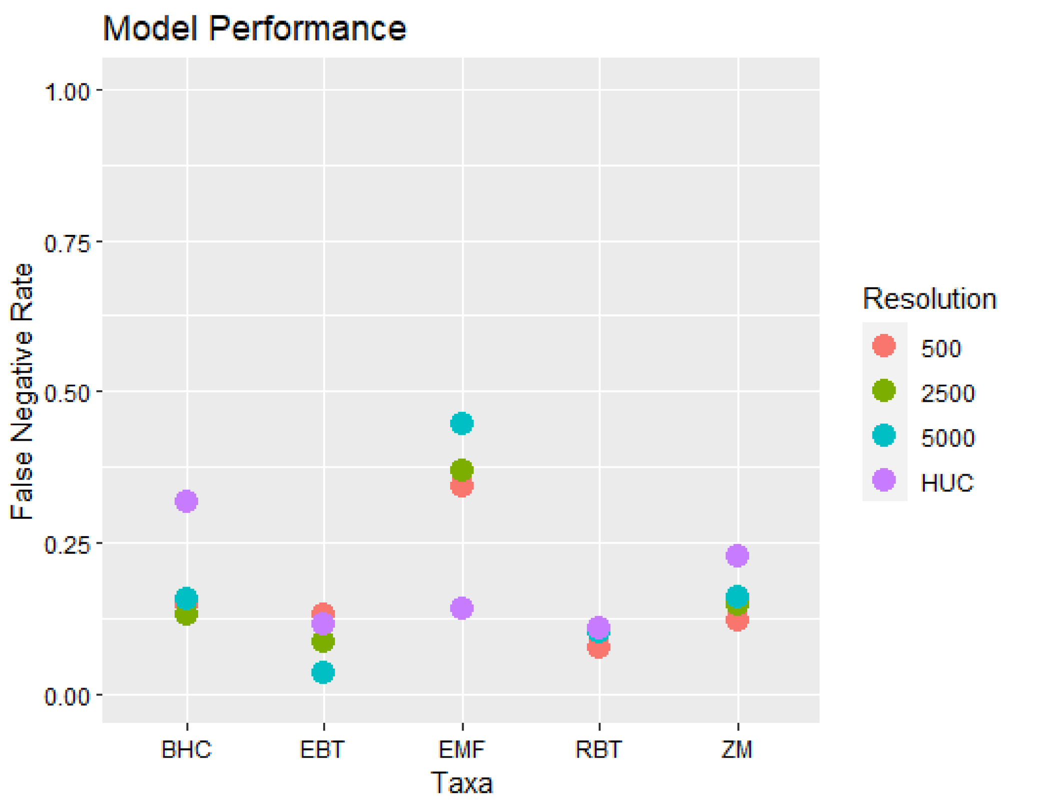

Figure 1: False negative rates for all 5 observed taxa at the three different pixel scales.

References/Acknowledgements

West AM, Kumar S, Brown CS, Stohlgren TJ, Bromberg J (2016) Field validation of an invasive species Maxent model. Ecological Informatics 36: 126–134.

Phillips SJ, Dud\’\ik M, Schapire RE (2004) A Maximum Entropy Approach to Species Distribution Modeling. In: Proceedings of the Twenty-First International Conference on Machine Learning. ICML ’04. Association for Computing Machinery, New York, NY, USA, 83.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17: 43–57.

Carter S, van Rees C, Hand B, Luikart G (2021) Testing a Generalizable Machine Learning Workflow for Aquatic Invasive Species on Rainbow Trout (Oncorhynchus mykiss) in Northwest Montana. Frontiers in Big Data 4: 734990.